

П. Эйкхофф

Основы
идентификации
систем
управления

Издательство
·Мир·

SYSTEM IDENTIFICATION

PARAMETER AND STATE ESTIMATION

Pieter Eykhoff

University of Technology
Eindhoven, the Netherlands

П. ЭЙХОФФ

ОСНОВЫ ИДЕНТИФИКАЦИИ СИСТЕМ УПРАВЛЕНИЯ

Оценивание параметров и состояния

ПЕРЕВОД С АНГЛИЙСКОГО
В. А. Лотоцкого и А. С. Манделя

ПОД РЕДАКЦИЕЙ
Н. С. Райбмана

ИЗДАТЕЛЬСТВО «МИР» • МОСКВА 1975

Книга П. Эйкхоффа, известного специалиста в области теории управления, посвящена одному из основных разделов теории управления — идентификации, получению математического описания объекта по результатам наблюдений. В монографии содержатся необходимые сведения из теории вероятностей, математической статистики, теории сигналов и динамических систем. Систематически излагаются основные методы идентификации, оценивания параметров и состояния систем управления. Теоретические положения иллюстрируются многочисленными примерами. Отдельная глава посвящена практическим применениям.

Книга представляет интерес для специалистов по теории управления, научных работников и инженеров, занимающихся вопросами метрологии, научных экспериментов, моделирования. Она может служить учебным пособием для аспирантов и студентов старших курсов при изучении основ теории управления.

Редакция литературы по новой технике

© 1974 John Wiley and Sons Ltd.

© Перевод на русский язык, «Мир», 1975

30502-410
Э 041(01)-75 149—74

ПРЕДИСЛОВИЕ К РУССКОМУ ИЗДАНИЮ

С давних пор человечество затрачивает огромные усилия на установление закономерностей происходящих в природе явлений. Первичным в процессе познания всегда являются результаты наблюдений. Они представляют собой отправной пункт к модели, к абстрактному мышлению, а уже от модели осуществляется переход к практической деятельности. Очевидно, что эта схема познания применима независимо от того, идет ли речь о естественном или искусственном объекте. Создание абстрактной модели обычно связано со «сжатием» информации, содержащейся в результатах наблюдений. Это объясняется тем, что каждый отдельный результат наблюдений является случайным, поэтому построение адекватной, изоморфной модели реального объекта может быть осуществлено только на основе многократных наблюдений. Случайность каждого результата наблюдений объясняется, с одной стороны, принципиальной невозможностью учесть все многообразие факторов, действующих на данный конкретный объект, каким бы простым он ни казался на первый взгляд, и сложными взаимосвязями этих факторов, а с другой стороны, несовершенством естественных или искусственных средств наблюдения.

Построение модели по результатам наблюдений представляет собой формализацию, необходимую для определения основных признаков, связей, закономерностей, присущих объекту-оригиналу, и отсеивания второстепенных признаков. Для одного и того же объекта в зависимости от конкретных требований практики и типа решаемой задачи может быть построен ряд моделей, осуществлена формализация различных функций этого объекта или внешних воздействий на этот объект. В этом в известной мере проявляется принцип декомпозиции, применение которого и дает возможность построить относительно простые модели. Так, например, при разработке какого-либо технологического процесса обычно учитываются потребности общества в данном продукте. Эта потребность выявляется по общей модели потребления, которая строится на основании изучения спроса, ресурсов и многих других факторов, а также с учетом достигнутого уровня технологии. Очевидно, что для каждого из рассматриваемых вопросов строится своя модель. Для самого технологического процесса в свою очередь строятся модели функционирования, отражающие технические, экономические, психологи-

ческие, социальные и другие аспекты процесса, обеспечивающие получение необходимого продукта. Если рассмотреть технические вопросы, то здесь также не представляется возможным ограничиться какой-либо одной моделью, так как требуется решать множество практических задач. Например, здесь могут быть рассмотрены модели превращения сырья в заданный продукт, модели качества, производительности, прочности, коррозионной стойкости и многие другие. Разработка нового технологического процесса была бы невозможной без наличия этих моделей, которые строятся на основе наблюдений и обобщения опыта. Очевидно, что в процессе накопления опыта созданные модели уточняются, изменяются, пересматриваются и это также осуществляется на основании новых наблюдений.

Следует отметить, что процесс формализации наблюдений осуществляется в связи с необходимостью передачи опыта, информации от одного поколения к другому, и абстрактная модель — наиболее удобный и компактный вид представления явления для возможности такой передачи. Однако большое разнообразие явлений привело к появлению многочисленных подходов и языков для осуществления процесса формализации и построения моделей. На первом этапе формализации осуществлялось в основном содержательное описание явления и обычно использовался разговорный язык, который оказался довольно удобным и для изложения результатов и их передачи. Однако, несмотря на универсальность этого способа описания, он оказался недостаточным для представления многих качественных и количественных сторон явлений. В дальнейшем процесс формализации сопровождался появлением различных языковых форм, оказавшихся более экономными, понятными и дополнявшими содержательное описание результатов наблюдений. Так появились многие языки схем, уравнений, карт, чертежей и т. д., которые и использовались для построения моделей и их формального описания. В настоящее время в различных областях человеческих знаний существует множество способов наблюдения, обобщения, формализации и представления результатов наблюдений.

Во второй половине XX в. значительно возросла роль науки об управлении в связи с установленными аналогиями в процессах управления целенаправленной деятельностью человеческого общества, живыми организмами и искусственными, созданными человеком машинами и механизмами. Появлению новых идей и методов управления способствовали, с одной стороны, обобщения высокоеффективных принципов теории автоматического управления, а с другой стороны, возросшие технические возможности в связи с широким развитием вычислительной техники. В области управления возникла необходимость в новых принципах построения моделей, формализации результатов наблюдений. В теории автоматического управления принципы построения системы управления разрабатывались на основе заданной модели. В дальнейшем оказалось, что во многих случаях модель, принятая при проектировании, существенно отличается от реального объекта, что значительно уменьшало или сводило на нет эффективность разработанной системы

управления. В связи с этим возникло одно из новых и важных направлений в теории управления, связанное с построением модели на основании наблюдений, полученных в условиях функционирования объекта по его входным и выходным переменным. Это направление известно в настоящее время как идентификация систем. Несмотря на то что термин «идентификация» появился около пяти лет назад, теории и методам идентификации посвящено большое число работ в отечественной и зарубежной литературе и в этом направлении разработаны свои принципы, подходы и методы. Эти подходы нашли широкое применение в различных областях науки и техники, и в том числе в биологии, медицине, аeronавтике, экономике.

Задача идентификации формулируется следующим образом: по результатам наблюдений над входными и выходными переменными системы должна быть построена оптимальная в некотором смысле модель, т. е. формализованное представление этой системы. Отсюда видна преемственность между задачей идентификации и указанной общей схемой установления закономерностей по результатам наблюдений. Задача идентификации базируется на современной теории управления. Для ее решения используются современные вычислительные машины. Последние, обладая большим быстродействием и практически неограниченным объемом памяти, создают предпосылки для получения, передачи и обработки больших массивов наблюдений, которые необходимы для построения адекватных моделей реальных объектов.

В зависимости от априорной информации об объекте управления различают задачи идентификации в узком и широком смысле. Задача идентификации в узком смысле состоит в оценивании параметров и состояния системы по результатам наблюдений над входными и выходными переменными, полученными в условиях функционирования объекта. При этом известна структура системы и задан класс моделей, к которому данный объект относится. Априорная информация об объекте достаточно велика.

Априорная информация об объекте при идентификации в широком смысле отсутствует или очень бедная, поэтому приходится предварительно решать большое число дополнительных задач. К этим задачам относятся: выбор структуры системы и задание класса моделей, оценивание степени стационарности и линейности объекта и действующих переменных, оценивание степени и формы влияния входных переменных на выходные, выбор информативных переменных и др. К настоящему времени накоплен большой опыт решения задач идентификации в узком смысле. Методы же решения задач идентификации в широком смысле начали разрабатываться только в последние годы, и здесь результаты значительно скромнее, что в первую очередь можно объяснить чрезвычайной трудностью задачи.

Данная книга посвящена основам теории идентификации в узком смысле. Ее автор — профессор П. Эйкхофф (электротехнический факультет технологического университета в г. Эйндховене, Голландия), один из создателей теории идентификации. Он является известным специалистом в области теории управления и принимает

активное участие в работе международной организации по автоматическому управлению (ИФАК). Эйхофф — почетный редактор ИФАК, один из организаторов и член программного комитета трех симпозиумов ИФАК по идентификации и оцениванию параметров и состояния (Прага, 1967 и 1970 гг.; Гаага, 1973 г.); он неоднократно выступал с обзорными докладами по идентификации на конгрессах (Лондон, 1966 г.) и симпозиумах ИФАК (1967 и 1970 гг.). Работу над книгой проф. П. Эйхофф начал еще в 1970 г., однако болезнь задержала завершение работы. По мере готовности отдельных глав он направлял их для ознакомления ряду специалистов, работающих в области идентификации в разных странах. Редактор перевода был среди тех, кто имел удовольствие ознакомиться с отдельными главами еще в 1971 г. и со всей книгой в конце 1972 г. по рукописи. Предлагаемый русский перевод выполнен по рукописи, и отрадно то, что русское и английское издания книги выходят почти одновременно.

В кратком предисловии трудно дать более или менее полный анализ достоинств предлагаемой книги. Как уже отмечалось, к настоящему времени по проблеме идентификации опубликовано много работ, однако не было книги, в которой бы систематически излагались различные методы и подходы, а также приложения в этой области. Как образно было отмечено на 2-м симпозиуме по идентификации, состояние в этой области таково, что каждого из авторов новой работы можно сравнить с фокусником, который из волшебного ящика достает новый алгоритм решения задачи. Систематическое изложение с единой точки зрения многочисленных методов и алгоритмов является большой удачей автора. В книге рассмотрен широкий круг вопросов. Читатель сможет познакомиться с основными понятиями и определениями модели, постановками задач оценивания параметров и состояния, методами решения задач для различных классов объектов, особенностями идентификации по данным нормального функционирования объектов, со статистическим и инженерным подходами, видами сигналов и методами их описания, с понятиями управляемости, идентифицируемости и наблюдаемости и связи между ними. Эти вопросы составляют первую часть книги. Краткие сведения по теории идентификации, основные детерминированные и вероятностные алгоритмы стохастической аппроксимации, регрессионные и корреляционные методы исследования непрерывных и импульсных систем для моделей, линейных и нелинейных по параметрам, функции чувствительности параметров, применяемые тестовые сигналы и задачи построения модели по различным критериям подробно рассмотрены в гл. 5—11. Оценки параметров и состояния, фильтры Винера и Калмана — Бьюси, выбор функции ошибок и ее минимизация обсуждаются в гл. 12 и 13.

В книге рассматриваются перспективы и пути развития идентификации. На одном из перспективных направлений следует остановиться особо. Речь идет о широком классе реальных объектов, для которых процессы идентификации и управления объединены в единой системе управления. По-видимому, этот подход следует рассматривать как практическую реализацию идеи дуального

управления Фельдбаума. При построении системы управления в реальном масштабе времени для объекта, характеристики которого меняются во времени, а априорная информация отсутствует или очень мала, невозможно ограничиться одноразовой моделью и в процессе функционирования объекта модель уточняется. Эта функция возлагается на идентификатор, который является составной частью системы управления. Особенно эффективен такой подход в системах управления технологическими процессами и летательными аппаратами. В настоящее время для систем этого типа применяется ряд алгоритмов, описанных в книге, которые представляют собой модификации алгоритмов стохастической аппроксимации.

Во всех частях книги автор приводит многочисленные примеры, хорошо иллюстрирующие основные теоретические положения. Описываемые алгоритмы и подходы могут быть использованы при построении аналогичных моделей. Не ограничиваясь только этими примерами, в последней главе автор приводит обзор практических применений. Из самого перечня областей применения — автоматическое управление, физические, механические и химические объекты, ядерные реакторы, энергетика, связь, авиация и космонавтика, биологические объекты, социально-экономические системы — следует вывод об универсальности методов и алгоритмов идентификации и возможности их применения во многих областях науки и техники.

В конце каждой главы приводится довольно полный список цитированной и дополнительной литературы, в том числе и советской, к которому автор отсылает читателя для более полного ознакомления с рассматриваемыми вопросами. Несомненно, что такой способ изложения дал возможность автору сосредоточить внимание на основных принципах, методах и результатах.

Систематическое изложение многочисленных методов идентификации, множество приведенных алгоритмов, а также примеров делают эту книгу интересной для широкого круга специалистов, работающих в различных областях науки и техники. Специалисты в области управления найдут в ней много полезных сведений, необходимых при проектировании систем управления, специалисты в области идентификации — подходы, методы и алгоритмы для решения задач построения модели объекта управления, экспериментаторы — методы планирования и обработки результатов научного эксперимента, технологи — методы описания технологических процессов. Книга может служить хорошим пособием по курсам идентификации и автоматического управления, поэтому она будет полезна преподавателям, аспирантам и студентам высших учебных заведений. Кроме того, инженеры, врачи, экономисты, биологи, а также программисты вычислительных центров могут использовать приведенные в книге алгоритмы для создания программ идентификации.

Н. Райбман

ПРЕДИСЛОВИЕ АВТОРА К РУССКОМУ ИЗДАНИЮ

Мне было очень приятно написать несколько вступительных слов к русскому переводу моей книги, опубликованной на английском языке под названием *System Identification; Parameter and State Estimation*.

Развитие науки и техники имеет, несомненно, международное значение. В полной мере это относится и к вопросам, охваченным в данной книге. Поскольку на обложке стоит мое имя, я ответствен за все возможные ошибки и неточности в тексте; в то же время эта книга является результатом международного сотрудничества и обмена. Она не могла бы появиться на свет без вклада многочисленных, искренне уважаемых мною коллег, без их статей и отчетов, докладов на симпозиумах и конгрессах. Многое я узнал при личных контактах. В связи с этим хотелось бы подчеркнуть ценность большой работы, проводимой в мировом масштабе Международной федерацией по автоматическому управлению (ИФАК) с помощью национальных организаций стран-участниц и ряда сотрудничающих с ней отдельных лиц.

Я глубоко признателен моему коллеге проф. Н. С. Райбману за его интерес к этой книге и усилия по организации ее перевода и издания на русском языке. Мне хочется поблагодарить переводчиков В. Лотоцкого и А. Манделя, издательство «Мир» и всех, кто внес свой вклад в осуществление этого издания.

Идентификация и оценивание могут найти широкие применения в технике, экономике, медицине и других областях. Я искренне надеюсь, что эта книга окажется полезной в разработке практических систем, способствующих росту благосостояния людей.

П. Эйкхофф

ПРЕДИСЛОВИЕ АВТОРА

В последние годы разные аспекты теории оценивания параметров и состояний рассматривались в ряде работ, на многих научных конференциях и в нескольких университетских курсах. Интерес к этому предмету объясняется, очевидно, различными причинами:

желанием инженеров разобраться в технологических процессах с тем, чтобы организовать процесс управления так, чтобы добиться более низкого уровня затрат или повышения эффективности;

исследованием динамики современных воздушных и космических летательных аппаратов и таких систем, как железнодорожный и водный транспорт;

изучением поведения человека при решении задач отслеживания и других задач управления или обучения;

исследованием биологических объектов, например таких нейромоторных систем, как система управления движением глазного яблока, система управления конечностями (рукой, ногой), система управления сокращением сердца и т. д.

Большой интерес к теории оценивания возник в результате необходимости повышения качества функционирования систем (расширения знаний о динамике систем), а также вследствие существенного изменения возможностей применения теории оценивания, связанного с наступлением эры современных вычислительных машин с совершенным математическим обеспечением. При обсуждении конкретных алгоритмов оценивания необходимость в подобной вычислительной технике станет очевидной.

Более или менее независимо от инженеров теорией оценивания заинтересовались эконометрики и статистики, занимавшиеся разработкой локальных и глобальных динамических экономико-математических моделей развития общества. Это привело к взаимообогащению двух разных наук и созданию общей математической теории

систем, применимой и к физическим и к социальным объектам. Непосредственным результатом этого развития стало появление большого числа работ, в которых описываются различные подходы к решению задач и приводятся разные постановки задач теории оценивания.

В этой книге сделана попытка с самых широких позиций последовательно изложить проблему оценивания. Книга рассчитана на ученых и инженеров, занимающихся динамическими системами, например технологическими процессами, самолетами, биологическими или экономическими объектами и т. д.

Из-за специфики предмета в книге пришлось обратиться к математической статистике, динамике систем, теории случайных сигналов, оптимизации, аналоговому, цифровому и гибридному моделированию. С некоторыми кратко изложенными вопросами (из-за ограниченного объема книги) подробнее можно познакомиться, воспользовавшись соответствующей литературой, на которую сделаны многочисленные ссылки.

При размещении материала принимались во внимание:

- 1) способ применения;
- 2) используемые статистические методы: метод наименьших средних квадратов, получение марковских оценок, оценивание по методу максимального правдоподобия, байесовский подход;
- 3) линейность или нелинейность модели по отношению к параметрам (заметим: это не то же самое, что линейность динамики объекта);
- 4) описание сигнала: импульсное и непрерывное;
- 5) реализация алгоритма: аналоговая, цифровая, гибридная;
- 6) тип алгоритма: оценивание по разомкнутому и замкнутому контуру регулирования.

Первые две главы посвящены наиболее важным подходам к решению задач оценивания.

В гл. 3—5 приводятся необходимые математические результаты. Здесь читатель найдет сжатое изложение теории сигналов, теории динамических систем, теории вероятностей, математической статистики и теории оптимизации.

Логическая схема части книги, посвященной оцениванию параметров (гл. 6—11), приведена в таблице на стр. 13.

В гл. 12 рассматривается задача оценивания состояний. Используется дедуктивный подход, так как предполагается, что читатель достаточно знаком с предметом или сможет без труда разобраться, опираясь на аналогию с задачей оценивания параметров. Глава 13

Таблица

Тип оценок	Тип сигналов	Тип алгоритмов	Тип моделей	Глава
			Разомкнутый контур или явные методы	Линейные по параметрам → 6
	Выборочные сигналы	Замкнутый контур или неявные методы		Линейные → 7
			Нелинейные по параметрам	
		Разомкнутый контур или явные методы	Линейные	→ 8
	Непрерывные сигналы			
		Замкнутый контур или неявные методы	Линейные	→ 9
			Нелинейные по параметрам	
Оценки по методу наименьших квадратов и марковские оценки				
Оценивание параметров		Сигналы периодические и большой мощности		→ 10
Оценки максимального правдоподобия				
Байесовские оценки				→ 11

посвящена описанию методов совместного оценивания параметров и состояний.

В последней главе представлены описания конкретных приложений методов оценивания при управлении в системах автоматического контроля, в биологических системах, в задачах обучения и распознавания образов.

В тех случаях, когда преобразование в частотной области удобнее, используются методы спектрального анализа.

При изложении материала автор стремился переходить к сложным вопросам от более простых, от частного случая к более общему. Несмотря на то что во многих научных работах заметна противоположная тенденция, по-видимому, для первого знакомства с предметом больше подходит индуктивный метод изложения, даже если студент или исследователь отдает предпочтение изяществу и другим достоинствам дедуктивного метода. На расположение материала в книге заметно повлияли методические соображения. Автор надеется, что благодаря этому читатель более глубоко разберется в обсуждаемых проблемах и получит ясное представление о перспективах возможного применения.

Книга рассчитана на инженеров и исследователей, аспирантов и преподавателей. Для понимания излагаемого материала достаточно:

- 1) знакомства с математикой в объеме вуза, включая математический анализ, дифференциальные уравнения, теорию функций комплексного переменного, элементы линейной алгебры и теорию матриц;
- 2) некоторого знакомства с теорией систем управления с обратной связью;
- 3) знания теории вероятностей, элементарной математической статистики и теории случайных процессов;
- 4) некоторых представлений о цифровой и аналоговой технике.

Для облегчения работы над книгой часть необходимых сведений приведена в приложениях. При проработке лекций весьма желательно, чтобы студенты закрепляли знания выполнением простых заданий. При этом необходимо прибегать к использованию вычислительных машин, так как в области идентификации решению вручную поддаются только чрезвычайно простые задачи. Однако подбор задач зависит от имеющейся вычислительной техники, и именно поэтому автор и не пытался включить в книгу задачи: любому лектору не составит труда сформулировать такие задачи и дать указания для их решения.

При работе над рукописью трудно было решить, следует ли включать в книгу машинные программы процедур оценивания. Пришлось взвешивать много факторов: выбор машинного языка (Алгол, Фортран и т. д.) и «диалекта», возможности различных типов компьютеров, периферийное оборудование и объем машинной памяти, скорость сходимости новых процедур оценивания, огра-

ниченный объем книги. В результате было принято решение не помещать в книгу такие программы, а возможному потребителю настоятельно рекомендуется обращаться непосредственно к авторам тех машинных программ, которые могут оказаться полезными при решении конкретной практической задачи. Приведено много таких ссылок.

Разнообразную информацию по идентификации и оцениванию можно найти в препринтах и трудах симпозиумов:

1-й симпозиум ИФАК «Идентификация в системах автоматического управления», Прага, 1967. Препринты в двух томах опубликованы издательством «Academia», Prague.

2-й симпозиум ИФАК «Идентификация и оценивание параметров», Прага, 1970. Препринты в двух томах опубликованы издательством «Academia», Prague.

3-й симпозиум ИФАК «Идентификация и оценивание параметров», Гаага, 1973. Препринты и труды опубликованы издательством «North-Holland Publishing Comp.», Amsterdam.

Имеется много ссылок на конгрессы ИФАК:

Первый конгресс ИФАК, Москва, 1960. Труды конгресса под названием «Автоматика и телемеханика» в четырех томах опубликованы издательством «Butterworths», London, 1961.

Второй конгресс ИФАК, Базель, 1963. Труды конгресса под названием «Автоматика и телемеханика» в двух томах опубликованы издательством «Butterworths», London, Oldenbourg, Munich, 1964.

Третий конгресс ИФАК, Лондон, 1966. Труды конгресса под названием «Автоматика и телемеханика III» в четырех томах опубликованы в «Inst. of Mechanic Engineers», London.

Четвертый конгресс ИФАК, Варшава, 1969. Препринты в 50 томах опубликованы в «Naczelna Organizacja Techniczna», Warsaw, Poland.

Пятый конгресс ИФАК, Париж, 1972. Препринты изданы «Instrument Society of America», Pittsburgh, Penna., USA.

Кроме периодической литературы полезными источниками информации могут служить препринты Объединенной конференции по автоматическому управлению (JACC), издаваемые американским советом по автоматическому управлению, и труды многих других конференций.

ГЛАВА 1

ПОНЯТИЯ И ИДЕИ. ПОСТРОЕНИЕ МОДЕЛИ, ОЦЕНИВАНИЕ ПАРАМЕТРОВ И СОСТОЯНИЙ

1.1. МОДЕЛИ

Наблюдение и измерение являются одними из основных понятий, используемых в естественных науках и технике. Опираясь на наблюдение, ученый строит физическую модель исследуемого явления, на основе которой методом проб и ошибок создается теория. Эта теория содержит в себе предполагаемое представление об исследуемом явлении природы. Руководствуясь таким представлением, ученый планирует новые эксперименты. Результаты экспериментов могут либо подтвердить теорию, либо потребовать ее частичной или полной переработки. Хотя сами идеи могут быть интересны и нравиться ученым, тем не менее фактические результаты сравнения теории с экспериментом являются решающими. С этой точки зрения можно утверждать, что в естественных науках и технике понятия эксперимента и наблюдения (измерения) являются фундаментальными.

Почти равнозначна им идея *построения модели* [20], которую довольно трудно отделить от упомянутых выше наблюдений и экспериментов. Разработка теории (как предполагаемого представления об исследуемом явлении природы) может быть названа построением модели; тогда теория становится как бы словесной или математической моделью действительности. Мы определим *модель как изображение существенных сторон реальной системы (или конструируемой системы), в удобной форме отражающее информацию о системе*.

В этом определении под системой можно понимать, например, совокупность упорядоченных объектов, в которой каким-то образом определены задачи или цели. Все, что не входит в систему, является частью среды. Система

характеризуется связями между входными и выходными сигналами, вид которых зависит от выбора границы между системой и средой. Что представляет собой система, зависит от точки зрения наблюдателя. «Системой» может быть и электронный усилитель, и контур регулирования, в котором этот усилитель может быть одним из многих элементов, и химический процесс с несколькими цепями обратной связи, и промышленное предприятие и несколько предприятий, которые функционируют как система в среде, определяемой всей экономикой страны.

Емкость понятия модели можно пояснить несколькими замечаниями:

1) Модель не должна быть описанием фактического устройства системы. Она может имитировать систему или «подражать» ее поведению. Можно сконструировать управляемый протез руки, не зная, как человек распоряжается своими конечностями. Искусство разработки состоит в том, чтобы скрыть необъясенные особенности поведения системы признаком модели таких почти алхимических качеств, как «расщепляемость», «сопрягаемость» и т. д.

2) Модели бывают *концептуальные, физические или математические* (другие названия: феноменологические, эмпирические и аналитические) в зависимости от того, какая сторона явления в данном случае наиболее существенна, от методов, которые можно использовать при построении модели, от количества и качества имеющейся информации. Хорошо понятные и известные примеры различных моделей можно найти в небесной механике. Птолемеевская модель Вселенной представляла собой достаточно адекватную концептуальную модель для наблюдений за движением планет. Модель Коперника, в которой наблюдаемые процессы объясняются как результат сложения движения Земли (наблюдателя) и наблюдаемой планеты, можно считать физической моделью. Законы Кеплера образуют математическую модель, которая делает возможным количественное предсказание. Примеры различных моделей читатель может привести сам на основе своего жизненного опыта.

Модели можно также различать в зависимости от используемого оборудования (приборы, устройства) и мате-

матического обеспечения (математические выражения, машинные программы).

Объект, для которого нужно построить модель, может проектироваться или реально существовать. Нет необходимости подчеркивать, насколько важно для инженера иметь модель системы на стадии ее проектирования. Для исследователя модель означает возможность дальнейшего планирования экспериментов.

3) Информация должна быть представлена в *удобной форме*. Это существенно, так как модель должна создать предпосылки для следующих решений. Если модель слишком сложна, ее полезность становится сомнительной. Относительная простота является главной характеристикой модели. Модель представляет собой упрощенное отображение действительности. Во многих случаях, для того чтобы модель была полезной, ее сложность должна находиться в определенном соотношении со сложностью описываемого объекта (пример: биологические системы).

Модели используются во многих областях исследований, в том числе в физике, биологии, астрономии, экономике, социологии и физиологии.

В этой книге основное внимание будет сконцентрировано на технических задачах, хотя излагаемые методы оценивания могут быть использованы и в других областях. В технике следует различать несколько целей использования моделей:

1) И с с л е д о в а н и е. Желательно дать интерпретацию полученной в результате измерений информации. Дальнейшее использование этой информации заранее не ясно, но само осмысливание и последующее уточнение формулировок имеют первостепенное значение (диагностика). Модель дает ключ к дальнейшему исследованию.

2) П р о е к т и р о в а н и е. Данные, характеризующие отдельные элементы или подсистемы, используются для того, чтобы построить модель системы, удовлетворяющую критерию проектирования (устойчивость, величина ошибки, доход, надежность и т. д.). С этой точки зрения адаптация реальной системы при повышении требований или необходимости обеспечить более экономичное функционирование может рассматриваться как часть процесса проектирования. По существу проектирование представ-

ляет собой процесс с обратной связью, в котором проектировщик оказывается элементом замкнутого контура.

3) Управление. Способы управления системой зависят от имеющейся информации. В этом случае следует различать:

а) условия нормального функционирования, например управление по каналам обратной и прямой связи, статическую и динамическую оптимизацию, адаптивное управление, групповое управление. Очевидно, существует тесная связь между пригодностью «непослушной» динамической системы и возможностями управления ею;

б) критические ситуации, такие, как частичная поломка, когда способ управления зависит от информации о типе и глубине отказа;

в) стартовые и финишные режимы, когда некоторые элементы программного управления могут зависеть от значений параметров и состояния системы.

Природа бросает вызов ученому, который пытается вывести какие-то общие закономерности, предлагая ему множество примеров применений систем управлений. В первую очередь это широкий набор искусственных систем управления в технике от простейших регуляторов до сложнейших вычислительных машин, управляющих технологическими процессами. Механизмы управления действуют также и в обществе (например, принуждение силой закона); имеется множество контуров регулирования в биологических системах. По-видимому, необходимость управления столь разнообразными системами можно объяснить желанием устраниТЬ имеющуюся «неопределенность».

Со временем применения знаменитого регулятора Уатта в бесчисленном множестве ситуаций стало ясно, что введение обратной связи можно использовать как эффективное средство для борьбы с неопределенностями. Эти неопределенности могут быть результатом непредсказуемых воздействий (возмущений) на систему со стороны среды или могут зарождаться внутри рассматриваемой системы (например, износ, старение, порча катализатора и т. д.). Обычная обратная связь может устранить влияние неопределенности или по крайней мере уменьшить эффект этого влияния до пренебрежимой величины. Конечно, это не панацея, и применение этого рецепта имеет свои гра-

ници. Эти ограничения проявляются в тех случаях, когда происходят большие изменения параметров; в этой ситуации возможно использование принципов *адаптивного управления*. С другим ограничением можно встретиться при такой оптимизации каких-либо стоимостных критериев, которая ведет к *оптимальным* или *субоптимальным* системам.

При проектировании обычной системы управления требуемая информация о поведении объекта исчерпывается данными (быть может, довольно неточными) об устойчивости и данными, необходимыми для оценки возможных улучшений, связанных с применением обратной связи, т. е. данными о характере нагрузок и возмущений, об ограничениях на управление и т. д. Из истории теории управления известно, что проектировщики редко располагают необходимой для синтеза системы управления полной априорной информацией об объекте и окружающей его среде. Даже если в принципе известны уравнения, описывающие поведение системы, то часто оказывается, что нет данных о величине отдельных параметров и к тому же нередко имеющиеся модели слишком сложны. Подобные ситуации естественны для самых разных областей. Имеются, однако, два признака, отличающие все задачи идентификации в автоматическом управлении:

- 1) часто для того, чтобы получить недостающую информацию, можно ставить эксперименты над системой;
- 2) цель идентификации состоит в выработке стратегии управления.

Одна из особенностей современной теории управления, несомненно связанная с успешным применением частотных методов в классической теории управления, состоит в том, что наряду с методами синтеза используются достаточно эффективные методы идентификации систем — методы частотного анализа. Эти методы сделали возможным точное определение передаточных функций, т. е. как раз тех характеристик, которые необходимы для применения методов синтеза, основанных на логарифмических диаграммах. Как правило, в современной теории управления применяются *параметрические* модели в пространстве состояний. Использование экспериментальных данных для построения таких моделей естественно возродило интерес к задачам оценивания параметров и родственным вопросам.

Для систем с большими вариациями параметров (что приводит к адаптивному управлению) и систем, функционирующих по чисто «стоимостным» критериям (что приводит к оптимальным или субоптимальным режимам), часто оказывается необходимой более точная и своевременная информация о параметрах и состояниях системы. Отсюда очевидна необходимость идентификации и оценки параметров и состояний. Подобную систему можно было бы рассматривать как знаменитый черный ящик, т. е. систему (подсистему) неизвестной структуры с одним или несколькими входами и выходами. Однако для большинства технических задач этот подход не слишком реалистичен. Во многих случаях экспериментатор располагает определенной априорной информацией о физической природе исследуемого процесса, пользуясь которой можно получить представление о структуре концептуальной модели, а может быть, и примерные значения ее коэффициентов (параметров). Поэтому ящик оказывается более или менее серым или прозрачным. Достаточно сказать, что рассматривается ректификационная колонна, а не лунная ракета, и это уже определенная априорная информация.

Характеристику некоторых особенностей задачи можно получить, задавая следующие вопросы:

1) Какую модель желательно построить — статическую (например, для оптимизации статического объекта) или динамическую (в частности, для оптимизации динамического объекта или управления полетом самолета), нелинейную или линеаризованную? Ответ на этот вопрос может определить желаемую точность и сложность модели, выбор математических методов и т. д.

2) Должна ли модель строиться вне контура регулирования на основании общих законов и каких-то отдельных экспериментов или ее можно строить в замкнутом контуре (в натуральном масштабе времени), когда допускаются непрерывные эксперименты с реальным объектом?

3) Должны ли приниматься во внимание финансовые соображения и какие?

Вслед за этим задается множество все более сложных вопросов. Приведем некоторые из них:

1) Как оценить качество модели?

2) Как учесть всю имеющуюся информацию?

3) В чем состоит оптимальная стратегия получения недостающей информации?

4) Как поступать с нелинейностями?

5) Можно ли аппроксимировать сложную систему простой моделью?

Ответы на эти вопросы зависят от конкретного класса систем. Некоторые общие соображения можно найти в гл. 4.

1.2. ПОСТРОЕНИЕ МОДЕЛИ

Во многих случаях построение модели начинается с использования основных физических законов (законов Ньютона, Максвелла или Кирхгоффа, законов сохранения массы, энергии и импульса, законов перераспределения количества тепла и энтропии) для описания исследуемого объекта, являющегося, например, механическим, электрическим или термодинамическим процессом. Из этих законов следуют различные соотношения между рассматриваемыми переменными и, в частности, связывающие их обыкновенные дифференциальные уравнения, дифференциальные уравнения в частных производных, разностные уравнения. Это соотношения между «силами» (интенсивные переменные) и «потоками» (экстенсивные переменные) (см. приведенную ниже таблицу).

	«Силы»	«Потоки»
Функции времени	Разность температур Разность концентраций Разность потенциалов	Поток тепла Поток массы Электрический ток
Функции времени и координат	Градиент температуры Градиент концентрации Градиент потенциала	Вектор плотности потока тепла Вектор плотности потока массы Вектор плотности электрического тока

Если известно состояние всех внешних и внутренних по отношению к системе условий и физическая модель системы является полной, то по крайней мере в принципе можно определить числовые значения всех коэффициентов уравнений. Такая ситуация, в частности, возможна при описании динамики космических аппаратов, но гораздо реже встречается в «земных» приложениях. Априорное знание об объекте обычно ограничено неопределенностью описания среды и неполнотой физической модели объекта. Даже если точно известны физические законы, управляющие поведением каждой элементарной ячейки системы, проклятье размерности может заставить нас воспользоваться упрощенным описанием поведения системы как целого. Это может означать, что значения параметров и координат модели не слишком поддаются априорному вычислению. Поэтому остается довериться идентификации и оцениванию. В результате построение модели сводится к следующим этапам:

- 1) выбор структуры модели из физических соображений;
- 2) подгонка параметров к имеющимся данным (оценивание);
- 3) проверка и подтверждение модели (диагностическая проверка);
- 4) использование модели по ее назначению.

Исходя из перечня научных направлений и разнообразия приложений, по этим этапам нельзя дать каких-либо общих рекомендаций. Структура модели выбирается на основе исходной (априорной) информации о системе и преследуемых целях¹⁾. На практике отыскание подходящей модели может быть достаточно трудной задачей даже для узкой прикладной области.

В этой книге основное внимание уделяется вопросам оценивания. Использование термина «оценивание» связано с тем, что в большинстве реальных ситуаций наблюдения над системой искажены случайными воздействиями (возмущениями, ошибками). Поэтому для того, чтобы получить «наилучшие» результаты, фильтруя влияние помех, приходится пользоваться статистическими методами.

¹⁾ Имеются также приближенные методы оценивания структуры модели при идентификации по данным вход — выход. — Прим. ред.

Проверка и подтверждение модели тесно связаны с оцениванием; по окончании процедуры оценивания следует выяснить, в какой степени модель действительно «объясняет» поведение системы. Схематично такую проверку можно представить как подачу на систему и модель одного и того же входного сигнала и изучение вида и величины разности обоих выходных сигналов (ошибки или невязки).

Применение модели существенно зависит от выбора конкретного объекта. Примеры для технологических процессов приведены в работе [8]. Примеры из других областей можно найти в работах, упоминаемых в гл. 14. Хороший пример выбора, подгонки, диагностирования и применения (в частности, для прогнозирования) модели описан в работе [4].

1.3. СТРУКТУРА, ПАРАМЕТРЫ И СОСТОЯНИЯ

Основной задачей системного анализа является определение выходного сигнала системы по известному входному сигналу и характеристикам системы. В этой книге обсуждается задача, которую иногда называют *обратной задачей* системного анализа, — по заданным входному и выходному сигналам определить уравнения (дифференциальные), описывающие поведение системы.

Заде [30] называет *идентификацией* «определение по входу и выходу системы из определенного класса систем, которой испытываемая система эквивалентна». Используя формулировку Заде, необходимо определить класс систем $\mathcal{S} = \{S\}$, класс входных сигналов \mathcal{U} и понятие «эквивалентность». В дальнейшем испытываемую систему будем называть просто *объектом*, а элементы \mathcal{S} — *моделями*. Эквивалентность часто понимается в смысле какого-либо критерия ошибки или функции потерь, являющейся функционалом от выхода объекта y и выхода модели y_M , т. е.

$$E = E(y, y_M).$$

Говорят, что две модели M_1 и M_2 *эквивалентны*, если значения функций потерь для этих моделей одинаковы, т. е.

$$E(y, y_{M_1}) = E(y, y_{M_2}).$$

Существует большая свобода в постановке задачи идентификации, что нашло отражение в литературе по идентификации. На выбор класса моделей \mathcal{S} , класса входных сигналов \mathcal{U} и критерия сильно влияет априорная информация об объекте и цель идентификации. Необходимо пояснить, что модель дает информацию трех видов, а именно:

- 1) о структуре в форме математических тождеств, блок-схем, сетей или графов, матриц связи;
- 2) значения параметров, т. е. величины, не зависящие от входов, или независимые величины;
- 3) значения зависимых переменных(состояний) в фиксированный момент времени или как функций времени.

Значение структуры нельзя переоценить. Ее выбор определяется типом применений модели и может оказаться решающим фактором успеха или неудачи принятой схемы оценивания.

Часто подразумевается, что идентификация начинается из «ничего» без всякой априорной информации об объекте. Но, как уже отмечалось, в большинстве технических и в ряде биологических задач такое предположение не реалистично; из структуры объекта и по крайней мере частичного понимания его функционирования можно извлечь определенную априорную информацию и, в частности, вид структуры модели. В этом случае остается только получить информацию о числовых значениях ряда параметров (коэффициентов дифференциальных уравнений, описывающих динамику объекта, коэффициентов линейной или нелинейной модели объекта и т. д.) и (или) состояний. В результате задача идентификации сводится к задаче оценивания параметров и (или) состояний.

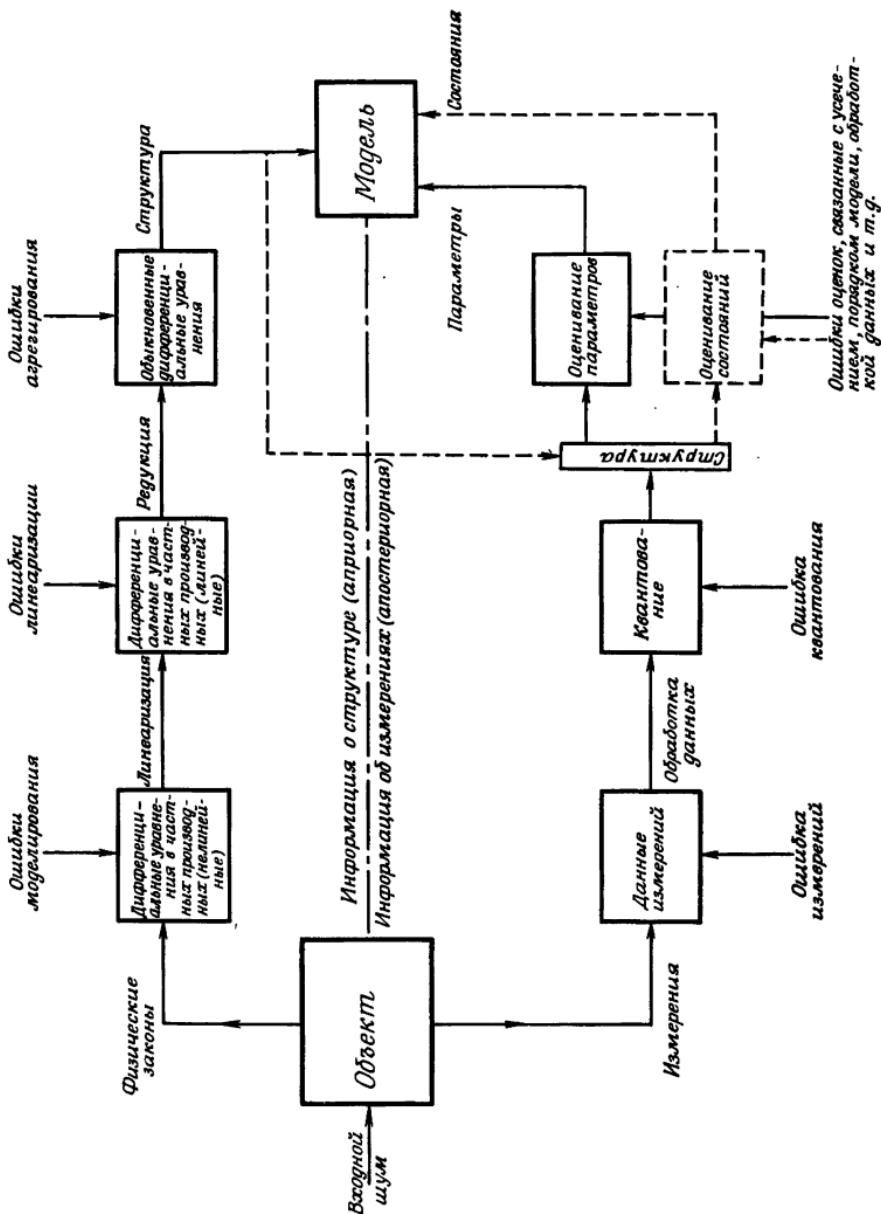
Под *оцениванием параметров* понимается экспериментальное определение значений параметров, характеризующих динамику поведения объекта, в предположении, что структура модели объекта известна. Следует отметить, что различие между знанием структуры модели и информацией о величине параметров не столь уж значительно, как это может показаться с первого взгляда. Изменение величины параметра от $\beta \neq 0$ до $\beta = 0$ может соответствовать такому упрощению структуры, как исчезновение ветви модели. С другой стороны, ясно, что дополнить модель таким

образом невозможно; потенциальные возможности модели ограничены исходным выбором структуры. Известно лишь несколько теоретических результатов по методам улучшения моделей добавлением новых структурных элементов. Здесь ограничениями по-прежнему являются возможности человеческого интеллекта и интуиции.

В некоторых случаях интересуются более подробными сведениями, в частности непрерывной информацией о состоянии объекта, что приводит к задаче *оценивания состояний*. Грубо говоря, состояние объекта — это переменная (например, вектор), которая вместе с входным сигналом объекта полностью определяет его дальнейшее поведение. Следовательно, если предоставить объект самому себе, т. е. когда нет входного сигнала, то знания состояния объекта в какой-то момент времени (вместе со знанием структуры и параметров объекта) достаточно для того, чтобы предсказать его будущее поведение. Так, например, начальные условия дифференциального уравнения можно рассматривать как вектор состояния (см. гл. 4). Задача определения наилучшей оценки состояния может встретиться в таких современных приложениях теории управления, как оптимальное управление. Если при этом не определено число оцениваемых параметров, то имеет место задача совместного оценивания параметров и состояний. Из этих объяснений понятно, что в большинстве случаев требуется оценивать состояние в условиях нормальной эксплуатации. Решение задачи оценивания параметров может быть основано как на активных экспериментах, так и на наблюдениях над исследуемым объектом.

На фиг. 1.1¹⁾ показана связь между априорной информацией (о структуре) и апостериорной информацией (об измерениях) при построении модели. Верхняя часть рисунка иллюстрирует процесс построения модели как конкретный пример использования физических законов с последующей линеаризацией и преобразованием к системе обыкновенных дифференциальных уравнений. Получающиеся уравнения определяют структуру модели. На каждом шаге возникают ошибки. В нижней части рисунка иллюстрируется процедура оценивания, основанная

¹⁾ Результат обсуждения с А. У. Каном.



на измерениях и включающая обработку данных и алгоритмы оценивания. Здесь также следует учесть различные виды ошибок.

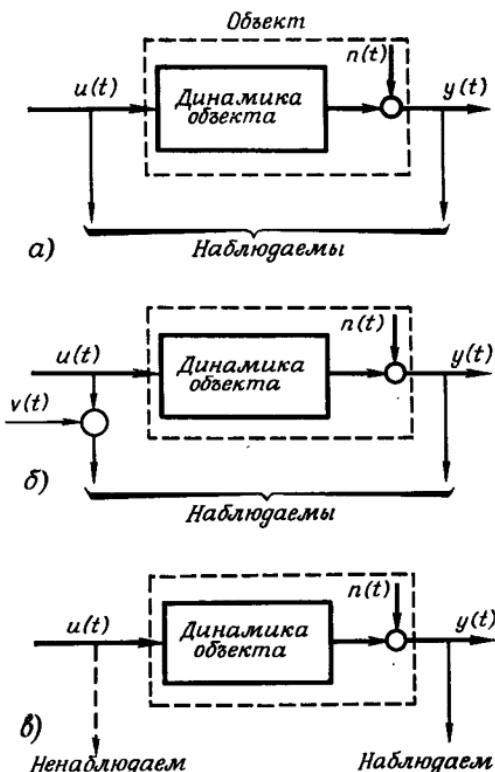
Отметим, что оценивание (параметров и (или) состояний) предполагает определенную структуру модели. Из-за неограниченного разнообразия возможных объектов и множества различных целей, требующих разных типов моделей, фиг. 1.1 можно рассматривать только как иллюстрацию; полное обсуждение проблемы построения модели выходит за рамки этой книги, в которой центральное место занимает последовательное изложение методов оценивания.

1.4. ОПИСАНИЕ ПРОБЛЕМЫ

Несколько простых примеров возможных задач представлено на фиг. 1.2, где $u(t)$ и $y(t)$ — входной и выходной сигналы объекта соответственно. Развличие между схемами, изображенными на фиг. 1.2, *a*, *b* и *c*, связано с наблюдаемостью входного сигнала, который может быть наблюдаемым точно, наблюдаемым в смеси с шумом и ненаблюдаемым [12]. Во всех случаях выход y искажен шумом. Необходимо извлечь информацию о параметрах и (или) состояниях из наблюдаемых сигналов и априорной информации о функциях плотности вероятности. Для случая систем, охваченных обратной связью, и многомерных систем могут применяться комбинации этих моделей. Возможные следствия будут рассмотрены в гл. 11.

Термин «объект» может использоваться в самом широком смысле и обозначать техническую систему (самолет, химический завод, коротковолновую линию радиопередач и другие каналы связи), биологическую систему (движение человека, медицинскую диагностику), а также экономическую и социальную системы.

Конечно, описание проблемы может быть усложнено несколькими обстоятельствами. Прежде всего возможны такие многомерные объекты, в которых каждый вход влияет на несколько выходов. В этом случае можно использовать векторное описание входного сигнала $u(t)$ и выходного сигнала $y(t)$; каждая компонента этих векторов представляет собой функцию времени, характеризующую входной или выходной сигнал. Другая трудность связана с тем,



Ф и г. 1.2.

что шум может быть не только аддитивным, но и мультипликативным (например, не представляющее обычно интереса затухание или переменный коэффициент усиления).

В разд. 1.1 мы подчеркнули значение моделей. Задаче оценивания можно дать следующее схематичное описание. На объект и модель действует один и тот же входной сигнал. Сравниваются искаженный помехой выходной сигнал объекта и выходной сигнал модели. Необходимо определить оптимальный в заданном смысле способ корректировки модели. Значения параметров непосредственному наблюдению не доступны. Таким образом, критерием выбора оптимума должен быть *функционал от выходных сигналов или от математического ожидания ошибок оценок параметров* (см. разд. 2.3). Выходной сигнал функцион-

нально связан с числовыми значениями параметров. Это приводит к усложнению алгоритмов оценивания и, вообще говоря, к необходимости использования вычислительных машин. Как уже указывалось, критерием часто является скалярная функция потерь. Тогда естественно задать несколько вопросов:

- 1) Достигается ли минимум?
- 2) Существует ли единственное решение?
- 3) Связана ли единственность решения с выбором входных сигналов?
- 4) Если решение не единственно, то как выглядят модели, обеспечивающие одинаковую величину функции потерь, и как следует ограничить класс \mathcal{S} , чтобы гарантировать единственность?

На некоторые из этих вопросов получены ответы для простого класса линейных систем, интерес к которым возник в связи с работами Беллмана и Острэма в медицине [3].

В том случае, когда задача идентификации ставится как задача оптимизации, функция потерь выбирается особым образом; если же ставится задача оценивания, то используются другие предположения.

Часто критерий записывается как функционал ошибки, например

$$E(y, y_M) = \int_0^T e^2(t) dt,$$

где y — выход объекта, y_M — выход модели, а e — ошибка; y , y_M и e рассматриваются как функции, определенные на $(0, T)$. Отметим, что этот критерий можно интерпретировать как критерий наименьших квадратов для ошибки e .

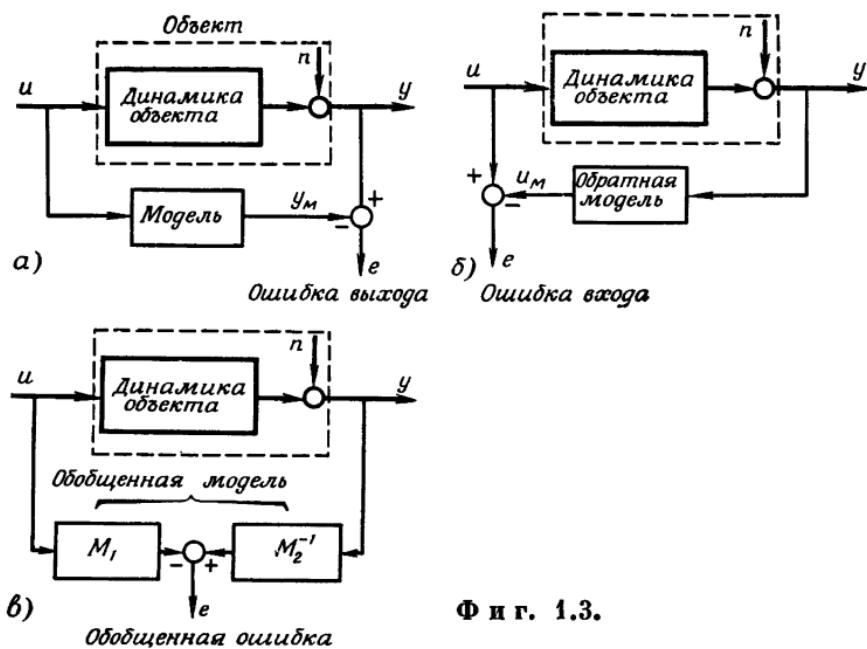
Будем говорить, что рассматривается *выходная ошибка*, если

$$e = y - y_M = y - M(u),$$

где через $M(u)$ обозначен выходной сигнал модели, на вход которой подается сигнал u . Это определение естественно только в том случае, когда единственной помехой является белый шум измерений выходного сигнала (фиг. 1.3, a).

Если

$$e = u - u_M = u - M^{-1}(y),$$



Фиг. 1.3.

где через u_M обозначен вход модели, выходной сигнал которой равен y , то говорят, что e — *входная ошибка* (фиг. 1.3, б). Обозначение M_2^{-1} предполагает, что модель обратима, т. е., грубо говоря, для заданного выхода всегда можно найти единственный входной сигнал¹). С точки зрения теории оценивания использование входной ошибки естественно тогда, когда помехи представляют собой белый шум на входе системы.

В общем случае ошибку можно определить как

$$e = M_2^{-1}(y) - M_1(u),$$

где M_2^{-1} представляет собой обратную модель (фиг. 1.3, в). Этот тип модели и ошибки будем называть *обобщенной моделью* и *обобщенной ошибкой* [7]. Специальный вид обобщенной ошибки — так называемая ошибка уравнения — использован в работе [16].

Другой тип задач идентификации получается при вложении решаемой задачи в семейство вероятностных задач.

¹⁾ Строгое определение обратимости приведено в работах [5, 22, 23].

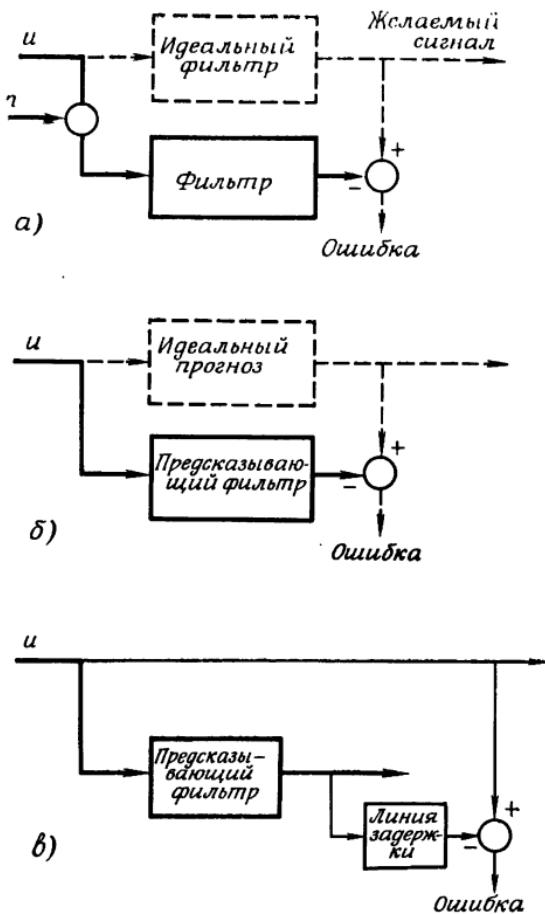
Если \mathcal{S} — параметрический класс, $\mathcal{S} = \{S_\beta\}$, где β — параметр, то задача идентификации сводится к задаче оценивания параметра. Такая постановка делает возможным использование теории оценивания и теории принятия решений. В частности, можно использовать такие методы, как метод максимального правдоподобия, байесовские методы, минимаксные методы. Можно задавать точность оценок параметров и проверять различные гипотезы.

Оказывается, однако, что многие вероятностные задачи оценивания можно свести к задаче оптимизации. Но в этом случае функция потерь определяется из вероятностных соображений. Заданной же функции потерь часто можно дать вероятностную интерпретацию.

Теория оценивания тесно связана со статистической теорией оптимальных систем [17]. Задача оптимальной фильтрации (фиг. 1.4, а) относится к тому же классу, что и задача предсказания (фиг. 1.4, б и в); возможны комбинированные постановки задач. Она связана также с задачами адаптации и обучения [23—26]. Задача становится еще более трудной, когда изучаемый объект является звеном контура регулирования. Фельдбаум [9] рассмотрел ситуацию, когда управляющее воздействие используется в двух целях:

- 1) для *обучения* или *изучения* характеристик объекта или способов управления объектом;
- 2) для повышения качества управления, *приведения* объекта в требуемое состояние.

Такое сочетание *изучения* и *управления* Фельдбаум предложил назвать *дуальным управлением*. Он показал, что иногда эти две функции разделяются в частотной или временной области. Однако такое разделение не является необходимым, Фельдбаум [9, 10] разработал общую теорию дуального управления. Он пишет: «В системах дуального управления возникает противоречие между двумя сторонами управляющего воздействия — изучающей и направляющей. Действительно, успешное управление возможно лишь при своевременном воздействии на объект. Запоздалое воздействие ухудшает процесс управления. Однако успешно управлять можно, лишь достаточно хорошо зная свойства объекта. Но, чтобы их узнать лучше, необходимо много времени. Слишком «поспешная»



Ф и г. 1.4.

управляющая часть будет производить направляющие действия без должного обоснования их полученной в результате изучения объекта информацией. Слишком «осторожная» система будет излишне долго выжидать, обрабатывая полученную информацию и не направляя своевременно объект к требуемому режиму. И в том и другом случае процесс управления может оказаться не наилучшим и даже неуспешным. Задача состоит в определении наивыгоднейшего в каком-либо смысле сочетания двух указанных выше сторон в управляющем воздействии. Воздействия следует подбирать так, чтобы какой-либо критерий каче-

ства приобрел наивысшее значение. Поэтому управляющая часть, принимая решение об определенной величине управляющего воздействия в данный момент времени, должна учитывать два фактора:

а) Потерю, происходящую в значении критерия качества вследствие того, что результат воздействия как в данный момент, так и в последующие вызовет отклонение объекта от требуемого или наилучшего возможного режима. Среднюю меру этой потери назовем *риском действия*.

б) Потерю, происходящую в значении критерия качества вследствие того, что значение управляющего воздействия в данный момент времени оказалось не наилучшим для получения информации о характеристиках объекта, в связи с чем последующие воздействия также оказываются не наилучшими из возможных. Среднюю меру такой потери назовем *риском изучения*.

Кроме того, представляет интерес связь между параметрами объекта и сигнала. Многие детерминированные сигналы, по крайней мере в принципе, можно получить как импульсную переходную характеристику или реакцию на скачок линейного или нелинейного фильтра. Многие случайные сигналы могут быть получены в результате фильтрации белого шума. Отсюда вытекает, что в подобных ситуациях характеристики сигнала можно выразить через параметры фильтра (объекта) и наоборот. Поэтому между параметрами сигнала и объекта существует тесная связь.

1.5. УСЛОВИЯ НОРМАЛЬНОГО ФУНКЦИОНИРОВАНИЯ

Как правило, постановка *эксперимента* подразумевает выполнение ряда существенных условий или возможность активного изменения условий, необходимых при повторении эксперимента. В то же время, например, в задачах небесной механики (до появления спутников) приходилось ограничиваться пассивными *наблюдениями*.

Иногда *продолжительность* эксперимента не имеет существенного значения. В этом случае количество получаемой информации не ограничено *продолжительностью* эксперимента или числом повторений эксперимента. В то же время могут существовать ограничения, связанные с эконо-

мическими и другими практическими соображениями. Конечно, имеются и такие фундаментальные ограничения, как принцип неопределенности Гейзенберга, утверждающий невозможность одновременного измерения положения и скорости частицы с любой желаемой точностью.

При оценивании параметров объекта время наблюдения часто ограничено экономическими или принципиальными техническими соображениями (например, в случае необходимости определения меняющихся во времени характеристик объекта). К тому же оценивание параметров должно производиться в условиях нормального функционирования. Иногда этого удается достичь, используя (случайные) сигналы естественного происхождения, в других случаях на систему подаются достаточно мощные пробные сигналы.

Следует отметить, что на практике ситуация может быть сложнее. Многие эксперименты ставятся в условиях, которые характерны и для нормального функционирования объекта. Возможен широкий диапазон ситуаций. На одном краю этого диапазона размещаются астрономические наблюдения, производимые, естественно, в условиях нормального функционирования. Использование генератора синусоидальных колебаний для определения передаточной функции цепи в лабораторных условиях представляет собой пример ситуации, которая находится на другом краю диапазона.

Большинство практических ситуаций, например при контроле технологических процессов, лежит между этими двумя крайними случаями, так как обычно состояние среды контролируется лишь частично. Иногда приходится проделывать что-то вроде эксперимента, наблюдая за изменением состояния объекта при подаче на вход объекта каких-либо возмущений. Однако на практике возможности эксперимента жестко ограничены. Для построения реалистичных моделей часто необходимо ставить эксперименты *в процессе нормального функционирования*. Это означает, что подаваемые возмущения должны быть достаточно малы для того, чтобы не слишком влиять на производство. Для удержания процесса в допустимых пределах, вероятно, может потребоваться несколько регуляторов. А это может заметно влиять на результаты оценивания.

При проведении таких экспериментов естественно возникает много вопросов:

1) Как планировать эксперимент? Стоит ли пользоваться последовательным методом, когда эксперимент ставится на основе априорной информации, полученный результат используется при планировании следующего эксперимента и т. д. Когда обрывать эксперимент?

2) Как анализировать результаты эксперимента для того, чтобы получить управление с желаемыми качествами? Насколько можно доверять результатам?

3) Какие возмущения следует использовать для того, чтобы получить в условиях эксперимента предельно возможные результаты?

4) Как выбрать интервал квантования, если используется цифровая вычислительная машина?

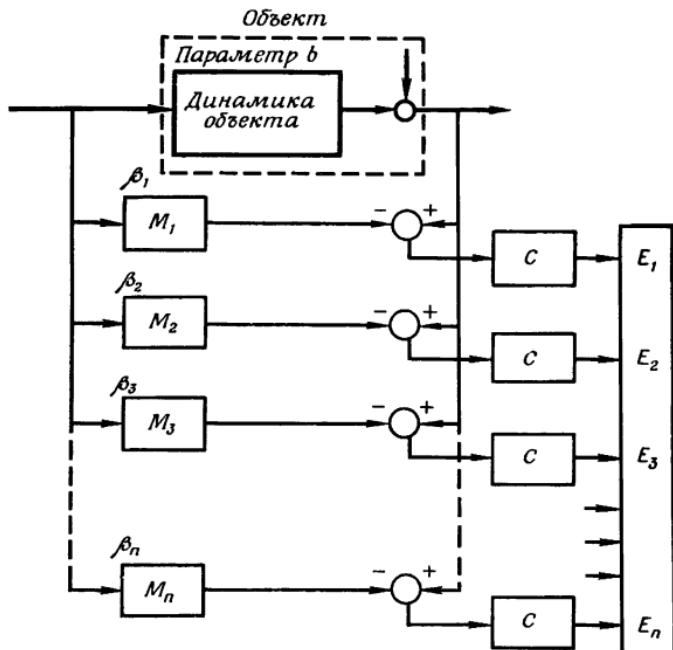
Несмотря на то что на развитие теории идентификации затрачено много труда, на эти вопросы до сих пор практически нет общих ответов. Однако применительно к конкретным ситуациям многие из этих задач решены. Следует иметь в виду, что обычно целью идентификации является создание системы управления, уже это может внести ясность в задачу идентификации. Приведем типичный пример. Как оценить точность идентификации — по отклонениям *параметров* модели или ее *отклика*? Если основная цель состоит в проектировании системы управления, то представляется логичным оценивать точность идентификации по *результатам функционирования* созданной на основе идентификации объекта системы управления.

В любом случае прежде всего важно, чтобы:

1) входной сигнал возбуждал все собственные колебания объекта, т. е. объект должен быть *управляемым* (см. гл. 4), а сигналы должны принадлежать к классу так называемых *постоянно возбуждающих* (см. разд. 11.4);

2) выходной сигнал содержал достаточно информации о том, «что происходит внутри объекта» (см. гл. 4): *наблюдаемость и идентифицируемость*.

Вообще говоря, интервал оценивания должен быть как можно более коротким, так как необходима достоверная информация о медленно изменяющихся параметрах. С другой стороны, минимальное время оценивания ограничено наличием шумов (ошибки измерений, влияние



Ф и г. 1.5.

аддитивных помех). В некотором смысле продолжительность интервала оценивания связана с затратами на измерение. Это можно пояснить по фиг. 1.5.

Допустим, что нужно определить только один параметр объекта b . Для этого можно было бы использовать n моделей одинаковой структуры с различными значениями параметра β_i ($i = 1, \dots, n$). Выход объекта сравнивается с выходом каждой модели. Мгновенные ошибки поступают в блоки C подсчета критерия, например интеграла от квадрата ошибки на интервале T . В этом случае параметр b определяется за один интервал измерения T , так как $b \approx \beta_i$, если $E_i < E_j$ для всех $j \neq i$. Для стационарного объекта такую оценку можно получить с помощью одной модели с переменным параметром, который в j -м интервале измерения выбирается равным β_j . Таким образом, в этом случае мы получаем необходимую информацию за время nT , но при гораздо более низких затратах. В связи с этим нужно осторожно сравнивать разные схемы оценивания. При большом числе параметров m общее число моделей

или общее число интервалов измерения растет как n^m . Уже отмечалось, что имеется континуальное множество различных ситуаций и измерения в условиях нормального функционирования принципиально отличаются от измерений в условиях пробных испытаний. Это отличие связано:

- 1) с информацией об объекте, содержащейся в текущих значениях его выходного сигнала;
- 2) с требованиями к времени измерений и оценивания;
- 3) с выбором пробных сигналов.

Поэтому необходимо подробное исследование эффективности системы обработки информации, используемой в рассматриваемой схеме оценивания. В результате такого анализа должны быть определены:

- 1) скорость сходимости к асимптотическим оценкам;
- 2) величина смещения асимптотических оценок по отношению к истинным значениям параметров;
- 3) дисперсия оценок параметров как функция времени;
- 4) оценка влияния аддитивных помех и нелинейностей и т. д.

Все эти характеристики следует сравнить со скоростью возможного изменения параметров объекта.

Рассмотрим кратко упомянутые выше особенности условий нормального функционирования объекта. Информация об объекте, содержащаяся в его выходном сигнале, зависит от предыстории входных сигналов. Это можно видеть на простейшем примере уравнения свертки для линейного стационарного объекта:

$$y(t) = \int_{-\infty}^t u(\theta) h(t - \theta) d\theta,$$

где u — входной сигнал объекта, y — неискаженный шумами выходной сигнал и $h(t)$ — весовая функция объекта. Мы хотим получить информацию о $h(t)$, наблюдая входной и выходной сигналы объекта, начиная с момента $t = T$. Выход $y(t)$ для $t > T$ также зависит от значений $u(t)$ для $t < T$. Эта часть выходного сигнала может быть записана в виде

$$y_1(t) = \int_{-\infty}^T u(\theta) h(t - \theta) d\theta.$$

Сигнал $y_1(t)$ действует как помеха, так как он не связан с наблюдаемой реализацией входного сигнала $u(t)$ при $t > T$. Эту трудность можно описать и в терминах пространства состояний (см. гл. 4). Грубо говоря, каждому объекту, описываемому дифференциальным уравнением n -го порядка, можно поставить в соответствие n -мерное пространство состояний. Координаты этого пространства представляют собой n независимых величин, например значения координат объекта и ее производных до $(n - 1)$ -го порядка. Состояние системы в произвольный момент времени t определяется точкой пространства состояний, называемой вектором состояния $x(t)$. Эта точка перемещается под влиянием сил, восстанавливающих положение равновесия (рассеяние энергии) и вынуждающих воздействий (входные сигналы). В момент T состояние $x(T)$ содержит всю возможную информацию о предыстории системы. При отсутствии вынуждающих воздействий (входных сигналов) вектор состояния устойчивой системы будет со временем стремиться к некоторому положению равновесия x_0 . Эта компонента движения вектора состояния, свободное движение, складывается с вынужденным движением, вызванным входными сигналами.

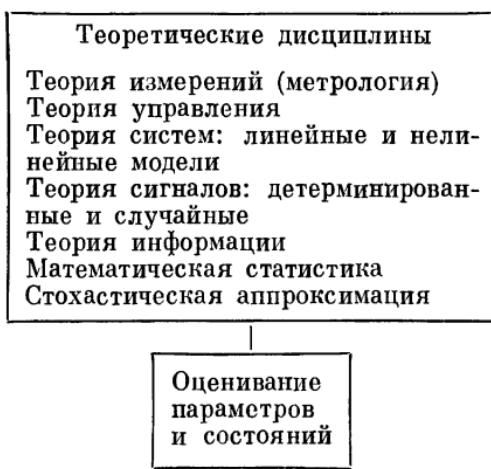
Таким образом, свободное движение действует как помеха, если входной и выходной сигналы наблюдаются, начиная с момента $t = T$. В результате обобщенная задача оценивания состоит в оценивании неизвестных параметров и неизвестного состояния при $t = T$.

Очевидно, что требования к времени измерения и оценивания связаны с тем, что при оценивании параметров должны использоваться характеристики сигналов, которые можно получить, только наблюдая сигнал на каком-то интервале времени. Вообще имеется тесная связь между продолжительностью интервала и точностью получаемой информации. Последнее замечание имеет отношение и к выбору пробных сигналов. В лабораторных условиях часто можно использовать мощные или периодические пробные сигналы. Иногда это можно сделать в условиях нормального функционирования. Если нельзя применить пробные сигналы, необходимо максимально использовать имеющиеся помехи. Если к входному сигналу можно добавить пробный сигнал низкого уровня, то необходимо рас-

смотреть отношение сигнал/шум, вопросы наилучшего выбора характеристик пробного сигнала, влияние вида сигнала на процесс получения информации и т. д. Изменение параметров объекта вносит принципиальную неопределенность в результаты оценивания.

1.6. НЕКОТОРЫЕ ОБЛАСТИ ПРИМЕНЕНИЯ

Ниже приводится схема, где перечислены дисциплины, имеющие отношение к оцениванию параметров и состояний, и некоторые области применения теории оценивания. В теории случайных сигналов и стохастической аппроксимации,



Цели применения	Области применения	Родственные задачи
Диагностика	Системы связи	Детектирование
Управление	Энергетика	Оптимальная фильтрация
Автоматический контроль	Машиностроение	Предсказание
Автоматизация принятия решений (в промышленности)	Авиация	
Автоматическая корректировка (в промышленности)	Химическая промышленность	Обучение
Распознавание образов	Физика	
	Геология	
	Экономика	
	Биология и медицина	

в математической статистике и теории информации широко применяются теоретико-вероятностные обозначения и теоремы теории вероятностей. Сводку необходимых результатов, которые предполагаются хорошо известными, можно найти в приложении Б. Теоретические аспекты описания сигналов и моделей обсуждаются в гл. 3—5.

Перечень применений показывает, что возможное использование оценивания параметров не сводится только к решению задач управления. В каждом случае цель оценивания ясна заранее. Эта цель определяет требуемую точность и допустимые затраты.

Оценивание в диагностических приложениях

Сюда относятся измерение передаточной функции объекта (например, построение диаграммы Боде), определение весовой функции, отыскание полюсов и нулей передаточной функции, оценивание нелинейностей. К этому же классу принадлежит знаменитый метод фильтрации Фурье, являющийся техническим воплощением следующей математической схемы:

$$\operatorname{Re}[H(j\omega)] = \lim_{k \rightarrow \infty} \frac{1}{k} \int_0^{kT} y(t) u \sin \omega t dt,$$

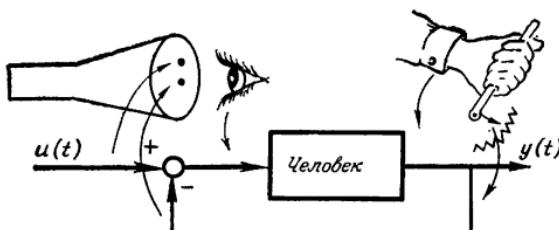
$$\operatorname{Im}[H(j\omega)] = \lim_{k \rightarrow \infty} \frac{1}{k} \int_0^{kT} y(t) u \cos \omega t dt,$$

где $u(t) = u \sin \omega t$ — входной сигнал объекта, а $y(t)$ — его выходной сигнал. Подобные задачи решаются не только в электротехнике и для электромеханических систем, но и в таких областях, как:

1) Химическая промышленность. Примеры: оценивание старения или загрязнения катализаторов, определение скорости протекания химических реакций.

2) Технологические процессы. Пример: определение коэффициентов теплопроводности.

3) Физические измерения. Пример: измерение импеданса ртутного электрода, представляющего собой стеклянную трубку, на одном конце которой набухает ртутная



Ф и г. 1.6.

капля. Достигая определенного размера, капля срывается и на ее месте начинает расти новая капля. Такой электрод является электродом с «чистой» поверхностью.

4) Машиностроение. Пример: измерение передаточных характеристик железнодорожных вагонов, ткацких станков.

5) Атомная энергетика. Пример: определение коэффициентов размножения тепловых нейтронов.

6) Аэронавтика и космонавтика. Пример: исследование динамики самолета в полете.

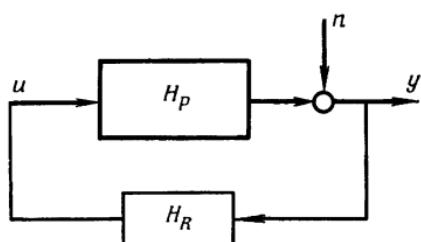
7) Биология и медицина. Пример: определение характеристик человека при отслеживании. Простое приспособление показано на фиг. 1.6. Задача состоит в максимально точном отслеживании движения светового пятна на экране катодной трубки. Можно исследовать изменения характеристик, связанные с «обучением», усталостью, воздействием среды, введением различных стимуляторов. Другие примеры — изучение реакции человека на неожиданную вспышку света, определение параметров модели системы кровообращения.

8) Экономика.

Оценивание для целей управления

Если целью идентификации является проектирование системы управления, то в зависимости от решаемой задачи управления может сильно меняться постановка задачи идентификации. Несколько примеров:

1. Конструирование устойчивого регулятора.
2. Синтез программы оптимального перехода из одного состояния в другое.



Фиг. 1.7.

3. Конструирование регулятора, который минимизирует отклонения координат объекта, вызванные помехами.

В первом случае может оказаться достаточным довольно грубое описание динамики системы. Для решения второй задачи управления

может потребоваться более точная модель. Для третьей задачи необходимо также иметь модель возмущающей среды.

Следует также отметить, что идентификация системы, охваченной обратной связью, требует осторожности. Рассмотрим классический пример, иллюстрируемый фиг. 1.7. Попытка идентифицировать H_P , измеряя u и y , дает оценку

$$\hat{H}_P = \frac{1}{H_R},$$

т. е. величину, обратную передаточной функции обратной связи. В то же время в промышленных приложениях способ введения обратной связи может быть достаточно тонким, например с помощью оператора, который время от времени корректирует систему. Интересный результат принадлежит Фишеру, который показал, что объект может быть идентифицирован, если обратную связь сделать нелинейной.

Системы оптимального управления. Как правило, оптимальность понимается в смысле какого-либо экономического критерия. Во многих случаях качество оптимального управления, по-видимому, сильно зависит от информации о динамике управляемого объекта. Если динамические характеристики изменяются во времени, их нужно определять непрерывно. Для оптимального управления очень важно также знать состояние объекта. Это приводит к совместному оцениванию параметров и состояний. Как показано Аоки [1], только для узкого класса систем (а именно линейных объектов с квадратичной функцией потерь) схема оптимального управления может быть разбита на оптимальное оценивание с последующим оптимальным регулированием (см. гл. 14). Для других классов

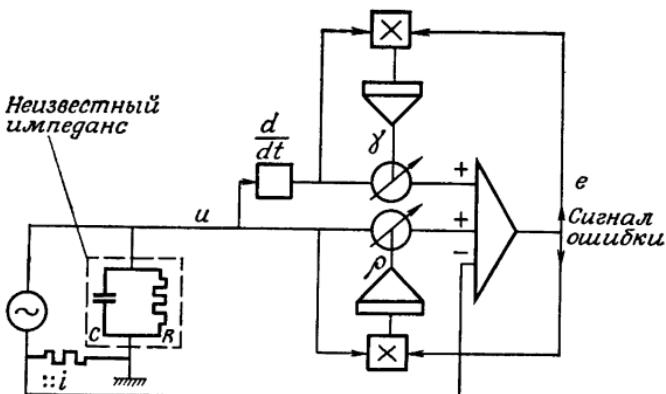
систем такое разделение ведет лишь к приближенно оптимальному управлению. Тем не менее такое расчленение задачи оптимального управления может оказаться очень полезным. В качестве примеров можно привести управление посадкой самолета, управление процессами групповой обработки деталей, управление непрерывными технологическими процессами, когда оптимальные рабочие точки органов управления выбираются на основе (частных) моделей объекта. Интересные примеры описаны в работе [15]:

1. Оптимальное управление электроэнергетическими установками.
2. Применение калмановской фильтрации для управления полетом самолета.
3. Применение в целлюлозно-бумажной промышленности.
4. Оптимальное управление запуском космической ракеты. Другие примеры читатель найдет в гл. 14.

Самонастраивающиеся и адаптивные системы. Для этого класса систем справедливо все, что было сказано об оптимальных системах. На основе информации о параметрах объекта параметры системы управления изменяют так, чтобы обеспечить оптимальность динамических характеристик объекта. Блестящим примером является полет самолета, динамика которого претерпевает большие изменения в результате вариаций плотности воздуха у поверхности Земли и на высоте 10 км. Другим примером адаптивного управления, основанного на оценивании параметров объекта, может служить вождение автомобиля. В плохую погоду дорога становится скользкой. Это можно заметить по небольшим возмущениям руля автомобиля. Оценивание результатов этих возмущений дает информацию, которая используется для регулирования скорости.

Оценивание при автоматическом контроле

Для этого класса задач исключительно важной характеристикой экономической эффективности является скорость измерений. Примером использования нового подхода к решению старой задачи может служить автоматическое измерение импеданса (фиг. 1.8). Импеданс может



Ф и г. 1.8.

быть описан следующим дифференциальным уравнением:

$$Cu' + \frac{1}{R} u - i = 0.$$

Уравнение модели имеет вид

$$\dot{yu} + \rho u - i = e.$$

В схеме, изображенной на фиг. 1.8, модель корректируется так, что сигнал ошибки стремится к нулю, иначе говоря, минимизируется среднеквадратичное отклонение. Этот принцип может быть использован при измерении более сложных импедансов нелинейных устройств, например пьезокристаллов, электромеханических преобразователей.

Оценивание при автоматизированном принятии решений (в промышленности)

В этом случае главное — это скорость обработки информации, особенно если эта скорость так велика, что превышает возможности человека. В качестве примера можно рассмотреть принятие решения о выключении генератора на электростанции при коротком замыкании на распределительном щите. Критерий принятия решения — перегорит или не перегорит генератор. Это определяется состоянием генератора и распределительной системы.

Оценивание при автоматической корректировке (в промышленности)

Автоматическую корректировку можно рассматривать как последовательность повторяющихся решений о продолжении и характере корректировки. Промышленную настройку широкополосных фильтров радио- и телевизионных приемников можно интерпретировать как корректировку модели по стандартной или эталонной цепи. Корректировка применяется также для устранения ошибок типа замещения символов при передаче информации по каналам связи. В этом случае статистические свойства помехи зависят от конфигурации линии связи, которая со временем меняется. Используя фильтр, состоящий из линии задержки и потенциометров, можно построить корректирующий фильтр (адаптивный фильтр), который будет автоматически подстраиваться, обеспечивая минимальный уровень помех.

Распознавание образов

Распознавание образов можно рассматривать как процедуру определения нескольких параметров исследуемого объекта и основанного на оценках этих параметров решения о том, к какому классу принадлежит объект. Основной вопрос при распознавании образов состоит в выборе наиболее важных параметров. В зависимости от приложений можно говорить о распознавании образов, с классификацией сигналов (анализ электрокардиограмм и т. д.), о классификации данных, полученных при поиске нефтеносных месторождений.

В интересном обзоре Клаймера [6] перечислены следующие области возможного применения методов оценивания параметров:

А. Авиация и транспорт.

- 1) Определение аэродинамических коэффициентов по:
 - а) результатам испытательных полетов, включающих сложные маневры;
 - б) динамическим испытаниям гибких моделей в аэродинамических трубах;
 - в) испытаниям ракетоносителей.

- 2) Определение скорости наступления флаттера по результатам испытаний в режимах, близких к критическим, путем выяснения характеристик затухания флаттера как функции частоты:
 - а) для самолетов в полете;
 - б) при продувке моделей в аэродинамической трубе;
 - в) для моделей ракетоносителей.
 - 3) Определение характеристик упругости различных конструкций.
 - 4) Определение внутренних характеристик серво-механизмов и их выходного импеданса.
 - 5) Определение баллистических свойств снарядов, ракет и т. п.
 - 6) Определение динамических характеристик колес.
 - 7) Определение нелинейных характеристик штурвалов, рулей высоты, рычагов управления.
 - 8) Определение динамических характеристик шасси и системы автоматического подъема шасси.
 - 9) Определение динамических характеристик турбореактивных, турбовинтовых, поршневых и других двигателей, компрессоров и т. п.
 - 10) Определение параметров, характеризующих прохождение воздушного потока через пористые материалы типа ткани.
 - 11) Определение направления и скорости движения цели в системах наведения.
 - 12) Определение траектории спутника и свойств верхних слоев атмосферы по данным о предыстории его движения (т. е. об орбите спутника и отклонениях от орбиты).
 - 13) Проектирование систем автоматической посадки самолетов, систем заправки в воздухе и других специализированных вычислительных устройств, использующих прогнозирование.
- Б. Отрасли промышленности (химическая промышленность, производство топлива, материалов и т. д.).**
- 1) Изучение кинетики химических реакций и других параметров, определяющих скорость реакции в дискретных или непрерывных процессах.

- 2) Определение времени полураспада компонент или состава радиоактивной смеси.
- 3) Определение динамики текучести материалов, подвергшихся деформации или жидкому формированию.
- 4) Определение конечного числа коэффициентов системы обыкновенных дифференциальных уравнений, которой аппроксимируются системы с распределенными параметрами, например теплообменник.
- 5) Изучение свойств процессов горения.
- 6) Изучение процессов размножения нейтронов в атомных реакторах.
- 7) Определение параметров, характеризующих динамические свойства различных инструментов, преобразователей и датчиков.

В. Другие применения.

- 1) Определение параметров динамики роста биологических популяций.
- 2) Изучение динамических характеристик систем, поведение которых нужно прогнозировать, например динамических экономических систем (распродажа какого-либо товара, развитие новой отрасли промышленности и т. д.).

Подводя итоги, можно сказать, что оценивание параметров и состояний можно рассматривать как развитие обычной теории измерений применительно к более сложным ситуациям и задачам, которое стало возможным благодаря созданию современных вычислительных машин.

1.7. ДАЛЬНЕЙШИЕ ПЕРСПЕКТИВЫ

Многие теоремы в физике и технике получены в расчете на некоторое предельное поведение или состояние системы. Приведем несколько примеров:

- 1) термодинамика: максимальный коэффициент полезного действия данной машины определяется соответствующим циклом Карно;
- 2) квантовая механика: соотношение неопределенности Гейзенберга указывает пределы точности при одновременном измерении координат и скорости;

3) теория связи: формулы Шеннона определяют максимальную пропускную способность канала связи.

В задачах оценивания также имеется необходимость в подобных предельных теоремах. Было бы не плохо определить максимальное количество информации, которое можно получить о параметрах и состояниях в какой-либо конкретной ситуации. В этом случае различные методы или схемы оценивания можно было бы сравнивать, как сравниваются различные способы передачи данных в рамках теории информации. Действительно, имеется довольно тесная связь между оцениванием параметров и состояний и теорией передачи информации. Оцениваемые величины представляют собой сообщения из некоторого источника; передатчик, канал и приемник соответствуют различным частям схемы получения и обработки информации. Шумы (в канале) соответствуют ошибкам измерений. Однако, по-видимому, пока еще нельзя говорить о наличии законченной теории, хотя имеются отдельные постановки задач и получены некоторые результаты (см., например, [2, 13, 14, 18, 19, 28]). Если попытаться описать изменение основных вероятностных характеристик в зависимости от времени оценивания, то наличие нелинейностей может привести к серьезным трудностям.

В настоящее время наиболее сильной предельной теоремой является неравенство Крамера — Рао (см. гл. 11), определяющее максимальную точность, которая может быть получена при некоторых конкретных условиях. Однако это неравенство не дает возможности учесть ту априорную информацию о параметрах, которая имеется до начала измерений и обработки данных.

1.8. ЗАКЛЮЧЕНИЕ

В этой главе показано, насколько сложен предмет идентификации (оценивания параметров) систем. Эта сложность обусловлена многообразием приложений, целей, условий и т. п. Однако в практических ситуациях обычно принимают во внимание следующие требования [11]:

- 1) точная формулировка конечной цели идентификации (оценивания) параметров;
- 2) отбор существенных признаков;

3) анализ достижимой точности.

Эти положения в сочетании с современным состоянием развития предмета приводят к тому, что до сих пор в каждом приложении еще проявляются элементы искусства и науки.

ЛИТЕРАТУРА

1. Aoki M., Optimization of Stochastic Systems, Academic Press, N.Y., 1967; есть русский перевод: Аоки М., Оптимизация стохастических систем, изд-во «Наука», 1971.
2. Barnes J., Information Theoretic Aspects of Feedback Control Systems, *Automatica*, 4, 165—185 (1968).
3. Bellman R., Åström K. J., On Structural Identifiability, *Math. Biosci.*, 1, 329—339 (1969).
4. Box G. E. P., Jenkins G. M., Discrete Models for Forecasting and Control, in: Encyclopaedia of Linguistics, Information and Control, Meetham A. R., ed., Pergamon Press, Oxford, 1969, pp. 162—168.
5. Brockett R. W., Mesarovic M. D., The Reproducibility of Multivariable Control Systems, *JMAA*, 11, 548—563 (1965).
6. Clymer A. B., System Synthesis by Dynamic Implicit Computation, Meeting Midwestern and Central Simulation Councils, St. Louis (Mo.), 1958.
7. Eykhoff P., Some Fundamental Aspects of Process-parameter Estimation, *IEEE Trans. Autom. Control*, AC-8, 347—357 (1963).
8. Eykhoff P., Van der Grinten P. M. E. M., Kwakernaak H., Veltman B. P. Th., Systems Modelling and Identification, Third IFAC Congress, London, 1966; есть русский перевод: Эйхhoff П., Ван-дер-Гринтен П. М. Е. М., Квакернаак Х., Вельтман Б. П., Моделирование и идентификация систем, в книге «Теория непрерывных автоматических систем и вопросы идентификации», изд-во «Наука», 1971, стр. 241—264.
9. Фельdbaum A. A., Теория дуального управления, *Автоматика и телемеханика*, № 9, 11 (1960); № 1, 2 (1961).
10. Фельdbaum A. A., Основы теории оптимальных автоматических систем, 2-е изд., изд-во «Наука», 1966.
11. Herles V., Limits of Precision in Identification Procedures, Second IFAC Symp. «Identification and Process Parameter Estimation», Prague, Paper 9.8, 1970.
12. Kerr R. B., Statistical, Structural, and Regression Models in Linear Process Identification, Joint Autom. Control Conf., 1965, pp. 458—466.
13. Kullback S., Information Theory and Statistics, Wiley, N.Y., 1959; есть русский перевод: Кульбак С., Теория информации и статистика, изд-во «Мир», 1967.
14. Lindley D. V. On a Measure of the Information Provided by an Experiment, *Ann. Math. Stat.*, 27, 986—1005 (1956).
15. Peschon J. et al., Case Studies in System Control, Joint Autom. Control Conf., Univ. of Michigan; Published by IEEE, N.Y., 1968.

16. Potts T. F., Ornstein G. N., Clymer A. B., The Automatic Determination of Human and Other System Parameters, Proc. West Joint Computer Conf., Los Angeles, 1961, pp. 645—660.
17. Пугачев В. С., Статистические методы в автоматическом управлении, в книге «Оптимальные системы. Статистические методы», Труды II Межд. конгресса ИФАК, Базель, 1963, изд-во «Наука», 1965, стр. 286—311.
18. Райбман Н. С. (см. литературу к гл. 2).
19. Rényi A., On the Amount of Missing Information and the Neyman-Pearson Lemma, in: Research Papers in Statistics, David F. N., ed., Festschrift for J. Neyman, Wiley, London, 1966.
20. Rosenblueth A., Wiener N., The Role of Models in Science, *Philosophy of Science*, 12, 316—321 (1945).
21. Sain M. K., Massey J. L., Invertibility of Linear, Time-Invariant Dynamical Systems, *IEEE Trans. Autom. Control*, AC-14, 141—149 (1969).
22. Silverman L. M., Inversion of Multivariable Linear Systems, *IEEE Trans. Autom. Control*, AC-14, 270—276 (1969).
23. Цыпкин Я. З., Адаптация, обучение и самообучение в автоматических системах, *Автоматика и телемеханика*, № 1, 23—61 (1966).
24. Tsypkin Ya. Z., Optimization, Adaptation, and Learning in Automatic Systems, *Computer and Inform. Sciences*, 2, 15—32 (1967).
25. Tsypkin Ya. Z., Self-learning — What is it?, *IEEE Trans. Autom. Control*, AC-13, 608—612 (1968).
26. Цыпкин Я. З., Обобщенные алгоритмы обучения, *Автоматика и телемеханика*, № 1, 97—104 (1970).
27. Weidemann H. L., Entropy Analysis of Feedback Control Systems, in: Leondes C. T., Advances in Control Systems, Academic Press, N.Y., 1969, pp. 225—255.
28. Weidemann H. L., Stear E. B., Entropy Analysis of Estimating Systems, *IEEE Trans. Inform. Theory*, IT-16, 264—270 (1970).
29. Zaborszky J., An Information Theory Viewpoint for the General Identification Problem (Correspondence), *IEEE Trans. Autom. Control*, AC-11, 130—131 (1966).
30. Zadeh L. A., From Circuit Theory to System Theory, *Proc. IRE*, 50, 856—865 (1962).

ГЛАВА 2

ОСНОВНЫЕ ПОДХОДЫ: СТАТИСТИЧЕСКИЙ И ИНЖЕНЕРНЫЙ

В этой главе дается обзор задач оценивания параметров, основная цель которого отметить сходство и особенности различных методов оценивания. Эта глава дает читателю ясное представление о структуре книги. Все изложенные здесь вопросы в последующих главах получат более полное и глубокое развитие. Поэтому не стоит беспокоиться, если какие-либо обозначения или рассуждения покажутся не слишком понятными. В то же время имеет смысл время от времени возвращаться к этой главе, чтобы получить полное представление о предмете.

Некоторые общие вопросы теории оценивания, включая свойства входных сигналов, априорную информацию об объекте и выбор схемы оценивания, рассмотрены в разд. 2.1. В разд. 2.2 отмечается наличие двух типов реализации процедур оценивания.

В разд. 2.3 описываются статистические методы, которые приводят к явным выражениям для оценок параметров (средние значения, дисперсия или доверительный интервал).

Разд. 2.4 посвящен методам оценивания по настраиваемой модели, использование которых связано с вычислением частных производных критерия ошибки по настраиваемым параметрам. Приводятся некоторые способы вычисления этих производных. Особенно интересен случай, когда ошибка является линейной функцией или линейным функционалом настраиваемых параметров. Это приводит к схеме обобщенной модели.

2.1. ОБЩИЕ АСПЕКТЫ ЗАДАЧИ

Входные сигналы объекта

Схема оценивания частично определяется видом имеющихся сигналов. Сигналы могут быть непрерывными во времени (невыборочными) и выборочными (импульсными). С помощью соответствующих приемов непрерывный сигнал может быть преобразован в импульсный и наоборот, если частота квантования f_s удовлетворяет условию Уиттакера — Шеннона: $f_s > 2f_h$, где f_h — максимальная частота ограниченного спектра непрерывного сигнала. При выполнении этого условия квантование не уменьшает содержащееся в сигнале количество информации. Различные характеристики сигналов и способы их описания рассматриваются в гл. 3. Список обозначений приведен в приложении А.

В тех случаях, когда допускается подача дополнительного пробного сигнала, свобода в его выборе является очень важной характеристикой. В этом случае необходимо изучать такие свойства сигнала, как ширина полосы, максимальная энергия и амплитуда и, кроме того, особенности генерирования и обработки сигнала (см. гл. 10).

Априорная информация об объекте

В гл. 1 отмечалось значение априорной информации об объекте. Эта информация отражает сведения о структуре объекта и о возможных значениях параметров физической или математической (априорной) модели. Наличие априорной информации может на порядок изменить сложность реализации модели и соответствующие затраты. То же самое относится к выбору между линейными и нелинейными моделями. Какое-то представление о *скорости изменения параметров* является совершенно необходимым, так как оно может определять реализацию метода.

Хотя теоретически идея черного ящика представляется привлекательной, она нередко приводит к очень продолжительным процедурам, так как пренебрегает частью имеющейся информации. На практике во многих ситуациях, кроме выходного сигнала объекта, можно измерять

и другие его характеристики. Неиспользование этих сигналов неизбежно ведет к потере информации (если только не известно, что между этими сигналами и выходом существует детерминированная связь, например, эти сигналы являются результатом пропускания выходного сигнала через инерционное звено с постоянной времени τ без помех). Возможность использования таких дополнительных «выходов» следует всегда иметь в виду. Исходя из априорной информации об объекте и цели оценивания, обычно ясно, что требовать от модели. Лучше всего, чтобы коэффициенты модели непосредственно давали оценки определяемых параметров объекта. Существует много различных типов моделей, наиболее привычны дифференциальные уравнения, передаточные функции, разложения в ряд во временной или частотной областях. Они обсуждаются в гл. 4. Выбор типа модели тесно связан с решаемой задачей.

Схема оценивания

Для любой модели требования к реализации процедуры оценивания состоят в том, чтобы:

- 1) оценки были точными, на их качество мало влияли аддитивные помехи и неучтенные нелинейности;
- 2) оценки получались быстро, что связано с требованием устойчивости системы;
- 3) оценки были экономически обоснованы, т. е. необходимо использовать доступное оборудование по сходным ценам.

Отсюда вытекают следующие требования к схеме оценивания, которая должна быть:

- 1) формализуемой;
- 2) легко реализуемой;
- 3) достаточно общего вида;
- 4) приводить к оптимальным оценкам;
- 5) обеспечивать приемлемую скорость сходимости.

Из этого перечня, а его можно продолжить, ясно, что выбор схемы оценивания определяется многими соображениями. Некоторые вопросы качества и сходимости оценок обсуждаются в гл. 5. Необходимые сведения из теории вероятностей читатель найдет в приложении Б.

2.2. ТИПЫ РЕАЛИЗАЦИИ

Решения задач теории оценивания можно дифференцировать по способу реализации процедур. В принципе можно выделить два основных способа реализации [8]:

1. Моделирование математических соотношений, выражающееся в задании набора числовых параметров (коэффициентов дифференциальных уравнений; частот и коэффициентов затухания полюсов и нулей передаточной функции; набора точек графика весовой функции). Этот тип реализации будем называть *реализацией, использующей явные математические выражения*.

2. Моделирование структуры (физической или математической) модели, параметры которой меняют так, чтобы характеристики модели были в каком-либо смысле близки к характеристикам исследуемой системы. В этом случае будем говорить о *реализациях по настраиваемой модели*.

На фиг. 2.1, *a*, *б* изображены блок-схемы двух типов реализации процедур оценивания, технические примеры приведены на фиг. 2.1, *в*, *г*.

Формально различие между указанными способами реализации состоит в следующем. Допустим, что выход объекта задается уравнением

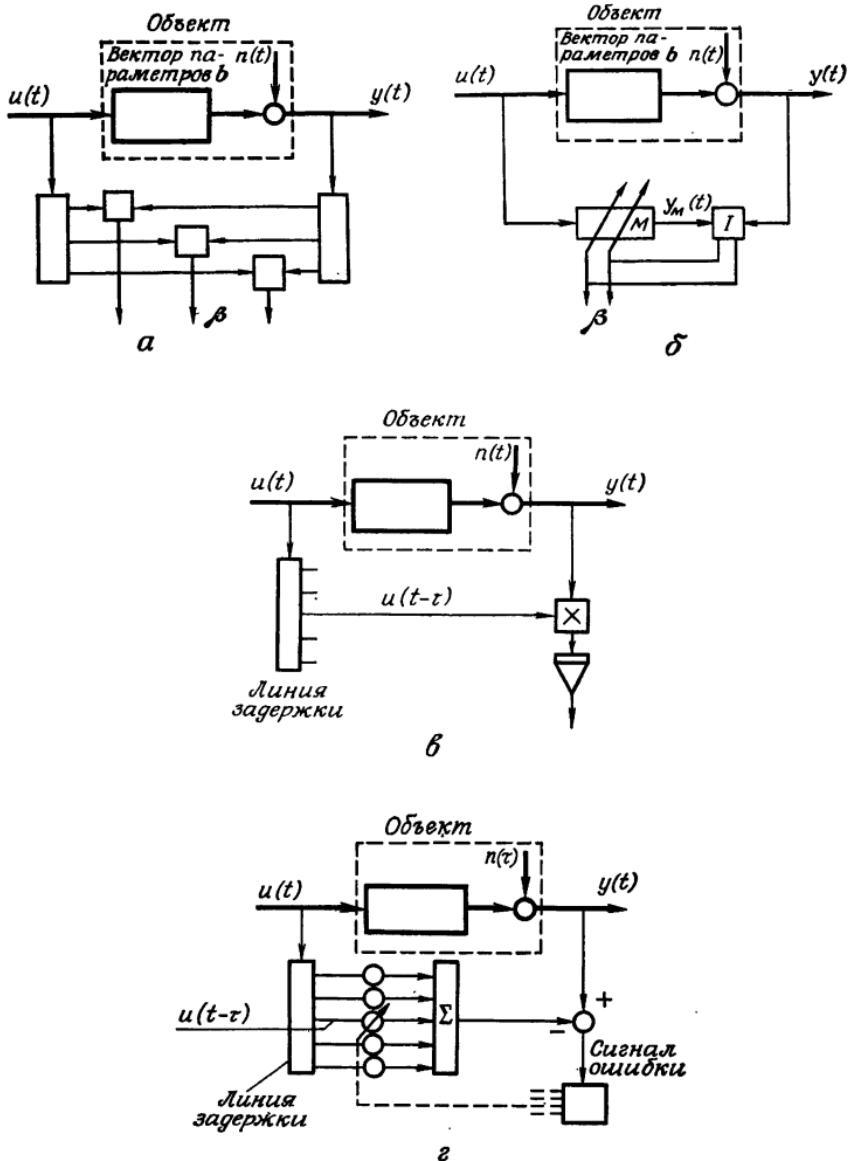
$$y(t) = f(u, b, n), \quad (2.1a)$$

а выход модели y_M описывается уравнением

$$y_M(t) = f(u, \beta, 0), \quad (2.1b)$$

где b — $(m + 1)$ -мерный вектор параметров объекта;
 β — $(m + 1)$ -мерный вектор оценок;
 $u = u(t)$ — входной сигнал объекта и модели;
 $n = n(t)$ — помехи (возмущения);
 f — функциональная связь между входной и выходной переменными.

Это означает, что объект и его модель предполагаются подобными по структуре и что помехи не меняют вида модели.



Ф и г. 2.1.

а – реализация, использующая явные математические выражения; **б** – реализация по настраиваемой модели; **в, г** – связь между двумя типами реализаций на примере исследования объекта с задержкой; **в** – корреляция; **г** – настройка модели линии задержки.

Допустим, например, что мера близости объекта и модели определяется функционалом ошибки

$$E\{y, y_M; \beta\} = \int_{t-T}^t \{y(t) - y_M(t; \beta)\}^2 dt. \quad (2.2)$$

Теперь можно поступить двояко.

1) Положить

$$\frac{\partial E}{\partial \beta_i} = 0 \quad \text{при } i = 0, 1, \dots, m.$$

Это необходимое условие достижения минимальной ошибки. Получаем систему из $m + 1$ уравнений с $m + 1$ неизвестными оценками (β_0, \dots, β_m), которую можно решить относительно β , использовав тем самым *реализацию первого типа* (подразумевается, что условие положительной определенности гессиана — матрицы из вторых производных $\|\partial^2 E / \partial \beta_i \partial \beta_j\|_{i,j=0}^m$ — выполнено).

2) Положить

$$\frac{\partial E}{\partial \beta_i} \rightarrow 0 \quad \text{при } i = 0, 1, \dots, m.$$

Стремления производных к нулю можно добиться, если имеются технические возможности их измерения и если эти производные используются для *настройки физической модели*.

Указанные типы реализации процедур оценивания получили несколько разных наименований:

Тип I

- 1) с использованием явных математических выражений;
- 2) явные методы;
- 3) методы по разомкнутому контуру;
- 4) прямые методы;
- 5) с накоплением;
- 6) с выдачей окончательного результата.

Здесь решение

- 1) получается в результате выполнения *конечного* числа элементарных операций;

Тип II

- 1) с настройкой модели;
- 2) неявные методы;
- 3) методы по замкнутому контуру;
- 4) итеративные методы.

Здесь решение

- 1) получается после (в принципе) выполнения бесконечного числа элементарных операций;

Продолжение

Тип II

- | | |
|--|--|
| Тип I | Тип II |
| 2) требует значительной памяти;
3) не может быть получено приближенно как результат промежуточных вычислений;
4) по отношению к оценке получается в разомкнутом контуре; | 2) требует меньшей памяти;
3) может быть получено приближенно как результат промежуточных вычислений;
4) по отношению к оценке получается в замкнутом контуре. |

Более подробное изложение этих вопросов содержится

- | | |
|---|---|
| в гл. 6 для импульсных сигналов;
в гл. 8 для непрерывных сигналов. | в гл. 7 для импульсных сигналов;
в гл. 9 для непрерывных сигналов. |
|---|---|

С инженерной точки зрения эти отличия имеют важное значение. Число операций и требуемая память зависят от выбора технической реализации решения: цифровой, аналоговой или гибридной. Доступность результатов промежуточных вычислений для некоторых приложений может быть существенной, если, например, поступают новые измерения, которые нужно использовать в реальном масштабе времени для уточнения оценок (отслеживание); такие промежуточные результаты важны, в частности, для целей управления. Таким образом, деление на методы *идентификации вне контура регулирования* и методы *идентификации в замкнутом контуре* обосновано. Метод идентификации по настраиваемой модели делает возможным использование априорной информации об объекте (его структуре, начальных оценках параметров). Точность получаемых решений для реализаций типа I сильнее зависит от качества отдельных элементов схемы, чем для реализаций типа II, так как в реализациях типа II сказываются корректирующие свойства обратной связи.

Схемы, изображенные на фиг. 2.1, являются достаточно общими. Они относятся к определенному описанию сигналов (гл. 3), но могут быть использованы для разных моделей объекта, в том числе для весовых функций, передаточных функций, ортогональных разложений и для моделей в пространстве состояний. В гл. 4 можно найти обширную библиографию, посвященную этим моделям

2.3. РАЗЛИЧНЫЕ СТАТИСТИЧЕСКИЕ ПОДХОДЫ

Среди вопросов, изучаемых в математической статистике, имеются следующие:

1) Сжатие информации, представленной большим набором данных, до нескольких числовых значений (*редукция данных*).

2) Извлечение выборки из большой совокупности с последующими статистическими выводами о свойствах совокупности по выборке. Эта процедура может преследовать следующие цели: а) проверку какого-либо вероятностного суждения, например гипотезы о том, что рассматриваемая совокупность имеет распределение вероятностей, принадлежащее определенному классу (*проверка гипотез*); б) нахождение оценки и дисперсии оценки неизвестного параметра функции распределения вероятности для данной совокупности (*оценивание параметров распределения*).

3) Анализ стохастической связи между двумя характеристиками, например отыскание числовых значений оценок и дисперсий оценок коэффициентов предполагаемой связи между величинами u и y (*оценивание параметров связи*).

Задачу оценивания параметров объекта можно рассматривать как исследование параметров распределения или параметров связи. Последнее естественно, когда u и y представляют входной и выходной сигналы объекта. Интерпретация оценивания параметров объекта как оценивания параметров распределения связана с использованием метода максимального правдоподобия.

Ожидаемые результаты оценивания могут быть самыми разными в зависимости от характера прикладной задачи и доступной априорной информации. Рассмотрим описание объекта в пространстве состояний¹⁾:

$$\begin{aligned}\dot{x} &= f(x, u, b), \\ y &= g(x, u, b, n),\end{aligned}\tag{2.3a}$$

¹⁾ Читатель, не знакомый с описанием объектов в пространстве состояний, может обратиться к обзору в гл. 4.

где

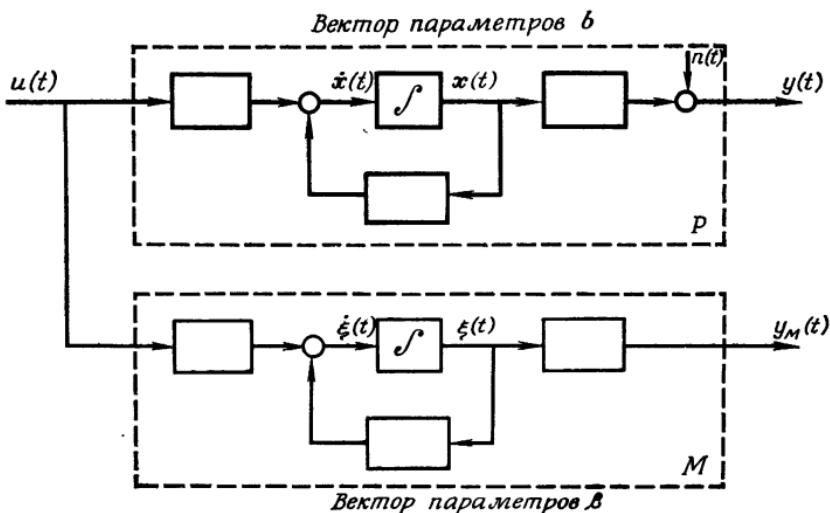
- $u = u(t)$ — входной сигнал,
- $y = y(t)$ — выходной сигнал,
- $n = n(t)$ — помеха,
- $\mathbf{x} = \mathbf{x}(t)$ — вектор состояния объекта,
- \mathbf{b} — вектор параметров объекта,

$$\mathbf{x}(t) = \begin{bmatrix} x_1(t) \\ \vdots \\ x_n(t) \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} b_0 \\ \vdots \\ b_m \end{bmatrix}.$$

Уравнения модели имеют тот же вид:

$$\begin{aligned} \dot{\xi} &= \mathbf{f}(\xi, u, \beta), \\ y_M(t) &= g(\xi, u, \beta, 0), \end{aligned} \tag{2.36}$$

где $\xi = \xi(t)$ — вектор состояния модели, β — вектор параметров модели. На фиг. 2.2 представлен упрощенный



Ф и г. 2.2.

линейный случай. Теперь задачу оценивания параметров можно сформулировать как задачу минимизации:

Минимизируемая величина	Параметр	Состояние
Функция (функционал) от измеримых величин, а именно от разности выходных сигналов объекта и модели	$E \{y(t; b) - y_M(t; \beta)\}$	$E \{y(x) - y_M(\xi)\}$
Математическое ожидание $E[\cdot]^1)$ функции (функционала) от неизмеримых величин, а именно параметров и (или) состояния объекта	$E[b - \beta]$	$E[x - \xi]$

1) См. приложение В.

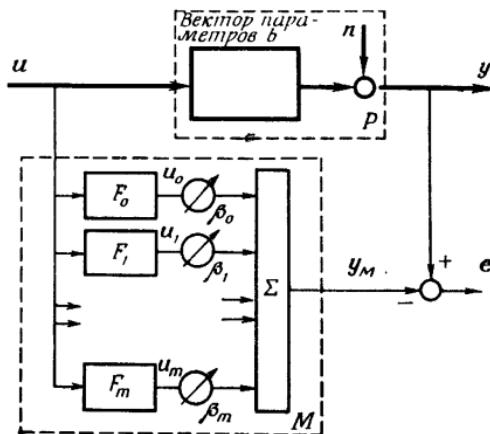
Некоторая функция ¹⁾ (функционал) от $e = y - y_M$, выходной ошибки, может быть минимизирована, так как выбором технической реализации или программированием модели на ЭВМ ошибка e может быть сделана измеримой.

В зависимости от модели и входного сигнала эта ошибка определяет некоторую меру соответствия между векторами параметров и (или) состояний объекта и модели. В некоторых случаях соответствие между входом и выходом важнее соответствия параметров, и в частности, если модель проще, например более низкого порядка, чем объект. Очевидно, что величины b и x нельзя непосредственно измерить. Таким образом, можно минимизировать только математические ожидания $E\{b - \beta\}$ или $E\{x - \xi\}$. Од-

1) Различие между функцией и функционалом можно проиллюстрировать простым примером:

функция $y(t) = a_1x(t) + a_2x^2(t) + a_3x^3(t)$, ее значение в момент t зависит только от $x(t)$;

функционал $y(t) = \int_0^t x(t - \theta) h(\theta) d\theta$, его значение в момент t зависит от x на всем интервале $(0, t)$.



Ф и г. 2.3.

нако это возможно только тогда, когда имеется какая-то априорная информация о распределениях вероятностей.

Как и во множестве других ситуаций, мы стремимся найти оптимальную или наилучшую процедуру оценивания. Однако определение наилучшей оценки зависит от вида прикладной задачи, требований, предъявляемых к оценке, априорной информации об объекте и помехах. В литературе по математической статистике описан ряд различных процедур оценивания. Эти методы отличаются прежде всего по используемому критерию оптимальности и имеющейся априорной информации. К сожалению, выбор критерия оптимальности более или менее субъективен, а получающиеся алгоритмы существенно зависят от сделанного выбора. Все эти процедуры оценивания и их взаимосвязь рассматриваются в гл. 5. Здесь приводится лишь краткое предварительное обсуждение. Следует еще раз подчеркнуть, что после первого чтения этой главы у читателя должно создаться лишь самое общее представление о методах оценивания и связанных с ними алгоритмах.

Один из простых подходов может быть иллюстрирован блок-схемой, представленной на фиг. 2.3. Рассматривается случай выборочных измерений функций непрерывного времени. Символами u , u_i , y , n и y_M обозначены векторы

соответствующих выборок, например¹⁾

$$\mathbf{y}' = [y(T_s), y(2T_s), \dots, y(kT_s)]. \quad (2.4)$$

Символами \mathbf{b} и β обозначены соответственно векторы параметров объекта и модели:

$$\mathbf{b}' = [b_0, b_1, \dots, b_m],$$

$$\beta' = [\beta_0, \beta_1, \dots, \beta_m].$$

Отсюда находим

$$\mathbf{y} = \mathbf{y}(\mathbf{u}, \mathbf{b}, \mathbf{n}), \quad (2.5a)$$

$$\mathbf{y}_M = \mathbf{U}\beta, \quad (2.5b)$$

где

$$\mathbf{U} = [\mathbf{u}_0, \mathbf{u}_1, \dots, \mathbf{u}_m] = \begin{bmatrix} u_0(1) & \dots & u_m(1) \\ \vdots & & \vdots \\ u_0(k) & \dots & u_m(k) \end{bmatrix} \quad (2.6)$$

Столбцы \mathbf{u}_i матрицы \mathbf{U} состоят из выборочных значений выходных сигналов блоков F_i . Эти блоки выбраны для построения требуемого описания объекта (см. гл. 4).

Теперь можно поставить задачу как отыскание оценки β вектора параметров объекта \mathbf{b} по наблюдениям \mathbf{y} . Основываясь на имеющейся априорной информации, можно выделить следующие виды оценок:

Оценивание по методу наименьших квадратов

В этом случае не требуется никакой априорной информации. Критерий ошибки определяется как

$$E = \mathbf{e}' \mathbf{I} \mathbf{e} = \sum_{j=1}^k e^2(j), \quad (2.7)$$

где

$$\mathbf{e} = \mathbf{y} - \mathbf{y}_M. \quad (2.8)$$

Здесь \mathbf{I} — единичная матрица. Минимизируя критерий ошибки, получаем так называемую систему нормальных

¹⁾ \mathbf{y}' означает вектор, транспонированный по отношению к \mathbf{y} . В дальнейшем предполагается, что выборочные измерения осуществляются с постоянным шагом T_s , а символ T_s будет опускаться.

уравнений для оценивания¹⁾ параметров:

$$\hat{\beta} = [U'U]^{-1}U'y \quad (\text{оценивание методом наименьших квадратов}). \quad (2.9)$$

Здесь, как и раньше, U' означает транспонированную матрицу U .

Марковские оценки или оценивание по обобщенному методу наименьших квадратов

Предположим теперь, что известна ковариационная матрица аддитивного шума

$$N = E[nn'] = \begin{bmatrix} E[n(1)n(1)] & \dots & E[n(1)n(k)] \\ \vdots & & \vdots \\ E[n(k)n(1)] & \dots & E[n(k)n(k)] \end{bmatrix}. \quad (2.10)$$

Можно доказать, что при $N \neq I$ наилучшая линейная оценка получается минимизацией выражения

$$E = e'N^{-1}e, \quad (2.11)$$

что приводит к следующему правилу оценивания:

$$\hat{\beta} = [U'N^{-1}U]^{-1}U'N^{-1}y \quad (2.12)$$

(марковская оценка или оценивание по обобщенному методу наименьших квадратов).

Оценивание методом максимального правдоподобия

Как уже отмечалось, задачу оценивания можно рассматривать как исследование параметров распределения вероятностей. Рассмотрим вектор выхода y . *Априори* (до измерений) известно, что выборочные значения y являются случайными величинами с совместным распределением вероятностей

$$p\{y(1), \dots, y(k); b\}, \quad (2.13)$$

¹⁾ Под *оцениванием* или *оператором оценивания* будем понимать правило получения оценок, а под *оценками* — числовые значения, к которым приводит применение правила оценивания в конкретном случае.

зависящим от вектора параметров \mathbf{b} (см. приложение В). *Апостериори* (после измерений) становятся известными реализации этих случайных величин: $y(1) = c_1, \dots, y(k) = c_k$. По ним определим оценку $\hat{\beta}$. Функциональная связь между $\mathbf{c}' = \{c_1, \dots, c_k\}$ и $\hat{\beta}$ определяется формулой (2.13). Для того чтобы подчеркнуть, что формула (2.13) связывает реализацию выборочных значений с вектором оценок параметров $\hat{\beta}$, перепишем ее в виде

$$L \{c_1, \dots, c_k; \hat{\beta}\}. \quad (2.14)$$

Эта функция называется *функцией правдоподобия*. Так как выборочные значения определяются в результате измерений, L является функцией только $\hat{\beta}$. Удобнее рассматривать $\ln L$; вследствие монотонности логарифма максимумы L и $\ln L$ достигаются при одном и том же значении $\hat{\beta}$. Это значение $\hat{\beta}$ можно найти из уравнения

$$\nabla_{\hat{\beta}} \ln L \{c; \hat{\beta}\} |_{\hat{\beta}=\hat{\beta}} = \frac{\partial}{\partial \hat{\beta}} \ln L \{c; \hat{\beta}\} |_{\hat{\beta}=\hat{\beta}} = 0. \quad (2.15)$$

Решение этого уравнения $\hat{\beta}$ называется оценкой максимального правдоподобия вектора \mathbf{b} , если $\hat{\beta}$ действительно зависит от выборочных значений c_1, \dots, c_k . В этом случае необходимо знать не только ковариационную матрицу помехи, но и вид совместного распределения вероятностей ее выборочных значений.

Оценивание по минимальному среднему риску

В этом случае необходима еще большая априорная информация. Как показано в гл. 5, должны быть заданы плотности распределения вероятностей u , p и b .

С теоретической и практической точек зрения интересно рассмотреть вопрос о том, какая априорная информация будет использоваться на практике. Вообще говоря, дополнительная априорная информация приводит к улучшению оценки, однако затраты на реализацию могут оказаться чрезмерными. Какие бы методы оценивания ни использовались, они должны опираться на случайные наблюдения, случайность которых связана с наличием помех. Следовательно, любая оценка, являющаяся функцией этих случайных наблюдений, также является случайной величиной.

2.4. МЕТОДЫ ОЦЕНИВАНИЯ ПО НАСТРАИВАЕМОЙ МОДЕЛИ

Нечетные и четные функции ошибок

Схема, изображенная на фиг. 2.1, б, может быть использована для такой настройки модели M , чтобы ее передаточные характеристики в некотором смысле приближались к передаточным характеристикам объекта P . Центральной частью фиг. 2.1, б является блок, помеченный символом I . В этом блоке обрабатывается поступающая информация. *Ошибка* можно определить как $e(t) = y(t) — y_m(t)$, а *критерий ошибки* — как минимизацию любого четного функционала E от этой ошибки, например

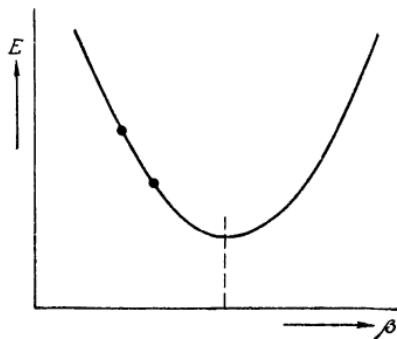
$$E = \int_{t_1}^{t_2} e^2(t, b, \beta) dt, \quad (2.16)$$

где, как и раньше, b — параметр рассматриваемого объекта, а β — соответствующий параметр модели. Выбор четного функционала приводит к такой зависимости между E и значением настраиваемого параметра β , которая показана на фиг. 2.4. Таким образом, мы подходим к различию между так называемыми управлениями по «нечетным» и по «четным» функциям [7]. Рассмотрим глаз человека как пример, иллюстрирующий оба механизма управления (фиг. 2.5, а).

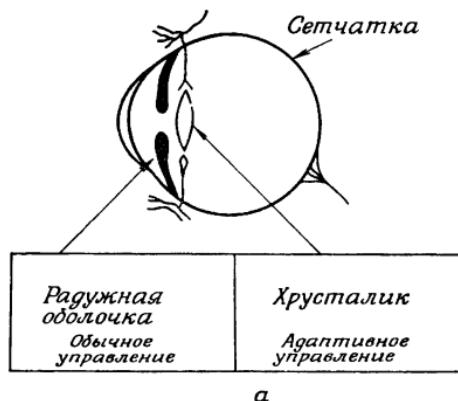
Радужная оболочка управляет уровнем освещенности сетчатки как самая обычная система управления: если уровень освещенности слишком высок, сигнал ошибки приводит к сужению зрачка, и наоборот. *Знак ошибки* непосредственно указывает направление изменения размеров зрачка (фиг. 2.5, б). Сигнал ошибки сводится к нулю.

Хрусталик управляет фокусировкой изображения. Здесь сигнал ошибки уже не меняет знака при перестройке фокусного расстояния хрусталика от наблюдения удаленных предметов для наблюдения близких предметов и наоборот. Четкость изображения как функция аккомодации хрусталика и, следовательно, сигнала ошибки имеет экстремум (фиг. 2.5, в)¹). Мы сталкиваемся с ситуацией,

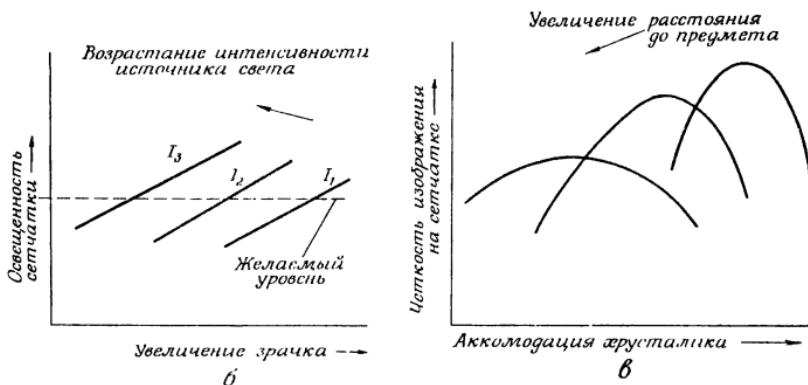
¹⁾ Для управления фокусным расстоянием хрусталика используется несколько механизмов.



Ф и г. 2.4.



a



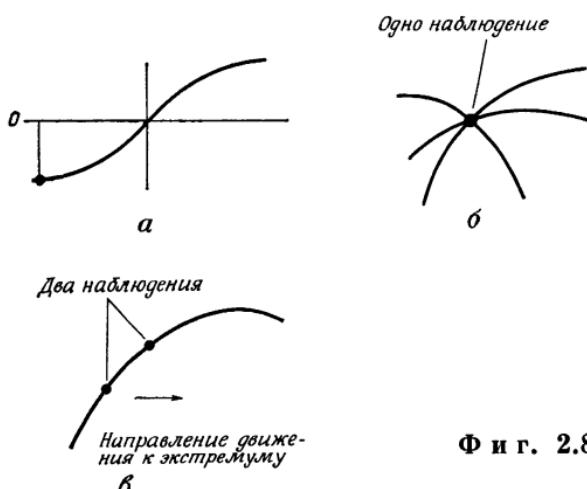
Ф и г. 2.5.

Оценивание параметров Мы хотим знать По наблюдениям мы хотим уточнить	<p><i>Адаптация</i></p> <p>Наблюда за динамической системой, мы хотим уточнить</p> <p>Помехи</p> <p>Линейка объекта</p> <p>?</p> <p>Показатель эффективности и т. п.</p> <p>Optimum</p> <p>Наблюдения</p>
	<p>Помехи</p> <p>Динамика объекта</p> <p>?</p> <p>Динамика объекта</p> <p>?</p>

Фиг. 2.6.

<i>I</i> <i>Обычное управление с обратной связью</i>	<i>II</i> <i>Адаптивное и экспрессимальное регулирование</i>
<i>Обычная система</i>	<i>Множество различных систем</i> <i>Показатель функционирования</i>
<i>Функция усиления ошибки</i>	
<i>Основное свойство</i>	<i>Первая производная не меняет знак</i>
<i>Априорная информация</i>	<i>Существует кульм функции.</i> <i>Известен знак второй производной</i>
<i>Информация, достаточная для управления с обратной связью</i>	<i>Одно наблюдение</i> <i>Минимум для нахождения (и редукция данных)</i>

Фиг. 2.7.



Ф и г. 2.8.

аналогичной схеме с настраиваемой моделью. Эта ситуация типична для адаптивных и экстремальных регуляторов. На фиг. 2.6 проводится сравнение целей оценивания параметров, оптимизации и адаптации. На фиг. 2.7 показано главное отличие обычного управления от управления экстремального типа. Будем называть эти два типа систем с обратной связью, подчеркивая их основное отличие, системами с *нечетной функцией* (ошибки) и системами с *четной функцией* (ошибки). Из фиг. 2.8 видно, что если известен знак второй производной (минимум или максимум функции), то для системы с четной функцией ошибки требуется по меньшей мере *два наблюдения*, чтобы определить по ним направление движения к экстремуму. Для того чтобы управлять, необходимо сжать информацию о двух или большем числе наблюдений, представив ее в виде одного числа. Такая редукция данных превращает четную функцию ошибки в нечетную функцию ошибки. После редукции, используя эту нечетную функцию ошибки, можно применять обычный принцип управления с обратной связью.

Направление оптимальной настройки указывается знаком первой производной по β . Для *непрерывных (по времени) сигналов*, используя выражение (2.16), находим

$$\nabla_{\beta} E = \frac{\partial E}{\partial \beta} = 2 \int_{t_1}^{t_2} e(t, b, \beta) \frac{\partial e}{\partial \beta} dt. \quad (2.17)$$

Ни $\partial E / \partial \beta$, ни $\partial e / \partial \beta = -\partial y_M / \partial \beta$ нельзя измерить непосредственно без дополнительных приспособлений. Некоторые из приемов получения желаемой информации будут рассмотрены в дальнейшем. Если удастся найти эти частные производные, то для определения наилучшего значения β можно использовать градиентные методы (см. гл. 5).

Для выборочных сигналов в том случае, когда выбранное описание адекватно представляет поведение исследуемого объекта, находим (см. фиг. 2.3)

$$\begin{aligned} y &= Ub + n \quad (\text{объект}), \\ y_M &= U\beta \quad (\text{модель}) \end{aligned} \quad (2.18a)$$

и, следовательно, ошибка определится как

$$e = y - y_M = U(b - \beta) + n. \quad (2.18b)$$

Таким образом, квадратичная форма

$$E = e'e \quad (2.19)$$

достигает минимума при $b = \beta$. Снова нужно определить $\nabla_\beta E = \partial E / \partial \beta$. Предполагая, что $\partial E / \partial \beta$ можно вычислить, приходим к задаче об отыскании способа автоматической настройки параметра β . Имеется несколько путей решения этой задачи. Настройка может быть непрерывной и дискретной.

Непрерывная схема настройки. Хорошая стратегия получается, если положить

$$\beta = \Gamma \nabla_\beta E. \quad (2.20)$$

Такой выбор приводит к градиентному методу. Γ — коэффициент усиления, от которого зависит скорость сходимости (настройки). Будет показано, что коэффициент Γ можно выбирать как постоянным, так и зависящим от времени, он может быть скалярным [Γ или $\Gamma(t)$] или матрицей [Γ или $\Gamma(t)$].

Строго говоря, при определении градиента величина β должна быть неизменной. В непрерывной схеме настройки это требование не выполняется. Поэтому уравнение (2.20) лишь приближенно описывает динамику настройки в том случае, когда настройка происходит относительно медленно.

Дискретная схема настройки. Упомянутая трудность определения градиента не возникает, если пользоваться дискретной схемой: измерение градиента при постоянном β , настройка β , новое измерение градиента и т. д. Подобная задача рассматривается в теории стохастической аппроксимации [21] и приводит к следующему алгоритму:

$$\hat{\beta}(i+1) = \hat{\beta}(i) - \Gamma(i)\nabla_{\beta} E(i), \quad (2.21)$$

где $\beta(i)$ — вектор параметров модели после i -й настройки; $E(i)$ — значение функционала на i -м наборе ошибок; $\Gamma(i)$ — коэффициент усиления, определяющий скорость сходимости алгоритма.

На простом примере снова можно продемонстрировать связь между методами с настраиваемой моделью и использованием явных математических выражений. Допустим, что критерием настройки является

$$\min e'e.$$

Из уравнения (2.9) видно, что после i -й настройки оптимальное значение параметра $\hat{\beta}$ определится как

$$\begin{aligned}\hat{\beta}(i+1) &= [U'U]^{-1}U'(y_M + e) = \\ &= [U'U]^{-1}U'U\beta(i) + [U'U]^{-1}U'e = \\ &= \beta(i) + [U'U]^{-1}U'e.\end{aligned}$$

Это оценка по методу наименьших квадратов при условии, что проведена априорная настройка модели и имеется последовательность выборочных значений ошибки. Для последующих настроек из уравнения (2.21) находим

$$\hat{\beta}(i+1) = \hat{\beta}(i) + [U'U]^{-1}U'e(i).$$

Также можно оценить степень «неопределенности» этой оценки $\text{cov}[\hat{\beta}(i)]$.

Точно так же в задачах настройки можно использовать критерий $\min e'N^{-1}e$. Сходимость получаемых процедур обсуждается в гл. 5.

Методы получения частных производных или градиентов

В принципе имеются следующие подходы:

а) Использование коэффициентов влияния параметров или коэффициентов чувствительности параметров. В последние годы этот способ вызвал большой интерес (см. гл. 9). Идею метода можно объяснить на простом примере, принадлежащем Мейссингеру [14]. Рассмотрим модель первого порядка с одним неизвестным параметром β и выходным сигналом $y_M = w$:

$$\frac{dw}{dt} + \beta w = u. \quad (2.22)$$

Ошибка $e = y - w$. Отсюда $de/d\beta = -\partial w/\partial\beta$. Так как нас интересует поведение модели на всей временной оси $-\infty < t < \infty$, начальные условия не имеют значения. Запишем формально

$w = w(t, \beta)$, тогда

$$w(t, \beta) = w(t, \beta_0) + \left. \frac{\partial w}{\partial \beta} \right|_{\beta_0} \Delta \beta + \frac{1}{2!} \left. \frac{\partial^2 w}{\partial \beta^2} \right|_{\beta_0} (\Delta \beta)^2 + \dots . \quad (2.23)$$

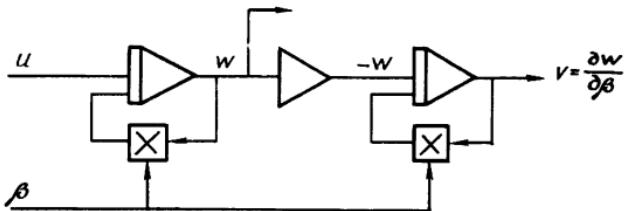
Оставим только члены первого порядка малости. Величину производной $\partial w/\partial\beta$ можно определить, продифференцировав по β дифференциальное уравнение (2.22):

$$\frac{\partial^2 w}{\partial \beta \partial t} + \beta \frac{\partial w}{\partial \beta} + w = 0. \quad (2.24)$$

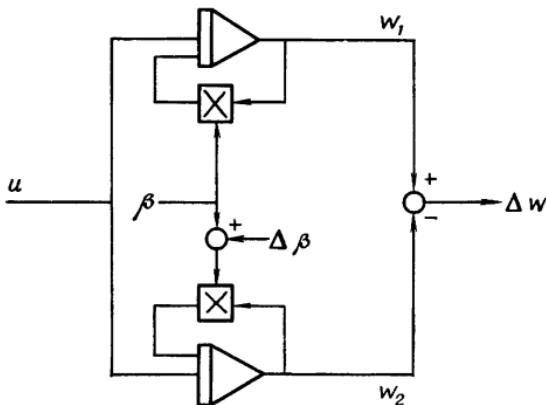
Меняя порядок дифференцирования и делая замену $v = \partial w/\partial\beta$, находим

$$\frac{\partial v}{\partial t} + \beta v = -w. \quad (2.25)$$

Это уравнение называется уравнением чувствительности по β . Для переменного β величина v является оценкой градиента $\partial w/\partial\beta$, который в этом случае зависит от времени. Уравнения (2.22) и (2.25) могут быть набраны в виде схемы генерации функции v , изображенной на фиг. 2.9. Можно также набрать схемы, соответствующие уравнениям (2.17) и (2.20) или (2.21), которые бы при некоторых



Ф и г. 2.9.



Ф и г. 2.10.

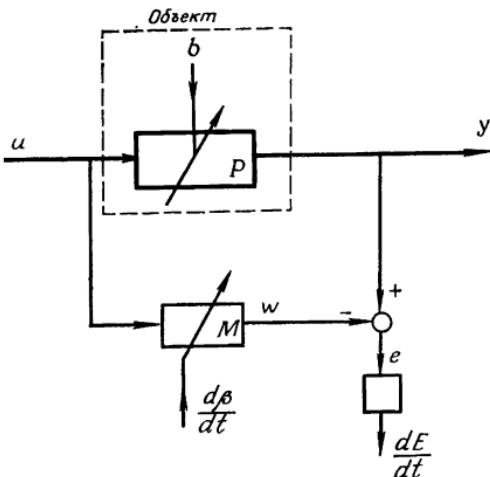
ограничениях обеспечивали сходимость выходного сигнала к параметру β . Подробно это обсуждается в гл. 9.

б) Использование двух моделей с параметрами β и $(\beta + \Delta\beta)$. Оценку первой производной по β можно найти, если использовать две модели с параметрами β и $(\beta + \Delta\beta)$ соответственно. Разность выходных сигналов этих моделей w_1 и w_2 дает величину $\Delta w / \Delta\beta = -\Delta e / \Delta\beta$, которую можно использовать как оценку $de/d\beta$ в уравнениях (2.17) и (2.20) или (2.21) (фиг. 2.10). Этот метод также рассматривается в гл. 9.

в) Использование модели с возмущаемым параметром. Этот метод основывается на соотношении

$$\frac{dE}{dt} = \frac{\partial E}{\partial \beta} \frac{d\beta}{dt} + \frac{\partial E}{\partial t}, \quad (2.26)$$

из которого следует, что если величина параметра модели определяется возмущением, заданным известной функци-



Ф и г. 2.11.

ей времени — пробным сигналом, а возникающее возмущение производной dE/dt может быть измерено, то можно получить информацию о величине $\partial E/\partial \beta$ (фиг. 2.11). Этот метод рассматривается в гл. 9.

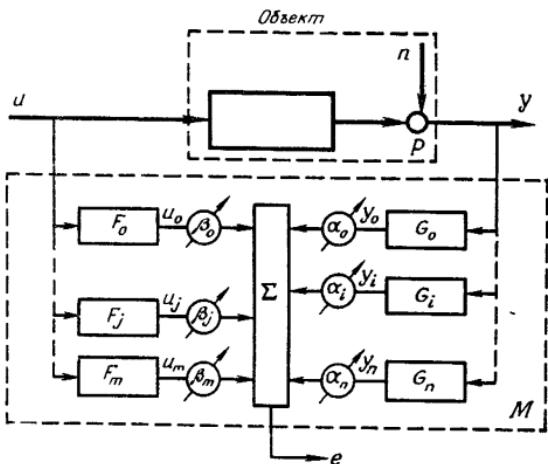
г) Использование обобщенной модели. Очевидно, что перечисленные методы вычисления $\partial E/\partial \beta$, $\partial w/\partial \beta$ или их оценок предъявляют серьезные требования к технической реализации или машинным процедурам. Эти трудности возрастают, если определению подлежит целый ряд параметров. Вообще говоря, для каждого оцениваемого параметра необходима еще одна модель или еще один возмущающий пробный сигнал. Поэтому полезно остановиться на случае, когда ошибка является линейной функцией настраиваемых параметров:

$$e = \sum_{i=0}^n \alpha_i y_i + \sum_{j=0}^m \beta_j u_j. \quad (2.27)$$

В этом случае имеем

$$\frac{\partial e}{\partial \alpha_i} = y_i, \quad \frac{\partial e}{\partial \beta_j} = u_j, \quad (2.28)$$

т. е. коэффициенты чувствительности входят в модель в явном виде. Этот случай представлен на фиг. 2.12. Входной сигнал $u(t)$ стационарен и в большинстве случа-



Ф и г. 2.12.

ев считается случайным процессом. Блоки F_0, \dots, F_m и G_0, \dots, G_n представляют собой операторы (например, передаточные функции для линейных систем, нелинейные операторы). Эти блоки вместе с «потенциометрами» $\alpha_0, \dots, \alpha_n$ и β_0, \dots, β_m и сумматором образуют обобщенную модель объекта. Потенциометры позволяют осуществлять настройку положительных и отрицательных значений коэффициентов. Выходную ошибку $e(t)$ будем называть обобщенной ошибкой.

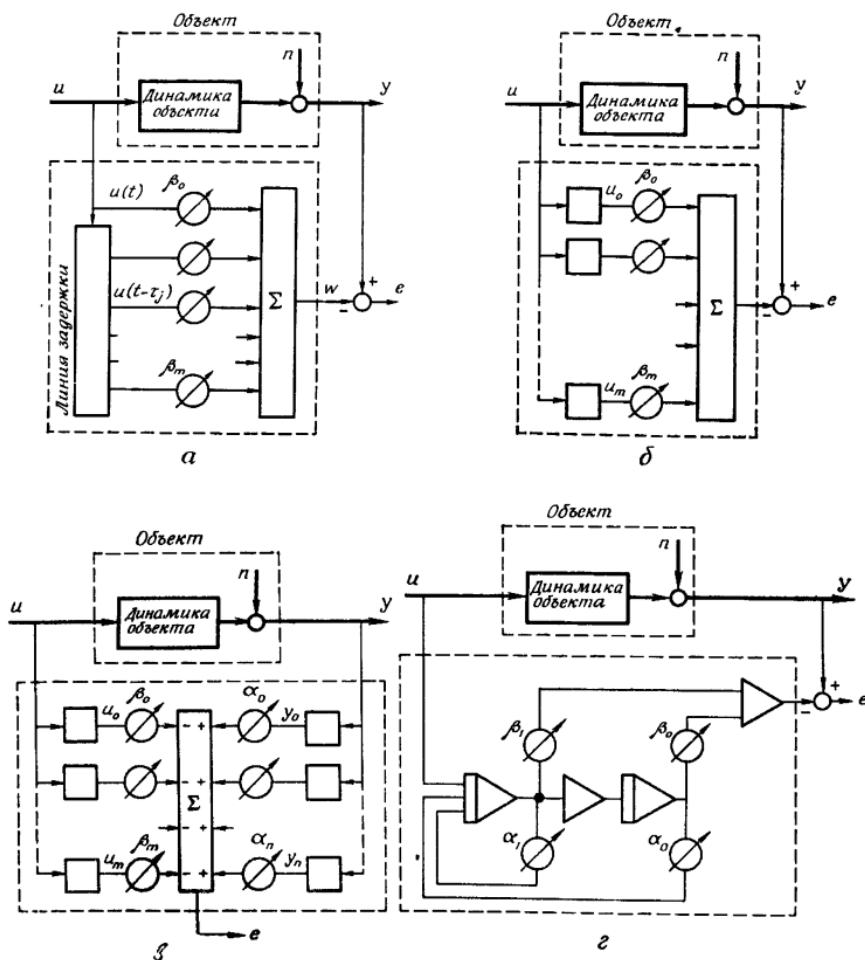
В дальнейшем читатель оценит удобство модели, которая приводит к ошибке, линейной по оцениваемым параметрам. На фиг. 2.13, а – в приводятся конкретные примеры:

$$a) e(t) = y(t) - \sum_j \beta_j u(t - \tau_j);$$

$$б) e(t) = y(t) - \sum_j \beta_j u_j(t);$$

$$в) e(t) = y(t) - \sum_i \alpha_i y_i(t) - \sum_j \beta_j u_j(t).$$

Отметим, что а) является частным случаем б), а б) — частным случаем в). Для обобщенной модели вовсе не обязательно, чтобы связь между входом и выходом модели была линейной. Уже из фиг. 2.12 видно, что операторы F_i и G_j могут быть нелинейными.



Фиг. 2.13.

На фиг. 2.13, в приводится пример системы, в которой ошибка нелинейна относительно оцениваемых параметров:

$$\text{г) } e(t) = y(t) - \mathcal{O}[u(t), \alpha, \beta].$$

Здесь $\mathcal{O}[\cdot]$ является динамическим оператором модели, а α и β — векторы параметров.

2.5. ОЦЕНИВАНИЕ ПАРАМЕТРОВ И СОСТОЯНИЙ

Как уже отмечалось в гл. 1, в ряде приложений понятие состояния играет важную роль. Это особенно справедливо для задач оптимального управления. Так как состоя-

ние объекта не поддается непосредственному измерению, а измеримые величины искажены помехами, состояние объекта может быть только оценено. Полное решение задачи оценивания получено для линейных объектов, динамические свойства которых полностью известны, включая числовые значения параметров объекта [12]. Для нелинейных объектов предложено много приближенных схем. Эти методы оценивания изложены в гл. 12.

Задача становится более сложной, если параметры объекта заранее неизвестны и должны оцениваться вместе с состояниями. Этот класс задач рассматривается в гл. 13.

Гл. 14 посвящена приложениям в различных областях науки и техники.

2.6. ЗАКЛЮЧЕНИЕ

Теперь читатель в состоянии оценить содержание всей книги. Настоятельно рекомендуем еще раз прочитать предисловие.

В периодической литературе появление статей по методам оценивания носит случайный характер. Изучая соответствующие разделы книги, читатель отметит связь различных методов оценивания. Данная глава поможет ему в этом.

С материалами обзорного характера можно познакомиться по работам [1—5, 9—11, 19, 20, 22]. Общий интерес представляют книги [6, 13, 16—18]. Обширная библиография содержится в работе [15].

ЛИТЕРАТУРА

1. Александровский Н. М., Дейч А. М., Методы определения динамических характеристик нелинейных объектов, *Автоматика и телемеханика*, № 1, 167—188 (1968).
2. Åström K. J., Eykhoff P., System Identification — a Survey, Second IFAC Symp. «Identification and Process Parameter Estimation», Prague; см. также *Automatica*, 7, 123—162 (1971).
3. Balakrishnan A. V., Peterka V., Identification in Automatic Control Systems, Fourth IFAC Congress, Warsaw, Survey Paper 9; см. также *Automatica*, 5, 817—829 (1969).
4. Bekey G. A., Identification of Dynamic Systems by Digital Computers, Joint Automatic Control Conf. (Computer Control Workshop), 1969, pp. 7—18.

5. Cuenod M., Sage A., Comparison of Some Methods Used for Process Identification IFAC Symp. «Identification in Autom. Control Systems», Prague; см. также *Automatica*, 4, 235—269 (1968).
6. Davies W. D. T., System Identification for Self-Adaptive Control, Wiley-Interscience, London, 1970.
7. Eykhoff P., Adaptive and Optimalizing Control Systems (Correspondence), *IRE Trans. Autom. Control*, AC-5, 148—151 (1960).
8. Eykhoff P. (см. литературу к гл. 1).
9. Eykhoff P., Process-Parameter Estimation, Progress in Control Engineering, Macmillan R. H., ed., 2, Heywood, London, 1964.
10. Eykhoff P., Van der Grinten P. M., Kwakernaak H., Veltman B. P. (см. литературу к гл. 1).
11. Eykhoff P., Process Parameter and State Estimation, IFAC Symp. «Identification in Autom. Control Systems», Prague; см. также *Automatica*, 4, 205—233 (1968).
12. Kalman R. E., Bucy R. S. (см. литературу к гл. 12).
13. Lee R. C. K., Optimal Estimation, Identification and Control, MIT Press, Cambridge, Mass., 1964; есть русский перевод: Ли Р. Оптимальные оценки, определение характеристик и управление, изд-во «Наука», 1966.
14. Meissinger H. F., The Use of Parameter Influence Coefficients in Computer Analysis of Dynamic Systems, Proc. West. Joint Computer Conf., San Francisco, 1960, pp. 181—192.
15. Nieman R. E., Fisher D. G., Seborg D. E., A Review of Process Identification and Parameter Estimation Techniques, *Int. J. Control*, 13, 209—264 (1971).
16. Райбман Н. С., Что такое идентификация?, изд-во «Наука», 1970.
17. Richalet J., Rault A., Poliquen R., Process Identification by the Model Method (на французском), Gordon and Breach, Paris, 1971.
18. Sage A. P., Melsa J. L., System Identification, Academic Press, N.Y., 1971; есть русский перевод: Сейдж Э., Мелса Дж., Идентификация систем управления, изд-во «Наука», 1974.
19. Strobel H., The Approximation Problem of the Experimental System Analysis (на немецком), *Zeitschr. Messen. Steuern, Regeln*, 10, 460—464 (1967); 11, 29—34; 73—77 (1968).
20. Strobel H., System Analysis Using Deterministic Test Signals (на немецком), VEB Verlag Technik, Berlin, 1968.
21. Пыпкин Я. З. (см. литературу к гл. 1).
22. Young P. C., Process Parameter Estimation, *Control*, 12, 931—936 (1968).

ПРЕДВАРИТЕЛЬНЫЕ СВЕДЕНИЯ

ГЛАВА 3

СИГНАЛЫ: ДЕТЕРМИНИРОВАННЫЕ И СЛУЧАЙНЫЕ

3.1. ТИПЫ СИГНАЛОВ

Теория оценивания и идентификации основывается на математическом описании связи между некоторыми (множеством) входами и (множеством) выходами объекта, заданными как функции времени. При выборе способа описания сигналов нужно учитывать:

- 1) представление об информационной емкости сигналов;
- 2) возможность использования сигналов для исследования динамических характеристик объекта.

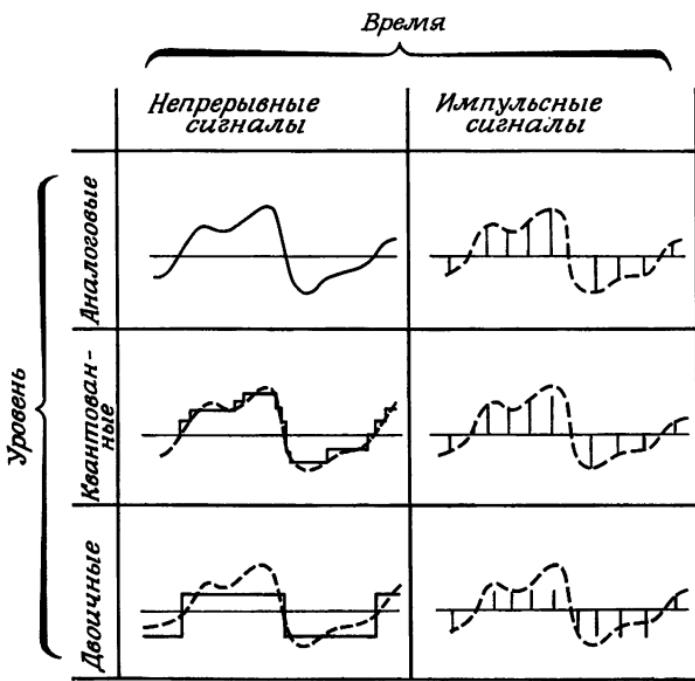
Сигналы, используемые при оценивании, можно разделить на *непрерывные* (более удачный термин *невыборочные*) и *выборочные* (импульсные). Эта классификация отражает свойства сигналов как функций *времени* (фиг. 3.1). Во многих случаях дискретизация сигналов оказывается полезной и даже необходимой, например, когда процесс обработки информации является дискретным (на цифровой машине) или при использовании разделения времени для передачи информации по каналу связи. При дискретизации функция времени $x(t)$ превращается в последовательность импульсов, которую можно представить как вектор:

$$\mathbf{x}' = [x(1), \dots, x(k)].$$

Операция дискретизации является линейной. Действительно, если обозначить оператор дискретизации через \mathcal{S} , то

$$\mathcal{S}[ax(t)] = a\mathcal{S}[x(t)] \text{ для всех } a,$$

$$\mathcal{S}[ax(t) + \beta y(t)] = a\mathcal{S}[x(t)] + \beta\mathcal{S}[y(t)] \text{ для всех } a \text{ и } \beta$$



Ф и г. 3.1.

при условии, что дискретизация обоих сигналов происходит синхронно во времени.

По уровню сигналы делятся на *аналоговые*, *квантованные* и *двоичные*. Это также показано на фиг. 3.1. Двоичный сигнал представляет собой предельный случай квантованного сигнала. Операция квантования не является линейной. Если обозначить оператор квантования через Q , то нельзя утверждать, что для всех α и β

$$Q[\alpha x(t) + \beta y(t)] = \alpha Q[x(t)] + \beta Q[y(t)].$$

Нелинейность операции квантования является источником многих математических трудностей (см. разд. 8.2).

3.2. ОРТОГОНАЛЬНЫЕ ФУНКЦИИ

Функции достаточно общего вида можно описать разложением в ряд по системе ортогональных функций $\{g_i(t)\}$. В этом случае можно получить приближенное выражение

функции, ограничиваясь конечным числом членов разложения. Другое преимущество ортогональных разложений будет выявлено позже в связи с процедурами настройки.

Счетный набор ортогональных функций $g_1(t), g_2(t), \dots$ на интервале (a, b) удовлетворяет условиям

$$\int_a^b g_i(t) g_k(t) dt = d_{ik} \begin{cases} = 0, & i \neq k, \\ \neq 0, & i = k. \end{cases} \quad (3.1)$$

Если $d_{ik} = \delta_{ik}$ (где δ_{ik} — символ Кронекера, $\delta_{ik} = 1$ для $i = k$), то такой набор функций называется *ортонормальным*. Набор ортонормальных функций называется *полным*, если не существует такой отличной от нуля функции $y(t)$, что

$$\int_a^b y(t) g_i(t) dt = 0 \quad \text{для всех } i. \quad (3.2)$$

Произвольную функцию времени $x(t)$ можно разложить в ряд по полному набору ортонормальных функций:

$$x(t) = \sum_{i=1}^{\infty} c_i g_i(t), \quad (3.3)$$

где

$$c_i = \int_a^b x(t) g_i(t) dt. \quad (3.4)$$

Примером таких разложений являются ряды Фурье.

Континуальный набор ортонормальных функций определяется условием

$$\int_a^b g_{\mu}(t) g_{\nu}(t) dt = \delta(\mu - \nu), \quad (3.5)$$

$$\delta(\xi) = 0 \text{ для } \xi \neq 0, \quad \int_{-\infty}^{\infty} \delta(\xi) d\xi = 1 \text{ (функция Дирака).}$$

В этом случае

$$x(t) = \int_{-\infty}^{\infty} c_{\nu} g_{\nu}(t) d\nu, \quad (3.6)$$

де

$$c_v = \int_a^b x(t) g_v(t) dt. \quad (3.7)$$

Одним из таких представлений является преобразование Фурье.

Можно отметить следующие интересные свойства счетных наборов ортонормальных функций:

1) Использование критерия наименьших средних квадратов при аппроксимации сигнала суммой конечного числа ортонормальных функций приводит к коэффициентам, определяемым формулой (3.4). Рассмотрим набор m ортонормальных функций. Пусть

$$E = \int_a^b \left\{ x(t) - \sum_{i=1}^m \hat{\gamma}_i g_i(t) \right\}^2 dt. \quad (3.8)$$

Запишем необходимое условие минимума E по γ_k

$$0 = \frac{dE}{d\gamma_k} \Big|_{\gamma_k = \hat{\gamma}_k} = -2 \int_a^b \left\{ x(t) - \sum_{i=1}^m \hat{\gamma}_i g_i(t) \right\} g_k(t) dt.$$

Отсюда вследствие ортонормальности

$$\int_a^b x(t) g_k(t) dt = \sum_{i=1}^m \hat{\gamma}_i \int_a^b g_i(t) g_k(t) dt = \hat{\gamma}_k. \quad (3.9)$$

Сравнение формул (3.9) и (3.4) показывает, что $\hat{\gamma}_i = c_i$. Отсюда следует, что $\sum_i c_i g_i(t)$ обеспечивает наименьшую среднеквадратичную ошибку аппроксимации по сравнению с любой другой линейной комбинацией тех же функций.

2) Величина $\hat{\gamma}_i$ не зависит от m — числа членов разложения. Таким образом, увеличение числа членов разложения с m до n ($n > m$) не меняет коэффициентов c_1, \dots, c_m . Это очень важное свойство.

3) Положив $\hat{\gamma}_i = c_i$ в формуле (3.8), получим

$$\min E = \int_a^b x^2(t) dt - \sum_{i=1}^m c_i^2 \geq 0,$$

или

$$\sum_{i=1}^m c_i^2 \leq \int_a^b x^2(t) dt. \quad (3.10)$$

Устремляя $m \rightarrow \infty$, получим неравенство Бесселя. Поэтому, если $\int_a^b x^2(t) dt < \infty$, то $c_i \rightarrow 0$ при $i \rightarrow \infty$.

4) Если при $m \rightarrow \infty$ минимальная ошибка стремится к нулю, то

$$\sum_{i=1}^{\infty} c_i^2 = \int_a^b x^2(t) dt.$$

Это соотношение называется равенством Парсеваля.

Используя последовательность функций $\{f_i(t)\}$, определенных на интервале (a, b) , с помощью известной процедуры Шмидта можно построить последовательность ортонормальных функций $\{g_i(t)\}$. Первая функция определяется как

$$g_1(t) = \frac{f_1(t)}{\|f_1(t)\|}, \quad (3.11)$$

где

$$\|f_1(t)\| \stackrel{\Delta}{=} \sqrt{\int_a^b f_1^2(t) dt}. \quad (3.12)$$

В результате функция оказывается нормированной, т. е. $\|g_1(t)\| = 1$. Если функция $f_2(t)$ не зависит от $f_1(t)$, то можно выбрать такой коэффициент α , что разность $f_2(t) - \alpha g_1(t)$ окажется ортогональной к $g_1(t)$:

$$\int_a^b \{f_2(t) - \alpha g_1(t)\} g_1(t) dt = 0.$$

Отсюда следует, что

$$\alpha = \int_a^b f_2(t) g_1(t) dt$$

и

$$g_2(t) = \frac{f_2(t) - \alpha g_1(t)}{\|f_2(t) - \alpha g_1(t)\|}. \quad (3.13)$$

Таблица 3.1

Наборы ортогональных функций

Наборы	Интервал (a, b)	Весовая функция $w(t)$	Ортогональные функции
<i>Дискретные</i>			
Фурье	$-\pi \leq t \leq \pi$	1	$g_k(t) = e^{jkt}$
полиномы Лежандра	$-1 \leq t \leq 1$	1	$P_k(t) = \frac{\sqrt{2k+1}}{2^k \sqrt{2k}!} \frac{d^k}{dt^k} (t^2 - 1)^k$
полиномы Чебышева	$-1 \leq t \leq 1$	$\frac{1}{\sqrt{1-t^2}}$	$T_k(t) = \sqrt{\frac{2}{11}} \cos(k \arccos t)$
<i>Полиномы</i> <i>Лагерра</i> <i>Функции</i>	$0 \leq t < \infty$	e^{-t}	$L_k(t) = \frac{1}{k!} \frac{e^t}{dt^k} (t^k e^{-t})$
<i>Полиномы</i> <i>Эрмита</i> <i>Функции</i>	$-\infty < t < \infty$	e^{-t^2}	$l_k(t) = L_k(t) e^{-t/2}$
<i>Непрерывные</i>			
Фурье	$-\infty < t < \infty$		$H_k(t) = \frac{(-1)^k}{\sqrt{2^k k!} \sqrt{\pi}} \frac{e^{t^2}}{dt^k} \frac{d^k}{dt^k} (e^{-t^2})$
Дирака	$-\infty < t < \infty$		$h_k(t) = H_k(t) e^{-t^2/2}$
			$g_\nu(t) = e^{-j\nu t}$
			$g_\nu(t) = \delta(t - \nu)$

Аналогично можно построить нормированную функцию $g_i(t)$, ортогональную к функциям $g_1(t), \dots, g_{i-1}(t)$.

Свойства ортогональности и ортонормальности можно обобщить введением весовой функции $w(t)$; это дает возможность заданным образом изменить вклад в среднеквадратичную ошибку отдельных значений функций $g_i(t)$. Аналогом формул (3.1) и (3.8) являются соотношения

$$\int_a^b g_i(t) g_k(t) w(t) dt = \delta_{ik},$$

$$E = \int_a^b \left\{ x(t) - \sum_{i=1}^m c_i g_i(t) \right\}^2 w(t) dt. \quad (3.14)$$

В зависимости от выбора интервала (a, b) и весовой функции $w(t)$ можно получить различные наборы ортогональных функций, приведенные в табл. 3.1. Для более глубокого изучения теории ортогональных функций следует обратиться к работам [7, 15, 20]. В литературе можно найти примеры практических приложений (см., в частности, [1, 13, 14, 17, 18] и др.). Одна из решенных задач состоит в идентификации ковариационной функции как коэффициентов ортогонального разложения. Новые результаты приведены в [8, 12, 19].

3.3. ОПИСАНИЕ ДЕТЕРМИНИРОВАННЫХ СИГНАЛОВ

Предполагается, что частотные представления сигналов с помощью преобразований Фурье, Лапласа и z -преобразования хорошо известны. Используются определения, приведенные в табл. 3.2¹⁾). Важно заметить, что по теореме Парсеваля ортогональность во временной области влечет за собой ортогональность в частотной области.

Один из способов описания сигналов с помощью ортогональных разложений рассмотрен в предыдущем разделе. Коэффициенты разложения можно найти по фор-

¹⁾ Эта таблица заимствована из *Hewlett — Packard Journal*, 19, № 2 (Oct. 1967).

Таблица 3.2

Различные преобразования и их обращения

Преобразование Фурье	Двустороннее преобразование Лапласа	Одностороннее преобразование Лапласа	Двустороннее Z-преобразование	Одностороннее Z-преобразование
$F(j\omega) = \int_{-\infty}^{\infty} f(t) e^{-j\omega t} dt,$ ω — действительная величина, $F(j\omega)$ существует, если существует	$F(s) = \int_{-\infty}^{\infty} f(t) e^{-st} dt,$ $s = \sigma + j\omega,$ $F(s)$ существует, если существует	$F(s) = \int_0^{\infty} f(t) e^{-st} dt,$ $s = \sigma + j\omega,$ $F(s)$ существует, если существует	$f*(t) = \sum_{n=-\infty}^{\infty} f(n) \delta(t - nT),$ $F(z) = \sum_{n=-\infty}^{\infty} f(n) z^{-n},$ или	$f*(t) = \sum_{n=0}^{\infty} f(n) \delta(t - nT),$ $F(z) = \sum_{n=0}^{\infty} f(n) z^{-n},$ или

$$f(t) =$$

$$= \frac{1}{2\pi} \int_{-\infty}^{\infty} F(j\omega) e^{j\omega t} d\omega$$

$$f(t) =$$

$$f(n) = \frac{1}{2\pi j} \oint_C F(z) z^{n-1} dz,$$

где C охватывает все полюсы $F(z)$ z^{n-1} .

$$= \frac{1}{2\pi j} \int_{\sigma_0-j\infty}^{\sigma_0+j\infty} F(s) e^{st} ds,$$

$$= \frac{1}{2\pi j} \int_{\sigma_0-j\infty}^{\sigma_0+j\infty} F(s) e^{st} ds,$$

$$\sigma_1 < \sigma_0 < \sigma_2$$

МОЖНО НАЙТИ ПО ТАБЛИЦАМ ИЛИ

$$f(t) = \frac{1}{2\pi j} \int_{-j\infty}^{j\infty} F(j\omega) \times$$

$$\times e^{j\omega t} d(j\omega) = \frac{1}{2\pi j} \times$$

$$\times \int_{-j\infty}^{j\infty} F(s) e^{st} ds |_{s=j\omega}$$

$$\sigma_0 > \sigma_1$$

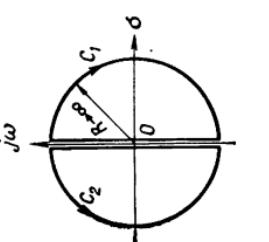
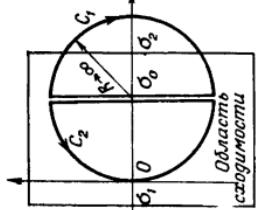
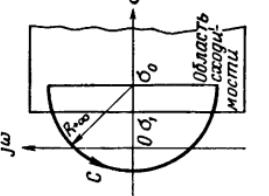
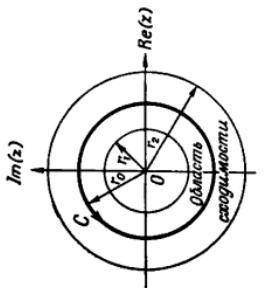
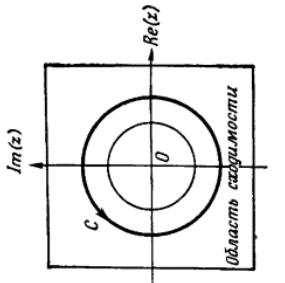
где C — окружность радиуса r_0 : $r_1 < r_0 < r_2$.

$$f(n) = \frac{1}{2\pi j} \oint_C F(z) z^{n-1} dz,$$

где C охватывает все полюсы $F(z)$ z^{n-1} .

Продолжение табл. 3.2

Преобразование Фурье	Двустороннее преобразование Лапласа	Одностороннее преобразование Лапласа	Двустороннее Z-преобразование	Одностороннее Z-преобразование
Для $t < 0$ $f(t) = \frac{1}{2\pi j} \oint_{C_1} F(s) e^{st} ds =$ = Сумма вычетов в полосах правой полу- плоскости	Для $t < 0$ $f(t) = \frac{1}{2\pi j} \oint_{C_1} F(s) e^{st} ds =$ = Сумма вычетов в полосах, расположенных справа от прямой $\sigma = \sigma_0$ (бю- действительная вели- чина)	Для $t < 0$ $f(t) = 0$	Для $n < 0$ $f(n) = \text{Сумма вычетов}$ $F(z) z^{n-1}$ в полосах вне кон- тура C	Для $n < 0$ $f(n) = 0$
Для $t > 0$ $f(t) = \frac{1}{2\pi j} \oint_{C_2} F(s) e^{st} ds =$ = Сумма вычетов в левой полу平面ности	Для $t > 0$ $f(t) = \frac{1}{2\pi j} \oint_{C_2} F(s) e^{st} ds =$ = Сумма вычетов в по- лосах, расположенных слева от прямой $\sigma = \sigma_0$	Для $t > 0$ $f(t) = \frac{1}{2\pi j} \oint_C F(s) e^{st} ds =$ = Сумма всех вычетов $F(s)$, (Контур C охва- тывает все полосы)	Для $n \geq 0$ $f(n) = \text{Сумма вычетов}$ $F(z) z^{n-1}$ в полюсах внутри C	Для $n \geq 0$ $f(n) = \text{Сумма вычетов}$ $F(z) z^{n-1}$
	Замечание. Вычеты определяются интегрированием против часо- вой стрелки. Контур C_1 ориентирован по часо- вой стрелке, поэтому приходится брать вы- четы с минусом	Замечание. По опре- делению односторонне- го преобразования f_0 автоматически лежит справа от полосов $F(s)$	Замечание. Формула для $f(n)$ является формулой для коэффициентов разло- жения Лорана $F(z)$ (см. дополнение)	Замечание. Формула для $f(n)$ является формулой для коэффициентов разло- жения Лорана $F(z)$ (см. дополнение)

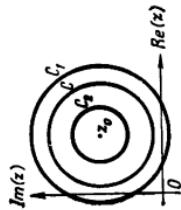


Дополнение

1. $\oint_C f(z) dz = 0$, если $f(z)$ – аналитическая функция внутри контура C и на самом контуре.
2. $\oint_C g(z) dz = 2\pi j \sum \{ \text{вычеты в полюсах } g(z) \text{ внутри } C \}$. Если $g(z)$ имеет в точке z_0 полюс n -го порядка, то $\text{Res } g(z_0) =$

$$= \frac{1}{(n-1)!} \left[\frac{d^{n-1}}{dz^{n-1}} \{(z - z_0)^n g(z)\} \right]_{z=z_0}.$$

3. Если $f(z)$ – аналитическая функция на C_1 и C_2 и в области между ними, то $f(z)$ можно в этой области представить разложением Лорана $f(z) = \sum_{n=-\infty}^{\infty} A_n (z - z_0)^n$, где $A_n = \frac{1}{2\pi j} \oint_C \frac{f(z)}{(z - z_0)^{n+1}} dz$, а C – произвольный контур между C_1 и C_2 , охватывающий точку z_0 .



муте (3.4):

$$c_i = \int_a^b x(t) g_i(t) dt. \quad (3.15)$$

Эту формулу можно реализовать несколькими способами:

1) Перемножая сигнал $x(t)$ и функции из набора ортогональных функций с последующим интегрированием выходного сигнала умножителя на интервале (a, b) (фиг. 3.2). Значения ортогональной функции можно запрограммировать или получить их на ЭВМ как решения дифференциальных уравнений.

2) Используя набор устойчивых фильтров. На фиг. 3.3 представлены примеры фильтров, весовые функции которых являются функциями ортогонального набора, которые можно получить, подавая на вход фильтра δ -функцию. Однако вследствие того, что на выходе фильтра наблюдается свертка входного сигнала и весовой функции фильтра

$$y_i(t) = \int_0^\infty x(t - \theta) g_i(\theta) d\theta, \quad (3.16)$$

непосредственное использование фильтра не приводит к желаемому результату. При $t = 0$ уравнение (3.16) дает

$$y_i(0) = \int_0^\infty x(-\theta) g_i(\theta) d\theta.$$

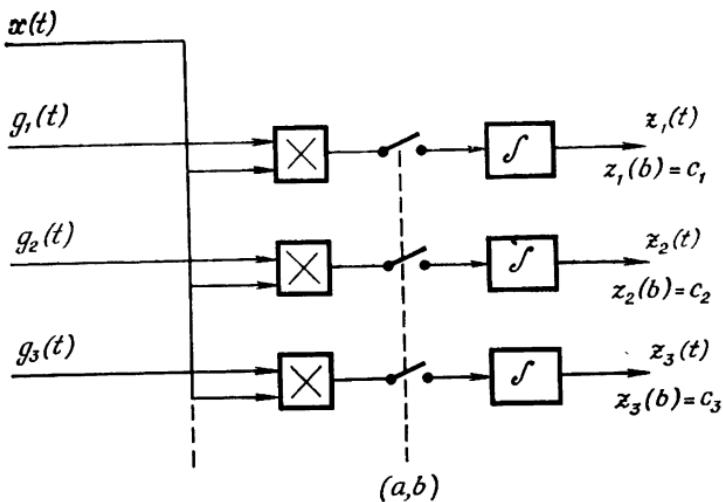
Сравнивая это выражение с формулой (3.4), находим

$$y_i(0) = c_i, \text{ если } x(-\theta) = x(t).$$

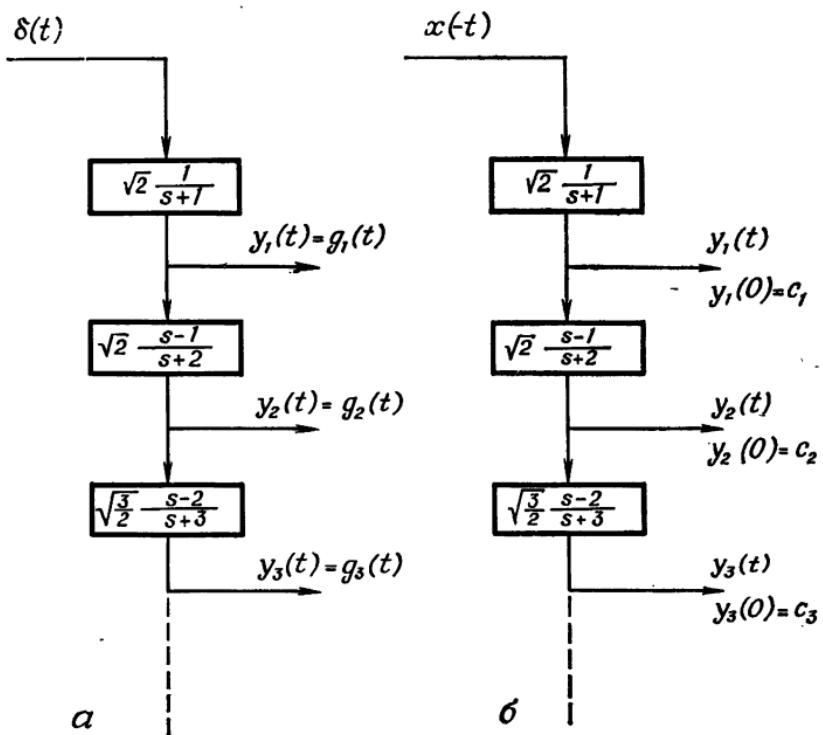
Следовательно, преобразуемый сигнал должен быть записан и пропущен через фильтр с конца записи (в обратном времени) (фиг. 3.3, б). Для оценивания параметров это, конечно, нежелательно.

3) Используя набор неустойчивых фильтров. Формулу (3.15) можно также переписать в виде

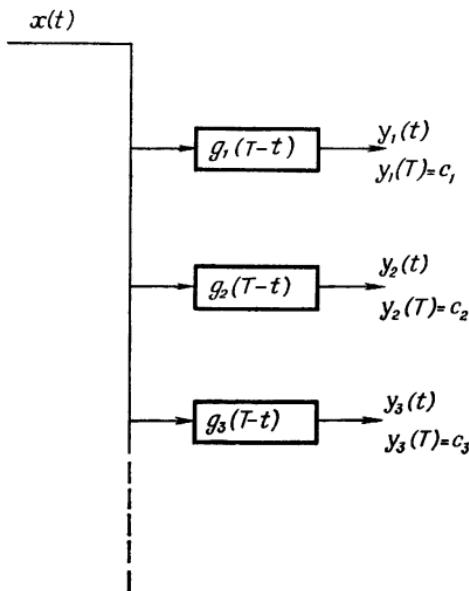
$$y_i(T) = \int_0^T x(\theta) g_i(T - \theta) d\theta,$$



Ф и г. 3.2.



Ф и г. 3.3.



Ф и г. 3.4.

если $x(t) = 0$ при $t < 0$, а свертка берется на интервале $(0, T)$ (фиг. 3.4). Сравнивая это выражение с формулой (3.4), находим, что

$$y_i(T) = c_i, \text{ если } g_i(T - \theta) = g_i(t).$$

Используемый фильтр является неустойчивым. Он может быть реализован на аналоговой машине, если принять меры для гашения возбудившихся усилителей.

Другой способ временного описания сигналов состоит в аппроксимации их функциями вида $a \exp(at)$ и $b \exp(\beta t) \sin(\omega t + \phi)$ и т. д. Эти функции удобны, так как их можно получить как решения линейных однородных дифференциальных уравнений с постоянными коэффициентами.

3.4. ОПИСАНИЕ СТОХАСТИЧЕСКИХ СИГНАЛОВ

На фиг. 3.5 приведено несколько обозначений и формул для случайных функций времени. Сводку формул можно найти в приложении Б. Наиболее важными являются следующие величины:

1) Корреляционная функция (общее определение)

$$\Psi_{xy}(t_1, t_2) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x(t_1) y(t_2) p(x, t_1; y, t_2) dx dy; \quad (3.17)$$

для стационарных процессов

$$\Psi_{xy}(\tau) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x(t) y(t - \tau) p(x, y; \tau) dx dy; \quad (3.18)$$

для эргодических процессов корреляционную функцию можно найти усреднением по времени

$$\Psi_{xy}(\tau) = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T x(t) y(t - \tau) dt; \quad (3.19)$$

для конечного интервала наблюдения

$$\tilde{\Psi}_{xy}(\tau, T) = \frac{1}{T} \int_0^T x(t) y(t - \tau) dt \quad (3.20)$$

является случайной величиной.

2) Спектральная плотность для стационарных процессов

$$\Phi_{xy}(s) = \mathcal{L}_2[\Psi_{xy}(\tau)] \quad (3.21)$$

или

$$\Phi_{xy}(j\omega) = \mathcal{F}[\Psi_{xy}(\tau)]. \quad (3.22)$$

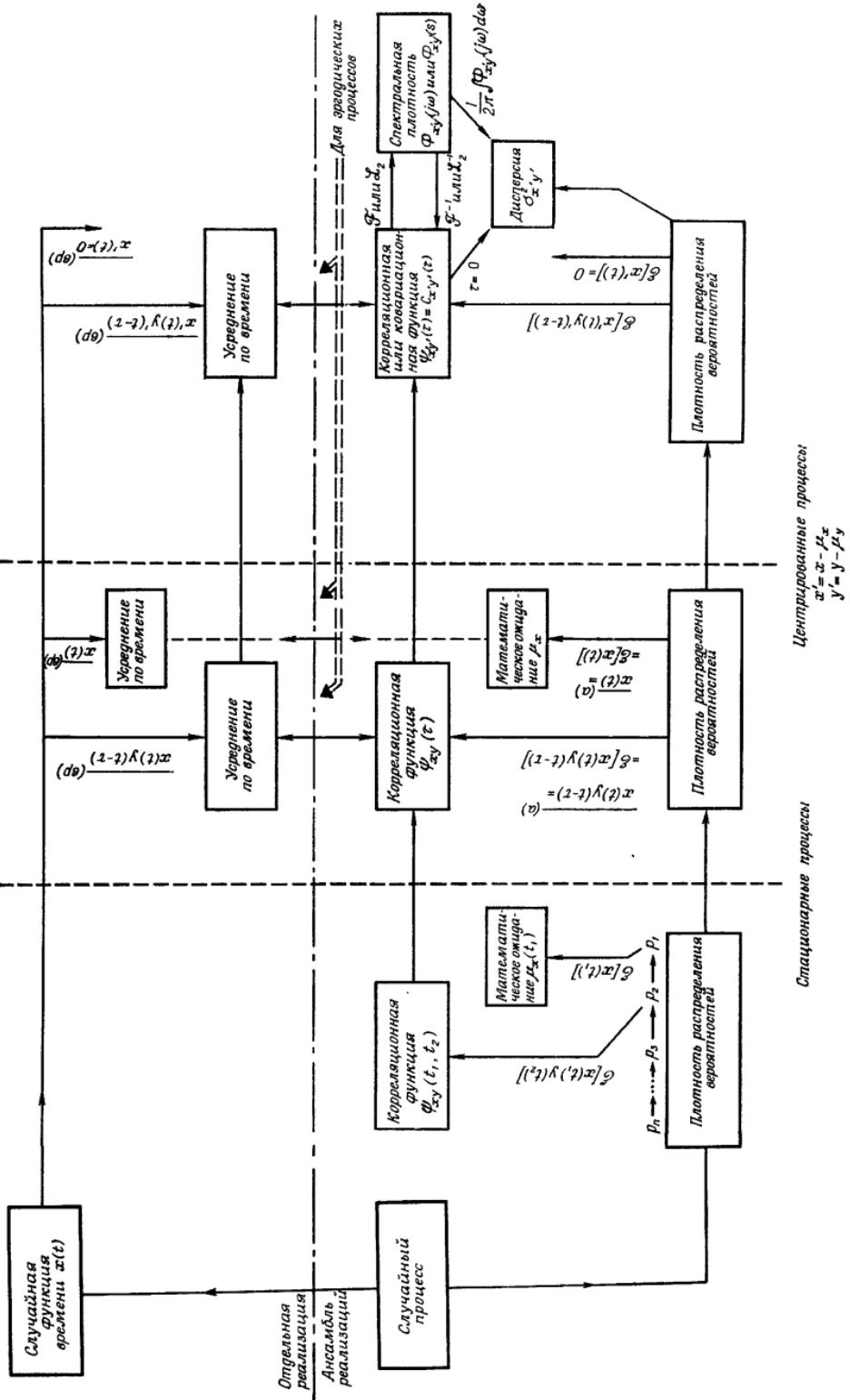
При изучении случайных сигналов в системах в большинстве случаев используется предположение о нормальности случайных процессов. Это связано с тем, что:

а) механизм формирования случайного сигнала по центральной предельной теореме обеспечивает выполнение этого предположения;

б) при подаче на вход линейного фильтра гауссовского случайного процесса на выходе также получается гауссовский процесс;

в) произвольные распределения вероятностей можно считать порождаемыми гауссовским процессом, пропущенным через нелинейный фильтр;

Ф и г. 3.5.



г) одномерное нормальное распределение полностью определяется его средним значением μ и дисперсией σ^2 :

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp \left[-\frac{1}{2} \left(\frac{x-\mu}{\sigma} \right)^2 \right]. \quad (3.23)$$

Для стационарного гауссовского процесса двумерная совместная плотность вероятности $x(t)$ и $y(t) = x(t - \tau)$ также является гауссовой и может быть выражена через $\Psi_{xx}(0)$ и $\Psi_{xx}(\tau)$. Аналогичное утверждение справедливо и для многомерных распределений.

В табл. 3.3 представлены преобразования математических ожиданий и дисперсий стационарных случайных сигналов динамическими системами, сумматорами и умножителями. Предполагается, что все эти формулы хорошо известны, за исключением, быть может, формулы для дисперсии на выходе умножителя, которая выведена в приложении Г.

3.5. ОБРАБОТКА СИГНАЛОВ; БЫСТРОЕ ПРЕОБРАЗОВАНИЕ ФУРЬЕ

Выбор способа описания сигнала существенно зависит от имеющихся возможностей по реализации процедур обработки сигналов. Частотное представление сигналов с помощью преобразования Фурье (см. табл. 3.2) имеет большое значение, несмотря на недостатки, связанные:

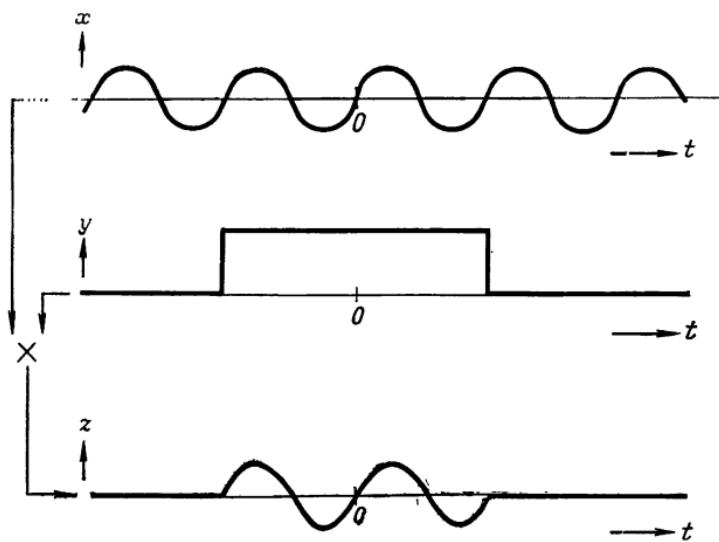
- 1) с бесконечностью интервала наблюдения;
- 2) с необходимостью интегрирования.

Ясно, что в практических ситуациях первый фактор представляет серьезную проблему, так как бесконечный интервал наблюдения исключительно неудобен. На фиг. 3.6 иллюстрируется эффект, связанный с конечностью интервала наблюдения. Здесь наблюдаемый отрезок синусоиды $z(t)$ рассматривается как произведение синусоидальной функции $x(t)$ на прямоугольную функцию $y(t)$:

$$z(t) = x(t) \times y(t). \quad (3.24)$$

Так как операции умножения во временной области соответствует операция свертки в частотной области, преобразование Фурье будет иметь вид

$$Z(j\omega) = X(j\omega) * Y(j\omega). \quad (3.25)$$



Ф и г. 3.6.

Это иллюстрируется фиг. 3.7. Так как ширина спектра $Y(j\omega)$ обратно пропорциональна длине T прямоугольной функции, естественно, возникают трудности, связанные с размыванием и искажением спектра исходной функции. С этим борются, выбирая более гладкую «функцию-окошко» $y(t)$. Это свойство преобразования Фурье и другие близкие вопросы рассмотрены в работах [2, 3, 16].

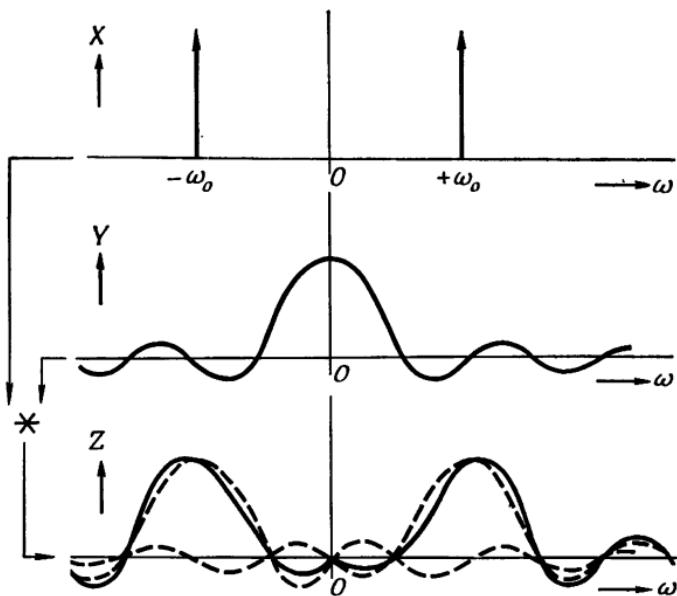
Второй недостаток преобразования Фурье состоит в необходимости интегрирования. Для реализации этой операции на ЦВМ необходимо записать дискретную форму преобразования Фурье.

Дискретное преобразование Фурье (ДПФ). ДПФ и обратное ДПФ для функции, заданной в N дискретных точках, могут быть записаны в виде

$$X(k) = \sum_{n=0}^{N-1} x(n) \exp \left(-j \frac{2\pi kn}{N} \right), \quad k = 0, \dots, N-1, \quad (3.26)$$

и

$$x(n) = \frac{1}{N} \sum_{k=0}^{N-1} X(k) \exp \left(j \frac{2\pi nk}{N} \right), \quad n = 0, \dots, N-1. \quad (3.27)$$



Ф и г. 3.7.

Отметим, что эти выражения очень похожи. Это означает, что оба преобразования могут быть реализованы с помощью одной процедуры. Однако реализация этой процедуры требует выполнения N^2 умножений. Для функций, заданных большим числом выборочных значений, это означает огромное с точки зрения реализации увеличение затрат машинного времени.

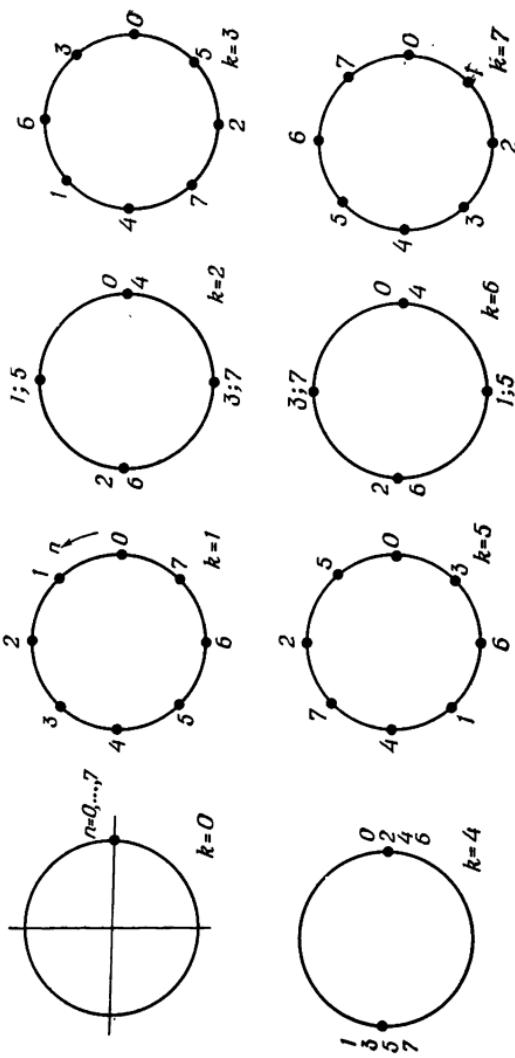
С появлением *быстрого преобразования Фурье (БПФ)* практическое использование частотных представлений заметно расширилось. БПФ обеспечивает значительное уменьшение затрат машинного времени и снижает ошибки округления. К тому же БПФ может быть реализовано при сравнительно небольшой машинной памяти.

БПФ основано на следующей идее. Число умножений для вычисления

$$\exp \left(-j \frac{2\pi k n}{N} \right) = W^{kn},$$

где

$$W = \exp \left(-j \frac{2\pi}{N} \right), \quad N = 2^l, \quad l \text{ — целое число,}$$



Ф и г. 3.8.

может быть сделано значительно меньше, чем N^2 , благодаря тому, что:

а) W^{kn} может принимать только N разных значений; это показано на фиг. 3.8 для случая $N = 8$; $k = 0, \dots, 7$; $n = 0, \dots, 7$;

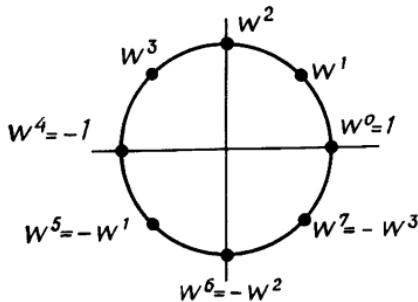
б) эти значения симметричны (фиг. 3.9);

в) одно и то же произведение встречается в БПФ неоднократно, так что можно сгруппировать соответствующие члены.

Можно показать, что число умножений может быть уменьшено до $N \log_2 N$, если $N = 2^l$, т. е. является целой степенью 2. Тогда для $N = 1024 = 2^{10}$ расходуемое время уменьшается до

$$\frac{N \log_2 N}{N^2} = \frac{10}{1024} \approx 1\%.$$

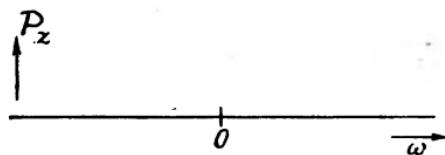
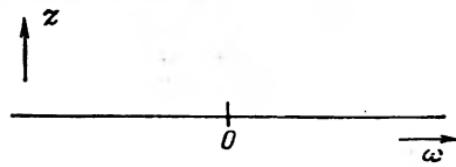
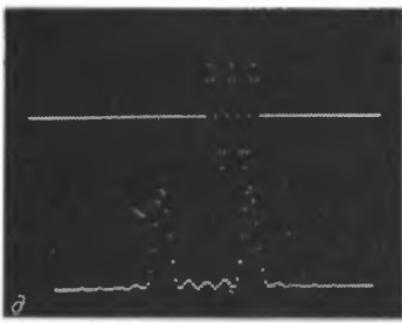
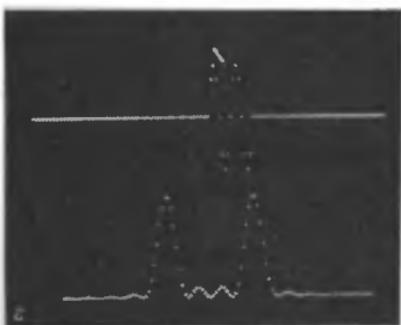
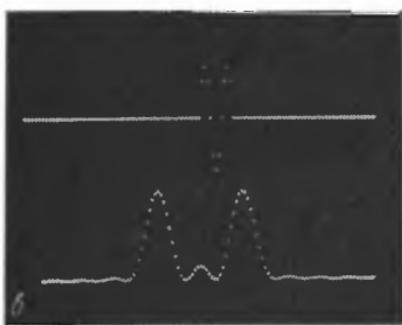
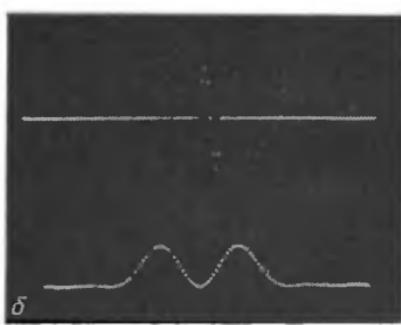
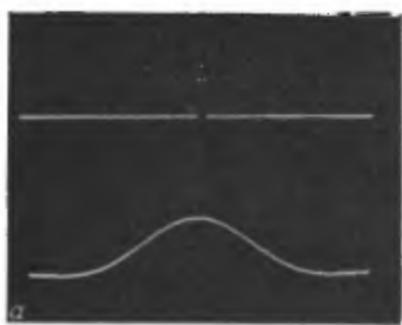
Введение в теорию БПФ можно найти в работах [4, 5]. Другие интересные публикации на эту тему — работы [6, 11] и [9, 10].



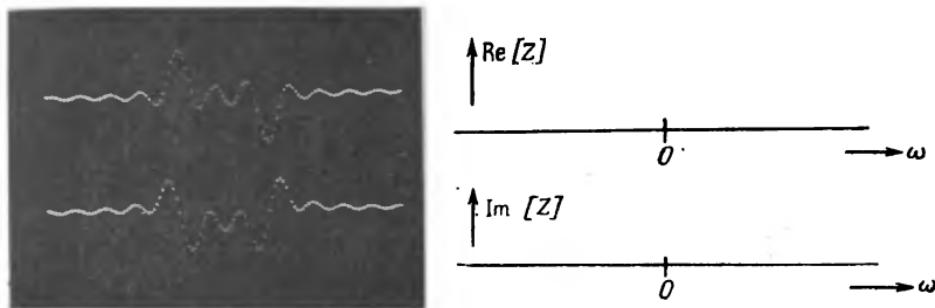
Ф и г. 3.9.

Пример. Использование малой вычислительной машины для вычисления БПФ иллюстрируется на фиг. 3.10 и 3.11¹⁾. На фиг. 3.10 верхний график на каждой фотографии соответствует дискретным значениям преобразуемой функции $z(k)$, нижний график изображает соответствующий энергетический спектр $P_z = Z^*(j\omega)Z(j\omega)$. На фиг. 3.10, а — ∂ виден рассмотренный в этом разделе эффект применения «окошка» из-за конечности реализации. В качестве предельных размеров «окошка» выбрана величина, кратная полупериоду синусоидальной волны — от полупериода до максимального в два с половиной полупериода. По графикам можно проследить «развитие» спектра и вза-

¹⁾ Вычисления проводились на PDP-8. За этот пример автор признателен мистеру Н. Дж. М. Коузенбергу, Технологический университет, Эйндховен, Голландия.



Ф и г. 3.10.



Ф и г. 3.11.

имосвязь между разрешимостью спектра и интервалом наблюдения (см. фиг. 3.6 и 3.7, где вместо P_z приводится Z). На фиг. 3.11 показаны вещественная и мнимая части Z для случая, которому соответствует фиг. 3.10, в.

ЛИТЕРАТУРА

1. Barker H. A., Hawley D. A., Orthogonalising and Optimization, Third IFAC Congress, London, Paper 36A; есть русский перевод: Баркер Х. А., Хоули Д. А., Методы ортогонализации для определения характеристик систем и оптимизации, в книге «Теория непрерывных автоматических систем и вопросы идентификации», изд-во «Наука», 1971.
2. Bendat J. S., Piersol A. G., Random Data Analysis and Measurement Procedures, Wiley-Interscience, N.Y., 1971; есть русский перевод: Бендат Дж., Пирсол А., Измерение и анализ случайных процессов, изд-во «Мир», 1971.
3. Blackman R. B., Tukey J. W., The Measurement of Power Spectra, Dover, N.Y., 1959; первая публикация в *Bell System Techn. J.*, 37 (1958).
4. Brigham E. O., Morrow R. E., The Fast Fourier Transform, *IEEE Spectrum*, 4, № 12, 63—70 (1967).
5. Cochran W. T. et al., What is Fast Fourier Transform?, *Proc. IEEE*, 55, 1664—1674 (1967).
6. Cooley J. W., Tukey J. W., An Algorithm for the Machine Calculation of Complex Fourier Series, *Math. of Comput.*, 19, 297—301 (1965).
7. Courant R., Hilbert D., Methods of Mathematical Physics, Part I (на немецком), 3rd ed., Springer, Berlin, 1968; см. также русский перевод 2-го изд.: Курант Р., Гильберт Д., Методы математической физики, т. I, ГИТТЛ, М.—Л., 1951.
8. Donati F., Milanese M., System Identification with Approximated Models, Second IFAC Symp. «Identification and Process Parameter Estimation», Prague, Paper 1.9, 1970.

9. Fast Fourier Transform, Special Issue *IEEE Trans. on Audio and Electroacoustics*, AU-15, № 2 (June 1967).
10. Fast Fourier Transform, Special Issue *IEEE Trans. on Audio and Electroacoustics*, AU-17, № 2 (June 1969).
11. Gentleman W. M., Sande C., Fast Fourier Transforms—for Fun and Profit, Proc. Fall Joint Computer Conf. 1966, pp. 563—578.
12. Gorecki H., Turowicz A., The Approximation Method of Identification, Fourth IFAC Congress, Warsaw, Paper 5.5, 1969.
13. Kitamori T., Application of Orthogonal Functions to the Determination of Process Dynamic Characteristics and to the Construction of Self-Optimizing Control Systems, First IFAC Congress, Moscow, 1960, pp. 613—618; есть русский перевод: Китамори Т., Применение ортогональных функций для определения динамических характеристик объектов и конструкции самооптимизирующихся систем автоматического управления, в книге «Теория дискретных, оптимальных и самонастраивающихся систем», Труды I Межд. конгресса ИФАК, Изд-во АН СССР, 1961, стр. 817—829.
14. Lampard D. G., A New Method of Determining Correlation Functions of Stationary Time Series, *Proc. IEE*, 102C, 35—41 (1955).
15. Lee Y. W., Statistical Theory of Communication, 3rd ed., Wiley, N.Y., 1963.
16. Richards P. I., Computing Reliable Power Spectra, *IEEE Spectrum*, 4, № 1, 83—90 (1967).
17. Roberts P. D., The Application of Orthogonal Functions to Self-Optimizing Control Systems Containing a Digital Compensator, Third IFAC Congress, London, paper 36C, 1966; есть русский перевод: Робертс П. Д., Применение ортогональных функций в самонастраивающихся системах управления, включаяющих дискретное корректирующее устройство, в книге «Теория непрерывных автоматических систем и вопросы идентификации», изд-во «Наука», 1971, стр. 252—304.
18. Roberts P. D., Orthogonal Transformations Applied to Control System Identification and Optimization, IFAC Symp. «Identification in Autom. Control Systems», Prague, Paper 5.12, 1967.
19. Schwarze G., Determination of Models for Identification with Quality Characteristics in the Time Domain, Fourth IFAC Congress, Warsaw, Paper 19.3, 1969.
20. Sneddon I. N., Special Functions of Mathematical Physics and Chemistry, Oliver and Boyd, Edinburgh, 1961.

ДОПОЛНИТЕЛЬНАЯ ЛИТЕРАТУРА

Anderson W. W., Optimum Estimation of the Mean of a Gaussian Random Process (Correspondence), *Proc. IEEE*, 53, 1640—1641 (1965).

Арато М., Оценка параметров стационарного гауссовского марковского процесса, *ДАН СССР*, 145, № 1, 13—16 (1962).

Balch H. T. et al., Estimation of the Mean of a Stationary Random Process by Periodic Sampling, *Bell Syst. Techn. J.*, 45, 733—741 (1966).

Bendat J. S., Piersol A. G., *Measurement and Analysis of Random Data*, Wiley, N.Y., 1966; есть русский перевод: Бендат Дж., Пирсол А., *Измерение и анализ случайных процессов*, изд-во «Мир», 1971.

Berger E. R., The Increase of Information during Observation of the Relative Frequency Distribution of an Ergodic Process (на немецком), *Nachrtech. Z.*, 17, 590—594 (1964).

Berndt H., Cooper G. R., An Optimum Observation Time for Estimates of Time-Varying Correlation Functions (Correspondence), *IEEE Trans. Inform. Theory*, IT-11, 307—310 (1965).

Berndt H., Cooper G. R., Estimates of Correlation Functions of Nonstationary Random Processes, *IEEE Trans. Inform. Theory*, IT-12, 70—72 (1966).

Bingulac S., The Orthonormalisation of Functions (Correspondence), *IEEE Trans. Autom. Control*, AC-10, 374—375 (1965).

Бочаров И. Н., Стаковский Р. И., Анализатор плотности распределения вероятностей случайных процессов, *Автоматика и телемеханика*, № 2, 169—175 (1962).

Boulton P. I., Kavanagh R. J., A Method of Producing Multiple Non-Correlated Random Signals from a Single Gaussian Noise Source, *IEEE Trans. Applic. and Ind.*, 82, № 65, 46—52 (1963).

Brown J. L., Jr., Generating Uncorrelated Random Outputs by Nonlinear Processing of a Single Noise Source, *IEEE Trans. Applic. and Ind.*, 83, № 75, 408—410 (1964).

Brubaker T. A., Korn G. A., Accurate Amplitude Distribution Analyzer Combining Analog and Digital Logic, *Rev. Scient. Instrum.*, 32, 317 (1961).

Bruni C., Analysis of Approximation of Linear and Time-Invariant Systems Pulse Response by Means of Laguerre Finite Term Expansion (Correspondence), *IEEE Trans. Autom. Control*, AC-9, 580—581 (1964).

Davenport W. B., Jr., Root W. L., *An Introduction to the Theory of Random Signals and Noise*, McGraw-Hill, N.Y., 1958; есть русский перевод: Давенпорт В. Б., Рут В. Л., *Введение в теорию случайных сигналов и шумов*, ИЛ, 1960.

Eier R., Signal Analysis Using Laguerre polynomials (на немецком), *Arch. Elekt. Ubertr.*, 20, 185—194 (1966).

Eine T., Johnson N., On the Estimation of the Mean of a Random Process (Correspondence), *Proc. IEEE*, 53, 187—188 (1965).

Горский А. А., Формулы подсчета математического ожидания и дисперсии оценки спектральной плотности стационарного случайного процесса, *Автоматика и телемеханика*, № 10, 1703—1708 (1965).

Hänel H., Possibilities of Producing Very Low Frequencies, *Zeitschr. Messen, Steuern, Regeln*, 5, 7—13 (1962).

Harvey I. K., A Probability Distribution Analyser Utilizing Electrostatic Stage, *Electron. Engng.*, 33, 432—436 (1961).

Havel J., An Electronic Generator of Random Sequences, *Trans. of the Second Prague Conf.—Information Theory, Statistical Decision Functions, Random Processes*, 1960, pp. 219—229; см. также: Гавел Я., Аппаратура для моделирования случайных процессов,

Труды I Всесоюзного симпозиума по статистическим проблемам в технической кибернетике. Идентификация и аппаратура для статистических исследований, изд-во «Наука», 1970, стр. 296—304.

Herschel R., Determination of the Coefficients of Expansions into Orthogonal Functions by Analog Computer (на немецком), *Elektron. Rechenanl.*, 3, 212—217 (1961).

Hofstetter E. M., Construction of Time-Limited Functions with Specified Autocorrelation Functions, *IEEE Trans. Inform. Theory*, IT-10, 119—126 (1964).

Ivanenko V. I., Hohel O. A., Stabilization Problems of Parameters of Artificially Generated Random Processes, IFAC Symp. «Identification in Autom. Control Systems», Prague, Paper 5.13, 1967; см. также: Иваненко В. И., Хохель О. А., Задачи стабилизации параметров искусственно генерируемых случайных процессов, *Автоматика и телемеханика*, № 6, 59—67 (1969).

Jaňač K., Direct Determination of Correlation Function for Generators of Random Processes, Proc. Third Internat. Conf. on Analog Computers, Opatija, 1962.

Kliger I., On the Determination of Laguerre's Spectrum from the Laplace Transform of a Given Function (Correspondence), *IEEE Trans. Autom. Control*, AC-9, 192—193 (1964).

Korn G. A., Random-Process Simulation and Measurements, McGraw-Hill, N.Y., 1966; есть русский перевод: Корн Г. А., Моделирование случайных процессов на аналоговых и аналого-цифровых машинах, изд-во «Мир», 1968.

Lee Y. W., Application of Statistical Methods to Communication Problems, Techn. Rep. 181, Res. Lab. of Electron. MIT, Cambridge (Mass.), 1950.

März K., The Mean Error of Short-Interval Correlation Functions (на немецком), *Nachrtechn. Z.*, 18, 485—487 (1965)

Пугачев В. С. (см. литературу к гл. 8).

Reissman H. C., Generation of a Noise Spectrum for Lowest Frequencies by Using Radio-Active Decay Processes (на немецком), *Wiss Z. Hochsch. Elektrotechn. Ilmenau*, 7, 185—191 (1961).

Rice S. O., Mathematical Analysis of Random Noise, Part 1—4, Bell Syst. Techn. J., 23, 282—332 (1944); 24, 46—156 (1945).

Rowe I. H., A Method for Generating a Large Number of Uncorrelated Random Signals from a Single Noise Source (Correspondence), *IEEE Trans. Autom. Control*, AC-10, 478—479 (1965).

Thelliez S. F., Note on the Experimental Design of a Random Symbol Generator, IFAC Symp. «Identification in Autom. Control Systems», Prague, Paper 6.2, 1967.

Thomas J. B., The Representation of Stochastic Processes by Orthogonal Expansions, Joint Autom. Control Conf., 1963.

Wax N., Selected Papers on Noise and Stochastic Processes, Dover, N.Y., 1954.

Werner G. W., Reissman H. C. The Generation of Very Low Frequency Noise Spectra and Methods for Their Measurement in Control Engineering (на немецком) *Zeitschr. Messen, Steuern, Regeln*, 6, 328—334 (1963).

Wold H. O. A., *Bibliography on Time Series and Stochastic Processes*, Oliver and Boyd, Edinburgh, 1965.

Яковлев В. П., Получение набора некоррелированных случайных процессов при помощи нелинейных преобразований единственного случайного сигнала, *Автоматика и телемеханика*, № 6, 1099—1104 (1965).

Young T. Y., Huggins W. H., Discrete Orthonormal Exponentials, *Proc. Natn. Electron. Conf.*, 18, 10—18 (1962).

Young T. Y., Huggins W. H., Complementary Signals and Orthogonalized Exponentials, *IRE Trans. Circuit Theory*, CT-9, 362—370 (1962).

ГЛАВА 4

МОДЕЛИ ОБЪЕКТОВ: ЛИНЕЙНЫЕ С ПОСТОЯННЫМИ И ПЕРЕМЕННЫМИ ПАРАМЕТРАМИ, НЕЛИНЕЙНЫЕ

Значение математического описания отмечалось в гл. 1. В этой главе рассматриваются возможности выбора тех или иных моделей. Первый из признаков, по которому различают модели, — их линейность или нелинейность, а также является модель параметрической или непараметрической. Эти типы моделей рассмотрены в разд. 4.2. — 4.4. Для теории идентификации важны такие понятия, как управляемость, наблюдаемость и идентифицируемость, рассмотренные в разд. 4.5. В разд. 4.6 основное внимание уделяется моделям, линейным по параметрам, широко используемым при оценивании, и каноническим представлениям, содержащим наименьшее число параметров.

4.1. КЛАССЫ МОДЕЛЕЙ

Описание поведения объектов с помощью дифференциальных уравнений заимствовано из классической механики. Рассматривается связь между входной величиной $u = u(t)$ и выходом $y = y(t)$ в виде уравнения

$$a_0 \frac{d^n y}{dt^n} + \dots + a_{n-1} \frac{dy}{dt} + a_n y = b_0 \frac{d^n u}{dt^n} + \dots \\ \dots + b_{n-1} \frac{du}{dt} + b_n u + c \quad (4.1)$$

вместе с начальными условиями для $\frac{dy^i}{dt^i}(0)$, $i = 0, 1, \dots, n-1$. Если в объекте имеется чистое запаздывание, то $u = u(t - \tau)$. Один из коэффициентов можно выбирать произвольно, например $a_0 = 1$.

В линейном случае коэффициенты a_i , b_j не зависят от u и y и их производных. Если, кроме того, они не зависят

от времени, то получается *уравнение с постоянными коэффициентами*. Это наиболее простой случай. Если эти коэффициенты зависят от времени, то уравнение называется *линейным уравнением с переменными коэффициентами*, или *нестационарнолинейным*. Если какие-либо из a_i или b_j зависят от u , y или их производных, то объект называют *нелинейным*.

Основное отличие между линейными и нелинейными объектами состоит в том, что для последних не выполняется принцип суперпозиции. Согласно этому принципу, если y_1 — выходной сигнал, обусловленный входом u_1 , а y_2 — выходной сигнал, обусловленный входом u_2 , то при подаче на объект сигнала $\alpha u_1 + \beta u_2$ на выходе наблюдается сигнал $\alpha y_1 + \beta y_2$.

В качестве иллюстрации можно привести следующий простой пример:

Линейное дифференциальное уравнение

$$\dot{y} + ay = u.$$

Отсюда имеем

$$\begin{aligned}\dot{y}_1 + ay_1 &\equiv u_1 \\ + &\\ \dot{y}_2 + ay_2 &\equiv u_2\end{aligned}$$

$$(\dot{y}_1 + \dot{y}_2) + a(y_1 + y_2) \equiv u_1 + u_2.$$

Если положить

$$u_1 + u_2 = u_3,$$

$$y_1 + y_2 = y_3, \text{ откуда}$$

$$\dot{y}_1 + \dot{y}_2 = \dot{y}_3,$$

$$\text{то } \dot{y}_3 + ay_3 = u_3.$$

Нелинейное дифференциальное уравнение

$$\dot{y} + ay^3 = u.$$

Отсюда имеем

$$\begin{aligned}\dot{y}_1 + ay_1^3 &\equiv u_1 \\ + &\\ \dot{y}_2 + ay_2^3 &\equiv u_2\end{aligned}$$

$$(\dot{y}_1 + \dot{y}_2) + a(y_1^3 + y_2^3) \equiv u_1 + u_2.$$

Если положить

$$u_1 + u_2 = u_3,$$

$$y_1 + y_2 = y_3, \text{ откуда}$$

$$\dot{y}_1 + \dot{y}_2 = \dot{y}_3,$$

$$\text{то } (y_1^3 + y_2^3) \not\equiv (y_1 + y_2)^3 = y_3^3$$

и

$$\dot{y}_3 + ay_3^3 \not\equiv u_3.$$

В линейном случае сумма сигналов подчиняется тому же дифференциальному уравнению, что и исходные сигналы, в нелинейном случае это не так. Принцип суперпозиции выполняется относительно как начальных условий

(общее решение), так и входных сигналов (частное решение). Отметим, что принцип суперпозиции выполняется и для линейных уравнений с переменными коэффициентами. Однако в инженерном плане наличие переменных коэффициентов приводит к существенным отличиям. Поэтому случай линейных систем с *переменными коэффициентами* будет рассмотрен отдельно в разд. 4.3.

Трудности решения *нелинейных* дифференциальных уравнений хорошо известны, поэтому набор методов их решения довольно ограничен. Классические подходы к изучению подобных систем, такие, как метод фазовой плоскости или использование приближенной теории, мало что дают для решения задачи описания объекта. В разд. 4.4 в качестве возможного подхода к изучению одного класса нелинейных систем рассматриваются ряды Вольтерра.

Для объектов важно знать следующие зависимости: соотношение между входным и выходным сигналами и связь этих сигналов с переменными состояниями объекта. Рассмотрим выражение, характеризующее связи первого типа:

$$F\{u, \dot{u}, \ddot{u}, \dots, u^{(m)}; y, \dot{y}, \ddot{y}, \dots, y^{(n)}\} = 0, \quad (4.2)$$

где F — аналитическая функция всех переменных $u, \dot{u}, \dots, \ddot{u}; y, \dot{y}, \dots, \ddot{y}$. Так как это дифференциальное уравнение выражает неявно связь между входным и выходным сигналами, то нужно найти явное выражение $y(t)$ как функции $u(t)$. Обозначим эту связь через

$$y(t) = \mathcal{O}[u(\tau)]_{-\infty}^t. \quad (4.3)$$

Оператор \mathcal{O} может быть функционалом, т. е. правилом, которое каждой функции в заданном интервале ставит в соответствие определенное число. Как видно из формулы (4.3), этот интервал может быть от $-\infty$ до t . Рассмотрим оператор \mathcal{O} для ряда частных случаев [57].

1) $F = y - bu = 0$. Это тривиальный случай линейной функциональной связи между $u(t)$ и $y(t)$; объект не содержит инерционных элементов, запасающих энергию. Поэтому такие системы называют *линейными системами без памяти*. В этом случае функционал $\mathcal{O}[u(\tau)]$ сводится

к линейной функции от u :

$$y(t) = \mathcal{O}[u(\tau)]_{-\infty}^t = bu(t).$$

2) $F = \ddot{y} + a_1\dot{y} + a_2y - \dot{u} - b_2u = 0$, a_1, a_2, b_2 — константы. Из теории дифференциальных уравнений известно, что $y(t)$ может быть записана в виде функционала

$$y(t) = \mathcal{O}[u(\tau)]_{-\infty}^t = \int_{-\infty}^t h(t-\tau)u(\tau)d\tau.$$

Объект содержит инерционные элементы, играющие роль памяти, так что имеется зависимость от предыдущих значений $u(t)$ (*линейные системы с памятью*). Один из способов определения весовой функции или импульсной переходной характеристики $h(t)$ по виду дифференциального уравнения состоит в использовании преобразования

$$h(t) = \mathcal{L}^{-1}\left[\frac{s+b_2}{s^2+a_1s+a_2}\right].$$

Другой подход к определению $h(t)$ рассмотрен в разд. 4.2.

3) $F = y + ay^3 - u = 0$, $a = \text{const}$. Между входом и выходом имеется нелинейная связь (например, усилитель в режиме насыщения; мы ограничимся рассмотрением случая монотонной функции). Функционал вновь сводится к функции

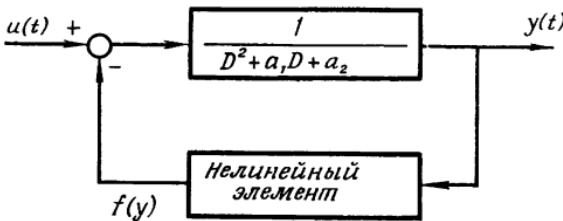
$$y(t) = \mathcal{O}[u(\tau)]_{-\infty}^t = g\{u(t)\}.$$

Эту функцию можно найти подстановкой, итеративным методом или разложением в ряд Тейлора. Для рассматриваемого примера имеем

$$y = u - au^3 + 3a^2u^5 - 12a^3u^7 + \dots$$

Это выражение справедливо для u , принадлежащих области сходимости. Хотя неявно связь задавалась очень простым уравнением, явное выражение оказалось достаточно громоздким. В некоторых случаях следует предпочесть графическое решение [42].

4) $F = \ddot{y} + a_1\dot{y} + a_2y + f(y) - u = 0$, a_1, a_2 — константы. Это уравнение описывает поведение нелинейного



Ф и г. 4.1.

динамического объекта, изображенного на фиг. 4.1 (*нелинейные системы с памятью*). Предполагается, что нелинейная часть безынерционна, что исключает нелинейности гистерезисного типа. В разд. 4.4 будет показано, что в этом случае можно записать следующее явное выражение для $y(t)$ (ряд Вольтерра):

$$\begin{aligned} y(t) = \mathcal{O}[u(\tau)]_{-\infty}^t &= \int_{-\infty}^t h_1(t-\tau) u(\tau) d\tau + \\ &+ \int_{-\infty}^t \int_{-\infty}^t h_2(t-\tau_1, t-\tau_2) u(\tau_1) u(\tau_2) d\tau_1 d\tau_2 + \\ &+ \int_{-\infty}^t \int_{-\infty}^t \int_{-\infty}^t h_3(t-\tau_1, t-\tau_2, t-\tau_3) u(\tau_1) u(\tau_2) \times \\ &\times u(\tau_3) d\tau_1 d\tau_2 d\tau_3 + \dots \end{aligned}$$

Эти примеры могут служить иллюстрациями при обсуждении возможных применений моделей [43]. В работе [43] вводятся понятия «описание в функциональном пространстве» и «описание в пространстве параметров». В первом случае используется идея преобразования, определенного на *функциональном пространстве*, в котором можно описывать входной сигнал объекта. На фиксированном интервале времени сигнал представляется *одной точкой* функционального пространства (например, гильбертова пространства). Можно также получить представление сигнала, используя «скользящий» интервал на временной оси; этому случаю соответствует кривая в *функциональном пространстве*. Примерами таких пред-

ствлений могут служить разложение в ряд Фурье, разложение по прямоугольным сигналам, разложение в ряд по функциям Лагерра. Преобразование в функциональном пространстве определяется динамикой объекта. Выходной сигнал объекта можно описать в таком пространстве или представить точкой в одномерном пространстве уровней сигнала. В этих терминах задача оценивания состоит в том, чтобы определить преобразование функционального пространства входных сигналов в функциональное пространство выходных сигналов. Этот подход относится к методам, использующим идею черного ящика, так как не учитывает информацию о физической природе объекта, о его гипотетической модели.

Описание *в пространстве параметров* основано на предполагаемой математической модели динамики объекта. Такое описание представляет собой параметрическую модель конечной размерности. Координатами пространства параметров являются числовые значения величин, определяющих выход модели. Если, например, предполагаемое описание сводится к обыкновенному дифференциальному уравнению, то координатами могут быть значения коэффициентов и начальных условий. Если внешние силы отсутствуют, то, зная точку пространства параметров, можно предсказать поведение выхода системы. При наличии внешних сил добавление неизвестных параметров этих сил увеличивает размерность пространства параметров. Размерность пространства остается конечной, тогда как при описании в функциональном пространстве пришлось бы в принципе добавить бесконечное число параметров. Таким образом, различают:

а) *непараметрические модели*, например весовые функции, передаточные функции, если заранее не задано число коэффициентов, ковариационные функции, спектральные плотности, ряды Вольтерра;

б) *параметрические модели*, в частности дифференциальные уравнения заданного порядка, модели в пространстве состояний.

Параметрические модели могут приводить к большим ошибкам, если порядок модели не соответствует порядку объекта. Преимущество непараметрических моделей состоит в том, что они не требуют явного знания порядка объ-

екта. Однако в этом случае описание является по существу бесконечномерным, а это часто означает возможность построения модели, выход которой точно повторяет выход объекта.

Интересное обсуждение взаимосвязи параметрических и непараметрических моделей можно найти в литературе по анализу временных рядов (см., например, работы [24, 30, 41, 59]). Следует отметить, что окончательная оценка модели связана, конечно, с основной целью идентификации и предполагаемым ее применением. Исчерпывающее обсуждение динамических моделей можно найти в работе [40].

Несмотря на относительно большой объем этой главы, все же не удалось рассмотреть:

- а) объекты с запаздыванием [13];
- б) объекты с распределенными параметрами [46].

4. 2. ЛИНЕЙНЫЕ МОДЕЛИ

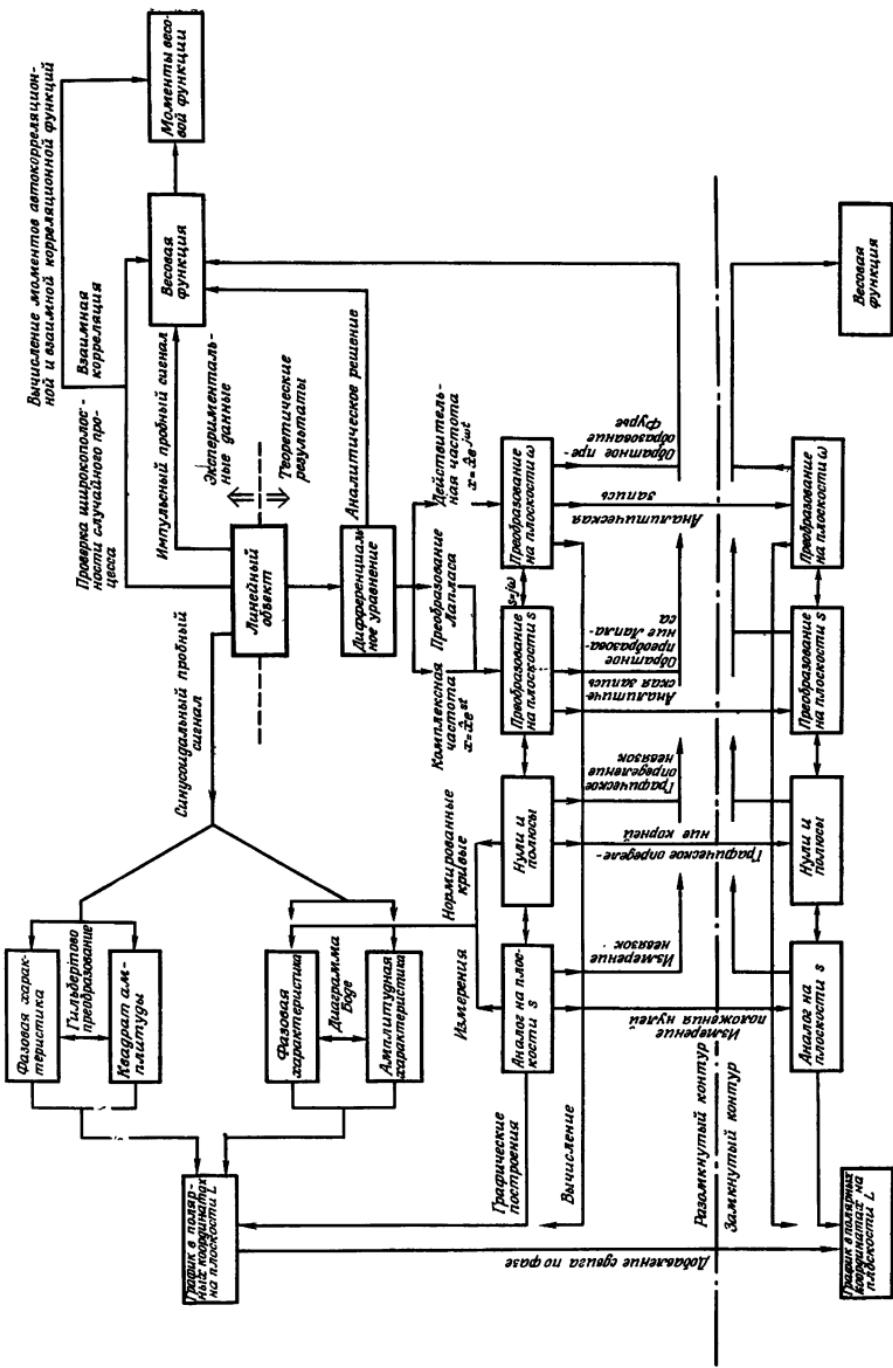
На фиг. 4.2 показаны связи между различными типами описания систем, функционирующих в непрерывном времени. Показаны также несколько видов тестовых проверок и дана информация, которую можно получить с помощью тестов. В последующих разделах некоторые из этих связей изучаются более детально.

Непрерывные детерминированные сигналы

Описание «вход — выход». В соответствии с принципом суперпозиции динамика линейного объекта может быть описана специальными функциями времени: *импульсной переходной функцией* или *весовой функцией* и (или) *реакцией на единичный скачок*. Эти функции являются решениями дифференциального уравнения (4.1) при нулевых начальных условиях и $u(t)$, равной соответственно

$$\left. \begin{array}{l} u(t) = \delta(t) \\ \delta(t) = 0 \text{ при } t \neq 0 \\ \int_{-\infty}^{\infty} \delta(t) dt = 1 \end{array} \right\} \text{единичная импульсная функция} \quad (4.4)$$

Ф и г. 4.2.



и

$$\left. \begin{array}{l} u(t) = 0 \text{ для } t < 0 \\ u(t) = 1 \text{ для } t \geq 0 \end{array} \right\} \text{функция единичного} \quad (4.5)$$

с скачком.

Физически любая функция времени $u(t)$ формально может быть представлена как комбинация таких импульсов или скачков. Это приводит к *интегралу свертки*, который для случая постоянных коэффициентов и нулевых начальных условий может быть записан в виде

$$y(t) = \int_0^{\infty} h(\tau) u(t - \tau) d\tau = \int_{-\infty}^t h(t - \tau) u(\tau) d\tau. \quad (4.6)$$

Если $u(t) = \delta(t)$, то

$$y(t) = \int_0^{\infty} h(\tau) \delta(t - \tau) d\tau = h(t), \quad (4.6a)$$

поэтому $h(t)$ называется весовой функцией. Условием физической реализуемости является $h(t) \equiv 0$ для $t < 0$, т. е. отклик не должен опережать воздействие. Это условие уже было использовано при выборе пределов интегрирования в (4.6).

Для определения $h(t)$ вместо решения дифференциального уравнения при $u(t) = \delta(t)$ можно также воспользоваться следующей процедурой. Рассмотрим выражение (4.6) как предполагаемую модель. Дифференцируя это выражение, найдём все производные, которые входят в уравнение (4.1). Подставляя эти производные в (4.1) и приравнивая коэффициенты, получим выражение для $h(t)$.

Пример. Рассмотрим линейное дифференциальное уравнение первого порядка

$$\dot{y} + ay = u, \quad y(0) = 0; \quad u(t) = 0 \text{ для } t < 0,$$

где

$$y(t) = \int_0^t h(t - \tau) u(\tau) d\tau.$$

Используя правило Лейбница¹⁾, получим

$$y(t) = h(0)u(t) + \int_0^t \frac{dh(t-\tau)}{dt} u(\tau) d\tau.$$

Выражения для y и y' подставляются в дифференциальное уравнение. Так как полученное уравнение должно быть справедливо для всех t и произвольного $u(t)$, то соответствующие члены в левой и правой частях уравнения должны быть равны:

$$h(0)u(t) = u(t) \Rightarrow h(0) = 1,$$

$$\int_0^t \left\{ \frac{dh(t-\tau)}{dt} + ah(t-\tau) \right\} u(\tau) d\tau = 0. \quad (4.7)$$

Последнее уравнение также должно быть справедливо при любой функции $u(t)$. Отсюда следует, что

$$\frac{dh(t)}{dt} + ah(t) = 0 \Rightarrow h(t) = ce^{-at}. \quad (4.8)$$

Объединяя (4.7) и (4.8), получим

$$h(t) = e^{-at}.$$

Таким образом, мы получили известный результат, не используя понятие импульсной функции $\delta(t)$. Также рассматривается случай $y(0) = b \neq 0$.

Описание в пространстве состояний. Понятие состояния играет важную роль в современной теории систем и теории управления [16, 54, 65]. В табл. 4.1 представлены известные соотношения, характеризующие связь между пред-

1) Если

$$y(t) = \int_{a(t)}^{b(t)} f\{\tau, t\} d\tau,$$

где a и b — дифференцируемые функции t , а $f\{\tau, t\}$ и $\partial f\{\tau, t\}/\partial t$ непрерывны по t и τ , то

$$\frac{dy}{dt} = f\{b(t), t\} \frac{db(t)}{dt} - f\{a(t), t\} \frac{da(t)}{dt} + \int_{a(t)}^{b(t)} \frac{\partial f\{\tau, t\}}{\partial t} d\tau$$

Таблица 4.1

Представление во временной области	Представление в частотной области
<p>Описание «вход — выход» (уравнение 2-го порядка)</p> $\ddot{y} + a_1 \dot{y} + a_2 y = u$ $\dot{y}(0), y(0)$ <p>Непосредственное решение</p> $y(t) = h_0(t) y(0) + h_1(t) \dot{y}(0) + \underbrace{\int_0^t h_1(t-\tau) u(\tau) d\tau}_{\text{Частное решение}}$	$\rightarrow \mathcal{L} []$ $\leftarrow \mathcal{L}^{-1} []$ $Y(s) = \frac{s+a_1}{s^2+a_1s+a_2} y(0) + \underbrace{\frac{1}{s^2+a_1s+a_2} \dot{y}(0)}_{\text{Передаточная функция}} + \underbrace{\frac{1}{s^2+a_1s+a_2} U(s)}$ <p>Общее решение</p>

Продолжение табл. 4.1

Представление во временной области

Представление в частотной области

*Описание в пространстве состояний
(общий пример)*

$$\left. \begin{array}{l} \dot{x} = Ax + Bu - \text{уравнение состояний} \\ y = Cx + Du - \text{уравнение выхода} \\ x(0) - \text{начальное условие} \end{array} \right\}$$

Непосредственное
решение

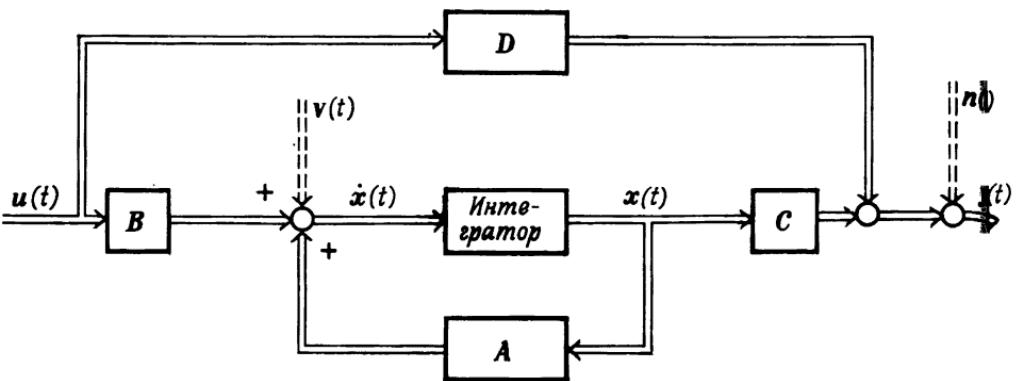
$$X(s) = [sI - A]^{-1} x(0) + [sI - A]^{-1} BU(s)$$

$$x(t) = \underbrace{\Phi(t)x(0)}_{\text{Общее решение}} + \underbrace{\int_0^t \Phi(t-\tau)Bu(\tau)d\tau}_{\text{Частное решение}}$$

$\Phi(t)$ — матрица переходов или матрица
фундаментальных решений

$$\Phi(t) = e^{At}$$

$$[sI - A]^{-1}$$



Ф и г. 4.3.

ставлениями состояний во временной и частотной областях. Помехи (шумы) здесь не учитываются.

Уравнения выходного сигнала и состояния системы схематически представлены на фиг. 4.3, где возможные помехи обозначены через $v(t)$ и $n(t)$. Используются следующие обозначения: A — матрица системы, B — распределительная матрица, C — выходная матрица или матрица наблюдений, D — матрица вход — выход, $v(t)$ — входной шум, $n(t)$ — шум измерений. Пренебрегая действием помех $v(t)$ и $n(t)$, непосредственно получим решение уравнения состояний

$$\dot{x} - Ax = Bu,$$

$$\frac{d}{dt} [e^{-At}x] = e^{-At} [\dot{x} - Ax] = e^{-At}Bu.$$

Определение экспоненциальной матричной функции см. в приложении В. Интегрирование по $(0, t)$ дает

$$e^{-At}x(t) - x(0) = \int_0^t e^{-A\tau}Bu(\tau) d\tau,$$

или

$$x(t) = e^{At}x(0) + \int_0^t e^{A(t-\tau)}Bu(\tau) d\tau. \quad (4.9)$$

Понятие «состояние» тесно связано с понятием «тип колебаний» или «гармоника». Для иллюстрации рассмотрим систему однородных уравнений

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x}, \quad (4.10)$$

здесь \mathbf{A} — матрица размерности $n \times n$. Уравнение (4.10) имеет решение вида

$$\mathbf{x}(t) = e^{\mathbf{At}} \mathbf{x}(0).$$

Вместо этого решения можно попробовать найти такие начальные условия, когда возбуждается одна и только одна гармоника, т. е. когда

$$\mathbf{x}(t) = \mathbf{r}e^{\lambda t}.$$

Подстановка в уравнение (4.10) дает

$$\lambda \mathbf{r}e^{\lambda t} = \mathbf{A}\mathbf{r}e^{\lambda t},$$

или

$$[\mathbf{A} - \lambda \mathbf{I}] \mathbf{r} = 0. \quad (4.11)$$

Ненулевое решение этой системы алгебраических уравнений существует только в том случае, когда определитель равен нулю, т. е. когда

$$|\mathbf{A} - \lambda \mathbf{I}| = 0. \quad (4.12)$$

Корни этого алгебраического уравнения $\lambda_1, \dots, \lambda_n$ называются *характеристическими* или *собственными значениями*. Допустим, что все λ_i различны, тогда каждому собственному значению λ_i соответствует *характеристический*, или *собственный*, *вектор* \mathbf{r}_i , являющийся решением уравнения (4.11). В табл. 4.2 приведен простой пример. Если все собственные значения различны, то можно выбрать такую систему координат, в которой система дифференциальных уравнений состояний распадается на независимые уравнения. Случай кратных собственных значений и обобщенных собственных векторов рассмотрен в работе [20].

Докажем эквивалентность матрицы переходов $\Phi(t)$ из табл. 4.1 и формально полученного решения $e^{\mathbf{At}}$ для примера из табл. 4.2. Используя теорему Кэли — Гамильтона [20, стр. 121] для объекта 2-го порядка с раз-

Таблица 4.2

Представление во временной области		Представление в частотной области
$\Phi(t) := \begin{cases} e^{At} & \text{(система второго порядка)} \\ \begin{bmatrix} h_0(t) & h_1(t) \\ \dot{h}_0(t) & \dot{h}_1(t) \end{bmatrix} & \end{cases}$	$\begin{array}{c} \rightarrow \mathcal{L}[\quad] \\ \leftarrow -\mathcal{L}^{-1}[\quad] \end{array}$	$sI - A = \begin{bmatrix} s & -1 \\ 3 & s+4 \end{bmatrix}$ $\begin{bmatrix} s+4 & 1 \\ s^2+4s+3 & s^2+4s+3 \end{bmatrix}$
<p>Пример:</p> $A = \begin{bmatrix} 0 & 1 \\ -3 & -4 \end{bmatrix}$	$\Phi(t) = \frac{1}{2} \begin{bmatrix} 3e^{-t} - e^{-3t} & e^{-t} - e^{-3t} \\ -3e^{-t} + 3e^{-3t} & -e^{-t} + 3e^{-3t} \end{bmatrix}$	$[sI - A]^{-1} = \begin{bmatrix} -3 & 1 \\ s^2+4s+3 & s^2+4s+3 \end{bmatrix}$

Собственные значения и собственные векторы:

$$|A - \lambda I| = \lambda^2 + 4\lambda + 3 = 0$$

$$\lambda_1 = -1 \Rightarrow \begin{bmatrix} 1 & 1 \\ -3 & -3 \end{bmatrix} r_1 = 0 \Rightarrow r_1 =$$

$$= \alpha_1 \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

$$\lambda_2 = -3 \Rightarrow \begin{bmatrix} 3 & 1 \\ -3 & -1 \end{bmatrix} r_2 = 0 \Rightarrow r_2 =$$

$$= \alpha_2 \begin{bmatrix} 1 \\ -3 \end{bmatrix}$$

Входных сигналов нет; $x(0)$ произвольно:

$$\begin{aligned} x(t) &= \gamma_1 r_1 e^{-t} + \gamma_2 r_2 e^{-3t}, \\ \text{если } x(0) &= \gamma_1 r_1 + \gamma_2 r_2 \end{aligned}$$

$$x(0) = \begin{cases} r_1 \Rightarrow X(s) = \alpha_1 \frac{1}{s+1} r_1 \\ r_2 \Rightarrow X(s) = \alpha_2 \frac{1}{s+3} r_2 \end{cases}$$

личными собственными значениями, можно записать

$$e^{At} = \alpha I + \beta A,$$

где α и β должны быть определены. Для этого по теореме Кэли — Гамильтона вместо матрицы A используются ее собственные значения

$$\begin{aligned} e^{-t} &= \alpha - \beta, \\ e^{-3t} &= \alpha - 3\beta. \end{aligned}$$

Отсюда следует, что

$$\alpha = \frac{1}{2} \{3e^{-t} - e^{-3t}\},$$

$$\beta = \frac{1}{2} \{e^{-t} - e^{-3t}\}.$$

Таким образом,

$$\begin{aligned} e^{\begin{bmatrix} 0 & 1 \\ -3 & -4 \end{bmatrix} t} &= \frac{1}{2} \{3e^{-t} - e^{-3t}\} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \\ &+ \frac{1}{2} \{e^{-t} - e^{-3t}\} \begin{bmatrix} 0 & 1 \\ -3 & -4 \end{bmatrix} = \\ &= \frac{1}{2} \begin{bmatrix} 3e^{-t} - e^{-3t} & e^{-t} - e^{-3t} \\ -3e^{-t} + 3e^{-3t} & -e^{-t} + 3e^{-3t} \end{bmatrix}, \end{aligned}$$

что совпадает с $\Phi(t)$ из табл. 4.2.

Выборочные детерминированные сигналы

Для описания динамики объектов, которые характеризуются выборочными значениями входных и выходных сигналов, вместо дифференциальных уравнений можно воспользоваться разностными уравнениями. Естественно, это возможно в случаях:

1) использования определенной измерительной техники (например, при относительно медленном химическом анализе);

2) измерения и (или) передачи информации по каналам связи с использованием принципа разделения времени;

3) использования устройства для сжатия или передачи цифровой информации;

4) когда информация имеет дискретный характер (как, например, в экономике).

Другие применения связаны с использованием выборочных значений непрерывных процессов и сигналов для упрощения модели. Для выборочных сигналов можно привести схему, аналогичную схеме на фиг. 4.2. В этом случае используются разностные уравнения, отклик на единичный импульс и z -преобразование. Отметим, что данное рассмотрение ограничивается ситуацией, когда как входной, так и выходной сигналы представлены в импульсной форме [16].

Описание «вход — выход». Обозначим через $u(k) = u(k\theta)$ и $y(k) = y(k\theta)$ выборочные значения входного и выходного сигналов соответственно. Тогда разностное уравнение может иметь следующий вид:

$$a_0 y(k) + a_1 y(k+1) + \dots + a_n y(k+n) = b_0 u(k) + \dots + b_m u(k+m). \quad (4.13)$$

Необходимо задать начальные условия. Как и прежде, коэффициент a_0 может быть выбран произвольно, например $a_0 = 1$. [Отметим, что разностное уравнение можно было бы записать через $y(k-i)$, $u(k-i)$.]

Снова рассматривается случай линейной системы с постоянными коэффициентами. В силу линейности справедлив принцип суперпозиции, поэтому динамика объекта может быть описана весовой функцией $h(k)$, *реакцией на единичный импульс*. Интеграл свертки заменяется суммой:

$$y(k) = \sum_{i=-\infty}^k h(k-i) u(i). \quad (4.14)$$

Выражение для $h(k)$ получается так же, как и в непрерывном случае¹⁾.

Пример. Рассмотрим простое разностное уравнение

$$y(k+1) + ay(k) = u(k), \quad y(0) = 0.$$

¹⁾ Заметим, что $h(k)$ не обязательно должно совпадать с $h(t)$ при $t = k T_s$ в непрерывном случае. Использование одинаковых обозначений подчеркивает лишь аналогию между дискретным и непрерывным случаями.

Из уравнения (4.14) имеем

$$\begin{aligned} y(k+1) &= h(0)u(k+1) + h(1)u(k) + h(2)u(k-1) + \dots \\ + ay(k) &= ah(0)u(k) + ah(1)u(k-1) + \dots \end{aligned}$$

$$\begin{aligned} u(k) = y(k+1) + ay(k) &= h(0)u(k+1) + \{h(1) + ah(0)\}u(k) + \\ &+ \{h(2) + ah(1)\}u(k-1) + \dots . \end{aligned}$$

Это уравнение справедливо для любых последовательностей $u(i)$, поэтому

$$h(0) = 0,$$

$$h(1) = 1,$$

$$h(i+1) + ah(i) = 0, \quad i = 1, 2, 3, \dots$$

Решение этой системы уравнений имеет вид

$$h(1) = 1,$$

$$h(2) = -ah(1) = -a,$$

$$h(3) = -ah(2) = a^2,$$

$$h(i) = (-a)^{i-1}, \quad i = 1, 2, 3, \dots$$

Для устойчивых объектов, если $h(i) \approx 0$ при $i > L$, достаточно ограничиться лишь конечным числом значений весовой функции.

Таким образом,

$$y(k) \approx \sum_{i=0}^L h(i)u(k-i), \quad k \geq L. \quad (4.14a)$$

В этом случае поведение системы характеризуется набором величин $\{h(0), \dots, h(L)\}$.

Введем оператор сдвига во времени z :

$$zy(k) = y(k+1)$$

и

$$Y(z) = \sum_{k=0}^{\infty} y(k) z^{-k}.$$

Тогда разностное уравнение (4.13) можно переписать в виде

$$\begin{aligned} \{a_0 z^h + a_1 z^{h+1} + \dots + a_n z^{h+n}\} Y(z) &= \\ = \{b_0 z^h + b_1 z^{h+1} + \dots + b_m z^{h+m}\} U(z). \end{aligned} \quad (4.15)$$

Таблица 4.3

Представление во временной области		Представление с помощью Z-преобразования
<p><i>Описание «ход — выход»</i> (система второго порядка)</p> $y(k+2) + a_1 y(k+1) + a_0 y(k) = u(k)$ $\left. \begin{array}{l} y(1), \quad y(0) \end{array} \right\}$	$\xrightarrow{\quad}$ $\xleftarrow{\quad}$ $\xrightarrow{\quad}$ $\xleftarrow{\quad}$	$\xrightarrow{\quad} Z[\quad]$ $\xleftarrow{\quad} Z^{-1}[\quad]$
<p>Непосредственное решение</p> $y(k) = h_0(k) y(0) + h_1(k) y(1) + \underbrace{\sum_{i=0}^k h_i(k-i) u(i)}_{\text{Общее решение}}$	$\xrightarrow{\quad}$	$Y(z) = \frac{z^2 + a_1 z + a_0}{z^2 + a_1 z + a_0} y(0) +$ $+ \frac{z}{z^2 + a_1 z + a_0} y(1) + \underbrace{\frac{1}{z^2 + a_1 z + a_0}}_{\substack{\text{Импульсная} \\ \text{передаточная} \\ \text{функция}}} U(z)$
<p><i>Описание в пространстве состояний</i> (общий пример)</p> $(k+1) = Ax(k) + Bu(k) \quad \text{— уравнение состояний}$ $y(k) = Cx(k) + Du(k) \quad \text{— уравнение выходного}$ сигнала $\left. \begin{array}{l} x(0) \\ \quad \text{— начальное условие} \end{array} \right\}$	$\xrightarrow{\quad}$ $\xrightarrow{\quad}$ $\xrightarrow{\quad}$ $\xrightarrow{\quad}$	$X(z) = z[zI - A]^{-1} x(0) + [zI - A]^{-1} BU(z)$
<p>Непосредственное решение</p> $x(k) = \Phi(k)x(0) + \underbrace{\sum_{i=0}^{k-1} \Phi(k-1-i)Bu(i)}_{\substack{\text{Общее} \\ \text{решение}}}$	\Downarrow	$\xrightarrow{\quad}$
<p>Частное решение</p> $\Phi(k) — \text{переходная, или фундаментальная матрица} \leftarrow$ $\Phi(k) = A^k$	$\xrightarrow{\quad}$	$z[zI - A]^{-1} \rightarrow$ z

Обе части этого уравнения можно разделить на z^k . Первые n значений y предполагаются равными нулю. Из уравнения (4.15) следует, что

$$H(z) = \frac{Y(z)}{U(z)} = \frac{b_0 + b_1 z + \dots + b_m z^m}{a_0 + a_1 z + \dots + a_n z^n} \quad (4.16)$$

Величину $H(z)$ называют импульсной (выборочной) передаточной функцией. Выполняя деление, находим

$$H(z) = h_0 + h_1 z^{-1} + h_2 z^{-2} + \dots, \quad (4.17)$$

где $h_i = h(i)$ те же, что и в формуле (4.14).

Описание в пространстве состояний. Вместо разностного уравнения (4.13) можно использовать векторное разностное уравнение. В табл. 4.3 сопоставлены представления во временной области и Z-преобразования моделей объектов. Непосредственное решение уравнения состояний во временной области не вызывает затруднений:

$$\mathbf{x}(k+1) = \mathbf{Ax}(k) + \mathbf{Bu}(k).$$

В результате получаем

$$\mathbf{x}(1) = \mathbf{Ax}(0) + \mathbf{Bu}(0),$$

$$\mathbf{x}(2) = \mathbf{Ax}(1) + \mathbf{Bu}(1) = \mathbf{A}^2\mathbf{x}(0) + \mathbf{ABu}(0) + \mathbf{Bu}(1),$$

• •

$$\mathbf{x}(k) = \mathbf{A}^k \mathbf{x}(0) + \sum_{i=0}^{k-1} \mathbf{A}^{k-1-i} \mathbf{B} \mathbf{u}(i).$$

В табл. 4.4 приведен простой пример.

Случайные сигналы

В принципе методы, изложенные в предыдущих разделах, могут быть использованы и для описания случайных сигналов. Текущее значение случайного сигнала не представляет особого интереса. Гораздо важнее знать такие статистические характеристики и параметры, как спектральная плотность, математическое ожидание, дисперсия, корреляционная функция и т. д. (см. приложение Б и разд. 3.4).

Пусть на вход системы поступает сигнал $u(t)$, а на выходе наблюдается сигнал $y(t)$. Эти сигналы связаны

Таблица 4.4

Представление во временной области

Представление с помощью Z-преобразования

$$\Phi(k) = \begin{cases} \mathbf{A}^k & (\text{система второго} \\ \begin{bmatrix} h_0(k) & h_1(k) \\ h_0(k+1) & h_1(k+1) \end{bmatrix} & \text{порядка}) \end{cases}$$

Пример:

$$\mathbf{A} = \begin{bmatrix} 0 & 1 \\ -3 & -4 \end{bmatrix}$$

$$z\mathbf{I} - \mathbf{A} = \begin{bmatrix} z & -1 \\ 3 & z+4 \end{bmatrix}$$

$$\Phi(k) = \frac{1}{2} \begin{bmatrix} 3(-1)^k - (-3)^k & (-1)^k - (-3)^k \\ -3(-1)^k + 3(-3)^k & -(-1)^k + 3(-3)^k \end{bmatrix} = \leftarrow z[\mathbf{z}\mathbf{I} - \mathbf{A}]^{-1} =$$

$$= \begin{bmatrix} \{1, 0, -3, \dots\} & \{0, 1, -4, \dots\} \\ \{0, -3, 12, \dots\} & \{1, -4, 13, \dots\} \end{bmatrix} = \begin{bmatrix} \frac{z^2 + 4z}{z^2 + 4z + 3} & \frac{z}{z^2 + 4z + 3} \\ \frac{-3z}{z^2 + 4z + 3} & \frac{z^2}{z^2 + 4z + 3} \end{bmatrix}$$

Система неустойчивая

Как и в непрерывном случае можно рассмотреть собственные значения и собственные векторы системы и соответствующие им гармоники.

интегралом свертки

$$y(t) = \int_{-\infty}^{+\infty} h(\theta) u(t - \theta) d\theta, \quad (4.18)$$

где

$$h(t) = 0 \text{ для } t < 0.$$

Найдем *первый* момент, усреднив по ансамблю (взяв математическое ожидание) обе части уравнения (4.18),

$$\mathbb{E}[y(t)] = \int_{-\infty}^{+\infty} h(\theta) \mathbb{E}[u(t - \theta)] d\theta,$$

т. е.

$$\begin{aligned} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} y(t) p(u, y) du dy &= \\ &= \int_{-\infty}^{+\infty} h(\theta) \left\{ \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} u(t - \theta) p(u, y) du dy \right\} d\theta \end{aligned}$$

или в общем случае нестационарного случайного процесса

$$\mu_y(t) = \int_{-\infty}^{+\infty} h(\theta) \mu_u(t - \theta) d\theta.$$

Если $u(t)$ — стационарный случайный процесс, то

$$\mu_y = \mu_u \int_{-\infty}^{+\infty} h(\theta) d\theta.$$

Если стационарный процесс $u(t)$ центрирован ($\mu_u = 0$), то

$$\mu_y = 0.$$

Для отыскания *второго* момента умножим обе части уравнения (4.18) на $u(t)$ и, взяв математическое ожидание от обеих частей, получим

$$\mathbb{E}[u(t - \tau) y(t)] = \int_{-\infty}^{+\infty} h(\theta) \mathbb{E}[u(t - \tau) u(t - \theta)] d\theta; \quad (4.19)$$

для общего случая нестационарного случайного процесса имеем

$$\Psi_{uy}(t - \tau, t) = \int_{-\infty}^{+\infty} h(\theta) \Psi_{uu}(t - \tau, t - \theta) d\theta,$$

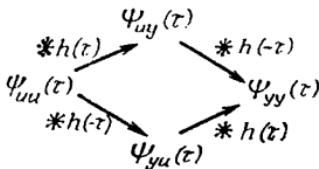
а для стационарного случая

$$\Psi_{uy}(\tau) = \int_{-\infty}^{+\infty} h(\theta) \Psi_{uu}(\tau - \theta) d\theta \stackrel{\Delta}{=} h(\tau) * \Psi_{uu}(\tau).$$

Просматривая полученные формулы, приходим к соотношениям

$$\begin{aligned}\Psi_{uy}(\tau) &= h(\tau) * \Psi_{uu}(\tau), \\ \Psi_{yu}(\tau) &= h(-\tau) * \Psi_{uu}(\tau), \\ \Psi_{yy}(\tau) &= h(\tau) * \Psi_{yu}(\tau), \\ \Psi_{yy}(\tau) &= h(-\tau) * \Psi_{uy}(\tau),\end{aligned}\tag{4.20}$$

которые можно представить следующей диаграммой:



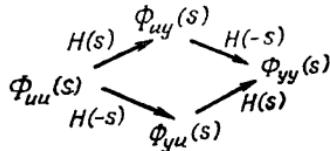
Отметим, что при $\mu_u = 0$

$$\Psi_{uu}(0) = \sigma_u^2 \text{ и } \Psi_{yy}(0) = \sigma_y^2.$$

Используя преобразование Лапласа, находим, что в частотной области

$$\begin{aligned}\Phi_{uy}(s) &= H(s) \Phi_{uu}(s), \\ \Phi_{yu}(s) &= H(-s) \Phi_{uu}(s), \\ \Phi_{yy}(s) &= H(s) \Phi_{yu}(s), \\ \Phi_{yy}(s) &= H(-s) \Phi_{uy}(s).\end{aligned}\tag{4.21}$$

Эти соотношения можно отобразить следующей диаграммой:



Весовую функцию $h(t)$ системы можно определить, решив уравнение свертки при известных автокорреляционной и взаимной корреляционной функциях входного и выходного сигналов. Однако, вообще говоря, это далеко не простая задача (см. разд. 8.1).

Вместо весовой функции объект можно характеризовать временными моментами весовой функции, определяемыми по формуле

$$m_i = \int_0^{\infty} t^i h(t) dt.$$

Момент m_0 представляет собой площадь под графиком весовой функции и равен статическому коэффициенту усиления объекта; первый момент весовой функции m_1 относительно оси $t = 0$ можно интерпретировать как среднее время запаздывания t_{μ} , а именно

$$t_{\mu} = \frac{m_1}{m_0};$$

m_2 — момент инерции весовой функции относительно оси $t = 0$ — можно интерпретировать как дисперсионное время t_{σ} :

$$t_{\sigma} = \sqrt{\frac{m_2}{m_0} - \left(\frac{m_1}{m_0}\right)^2}.$$

Передаточную функцию можно разложить в ряд:

$$\begin{aligned} H(s) &= \int_0^{\infty} h(t) e^{-st} dt = \int_0^{\infty} h(t) \left\{ 1 - st + \frac{s^2 t^2}{2} - \dots \right\} dt = \\ &= m_0 - m_1 s + \frac{m_2 s^2}{2} - \dots \end{aligned}$$

Если этот ряд сходится, то передаточная функция полностью характеризуется своими моментами [23]. Передаточную функцию можно также определить по спектральным плотностям:

$$H(j\omega) = \frac{\Phi_{uy}(j\omega)}{\Phi_{uu}(j\omega)}. \quad (4.22)$$

4.3. НЕСТАЦИОНАРНЫЕ ЛИНЕЙНЫЕ МОДЕЛИ

Как уже отмечалось в разд. 4.1, зависимость от времени коэффициентов линейных дифференциальных или разностных уравнений не отражается на свойстве линейности. Таким образом, можно опять пользоваться аппаратом, изложенным в предыдущем разделе. Однако обычное преобразование Лапласа становится бесполезным. Приведем сводку основных формул.

Описание «вход — выход»

$$y(t) = \int_{-\infty}^t h(t, \tau) u(\tau) d\tau, \quad (4.23)$$

где $h(t, \tau)$ может быть найдено подстановкой в дифференциальное уравнение y , \dot{y} и т. д. с последующим приравниванием коэффициентов. Обзор методов описания нестационарных систем содержится в работе [64], в которой приводится список литературы, содержащий 138 ссылок.

Описание в пространстве состояний

$$\begin{aligned} \dot{\mathbf{x}}(t) &= \mathbf{A}(t) \mathbf{x}(t) + \mathbf{B}(t) \mathbf{u}(t), \\ \mathbf{y}(t) &= \mathbf{C}(t) \mathbf{x}(t) + \mathbf{D}(t) \mathbf{u}(t). \end{aligned} \quad (4.24)$$

Отсюда имеем

$$\mathbf{x}(t) = \Phi(t, t_0) \mathbf{x}(t_0) + \int_{t_0}^t \Phi(t, \tau) \mathbf{B}(\tau) \mathbf{u}(\tau) d\tau, \quad (4.25)$$

где

$$\begin{aligned} \Phi(t_0, t_0) &= \mathbf{I}, \quad [\Phi(t, \tau)]^{-1} = \Phi(\tau, t), \\ \frac{d}{dt} \Phi(t, \tau) &= \mathbf{A}(t) \Phi(t, \tau), \end{aligned} \quad (4.26)$$

$$\frac{d}{d\tau} \Phi'(t, \tau) = -\mathbf{A}'(\tau) \Phi'(t, \tau) \quad (\text{сопряженное уравнение}), \quad (4.26a)$$

$$\Phi(t, \tau) \Phi(\tau, \theta) = \Phi(t, \theta). \quad (4.27)$$

Для выборочных сигналов представление в пространстве состояний имеет вид

$$\begin{aligned} \mathbf{x}(k+1) &= \mathbf{A}(k) \mathbf{x}(k) + \mathbf{B}(k) \mathbf{u}(k), \\ \mathbf{y}(k) &= \mathbf{C}(k) \mathbf{x}(k) + \mathbf{D}(k) \mathbf{u}(k). \end{aligned} \quad (4.28)$$

4.4. НЕЛИНЕЙНЫЕ МОДЕЛИ

Значение свойства линейности и связанного с ним принципа суперпозиции отмечено в разд. 4.1. В случае линейной модели знания одной функции времени, характеризующей объект (весовой функции), достаточно для определения выходного сигнала объекта при произвольных входных сигналах. Это обеспечивает преимущества, которые хотелось бы сохранить и при описании нелинейных объектов, а именно:

- 1) возможность явно записать связь между входом и выходом;
 - 2) простота описания соединений систем;
 - 3) возможность рассмотрения случайных сигналов.
- Для некоторых классов нелинейных объектов эти требования выполняются при использовании рядов Вольтерра.

Разложение Вольтерра¹⁾

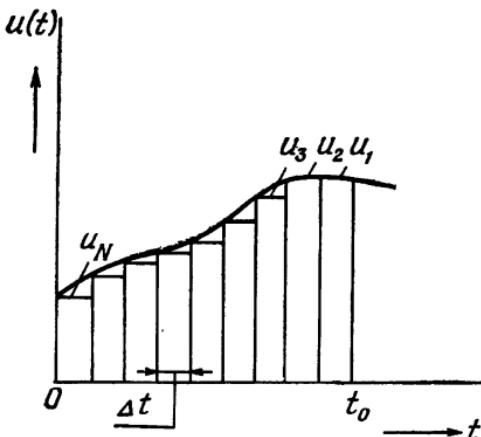
Используя ряды Вольтерра, ядра которых представляют собой весовые функции высших порядков, можно получить описание нелинейного объекта, допускающее ясную физическую интерпретацию. Этот метод имеет большое достоинство, связанное с тем, что нелинейная система рассматривается как непосредственное обобщение линейного случая, хотя сам объект может существенно отличаться от линейного. Иначе говоря, метод с использованием рядов Вольтерра интерпретирует линейные объекты как подкласс нелинейных объектов [63].

Это достоинство вряд ли можно переоценить, так как на временной и спектральный анализ линейных систем затрачено немало времени и усилий, а ряды Вольтерра позволяют использовать накопленные знания для исследования нелинейных систем.

Анализ и синтез с помощью рядов Вольтерра является наиболее удобным из существующих методов изучения нелинейных систем. Изложению этого метода посвящена обширная литература.

Рассмотрим нелинейный объект с памятью. Пусть $u(t) = 0$ при $t < 0$. Кроме того, допустим, что память

¹⁾ В этом разделе изложение в основном следует работе [57].



Ф и г. 4.4.

конечна, т. е. величина $u(t_0 - \tau)$ при достаточно большом τ ничего не добавляет к выходному сигналу $u(t_0)$. Эти допущения естественно вытекают из требований устойчивости и физической реализуемости объекта (у идеального интегратора бесконечная память). Входной сигнал можно аппроксимировать конечным числом прямоугольных импульсов. Обозначим выборочные значения $u(t)$ при $t \leq t_0$, взятые с интервалом Δt , через u_1, u_2, \dots, u_N соответственно (фиг. 4.4). Величина N выбирается достаточно большой, чтобы u_n при $n > N$ ничего не добавляло к $u(t_0)$. Таким образом, выходной сигнал $\tilde{y}(t_0)$ аппроксимируется функцией N переменных $f(u_1, u_2, \dots, u_N)$. Используя разложение в ряд Тейлора функции многих переменных [21], получим

$$\begin{aligned}
 \tilde{y}(t) = & (a_1 u_1 + a_2 u_2 + \dots + a_N u_N) + \\
 & + (a_{11} u_1^2 + a_{12} u_1 u_2 + \dots + a_{NN} u_N^2) + \\
 & + (a_{111} u_1^3 + a_{123} u_1 u_2 u_3 + \dots + a_{NNN} u_N^3) + \dots = \\
 & = \sum_{i=1}^N a_i u_i + \sum_{i=1}^N \sum_{j=1}^N a_{ij} u_i u_j + \\
 & + \sum_{i=1}^N \sum_{j=1}^N \sum_{k=1}^N a_{ijk} u_i u_j u_k + \dots = \\
 & = \tilde{y}_{\text{лин}} + \tilde{y}_{\text{квадр}} + \tilde{y}_{\text{куб}} + \dots \quad (4.29)
 \end{aligned}$$

(многочлены Колмогорова—Габора)

Начальные условия предполагаются нулевыми. При выполнении этих условий и если Δt мало по сравнению с постоянной времени системы, $\tilde{y}(t)$ является хорошей аппроксимацией $y(t)$.

Если все a_{ij} и a_{ijk} положить равными нулю, то остается только один член $y_{\text{лин}} = \sum_{i=1}^N a_i u_i$ — аппроксимация интеграла свертки для линейного объекта. Покажем это. Пусть $a_i = h_i \Delta t$, где h_i — амплитуда, а Δt — ширина импульса. Тогда

$$\tilde{y}_{\text{лин}}(t_0) = h_1 u_1 \Delta t + h_2 u_2 \Delta t + \dots = \sum_{i=1}^N h_i u_i \Delta t. \quad (4.30)$$

Учитывая порядок нумерации u_i (см. фиг. 4.4), покажем, что эта сумма при $\Delta t \rightarrow 0$, $N \rightarrow \infty$ и $N \Delta t = t_0$ стремится к

$$y_{\text{лин}}(t_0) = \int_0^{t_0} h_1(\tau) u(t_0 - \tau) d\tau. \quad (4.31)$$

При подаче на систему двух импульсов вид линейного члена разложения определяется принципом суперпозиции:

$$h_i u_i \Delta t + h_j u_j \Delta t.$$

Если учесть квадратичные члены или члены более высоких порядков, то становится очевидным взаимодействие двух импульсов. Пусть $a_{ij} = h_{ij} \Delta t \Delta t$, тогда сумма двух импульсов даст в квадратичном члене добавку вида $h_{ii} u_i^2 \Delta t \Delta t + 2h_{ij} u_i u_j \Delta t \Delta t + h_{jj} u_j^2 \Delta t \Delta t$. (4.32)

В пределе $\tilde{y}_{\text{квадр}}$ принимает вид

$$y_{\text{квадр}}(t_0) = \int_0^{t_0} \int_0^{t_0} h_2(\tau_1, \tau_2) u(t_0 - \tau_1) u(t_0 - \tau_2) d\tau_1 d\tau_2. \quad (4.33)$$

Таким образом,

$$\tilde{y}(t) = \tilde{y}_{\text{лин}}(t) + \tilde{y}_{\text{квадр}}(t) + \tilde{y}_{\text{куб}}(t)$$

преобразуется в ряд Вольтерра вида

$$\begin{aligned}
 y(t) = & \int_0^t h_1(\tau) u(t-\tau) d\tau + \\
 & + \int_0^t \int_0^t h_2(\tau_1, \tau_2) u(t-\tau_1) u(t-\tau_2) d\tau_1 d\tau_2 + \\
 & + \int_0^t \int_0^t \int_0^t h_3(\dots) \dots \quad (4.34)
 \end{aligned}$$

Таким образом, разложение в ряд Вольтерра является непосредственным обобщением модели линейного объекта в форме интеграла свертки. Весовая функция $h(\tau)$ линейной системы заменяется весовыми функциями (ядрами)

$$h_1(\tau), \quad h_2(\tau_1, \tau_2), \quad h_3(\tau_1, \tau_2, \tau_3), \dots$$

Функция h_i называется весовой функцией i -го порядка. Интеграл

$$\underbrace{\int_0^t \int_0^t \int_0^t \dots \int_0^t}_{i} h_i(\tau_1, \tau_2, \dots, \tau_i) u(t-\tau_1) \dots$$

$$\dots u(t-\tau_i) d\tau_1 \dots d\tau_i$$

называется функционалом i -го порядка. Ряд Вольтерра известен как функциональный степенной ряд, или функциональное разложение. Если объект не содержит динамических элементов или при периодическом входном сигнале частота сигнала стремится к нулю, то ряд (4.34) превращается в степенной ряд по u . Допустим, что динамические элементы отсутствуют, т. е.

$$h_1(\tau) = c_1 \delta(\tau),$$

$$h_2(\tau_1, \tau_2) = c_2 \delta(\tau_1) \delta(\tau_2).$$

Тогда функциональный ряд принимает вид

$$y(t) = c_1 u(t) + c_2 u(t) u(t) + \dots,$$

Ниже будет показано, что поведение степенного ряда определяет характер нелинейности. В некоторых случаях, используя степенные ряды, можно исследовать сходимость ряда Вольтерра. Если перейти к ряду Вольтерра с бесконечным числом членов, то по теореме Фреше функциональным степенным рядом можно описать связь между входом и выходом произвольного непрерывного нелинейного объекта. Функциональный ряд единствен и сходится при $|u(t)| < r$, r называется радиусом сходимости. Для функционального ряда редко удается записать замкнутое выражение, обычно используются разложения в ряд с конечным числом членов или разложения, которые дают удовлетворительную оценку по нескольким членам.

Ядра рядов Вольтерра для физически реализуемых объектов обладают следующими свойствами:

а) $h_n(\tau_1, \tau_2, \dots, \tau_n) = 0$ для $\tau_i < 0$, $i = 1, 2, \dots, n$;

б) $\lim_{\tau_i \rightarrow \infty} h_n(\tau_1, \tau_2, \dots, \tau_n) = 0$, $i = 1, 2, \dots, n$;

в) $h_n(\tau_1, \dots, \tau_n)$ является симметричной функцией или может быть симметризована.

Первые два свойства тривиальны. Выходной сигнал физически реализуемой системы зависит только от предыстории входного сигнала. Все устойчивые объекты имеют конечную память, так что для больших τ_i весовые функции стремятся к нулю. Отметим, что интегратор представляет собой пример объекта с бесконечной памятью, характеристики которого находятся на границе устойчивости.

Перейдем к третьему свойству. Ядра являются симметричными или могут быть представлены в симметричном виде циклической перестановкой аргументов несимметричного ядра. Покажем это для двумерного случая на примере

$$h_2 \text{ симм } (\tau_1, \tau_2) \text{ и } \frac{1}{2} \{h_2 \text{ асимм } (\tau_1, \tau_2) + h_2 \text{ асимм } (\tau_2, \tau_1)\}.$$

Левое и правое выражения не равны между собой, но если проинтегрировать эти выражения в одних и тех же пределах по первому квадранту плоскости (τ_1, τ_2) , то величина

интеграла

$$\int_0^t \int_0^t h_2 \text{симм} (\tau_1, \tau_2) u(t - \tau_1) u(t - \tau_2) d\tau_1 d\tau_2 \quad (4.35)$$

совпадает с интегралом

$$\frac{1}{2} \int_0^t \int_0^t \{h_2 \text{асимм} (\tau_1, \tau_2) + h_2 \text{асимм} (\tau_2, \tau_1)\} \times \\ \times u(t - \tau_1) u(t - \tau_2) d\tau_1 d\tau_2. \quad (4.36)$$

Таким образом, в общем случае симметризация проводится по формуле

$$h_n \text{симм} (\tau_1, \tau_2, \dots, \tau_n) \Leftrightarrow$$

$$\Leftrightarrow \frac{1}{n!} \{h_n \text{асимм} (\tau_1, \dots, \tau_n) + \dots + h_n \text{асимм} (\tau_n, \dots, \tau_1)\}.$$

Определение ядер по дифференциальным уравнениям. В принципе эту задачу можно решить, подставляя ряд Вольтерра для $y(t)$ и его производные в дифференциальное уравнение, как это было сделано в разд. 4.2 при исследовании линейных систем. В работе [19] это проделано для дифференциальных уравнений вида

$$\Theta [y(t)] + \mathcal{K} [y(t), \dot{y}(t), \dots, y^{(n-1)}(t)] = u(t), \quad (4.37)$$

где $\Theta []$ — линейный стационарный динамический оператор, например $\Theta [y(t)] = y^{(n)} + \dots + a_1 \dot{y} + a_0 y$, а $\mathcal{K} []$ — полином от y и производных y . Производная высшего порядка входит в Θ , порядок полинома \mathcal{K} не ниже второго. Метод состоит в определении $\dot{y}, \ddot{y}, \dots, y^{(n)}$ дифференцированием ряда Вольтерра (4.34) с последующей подстановкой в уравнение (4.37). Полученное выражение должно быть справедливо при произвольных входных воздействиях, поэтому можно приравнять нулю коэффициенты при всех функционалах от $u(t)$. Это приводит к системе уравнений в частных производных, решив которые можно определить ядра Вольтерра. В работе [19] процедура подстановки показана на примере дифферен-

циального уравнения

$$\dot{y} + y + ky^2 = u(t).$$

Эта процедура здесь не воспроизводится, так как она достаточно трудоемка, особенно для более сложных дифференциальных уравнений. Удобнее воспользоваться методом многомерного преобразования Лапласа.

Использование преобразования Лапласа. Многомерное преобразование Лапласа определяется следующей формулой:

$$H_n(s_1, \dots, s_n) = \underbrace{\int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty}}_n h_n(\tau_1, \dots, \tau_n) \times \exp \left[- \sum_{i=1}^n s_i \tau_i \right] d\tau_1 \dots d\tau_n \quad (4.38)$$

и называется преобразованием Лапласа n -й степени. Для того чтобы преобразовать функционал

$$y_n(t) = \int_0^t \dots \int_0^t h_n(\tau_1, \dots, \tau_n) u(t - \tau_1) \dots u(t - \tau_n) d\tau_1 \dots d\tau_n,$$

введем новые переменные t_1, t_2, \dots, t_n вместо одной переменной t . Если

$$h_n(\tau_1, \dots, \tau_n) \stackrel{\text{def}}{=} H(s_1, \dots, s_n)^1) \quad (4.39)$$

и

$$u(t) \stackrel{\text{def}}{=} U(s), \quad (4.40)$$

то

$$Y_n(s_1, \dots, s_n) = H_n(s_1, \dots, s_n) U(s_1) \dots U(s_n). \quad (4.41)$$

Теперь функцию $y_n(t)$ можно найти, применяя обратное преобразование Лапласа к $Y_n(s_1, \dots, s_n)$ с последующей подстановкой $t_1 = t_2 = \dots = t_n = t$. Этот прием

¹⁾ Символ $\stackrel{\text{def}}{=}$ означает соответствие, устанавливаемое преобразованием Лапласа второй степени, $\stackrel{\text{def}}{=}$ — третьей степени и т. д.

довольно трудоемок, и лучше воспользоваться методом группировки переменных [22], который устанавливает связь между $Y_n(s_1, s_2, \dots, s_n)$ и $\bar{Y}_n(s)$. Рассмотрим, например,

$$y_2(t_1, t_2) = \frac{1}{(2\pi j)^2} \int_{c_1-j\infty}^{c_1+j\infty} \int_{c_2-j\infty}^{c_2+j\infty} Y_2(s_1, s_2) e^{s_1 t_1 + s_2 t_2} ds_1 ds_2. \quad (4.42)$$

Положим $t_1 = t_2 = t$ и $s_1 + s_2 = s$. Тогда

$$y_2(t, t) = \frac{1}{(2\pi j)^2} \int_{c_1-j\infty}^{c_1+j\infty} \int_{c_2-j\infty}^{c_2+j\infty} Y_2(s - s_2, s_2) e^{st} ds_2 ds. \quad (4.43)$$

Следовательно,

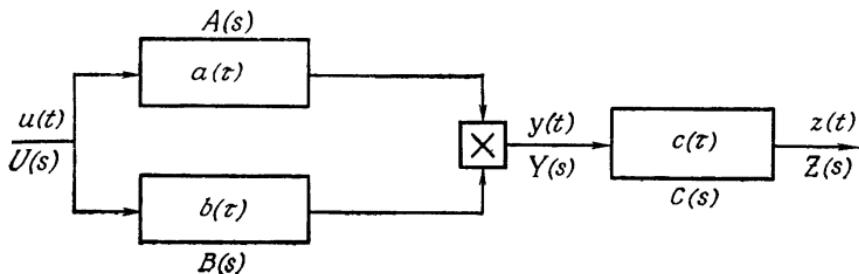
$$\begin{aligned} y_2(t) &\stackrel{\text{д}}{=} Y_2(s), \quad \text{т. е.} \quad y_2(t) = \frac{1}{2\pi j} \int_{c_2-j\infty}^{c_2+j\infty} Y_2(s) e^{st} ds, \\ Y_2(s) &= \frac{1}{2\pi j} \int_{c_1-j\infty}^{c_1+j\infty} Y_2(s - s_2, s_2) ds_2. \end{aligned} \quad (4.44)$$

$Y_n(s)$ можно найти по функции $Y_n(s_1, s_2, \dots, s_n)$, применяя преобразование (4.44) последовательно к (s_1, s_2) и т. д. Способ группировки переменных может подсказать и вид функции $Y_n(s_1, s_2, \dots, s_n)$.

В пространстве (s_1, \dots, s_n) ядра обладают симметрией, определяемой следующим образом:

$$\begin{aligned} H_n \text{ симм } (s_1, s_2, \dots, s_n) &\Leftrightarrow \\ &\Leftrightarrow \frac{1}{n!} \{ H_n \text{ асимм } (s_1, s_2, \dots, s_{n-1}, s_n) + \dots + \\ &\quad + H_n \text{ асимм } (s_n, s_{n-1}, \dots, s_2, s_1) \}. \end{aligned} \quad (4.45)$$

Объекты с линейными звеньями и множительными устройствами хорошо описываются рядами Вольтерра. Вид ядер можно непосредственно определить из структуры такого нелинейного объекта. Это оказывается полезным при решении задачи определения ядер, удовлетворяющих дифференциальному уравнению. Кроме того, в этом случае по ядрам легче строить модель объекта.



Ф и г. 4.5.

Рассмотрим схему, изображенную на фиг. 4.5.

$$\begin{aligned}
 y(t) &= \int_0^t a(\tau) u(t-\tau) d\tau \int_0^t b(\tau) u(t-\tau) d\tau = \\
 &= \int_0^t \int_0^t a(\tau_1) b(\tau_2) u(t-\tau_1) u(t-\tau_2) d\tau_1 d\tau_2, \\
 z(t) &= \int_0^t c(\tau_3) y(t-\tau_3) d\tau_3 = \\
 &= \int_0^t c(\tau_3) \int_0^{t-\tau_3} \int_0^{t-\tau_3} a(\tau_1) b(\tau_2) u(t-\tau_3-\tau_1) \times \\
 &\quad \times u(t-\tau_3-\tau_2) d\tau_1 d\tau_2 d\tau_3.
 \end{aligned}$$

Это функционал второй степени по $u(t)$. Преобразование Лапласа определится как

$$a(\tau_1) b(\tau_2) \stackrel{\text{def}}{=} A(s_1) B(s_2), \quad a(\tau_1) \stackrel{\text{def}}{=} A(s_1), \quad b(\tau_2) \stackrel{\text{def}}{=} B(s_2),$$

$$y(t_1, t_2) \stackrel{\text{def}}{=} Y(s_1, s_2) = A(s_1) B(s_2) U(s_1) U(s_2).$$

Использование формулы (4.44) для группировки переменных приводит к выражению

$$Y(s) = \frac{1}{2\pi j} \int_{c-j\infty}^{c+j\infty} A(s-s_2) B(s_2) U(s-s_2) U(s_2) ds_2.$$

Следовательно,

$$\begin{aligned} Z(s) &= C(s) Y(s) = \\ &= C(s) \frac{1}{2\pi j} \int_{c-j\infty}^{c+j\infty} A(s-s_2) B(s_2) U(s-s_2) U(s_2) ds_2 = \\ &= \frac{1}{2\pi j} \int_{c-j\infty}^{c+j\infty} A(s-s_2) B(s_2) C(s-s_2+s_2) U(s-s_2) U(s_2) ds_2. \end{aligned}$$

Это приводит к формулам

$$Z(s_1, s_2) = A(s_1) B(s_2) C(s_1 + s_2) U(s_1) U(s_2)$$

и

$$z(t) = \int_0^t \int_0^t h(\tau_1, \tau_2) u(t-\tau_1) u(t-\tau_2) d\tau_1 d\tau_2,$$

где

$$h(\tau_1, \tau_2) \doteqdot A(s_1) B(s_2) C(s_1 + s_2).$$

Этот результат можно обобщить следующими утверждениями:

1) если рассматривать линейные системы, выходы которых перемножаются, то преобразование ядра объекта имеет вид

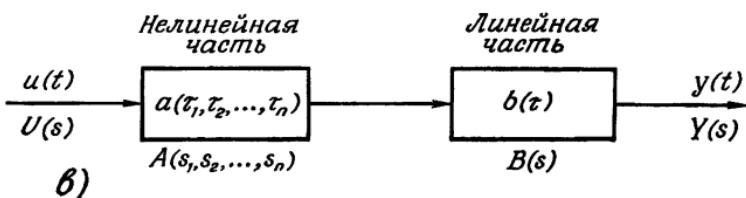
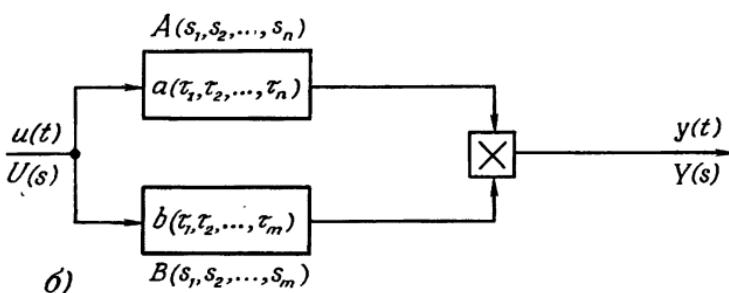
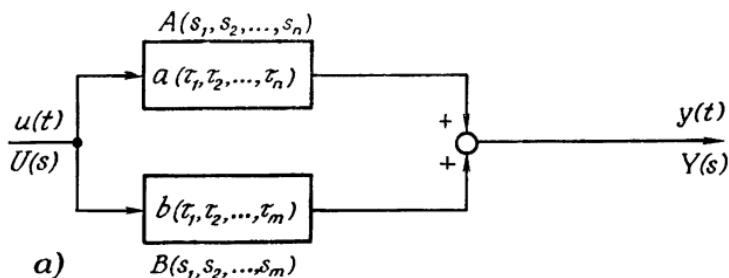
$$A_1(s_1) A_2(s_2) \dots A_n(s_n); \quad (4.46)$$

2) если такой объект соединить последовательно с линейной системой, передаточная функция которой $B(s)$, то преобразование ядра нового объекта имеет вид

$$A_1(s_1) A_2(s_2) \dots A_n(s_n) B(s_1 + s_2 + \dots + s_n). \quad (4.47)$$

Это функция n переменных, которая во временной области соответствует функционалу n -й степени.

Временное представление сводится к сложным интегралам. Преобразование Лапласа имеет простую структуру и может быть сразу переведено на язык схем. Правила перевода приведены на фиг. 4.6.



Ф и г. 4.6.

а) Если $m \neq n$, то систему можно описать уравнением

$$\begin{aligned}
 y(t) = & \underbrace{\int_0^t \dots \int_0^t}_{n} a(\tau_1, \dots, \tau_n) u(t-\tau_1) \dots u(t-\tau_n) d\tau_1 \dots d\tau_n + \\
 & + \underbrace{\int_0^t \dots \int_0^t}_{m} b(\tau_1, \dots, \tau_m) u(t-\tau_1) \dots u(t-\tau_m) d\tau_1 \dots d\tau_m. \quad (4.48)
 \end{aligned}$$

Если $m = n$, то

$$a(\tau_1, \tau_2, \dots, \tau_n) + b(\tau_1, \tau_2, \dots, \tau_n) \stackrel{\text{def}}{=} A(s_1, s_2, \dots, s_n) + B(s_1, s_2, \dots, s_n). \quad (4.49)$$

б) $y(t) = \int_0^t \dots \int_0^t a(\tau_1, \tau_2, \dots, \tau_n) b(\tau_{n+1}, \tau_{n+2}, \dots, \tau_{m+n}) u(t - \tau_1) \dots u(t - \tau_{m+n}) d\tau_1 \dots d\tau_{m+n}, \quad (4.50)$

где

$$a(\tau_1, \tau_2, \dots, \tau_n) b(\tau_{n+1}, \dots, \tau_{n+m}) \stackrel{\text{def}}{=} \stackrel{\text{def}}{=} A(s_1, s_2, \dots, s_n) B(s_{n+1}, \dots, s_{n+m});$$

в) $y(t) = \int_0^t \dots \int_0^t h(\tau_1, \tau_2, \dots, \tau_n) u(t - \tau_1) u(t - \tau_2) \dots u(t - \tau_n) d\tau_1 \dots d\tau_n, \quad (4.51)$

где

$$h(\tau_1, \tau_2, \dots, \tau_n) \stackrel{\text{def}}{=} A(s_1, s_2, \dots, s_n) B(s_1 + s_2 + \dots + s_n).$$

Пример 1. В соответствии с фиг. 4.7

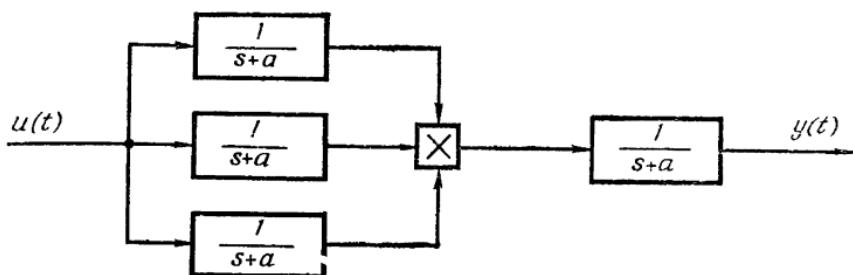
$$y(t) = \int_0^t \int_0^t \int_0^t h_3(\tau_1, \tau_2, \tau_3) u(t - \tau_1) u(t - \tau_2) \times \\ \times u(t - \tau_3) d\tau_1 d\tau_2 d\tau_3,$$

где

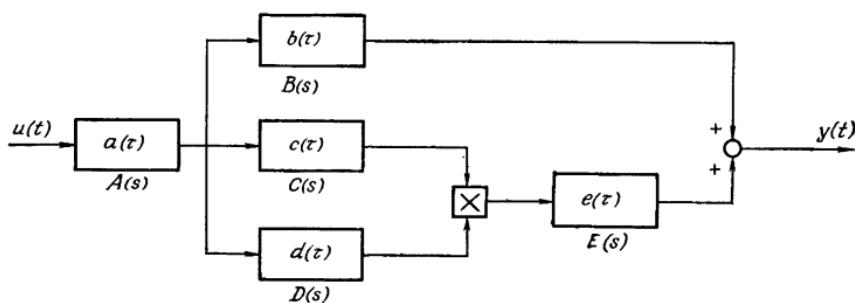
$$H_3(s_1, s_2, s_3) = \frac{1}{s_1 + a} \frac{1}{s_2 + a} \frac{1}{s_3 + a} \frac{1}{s_1 + s_2 + s_3 + a} \stackrel{\text{def}}{=} \\ \stackrel{\text{def}}{=} h_3(\tau_1, \tau_2, \tau_3).$$

Использованы правила б) и в).

Пример 2. На фиг. 4.8 представлен объект, состоящий из линейной и нелинейной частей. Динамика объекта



Ф и г. 4.7.



Ф и г. 4.8.

описывается уравнением

$$\begin{aligned} y(t) = & \int_0^t h_1(\tau) u(t-\tau) d\tau + \\ & + \int_0^t \int_0^t h_2(\tau_1, \tau_2) u(t-\tau_1) u(t-\tau_2) d\tau_1 d\tau_2, \end{aligned}$$

где

$$h_1(\tau) \doteq A(s) B(s),$$

$$h_2(\tau_1, \tau_2) \doteq A(s_1) C(s_1) A(s_2) D(s_2) E(s_1 + s_2).$$

Этот пример демонстрирует изящество многомерного преобразования Лапласа; формула для $h(\tau_1, \tau_2)$ достаточно сложна, а преобразование имеет простой вид и легко может быть интерпретировано. Простота записи в частотной области может быть использована при определении ядер ряда Вольтерра по дифференциальным уравнениям.

Пример 3. Рассмотрим уравнение

$$\dot{y} + y + ay^2 = u(t).$$

Дифференцирование ряда Вольтерра является достаточно трудоемкой операцией. Таким образом, желательно заменить операцию дифференцирования соответствующей операцией для преобразований. Для одномерного преобразования Лапласа это сводится к умножению на s , а в случае n -мерного преобразования — к умножению на $(s_1 + s_2 + \dots + s_n)$. В рассматриваемом примере преобразование Лапласа от ядра второго порядка $H_2(s_1, s_2) = \mathcal{L}_2[h_2(\tau_1, \tau_2)]$ в результате дифференцирования ядра заменяется на $(s_1 + s_2) H_2(s_1, s_2)$.

При этом исходное уравнение принимает вид

$$\underbrace{s_1 H_2(s_1, s_2)}_{\dot{y}} + \underbrace{s_2 H_2(s_1, s_2)}_{y} + \underbrace{H_2(s_1, s_2)}_{u} + a \frac{1}{s_1 + 1} \frac{1}{s_2 + 1} = 0,$$

откуда

$$H_2(s_1, s_2) = -a \frac{1}{s_1 + s_2 + 1} \frac{1}{s_1 + 1} \frac{1}{s_2 + 1}.$$

Применение обратного преобразования дает

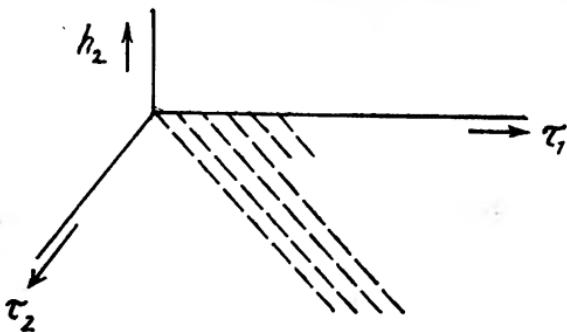
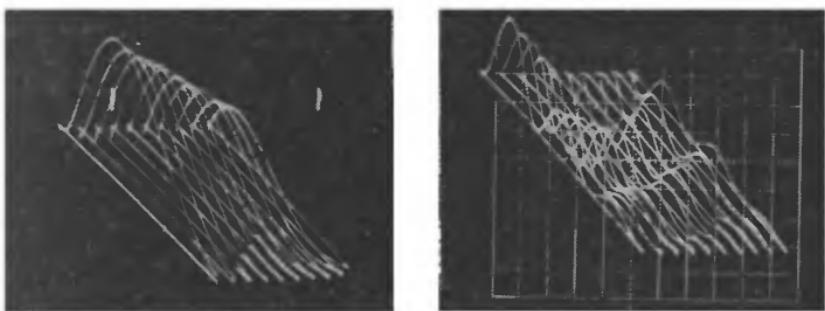
$$h_2(\tau_1, \tau_2) = a \{e^{-(\tau_1+\tau_2)} - e^{-1/2[(\tau_1+\tau_2)-|\tau_1-\tau_2|]}\}.$$

Для $\tau_1 = \tau_2 = t$ имеем

$$h_2(t) = a \{e^{-2t} - e^{-t}\}.$$

Члены более высокого порядка могут быть вычислены точно так же. Без априорной информации об объекте невозможно определить, сколько членов потребуется для адекватного описания объекта. Этот пример показывает, что для некоторых типов дифференциальных уравнений довольно просто можно определить соответствующие им ряды Вольтерра.

Ядра $h_1(\tau)$, $h_2(\tau_1, \tau_2), \dots$ определяются через преобразования Лапласа. Ряды Вольтерра можно реализовать схемами, состоящими из линейных звеньев и умножителей. С другой стороны, для нелинейных объектов,



Ф и г. 4.9.

составленных из линейных звеньев и умножителей, могут быть определены переходные характеристики.

Из приведенных примеров видно, что модели, основанные на рядах Вольтерра, могут быть легко набраны на аналоговых вычислительных машинах. На фиг. 4.9 представлены две осциллограммы ядра второго порядка для простого нелинейного объекта. Эта «весовая функция» была определена экспериментально [57]. Другие результаты в теории и приложениях рядов Вольтерра можно найти в работах [45, 51]. В работе [19] рассматривается также случай ненулевых начальных условий. Для нелинейных импульсных объектов можно воспользоваться рядами Вольтерра в дискретной форме. Применяя многомерное z -преобразование, приходим к схеме, которая аналогочна схеме, полученной в непрерывном случае (см. [1]).

Другие способы описания нелинейных объектов

Кроме разложения в ряды Вольтерра, существует также несколько других способов описания нелинейных объектов, которые будут даны ниже в сокращенном изложении. Своим происхождением эти методы в большей или меньшей степени обязаны исследованиям, направленным на разработку способов *идентификации* нелинейных объектов. В работе [38] предлагается функциональное разложение с ортогональными свойствами. Входной сигнал представляет собой случайный шум с равномерным энергетическим спектром (белый шум). Разложение имеет следующий вид:

$$y(t) = \sum_{i=1}^n G_i [g_i, u(t)], \quad (4.52)$$

где $G_i [g_i, u(t)]$ является функционалом от $u(t)$, а g_i есть функции n переменных, которые, не совпадая с ядрами ряда Вольтерра, могут быть сопоставлены им. Если входной сигнал $u(t)$ — белый шум, то G_j является функционалом j -й степени, ортогональным ко всем функционалам G_i с номерами $i < j$:

$$\overline{G_i [g_i, u(t)] G_j [g_j, u(t)]} = 0, \quad i < j. \quad (4.53)$$

Черта сверху означает усреднение. Несколько первых членов функционального ряда имеют вид

$$G_1 [g_1, u(t)] = \int_{-\infty}^{+\infty} g_1(\tau_1) u(t - \tau_1) d\tau_1,$$

$$G_2 [g_2, u(t)] = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} g_2(\tau_1, \tau_2) u(t - \tau_1) u(t - \tau_2) d\tau_1 d\tau_2 -$$

$$- k \int_{-\infty}^{+\infty} g_2(\tau_1, \tau_1) d\tau_1, \quad (4.54)$$

$$G_3 [g_3, u(t)] = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} g_3(\tau_1, \tau_2, \tau_3) u(t - \tau_1) \times \\ \times u(t - \tau_2) u(t - \tau_3) d\tau_1 d\tau_2 d\tau_3 - \\ - 3k \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} g_3(\tau_1, \tau_2, \tau_3) u(t - \tau_1) u(t - \tau_2) d\tau_1 d\tau_2.$$

Здесь k — мощность белого шума. Ортогональность ряда приводит к простым методам идентификации неизвестных нелинейных объектов. Так, ядра $g_j(\tau_1, \tau_2, \dots, \tau_j)$ можно определить, используя корреляционные методы.

Винер [60] разработал другой способ описания нелинейных объектов, и в этом случае цель разработки была продиктована необходимостью идентификации неизвестных объектов. Были использованы ортогональные функции Лагерра и полиномы Эрмита. Ортогональность привела к простой общей модели. Разложение входного сигнала $u(t)$ по функциям Лагерра имеет вид

$$u(-t) = \sum_{i=1}^{\infty} a_i l_i(t), \quad t \geq 0, \quad (4.55)$$

где $l_i(t)$ — функции Лагерра. Таким образом, предыстория сигнала известна, если заданы коэффициенты a_i (фиг. 4. 10). Если $u(t)$ разложить в ряд по ортогональным полиномам, то отклик $y(t)$ примет вид

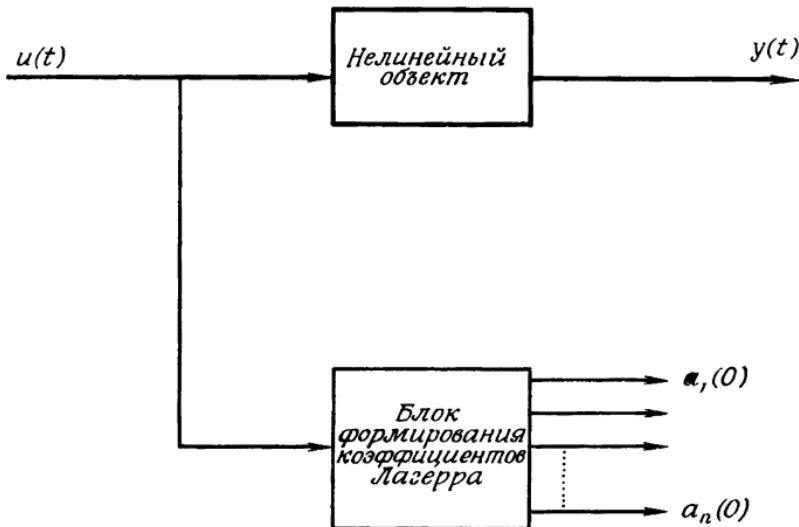
$$y(t) = \lim_{n \rightarrow \infty} F(a_1, a_2, \dots, a_n), \quad (4.56)$$

где a_i являются коэффициентами разложения в ряд сигнала $u(t)$.

Следующий шаг состоит в разложении F в ряд по полиномам Эрмита, так что

$$y(t) = \lim_{n \rightarrow \infty} \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \dots \sum_{k=1}^{\infty} b_{ij\dots k} H_i(a_1) H_j(a_2) \dots \\ \dots H_k(a_n) \exp \left[-\frac{a_1^2 + a_2^2 + \dots + a_n^2}{2} \right]. \quad (4.57)$$

Так как это разложение отражает связь между входным и выходным сигналами произвольного нелинейного объекта,



Ф и г. 4.10

та, то по такой модели можно идентифицировать неизвестный объект. Если входной сигнал является белым шумом, то благодаря ортогональности такая модель в принципе может быть легко построена.

Эта модель описывается следующим образом. Без особого изменения по сравнению с табл. 3.1 функции Лагерра можно записать как

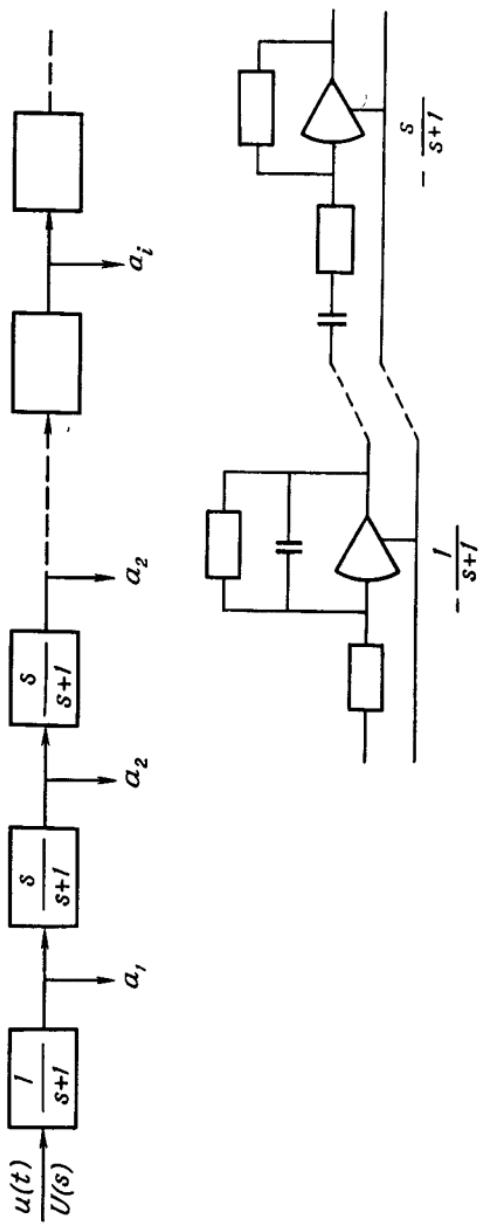
$$l_k(t) = \frac{1}{k!} \frac{d^k}{dt^k} (t^k e^{-t}). \quad (4.58)$$

Преобразование Лапласа функций Лагерра имеет вид

$$L_k(s) = \frac{s^k}{(s+1)^{k+1}}. \quad (4.59)$$

Коэффициенты разложения в ряд по функциям Лагерра могут быть легко найдены с помощью простых ступенчатых схем или аналоговых элементов. Соответствующий фильтр вместе с его реализацией показан на фиг. 4.11. Коэффициенты Лагерра равны

$$a_i(t) = \int_0^\infty u(t-\tau) l_i(\tau) d\tau$$



Ф и г. 4.41.

или при $t = 0$

$$a_i(0) = \int_0^{\infty} u(-\tau) l_i(\tau) d\tau.$$

Следовательно, коэффициенты $a_i(0)$ определяют предысторию $u(t)$.

В качестве последнего примера рассмотрим модели, использующие разложения в ряд по полиномам Чебышева. Эти разложения можно использовать только при ограниченном входном сигнале. Для нелинейного безынерционного объекта

$$y(t) = f[u(t)]. \quad (4.60)$$

В этом случае

$$y(t) = \sum_{i=0}^{\infty} a_i T_i \left(\frac{u(t)}{u_0} \right), \quad (4.61)$$

где u_0 представляет собой максимальную величину входного сигнала. Это разложение хорошо подходит для синусоидальных входных сигналов. Если $u = u_0 \cos \omega t$, то для безынерционного объекта имеем

$$y(t) = \sum_{i=1}^n a_i T_i(\cos \omega t) = \sum_{i=1}^n a_i \cos i\omega t. \quad (4.62)$$

Коэффициенты полиномиального разложения подобны амплитудам высших гармоник входного сигнала.

Такое представление основано на следующем рекуррентном соотношении для полиномов Чебышева

$$T_k(u) = 2uT_{k-1}(u) - T_{k-2}(u). \quad (4.63)$$

Допустим, что T_{k-1} и T_{k-2} равны соответственно $\cos(k-1)\omega t$ и $\cos(k-2)\omega t$. Тогда

$$\begin{aligned} T_k(u) &= 2\cos \omega t \cos(k-1)\omega t - \cos\{(k-1)\omega t - \omega t\} = \\ &= \cos \omega t \cos(k-1)\omega t - \sin(k-1)\omega t \sin \omega t = \\ &= \cos k\omega t. \end{aligned}$$

Подстановка $u = \cos \omega t$ в $T_0(u)$ и $T_1(u)$ показывает, что эта формула справедлива в случае $k = 0$ и $k = 1$, а следовательно, и для любого k .

Для анализа систем наиболее удобны разложения в ряды Вольтерра. Ортогональные ряды имеют преимущества при идентификации нелинейных объектов. Они приводят к достаточно простым моделям, описывающим поведение системы. Однако ортогональные ряды трудно преобразовать в ряды Вольтерра.

Описание в пространстве состояний. Как и в предыдущем разделе, представление нелинейного объекта в пространстве состояний может иметь вид

$$\begin{aligned}\dot{\mathbf{x}}(t) &= \mathbf{f}[\mathbf{x}(t), \mathbf{u}(t), \mathbf{a}(t), \mathbf{v}(t), t], \\ \mathbf{y}(t) &= \mathbf{g}[\mathbf{x}(t), \mathbf{u}(t), \mathbf{n}(t), t],\end{aligned}\quad (4.64)$$

где $\mathbf{x}(t)$ — вектор состояния; $\mathbf{u}(t)$ — наблюдаемый входной сигнал; $\mathbf{a}(t)$ — вектор параметров объекта; $\mathbf{v}(t)$ — входной шум (неизмеряемый входной сигнал); $\mathbf{n}(t)$ — вектор помех наблюдений (или измерений); $\mathbf{y}(t)$ — наблюдаемый выходной сигнал.

4.5. УПРАВЛЯЕМОСТЬ, НАБЛЮДАЕМОСТЬ, ИДЕНТИФИЦИРУЕМОСТЬ

Эти понятия были введены в работе Калмана [32] и затем рассматривались многими авторами (см., например, [16, 37]). Для более полного ознакомления с предметом читателю следует обратиться к работам, список которых приведен в конце этой главы. Здесь изложены только основные результаты для объектов, описываемых линейными дифференциальными уравнениями с постоянными коэффициентами. Случай систем с переменными коэффициентами рассмотрен в работе [36]. Пусть объект описывается следующими уравнениями:

$$\begin{aligned}\mathbf{x}(k+1) &= \mathbf{A}\mathbf{x}(k) + \mathbf{B}\mathbf{u}(k), \\ \mathbf{y}(k) &= \mathbf{C}\mathbf{x}(k),\end{aligned}\quad (4.65)$$

где \mathbf{x} — вектор размерности n .

Объект называется *управляемым*, если можно найти такой (быть может, неограниченный) вектор управления, который из произвольного начального состояния переводит систему в произвольное конечное состояние за огра-

ниченнное время. Таким образом, необходимо найти условие, при котором можно определить управление, которое переводит систему из состояния $x(0)$ в заданное состояние $x(n)$.

$$x(1) = Ax(0) + Bu(0),$$

$$x(2) = Ax(1) + Bu(1) = A^2x(0) + ABu(0) + Bu(1),$$

.....

$$x(n) = A^n x(0) + A^{n-1}Bu(0) + \dots + Bu(n-1),$$

или

$$x(n) - A^n x(0) = [A^{n-1}B | A^{n-2}B | \dots | B] \begin{bmatrix} u(0) \\ u(1) \\ \vdots \\ u(n-1) \end{bmatrix}. \quad (4.66)$$

Поскольку A , $x(0)$ и $x(n)$ известны, левая часть уравнения (4.66) определена. Единственное решение u существует только тогда, когда матрица

$$[A^{n-1}B | A^{n-2}B | \dots | B]$$

имеет ранг n . В этом случае (A, B) называют управляемой парой.

Объект называется *наблюдаемым*, если по измерениям выходного сигнала объекта можно определить его состояние. Таким образом, необходимо найти условие, при котором по измерениям y можно определить $x(0)$.

$$y(0) = Cx(0),$$

$$y(1) = Cx(1) = CAx(0),$$

.....

$$y(n-1) = CA^{n-1}x(0),$$

или, транспонируя, имеем

$$[y'(0) | y'(1) | \dots | y'(n-1)] =$$

$$= x'(0) [C' | A'C' | \dots | A'^{n-1}C']. \quad (4.67)$$

Так как векторы y известны, единственное решение $x(0)$ существует только тогда, когда матрица

$$[C' | A'C' | \dots | A'^{n-1}C']$$

имеет ранг n . В этом случае (A , C) называются наблюдаемой парой. Как показано Брокеттом [10], введение обратной связи может отразиться на наблюдаемости объекта.

Объект называется *идентифицируемым*, если по измерениям координат состояния объекта можно определить матрицу системы A .

$$\begin{aligned}x(1) &= Ax(0), \\x(2) &= A^2x(0), \\&\dots \dots \dots \\x(n) &= A^n x(0),\end{aligned}$$

или

$$\begin{aligned}[x(1) | x(2) | \dots | x(n)] &= \\&= A[x(0) | Ax(0) | \dots | A^{n-1}x(0)].\end{aligned}\quad (4.68)$$

Так как векторы x известны, единственное решение для A существует только тогда, когда матрица

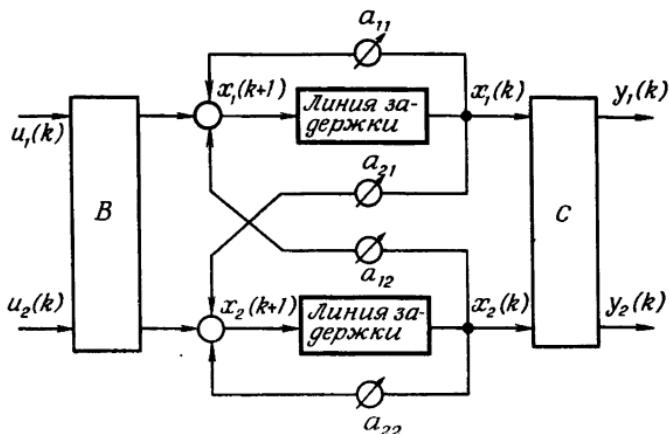
$$[x(0) | Ax(0) | \dots | A^{n-1}x(0)]$$

имеет ранг n .

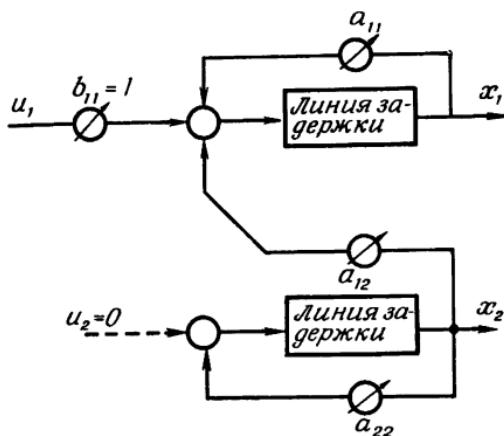
В литературе можно найти более детальное рассмотрение этих понятий. Так, например, различают управляемость по состоянию и по выходу, полную и общую, сильную и слабую. Наблюдаемость может быть полной и общей [36]. В работе [18] введено понятие δ -управляемости, когда система из начального состояния $x(0)$ может быть переведена в δ -окрестность произвольного заданного состояния. Аоки [2] распространил понятия управляемости, наблюдаемости и идентифицируемости на стохастический случай. Вопросы идентифицируемости рассматривались также в работах [7, 52]. С идентифицируемостью тесно связано понятие различных классов [47—49].

Примеры. Для иллюстрации этих понятий выберем простую систему второго порядка

$$\begin{aligned}x(k+1) &= Ax(k) + Bu(k), \\y(k) &= Cx(k)\end{aligned}$$



Ф и г. 4.12.



Ф и г. 4.13.

(фиг. 4.12), где

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \quad \mathbf{C} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}.$$

Рассмотрим управляемость системы. Имеем

$$[\mathbf{AB} : \mathbf{B}] = \begin{bmatrix} a_{11} & 0 & 1 & 0 \\ a_{21} & 0 & 0 & 0 \end{bmatrix}.$$

Система управляема, если ранг равен 2, т. е. когда $a_{21} \neq 0$, и неуправляема, если $a_{21} = 0$. Из фиг. 4.13 видно, что,

когда $a_{21} = 0$, отсутствует управление координатой x_2 . Отметим, что все остальные параметры a_{ij} могут равняться нулю, но система останется управляемой.

Перейдем к наблюдаемости. Имеем

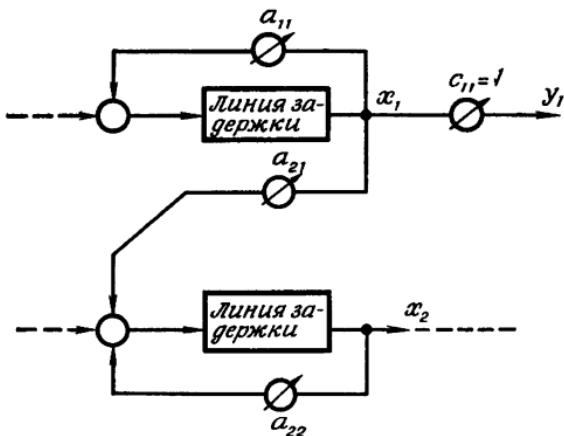
$$[\mathbf{C}' : \mathbf{A}'\mathbf{C}] = \begin{bmatrix} 1 & 0 & a_{11} & 0 \\ 0 & 0 & a_{12} & 0 \end{bmatrix}.$$

Система наблюдаема, если ранг равен 2, т. е. когда $a_{12} \neq 0$, и ненаблюдаема, когда $a_{12} = 0$. В этом случае выходная координата y не содержит информации о x_2 (фиг. 4.14). Снова отметим, что все остальные параметры могут равняться нулю, но система останется наблюдаемой.

Идентифицируемость:

$$[\mathbf{x}(0) : \mathbf{Ax}(0)] = \begin{bmatrix} x_1(0) & a_{11}x_1(0) + a_{12}x_2(0) \\ x_2(0) & a_{21}x_1(0) + a_{22}x_2(0) \end{bmatrix}.$$

Система идентифицируема, если ранг матрицы равен 2, и неидентифицируема, если определитель матрицы равен нулю. Для этого оба столбца матрицы должны быть линейно зависимы. Различают простейший случай, когда $x_1(0) = x_2(0) = 0$, т. е. объект, который находится в состоянии покоя, не может быть идентифицирован,



Ф и г. 4.14.

и нетривиальный случай, когда

$$\lambda \begin{bmatrix} x_1(0) \\ x_2(0) \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} x_1(0) \\ x_2(0) \end{bmatrix}, \text{ или}$$

$$[A - \lambda I] \begin{bmatrix} x_1(0) \\ x_2(0) \end{bmatrix} = 0.$$

В этом случае нужно найти собственные значения λ_1 и λ_2 и соответствующие собственные векторы r_1 и r_2 . Если $x(0) = \alpha r_1$, то возбуждается только одна гармоника объекта $\exp[\lambda_1 t]$, а гармоника $\exp[\lambda_2 t]$ не идентифицируется. Если $x(0) = \beta r_2$, то может быть идентифицирована одна только гармоника $\exp[\lambda_2 t]$. Таким образом, объект идентифицируем только тогда, когда начальное условие $x(0)$ возбуждает все гармоники объекта.

Другие вопросы, связанные с идентифицируемостью, рассмотрены в гл. 11.

4.6. МОДЕЛИ

В предыдущих разделах приведено несколько способов описания характеристик объектов. Здесь будут кратко рассмотрены основанные на этих способах математические модели, или моделирующие схемы. Рассмотрение ведется в общем виде, так как выбор конкретной модели существенно зависит от объекта и цели идентификации.

Линейность динамики и линейность по параметрам

В разд. 4.1 различие между линейными и нелинейными моделями было проведено на основе их динамических характеристик, т. е. связи между *зависимыми (выходами) и независимыми (входами) функциями времени*. Для задач оценивания параметров более существенно исходить при этом из связи между *зависимыми переменными и параметрами*. Модель называется линейной по параметрам, если ошибка (обобщенная) является линейной функцией параметров.

Как видно из следующих примеров, эти два определения линейности непосредственно не связаны. Пусть u — входной сигнал объекта, а y — его выходной сигнал. Тогда «модель» формирования ошибки выходного сигнала модели по сравнению с выходным сигналом объекта может иметь следующий вид:

Динамические характеристики	Модель		Объект
	линейная	нелинейная	
линейные	$e = \dot{y} + \alpha y - u$	$e = y - w = y - \frac{u}{D + \alpha}$ $\dot{w} + \alpha w = u$	$\dot{y} + \alpha y = u$
нелинейные	$e = \dot{y} + \alpha y^3 - u$	$e = y - w = y - \mathcal{O}[u, \alpha]$ $\dot{w} + \alpha w^3 = u$	$\dot{y} + \alpha y^3 = u$

Использование таких двух подходов при различении линейных и нелинейных моделей может стать источником недоразумений, которые связаны со смешиванием понятий, относящихся к теории систем и регрессионному анализу¹⁾. Поэтому мы в дальнейшем под *линейными системами* будем понимать системы с линейными динамическими характеристиками.

Важность линейности по параметрам станет очевидной при рассмотрении задач оценивания. Рассмотрим типичный пример идентификации импульсной линейной системы, выходной сигнал которой искажен белым шумом измерений. Описание системы с помощью коэффициентов передаточной функции приводит к нелинейной регрессионной

¹⁾ Отметим, что другим источником недоразумений может стать использование термина «порядок». В регрессионном анализе этот термин означает наивысший показатель степени независимой переменной:

$$\begin{aligned} y &= \beta_0 + \beta_1 u_1 + n \\ y &= \beta_0 + \beta_1 u_1 + \beta_2 u_2 + \dots + \beta_m u_m + n \end{aligned} \quad \left. \right\} \text{модель первого порядка,}$$

$$y = \beta_0 + \beta_1 u_1 + \beta_2 u_1^2 \quad \text{модель второго порядка.}$$

модели, тогда как использование коэффициентов обобщенной модели или ординат весовой функции ведет к линейной по параметрам задаче оценивания. Имеет смысл попробовать найти такое преобразование переменных, которое приводит к линейной по параметрам модели. Продемонстрируем это на простых примерах:

$$z = \frac{\alpha_2 x_2 + x_1}{\alpha_1 x_1 x_2} \Rightarrow \left\{ \begin{array}{l} z = y \\ \frac{1}{x_1} = u_1 \quad \frac{\alpha_2}{\alpha_1} = \beta_2 \\ \frac{1}{x_2} = u_2 \quad \frac{1}{\alpha_1} = \beta_1 \end{array} \right\} \quad \text{Преобразование обращения}$$

↓

$$y = \beta_1 u_1 + \beta_2 u_2,$$

$$z = c x_1^{\alpha_1} x_2^{\alpha_2} \Rightarrow \left\{ \begin{array}{l} \log z = y \quad \log c = \beta_0 \\ \log x_1 = u_1 \quad \alpha_1 = \beta_1 \\ \log x_2 = u_2 \quad \alpha_2 = \beta_2 \end{array} \right\} \quad \text{Логарифмическое преобразование}$$

↓

$$y = \beta_0 + \beta_1 u_1 + \beta_2 u_2.$$

Нелинейные выражения, которые путем преобразования переменных могут быть сделаны линейными по параметрам, называются *скрыто линейными*. Если линеаризация невозможна, говорят о *существенно нелинейной модели*. Использование преобразований может оказаться полезным даже в случае существенно нелинейной системы [14].

Линейные модели «вход — выход»

Линейные модели, естественно, представляют собой наиболее разработанный раздел теории идентификации. Здесь будут рассматриваться линейные объекты в линейной среде, т. е. в среде, которую можно описать линейными стохастическими моделями. Во многих задачах теории управления свойства среды имеют такое же значение, как и динамика системы, поскольку сама постановка задачи управления в первую очередь связана с наличием помех.

В рамках классического подхода к описанию систем модель может задаваться передаточной или весовой функциями. Однако во многих случаях в настоящее время применяется модель состояний, т. е. параметрическая модель. Естественно, возникает ряд вопросов:

1) Пусть требуется оценить весовую функцию. Что лучше, идентифицировать ее непосредственно или построить параметрическую модель и по ней рассчитать весовую функцию?

2) Допустим, что необходима параметрическая модель. Следует ли непосредственно подгонять ее параметры или лучше сначала определить весовую функцию, а уже затем по этой весовой функции построить модель?

3) Поскольку в параметрическую модель явно входит порядок системы, что произойдет, если порядок выбран неправильно?

Общих ответов на эти вопросы пока еще нет. Некоторые частные случаи рассмотрены Густавссоном (1969 г.) в связи с задачами идентификации атомного реактора и динамики ректификационной колонны, а также на модельных примерах.

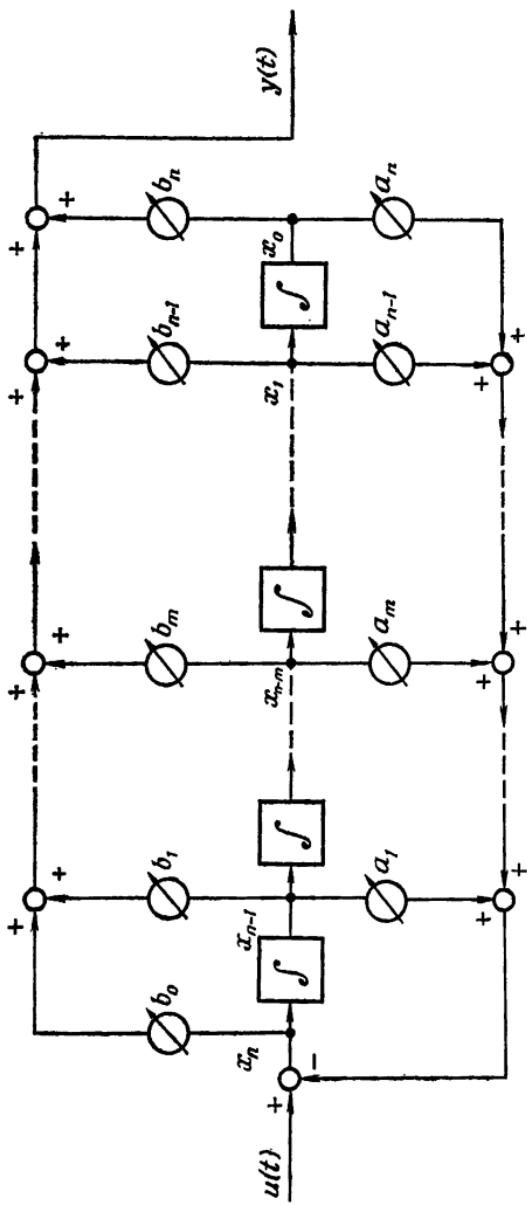
Для линейных объектов существуют хорошо известные аналоговые методы непосредственного моделирования обыкновенных дифференциальных уравнений, например уравнения вида

$$\frac{d^n y}{dt^n} + \dots + a_{n-1} \frac{dy}{dt} + a_n y = b_0 \frac{d^n u}{dt^n} + \dots + b_{n-1} \frac{du}{dt} + b_n u.$$

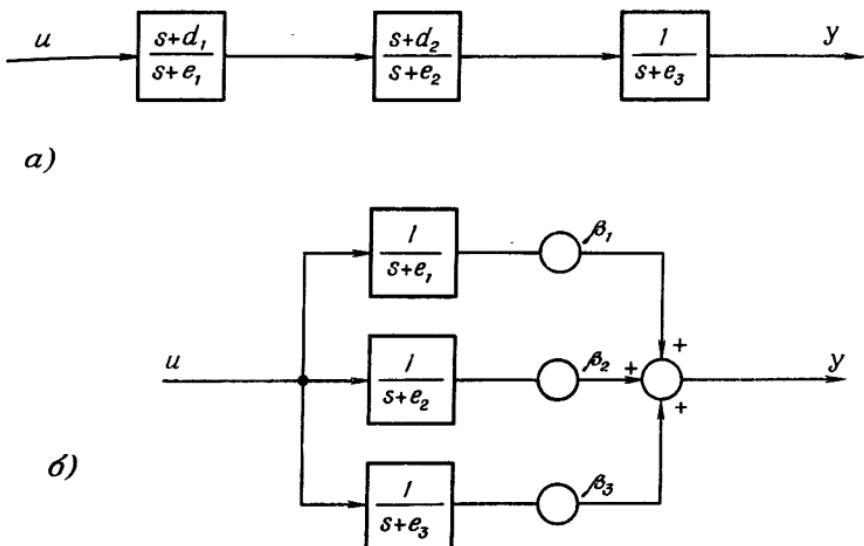
Этому уравнению соответствует передаточная функция

$$H(s) = \frac{Y(s)}{U(s)} = \frac{b_0 s^n + b_1 s^{n-1} + \dots + b_{n-1} s + b_n}{s^n + a_1 s^{n-1} + \dots + a_{n-1} s + a_n} \quad (4.69)$$

и блок-схема, приведенная на фиг. 4.15. Если в качестве переменных, описывающих состояние, выбраны x_{n-1}, \dots, x_0 , то их величины могут быть измерены непосредственно на модели. Если же состояние системы описывается выходом y и его производными, то их можно представить как линейные комбинации x -переменных. Если $a_i = 0$ для $i = 1, \dots, m$, то $x_i = y^{(i)}$. Заметим, что $H(s)$ нелинейна по параметрам a_i .



Ф и г. 4.15



Ф и г. 4.16.

Линейный (линеаризованный) объект, например, с передаточной функцией

$$H(s) = \frac{(s+d_1)(s+d_2)}{(s+e_1)(s+e_2)(s+e_3)} \quad (4.70)$$

можно представить в виде соединений более элементарных звеньев (фиг. 4.16):

1) последовательное соединение звеньев с простыми передаточными функциями, каждая из которых имеет один или два полюса и, быть может, нуль, совпадающие с полюсами и нулем передаточной функции системы

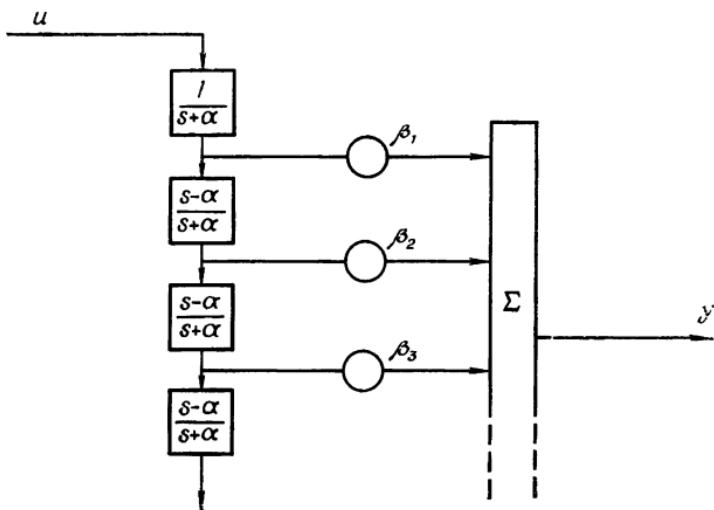
$$H(s) = \frac{s+d_1}{s+e_1} \frac{s+d_2}{s+e_2} \frac{1}{s+e_3}; \quad (4.71)$$

2) параллельное соединение звеньев с простыми передаточными функциями с весовыми коэффициентами β_i

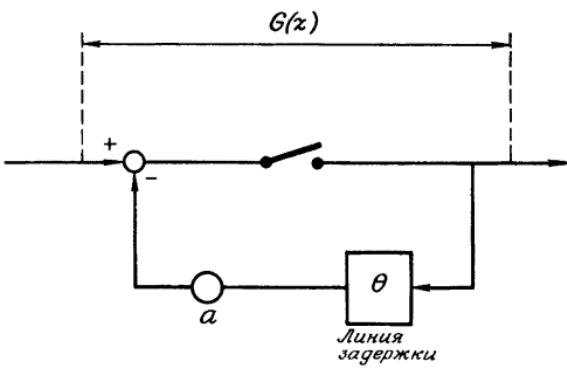
$$H(s) = \frac{\beta_1}{s+e_1} + \frac{\beta_2}{s+e_2} + \frac{\beta_3}{s+e_3}; \quad (4.72)$$

3) параллельное соединение звеньев с ортогональными передаточными функциями

$$H(s) = \sum_{i=1}^n \beta_i F_i(s), \quad (4.73)$$



Ф и г. 4.17.



Ф и г. 4.18.

например функциями Лагерра:

$$F_i(s) = \frac{1}{s+\alpha} \left(\frac{s-\alpha}{s+\alpha} \right)^i.$$

Этому случаю соответствует блок-схема, изображенная на фиг. 4.17.

Рассмотренный метод моделирования можно также применить для описания систем, в которых информация об имеющихся сигналах представлена их выборочными значениями. На фиг. 4.18 показан пример простой системы

первого порядка с передаточной функцией

$$H(z) = \frac{z}{z+a}, \quad z = e^{s\theta}.$$

Объект с передаточной функцией высокого порядка

$$H(z) = \frac{b_0 + b_1 z + \dots + b_n z^n}{1 + a_1 z + \dots + a_n z^n} \quad (4.74)$$

можно представить, используя линии задержки, как на фиг. 4.15, и, кроме того, в виде разложения на элементарные звенья, как на фиг. 4.16. Легко учесть наличие чистого запаздывания:

$$G(z) = z^{-k} H(z).$$

Этот способ описания хорошо подходит для решения задач управления. Если на объект действует коррелированный аддитивный шум, то свойства шума также могут быть описаны с помощью линейной системы, на вход которой подается белый шум.

Весовая функция $h(t)$ дает общее непараметрическое описание линейного объекта, так как вся необходимая априорная информация сводится к тому, что объект линеен. Легко учесть чистое запаздывание, которое можно реализовать несколькими способами. Однако модель должна позволять строить свертку $h(t)$ с $u(t)$. Это приводит к следующим возможностям:

1) $h(t)$ представлена фотоснимком, а $u(t)$ имеется в виде записи на фотопленке.

2) $h(t)$ определяется набором коэффициентов ортогонального разложения (см. гл. 3):

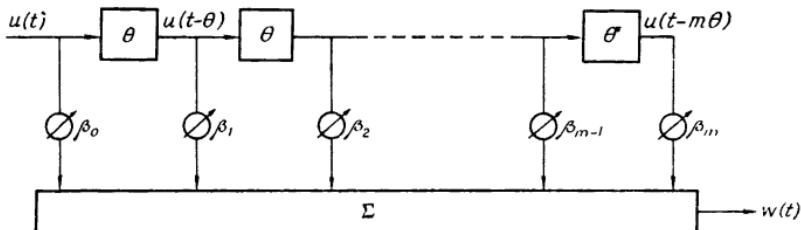
$$h(t) = \sum \beta_i g_i(t), \quad (4.75)$$

где

$$\int_a^b g_i(t) g_j(t) dt = \delta_{ij},$$

$$\delta_{ij} = \begin{cases} 1, & \text{если } i = j, \\ 0, & \text{если } i \neq j. \end{cases}$$

Пригодность описания, задаваемого формулой (4.75), в большой степени зависит от того, насколько выбранная



Ф и г. 4.19.

система ортогональных функций «подходит» к весовой функции. В благоприятном случае уже несколько отличных от нуля коэффициентов β_i могут дать хорошую аппроксимацию, в противном случае может потребоваться достаточно много членов разложения. Было предложено [29, 61] выбирать функции $g_i(t)$ зависящими от дополнительных параметров $g_i(t; \alpha)$. При этом необходимо, усложнив процедуру оценивания, определить наилучшие значения параметров β и α . Были предприняты попытки распространить ортогональные модели на многомерные объекты [8].

3) $h(t)$ задается набором своих значений $\beta_0 = h(0)$, $\beta_1 = h(\theta)$, Это сводится к синтезу с использованием линии задержки [23]. В соответствии с блок-схемой, изображенной на фиг. 4.19, интеграл свертки аппроксируется выражением

$$w(t) = \sum_{i=0}^n \beta_i u(t - i\theta). \quad (4.76)$$

Недостатком этого способа является необходимость определения достаточно большого числа параметров.

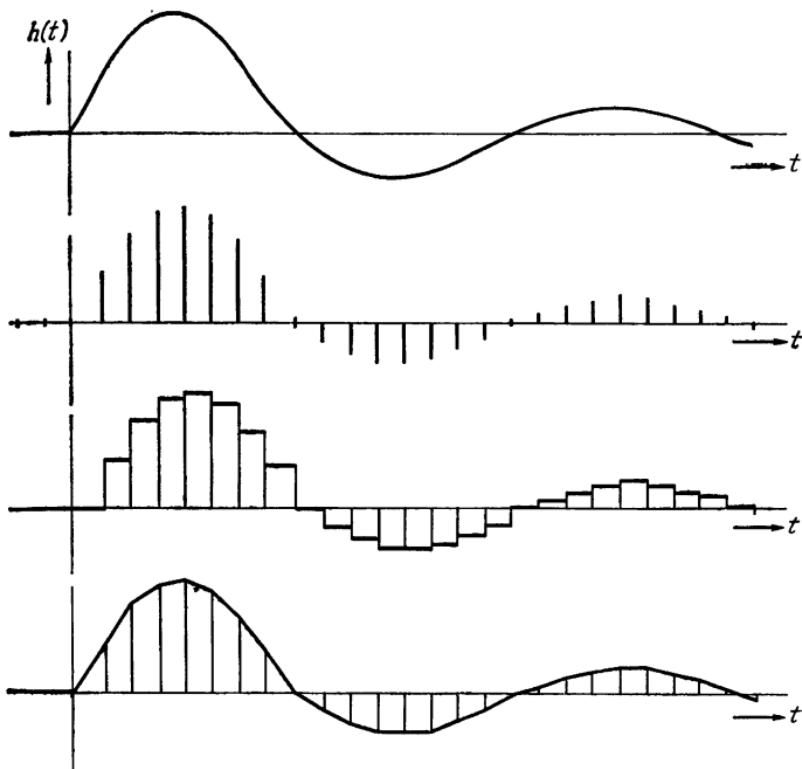
Весовую функцию $h(t)$ можно аппроксимировать по-разному (фиг. 4.20):

как непрерывную функцию $h(t)$;

$$h_1(t) = \sum h(t) \delta(t - i\theta);$$

$$h_2(t) = h(i\theta) \text{ при } 0 \leq t - i\theta < \theta;$$

$$h_3(t) = h(i\theta) + \frac{t - i\theta}{\theta} \{h[(i+1)\theta] - h[i\theta]\} \text{ при } 0 \leq t - i\theta < \theta.$$



Ф и г. 4.20.

Линейность по параметрам дает возможность представить выборочные данные в виде

$$\mathbf{y} = \mathbf{U}\mathbf{b}, \quad (4.77)$$

где

$$\mathbf{y} = \begin{bmatrix} y(1) \\ \vdots \\ y(k) \end{bmatrix}, \quad \mathbf{U} = \begin{bmatrix} u_0(1) \dots u_m(1) \\ \vdots \\ u_0(k) \dots u_m(k) \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} b_0 \\ \vdots \\ b_m \end{bmatrix}.$$

Отметим, что

$$\mathbf{U} = [u_0, \dots, u_m],$$

причем не было сделано никаких специальных предположений относительно связи между u_i и u_j . Они могут представлять собой одинаковые последовательности, сдвинутые во времени относительно друг друга, а могут быть и выходными сигналами нелинейных импульсных фильтров.

тров. В первом случае

$$\mathbf{U} = \begin{bmatrix} u(1) & 0 & 0 & \dots & 0 \\ u(2) & u(1) & 0 & \dots & 0 \\ u(3) & u(2) & u(1) & \dots & 0 \\ \vdots & & & & \vdots \\ u(k) & u(k-1) & \dots & u(k-m) & \end{bmatrix}. \quad (4.78)$$

Здесь предполагается, что $u(i) = 0$ для $i \leq 0$.

Кроме того, можно использовать другую формулу

$$\mathbf{y} = \mathbf{B}\mathbf{u}, \quad (4.79)$$

где

$$\mathbf{y} = \begin{bmatrix} y(1) \\ \vdots \\ y(k) \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} b_0 & 0 & 0 & \dots & 0 \\ b_1 & b_0 & 0 & \dots & 0 \\ b_2 & b_1 & b_0 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ b_m & & & \ddots & \vdots \\ 0 & b_m & & & \vdots \\ 0 & 0 & b_m & \dots & b_0 \end{bmatrix}, \quad \mathbf{u} = \begin{bmatrix} u(1) \\ \vdots \\ u(k) \end{bmatrix}.$$

Это дает возможность сразу записать выражение для ковариации \mathbf{y} , а именно

$$\text{cov}[\mathbf{y}] = \mathbf{B} \text{cov}[\mathbf{u}] \mathbf{B}' . \quad (4.80)$$

Линейные модели состояний в канонической форме¹⁾

В заключение наше внимание будет сконцентрировано на важнейшем типе моделей состояний $S(A, B, C, D)$, определяемых уравнениями

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{u}, \quad (4.81)$$

$$\mathbf{y} = \mathbf{C}\mathbf{x} + \mathbf{D}\mathbf{u},$$

где \mathbf{x} — вектор размерности n , входной сигнал \mathbf{u} — вектор размерности p и выходной сигнал \mathbf{y} — вектор размерности r . Известно, что системы $S(A, B, C, D)$ и $S(TAT^{-1}, TB, CT^{-1}, D)$, где T — невырожденная матри-

¹⁾ Этот раздел взят из работы [6].

ца, эквивалентны в том смысле, что они имеют одинаковое соотношение между входом и выходом. Легко проверить, что две системы $S(A, B, C, D)$ и $S(\tilde{A}, \tilde{B}, \tilde{C}\tilde{D})$ эквивалентны в том же смысле, если

$$D = \tilde{D}, \\ CA^k B = \tilde{C} \tilde{A}^k \tilde{B}, \quad k = 0, 1, \dots, n. \quad (4.82)$$

Связь между различными представлениями систем была выяснена в работе Калмана [33]. Весовая и передаточная функции описывают только часть системы S , которая является полностью управляемой. Отсюда понятно, что только полностью управляемая и полностью наблюдаемая часть модели состояний $S(A, B, C, D)$ может быть определена по результатам измерений входного и выходного сигналов. По модели состояний легко можно получить весовую и передаточную функции. Задача определения модели состояний по весовой функции гораздо сложнее, даже если не обращать внимания на то, что по весовой функции можно определить только модель управляемой и наблюдаемой части системы. Задача о выборе модели состояний наименьшего порядка по заданной весовой функции была решена Хо и Калманом [25] (см. также [34, 11]). Решение опять оказалось не единственным. Модель $S(A, B, C, D)$ включает

$$N_1 = n^2 + np + nr + pr \quad (4.83)$$

параметров. То, что соотношение между входом и выходом инвариантно относительно линейных преобразований координат, означает, что по измерениям входных и выходных сигналов нельзя оценить все N_1 параметров. Таким образом, для получения единственного решения, а также для построения эффективных алгоритмов важно найти описание системы, содержащее минимальное число параметров, т. е. *каноническое представление*.

Канонические представления линейных детерминированных моделей. Каноническая запись линейных систем рассматривалась Калманом [33]. В том случае, когда все собственные значения матрицы A различны, каноническая форма записи получается следующим образом. Соответствующим выбором координат матрица A преоб-

разуется к диагональному виду

$$\dot{\mathbf{x}} = \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{bmatrix} \mathbf{x} + \begin{bmatrix} b_{11} & b_{12} & \dots & b_{1p} \\ b_{21} & b_{22} & \dots & b_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ b_{n1} & b_{n2} & \dots & b_{np} \end{bmatrix} \mathbf{u}, \quad (4.84)$$

$$\mathbf{y} = \begin{bmatrix} c_{11} & c_{12} & \dots & c_{1n} \\ c_{21} & c_{22} & \dots & c_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ c_{r1} & c_{r2} & \dots & c_{rn} \end{bmatrix} \mathbf{x} + \begin{bmatrix} d_{11} & d_{12} & \dots & d_{1p} \\ d_{21} & d_{22} & \dots & d_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ d_{r1} & d_{r2} & \dots & d_{rp} \end{bmatrix} \mathbf{u}.$$

Это представление содержит $n + np + nr + pr$ параметров, n из которых являются избыточными, так как все координаты в пространстве состояний могут быть нормированы без изменения уравнений состояний. Таким образом, уравнение состояний может быть охарактеризовано

$$N_2 = n(p + r) + pr \quad (4.85)$$

параметрами. Так как система полностью управляема и наблюдаема, в каждой строке матрицы **B** и в каждом столбце матрицы **C** имеется по меньшей мере один ненулевой элемент. Таким образом, избыточность в (4.84) можно уменьшить, наложив дополнительные условия

$$\max_j b_{ij} = 1, \quad i = 1, 2, \dots, n, \quad (4.86)$$

$$\sum_j |b_{ij}| = 1, \quad i = 1, 2, \dots, n, \quad (4.87)$$

или аналогичные условия на матрицу **C**. Если матрица **A** имеет кратные собственные значения, то задача отыскания параметрического описания с минимальным числом параметров усложняется. Если **A**, кроме того, еще и является циклической (т. е. существует такой вектор \mathbf{x} , что векторы $\mathbf{x}, \mathbf{Ax}, \mathbf{A}^2\mathbf{x}, \dots, \mathbf{A}^{n-1}\mathbf{x}$ образуют базис в n -мерном пространстве), то параметрическое описание с минимальным

числом параметров задается уравнениями

$$\begin{aligned} \dot{x} &= \begin{bmatrix} -a_1 & 1 & 0 & \dots & 0 \\ -a_2 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -a_{n-1} & 0 & 0 & \dots & 1 \\ -a_n & 0 & 0 & \dots & 0 \end{bmatrix} x + \begin{bmatrix} b_{11} & \dots & b_{1p} \\ \vdots & & \vdots \\ b_{n1} & \dots & b_{np} \end{bmatrix} u, \\ y &= \begin{bmatrix} c_{11} & \dots & c_{1n} \\ \vdots & & \vdots \\ c_{r1} & \dots & c_{rn} \end{bmatrix} x + \begin{bmatrix} d_{11} & \dots & d_{1p} \\ \vdots & & \vdots \\ d_{r1} & \dots & d_{rp} \end{bmatrix} u, \quad (4.88) \end{aligned}$$

где на элементы матриц B и C наложены n дополнительных условий типа (4.86) или (4.87). В случае объекта с одним выходом эти дополнительные условия вводятся заданием всех компонент вектора C , например $C' = [1 \ 0 \ \dots \ 0]$. Каноническое представление тогда принимает вид

$$\begin{aligned} Y(s) &= \left[d_{11} + \frac{b_{11}s^{n-1} + b_{21}s^{n-2} + \dots + b_{n1}}{s^n + a_1s^{n-1} + \dots + a_n} \right] U_1(s) + \dots + \\ &\quad + \left[d_{1p} + \frac{b_{1p}s^{n-1} + b_{2p}s^{n-2} + \dots + b_{np}}{s^n + a_1s^{n-1} + \dots + a_n} \right] U_p(s), \quad (4.89) \end{aligned}$$

где через Y и U_i обозначены преобразования Лапласа y и u_i . Таким образом, каноническое представление объекта n -го порядка с p входами и одним выходом может быть записано в виде

$$\begin{aligned} \frac{d^n y}{dt^n} + a_1 \frac{d^{n-1} y}{dt^{n-1}} + \dots + a_n y &= \\ &= \left[b'_{01} \frac{d^n u_1}{dt^n} + \dots + b'_{n1} u_1 \right] + \dots + \\ &\quad + \left[b'_{0p} \frac{d^n u_p}{dt^n} + \dots + b'_{np} u_p \right]. \quad (4.90) \end{aligned}$$

Аналогичный вид имеет представление для систем с несколькими выходами:

$$\begin{aligned} \frac{d^n y}{dt^n} + A_1 \frac{d^{n-1} y}{dt^{n-1}} + \dots + A_n y &= \\ &= \left[B_{01} \frac{d^n u_1}{dt^n} + \dots + B_{n1} u_1 \right] + \dots + \\ &\quad + \left[B_{0p} \frac{d^n u_p}{dt^n} + \dots + B_{np} u_p \right]. \quad (4.91) \end{aligned}$$

Эта форма записи уравнений состояний наряду с другими использована в работах [50, 62]. Определение порядка объекта (4.91), в общем случае отличного от n , и преобразование (4.91) для уравнения состояний проделано в работе [55]. Канонические представления линейных многомерных систем рассматривались также в работах [12, 39]. Аналогичные результаты справедливы и для импульсных систем.

В том случае, когда матрица A имеет кратные собственные значения и не циклическая, не ясно, что означает «минимально параметрическое представление». Конечно, матрица A всегда может быть преобразована к жордановой форме [20]. Так как среди собственных значений матрицы A есть совпадающие, то, строго говоря, матрица A может быть описана менее чем n параметрами. Единицы над диагональю в жордановой форме могут быть записаны многими способами в зависимости от способа объединения, что ведет к различным структурам.

Канонические представления линейных стохастических моделей. Рассмотрим канонические представления стохастических систем. Для того чтобы избежать трудностей, связанных с непрерывным белым шумом, приведем результаты для импульсных систем. Однако аналогичные результаты справедливы и для непрерывных систем. Рассмотрим систему

$$\begin{aligned} \mathbf{x}(k+1) &= \mathbf{Ax}(k) + \mathbf{Bu}(k) + \mathbf{v}(k), \\ (4.92) \end{aligned}$$

$$\mathbf{y}(k) = \mathbf{Cx}(k) + \mathbf{Du}(k) + \mathbf{n}(k),$$

где k принимает целочисленные значения. Вектор состояний \mathbf{x} , входной сигнал \mathbf{u} и выходной сигнал \mathbf{y} имеют размерности n , p и r соответственно; $\{\mathbf{v}(k)\}$ и $\{\mathbf{n}(k)\}$ представляют собой последовательности независимых случайных векторов с нулевыми средними и ковариациями \mathbf{V} и \mathbf{N} . В силу симметрии ковариационных матриц модель (4.92) содержит

$$\begin{aligned} N_3 &= n^2 + np + nr + pr + \frac{1}{2}n(n+1) + \frac{1}{2}r(r+1) = \\ &= n\left(\frac{3}{2}n + \frac{1}{2} + p + r\right) + r\left(p + \frac{r}{2} + \frac{1}{2}\right) \quad (4.93) \end{aligned}$$

параметров. Говорят, что две модели типа (4.92) эквивалентны, если: 1) уравнения состояний совпадают при $n = 0$ и $v = 0$ и 2) стохастические свойства выходных сигналов одинаковы при $u = 0$. Число параметров в А, В и С может быть сокращено, как было описано выше. Остается уменьшить число параметров, описывающих помехи. Это достигается использованием теоремы фильтрации Калмана. Из этой теоремы вытекает, что объект с одним выходом может быть описан формулами

$$\begin{aligned}\hat{x}(k+1) &= Ax(k) + Bu(k) + Kr(k), \\ y(k) &= Cx(k) + Du(k) + r(k),\end{aligned}\quad (4.94)$$

где $\hat{x}(k)$ означает условное среднее $x(k)$ при заданных $y(k-1)$, $y(k-2)$, ... и $\{r(k)\}$ — последовательность независимых одинаково распределенных случайных величин с нулевым средним и ковариацией R .

Модель с одним выходом рассмотрена в работе [3]. В работе [31] представление (4.94) названо *обновленным представлением* объекта. Подробное обсуждение можно найти у Острэма [5]. Модель (4.94) использована также в [44].

Отметим, что, если модель (4.94) известна, задачи устойчивой фильтрации и оценивания решаются очень легко. Поскольку K — усиление фильтра, нет необходимости решать уравнение Риккати. Отметим также, что физически состояние модели (4.94) интерпретируется как условное математическое ожидание состояния модели (4.92). Если матрица А представлена в диагональной форме, а на матрицы В и С наложены условия типа (4.86), то модель (4.94) является каноническим представлением, которое содержит N_4 параметров:

$$N_4 = n(p + 2r) + r \left(p + \frac{r}{2} + \frac{1}{2} \right). \quad (4.95)$$

Для систем с одним выходом, когда $C' = [1 \ 0 \ \dots \ 0]$, уравнение (4.94) преобразуется к виду

$$\begin{aligned}y(k) + a_1y(k-1) + \dots + a_ny(k-n) &= \\ &= [b'_{01}u_1(k) + \dots + b'_{n1}u_1(k-n)] + \dots + \\ &\quad + [b'_{0p}u_p(k) + \dots + b'_{np}u_p(k-n)] + \\ &\quad + r(k) + c_1r(k-1) + \dots + c_nr(k-n).\end{aligned}\quad (4.96)$$

Вводя оператор сдвига z , определяемый формулой

$$zy(k) = y(k+1), \quad (4.97)$$

полиномиальные операторы

$$\begin{aligned} A(z) &= z^n + a_1 z^{n-1} + \dots + a_n, \\ B_i(z) &= b'_0 z^n + b'_1 z^{n-1} + \dots + b'_{ni} \quad (i=1, 2, \dots, p), \\ C(z) &= z^n + c_1 z^{n-1} + \dots + c_n \end{aligned} \quad (4.98)$$

и соответствующие обратные полиномы

$$\begin{aligned} A^*(z^{-1}) &= 1 + a_1 z^{-1} + \dots + a_n z^{-n} = z^{-n} A(z), \\ B_i^*(z^{-1}) &= b'_0 + b'_1 z^{-1} + \dots + b'_{ni} z^{-n} = z^{-n} B_i(z), \\ C^*(z^{-1}) &= 1 + c_1 z^{-1} + \dots + c_n z^{-n} = z^{-n} C(z), \end{aligned} \quad (4.99)$$

можно переписать уравнение (4.96) в виде

$$A^*(z^{-1}) y(k) = \sum_{i=1}^p B_i^*(z^{-1}) u_i(k) + C^*(z^{-1}) r(k) \quad (4.96')$$

или

$$A(z) y(k) = \sum_{i=1}^p B_i(z) u_i(k) + C(z) r(k). \quad (4.96'')$$

Такая каноническая форма записи системы n -го порядка была введена в работе [4] и с тех пор широко используется. Соответствующая запись для многомерного случая получается, если интерпретировать y и u_i как векторы, а A , B_i и C — как полиномы с матричными коэффициентами. Такие модели рассмотрены в работах [15, 35, 50, 56]. Как альтернатива модели (4.96) в работах [9, 53] использована каноническая форма вида

$$\begin{aligned} y(k) &= \frac{B_1(z)}{A_1(z)} u_1(k) + \frac{B_2(z)}{A_2(z)} u_2(k) + \dots + \\ &\quad + \frac{B_p(z)}{A_p(z)} u_p(k) + \frac{C(z)}{A(z)} r(k). \end{aligned} \quad (4.100)$$

В каждом конкретном случае выбор структуры модели может заметно влиять на объем работы, связанной с решением задачи. Это подтверждается следующим примером.

Пример. Предположим, что конечной целью идентификации является получение прогноза с использованием калмановской фильтрации. Объект описывается уравнениями

$$\begin{aligned} \mathbf{x}(k+1) &= \mathbf{A}\mathbf{x}(k) + \mathbf{v}(k), \\ \mathbf{y}(k) &= \mathbf{C}\mathbf{x}(k) + \mathbf{n}(k), \end{aligned} \quad (4.101)$$

где $\{\mathbf{n}(k)\}$ и $\{\mathbf{v}(k)\}$ — дискретные белые шумы с ковариациями \mathbf{N} и \mathbf{V} . Тогда функция правдоподобия для задачи оценивания может быть записана в виде

$$\begin{aligned} -\log L = \frac{1}{2} \sum_{k=1}^n [\mathbf{v}'(k) \mathbf{V}^{-1} \mathbf{v}(k) + \mathbf{n}'(k) \mathbf{N}^{-1} \mathbf{n}(k)] + \\ + \frac{n}{2} \log (\det \mathbf{V}) (\det \mathbf{N}) + \text{const.} \end{aligned} \quad (4.102)$$

Здесь уравнения объекта рассматриваются как ограничения. Оценивание градиентов функции потерь приводит к двухточечной краевой задаче. К тому же в процессе идентификации для решения задачи калмановской фильтрации требуется решить уравнение Риккати. Допустим теперь, что структура идентифицируемого объекта имеет вид

$$\begin{aligned} \mathbf{w}(k+1) &= \mathbf{Aw}(k) + \mathbf{Kr}(k), \\ \mathbf{y}(k) &= \mathbf{Cw}(k) + \mathbf{r}(k). \end{aligned} \quad (4.103)$$

Тогда функция правдоподобия принимает вид

$$-\log L = \frac{1}{2} \sum_{k=1}^n \mathbf{r}'(k) \mathbf{R}^{-1} \mathbf{r}(k) + \frac{n}{2} \log \det \mathbf{R}. \quad (4.104)$$

В этом случае оценивание градиентов функции потерь сводится к простой задаче с начальным условием. После завершения идентификации устойчивый фильтр Калмана описывается уравнением

$$\hat{\mathbf{x}}(k+1) = \hat{\mathbf{A}}\hat{\mathbf{x}}(k) + \mathbf{K}[\mathbf{y}(k) - \hat{\mathbf{A}}\hat{\mathbf{x}}(k)]. \quad (4.105)$$

Таким образом, если известно, что модель имеет структуру (4.103), то для построения устойчивого фильтра Калмана не нужно решать уравнение Риккати.

Выбор конкретного (канонического) способа описания определяется рядом соображений, и в том числе собственными значениями, коэффициентами дифференциальных уравнений, некоторыми критическими параметрами, минимальным числом параметров и т. д. В обзорном докладе на III конгрессе ИФАК Ван-дер-Гринтен привел примеры моделей промышленных объектов. Вот выдержка из этого доклада, принадлежащая Квакернааку: «Для того чтобы показать, что понятие состояния можно успешно использовать в моделях самой разной сложности, приведем несколько случайно отобранных примеров из различных областей применения:

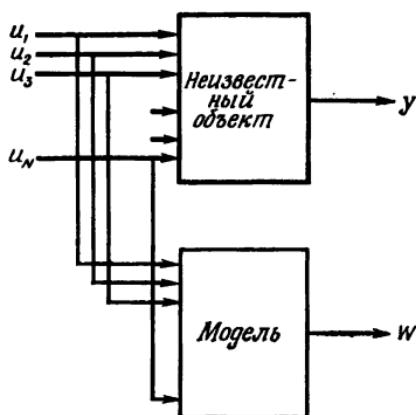
- 1) Состояние электрических цепей может быть описано емкостными напряжениями и индуктивными токами.
 - 2) Механические системы могут быть охарактеризованы моделями, в которых состояние определяется скоростью и пространственными координатами.
 - 3) Переменными состояний термодинамических систем являются термодинамические функции.
 - 4) В простейшей модели управления посадкой самолета используются угол тангажа, скорость изменения угла тангажа, высота и скорость подъема.
 - 5) Переменными состояний многоагрегатной энергетической системы являются магнитные потоки, фазы и скорость вращения.
 - 6) Известны примеры использования понятия состояния в технике. В классе моделей с сосредоточенными параметрами пароперегреватель описывается заданием теплосодержания выходной струи пара и температурой стенок трубы.
 - 7) Переменными состояний ректификационной колонны являются состав жидкости и теплосодержание насыщенной жидкости на каждой тарелке. Для проверки можно использовать численные методы.
- Понятие состояния находит применение и в других областях при построении динамических моделей.
- 1) Моделирование транспортных потоков на «микроуровне» заданием положений и скоростей отдельных автомашин.
 - 2) В динамических моделях производственных процессов также используется понятие состояния. В модели

распределения по Форрестеру в качестве переменных состояния (называемых уровнями) выбраны запасы, число служащих, невыполненные заказы, размеры капиталоизложений, банковские балансы, заказы в пути, товары в процессе доставки и потребность в рабочей силе.

Из этого следует, что понятие состояния с успехом может быть использовано для описания систем самой разной природы. Оно оказывает большую помощь при анализе поведения системы и построении ее модели. Для осуществления технического аспекта моделирования могут быть использованы цифровые вычислительные машины.»

Нелинейные модели

Для широкого класса нелинейных объектов достаточно общее непараметрическое описание дается рядами Вольтерра, использующими весовые функции высших порядков. Аппроксимация этих функций в конечном числе точек приводит к линейной по параметрам модели [1, 17, 51]. В большинстве практических ситуаций необходимое число параметров слишко велико. Допустим, что весовая функция первого порядка $h_1(\tau)$ аппроксимируется своими значениями в 10 точках. Так как функция $h_2(\tau_1, \tau_2)$ симметрична, для ее описания потребуется 55 параметров, для описания $h_3(\tau_1, \tau_2, \tau_3)$ — 220 параметров и т. д., если весовые функции высших порядков должны иметь одну и ту же «разрешающую способность». Очевидно, что размерность задачи оценивания параметров становится чрезмерной. Более того, физическая интерпретация этих параметров, по-видимому, будет затруднительна. Таким образом, при исследовании нелинейных объектов еще более, чем в случае линейных объектов, для которых применимы непараметрические модели, ощущается необходимость в подборе параметрических моделей. Большое значение имеет использование априорной информации о физической природе объекта. Хотя в работах [26—28] предложен способ построения параметрической модели без использования априорной информации об объекте под названием «эвристическая самоорганизация» или «метод группового учета аргументов (МГУА)».



Ф и г. 4.21.

На фиг. 4.21 изображен неизвестный объект, который характеризуется функциональным соотношением между выходной переменной y и набором входных переменных $\{u_1, \dots, u_N\}$:

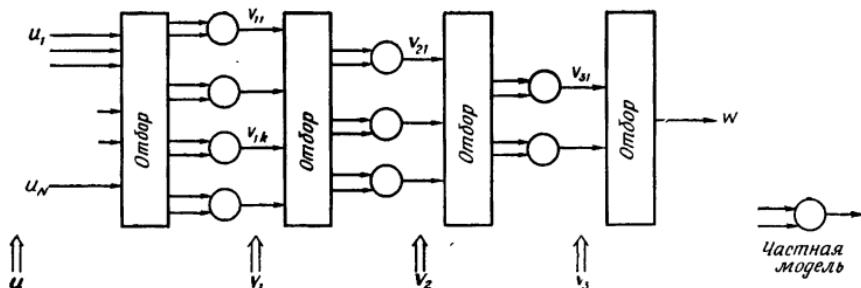
$$y = \Theta [u_1, \dots, u_N].$$

Не используя никакой априорной информации о $\Theta []$, необходимо построить такую модель, чтобы w было хорошей аппроксимацией y . Это можно было бы сделать с помощью полиномов Колмогорова — Габора (4.29). Здесь возникают следующие проблемы:

- 1) проблема размерности, вызываемая необходимостью оценивать большое число коэффициентов;
- 2) для оценивания необходимо располагать большим числом значений входного и выходного сигналов.

Метод Ивахненко состоит в выборе иерархии частных моделей вместо одной общей модели. На фиг. 4.22 изображено несколько «столбцов» частных моделей. Каждая частная модель из первого столбца имеет несколько входов u_i и только один выход v_{1k} . Параметры частных моделей подбираются так, чтобы v_{1k} как можно точнее аппроксимировало y .

У каждой частной модели из второго столбца имеется несколько входов v_{1i} и один выход v_{2k} . Параметры этих



Ф и г. 4.22

частных моделей подбираются так, чтобы каждое v_{2k} по возможности лучше приближало y и т. д. Таким образом, МГУА представляет собой алгоритм построения модели из набора идентичных частных моделей, объединенных в иерархическую структуру. Последовательно образуя новые переменные, стараются в конце концов получить w , являющееся хорошей аппроксимацией y . Новые переменные образуются как комбинации переменных предыдущего столбца (уровня). На каждом уровне для последующих преобразований отбираются только лучшие из переменных.

Осталось обсудить только два вопроса:

- 1) выбор типа частных моделей;
- 2) способ отбора переменных для последующих преобразований.

Частная модель может быть простой, скажем, с двумя входами u_1 и u_2 :

- 1) $v = \beta_0 + \beta_1 u_1 + \beta_2 u_2$, линейная комбинация с тремя параметрами;
- 2) $v = \beta_0 + \beta_1 u_1 + \beta_2 u_2 + \beta_{11} u_1^2 + \beta_{22} u_2^2 + \beta_{12} u_1 u_2$, квадратичный многочлен с шестью параметрами;
- 3) $v = \beta_0 + \beta_1 u_1 + \beta_2 u_2 + \beta_{12} u_1 u_2$, четыре параметра;
- 4) $v = \beta_1 u_1 u_2 + \beta_2 \bar{u}_1 u_2 + \beta_3 u_1 \bar{u}_2 + \beta_4 \bar{u}_1 \bar{u}_2$, для бинарных переменных, четыре параметра и т. д.

Частные модели первого типа не увеличивают порядок модели; в этом случае высшие степени и перекрестные члены входных сигналов объекта должны рассматриваться как входные сигналы частных моделей. Отбор переменных на каждом уровне необходим из-за многообразия комбинаций (u_i, u_j) на множестве (u_1, \dots, u_N), всего $N!$ Лучшие переменные для последующих преобразований отбираются на основе: 1) близости к y в смысле наименьших квадратов и 2) несходства группируемых переменных. Первый критерий очевиден и целенаправлен. Второй критерий связан с тем, что отсекаемые по методу наименьших квадратов переменные могут содержать существенную информацию. Это может выясниться при объединении с другими переменными, на более высоком уровне.

Пока еще не существует общих правил отбора промежуточных переменных на следующий уровень. Поскольку каждая частная модель содержит небольшое число пара-

метров, для их оценки требуется относительно небольшой объем данных. Для нелинейных объектов с памятью входные переменные модели u_i включают также входные сигналы объекта, сдвинутые по времени. В своих работах Ивахненко приводит несколько примеров приложений с хорошим соответствием между моделью и объектом.

Несмотря на довольно большой объем этой главы, здесь рассмотрены лишь некоторые вопросы математического описания систем. В литературе, указанной в конце гл. 14, можно найти примеры построения моделей в конкретных приложениях.

ЛИТЕРАТУРА

1. Alper P., A Consideration of the Discrete Volterra Series, *IEEE Trans. Autom. Control*, AC-10, 322—327 (1965).
2. Aoki M., Optimization of Stochastic Systems; Topics in Discrete-Time Systems, Academic Press, N.Y., 1967; есть русский перевод: Аоки М., Оптимизация стохастических систем, изд-во «Наука», 1971.
3. Åström K. J., Control Problems in Papermaking, Proc. IBM Scientific Computing Symp. on Control Theory and Applications, Yorktown Heights, N.Y., 135, 1964.
4. Åström K. J., Bohlin T., Wensmark S., Automatic Construction of Linear Stochastic Dynamic Models for Industrial Processes with Random Disturbances Using Operating Records. Rep. IBM Nordic Lab., Stockholm, 1965.
5. Åström K. J., Introduction to Stochastic Control Theory, Academic Press, N.Y., 1970.
6. Åström K. J., Eykhoff P., (см. литературу к гл. 2).
7. Balakrishnan A. V., Techniques of System Identification, Lecture Notes, Instituto Electrotecnico, Univ. of Rome, 1969.
8. Barker H. A., Multivariable Models for Control Systems, IFAC Symp. «Theory of Self-Adaptive Control Systems», Teddington, 1965, pp. 112—117; есть русский перевод: Баркер Г. А., Многосвязные модели для систем регулирования, в книге «Теория самонастраивающихся систем управления», Труды II межд. конгресса ИФАК по самонастраивающимся системам, изд-во «Наука», 1969.
9. Bohlin T., Real-Time Estimation of Time-Variable Processes Characteristics, Rep. IBM Nordic Lab., Stockholm, 1968.
10. Brockett R. W., Poles, Zeros and Feedback: State Space Interpretation, *IEEE Trans. Autom. Control*, AC-10, 129—135 (1965).
11. Bruni C., Isidori A., Ruberti R., Order and Factorization of the Impulse Response Matrix, Fourth IFAC Congress, Warsaw, Paper 19.4, 1969.
12. Bucy R. S., Canonical Forms for Multivariable Systems (Short Paper), *IEEE Trans. Autom. Control*, AC-13, 567—569 (1968).

13. Choksy N. H., Time Lag Systems, in: Progress in Control Engineering, Macmillan R. H., ed., Heywood, London, 1962, pp. 19—38.
14. Diskind T., Parameter Evaluation by Complex Domain Analysis, Joint Autom. Control Conf., 1969, pp. 770—781.
15. Eaton J., Identification for Control Purposes, IEEE Winter Meeting, N.Y., 1967.
16. Elgerd O. I., Control Systems Theory, McGraw-Hill, N.Y., 1967.
17. Eykhoff P., Some Fundamental Aspects of Process-Parameter Estimation, *IEEE Trans. Autom. Control*, AC-8, 347—357 (1963).
18. Fischer P., Vámos T., A Generalization of the Controllability and Observability Concept from the Point of View of Systems Identification, IFAC Symp. «Identification in Autom. Control Systems», Prague, Paper 1.10, 1967.
19. Flake R. H., Volterra Series Representation of Nonlinear systems, *AIEE Trans.*, 81, Part II (Applic. and Ind., № 64), 330—335 (1963).
20. Friedman B., Principles and Techniques of Applied Mathematics, Wiley, N.Y., 1956.
21. Gabor D., Wilby W. P. L., Woodcock R., A Universal Nonlinear Filter, Predictor and Simulator which Optimizes Itself by a Learning Process, *Proc. IEE*, 108B, 422—438 (1961).
22. George D. A., Continuous Nonlinear Systems, Techn. Rep. 355, Res. Lab. Electron. MIT, Cambridge (Mass.), 1959.
23. Goodman T. R., Hillsley R. H., Continuous Measurement of the Characteristics of Systems with Random Inputs: a Step Towards Self Optimizing Control, *Trans. ASME*, 80, 1839—1848 (1958).
24. Grenander U., Rosenblatt M., Statistical Analysis of Stationary Time Series, Wiley, N.Y., 1957.
25. Ho B. L., Kalman R. E., Effective Construction of Linear State-Variable Models from Input/Output Functions, *Regelungstechnik*, 14, 545—548 (1966).
26. Ивахненко А. Г., Метод группового учета аргументов — конкурент метода стохастической аппроксимации, *Автоматика*, № 3, 57—74 (1968).
27. Ivakhnenko A. G., Koppa Ju. V., The Group Method of Data Handling for the Solution of the Various Interpolation Problems of Cybernetics, Second IFAC Symp. «Identification and Process Parameter Estimation», Prague, Paper 2.1, 1970.
28. Ivakhnenko A. G., Heuristic Self-Organization in Problems of Engineering Cybernetics, *Automatica*, 6, 207—219 (1970).
29. Izawa K., Furuta K., Process Identification Using Correlation Coefficient, IFAC Symp. «Identification in Autom. Control Systems», Prague, Paper 3.9, 1967.
30. Jenkins G., Watts D., Spectral Analysis and its Applications, Holden Day, San Francisco, 1968; есть русский перевод: Дженкинс Г., Ваттс Д., Спектральный анализ и его применения, вып. 1, 2, изд-во «Мир», 1971, 1972.
31. Kailath T., Frost P., An Innovations Approach to Least-Squares Estimation, *IEEE Trans. Autom. Control*, Part I, AC-13, 646—660 (1968).

32. Kalman R. E., On the General Theory of Control Systems, First IFAC Congress, Moscow, 1960, pp. 481—492; есть русский перевод: Калман Р. Е., Об общей теории систем управления, в сб. «Теория дискретных оптимальных и самонастраивающихся систем», Труды I Межд. конгресса ИФАК, Изд-во АН СССР, т. 2, 1961, стр. 521—547.
33. Kalman R. E., Mathematical Description of Linear Dynamical Systems, *SIAM J. Control*, 1, 152—192 (1963).
34. Kalman R. E., Falb P. L., Arbib M. A., Topics in Mathematical System Theory, McGraw-Hill, N.Y., 1969; есть русский перевод: Калман Р., Фалб П., Арбид М., Очерки по математической теории систем, изд-во «Мир», 1971.
35. Kashyap R. L., Maximum Likelihood Identification of Stochastic Linear Systems, *IEEE Trans. Autom. Control*, AC-15, 25—34 (1970).
36. Kreindler E., Sarachik P. E., On the Concepts of Controllability and Observability of Linear Systems, *IEEE Trans. Autom. Control*, AC-9, 129—136 (1964).
37. Lee R. C. K., Optimal Estimation, Identification, and Control, MIT Press, Cambridge (Mass.), 1964 (Research Monograph 28); есть русский перевод: Ли Р., Оптимальные оценки, определение характеристик и управление, изд-во «Наука», 1966.
38. Lee Y. W., Schetzen M., Measurement of the Wiener Kernels of a Nonlinear System by Cross Correlation, *Int. J. Control*, 3, 237—254 (1965).
39. Luenberger D. G., Canonical Forms for Linear Multivariable Systems (Short Paper), *IEEE Trans. Autom. Control*, AC-12, 290—293 (1967).
40. MacFarlane A. G. J., Dynamical System Models, Harrap, London, 1970.
41. Mann H. B., Wald A., On the Statistical Treatment of Linear Stochastic Difference Equations, *Econometrica*, 11, № 3, 4 (1943).
42. McFee R., Determining the Response of Nonlinear Systems to Arbitrary Inputs, *AIEE Transact.*, 80, Part II (Applic. and Ind., № 56), 189—193 (1961).
43. McGhee R. B., Identification of Nonlinear Dynamic Systems by Regression Analysis Methods, Univ. of Southern California, Los Angeles (Cal.), EE Dept., Ph. D. Dissert., 1963.
44. Mehra R. K., Identification of Stochastic Linear Dynamic Systems, Proc. 1969 IEEE Symp. on Adaptive Processes, Pennsylvania State Univ.
45. MIT, Several References and Notes, Q. Prog. Rep. Res. Lab. Electron. MIT, Cambridge (Mass.), № 71, 75, 76, 77, 1963—1965.
46. Phillipson G. A., Identification of Distributed Systems, Elsevier, Amsterdam, 1971.
47. Prosser R. T., Root W. L., Determinable Classes of Channels, *J. Math. Mech.*, 16, 365—398 (1966).
48. Root W. L., On System Measurement and Identification, Symp. on System Theory, Polytechnic Institute of Brooklyn, 1965.
49. Root W. L., On the Structure of a Class of System Identification Problems, Second IFAC Symp. «Identification and Process Parameter Estimation», Prague, Paper 8.1, 1970.

50. Rowe I. H., A Statistical Model for the Identification of Multivariable Stochastic Systems, Proc. IFAC Symp. on Multivariable Systems, Düsseldorf, 1968.
51. Roy R., Sherman J., System Identification and Pattern Recognition, IFAC Symp. «Identification in Autom. Control Systems», Prague, Paper 1.6, 1967.
52. Staley R. M., Yue P. C., On System Parameter Identifiability, *Inform. Sci.*, 2, 127—138 (1970).
53. Steiglitz K., McBride L. E., A Technique for the Identification of Linear Systems (Short Paper), *IEEE Trans. Aut. Control*, AC-10, 461—464 (1965).
54. Tou J. T., Modern Control Theory, McGraw-Hill, N.Y., 1964; есть русский перевод: Ту Ю., Современная теория управления, изд-во «Машиностроение», 1971.
55. Tuel W. G., Jr., Canonical Forms for Linear Systems, I and II, IBM Res. Rep. RJ 375 and 390, 1966.
56. Vališ J., On-Line Identifications of Multivariable-Linear Systems of Unknown Structure from Input Output Data, Second IFAC Symp. «Identification and Process Parameter Estimation», Prague, paper 1.5, 1970.
57. Van Harten J., The Description of Nonlinear Systems (на голландском), Thesis, E. E. Dept., Univ. of Technology, Eindhoven, Netherlands, 1966.
58. Volterra V., Theory of Functionals and of Integral and Integro-Differential Equations, Dover Publ., N.Y., 1959.
59. Whittle P., Prediction and Regulation, Van Nostrand, Princeton, N.J., 1963.
60. Wiener N., Nonlinear Problems in Random Theory, MIT Press, Cambridge (Mass.), 1958, 1962; есть русский перевод: Винер Н., Нелинейные задачи в теории случайных процессов, ИЛ, 1961.
61. Wolff A. A., Dietz J. H., An Adaptive Correlator for Underwater Measurements, *Marine Sciences Instrumentation*, 1, 347—354 (1962).
62. Wong K. Y., Wiig K. M., Allbritton E. J., Computer Control of the Clarksville Cement Plant by State Design Method, IEEE Cement Ind., Conf., St. Louis, Mo., 1968.
63. Zadeh L. A., A Contribution to the Theory of Nonlinear Systems, *J. Franklin Inst.*, 255, 387—408 (1953).
64. Zadeh L. A., Time-Varying Networks, *Proc. IRE*, 49, 1488—1503 (1961).
65. Zadeh L. A., Desoer C. A., Linear System Theory; The State Space Approach, McGraw-Hill, N.Y., 1963; есть русский перевод: Заде Л., Дезоэр Ч., Теория линейных систем. Метод пространства состояний, изд-во «Наука», 1970.

ДОПОЛНИТЕЛЬНАЯ ЛИТЕРАТУРА

4.2. Линейные модели

Anderson B. D. O. et al., Equivalence of Linear Time-Invariant Dynamical Systems, *J. Franklin Inst.*, 281, 371—378 (1966).

Balakrishnan A. V., Foundations of State-Space Theory of Continuous Systems, *Journal of Computer and System Sciences*, 1, 91—116 (1967).

Barker H. A., Hawley D. A. (см. литературу к гл. 3).

Bolte W., An Approximation Procedure for Determining the Transient Response from the Frequency Response (на немецком), *Zeitschr. Messen, Steuern, Regeln*, 13, 248—249 (1965).

Brown B. M., The Mathematical Theory of Linear Systems, Chapman and Hall, London, 1961.

Budin M. A., Minimal Realization of Discrete Linear Systems from Input-Output Observations, *IEEE Trans. Autom. Control*, AC-16, 395—401 (1971).

Chen C. F., Philip B. L., Graphical Determination of Transfer Function Coefficients of a System From its Frequency Response, *IEEE Trans. Applic. and Ind.*, 82, № 65, 42—45 (1963).

Davies W. D. T., Fishwick W., The Synthesis of Dynamic Models of Plants and Processes, *J. Mech. E. Proceedings*, 179 (1964).

DeRusso P. M., Roy R. J., Close C. M., State Variables for Engineers, Wiley, N.Y., 1965; есть русский перевод: Деруссо П. и др., Пространство состояний в теории управления. Для инженеров, изд-во «Наука», 1970.

Douce J. L., Roberts P. D. (см. литературу к гл. 6).

Kautz W. H., Transient Synthesis in the Time Domain, *IRE Trans. Circuit Theory*, CT-1, 29—39 (1954).

Kitamori T. (см. литературу к гл. 3).

Koenig J. F., A Method for Obtaining the Time Response of Any Linear System (Short Paper) *IEEE Trans. Autom. Control*, AC-9, 556—565 (1964).

Krishnamurthy E. V., Solution of Linear Systems as Eigen Value Problems (Correspondence), *Proc. IEEE*, 52, 1065—1066 (1964).

Kuh E. S., Rohrer R. A., The State-Variable Approach to Network Analysis, *Proc. IEEE*, 53, 672—686 (1965).

Lovering W. F., Analog Computer Simulation of Transfer Functions, *Proc. IEEE*, 53, 306—307 (1965).

Lubbock J. K., Barker H. A., A Solution of the Identification Problem, *IEEE Trans. Applic. and Ind.*, 83, № 72, 166—173 (1964).

Milsum J. H., Determining Transfer Functions with Analog Computers, *ISA Journal*, 7, 54—57 (1960).

Нелепин Р. А., Об исследовании методом сечений пространства параметров одного класса систем управления, *Изв. АН СССР, мехн. кибернетика*, № 4, 126—133 (1965).

Ogata K., State Space Analysis of Control Systems, Prentice-Hall, Englewood Cliffs, N.Y., 1967.

Peterka V., Vidinčev P., Rational-Fraction Approximation of Transfer Functions, IFAC Symp. «Identification in Autom. Control Systems», Prague, Paper 2.1, 1967.

Полонников Д. Е., Метод определения фазочастотных характеристик систем по логарифмической характеристике при произвольных наклонах ее отдельных участков, *Автоматика и телемеханика*, № 4, 721—723 (1965).

Radtke M., On the Approximation of Linear Aperiodic Transient Responses (на немецком), *Zeitschr. Messen, Steuern, Regeln*, 9, 192—196 (1966).

Roberts P. D. (см. литературу к гл. 3 и 6).

Roberts P. D. (см. литературу к гл. 3).

Sauer W., A Survey of System Theoretical Parameters (на немецком), *Nachr. Technik*, 16, 194—197 (1966).

Seide W., Determination of Characteristic Values from the Weight Function (на немецком), *Zeitschr. Messen, Steuern, Regeln*, 6, 323—325 (1963).

Storer J. E., Passive Network Synthesis, McGraw-Hill, N.Y., 1957.

Thal-Larsen H., Frequency Response from Experimental Non-oscillatory Transient Response Data, *Trans. AIEE*, 74, Part II (Appl. and Ind.), 109—114 (1955).

Timothy L. K., Bona B. E., State Space Analysis, an Introduction, McGraw-Hill, N.Y., 1968.

Wang P. K. C., Mathematical Models for Time-Domain Design of Electrohydraulic Servomechanisms, *Trans. AIEE*, 80, Part II (Appl. and Ind., № 57), 252—260 (1961).

Woodside C. M., The Interpretation of Time Domain Models in the Frequency Domain (Correspondence), *IEEE Trans. Autom. Control*, AC-12, 110—111 (1967).

Zakian V., Relation Between System Parameters and Response Duration, *Electron. Letters*, 1, 293—294 (1965).

Zorn J., Methods of Evaluating Fourier Transforms with Application to Control Engineering, Univ. of Technology, Delft, Netherlands, Doct. Dissertation, 1963.

4.3. Нестационарные линейные модели

Bishop K. A., Sliepcevich C. M., Puckett T. H., Techniques for Identification of Linear and Linear Timevarying Processes, Joint Autom. Control Conf., 1962.

Calahan D. A., Equivalence of Timevarying Systems (Correspondence), *Proc. IEEE*, 52, 1372 (1964).

Литовченко Ц. Г., Анализ решения линейных уравнений, описывающих динамические системы с переменными параметрами одного класса, *Автоматика и телемеханика*, № 4, 457—465 (1961).

Мееров М. В., Методы синтеза структур, эквивалентных самонастраивающимся системам для объектов с переменными параметрами, в книге «Теория самонастраивающихся систем управления», Труды II Межд. конгресса по самонастраивающимся системам, изд-во «Наука», 1969, стр. 181—189.

Peschel M., On the Transfer Characteristics of Timevarying Linear Blocks (на немецком), *Zeitschr. Messen, Steuern, Regeln*, 6, 310—315 (1963).

Schweizer G., A General Method for Studying Systems with Periodically Changing Parameters (на немецком), *Arch. Elekt. Ubertr.*, 19, 469—482 (1965).

Солодов А. В., О преобразовании начальных условий на выходе линейной системы с переменными параметрами в эквивалентный входной сигнал, *Автоматика и телемеханика*, № 7, 654—660 (1958).

Солодов А. В., Линейные системы автоматического управления с переменными параметрами. Физматгиз, 1962.

Tucker D. G., Circuits with Periodically-Varying Parameters, MacDonald, London, 1964.

Wunsch G., The Mathematical Treatment of Linear Systems with Variable Parameters, *Nachr. Technik*, 11, 221—222 (1962).

Wunsch G., Analysis of Networks with Changing Parameters by Means of Z-Transformation (на немецком), *Arch. Elekt. Übertr.*, 20, 289—293 (1966).

Zadeh L. A., Correlation Functions and Power Spectra in Variable Networks, *Proc. IRE*, 38, 1342—1345 (1950).

Zadeh L. A., The Determination of the Impulsive Response of Variable Networks, *J. Appl. Phys.*, 21, 642—645 (1950).

4.4. Нелинейные модели

Alper P., Higher-Dimensional Z-Transforms and Nonlinear Discrete Systems, *Revue A*, 6, 199—212 (1964).

Balakrishnan A. V., State-Space Theory of Nonlinear Systems, in: Caianiello E., Functional Analysis and Optimization, Academic Press, N.Y., 1966.

Barrett J. F., The Use of Functionals in the Analysis of Non-Linear Physical Systems, Rep. Statistical Advisory Unit. № 1/57, Ministry of Supply, England; см. также *J. Electron. Control*, 15, 567—617 (1963).

Bose A. G., A Theory of Nonlinear Systems, Techn. Rep. 309, Res. Lab. Electron. MIT, Cambridge (Mass.), 1956.

Brilliant M. B., Theory of the Analysis of Nonlinear Systems, Techn. Rep. 345, Res. Lab. Electron. MIT, Cambridge (Mass.), 1958.

Bruni C., Pillo G. Di., Koch G., On the Mathematical Models of Bilinear Systems, *Ricerche di Automatica*, 2, 11—26 (1971).

Cameron R. H., Martin W. T., The Orthogonal Development of Nonlinear Functionals in Series of Fourier-Hermite Functionals, *Ann. Math. Statist.*, 18, 385—389 (1947).

Flake R. H., Volterra Series Representation of Time Varying Nonlinear Systems, Second IFAC congress, Basle, 1963, pp. 91—99; есть русский перевод: Флейк Р. Г., Теория рядов Вольтерра и ее приложения к нелинейным системам с переменными параметрами, в книге «Оптимальные системы. Статистические методы», Труды I Межд. конгресса ИФАК, изд-во «Наука», 1965, стр. 453—468.

Frank P., McFee R., Determining Input-Output Relationships of Nonlinear Systems by Inversions, *IEEE Trans. Circuit Theory*, CT-10, 168—180 (1963).

Hemani H., Cosgriff R. L., Graph Theory in the Analysis of Non-Linear Systems, *IEEE Trans. Autom. Control*, AC-11, 550—552 (1966).

Kohr R. H., (см. литературу к гл. 9).

Ku Y. H., Wolf A. A., Volterra-Wiener Functionals for the Analysis of Nonlinear Systems, *J. Franklin Inst.*, 281, 9 (1966).

Lubbock J. K., Optimization of a Class of Nonlinear Systems, First IFAC Congress, Moscow, 1960, pp. 761—769; есть русский перевод: Лаббок Дж. К., Оптимизация одного класса нелинейных фильтров, в сб. «Статистические методы исследования. Теория структур. Моделирование, терминология, образование», Труды I Международного конгресса ИФАК, т. 3, 1961, стр. 159—180.

Oppenheim A. V., Superposition in a Class of Nonlinear Systems, Techn. Rep. 432, Res. Lab. Electron. MIT, Cambridge (Mass.), 1965.

Pugachev V. S., Statistical Theory of Systems Reducible to Linear, *IRE Trans. Circuit Theory*, CT-7, 506—513 (1960).

Radtke M., On the Determination of the Dynamic Characteristic Values of Nonlinear Transmission Elements from Graphically Represented Transition Functions, Third IFAC Congress, London, paper 12E, 1966; есть русский перевод: Радтке М., Определение динамических характеристик нелинейных объектов по графикам переходных функций, в книге «Теория непрерывных автоматических систем и вопросы идентификации», изд-во «Наука», 1971, стр. 217—229.

Roy R. J., DeRusso P. M., A Digital Orthogonal Model for Nonlinear Processes with Two-Level Inputs, *IRE Trans. Autom. Control*, AC-7, 93—101 (1962).

Simpson H. R., A Sampled Data Non-Linear Filter, *Proc. IEE*, 112, 1187—1196 (1965).

Singleton H. B., The Theory of Nonlinear Transducers. Techn. Rep. 160, Res. Lab. Electron. MIT, Cambridge (Mass.), 1950.

Smets H. B., Analysis and Synthesis of Nonlinear Systems, *IRE Trans. Circuit Theory*, CT-7, 459—469 (1960).

Van Trees H. L., Synthesis of Optimum Non-Linear Control Systems, MIT Press, Cambridge (Mass.), 1962; есть русский перевод: Ван-Трис Г. Л., Синтез оптимальных нелинейных систем управления, изд-во «Мир», 1964.

Wolf A. A., Some Recent Advances in the Analysis and Synthesis of Nonlinear Systems, *AIEE Trans.*, 80, Part II (Applic. and Ind.), 289—300 (1961).

Zadeh L. A., Nonlinear Multipoles, *Proc. Nat. Acad. of Sciences*, Wash., DC., 39, 274—280 (1953).

Zadeh L. A., Optimum Nonlinear filters, *J. Appl. Phys.*, 24, 396—404 (1953).

Zadeh L. A., On the Identification Problem, *IRE Trans. Circuit Theory*, CT-3, 277—281 (1956).

Zadeh L. A., On the Representation of Nonlinear Operators, *IRE Conv. Rec.*, Part II, 105—113 (1957).

4.5. Управляемость, наблюдаемость, идентифицируемость

Chen C. T. et al., Simplified Conditions for Controllability and Observability of Linear Time-Invariant Systems (Correspondence), *IEEE Trans. Autom. Control*, AC-11, 613—614 (1966).

Chen C. T., Desoer C. A., Controllability and Observability of Composite Systems, *IEEE Trans. Autom. Control*, AC-12, 402—409 (1967).

Gilbert E. G., Controllability and Observability in Multivariable Control Systems, *SIAM J. Control*, 1, 128—151 (1963).

Sekiguchi T., Observability of Linear Dynamic Measuring Systems and Some Applications, Fourth IFAC Congress, Warsaw, Paper 12.5, 1969.

Tse E., Anton J. J., On the Identifiability of Parameters, *IEEE Trans. Autom. Control*, AC-17, 637—646 (1972).

Wang P. K. C., Control of Distributed Parameter Systems, in: Advances in Control Systems, Leondes C. T., ed., 1, Academic Press, N.Y., 1964, pp. 75—172.

Weiss L., Kalman R. E., Contributions to Linear System Theory, *Internat. J. Engng. Sci.*, 3, 141—171 (1965).

4.6. Модели

Davison E. J., A New Method for Simplifying Large Linear Dynamic Systems (Correspondence), *IEEE Trans. Autom. Control*, AC-13, 214—215 (1968).

Kashyap R. L., A New Method if Recursive Estimation in Discrete Linear Systems, *IEEE Trans. Autom. Control*, AC-15, 34—43 (1970).

Koepcke R. W., Solution to Sampled, Minimum-Time Problem, Joint Autom. Control Conf., 1963, pp. 94—100; см. также *Trans. ASME (J. Basic Engng)*, 86D, 145—150 (1964).

ГЛАВА 5

НЕКОТОРЫЕ ВЕРОЯТНОСТНЫЕ ПОНЯТИЯ. ТЕОРИЯ ОЦЕНИВАНИЯ, СХОДЯЩИЕСЯ АЛГОРИТМЫ, СТОХАСТИЧЕСКАЯ АППРОКСИМАЦИЯ

В этой главе обсуждаются идеи, которые имеют большое значение для оценивания параметров. Сначала оценки рассматриваются со статистической точки зрения. Очевидно, что существует тесная связь между количеством априорной информации и типом оптимальной процедуры оценивания. Рассматриваются различные виды оценок:

- 1) оценки по методу наименьших квадратов;
- 2) марковские оценки;
- 3) оценки максимального правдоподобия;
- 4) байесовские оценки.

Эти оценки упорядочены по возрастанию объема исходной информации об объекте. При оценивании методом наименьших квадратов предполагается, что динамика объекта может быть аппроксимирована выбранной моделью. При получении марковских оценок считается также известной ковариационная матрица шума. Для вычисления оценок максимального правдоподобия необходимо знание плотности вероятности измеряемого случайного процесса. Байесовские оценки, или оценки с минимальным риском, требуют знания априорных плотностей вероятности неизвестных параметров и величины штрафа за ошибки. Описывается такие свойства оценок, как эффективность, асимптотическая эффективность, состоятельность, смещенность и т. п. Оценки 1—3 можно рассматривать как частные случаи байесовских оценок с меньшим объемом априорной информации. Так выясняется связь между различными типами оценок.

Во многих случаях, например в явных методах или методах с настраиваемой моделью, алгоритмы оценивания представляют собой способы оптимизации. Эта оптимиза-

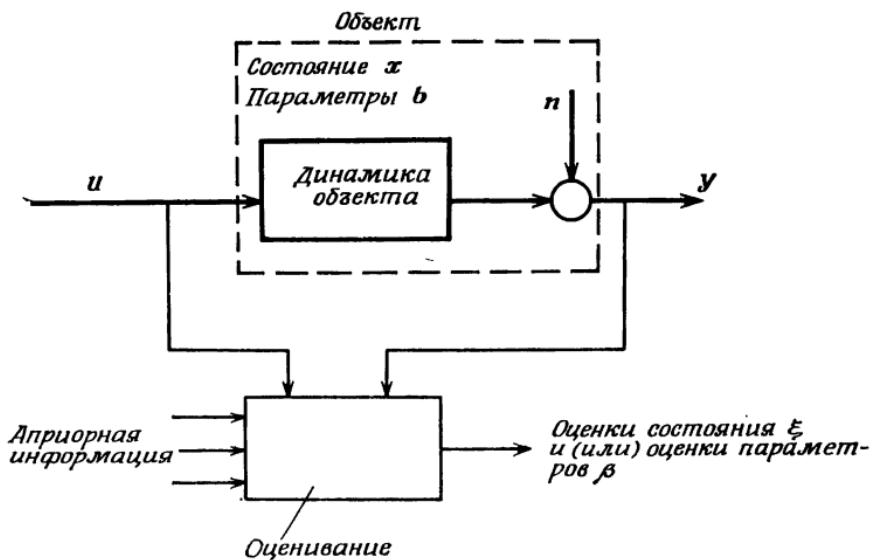
ция может быть поиском минимума функций затрат, максимума вероятности, минимума среднеквадратической ошибки и т. д. Поэтому в этой главе рассматриваются также оптимальные поисковые, или сходящиеся, алгоритмы. В зависимости от отсутствия или наличия помех эти алгоритмы могут быть названы детерминированными (разд. 5.2) или стохастическими (разд. 5.3).

5.1. ТЕОРИЯ ОЦЕНИВАНИЯ

Рассматриваемая ситуация изображена на фиг. 5.1. В этом разделе излагаются только основные идеи, более подробное обсуждение содержится в гл. 11. Сосредоточим внимание на оценивании параметров. Необходимо получить правило (*оператор оценивания*), т. е. такую связь

$$\beta = \beta \{u(1), \dots, u(k); y(1), \dots, y(k)\} = \\ = \beta(u', y'), \quad (5.1)$$

которая позволяла бы приписать неизвестному параметру b рассматриваемого объекта некоторое числовое значение (оценку) β . Эта оценка зависит от последовательности



Ф и г. 5.1.

наблюдений. Если необходимо найти набор оценок $[\beta_0, \dots, \beta_m] = \beta'$ для вектора параметров объекта \mathbf{b} , то эта связь описывается векторным соотношением

$$\beta = \beta(u', y').$$

Свойства оценок

При оценивании существенный интерес представляет информация о рассматриваемых параметрах, в частности плотность вероятности $p(\beta; T)$ или $p(\beta; k)$. Эта функция зависит от продолжительности интервала наблюдений T или, что то же самое, от размера выборки k . Это показано на фиг. 5.2 для одномерного случая; $T_1 < T_2 < T_3$. Плотность вероятности представляет собой наиболее полную информацию, которую можно получить, применяя статистические методы. Однако работать с такой информацией довольно сложно и непрактично, особенно если размерность β больше двух, когда для описания результатов эксперимента требуется многомерное представление. Поэтому во многих случаях плотность вероятности заменяется ее важнейшими характеристиками¹⁾:

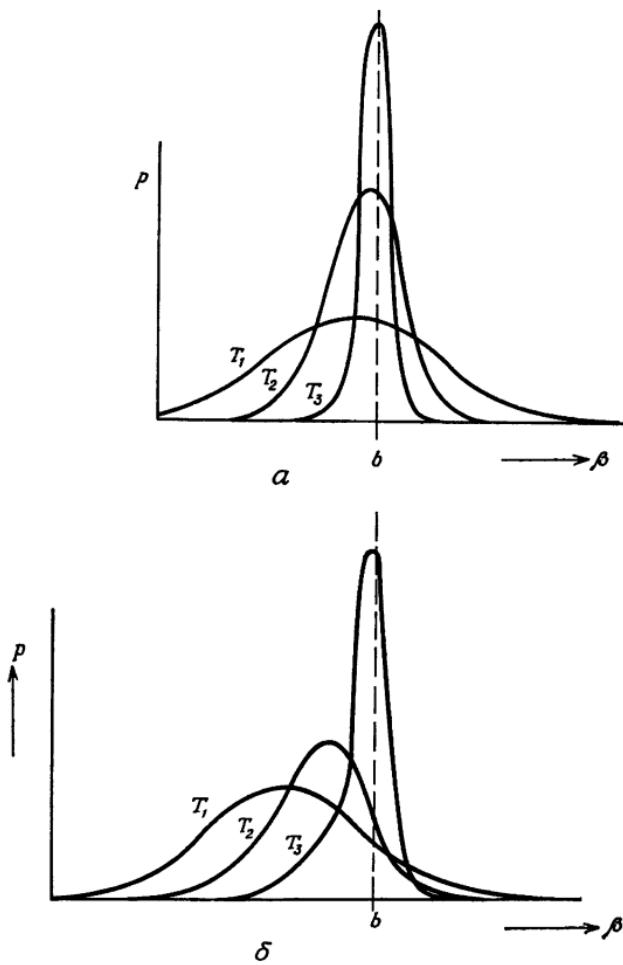
математическое ожидание $E[\beta]$,

смещение $E[\beta] - b$,

ковариация $\text{cov}[\beta] = E[(\beta - E[\beta])(\beta - E[\beta])']$.

Упрощение, связанное с заменой функций их числовыми характеристиками, можно оценить на примерах, приведенных на фиг. 5.2, а и б. Одновременно можно получить представление об информации, сохраняющейся в результате такого упрощения. Отметим, что если бы плотность вероятности была (многомерной) плотностью гауссовского распределения вероятностей, то замена плотности вероятности на $E[\beta]$ и $\text{cov}[\beta]$ не привела бы к потери информации, так как гауссовское распределение полностью характеризуется своими первыми и вторыми моментами.

¹⁾ Основные понятия из теории вероятностей приведены в приложении Б.



Ф и г. 5.2.

В литературе по математической статистике (см., например, [5, 16]) перечисляются некоторые желательные свойства оценок β параметров b , а именно:

1) несмещенность, когда для каждого k

$$\mathcal{E}[\beta] = b;$$

2) состоятельность, если с ростом k

$$\lim_{k \rightarrow \infty} P\{|b - \beta| > \varepsilon\} = 0$$

для сколь угодно малого $\varepsilon > 0$, т. е. β сходится по вероятности (или стохастически) к истинному значению b ;

3) эффективность, если для всех несмешанных оценок γ

$$\text{cov}[\beta] = \mathbb{E}[(\beta - b)(\beta - b)'] \leq \mathbb{E}[(\gamma - b)(\gamma - b)'] = \text{cov}[\gamma]$$

или

$$\det[\text{cov}[\gamma] - \text{cov}[\beta]] \geq 0;$$

4) достаточность, если для всех остальных оценок γ $p(\gamma | \beta)$ не зависит от b .

Если первое и третье свойства выполняются только в пределе при $k \rightarrow \infty$, то они называются асимптотической несмешанностью и асимптотической эффективностью. Снова обращаясь к фиг. 5.2, заметим, что там представлены случай несмешанной оценки (*a*) и случай смешанной, но, быть может, асимптотически несмешанной оценки (*b*). Следует отметить, что оценки отличаются еще и тем, что в них по-разному учитывается априорная информация, основанная на физической природе объекта или на априорных измерениях, включая наше «доверие» к этой информации.

Некоторые типы оценок

Для рассмотрения оценок в качестве исходной выберем ситуацию, когда имеется много априорной информации [15], а именно:

а) Известна плотность вероятности шума n . По ней можно рассчитать плотность вероятности измерений y , которая зависит от параметров объекта b и обозначается через $p(y | b)$. Понятно, что $p(y)$ также зависит и от n .

б) Известна плотность распределения вероятностей параметров b , обозначаемая через $q(b)$.

в) Известны потери, связанные с численной величиной оценки β при истинном значении вектора параметров b . Эта функция потерь $C(\beta, b)$ имеет минимум при $\beta = b$. После обсуждения этого наиболее сложного случая будут последовательно отброшены предположения в), б) и а).

Байесовские оценки. Априорная информация: $p(y | b)$, $q(b)$, $C(\beta, b)$. Существо метода сконцентрировано

в формуле Байеса

$$p(\mathbf{y} \mid \mathbf{b}) q(\mathbf{b}) = p(\mathbf{y}, \mathbf{b}) = p(\mathbf{b} \mid \mathbf{y}) p(\mathbf{y}).$$

Рассмотрим также

$$p(\mathbf{b} \mid \mathbf{y}) = \frac{p(\mathbf{y}, \mathbf{b})}{p(\mathbf{y})} = \frac{p(\mathbf{y} \mid \mathbf{b}) q(\mathbf{b})}{p(\mathbf{y})}, \quad (5.2)$$

где

$$p(\mathbf{y}) = \int_{m+1} p(\mathbf{y}, \mathbf{b}) d^{m+1}\mathbf{b},$$

\int_{m+1} обозначает $(m + 1)$ -кратный интеграл; $d^{m+1}\mathbf{b}$ обозначает $db_0 db_1 \dots db_m$. Таким образом, при имеющейся априорной информации все члены в правой части известны.

Эту условную плотность вероятности $p(\mathbf{b} \mid \mathbf{y})$ можно интерпретировать как *апостериорную плотность вероятности вектора параметров \mathbf{b} при заданных результатах измерений $\mathbf{y} : \mathbf{y} = \mathbf{c}$* .

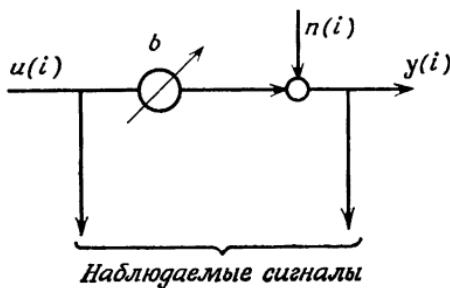
Пример. Полезно иллюстрировать эти идеи на крайне простом примере. Обратимся к фиг. 5.3; нужно найти оценку β параметра b . Эта оценка должна основываться на выборочных значениях сигналов $u(i)$ и $y(i)$, где

$$y(i) = bu(i) + n(i), \quad i = 1, 2, \dots, k,$$

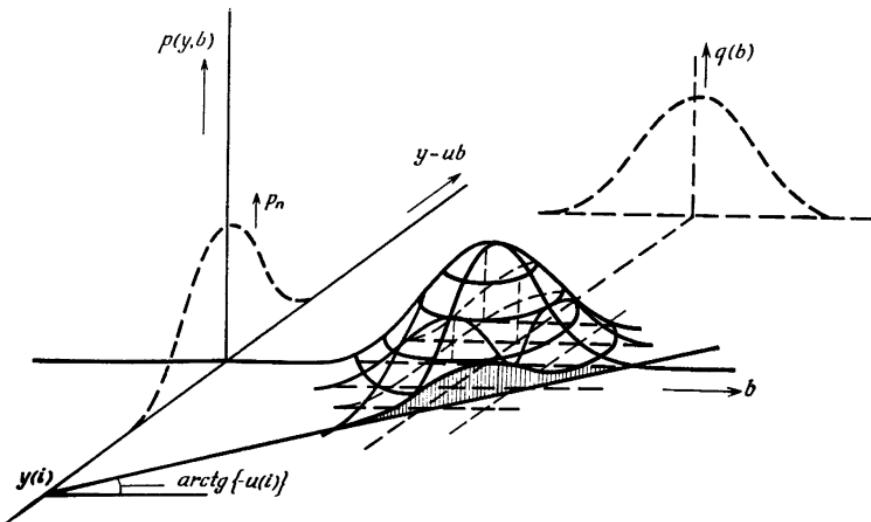
а среднее значение $n(i)$ равно нулю.

Пусть плотности распределений вероятностей n и b есть p_n и q . Если n и b статистически независимы, то

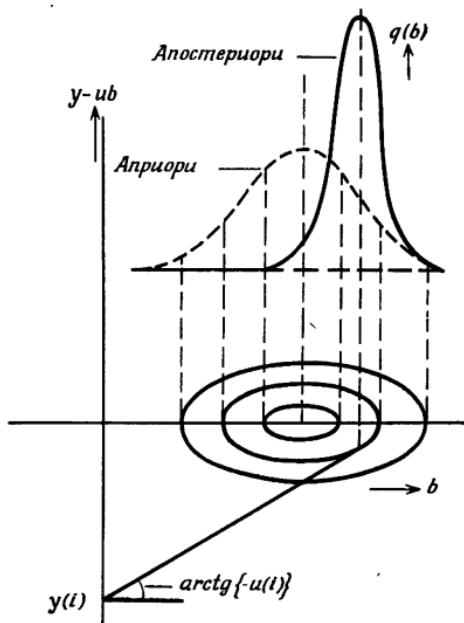
$$p(y, b) = p(y - ub, b) \underbrace{\left| \frac{\partial(y - ub)}{\partial y} \right|}_1 = p_n(y - ub) q(b)$$



Фиг. 5.3.



Ф и г. 5.4.



Ф и г. 5.5.

(фиг. 5.4). Отметим, что это *априорная* информация, имевшаяся до того, как были сделаны измерения u (i) и y (i). После выполнения этих измерений выявляется новый «разрез» плотности распределения вероятностей, который определяет *апостериорную* плотность вероятности параметра b . Эта новая функция может использоваться как априорная информация для следующих измерений. Таким способом можно построить $q(b)$ как функцию числа наблюдений (фиг. 5.5). Отметим, что при стационарном аддитивном шуме функция p_n не меняется.

Читателю предлагается рассмотреть частные случаи $u(k) = 0$ и $u(k) \rightarrow \infty$. Опять метод может быть обобщен на большее число параметров, однако имеются серьезные ограничения.

Опираясь на апостериорную плотность вероятности $p(b|y)$, нужно решить, какое значение приписать оценке. На фиг. 5.6 иллюстрируется одномерный случай функции потерь вида $C(\beta, b) = (\beta - b)^2$. Если β — величина оценки, то ожидаемые потери составят

$$\int_{-\infty}^{+\infty} (\beta - b)^2 p(b|y) db, \quad y = c. \quad (5.3)$$

Это скалярная величина, которую можно изобразить на графике как функцию β (фиг. 5.6). Таким образом, наилучшая оценка представляет собой такое $\hat{\beta}$, при котором эта функция принимает минимальное значение.

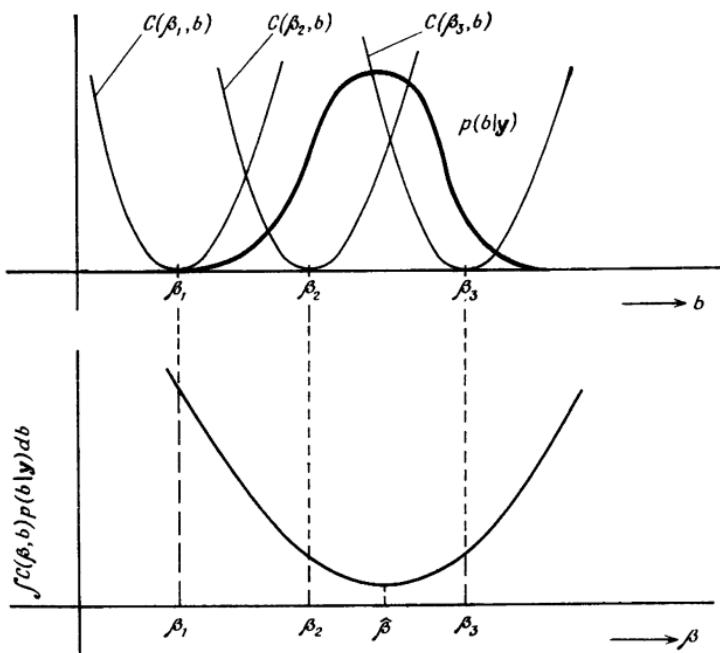
Приведенные рассуждения можно представить также следующим образом. *Условный риск* выбора β (y) при истинном значении вектора параметров b может быть записан как математическое ожидание¹⁾ функции потерь по

¹⁾ Отметим, что

$$\begin{aligned} \mathcal{E}_y [\dots] &= \int_k \dots p(y) d^k y, \\ \mathcal{E}_{y|b} [\dots] &= \int_k \dots p(y|b) d^k y \end{aligned}$$

и

$$\int_k p(y|b) d^k y = 1.$$



Ф и г. 5.6.

наблюдениям \mathbf{y} :

$$\mathcal{E}_{\mathbf{y} \mid \mathbf{b}} [C(\beta, \mathbf{b})] = \int_k C(\beta, \mathbf{b}) p(\mathbf{y} \mid \mathbf{b}) d^k \mathbf{y}. \quad (5.4)$$

Средний риск представляет собой математическое ожидание условного риска по распределению значений параметра объекта \mathbf{b} :

$$\begin{aligned} R(\beta) &= \mathcal{E}_{\mathbf{b}} [\mathcal{E}_{\mathbf{y} \mid \mathbf{b}} [C(\beta, \mathbf{b})]] = \\ &= \int_{m+1} \int_k C(\beta, \mathbf{b}) p(\mathbf{y} \mid \mathbf{b}) q(\mathbf{b}) d^k \mathbf{y} d^{m+1} \mathbf{b}. \end{aligned} \quad (5.5)$$

Оценка, которая минимизирует это выражение, называется оценкой *минимального риска*. По формуле Байеса средний риск можно переписать в виде

$$R(\beta) = \int_k p(\mathbf{y}) d^k \mathbf{y} \int_{m+1} C(\beta, \mathbf{b}) p(\mathbf{b} \mid \mathbf{y}) d^{m+1} \mathbf{b}. \quad (5.6)$$

Так как $p(y) \geq 0$, средний риск можно минимизировать, сделав внутренний интеграл как можно меньше при $y = c$:

$$\min_{\beta} \int_{m+1} C(\beta, b) p(b | c) d^{m+1}b. \quad (5.7)$$

Необходимое условие этого минимума имеет простой вид

$$\frac{\partial}{\partial \beta} \int_{m+1} C(\beta, b) p(b | c) d^{m+1}b \Big|_{\beta=\hat{\beta}} = 0.$$

Получаемая оценка называется *байесовской*, поскольку используется формула Байеса.

Пример. В качестве простого примера рассмотрим снова задачу оценивания одного параметра b с использованием квадратичной функции потерь $C(\beta, b) = (\beta - b)^2$. Тогда условие (5.7) принимает вид

$$\hat{\beta} = \frac{\int b p(b | c) db}{\int p(b | c) db} = \int b p(b | c) db = E_{b|c}[b], \quad (5.8)$$

т.е. оптимальная оценка представляет собой условное математическое ожидание величины b .

Так как правильно выбранная функция потерь $C(\beta, b)$ имеет минимум при $\beta = b$ и, по-видимому, должна быть достаточно мала как функция β в малой окрестности b и поскольку

$$\int_{m+1} p(b | y) d^{m+1}b = 1,$$

ясно, что условие (5.7) выполняется, если β выбрано в окрестности того значения b , которое максимизирует условную вероятность $p(b | y)$ (см. фиг. 5.6).

Отбросим теперь предположение в), т. е. информация о виде функции потерь $C(\beta, b)$ отсутствует. В этом случае разумно выбирать β , совпадающее со значением b , максимизирующим $p(b | y)$. В соответствии с формулой Байеса

$$p(b | y) = \frac{p(y | b) q(b)}{p(y)},$$

где

$$p(\mathbf{y}) = \int_{\mathbf{b}} p(\mathbf{y} | \mathbf{b}) q(\mathbf{b}) d^{m+1}\mathbf{b}.$$

По-прежнему используются предположения а) и б).

Априорная информация: $p(\mathbf{y} | \mathbf{b})$, $q(\mathbf{b})$.

Оценки максимального правдоподобия. Теперь отбросим предположения б) и в). Неизвестным считается и априорное распределение вероятностей $q(\mathbf{b})$ параметра \mathbf{b} . При этом можно положить $q(\mathbf{b}) = \text{const}$ в области возможных значений параметра \mathbf{b} . В этом случае для произвольного \mathbf{y} имеем

$$\max_{\mathbf{b}} p(\mathbf{b} | \mathbf{y}) = \frac{\text{const}}{p(\mathbf{y})} \max_{\mathbf{b}} p(\mathbf{y}; \mathbf{b}). \quad (5.9)$$

Здесь \mathbf{b} уже не случайная величина, а неизвестный постоянный параметр; вместо вертикальной черты появляется точка с запятой. Априорная плотность вероятности имеет вид

$$p\{y(1), \dots, y(k); \mathbf{b}\}.$$

Апостериори после измерений выборочных значений $y(1) = c_1, \dots, y(k) = c_k$, или $\mathbf{y} = \mathbf{c}$, плотность вероятности обозначается как

$$L\{c_1, \dots, c_k; \beta\}$$

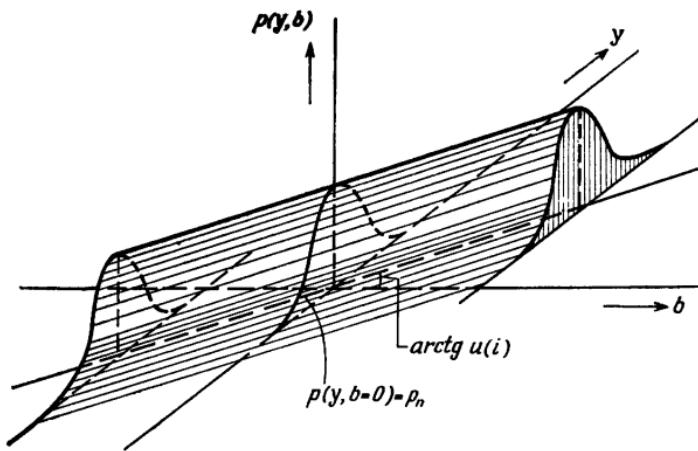
и называется *функцией правдоподобия* (см. разд. 2.3).

Пример. Опять обратимся к задаче, иллюстрированной фиг. 5.3. Известна плотность распределения вероятностей p_n помехи n . В этом случае измерение $u(i)$ дает априорную плотность вероятности $y(i)$ как функцию параметра b (фиг. 5.7). Измерение $y(i) = c_i$ приводит к ситуации, показанной на фиг. 5.8. По функции правдоподобия находится оценка $\hat{\beta}$. Логично выбирать такое значение $\hat{\beta}$, которое максимизирует L .

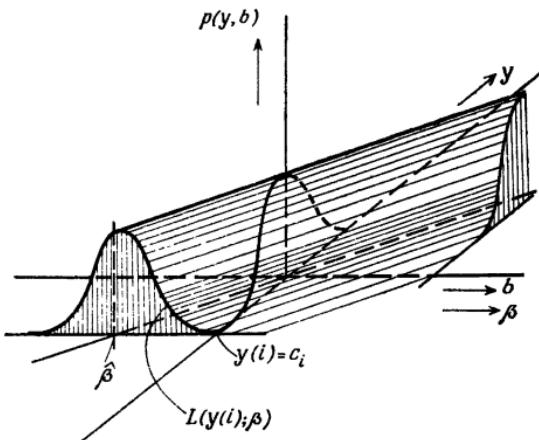
Априорная информация: $p(\mathbf{y}; \mathbf{b})$.

В многомерном случае необходимое условие максимума имеет вид

$$\frac{\partial}{\partial \beta} L(\mathbf{c}; \beta) |_{\beta=\hat{\beta}} = 0 \quad (5.10a)$$



Ф и г. 5.7.



Ф и г. 5.8.

или в силу монотонности логарифма

$$\frac{\partial}{\partial \beta} \ln L(\mathbf{c}; \boldsymbol{\beta}) |_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}} = 0. \quad (5.106)$$

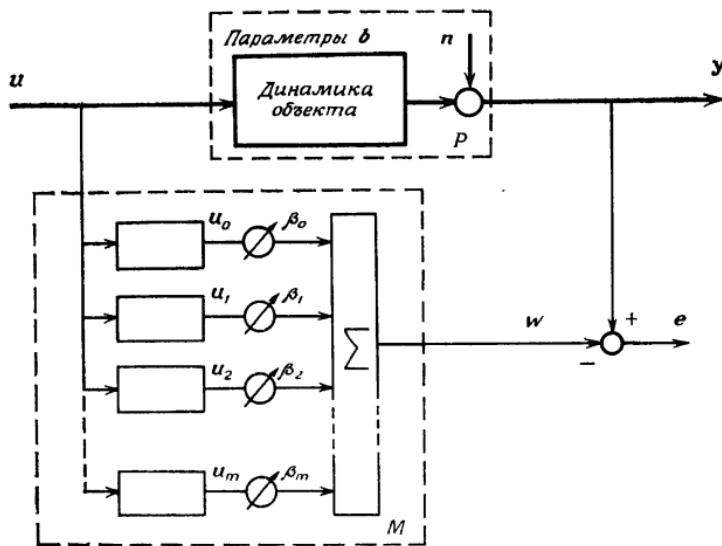
Эту формулу называют уравнением правдоподобия. Отыскивая решение этой системы уравнений, которое обеспечивает наибольшее значение $L(\mathbf{c}; \boldsymbol{\beta})$ или $\ln L(\mathbf{c}; \boldsymbol{\beta})$, находим *оценку максимального правдоподобия* (ОМП), которая широко обсуждалась в литературе. Некоторые из ее интересных свойств:

- 1) асимптотическая нормальность: $p(\hat{\beta}; \mathbf{b})$ приближается к нормальному распределению при $k \rightarrow \infty$ [см. (5.1)];
- 2) асимптотическая несмещенностъ; $E[\hat{\beta}] \rightarrow \mathbf{b}$ при $k \rightarrow \infty$;
- 3) асимптотическая эффективность: достижение наилучшей точности или минимальной дисперсии, определяемой неравенством Крамера — Рао (см. гл. 11);
- 4) состоятельность: см. определение;
- 5) инвариантность: если $\hat{\beta}$ есть ОМП параметра \mathbf{b} , то $g(\hat{\beta})$ — ОМП $g(\mathbf{b})$.

Связь разных оценок

Покажем, как связаны марковские оценки и оценки метода наименьших квадратов из разд. 2.2, которые обсуждаются в гл. 6, и оценки, рассматриваемые в этом разделе. Обратимся к фиг. 5.9 и предположим, что компоненты \mathbf{n} имеют *k-мерное гауссовское распределение*, т. е.

$$p(\mathbf{n}) = \frac{1}{(2\pi)^{k/2} |\mathbf{N}|^{1/2}} \exp \left[-\frac{1}{2} (\mathbf{n}' \mathbf{N}^{-1} \mathbf{n}) \right], \quad (5.11)$$



Ф и г. 5.9.

где

$$\mathcal{E}[n] = 0 \text{ и } \mathcal{E}[nn'] = N.$$

Отметим, что n является единственной случайной составляющей ошибки $e = y - w = y - U\beta$. Следовательно,

$$\ln p(y - U\beta) = -\frac{1}{2} \ln \{(2\pi)^k |N|\} - \\ -\frac{1}{2} \{(y - U\beta)' N^{-1} (y - U\beta)\}.$$

Максимизация этой функции дает

$$\frac{\partial}{\partial \beta} (y - U\beta)' N^{-1} (y - U\beta) |_{\beta=\hat{\beta}} = 0, \quad (5.12)$$

или

$$U'N^{-1}U\hat{\beta} - U'N^{-1}y = 0. \quad (5.13)$$

Если матрица $U'N^{-1}U$ имеет обратную, то уравнение (5.13) можно переписать в виде

$$\hat{\beta} = [U'N^{-1}U]^{-1}U'N^{-1}y. \quad (5.14)$$

Это *марковская оценка*. Она имеет следующие свойства:

1) линейность: $\hat{\beta} = Qy$;

2) несмещенность: $\mathcal{E}[\hat{\beta}] = b$;

3) минимальная дисперсия: среди всех линейных несмешенных оценок. Эту дисперсию можно найти по формуле

$$\text{cov}[\hat{\beta}] = \mathcal{E}[(\hat{\beta} - b)(\hat{\beta} - b)'].$$

Априорная информация: $N = \mathcal{E}[nn']$.

Если неизвестна и ковариационная матрица шума, то удобно выбрать $N^{-1} = \sigma^{-2}I$, где σ^2 — дисперсия шума. Это неявно означает, что шум белый. Следовательно,

$$\hat{\beta} = [U'U]^{-1}U'y. \quad (5.15)$$

Это формула *оценки по методу наименьших квадратов*.

Априорная информация отсутствует.

Марковские оценки и оценки по методу наименьших квадратов получены из оценок максимального правдоподобия в случае гауссовских шумов. (Предположение о нормальности распределения не является обязательным. В гл. 6 эти оценки получаются независимо от этого предположения.)

Пример. Связь между разными оценками можно продемонстрировать на простом примере, иллюстрированном фиг. 5.3. Оценка метода наименьших квадратов минимизирует следующую функцию потерь:

$$E(\beta) = \sum_{i=1}^k \{y(i) - u(i)\beta\}^2 = [\mathbf{y} - \mathbf{u}\beta]' [\mathbf{y} - \mathbf{u}\beta].$$

На фиг. 5.10 показаны разности между наблюдениями y и значениями выхода модели $u\beta$. Минимизацию можно осуществить аналитически. Необходимое условие минимума имеет вид

$$\frac{d}{d\beta} \sum_{i=1}^k \{y(i) - u(i)\beta\}^2 |_{\beta=\hat{\beta}} = 0,$$

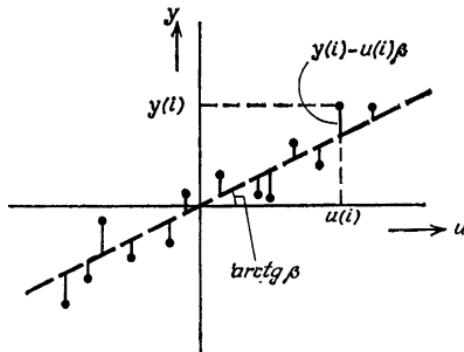
или

$$\sum_{i=1}^k u(i) \{y(i) - u(i)\hat{\beta}\} = 0, \quad (5.16)$$

или

$$\hat{\beta} = \frac{\sum_{i=1}^k u(i) y(i)}{\sum_{i=1}^k u(i) u(i)}. \quad (5.17)$$

При этом условии $\hat{\beta}$ — оптимальная оценка. Отметим, что в формуле (5.16) $y(i) - u(i)\beta$ имеют веса $u(i)$. Есте-



Ф и г. 5.10.

ственno, что, чем больше входной сигнал, тем больше вес расхождения между прогнозом $u(i)$ и наблюдением y . Формула (5.17) относится к корреляционной теории.

Некоторые другие вопросы

Простой пример на фиг. 5.3 очень хорошо иллюстрирует также так называемую проблему шума на входе. Рассмотрим систему, блок-схема которой изображена на фиг. 5.11. Ни входной, ни выходной сигналы этой системы не являются полностью наблюдаемыми. Известно [14], что, если относительно изменений u , n и v не сделано специальных предположений, оценить динамику системы невозможно. Рассмотрим, например, систему вида

$$\begin{aligned} y(i) &= bu(i) + n(i), \\ \tilde{u}(i) &= u(i) + v(i). \end{aligned}$$

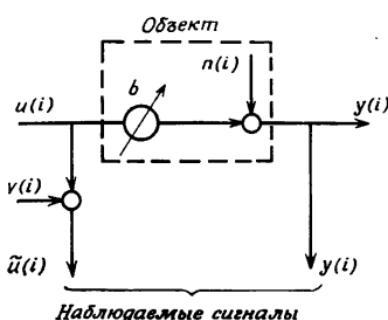
Допустим, что $v(i)$ и $n(i)$ — независимые случайные величины с нулевыми средними значениями. Если $v(i) = 0$ и $n(i) \neq 0$, то в соответствии с (5.17) оценка b определяется как

$$\hat{\beta} = \frac{\sum_{i=1}^k \tilde{u}(i) y(i)}{\sum_{i=1}^k \tilde{u}(i) \tilde{u}(i)}.$$

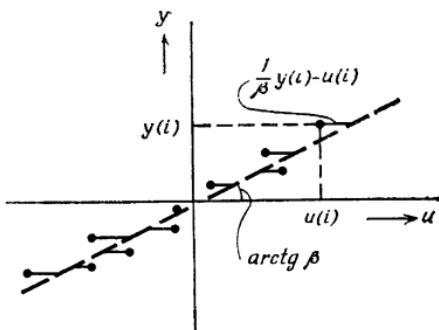
Однако если $n(i) = 0$ и $v(i) \neq 0$, то, рассуждая точно так же, получим оценку b в виде

$$\hat{\beta} = \frac{\sum_{i=1}^k y(i) y(i)}{\sum_{i=1}^k y(i) \tilde{u}(i)}.$$

Это соответствует такому выбору β , что средний квадрат разности между наблюдениями $\tilde{u}(i)$ и прогнозом $(1/\beta)y(i)$ оказывается минимальным (фиг. 5.12). Без дополнительной информации, конечно, невозможно выбрать лучшую из двух оценок.



Ф и г. 5.11.



Ф и г. 5.12.

Из формулы (5.17) для оценки метода наименьших квадратов, которая определяет явное или одноразовое решение, довольно просто получить *итеративное* решение, или решение с настройкой. Запишем выражения для $k + 1$ и k пар наблюдений:

$$\left\{ \sum_{i=1}^{k+1} u^2(i) \right\} \hat{\beta}(k+1) = \sum_{i=1}^{k+1} u(i) y(i),$$

$$\left\{ \sum_{i=1}^k u^2(i) \right\} \hat{\beta}(k) = \sum_{i=1}^k u(i) y(i).$$

Эти выражения для оценок $\hat{\beta}(k)$ и $\hat{\beta}(k+1)$ можно переписать в виде

$$\begin{aligned} & \left\{ \sum_{i=1}^{k+1} u^2(i) \right\} \hat{\beta}(k+1) = \sum_{i=1}^k u(i) y(i) + u(k+1) y(k+1) \\ & - \left\{ \sum_{i=1}^{k+1} u^2(i) \right\} \hat{\beta}(k) = \sum_{i=1}^k u(i) y(i) + u^2(k+1) \hat{\beta}(k) \end{aligned}$$

$$\left\{ \sum_{i=1}^{k+1} u^2(i) \right\} \{\hat{\beta}(k+1) - \hat{\beta}(k)\} = u(k+1) \{y(k+1) - u(k+1) \hat{\beta}(k)\},$$

или

$$\begin{aligned} \hat{\beta}(k+1) = & \hat{\beta}(k) + \left\{ \sum_{i=1}^{k+1} u^2(i) \right\}^{-1} \times \\ & \times u(k+1) \underbrace{\{y(k+1) - u(k+1) \hat{\beta}(k)\}}_{e(k+1)}. \end{aligned}$$

Таблица 5.1

Байесовские оценки	Безусловные оценки максимального правдоподобия	Условные оценки максимального правдоподобия	Марковские оценки	Оценки метода наименьших квадратов
$p(\mathbf{y} \mathbf{b})$ $q(\mathbf{b})$ $C(\mathbf{p}, \mathbf{b})$	$p(\mathbf{y} \mathbf{b})$ $q(\mathbf{b})$	$p(\mathbf{y}; \mathbf{b})$	$\mathbf{N} = \mathcal{E} [\mathbf{n} \mathbf{n}']$	Отсутствует
$\min_{\mathbf{b}} \int_{m+1}^{\infty} C(\mathbf{p}, \mathbf{b}) p(\mathbf{b} \mathbf{y}) d\mathbf{m+1} \mathbf{b}$	$\max_{\mathbf{b}} p(\mathbf{b} \mathbf{y}) =$ $= \max_{\mathbf{b}} \frac{p(\mathbf{y} \mathbf{b}) q(\mathbf{b})}{p(\mathbf{y})}$	$\max_{\mathbf{b}} p(\mathbf{y}; \mathbf{b})$ или $\max_{\mathbf{b}} \ln p(\mathbf{y}; \mathbf{b})$	$\hat{\mathbf{p}} = [\mathbf{U}' \mathbf{N}^{-1} \mathbf{U}]^{-1} \mathbf{U}' \mathbf{y}$	
Наблюдения $\mathbf{y} = \mathbf{c}$				
Минимальный риск или минимальные средние потери	Асимптотическая нормальность	Линейность Несмешенность Минимальная персия	Линейность Несмешенность	Линейность Несмешенность
См. текст главы				
При гауссовском шуме				При белом шуме

Последний член $e(k+1)$ представляет собой ошибку наблюдения $y(k+1)$ по сравнению с выходом модели $u(k+1)\hat{\beta}(k)$. Новая оценка состоит из старой оценки, скорректированной пропорционально ошибке $e(k+1)$. Величина веса, придаваемого ошибке, зависит от величины $u(k+1)$ и суммы квадратов всех предыдущих выборочных значений входного сигнала («энергетическая» характеристика входного сигнала).

Выше рассматривались только выборочные сигналы. Увеличивая число точек и уменьшая шаг дискретизации времени, можно вывести соответствующие формулы для непрерывных сигналов.

Некоторые из соотношений, связывающих различные типы оценок, представлены в табл. 5.1.

5.2. ДЕТЕРМИНИРОВАННЫЕ СХОДЯЩИЕСЯ АЛГОРИТМЫ¹⁾

Как указывалось в гл. 2, задача оценивания может быть сформулирована как задача минимизации ошибки по некоторому критерию. Критерий ошибки должен быть таким, чтобы его величина одинаково зависела от положительных и отрицательных значений ошибки. Такой критерий условно называется критерием четной функции ошибки. Примером подобного критерия может служить функционал вида

$$E\{\beta\} = \frac{1}{2} \int_0^T (y - w)' G (y - w) dt, \quad (5.18)$$

где

$y = y(b, t)$ — выход объекта,

$w = w(\beta, t)$ — соответствующий выход модели,

G — положительно определенная матрица весов различных y_i .

При оценивании параметров по настраиваемой модели, как и в случае оптимальных и самонастраивающихся систем, необходимо найти экстремум функции или функционала.

¹⁾ Этот раздел переработан и дополнен А. А. Г. Даменом (Технологический университет, г. Эйндховен, Голландия).

Пусть минимизируемый критерий ошибки есть

$$E = E(\beta_0, \beta_1, \dots, \beta_m) = E(\beta), \quad (5.19)$$

где β_j — настраиваемые параметры модели. В гл. 2 упоминались некоторые методы вычисления частных производных по этим параметрам. Эти производные можно использовать по-разному. Приведем обоснование достаточно общего подхода. Функция ошибок (5.19) может рассматриваться как гиперповерхность, определяемая уравнением

$$E(\beta) - E_0 = 0, \quad (5.20)$$

где E_0 — минимальное значение функции ошибок, которое нужно определить: $E_0 = E|_{\beta=b}$.

Функцию ошибок можно разложить в ряд Тейлора в окрестности точки минимума

$$E(\beta) = E_0 + \frac{\partial E}{\partial \beta} \Big|_{\beta=b} (\beta - b) + \frac{1}{2} (\beta - b)' \frac{\partial^2 E}{\partial \beta \partial \beta'} \Big|_{\beta=b} \times \quad (5.21a)$$

$\times (\beta - b) + \text{Члены более высокого порядка.}$
Поскольку E имеет минимум в точке $\beta = b$, то

$$\frac{\partial E}{\partial \beta} \Big|_{\beta=b} = 0$$

и величина

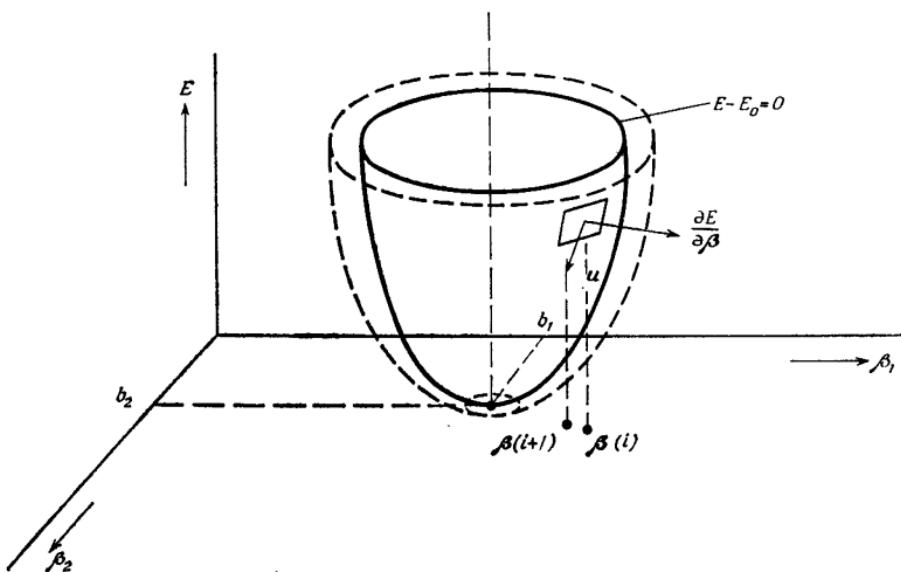
$$\frac{\partial^2 E}{\partial \beta \partial \beta'} \Big|_{\beta=b}$$

положительно определена. Кроме того, в большинстве случаев членами высшего порядка можно пренебречь (в противном случае см. ниже метод з). Следовательно,

$$E(\beta) \approx E_0 + \frac{1}{2} (\beta - b)' \frac{\partial^2 E}{\partial \beta \partial \beta'} \Big|_{\beta=b} (\beta - b). \quad (5.21b)$$

После подстановки (5.21б) уравнение (5.20) описывает гиперпараболоид (фиг. 5.13).

Для исследования сходимости рассматриваемых процедур важно знать поверхности уровня $E = \text{const}$. В квадратичном случае эти гиперповерхности представляют



Ф и г. 5.13.

собой многомерные эллипсоиды. Прежде чем перейти к обсуждению деталей, приведем пример квадратичной функции ошибок.

Пример. Рассмотрим простой пример довольно грубой аппроксимации. Допустим, что объект описывается следующим дифференциальным уравнением:

$$(a_n D^n + \dots + a_1 D + 1) y(t) = (b_m D^m + \dots + b_1 D + b_0) u(t), \quad (5.22)$$

где $D \triangleq d/dt$. Модель описывается уравнением

$$(\alpha_n D^n + \dots + \alpha_1 D + 1) w(t) = (\beta_m D^m + \dots + \beta_1 D + \beta_0) u(t). \quad (5.23)$$

Эти уравнения можно разрешить явно относительно $y(t)$ и $w(t)$. Когда помехи отсутствуют, в окрестности экстремума ошибки $e(t) = y(t) - w(t)$ мала. Если $e(t)$ достаточно мала, то $w(t)$ можно приближенно записать как

$$\tilde{w}(t) = -(a_n D^n + \dots + a_1 D) y(t) + (\beta_m D^m + \dots + \beta_1 D + \beta_0) u(t). \quad (5.24)$$

Отсюда

$$\begin{aligned} e(t) \approx y(t) - \tilde{w}(t) &= [(a_n - a_n)D^n + \dots \\ &\dots + (\alpha_1 - a_1)D]y(t) - [(\beta_m - b_m)D^m + \dots \\ &\dots + (\beta_1 - b_1)D + (\beta_0 - b_0)]u(t). \end{aligned} \quad (5.25)$$

Вблизи оптимума $e(t)$ является приближенно линейной функцией α и β . Вводя обозначения¹⁾

$$\left. \begin{array}{l} \alpha_1 - a_1 = \gamma_1, \\ \dots \dots \dots \dots \dots \\ \alpha_n - a_n = \gamma_n, \\ \beta_0 - b_0 = \gamma_{n+1}, \\ \dots \dots \dots \dots \dots \\ \beta_m - b_m = \gamma_{n+m+1} \end{array} \right\} \text{и} \left. \begin{array}{l} Dy(t) = v_1(t), \\ \dots \dots \dots \dots \dots \\ D^n y(t) = v_n(t), \\ u(t) = v_{n+1}(t), \\ Du(t) = v_{n+2}(t), \\ \dots \dots \dots \dots \dots \\ D^m u(t) = v_{n+m+1}(t), \end{array} \right.$$

получаем простые формулы

$$\begin{aligned} e(t) &\approx \sum_{i=1}^N \gamma_i v_i, \quad N = n + m + 1, \\ e^2(t) &\approx \sum_{i=1}^N \sum_{j=1}^N \gamma_i \gamma_j v_i v_j, \\ E &= \int_{t-T}^t e^2(t) dt = \sum_{i=1}^N \sum_{j=1}^N \gamma_i \gamma_j r_{ij}, \end{aligned} \quad (5.26)$$

где

$$r_{ij} = \int_{t-T}^t v_i(t) v_j(t) dt = r_{ji}.$$

Формулу (5.26) можно также переписать в виде

$$E = \boldsymbol{\gamma}' \mathbf{R} \boldsymbol{\gamma}, \quad (5.27)$$

где

$$\boldsymbol{\gamma} = \begin{pmatrix} \gamma_1 \\ \vdots \\ \gamma_N \end{pmatrix}, \quad \mathbf{R} = \begin{pmatrix} r_{11} & \dots & r_{1N} \\ \vdots & & \vdots \\ r_{N1} & \dots & r_{NN} \end{pmatrix}.$$

¹⁾ Практически производные определить трудно. Однако операция интегрирования функции или уравнения первого порядка $(1/(s + \sigma)$ вместо $1/s$, где s — оператор Лапласа) может быть реализована на аналоговой схеме (см. гл. 9).

Сравнение уравнений (5.27) и (5.21б) показывает, что они эквивалентны при $E_0 = 0$ и

$$R = \frac{1}{2} \left. \frac{\partial^2 E}{\partial \beta \partial \beta'} \right|_{\beta=b}.$$

Алгоритмы настройки делятся на *поисковые* и *градиентные*. В первом случае не требуется явного выражения для градиента функции ошибок или функции потерь, тогда как во втором случае используется допущение о возможности вычисления градиента. О методах оптимизации см. [2, 3, 7, 13, 18, 27].

Поисковые методы можно разбить на два подкласса:

а) *Табличные методы*. Функция ошибок оценивается в ряде точек пространства параметров, например в узлах регулярной решетки или в случайно выбранных точках.

б) *Итеративные методы*. Эти алгоритмы обеспечивают движение в направлении спуска, указываемом градиентом. Одним из примеров является симплекс-метод, при использовании которого вычисления начинаются с n равноудаленных одна от другой точек n -мерного пространства параметров. На каждой итерации новая точка находится «отражением» точки, в которой функция принимает наибольшее значение. Новая точка также равноудалена от всех остальных. После нескольких итераций расстояния уменьшаются во избежание возвращения в старые точки.

Другой подход связан с использованием методов случайного поиска, когда в каждой точке случайно ищется направление спуска к меньшим значениям функции. Существует много итеративных алгоритмов случайного поиска. Обзор различных методов можно найти в [25]. Случайный поиск особенно удобен, когда число параметров больше 20.

Градиентные методы бывают непрерывными и дискретными. Дискретность может быть связана с измерениями и настройкой.

Из *непрерывных* методов чаще всего применяется метод наискорейшего спуска. В этом случае движение происходит по траектории, которая при фиксированной скорости настройки обеспечивает наиболее быстрое уменьшение ошибки. Оказывается, что эта траектория в каждой точке ортогональна к изоповерхностям критерия E . В двумерном случае уравнение поверхности уровня $E = \text{const}$

имеет вид

$$dE = \frac{\partial E}{\partial \beta_1} d\beta_1 + \frac{\partial E}{\partial \beta_2} d\beta_2 = 0. \quad (5.28)$$

Уравнение касательной к линии уровня в точке $P = (\beta_{1p}, \beta_{2p})$ запишется как

$$\left(\frac{\partial E}{\partial \beta_1} \right)_P (\beta_1 - \beta_{1p}) + \left(\frac{\partial E}{\partial \beta_2} \right)_P (\beta_2 - \beta_{2p}) = 0. \quad (5.29)$$

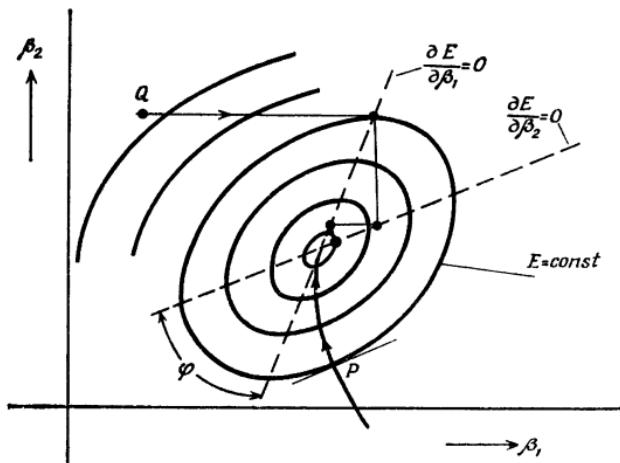
Вектор

$$\left(\frac{\partial E}{\partial \beta_1}, \frac{\partial E}{\partial \beta_2} \right)'$$

ортогонален к касательной и, следовательно, к линии уровня. В методе наискорейшего спуска настройка параметров производится по формуле

$$\begin{pmatrix} \dot{\beta}_1 \\ \dot{\beta}_2 \end{pmatrix} = -\Gamma \begin{pmatrix} \frac{\partial E}{\partial \beta_1} \\ \frac{\partial E}{\partial \beta_2} \end{pmatrix} = -\Gamma \frac{\partial E}{\partial \beta}. \quad (5.30)$$

Таким образом, траектория движения в каждой точке ортогональна к линиям уровня $E = \text{const}$ (фиг. 5.14). Здесь Γ — константа, которая вместе с частными производными определяет скорость изменения параметров. А так как частные производные не измеряются мгновенно, эту скорость



Ф и г. 5.14.

необходимо как-то ограничить. Если усиление Γ слишком велико, производные будут вычислены с большой ошибкой и движение будет происходить не в направлении наискорейшего спуска. При очень большом коэффициенте усиления система может даже потерять устойчивость. Для произвольных систем пока еще не существует общих методов анализа устойчивости, а рассматриваемые системы нестационарны и нелинейны. Однако для простых систем успешно применяются методы Ляпунова. Если же усиление невелико, то движение к оптимуму будет медленным.

Дискретные методы делятся на методы, в которых параметры настраиваются поочередно, и методы, в которых все параметры настраиваются одновременно.

а) **Циклическая настройка параметров.** Допустим, что β_1 — параметр, который настраивается первым. Одна из возможностей состоит в такой настройке параметра β_1 , которая обеспечивает равенство нулю $\partial E / \partial \beta_1$. Затем эта процедура повторяется для остальных параметров модели. В случае необходимости для удовлетворительной настройки модели циклы настройки повторяются несколько раз. В двумерном случае плоскость (β_1, β_2) покрывается сеткой линий уровня $E = \text{const}$. Как уже отмечалось, в окрестности оптимума линии уровня образуют семейство концентрических эллипсов, главные оси которых могут быть ориентированы произвольно (фиг. 5.14). Таким образом, при настройке параметры модели могут взаимодействовать, поэтому для перевода модели в окрестность оптимума в общем случае требуется несколько циклов. Если система ортогонализована, то достаточно однократной настройки каждого из параметров. Однако проблема состоит в том, что заранее не известно, какое линейное преобразование ортогонализует систему. Семейство эллипсов характеризуется углом между главными осями, которые являются геометрическим местом точек, удовлетворяющих одному из двух уравнений:

$$\frac{\partial E}{\partial \beta_1} = 0, \quad \frac{\partial E}{\partial \beta_2} = 0. \quad (5.31)$$

Чем меньше этот угол, тем больше циклов требуется для обеспечения заданной точности. На фиг. 5.14 показана возможная траектория настройки Q . Если линии уровня

представляют собой «овраги», то настройка может прекратиться при $E > E_0$. В этом недостаток метода.

В дальнейшем будут рассматриваться схемы с одновременной настройкой параметров. Более детальный анализ этих схем содержится в работе [20].

б) Метод наискорейшего спуска. Алгоритм настройки имеет вид

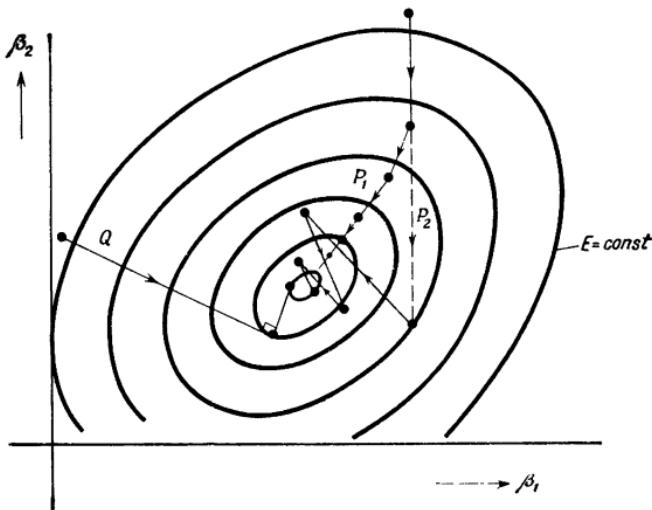
$$\beta(i+1) = \beta(i) - \Gamma \frac{\partial E}{\partial \beta} \Big|_{\beta=\beta(i)}, \quad \Gamma > 0,$$

где $\Gamma = \text{const}$. В зависимости от выбранного значения Γ приближение к оптимуму может быть медленным и гладким (Γ мало, см. траекторию P_1 на фиг. 5.15) или осциллирующим (траектория P_2). Если вычислить градиент невозможно, то его следует определить «экспериментально» путем малых вариаций в окрестности $\beta(i)$.

в) Наискорейший спуск с минимизацией вдоль направления движения. В этом случае имеем алгоритм

$$\beta(i+1) = \beta(i) - \Gamma(i) \frac{\partial E}{\partial \beta} \Big|_{\beta=\beta(i)}, \quad \Gamma(i) > 0.$$

Здесь $\Gamma(i)$ выбирается так, чтобы движение вдоль направления градиента продолжалось до достижения минимума. После этого вновь определяется направление градиента; траектория Q на фиг. 5.15 показана как пример такой



Ф и г. 5.15.

настройки. Отметим, что соседние участки траектории обязательно ортогональны. Вблизи минимума сходимость может быть медленной. Естественно поставить вопрос: в каких случаях при эллиптических линиях уровня обеспечивается попадание в оптимум за один шаг? В двумерном случае из формулы (5.26) следует, что

$$E = r_{11}\gamma_1^2 + 2r_{12}\gamma_1\gamma_2 + r_{22}\gamma_2^2, \quad (5.32)$$

$$\frac{\partial E}{\partial \beta_i} = \frac{\partial E}{\partial \gamma_i}, \quad \text{так как } \gamma_i = \beta_i - b_i,$$

$$\frac{\partial E}{\partial \gamma_1} = 2r_{11}\gamma_1 + 2r_{12}\gamma_2 \approx 2r_{11}\gamma_1, \quad \text{если } r_{12} \ll r_{11},$$

$$\frac{\partial E}{\partial \gamma_2} = 2r_{12}\gamma_1 + 2r_{22}\gamma_2 \approx 2r_{22}\gamma_2, \quad \text{если } r_{12} \ll r_{22}.$$

Для попадания в оптимум за один шаг необходимо, чтобы

$$-\gamma_1 = \Delta\gamma_1 = -\Gamma_1 \frac{\partial E}{\partial \gamma_1} \approx -2\Gamma_1 r_{11}\gamma_1, \quad \text{или } \Gamma_1 = \frac{1}{2r_{11}}, \quad (5.33)$$

$$-\gamma_2 = \Delta\gamma_2 = -\Gamma_2 \frac{\partial E}{\partial \gamma_2} \approx -2\Gamma_2 r_{22}\gamma_2, \quad \text{или } \Gamma_2 = \frac{1}{2r_{22}}.$$

Равенства $\Gamma_1 = \Gamma_2 = \Gamma$ приемлемы, если только $r_{11} \approx r_{22}$, в противном случае коэффициенты усиления нужно подбирать по r_{ij} .

Возвращаясь к геометрической интерпретации, отметим, что условия $r_{12} = r_{21} = 0$ и $r_{11} = r_{22}$ означают, что линии уровня — окружности, а в общем случае метод наискорейшего спуска оптимальен, если поверхности уровня являются гиперсферами. По формуле (5.216) поверхности уровня в окрестности минимума близки к гиперэллипсоидам, поэтому естественно попытаться преобразовать эти эллипсоиды в сферы, а затем в новом пространстве использовать метод наискорейшего спуска. В силу симметрии и положительной определенности матрицы $\partial^2 E / \partial \beta \partial \beta'$

$$\left. \frac{\partial^2 E}{\partial \beta \partial \beta'} \right|_{\beta=b} = S \Lambda S', \quad (5.34)$$

где Λ — диагональная матрица собственных значений, а S -матрица состоит из столбцов, образованных ортогональными собственными векторами. Подстановка (5.34) в (5.216) дает

$$E(\mathbf{z}^*) = E_0 + \frac{1}{2} \mathbf{z}^{*'} \Lambda \mathbf{z}^*, \quad (5.35)$$

где

$$\mathbf{z}^* = \mathbf{S}' (\boldsymbol{\beta} - \mathbf{b}).$$

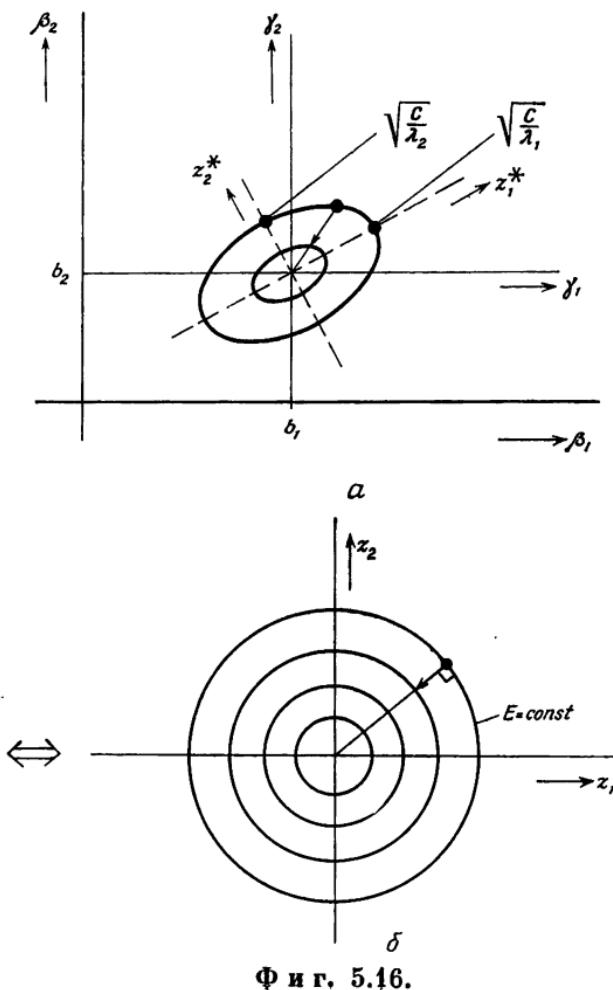
Определим

$$\mathbf{z} = \Lambda^{1/2} \mathbf{z}^*, \quad \mathbf{z}^* = \Lambda^{-1/2} \mathbf{z},$$

тогда

$$E(\mathbf{z}) = E_0 + \frac{1}{2} \mathbf{z}' \mathbf{z}. \quad (5.36)$$

В пространстве \mathbf{z} поверхности уровня являются сферами (фиг. 5.16). Алгоритм наискорейшего спуска в простран-



стве \mathbf{z} с $\Gamma = 1$ имеет вид

$$\mathbf{z}(i+1) = \mathbf{z}(i) - \frac{\partial E}{\partial \mathbf{z}} \Big|_{\mathbf{z}=\mathbf{z}(i)} = \mathbf{z}(i) - \mathbf{z}(i) = 0,$$

и в силу того, что $\mathbf{b} - \mathbf{b} = S\Lambda^{-1/2}\mathbf{z}$ с положительно определенными S и Λ , $\mathbf{b} - \mathbf{b} = 0$. Обратное преобразование к пространству β дает

$$\beta(i+1) = \beta(i) - \left[\frac{\partial^2 E}{\partial \beta \partial \beta'} \right]^{-1} \Big|_{\beta=\beta(i)} \frac{\partial E}{\partial \beta} \Big|_{\beta=\beta(i)}. \quad (5.37)$$

Вторая производная не зависит от β только для строго квадратичной функции ошибок, но в большинстве случаев можно предположить, что она почти не зависит от β , а это приводит к следующему методу.

г) Метод Ньютона — Рафсона. Алгоритм имеет вид

$$\beta(i+1) = \beta(i) - \left[\frac{\partial^2 E}{\partial \beta \partial \beta'} \right]^{-1} \Big|_{\beta=\beta(i)} \frac{\partial E}{\partial \beta} \Big|_{\beta=\beta(i)}. \quad (5.38)$$

Из этой формулы вытекает, что в методе Ньютона — Рафсона движение происходит по направлению к центру гиперэллипсоида, так как используется матрица вторых производных в точке $\beta(i)$. Это движение можно также интерпретировать как движение к минимуму гиперпараболоида, который является аппроксимацией второго порядка гиперповерхности, определяемой уравнением (5.20) (фиг. 5.17).

Разлагая критерий в ряд Тейлора в окрестности $\beta(i)$, получим

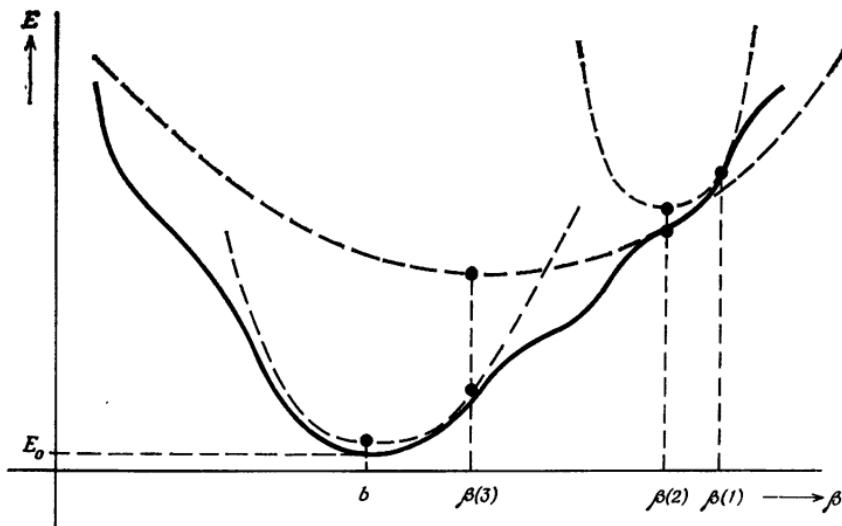
$$\begin{aligned} E(\beta) &= E[\beta(i)] + \frac{\partial E}{\partial \beta} \Big|_{\beta=\beta(i)} [\beta - \beta(i)] + \\ &+ \frac{1}{2} [\beta - \beta(i)]' \frac{\partial^2 E}{\partial \beta \partial \beta'} \Big|_{\beta=\beta(i)} [\beta - \beta(i)] + \dots \end{aligned} \quad (5.39)$$

Минимум достигается в точке $\beta = \beta(i+1)$, в которой $\partial E / \partial \beta \Big|_{\beta=\beta(i+1)} = 0$. Следовательно,

$$\frac{\partial E}{\partial \beta} \Big|_{\beta=\beta(i)} + \frac{\partial^2 E}{\partial \beta \partial \beta'} \Big|_{\beta=\beta(i)} [\beta(i+1) - \beta(i)] = 0,$$

что приводит к уравнению (5.38).

Оценивание вторых производных является достаточно трудной задачей, допускающей, правда, некоторые упрощения, когда E имеет вид критерия наименьших квадра-



Ф и г. 5.17.

тов (5.18). В этом случае

$$\frac{\partial E}{\partial \beta} = - \int_0^T \frac{\partial w'}{\partial \beta} G (y - w) dt,$$

$$\frac{\partial^2 E}{\partial \beta \partial \beta'} = \int_0^T \frac{\partial w'}{\partial \beta} G \frac{\partial w}{\partial \beta'} dt - \int_0^T \frac{\partial^2 w}{\partial \beta \partial \beta'} G (y - w) dt. \quad 1)$$

Вблизи минимума $y \approx w$ и вторым членом можно пренебречь. Подстановка в (5.38) приводит к следующему методу.

д) Метод Гаусса — Ньютона. Алгоритм имеет вид

$$\begin{aligned} \beta(i+1) &= \beta(i) + \left[\int_0^T \frac{\partial w'}{\partial \beta} G \frac{\partial w}{\partial \beta'} dt \right]_{\beta=\beta(i)}^{-1} \times \\ &\times \left[\int_0^T \frac{\partial w'}{\partial \beta} G (y - w) dt \right]_{\beta=\beta(i)}. \end{aligned} \quad (5.40)$$

¹⁾ Обозначения уже не являются правильными, так как $\partial^2 w / \partial \beta \partial \beta'$ — тензор.

Легко проверить, что это обычная оценка метода наименьших квадратов, если w линейна по β .

Метод Гаусса — Ньютона предпочтительнее метода наискорейшего спуска в силу квадратичной сходимости, хотя в отличие от метода наискорейшего спуска сходимость не гарантируется. Поэтому желательна комбинация этих методов, соединяющая их достоинства. Метод Маркуардта удовлетворяет этому условию.

е) Метод Маркуардта. Пусть $\beta(i)$ — центр гиперсферы в пространстве параметров. Воспользуемся методом множителей Лагранжа для решения задачи поиска минимума на гиперсфере при ограничении

$$\Delta\beta = \beta(i+1) - \beta(i), \quad \Delta\beta' \Delta\beta = r^2 = \text{const}.$$

Тогда

$$\frac{\partial E}{\partial(\Delta\beta)} + \mu \Delta\beta = 0. \quad (5.41)$$

Так как поверхность уровня не является идеальной гиперсферой, разложим w в ряд Тейлора

$$w(i+1) = w(i) + \left. \frac{\partial w}{\partial \beta'} \right|_{\beta=\beta(i)} \Delta\beta \text{ или } \left. \frac{\partial w(i+1)}{\partial(\Delta\beta)'} \right| \approx \left. \frac{\partial w}{\partial \beta'} \right|_{\beta=\beta(i)}.$$

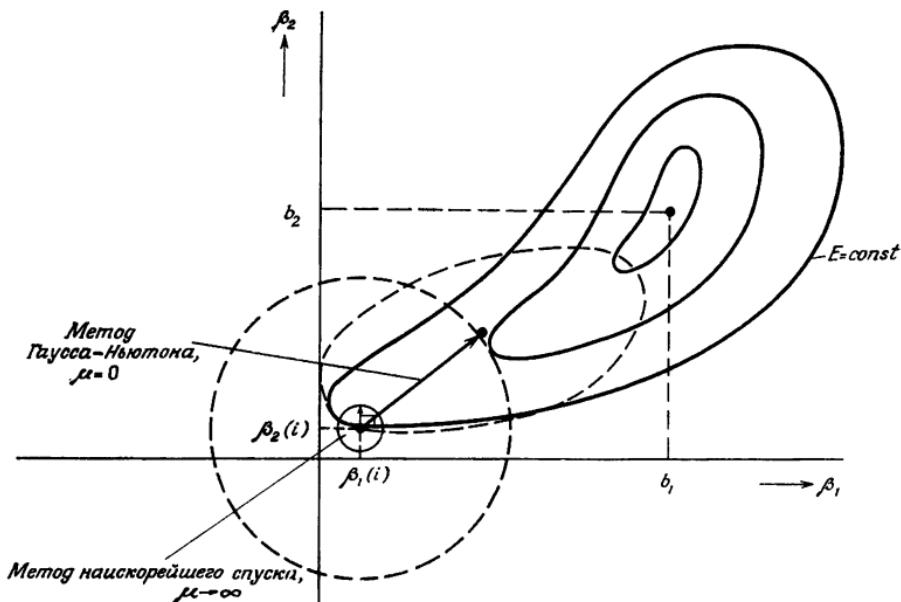
Следовательно, уравнение (5.41) запишется как

$$-\int_0^T \left. \frac{\partial w'}{\partial \beta} \right|_{\beta=\beta(i)} G \left(y - w(i) - \left. \frac{\partial w}{\partial \beta'} \right|_{\beta=\beta(i)} \Delta\beta \right) dt + \mu \Delta\beta = 0.$$

Отсюда следует, что

$$\begin{aligned} \beta(i+1) &= \beta(i) + \left[\int_0^T \left. \frac{\partial w'}{\partial \beta} \right|_{\beta=\beta(i)} G \left. \frac{\partial w}{\partial \beta'} \right|_{\beta=\beta(i)} dt + \mu I \right]^{-1} \times \\ &\quad \times \int_0^T \left. \frac{\partial w'}{\partial \beta} \right|_{\beta=\beta(i)} G [y - w(i)] dt. \end{aligned} \quad (5.42)$$

Дальнейший анализ этой формулы показывает, что при $\mu = 0$ имеем метод Гаусса — Ньютона, а при $\mu = \infty$ — метод наискорейшего спуска с нулевым шагом (фиг. 5.18). Изменяя μ , можно управлять сходимостью алгоритма. Недостаток состоит в необходимости приближенной оценки матрицы вторых производных E и последующего ее



Ф и г. 5.18.

обращения, особенно когда число параметров велико. Желательно строить обратную матрицу по информации, полученной на итерациях более простых методов. Это приводит к следующему методу.

ж) Метод сопряженных градиентов. Заметим, что для удобства сначала формулу (5.21б) можно записать как

$$E = E^* + \mathbf{a}'\beta + \frac{1}{2}\beta'\mathbf{A}\beta, \quad (5.43)$$

где

$$\mathbf{A} = \frac{\partial^2 E}{\partial \beta \partial \beta'}, \Big|_{\beta=\mathbf{b}}, \quad \mathbf{a} = -\mathbf{Ab}, \quad E^* = E_0 + \frac{1}{2}\mathbf{b}'\mathbf{Ab},$$

поэтому

$$\frac{\partial E}{\partial \beta} \Big|_{\beta=\beta(i)} = \mathbf{g}(i) = \mathbf{a} + \mathbf{A}\beta(i).$$

Направления \mathbf{u} и \mathbf{v} называются сопряженными по отношению к положительно определенной матрице \mathbf{A} , если

$$\mathbf{u}'\mathbf{A}\mathbf{v} = 0.$$

(Заметим, что ортогональные векторы являются сопряженными по отношению к единичной матрице.)

Положительно определенная матрица A размерности $n \times n$ имеет по меньшей мере один набор n попарно сопряженных векторов, в качестве которого можно выбрать собственные векторы A . Таким образом, в пространстве параметров вектор v , соединяющий начальную точку с точкой минимума, можно разложить по n независимым сопряженным направлениям d_i :

$$v = \sum_{i=1}^n \alpha_i d_i.$$

Модификация метода наискорейшего спуска заменой градиентного направления движения на сопряженное, как мы увидим, обеспечивает достижение минимума за n шагов для n -мерной квадратичной функции ошибок. Формула для вычисления итераций имеет вид

$$\beta(i+1) = \beta(i) + \lambda_i d(i), \quad (5.44)$$

где λ_i определяется положением локального минимума вдоль направления $d(i)$, что означает

$$g'(i+1)d(i) = 0. \quad (5.45a)$$

Так как

$$\beta(i+1) = \beta(s+1) + \sum_{j=s+1}^i \lambda_j d(j),$$

справедливо также равенство

$$g'(i+1)d(s) = 0, \quad s = 1, 2, \dots, i, \quad (5.45b)$$

поскольку

$$\begin{aligned} g(i+1) &= A\beta(i+1) + a = A \left\{ \beta(s+1) + \sum_{j=s+1}^i \lambda_j d(j) \right\} + a = \\ &= g(s+1) + \sum_{j=s+1}^i \lambda_j A d(j) \end{aligned}$$

и

$$d'(s) g(i+1) = d'(s) g(s+1) + \sum_{j=s+1}^i \lambda_j d'(s) A d(j).$$

Таким образом, на последнем шаге градиент ортогонален ко всем ранее использованным направлениям. Поэтому, когда на очередном шаге при определении сопряженного направления $\mathbf{d}(i+1)$ учитывается градиент $\mathbf{g}(i+1)$, это придает уверенность в том, что исследуется новое измерение, и поясняет гипотетическую формулу

$$\mathbf{d}(i+1) = -\mathbf{g}(i+1) + \gamma_{i+1} \mathbf{d}(i). \quad (5.46)$$

Подстановкой эту гипотезу легко проверить, что дает

$$\gamma_{i+1} = \frac{\mathbf{g}'(i+1) \mathbf{g}(i)}{\mathbf{g}'(i) \mathbf{g}(i)}. \quad (5.47)$$

Начиная на первом шаге с направления, определяемого градиентом, после n итераций имеем n независимых сопряженных направлений. На последнем шаге градиент в соответствии с (5.45б) должен быть ортогонален ко всем этим независимым направлениям, но в n -мерном пространстве это невозможно. Следовательно, градиент равен нулю и минимум достигнут. Этот алгоритм получен в работе [9].

Чаще используется алгоритм Дэвидона [8]. Как уже отмечалось, идея этого метода состоит в уточнении обратной матрицы \mathbf{A}^{-1} на каждом шаге. Явное выражение для обратной матрицы можно записать с помощью сопряженных направлений:

$$\mathbf{H} = \sum_{i=1}^n \frac{\mathbf{d}(i) \mathbf{d}'(i)}{\mathbf{d}'(i) \mathbf{Ad}(i)} \Rightarrow \mathbf{H} \mathbf{Ad}(s) = \mathbf{d}(s), \quad s = 1, 2, \dots, n,$$

или

$$\mathbf{H} \mathbf{A} = \mathbf{I} \Rightarrow \mathbf{H} = \mathbf{A}^{-1}.$$

Можно доказать, что

$$\mathbf{d}(i) = -\mathbf{H}(i-1) \mathbf{g}(i) \quad (5.48)$$

и

$$\mathbf{H}(i) = \mathbf{H}(i-1) + \lambda_i \frac{\mathbf{d}(i) \mathbf{d}'(i)}{\mathbf{g}'(i) \mathbf{H}(i-1) \mathbf{g}(i)} - \frac{\mathbf{H}(i-1) \mathbf{y}(i) \mathbf{y}'(i) \mathbf{H}(i-1)}{\mathbf{y}'(i) \mathbf{H}(i-1) \mathbf{y}(i)},$$

где

$$\mathbf{y}(i) = \mathbf{g}(i+1) - \mathbf{g}(i).$$

На n -м шаге $\mathbf{H} = \mathbf{A}^{-1}$, что в случае квадратичного критерия означает достижение минимума. Недостаток этого метода состоит в том, что на каждом шаге нужно решать задачу минимизации по направлению. В работе [4] предложен подобный метод, но без минимизации. Этот алгоритм приводит в точку минимума за столько же шагов, но на каждом шаге используется одна функция и один градиент.

Все перечисленные методы хорошо применимы для случая почти квадратичных поверхностей, а также, по-видимому, и для многих других гладких поверхностей, в том числе на наборах однородных функций:

$$E(\beta) = \frac{1}{\gamma} (\beta - \mathbf{b})' \mathbf{g}(\beta) + E_0^*, \quad (5.49)$$

где

$$\mathbf{g}(\beta) = \frac{\partial E}{\partial \beta}.$$

Многообещающий алгоритм минимизации таких однородных функций основан на следующем методе.

3) Метод Якобсона — Оксмана [11]. Идея метода удивительно проста. Предположим, что в точках $\beta(i)$ известны значения функции и ее градиента и в этих точках выражение (5.49) можно записать в виде

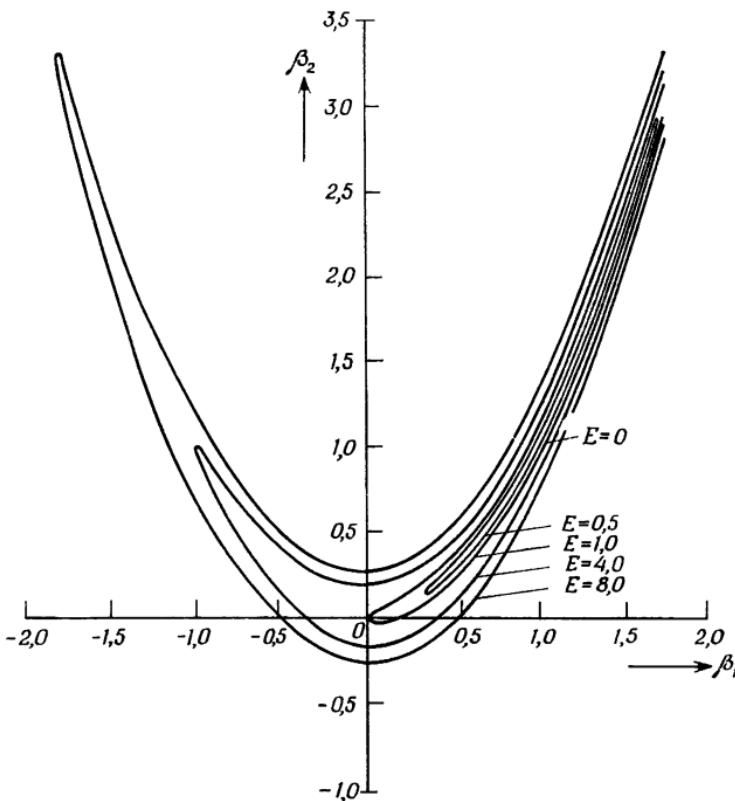
$$\mathbf{g}'(\beta(i))\mathbf{b} + E(\beta(i))\gamma - \gamma E_0^* = \mathbf{g}'(\beta(i))\beta(i),$$

$$i = 1, 2, \dots, n+2.$$

Это система $n+2$ (или большего числа) линейных уравнений относительно $n+2$ параметров \mathbf{b} , γ и γE_0^* . Если подходящим выбором вектора $\beta(i)$ можно регулировать матрицу коэффициентов, то решение получается с помощью обратной или псевдообратной матрицы. Конечно, эта идея может быть реализована схемой корректировки, что и было сделано Якобсоном и Оксманом. На многих тестовых задачах они получили лучшие результаты, чем с помощью методов Паузлла, Флетчера и Дэвидона.

Одним из наиболее серьезных тестов для проверки градиентных алгоритмов является так называемый «овраг» Розенброка, определяемый формулой (см. [20])

$$E = 100(\beta_1^2 - \beta_2)^2 + (1 - \beta_1)^2.$$



Ф и г. 5.19 (из работы [20]).

Эта поверхность изображена на фиг. 5.19. В работе [20] проведено сравнение нескольких алгоритмов, проверенных на ряде разных «оврагов».

Примеры.

Для того чтобы дать наглядное представление о возможностях различных методов, было предпринято моделирование этих методов на небольшой вычислительной машине¹⁾. Рассматривались квадратичная функция

$$E = \frac{1}{2} \{ 5\beta_1^2 + 6\beta_1\beta_2 + 5\beta_2^2 \}$$

¹⁾ Автор признателен И. Крегтинг-Янсен, А. А. Дамену, П. Дамману и К. де Грооту за эту работу.

и «овраг» Розенброка. Для каждого алгоритма проводилось по 20 итераций. На фиг. 5.20—5.29 представлены результаты, полученные при рассмотрении квадратичной функции (а) и «оврага» Розенброка (б).

Поисковые итеративные методы

Метод Хука — Дживса [13].

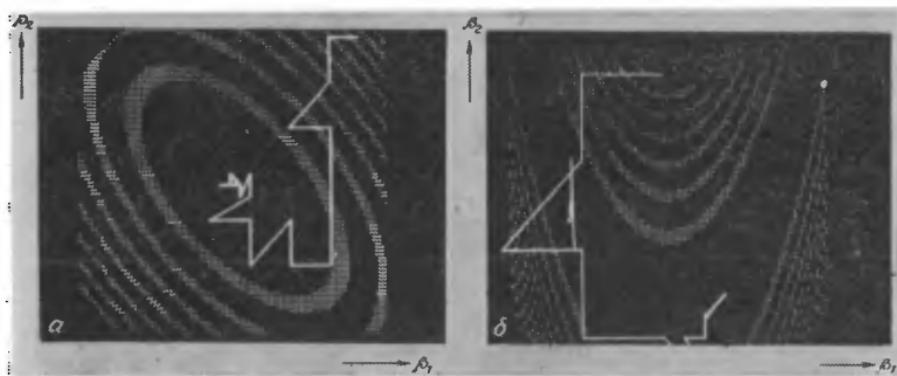
Начальная точка:

$$\text{Фиг. 5.20, а} \begin{cases} \beta_1 = 10; \\ \beta_2 = 10. \end{cases} \quad \text{Фиг. 5.20, б} \begin{cases} \beta_1 = -1,2; 0; \\ \beta_2 = 1; 3. \end{cases}$$

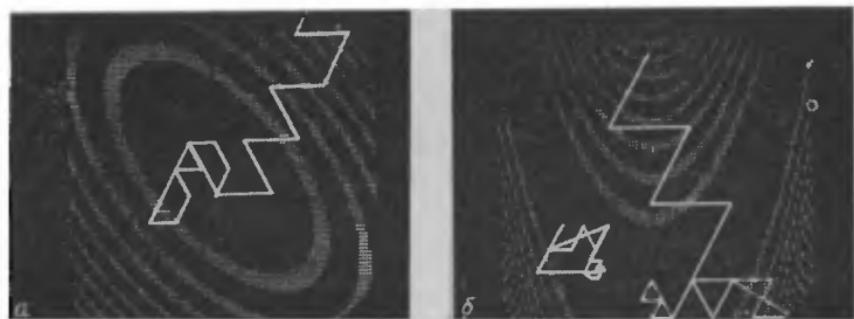
Симплекс-метод [3].

Начальная точка:

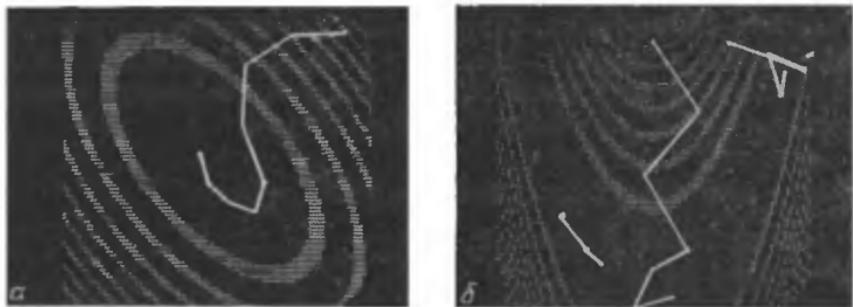
$$\text{Фиг. 5.21, а} \begin{cases} \beta_1 = 8; \\ \beta_2 = 12. \end{cases} \quad \text{Фиг. 5.21, б} \begin{cases} \beta_1 = -1,2; 0; \\ \beta_2 = 1; 3. \end{cases}$$



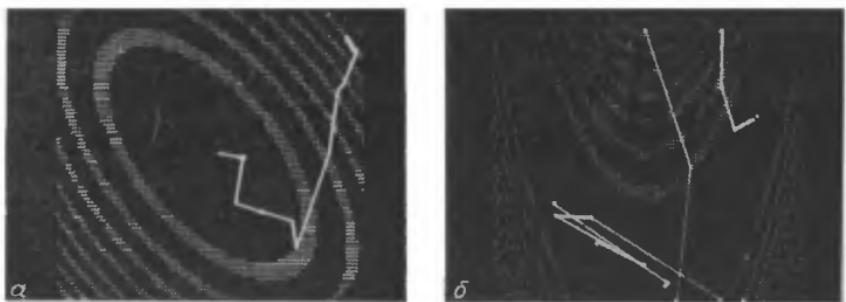
Ф и г. 5.20.



Ф и г. 5.21.



Ф и г. 5.22.



Ф и г. 5.23.

Случайный поиск [25].

Начальная точка:

$$\text{Фиг. 5.22, } a \begin{cases} \beta_1 = 10; \\ \beta_2 = 8. \end{cases} \quad \text{Фиг. 5.22, } b \begin{cases} \beta_1 = -1,2; 0; 1; \\ \beta_2 = 1; 3; 3. \end{cases}$$

Адаптивный случайный поиск [25].

Начальная точка:

$$\text{Фиг. 5.23, } a \begin{cases} \beta_1 = 10; \\ \beta_2 = 8. \end{cases} \quad \text{Фиг. 5.23, } b \begin{cases} \beta_1 = -1,2; 0; 1; \\ \beta_2 = 1; 3; 3. \end{cases}$$

Градиентные методы с дискретной настройкой

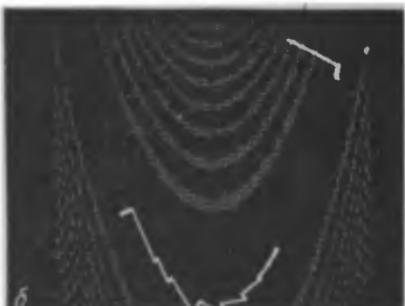
Наискорейший спуск.

Начальная точка:

$$\text{Фиг. 5.24, } a \begin{cases} \beta_1 = -6; \\ \beta_2 = 12. \end{cases} \quad \text{Фиг. 5.24, } b \begin{cases} \beta_1 = -1,2; 0; 1; \\ \beta_2 = 1; 3; 3. \end{cases}$$



Ф и г. 5.24.



Ф и г. 5.25.

Наискорейший спуск с минимизацией вдоль направления движения.

Начальная точка:

$$\text{Фиг. 5.25, } a \begin{cases} \beta_1 = -8; \\ \beta_2 = 12. \end{cases} \quad \text{Фиг. 5.25, } b \begin{cases} \beta_1 = -1,2; & 1^1); \\ \beta_2 = 1; & 3. \end{cases}$$

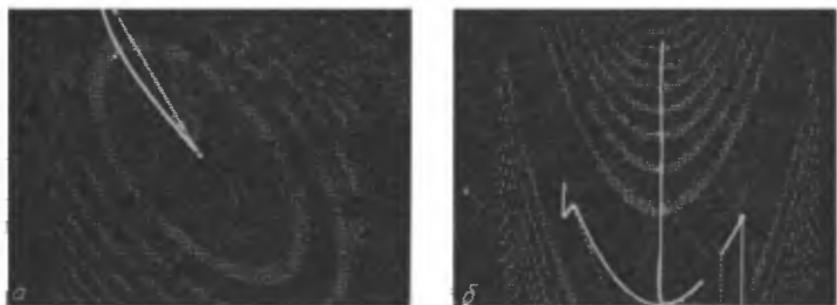
Метод Маркуардта.

Начальная точка:

$$\text{Фиг. 5.26, } a \begin{cases} \beta_1 = -8; 8; \\ \beta_2 = 12; 12; \\ \mu = 100; 0^2). \end{cases}$$

¹⁾ Последовательные шаги не ортогональны, так как минимизация проводится по интерполированной зависимости третьего порядка и, следовательно, является приближенной.

²⁾ При $\mu = 0$ переходит в метод Ньютона!



Ф и г. 5.26.



Ф и г. 5.27.

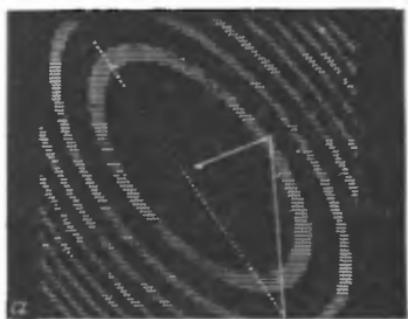
$$\text{Фиг. 5.26, б} \quad \begin{cases} \beta_1 = -1,2; & 0; -1,2; 0; \\ \beta_2 = 1; & 3; 1; 3; \\ \mu = 100; & 1000; 0; 0. \end{cases}$$

Метод Флетчера — Пауэлла (см. [18]).
Начальная точка:

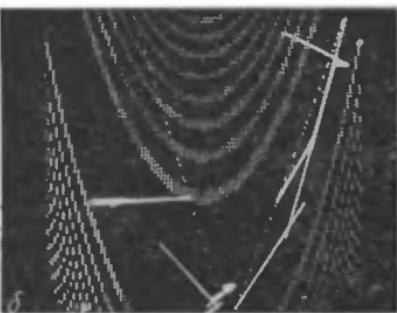
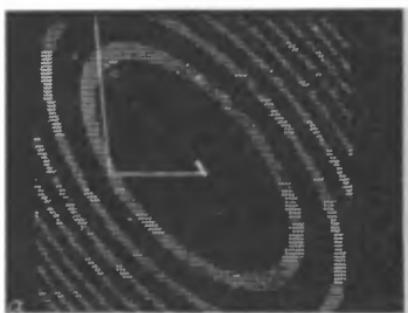
$$\text{Фиг. 5.27, а} \quad \begin{cases} \beta_1 = -8; \\ \beta_2 = 12. \end{cases} \quad \text{Фиг. 5.27, б} \quad \begin{cases} \beta_1 = -1,2; 0; 1; \\ \beta_2 = 1; 3; 3. \end{cases}$$

Метод Дэвидона (см. [4]).
Начальная точка:

$$\text{Фиг. 5.28, а} \quad \begin{cases} \beta_1 = -8; \\ \beta_2 = 8,5. \end{cases} \quad \text{Фиг. 5.28, б} \quad \begin{cases} \beta_1 = -1,2; 1; \\ \beta_2 = 1; 3. \end{cases}$$



Ф и г. 5.28.



Ф и г. 5.29.

Метод Якобсона — Оксмана (см. [11]).

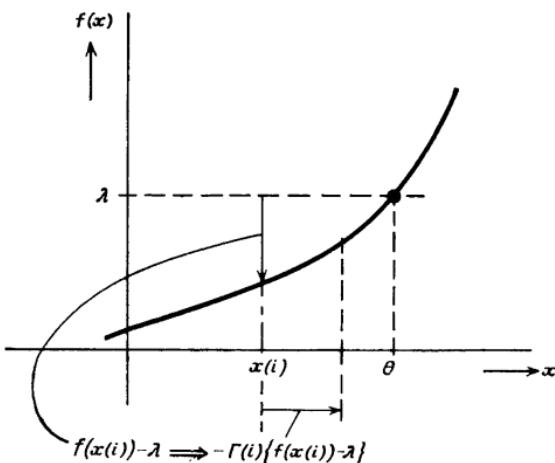
Начальная точка:

$$\text{Фиг. 5.29, } a \begin{cases} \beta_1 = -8; \\ \beta_2 = 12. \end{cases} \quad \text{Фиг. 5.29, } b \begin{cases} \beta_1 = -1,2; \\ \beta_2 = 1; \end{cases} \begin{matrix} 1 \\ 3. \end{matrix}$$

Читателю рекомендуется снова просмотреть все алгоритмы настройки, используя в качестве иллюстраций предыдущие примеры.

5.3. ВЕРОЯТНОСТНЫЕ АЛГОРИТМЫ; СТОХАСТИЧЕСКАЯ АППРОКСИМАЦИЯ

В разд. 5.2 при вычислении градиентов не учитывались возмущения аддитивного характера. Рассматриваемая в этом разделе стохастическая аппроксимация может



Ф и г. 5.30.

быть использована для решения различных задач, связанных с оптимизацией, адаптацией, настройкой моделей и обучением при наличии шумов [24].

В 1951 г. Роббинс и Монро [19] предложили подход к решению задачи об отыскании корня уравнения регрессии. Рассматривается условная плотность вероятности $p(y|x)$ или условная функция распределения $F(y|x)$ и предполагается, что:

- существует функция регрессии $f(x)$;
- уравнение $f(x) = \lambda$ имеет единственное решение $x = \theta$. Последнее условие означает, что $f(x)$ — монотонная функция.

Корень θ уравнения $f(x) = \lambda$ предлагается искать с помощью следующего рекуррентного алгоритма:

$$x(i+1) = x(i) - \Gamma(i)\{f(x(i)) - \lambda\}, \quad (5.50)$$

где $x(i)$ — оценка на i -м шаге, наблюдение $f(x(i))$ искажено *аддитивным* случайным шумом (фиг. 5.30), а $\Gamma(i)$, $i = 1, 2, \dots$ — последовательность действительных чисел, свойства которой будут определены ниже. Эта последовательность должна быть выбрана так, чтобы $x(i)$ сходились по вероятности к θ , т. е. для всех $\varepsilon > 0$:

$$\lim_{i \rightarrow \infty} P\{|x(i) - \theta| \geq \varepsilon\} = 0. \quad (5.51)$$

Из этого условия вытекают следующие ограничения. Рассмотрим выражение $\{x(i+1) - \theta\}^2$ совместно с уравнением (5.50). Вычислим математическое ожидание

$$\begin{aligned}\mathbb{E}[\{x(i+1) - \theta\}^2] &= \mathbb{E}[\{x(i) - \theta\}^2] + \\ &+ \Gamma^2(i) \mathbb{E}[\{f(x(i)) - \lambda\}^2] - 2\Gamma(i) \mathbb{E}[\{x(i) - \theta\} \{f(x(i)) - \lambda\}]\end{aligned}\quad (5.52)$$

или

$$b(i+1) = b(i) + \Gamma^2(i)e(i) - 2\Gamma(i)d(i), \quad (5.52a)$$

где

$$\begin{aligned}b(i) &= \mathbb{E}[\{x(i) - \theta\}^2] \geq 0, \\ e(i) &= \mathbb{E}[\{f(x(i)) - \lambda\}^2] \geq 0, \\ d(i) &= \mathbb{E}[\{x(i) - \theta\} \{f(x(i)) - \lambda\}], \\ \Gamma(i) &> 0.\end{aligned}$$

Для существования $e(i)$ необходимо, чтобы $f(x(i))$ была конечной, за исключением, быть может, конечного числа точек. Потребуем, чтобы $b(i+1) < b(i)$. Отсюда $d(i) > 0$. Это означает, что $f(x)$ должна быть монотонной по крайней мере в окрестности θ . Записывая выражения для $b(2)$, $b(3)$, ..., $b(i+1)$, находим, что

$$b(i+1) = b(1) + \sum_{j=1}^i \Gamma^2(j) e(j) - 2 \sum_{j=1}^i \Gamma(j) d(j) \quad (5.53)$$

или

$$\begin{aligned}\sum_{j=1}^i \Gamma(j) d(j) &= \frac{1}{2} \left[b(1) + \sum_{j=1}^i \Gamma^2(j) e(j) - b(i+1) \right] \leqslant \\ &\leqslant \frac{1}{2} \left[b(1) + \sum_{j=1}^i \Gamma^2(j) e(j) \right].\end{aligned}\quad (5.54)$$

При условии, что $e(j)$ конечна и

$$\sum_{j=1}^{\infty} \Gamma^2(j) < \infty, \quad (5.55)$$

находим, что

$$\sum_{j=1}^{\infty} \Gamma(j) d(j) < \infty. \quad (5.55a)$$

Если $\Gamma(j)$ выбрано так, что

$$\sum_{j=1}^{\infty} \Gamma(j) = \infty, \quad \text{то } \lim_{j \rightarrow \infty} d(j) = 0. \quad (5.56)$$

Следовательно,

$$\lim_{i \rightarrow \infty} \mathcal{E} [\{x(i) - \theta\} \{f(x(i)) - \lambda\}] = 0. \quad (5.57)$$

Разлагая $f(x(i))$ в ряд в окрестности $x = \theta$, находим, что

$$f(x(i)) = f(\theta) + \frac{f'(\theta)}{1!} \{x(i) - \theta\} + O \{x(i) - \theta\}^2, \quad (5.58)$$

или

$$f(x(i)) - \lambda = \frac{f'(\theta)}{1!} \{x(i) - \theta\} + O \{x(i) - \theta\}^2. \quad (5.59)$$

Подстановка в уравнение (5.57) дает

$$\lim_{i \rightarrow \infty} \mathcal{E} [f'(\theta) \{x(i) - \theta\}^2] = 0, \quad (5.60)$$

или

$$\lim_{i \rightarrow \infty} P \{|x(i) - \theta| \geq \varepsilon\} = 0. \quad (5.61)$$

Условиям (5.55), (5.55a) и (5.56) удовлетворяет, например, последовательность $\Gamma(j) = c/j^\alpha$ при $0.5 < \alpha \leq 1$.

Как уже отмечалось, первоначально этот метод был предложен для решения уравнений с монотонной функцией в левой части. Затем Кифер и Вольфович [12] применили этот метод для отыскания экстремума функции регрессии. Так как функция регрессии не известна, ее производная оценивается по формуле

$$\nabla_x f(x(i)) \approx \frac{f(x(i) + a(i)) - f(x(i) - a(i))}{2a(i)}. \quad (5.62)$$

При этом преобразовании четная функция заменяется на нечетную (см. гл. 2). Алгоритм определения точки минимума имеет вид

$$x(i+1) = x(i) - \Gamma(i) \nabla_x f(x(i)). \quad (5.63)$$

Этот алгоритм сходится, если выполнены все условия, накладываемые на $f(x)$ и $\Gamma(i)$. Кифер и Вольфович распространяли этот метод на многомерный и непрерывный случай. Задача в многомерном случае формулируется

следующим образом. Необходимо максимизировать (минимизировать) по x функцию

$$F(x) = F(x_1, x_2, \dots, x_N). \quad (5.64)$$

Допустим, что \hat{x} — точка максимума. Если $F(x)$ известна и нет ограничений и седловых точек, то \hat{x} определяется из следующей системы уравнений:

$$\nabla_x F(x) = \left\{ \frac{\partial F(x)}{\partial x_1}, \dots, \frac{\partial F(x)}{\partial x_N} \right\}' = 0 \text{ для } x = \hat{x}.$$

В этом случае алгоритм вычисления \hat{x} имеет вид

$$x(i+1) = x(i) - \Gamma(i) \nabla_x F(x(i)). \quad (5.65)$$

Выбор $\Gamma(i)$ определяет итеративный метод. Если выполнены условия сходимости, то

$$x(i) \rightarrow \hat{x} \text{ при } i \rightarrow \infty.$$

Если $F(x)$ не известна, то градиент можно оценить по формуле

$$\begin{aligned} \nabla_x F(x) &\approx \Delta F'(x, \pm a) = \\ &= \left\{ \frac{F(x+a e_1) - F(x-a e_1)}{2a}, \dots, \frac{F(x+a e_N) - F(x-a e_N)}{2a} \right\}', \end{aligned} \quad (5.66)$$

где e_1, \dots, e_N — базис, а a — небольшое вещественное число, которое может зависеть от номера итерации и от компонент вектора $x : x_j$.

Если, кроме того, $F(x)$ — случайная и задана функция распределения $F(n | x)$, где n — случайный вектор, то лучшее, чего можно в этом случае добиться, — это максимизировать (минимизировать) математическое ожидание

$$\int_n F(n | x) p(n) dn = E_n [F(n | x)]. \quad (5.67)$$

Тогда $x = \hat{x}$ есть корень уравнения

$$\nabla_x E_n [F(n | x)]|_{x=\hat{x}} = 0,$$

который определяется по следующему алгоритму:

$$x(i+1) = x(i) - \Gamma(i) \nabla_x E_n [F(n | x(i))]. \quad (5.68)$$

Если $p(n)$ и $F(x)$ заранее не известны, то, используя вместо градиента его оценку, получаем

$$\mathbf{x}(i+1) = \mathbf{x}(i) - \Gamma(i) \frac{F(\mathbf{n} | \mathbf{x}(i) + \mathbf{a}(i)) - F(\mathbf{n} | \mathbf{x}(i) - \mathbf{a}(i))}{2\mathbf{a}(i)}. \quad (5.69)$$

Здесь $\mathbf{x}(i) \rightarrow \hat{\mathbf{x}}$, т. е.

$$\lim_{i \rightarrow \infty} \mathcal{E}[|\mathbf{x}(i) - \hat{\mathbf{x}}|^2] = 0, \quad (5.70)$$

если

$$\sum_{i=1}^{\infty} \Gamma(i) = \infty, \quad \sum_{i=1}^{\infty} \frac{\Gamma^2(i)}{\mathbf{a}^2(i)} < \infty, \quad \sum_{i=1}^{\infty} \Gamma(i) \mathbf{a}(i) < \infty.$$

Эти условия означают, что для того чтобы дисперсия оценок стремилась к нулю, $\Gamma(i)$ должны достаточно быстро убывать, но не слишком быстро, чтобы не утратить в процессе обработки полезной информации. Если объект детерминирован, то можно опустить требования к $\Gamma(i)$ и выбирать Γ равной константе.

Пример. Возможности стохастической аппроксимации иллюстрируются результатами простого эксперимента ¹⁾, приведенными на фиг. 5.31 и 5.32. Функция F выбрана одномерной,

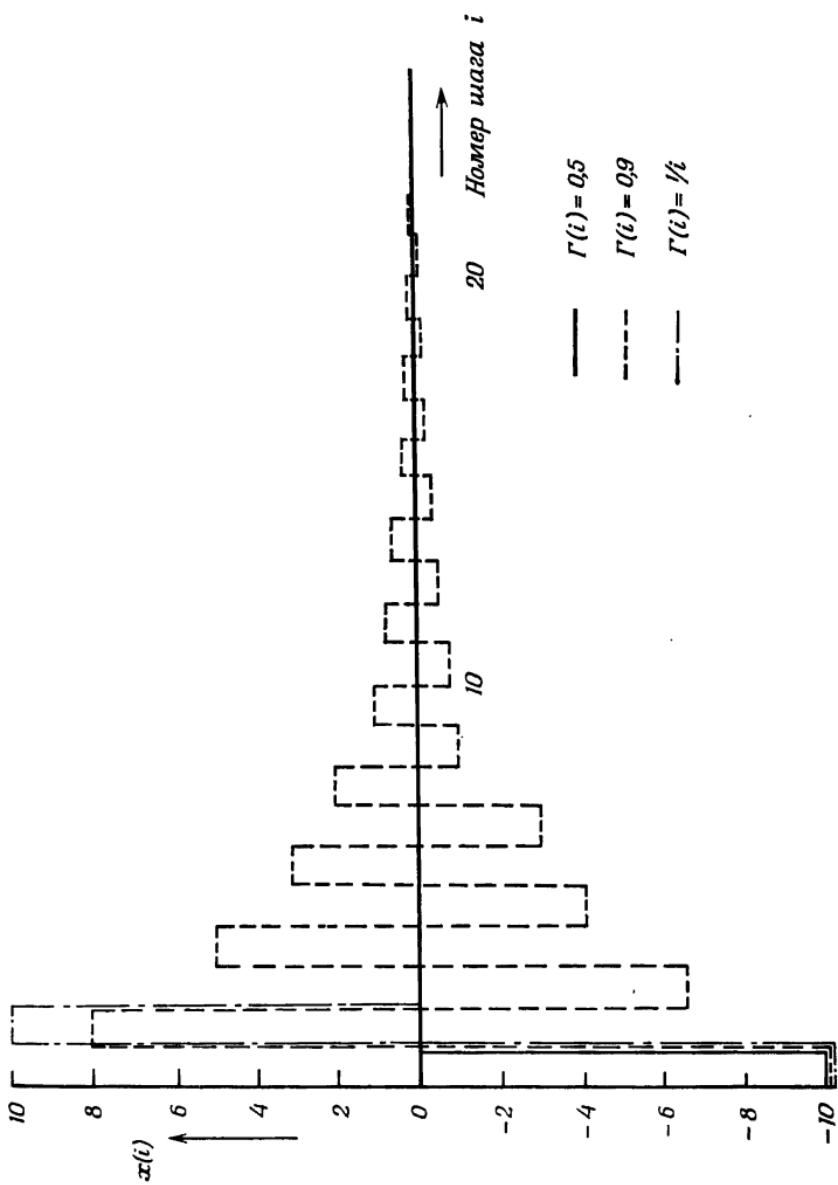
$$F(n(i) | x(i)) = 1 + x^2(i) + \lambda n(i),$$

$$p(n) = 1 \text{ для } 0 \leq n \leq 1,$$

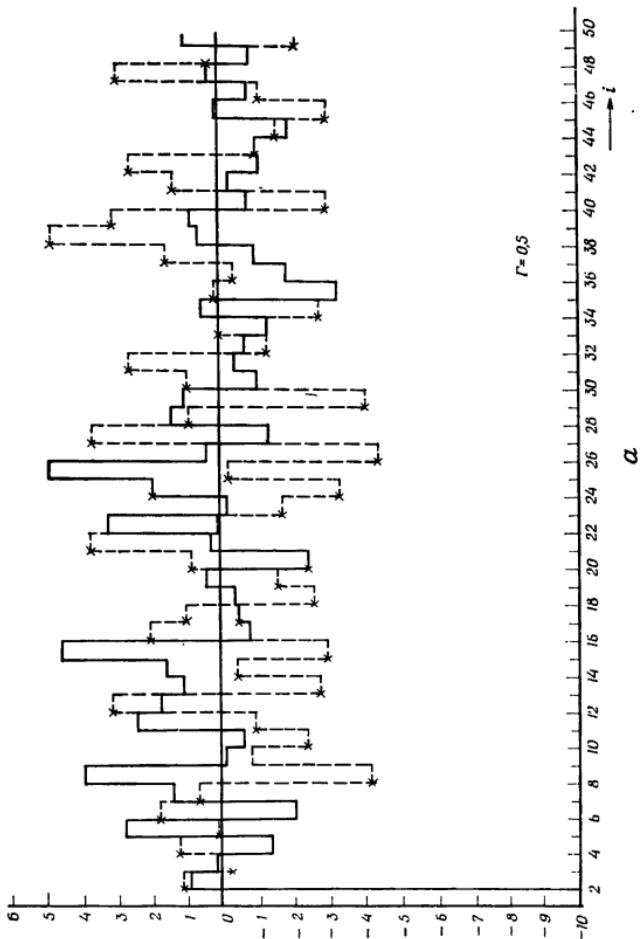
$$a(i) = a = 0,5.$$

Для нижнего уровня шума $\lambda = 0,1$ сходимость к истинному значению $x = 0$ показана на фиг. 5.31. Выбор $\Gamma = 0,5$ приводит к истинному значению за один шаг; при $\Gamma = 0,9$ имеет место движение осцилляторного типа; выбор $\Gamma = 1/i$ обеспечивает сходимость за два шага. На фиг. 5.32, а, б для верхнего уровня шума $\lambda = 10$ при $\Gamma = 0,5$ и $\Gamma = 1/i$ соответственно показаны итерации при оценке x для двух разных последовательностей (сплошная и пунк-

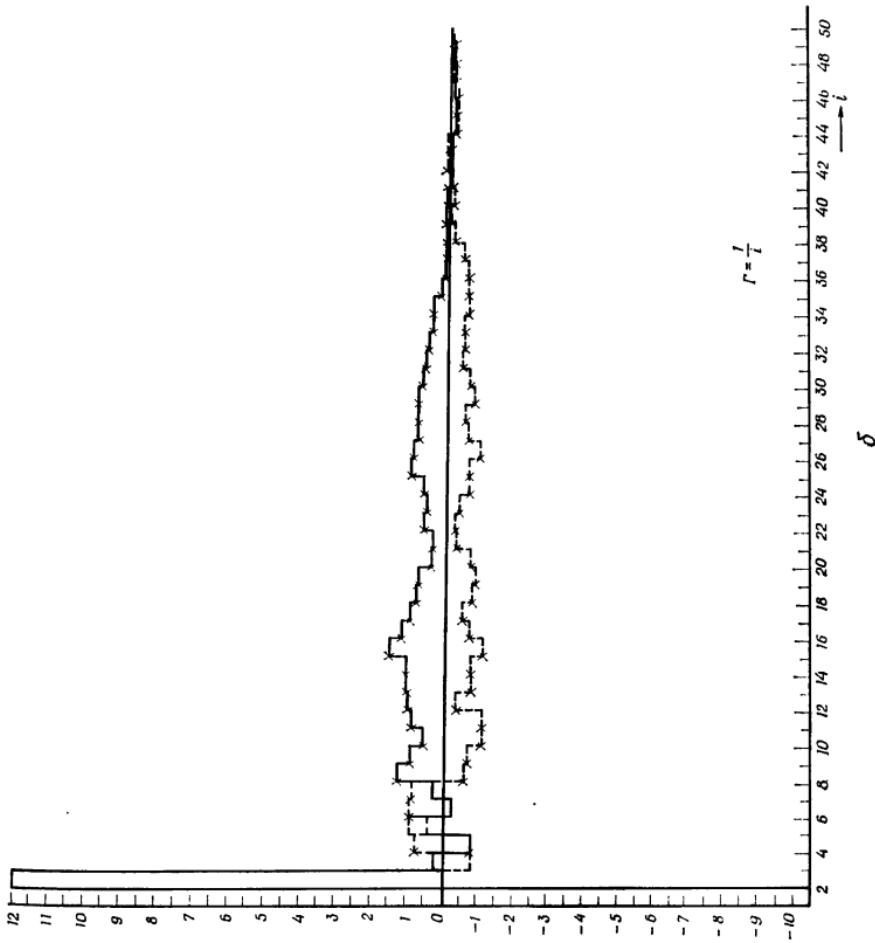
¹⁾ Выполнен Дж. П. М. де Веердом, Технологический университет, г. Эйндховен, Голландия.



Ф и г. 5.34.



Ф и г. 5.32.



тирная кривые) помех $n(i)$. При постоянном Γ сходимости к $x = 0$ нет. При тех же помехах алгоритм стохастической аппроксимации при $\Gamma = 1/i$ сходится удовлетворительно.

Некоторые из этих алгоритмов могут быть сделаны непрерывными. Детерминированный вариант:

$$\frac{dx(t)}{dt} = -\Gamma(t) \nabla_x F\{x(t)\}. \quad (5.71)$$

В стохастическом случае алгоритм имеет вид

$$\frac{dx(t)}{dt} = -\Gamma(t) \nabla_x F\{n(t) | x(t)\}. \quad (5.72)$$

Более обширную информацию о теории и приложениях стохастической аппроксимации можно почерпнуть из нескольких хороших обзоров (см., например, [1, 24]), а приложения описаны в работах [6, 10, 17, 21–23].

ЛИТЕРАТУРА

1. Albert A. E., Gardner L. A., Jr., *Stochastic Approximation and Nonlinear Regression*, MIT Press, Cambridge, Mass., 1967.
2. Bekey G. A., McGee R. B., Gradient Methods for the Optimization of Dynamic System Parameters by Hybrid Computation, in: *Computing Methods in Optimization Problems* Balakrishnan, Neustadt, eds., Academic Press, N.Y., 1964, pp. 305–327.
3. Box M. J., Davies D., Swann W. H., *Non-linear Optimization Techniques*, Oliver and Boyd, Edinburgh, 1969.
4. Davidon W. C., Variance Algorithm for Minimization, *Computer J.*, **10**, 406–410 (1968).
5. Deutsch R., *Estimation Theory*, Prentice Hall, Englewood Cliffs, N.J., 1965.
6. Elliot D. F., Swarder D. D., Applications of a Simplified Multidimensional Stochastic Approximation Algorithm, Joint Autom. Control Conf., 1969, pp. 148–154.
7. Dixon L. C. W., *Nonlinear Optimization*, English Univ. Press, London, 1972.
8. Fletcher R., Powell M. J. D., A Rapidly Convergent Descent Method for Minimization, *Computer J.*, **6**, 163–168 (1963).
9. Fletcher R., Reeves C. M., Function Minimization by Conjugate Gradients, *Computer J.*, **7**, 149–154 (1964).
10. Holmes J. K., System Identification from Noise Corrupted Measurements, *J. Optimizat. Theory and Applicat.*, **2**, 103–116 (1968).
11. Jacobson D. H., Oksman W., New Algorithms for Function Minimization, IEEE Symp. on Adaptive Processes, Decision and Control, Univ. of Texas, Austin, December 7–9, 1970, page XXI. 1.1–4.

12. Kiefer J., Wolfowitz J., Stochastic Estimation of the Maximum of a Regression Function, *Ann. Math. Statist.*, 23, 462—466 (1952).
13. Kowalik J., Osborne M. R., Methods for Unconstrained Optimization Problems, American Elsevier Publ. Co., N.Y., 1968.
14. Lindley D. V., Introduction to Probability and Statistics, from a Bayesian View-Point, Cambridge Univ. Press, London, 1965.
15. Маслов Е. П., Применение теории статистических решений к задачам оценки параметров объекта, *Автоматика и телемеханика*, № 10, 1338—1350 (1963).
16. Nahi N. E., Estimation Theory and Application, Wiley, N.Y., 1969.
17. Neal C. B., Bekey G. A. (см. литературу к гл. 7).
18. Powell M. J. D., A Survey of Numerical Methods for Unconstrained Optimization, *SIAM Review*, 12 (1970).
19. Robbins H., Monroe S., A Stochastic Approximation Method, *Ann. Math. Statist.*, 22, 400—407 (1951).
20. Rosenbrock H. H., Storey C., Computational Techniques for Chemical Engineers, Pergamon, Oxford, 1966.
21. Sakrison D. J., The Use of Stochastic Approximation to Solve the System Identification Problem, *IEEE Trans. Autom. Control*, AC-12, 563—567 (1967).
22. Saridis G. N., Stein G., Stochastic Approximation Algorithms for Linear Discrete-Time System Identification, *IEEE Trans. Autom. Control*, AC-13, 515—523 (1968).
23. Saridis G. N., Stein G., A New Algorithm for Linear System Identification (Correspondence), *IEEE Trans. Autom. Control*, AC-13, 592—594 (1968).
24. Цыпкин Я. З. (см. литературу к гл. 1).
25. White R., A Survey of Random Methods for Parameter Optimization, Rep., EE Dept., Univ. of Technology, Eindhoven, Netherlands, 1970.
26. Wilde D. J., Optimum Seeking Methods, Prentice-Hall, Englewood Cliffs, N.J., 1964; есть русский перевод: Уайлд Д. Дж., Методы поиска экстремума, изд-во «Наука», 1967.
27. Wilde D. J., Beightler C. S., Foundation of Optimization, Prentice-Hall, Englewood Cliffs, N.J., 1967.

ДОПОЛНИТЕЛЬНАЯ ЛИТЕРАТУРА

5.2. Детерминированные градиентные алгоритмы

- Bryson A. E., Denham W., A Steepest Ascent Method for Solving Optimum Programming Problems, *J. Appl. Mech.*, 29, 247—257 (1962).
 Goldstein A. A., On Steepest Descent, *SIAM J. Control*, 3, 147—151 (1965).
 Goldstein A. A., Minimizing Functionals on Normed Linear Spaces, *SIAM J. Control*, 4, 81—89 (1966).

- Goodwin G. C., Application of Curvature Methods to Parameter and State Estimation, *Proc. IEE*, 116, 1107—1110 (1969).
- Hestenes M. R. Stiefel E., Methods of Conjugate Gradients for Solving Linear Systems, *J. Research NBS*, 49, 409 (1952).
- Kelley H. J., Methods of Gradients, in: Optimization, Leitmann G., ed., Academic Press, N.Y., 1962.
- Kopp R. E., McGill R., Several Trajectory Optimization Techniques, in: Computation Methods in Optimization Problems, Balakrishnan A. V., Neustadt L. W., eds., Academic Press, N.Y., 1964, p. 65—89.
- Lasdon L. S., Mitter S. K., Warren A. D., Conjugate Gradient Method for Optimal Control Problems, *IEEE Trans. Autom. Control*, AC-12, 132—138 (1967).
- Leitmann G., ed., Optimization Techniques, Academic Press, N.Y., 1962; есть русский перевод: Лейтманн Дж., ред., Методы оптимизации с приложениями к механике космического полета, изд-во «Мир», 1965.
- Ortega J. M., Rheinholdt W. C., Iterative Solution of Nonlinear Equations in Several Variables, Academic Press, N.Y., 1970.
- Rosenbrock H. H., An Automatic Method for Finding the Greatest or Least Value of a Function, *Computer J.*, 3, 175—184 (1960).
- Saaty T. L., Bram J., Nonlinear Mathematics, McGraw-Hill, N.Y., 1964.
- Sorenson H. W., Comparison of Some Conjugate Direction Procedures for Function Minimization, *J. Franklin Inst.*, 288, 421—441 (1969).
- Воронов А. А., Игнатьев М. Б., Об отыскании экстремумов функций в автоматических системах, в книге «Дискретные и самонастраивающиеся системы», Труды II Межд. конгресса ИФАК по автоматическому управлению. Базель, изд-во «Наука», 1965, стр. 351—360.

5.3. Стохастическая аппроксимация

- Blum J. A., Multidimensional Stochastic Approximation Methods, *Ann. Math. Statist.*, 25, 737—744 (1954).
- Blum J. R., Approximation Methods which Converge with Probability One, *Ann. Math. Statist.*, 25, 382—386 (1954).
- Божанов Э. С., Применение метода стохастической аппроксимации для восстановления характеристик объектов, *Автоматика и телемеханика*, № 6, 95—103 (1967).
- Chung K. L., On a Stochastic Approximation Method, *Ann. Math. Statist.*, 25, 463—483 (1954).
- Derman C., An Application of Chun's Lemma to the Kiefer-Wolfowitz Stochastic Approximation Procedure, *Ann. Math. Statist.*, 27, 532—536 (1956).
- Derman C., Sacks J., On Dvoretzky's Stochastic Approximation Theorem, *Ann. Math. Statist.*, 30, 601—606 (1959).
- Driml M., Nedoma T., Stochastic Approximations for Continuous Random Processes, *Trans. of the Second Prague Conference on Infor-*

mation Theory, Statistical Decision Function, Random Processes, Prague, 1959.

Dvoretzky A., On Stochastic Approximation in: Proceedings of the Third Berkeley Symp. Math. Statist. Probab., Neyman J., ed., Univ. of California Press, Berkeley (Cal.), 1956, pp. 39—55.

Гладышев Е. Г., О стохастической аппроксимации. Теория вероятностей и ее применения, 10, № 2, 297—300 (1965).

Holmes J. K., System Identification with Noisy Measurements, IFAC Symp. «Identification in Autom. Control Systems», Prague, Paper 3.3, 1967.

Holmes J. K., Two Stochastic Approximation Procedures for Identifying Linear Systems (Short Paper), *IEEE Trans. Autom. Control*, AC-14, 292—295 (1969).

Imedadze V. V., Lelashvili Sh. G., Investigation of Some Algorithms of Stochastic Approximation in Identification of Multidimensional Plants, IFAC Symp. «Identification in Autom. Control Systems», Prague, Paper 3.1, 1967.

Jury E. I., Oza K. G., System Identification and the Principle of Random Contraction Mapping, *SIAM J. Control*, 6, 244—257 (1968).

Kesten H., Accelerated Stochastic Approximation, *Ann. Math. Statist.*, 29, 41—59 (1958).

Panuska V., A Stochastic Approximation Method for Identification of Linear Systems Using Adaptive Filtering, Joint Autom. Control Conf., 1968.

Popkov Yu. S., Restoration of Dynamic Characteristics of Non-linear Plants in the Presence of Noise, IFAC Symp. «Identification in Autom. Control Systems», Prague, Paper 3.15.

Roberts J. D., Extremum or Hill Climbing Regulation: a Statistical Theory Involving Lags, Disturbances and Noise, *Proc. IEE*, 112, 137—150, 1646—1650 (1965).

Sacks J., Asymptotic Distribution of Stochastic Approximation Procedures, *Ann. Math. Statist.*, 29, 373—405 (1958).

Sebestyen G. S., Decision-Making Processes in Pattern Recognition, MacMillan, N.Y., 1962; есть русский перевод: Себестиан Г. С., Процессы принятия решений при распознавании образов, изд-во «Техника», Киев, 1965.

Wolfowitz J., On the Stochastic Approximation method of Robbins and Monro, *Ann. Math. Statist.*, 23, 457—461 (1952).

Wolfowitz J., On Stochastic Approximation Methods, *Ann. Math. Stat.*, 27, 1151—1155 (1956).

Живоглядов В. П., Каипов В. Х., О применении метода стохастических аппроксимаций в проблеме идентификации, *Автоматика и телемеханика*, № 10, 54—58 (1966).

ОЦЕНИВАНИЕ ПАРАМЕТРОВ

ГЛАВА 6

ВЫБОРОЧНЫЕ СИГНАЛЫ; ЯВНЫЕ МЕТОДЫ

В гл. 2 рассматривались метод наименьших квадратов и обобщенный метод наименьших квадратов. В гл. 5 эти методы были получены из метода максимального правдоподобия при выполнении некоторых специальных условий. Это было сделано для того, чтобы показать связь между разными процедурами оценивания.

Здесь эти оценки получаются методами регрессионного анализа. Обсуждаются такие свойства оценок, как линейность, несмещенность, эффективность. Рассматривается реализация этих оценок и описываются различные виды встречающихся на практике ошибок: ошибки, вызванные помехами, ошибки усечения, ошибки из-за неправильного определения состояния, ошибки, связанные с упрощением при реализации, ошибки выборочной аппроксимации. Невязки, получающиеся в процессе оценивания, можно использовать как источник дополнительной информации о дисперсии шума и количестве параметров объекта. Такая процедура рассматривается в разд. 6.4. Попытка использования этих явных методов на классе обобщенных моделей приводит обычно к смещенным оценкам. Получено асимптотическое выражение для величины смещения. Некоторые методы решения этой задачи рассмотрены в разд. 6.5.

6.1. РЕГРЕССИОННЫЙ АНАЛИЗ

Линии и плоскости регрессии

Элементарным введением в регрессионный анализ может служить случай двух переменных [1, 8]. Пусть x и y — случайные величины с совместной плотностью

вероятности $p(x, y)$. Если эта функция непрерывна по y , то условную плотность вероятности y по x можно записать в виде

$$p(y|x) = \frac{p(x, y)}{\int_{-\infty}^{+\infty} p(x, y) dy} = \frac{p(x, y)}{p(x)}. \quad (6.1)$$

По этой функции необходимо определить функциональные зависимости, например условное математическое ожидание y при фиксированном x

$$y_x = \mathbb{E}[y|x] = \frac{\int_{-\infty}^{+\infty} y p(x, y) dy}{p(x)} = f(x). \quad (6.2)$$

Эта функция называется *линией регрессии* y по x . Можно показать ([1], стр. 96), что линия регрессии дает наилучшую оценку y в смысле минимума среднеквадратической ошибки.

Точно так же определяется любая функциональная зависимость без ограничений на f :

$$y_x = f(x).$$

Однако можно потребовать, чтобы эта функция принадлежала к какому-то классу функций, например к классу всех линейных функций или к классу полиномов заданной степени. Выбор такого класса и минимизация среднеквадратической ошибки ведут к *среднеквадратическим линиям регрессии*.

Пример. Пусть нужно найти наилучшую аппроксимацию y линейной функцией

$$w = \alpha + \beta x. \quad (6.3)$$

Такая аппроксимация называется *линейной регрессионной моделью*, а α и β — коэффициентами регрессии. Нужно минимизировать

$$\mathbb{E}[(y - w)^2] = \mathbb{E}[(y - \alpha - \beta x)^2], \quad (6.4)$$

где оператор математического ожидания $\mathbb{E}\{\dots\}$ означает интеграл $\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \dots p(x, y) dx dy$. В дальнейшем будут использоваться следующие обозначения (см. приложение Б):

$$\mu_x = \mathbb{E}[x],$$

$$\mu_y = \mathbb{E}[y],$$

$$\sigma_x^2 = \mathbb{E}[x^2] - \mathbb{E}^2[x],$$

$$\sigma_{xy}^2 = \mathbb{E}[xy] - \mathbb{E}[x]\mathbb{E}[y],$$

$$\rho = \frac{\sigma_{xy}^2}{\sigma_x \sigma_y} \text{ — коэффициент корреляции.}$$

Дифференцирование (6.4) по α и β с приравниванием полученных результатов нулю дает

$$\begin{aligned} -2\mathbb{E}[y - \alpha - \beta x] &= -2\mathbb{E}[y] + 2\alpha + 2\beta\mathbb{E}[x] = 0, \\ -2\mathbb{E}[x(y - \alpha - \beta x)] &= -2\mathbb{E}[xy] + 2\alpha\mathbb{E}[x] + 2\beta\mathbb{E}[x^2] = 0. \end{aligned} \quad (6.5)$$

Отсюда следует, что

$$\alpha = \mu_y - \mu_x \frac{\sigma_{xy}^2}{\sigma_x^2},$$

$$\beta = \frac{\sigma_{xy}^2}{\sigma_x^2}.$$

Подстановка этих выражений для α и β в (6.3) приводит к уравнению линии регрессии

$$w = \mu_y + \frac{\sigma_{xy}^2}{\sigma_x^2} (x - \mu_x) = \mu_y + \rho \frac{\sigma_y}{\sigma_x} (x - \mu_x). \quad (6.6)$$

Читатель может самостоятельно записать формулу для случая двумерного нормального распределения $p(x, y)$.

Эти идеи можно использовать для отыскания наилучших нелинейных регрессионных моделей, например полиномиальных моделей вида $w = a + \beta x + \gamma x^2 + \dots$. Может оказаться полезным набор ортонормальных полиномов (см. [1], стр. 115). Это полиномы π_i степени i ($i = 0, 1, \dots, m$), для которых выполняются условия ортогональ-

ности

$$\int_{-\infty}^{+\infty} \pi_i(x) \pi_j(x) p(x) dx = \begin{cases} 1 & \text{для } i=j, \\ 0 & \text{для } i \neq j, \end{cases} \quad (6.7)$$

где

$$p(x) = \int_{-\infty}^{+\infty} p(x, y) dy$$

представляет собой маргинальную плотность распределения вероятностей x . Если теперь положить

$$w = \sum_{i=0}^m \beta_i \pi_i(x), \quad (6.8)$$

то нужно минимизировать

$$\mathcal{E} \left[\left(y - \sum_{i=0}^m \beta_i \pi_i(x) \right)^2 \right]. \quad (6.9)$$

Дифференцируя по β_j , получаем

$$\begin{aligned} \mathcal{E} \left[\pi_j(x) \left\{ y - \sum_{i=0}^m \beta_i \pi_i(x) \right\} \right] &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} y \pi_j(x) p(x, y) dx dy - \\ &- \sum_{i=0}^m \beta_i \int_{-\infty}^{+\infty} \pi_j(x) \pi_i(x) p(x) dx. \end{aligned} \quad (6.10)$$

Приравнивая производные нулю при $\beta_j = \hat{\beta}_j$, в силу ортогональности будем иметь

$$\hat{\beta}_j = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} y \pi_j(x) p(x, y) dx dy. \quad (6.11)$$

Таким образом, не нужно решать систему линейных алгебраических уравнений. Это особенно удобно, так как по мере увеличения порядка модели система линейных уравнений часто становится почти вырожденной.

Эти идеи можно распространить на многомерный случай, когда условная плотность вероятности имеет вид

$$p(y | x_1, x_2, \dots, x_m) = p(y | x). \quad (6.12)$$

Класс функций точно так же может быть ограничен, например, классом линейных функций. В этом случае выбор

$$w = \alpha + \beta_1 x_1 + \dots + \beta_m x_m \quad (6.13)$$

приводит к среднеквадратической гиперплоскости регрессии для y как функции x_1, \dots, x_m , если минимизируется математическое ожидание

$$\mathcal{E} [(y - w)^2] = \mathcal{E} [(y - \alpha - \beta_1 x_1 - \dots - \beta_m x_m)^2]. \quad (6.14)$$

Если все случайные величины предполагаются центрированными, то коэффициент α может быть опущен.

Дифференцирование по β_1, \dots, β_m и приравнивание производных нулю при $\beta_1 = \hat{\beta}_1, \dots, \beta_m = \hat{\beta}_m$ приводят к следующей системе уравнений:

$$\begin{aligned} \hat{\beta}_1 \mathcal{E} [x_1^2] + \hat{\beta}_2 \mathcal{E} [x_1 x_2] + \dots + \hat{\beta}_m \mathcal{E} [x_1 x_m] &= \mathcal{E} [x_1 y], \\ \dots &\dots \\ \hat{\beta}_1 \mathcal{E} [x_m x_1] + \hat{\beta}_2 \mathcal{E} [x_m x_2] + \dots + \hat{\beta}_m \mathcal{E} [x_m^2] &= \mathcal{E} [x_m y]. \end{aligned} \quad (6.15)$$

Если ковариационная матрица невырождена, то эту систему линейных алгебраических уравнений можно решить относительно коэффициентов регрессии $\hat{\beta}_1, \dots, \hat{\beta}_m$. Если же w сложнее полинома (например, включает экспоненты), то соответствующую систему уравнений трудно разрешить относительно коэффициентов регрессии.

Оценивание по конечному числу наблюдений

До сих пор предполагалось, что все математические ожидания могут быть вычислены, т. е. известна совместная плотность распределения $p(x_1, \dots, x_m, y)$. Так бывает довольно редко. Обычно необходимо оценивать параметры, используя *конечное* число наблюдений, а именно выборочные значения. Таким образом, оценка должна быть функцией этих выборочных значений, которые фактически представляют собой наблюдаемые значения реализаций случайных величин. Это означает, что оценка тоже случайная величина и может быть охарактеризована плотностью вероятности. Качество оценки зависит от этой функции и, в частности, от среднего значения и дисперсии.

Излагаемые методы имеют длинную историю. Уже в 1795 г. Гаусс использовал их при исследовании движения планет. В наши дни они применяются, например, при определении параметров орбит спутников. Следует отметить, что, помимо обычных регрессионных моделей

$$y_i = \beta_1 x_{1i} + \dots + \beta_m x_{mi} + n_i,$$

где n_i — случайная величина, в литературе [12, 13] рассматриваются также *авторегрессионная модель*

$$y_i + a_1 y_{i-1} + \dots + a_k y_{i-k} = \beta_1 x_{1i} + \dots + \beta_m x_{mi} + n_i$$

и *обобщенная регрессионная модель*

$$y_i + \alpha_1 y_{i-1} + \dots + \alpha_k y_{i-k} = \beta_1 x_{1i} + \dots + \beta_m x_{mi} + n_i.$$

(См. также [9, 11, 14, 16, 17].)

Обозначения. Теперь посмотрим, как получаются оценки. Пусть наблюдается выходной сигнал объекта y . Как и в гл. 2, y состоит из отклика, на входное воздействие u , шума объекта и ошибок измерений. В момент j -го измерения выходной сигнал имеет вид

$$y(j) = y(u, b, n, j), \quad j = 1, \dots, k. \quad (6.16)$$

Вектором b обозначена зависимость выборочных значений от компонент вектора параметров объекта b_0, b_1, \dots, b_m . Определим

$$\mathbf{y}' = [y(1), y(2), \dots, y(k)]. \quad (6.17)$$

Шум зададим его математическим ожиданием и ковариационной матрицей:

$$\mathcal{E}[n] = 0, \quad (6.18)$$

$$\mathcal{E}[nn'] = \begin{bmatrix} \mathcal{E}[n(1)n(1)] & \dots & \mathcal{E}[n(k)n(1)] \\ \vdots & & \vdots \\ \mathcal{E}[n(1)n(k)] & \dots & \mathcal{E}[n(k)n(k)] \end{bmatrix} = N. \quad (6.19)$$

Задача состоит в том, чтобы определить оценку β вектора параметров b . Для этого используется теоретически предсказываемый выходной сигнал w , т. е. выход модели, который зависит от вектора коэффициентов $\beta = (\beta_0, \beta_1, \dots, \beta_m)$. Эта функциональная зависимость может быть выбрана различными способами. Простейшей являет-

ся линейная функциональная связь между w и β (линейная по параметрам модель, см. разд. 4.6)

$$w(j) = \sum_{i=0}^m \beta_i u_i(j), \quad j = 1, \dots, k,$$

где $u_i(j)$ — известные линейно независимые функции. Запишем w в виде

$$w = U\beta, \quad (6.20)$$

где

$$U = \begin{bmatrix} u_0(1) & \dots & u_m(1) \\ \vdots & & \vdots \\ u_0(k) & \dots & u_m(k) \end{bmatrix}. \quad (6.21)$$

Снова заметим, что такой выбор линейной связи между w и β не означает того, что связь между входом и выходом модели должна быть линейной. Предполагается, что матрица U полностью известна, т. е. может быть измерена без ошибок. Кроме того, предполагается, что число наблюдений k превышает число $m + 1$ неизвестных параметров.

Линейные несмешанные оценки

Класс линейных несмешанных оценок определяется следующими свойствами:

$$\beta = Qy \quad (\text{линейность}), \quad (6.22)$$

где $Q = (m + 1) \times k$ -матрица, и

$$\mathbb{E}[\beta] = b \quad (\text{несмешанность}). \quad (6.23)$$

Предполагается, что равенство (6.20) может дать полное описание объекта, т. е.

$$y = Ub + n. \quad (6.24)$$

Допустим сначала, что U и n статистически независимы. Теперь вектор ошибки e можно определить как

$$e = y - w = y - Ub. \quad (6.25)$$

В качестве функции ошибок или функции потерь можно выбрать положительно определенную форму

$$E = e' Re \stackrel{\Delta}{=} \|e\|_R^2, \quad (6.26)$$

где R — матрица весовых коэффициентов r_{ij} . Без потери общности можно предположить, что эта матрица симметрична. Функция ошибок может быть записана в виде

$$E = \|y - U\beta\|_R^2 = [y - U\beta]' R [y - U\beta] = \\ = y' Ry + [U\beta]' R [U\beta] - y' R [U\beta] - [U\beta]' Ry. \quad (6.27)$$

Так как $[U\beta]' = \beta' U'$, а R — симметричная матрица, то

$$E = y' Ry + \beta' U' RU\beta - 2\beta' U' Ry. \quad (6.28)$$

Дифференцирование этого выражения по β дает (см. приложение В)

$$\frac{\partial E}{\partial \beta} = \nabla_{\beta} E = 2U'RU\beta - 2U' Ry. \quad (6.29)$$

Последнее выражение можно записать в виде

$$-2U' R [y - U\beta] = -2U' Re.$$

При некотором $\hat{\beta}$ выражение (6.29) обращается в нуль. Отсюда находим $\hat{\beta}$, обеспечивающее экстремум функции ошибок E :

$$U' RU\hat{\beta} = U' Ry. \quad (6.30)$$

Эту систему называют *системой нормальных уравнений*. Если $U' RU$ — невырожденная матрица, то

$$\hat{\beta} = [U' RU]^{-1} U' Ry. \quad (6.31)$$

Нетрудно показать, что при $\beta = \hat{\beta}$ функция ошибок E принимает минимальное значение. Это значение $E(\hat{\beta})$ называется *остаточной ошибкой* (основанной на k наблюдениях).

Здесь уместно сделать несколько замечаний:

1) Конечно, уравнение (6.31) можно решить методами вариационного исчисления:

$$\frac{d}{d\alpha} E(\hat{\beta} + \alpha\Delta\beta) |_{\alpha=0} = 0,$$

или

$$\Delta\beta' U' R [y - U\hat{\beta}] = 0$$

при произвольном $\Delta\beta$ (принцип ортогональности).

2) Прямое доказательство того, что E достигает минимума, может быть основано на стандартном приеме ана-

$$\begin{aligned} \text{лиза членов второго порядка по } \beta. \text{ Из формулы (6.27) имеем} \\ E = y' R y + \beta' U' R U \beta - y' R U \beta - \beta' U' R y = \\ = [\beta - [U' R U]^{-1} U' R y]' U' R U [\beta - [U' R U]^{-1} U' R y] - \\ - y' R U [U' R U]^{-1} U' R U [U' R U]^{-1} U' R y + y' R y. \end{aligned}$$

Очевидно, что при β , удовлетворяющем уравнению (6.31), E достигает минимума.

3) В качестве мнемонического правила может оаться удобным использовать то, что

$$y = Ub + n$$

умножается на $U' R$:

$$U' R y = U' R U b + U' R n.$$

Так как второе слагаемое неизвестно, не измеряется и предполагается, что U и n статистически независимы, то это слагаемое отбрасывается. В результате получается оценка β истинного значения b [см. формулу (6.30)]. Естественно, такой способ вывода уравнения (6.31) не показывает, в каком смысле оценка оптимальна.

Эта оценка обладает свойством *линейности*, поскольку

$$\hat{\beta} = Qy, \text{ где } Q = [U' R U]^{-1} U' R. \quad (6.32)$$

Из формул (6.31) и (6.24) следует, что

$$\hat{\beta} = [U' R U]^{-1} U' R [Ub + n] = b + [U' R U]^{-1} U' R n.$$

Поскольку входной сигнал и шум статистически независимы,

$$\mathbb{E}[\hat{\beta}] = b + \mathbb{E}[[U' R U]^{-1} U' R] \mathbb{E}[n]. \quad (6.33)$$

А так как уже предполагалось, что $\mathbb{E}[n] = 0$, то оценка является и *несмещенной*:

$$\mathbb{E}[\hat{\beta}] = b.$$

Отсюда следует, что

$$\mathbb{E}[w] = \mathbb{E}[Ub] = Ub,$$

т. е. математическое ожидание выхода модели равно выходу объекта без аддитивного шума.

Желательно определить еще одну характеристику оценки $\hat{\beta}$ [формула (6.31)] — ее дисперсию. Интересно также оценить корреляцию между компонентами вектора $\hat{\beta}$. Все эти характеристики можно определить с помощью ковариационной матрицы

$$\text{cov} [\hat{\beta}] = \mathcal{E} [[\hat{\beta} - b] [\hat{\beta} - b]'] = \mathcal{E} [[Qy - b] [Qy - b]'] \quad (6.34)$$

По-прежнему предполагается, что справедливо соотношение (6.24) и U и n статистически независимы. Тогда, используя формулу (6.32), находим

$$\begin{aligned} \text{cov} [\hat{\beta}] &= \mathcal{E} [[QUb + Qn - b] [QUb + Qn - b]'] = \\ &= \mathcal{E} [[Qn] [Qn]'] = Q \mathcal{E} [nn'] Q' = QNQ'. \end{aligned} \quad (6.35)$$

Следовательно,

$$\text{cov} [\hat{\beta}] = [U'RU]^{-1}U'RNRU[U'RU]^{-1}.$$

Будет показано, что в нескольких практически интересных случаях это выражение можно существенно упростить. Главная диагональ матрицы состоит из оценок дисперсий оцениваемых параметров.

Оценки по методу наименьших квадратов [26, 27]

При использовании метода наименьших квадратов минимизируется выражение

$$E = e'e = ||e||^2. \quad (6.36)$$

Таким образом, в уравнении (6.26) и вытекающих из него уравнениях

$$R = I,$$

и из формул (6.30), (6.31) и (6.35) получаем

$$U'U \hat{\beta}_s = U'y, \quad (6.37)$$

или

$$\hat{\beta}_s = [U'U]^{-1}U'y \quad (6.38)$$

и

$$\text{cov} [\hat{\beta}_s] = [U'U]^{-1}U'NU[U'U]^{-1}. \quad (6.39)$$

Если U — квадратная матрица, т. е. если размер выборки равен числу оцениваемых параметров, и если матрица U

имеет обратную, то

$$[U'U]^{-1}U' = U^{-1}, \quad (6.40)$$

откуда

$$\hat{\beta}_s = U^{-1}y. \quad (6.41)$$

С инженерной точки зрения этот случай не представляет особого интереса, поскольку случайные возмущения не учитываются. Для уменьшения влияния шумов размер выборки должен быть гораздо больше числа параметров.

Если в уравнении (6.37) выразить все величины через $u_i(j)$, то нетрудно получить

$$\begin{bmatrix} \sum_{j=1}^k u_0^2(j) & \dots & \sum_{j=1}^k u_0(j) u_m(j) \\ \vdots & \ddots & \vdots \\ \sum_{j=1}^k u_m(j) u_0(j) & \dots & \sum_{j=1}^k u_m^2(j) \end{bmatrix} \begin{bmatrix} \hat{\beta}_0 \\ \vdots \\ \hat{\beta}_m \end{bmatrix} = \begin{bmatrix} \sum_{j=1}^k u_0(j) y(j) \\ \vdots \\ \sum_{j=1}^k u_m(j) y(j) \end{bmatrix}. \quad (6.42)$$

Ортогональность или ортонормальность пробных сигналов может привести к существенным упрощениям (см. гл. 4). В случае ортонормальности

$$U'U = I — единичная матрица$$

и

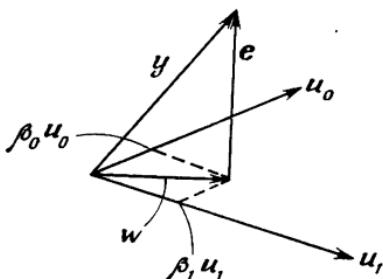
$$\hat{\beta} = U'y,$$

или

$$\hat{\beta}_i = \sum_{j=1}^k u_i(j) y(j), \quad i = 0, 1, \dots, m.$$

Можно дать простую геометрическую интерпретацию оценок метода наименьших квадратов для случая двумерного вектора параметров β (фиг. 6.1). Необходимо минимизировать длину вектора

$$e = y - w.$$



Ф и г. 6.1.

Если вектор \mathbf{e} ортогонален к \mathbf{u}_1 и \mathbf{u}_2 ,

$$\left. \begin{array}{l} \mathbf{u}'_1 \mathbf{e} = 0 \\ \mathbf{u}'_2 \mathbf{e} = 0 \end{array} \right\}, \quad \text{или} \quad \mathbf{U}' \mathbf{e} = \mathbf{0}.$$

Следовательно,

$$\mathbf{U}' (\mathbf{y} - \mathbf{w}) = \mathbf{U}' (\mathbf{y} - \mathbf{U} \hat{\beta}_s) = \mathbf{0},$$

или

$$\mathbf{U}' \mathbf{U} \hat{\beta}_s = \mathbf{U}' \mathbf{y},$$

т. е. имеем уравнение (6.37).

Марковские оценки (оценки обобщенного метода наименьших квадратов)

При использовании марковских оценок минимизируется выражение

$$E = \mathbf{e}' \mathbf{N}^{-1} \mathbf{e} = \|\mathbf{e}\|_{\mathbf{N}^{-1}}^2. \quad (6.43)$$

Следовательно,

$$\mathbf{R} = \mathbf{N}^{-1},$$

и из формул (6.30), (6.31) и (6.35) находим

$$\mathbf{U}' \mathbf{N}^{-1} \mathbf{U} \hat{\beta}_M = \mathbf{U}' \mathbf{N}^{-1} \mathbf{y}, \quad (6.44)$$

или

$$\hat{\beta}_M = [\mathbf{U}' \mathbf{N}^{-1} \mathbf{U}]^{-1} \mathbf{U}' \mathbf{N}^{-1} \mathbf{y} \quad (6.45)$$

и

$$\begin{aligned} \text{cov} [\hat{\beta}_M] &= [\mathbf{U}' \mathbf{N}^{-1} \mathbf{U}]^{-1} \mathbf{U}' \mathbf{N}^{-1} \mathbf{N} \mathbf{N}^{-1} \mathbf{U} [\mathbf{U}' \mathbf{N}^{-1} \mathbf{U}]^{-1} = \\ &= [\mathbf{U}' \mathbf{N}^{-1} \mathbf{U}]^{-1} [\mathbf{U}' \mathbf{N}^{-1} \mathbf{U}] [\mathbf{U}' \mathbf{N}^{-1} \mathbf{U}]^{-1} = [\mathbf{U}' \mathbf{N}^{-1} \mathbf{U}]^{-1}. \end{aligned} \quad (6.46)$$

Так как ковариационная матрица (6.46) входит в формулу для оценки (6.45), оценка и ковариация определяются одновременно. Диагональные элементы (6.46) представляют собой дисперсии оценок параметров, остальные элементы матрицы — коэффициенты корреляции между оценками разных параметров. Если $k \rightarrow \infty$ и $\sum_{j=1}^k u_i^2(j) \rightarrow \infty$, то дисперсии оценок стремятся к нулю и, следовательно, оценки являются *состоятельными* (см. разд. 5.1).

Используя матричное неравенство Шварца (см. приложение B), можно легко показать, что эта линейная несмещенная оценка обладает минимальной дисперсией. Неравенство Шварца имеет вид

$$[B'B] \geq [B'A][A'A]^{-1}[A'B].$$

Выбирая $A = N^{-1/2}U$ и $B' = QN^{1/2}$, получаем

$$\begin{aligned} &QN^{1/2}(N^{1/2})'Q' = QNQ' \geq \\ &\geq [QN^{1/2}N^{-1/2}U][U'(N^{-1/2})'N^{-1/2}U]^{-1}[U'(N^{-1/2})'(N^{1/2})'Q']. \end{aligned}$$

Так как $QU = I$, то

$$QNQ' \geq [U'N^{-1}U]^{-1}$$

или

$$\text{cov}[\beta] \geq \text{cov}[\hat{\beta}_M], \quad (6.47)$$

т. е. матрица

$$\text{cov}[\beta] = \text{cov}[\hat{\beta}_M]$$

не отрицательно определена. Здесь β — любая другая линейная оценка. Это означает, что оценка *эффективна* (см. разд. 5.1).

Из предыдущих разделов ясно, что если аддитивная помеха — белый шум, т. е.

$$N = E[nn'] = \sigma_n^2 I, \quad (6.48)$$

то

$$\beta_s = \hat{\beta}_M \quad (6.49)$$

и

$$\text{cov}[\hat{\beta}_M] = \sigma_n^2 [U'U]^{-1}. \quad (6.50)$$

Как отмечалось в гл. 5, для гауссовских шумов оценки метода наименьших квадратов и марковские оценки сов-

падают с оценками максимального правдоподобия, все они обладают минимальной дисперсией в классе несмешанных оценок.

Класс линейных несмешанных оценок порождается минимизацией положительно определенной формы

$$E = \| \mathbf{e} \|_{\mathbf{R}}^2.$$

Таким образом, выбор $\mathbf{R} = \mathbf{I}$ и $\mathbf{R} = \mathbf{N}^{-1}$ приводит к оценкам метода наименьших квадратов и марковским оценкам. Можно было бы выбрать \mathbf{R} иначе, например, как матрицу весовых коэффициентов, учитывающую возможную нестационарную аддитивную помеху. В этом случае необходимо минимизировать следующую функцию ошибок:

$$E = \sum_{j=1}^k \frac{1}{\sigma_n^2(j)} \{y(j) - \mathbf{u}'(j) \boldsymbol{\beta}\}^2, \quad (6.51)$$

где $\sigma_n^2(j)$ — дисперсия шума в момент времени j .

6.2. РЕАЛИЗАЦИЯ РАЗОМКНУТЫХ СХЕМ ОЦЕНИВАНИЯ

В разд. 6.1 приводятся следующие формулы:

Метод наименьших квадратов:

$$\hat{\boldsymbol{\beta}}_s = [\mathbf{U}' \mathbf{U}]^{-1} \mathbf{U}' \mathbf{y}, \quad (6.52)$$

$$\text{cov}[\hat{\boldsymbol{\beta}}_s] = [\mathbf{U}' \mathbf{U}]^{-1} \mathbf{U}' \mathbf{N} \mathbf{U} [\mathbf{U}' \mathbf{U}]^{-1}. \quad (6.53)$$

Марковские оценки:

$$\hat{\boldsymbol{\beta}}_M = [\mathbf{U}' \mathbf{N}^{-1} \mathbf{U}]^{-1} \mathbf{U}' \mathbf{N}^{-1} \mathbf{y}, \quad (6.54)$$

$$\text{cov}[\hat{\boldsymbol{\beta}}_M] = [\mathbf{U}' \mathbf{N}^{-1} \mathbf{U}]^{-1}. \quad (6.55)$$

В этом разделе будут рассмотрены некоторые вопросы реализации этих формул.

Начнем с простейшего случая *оценок по методу наименьших квадратов*, связанных с операцией $\min_{\boldsymbol{\beta}} E$, где

$$E = \mathbf{e}' \mathbf{e} = \sum_{j=1}^k e^2(j), \text{ которая приводит к уравнению (6.52).}$$

Это уравнение реализуется с помощью известных корреляционных методов.

Пример. Пусть требуется определить только один параметр β_i . В этом случае уравнение (6.52) имеет вид

$$\hat{\beta}_i = (\mathbf{u}'_i \mathbf{u}_i)^{-1} \mathbf{u}'_i \mathbf{y} = \frac{\sum_{j=1}^k u_i(j) y(j)}{\sum_{j=1}^k u_i^2(j)}. \quad (6.56)$$

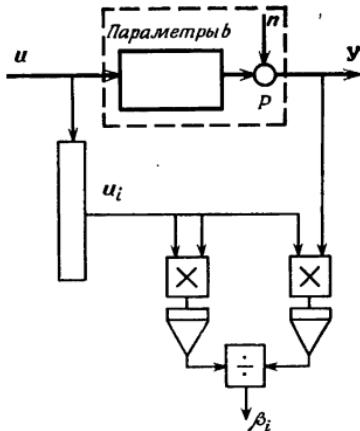
Числитель и знаменатель представляют собой средние по времени значения произведений, которые являются аппроксимациями взаимной корреляционной и автокорреляционной функций (фиг. 6.2). В этом разделе обсуждаются возможные ошибки.

В случае многих параметров уравнение (6.52) можно переписать в виде

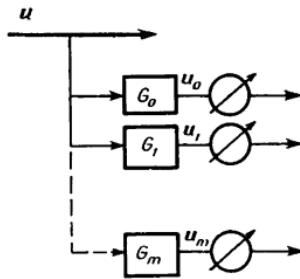
$$\hat{\beta} = \begin{bmatrix} \hat{\beta}_0 \\ \vdots \\ \hat{\beta}_m \end{bmatrix} = \begin{bmatrix} \mathbf{u}'_0 \mathbf{u}_0 \dots \mathbf{u}'_0 \mathbf{u}_m \\ \vdots \\ \mathbf{u}'_m \mathbf{u}_0 \dots \mathbf{u}'_m \mathbf{u}_m \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{u}'_0 \mathbf{y} \\ \vdots \\ \mathbf{u}'_m \mathbf{y} \end{bmatrix}. \quad (6.57)$$

Уместно сделать несколько замечаний:

1) Матрица $\mathbf{U}'\mathbf{U}$ симметрична относительно главной диагонали, поскольку $\mathbf{u}'_i \mathbf{u}_j = \mathbf{u}'_j \mathbf{u}_i$. Так как элементы этой матрицы действительные, то она самосопряженная и мо-



Ф и г. 6.2.



Ф и г. 6.3.

жет быть приведена к диагональной форме преобразованием подобия.

2) Одной из практических задач является вычисление обратной матрицы (см. [30, 39]). С помощью обратных матриц нетрудно непосредственно получить оценки параметров. Можно определить величину невязок и функцию ошибок. При известной ковариационной матрице шума N можно вычислить ковариационную матрицу оценок.

3) Если $u_i u_j = \delta_{ij}$ (символ Кронекера), то $U'U = I$ — единичная матрица. Это условие ортонормальности можно аппроксимировать соответствующим выбором входного сигнала u и передаточных функций G_i (фиг. 6.3). Чаще всего используется сочетание u (белый шум) и G_i (чистое запаздывание). Довольно часто предлагается использовать систему ортонормальных фильтров G_i , например фильтров Лагерра (см. гл. 4).

Синтез фильтров, обеспечивающих ортогонализацию компонент u_j при фиксированном входном сигнале U , рассмотрен в работах [10, 33].

4) Вместо выборочных значений можно использовать соответствующие реализации непрерывных сигналов (см. гл. 8).

5) Аналоговые сигналы можно заменить квантованными сигналами с конечным числом заданных уровней. Предельное число уровней определяется двоичным сигналом (см. гл. 8).

6) Если воспользоваться пробным сигналом, то при выборе u или $u(t)$ можно сосредоточить внимание на приятии сигналу желаемых свойств. Широко распространены пробные сигналы в виде набора синусоид или псевдослучайной двоичной последовательности. Каждому из этих типов сигналов можно придать полезные свойства в отношении к ортогональности и способу генерирования (см. гл. 10).

7) При получении оценок не делалось предположений о линейности объекта P . Преобразователи G_i на фиг. 6.3 могут быть нелинейными. Единственным требованием к модели w объекта $y = Ub$ является линейность ее выхода по определяемым параметрам ($w = U\beta$). Таким образом, к некоторым классам нелинейных систем можно применить корреляционные методы (см., например, в гл. 4).

винеровскую характеристику нелинейных систем, ряды Вольтерра, системы, приводимые к линейным).

8) Если

$$\mathbf{u}'_i = [u_i(1), u_i(2), \dots, u_i(k)], i = 0, 1, \dots, m,$$

то

$$\mathbf{u}'_i \mathbf{u}_j = \sum_{l=1}^k u_i(l) u_j(l) = k \tilde{\Psi}_{u_i u_j}(0, k), \quad (6.58)$$

где $\tilde{\Psi}(0, k)$ — аппроксимация одной из точек корреляционной функции, основанная на k наблюдениях. Если \mathbf{u}_i и \mathbf{u}_j — временные сдвиги стационарного входа \mathbf{u} , то

$$\mathbf{u}'_i \mathbf{u}_j = \sum_{l=1}^k u(l-i) u(l-j) = k \tilde{\Psi}_{uu}(j-i, k).$$

Для эргодического (стационарного) временного ряда

$$\tilde{\Psi}_{uu}(j, k) \rightarrow \Psi_{uu}(j) \text{ при } k \rightarrow \infty.$$

Следующая ступень усложнения возникает при реализации *марковских оценок*, которые определяются формулой (6.54), вытекающей из условия

$$\min_{\beta} E, \text{ где } E = \mathbf{e}' \mathbf{N}^{-1} \mathbf{e}. \quad (6.59)$$

Матрицу \mathbf{N}^{-1} можно разложить на произведение нижней треугольной матрицы \mathbf{D} и транспонированной к ней матрицы \mathbf{D}' :

$$\mathbf{N}^{-1} = \mathbf{D}' \mathbf{D}.$$

Используя это соотношение, формулы (6.59) и (6.54) можно переписать в виде

$$E = [\mathbf{D}\mathbf{e}]' \mathbf{D}\mathbf{e},$$

$$\hat{\beta} = [[\mathbf{D}\mathbf{U}]' \mathbf{D}\mathbf{U}]^{-1} [\mathbf{D}\mathbf{U}]' \mathbf{D}\mathbf{y}.$$

Матрица \mathbf{D} соответствует винеровскому фильтру, на вход которого поступает сигнал (последовательность \mathbf{n} , а на выходе получается белый шум. На фиг. 6.4 показано, каким образом подобные фильтры могут быть использованы при реализации корреляционных методов анализа систем. Как уже упоминалось, при этом обеспечивается минимальная дисперсия оценки (в классе несмещенных

оценок). Но, несмотря на это и относительную простоту реализации, метод, по-видимому, не нашел большого применения. По поводу матрицы $[(DU)'DU]^{-1}$ можно сделать те же замечания, что и по поводу матрицы $[U'U]^{-1}$, используемой в методе наименьших квадратов.

На практике априорная информация о возмущениях редко бывает достаточной для того, чтобы можно было пользоваться марковскими оценками. Таким образом, может оказаться желательным оценивать статистику возмущений по модели авторегрессии невязок. На основе этой информации можно построить фильтры D , показанные на фиг. 6.4.

Опишем один из таких методов. Пусть

$$\hat{\beta} = [U'U]^{-1}U'y,$$

$$e = y - U\hat{\beta}.$$

Теперь невязки e рассматриваются как аппроксимация возмущений δ . Допустим, что эти возмущения образуются в результате пропускания белого шума через фильтр m -го порядка. Отсюда следует, что минимизируется следующее выражение:

$$e - E\delta,$$

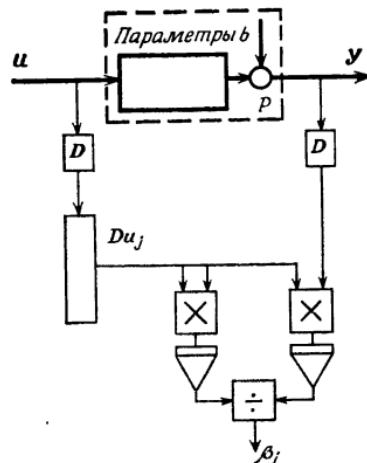
или

$$\begin{bmatrix} e(m+1) \\ \vdots \\ e(k) \end{bmatrix} - \begin{bmatrix} e(m) & \dots & e(1) \\ \vdots & & \vdots \\ e(k-1) & \dots & e(k-m) \end{bmatrix} \begin{bmatrix} \delta_1 \\ \vdots \\ \delta_m \end{bmatrix}.$$

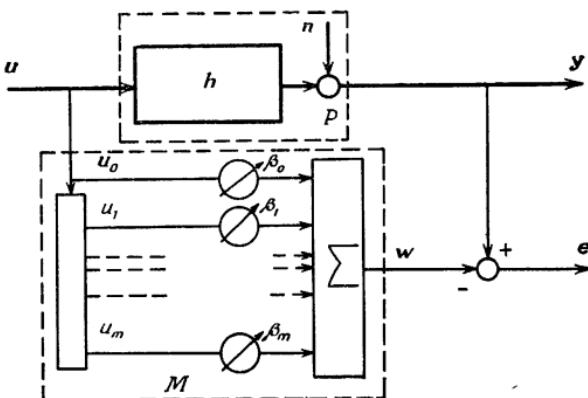
Тогда из системы нормальных уравнений находим

$$\delta = [E'E]^{-1}E'e.$$

С помощью этих оценок можно (приближенно) настроить модель \mathcal{D} фильтра D . Тогда определяется новая оценка $\hat{\beta}$, а по ней новая оценка $\hat{\delta}$ и т. д. (см. разд. 7.2).



Фиг. 6.4.



Ф и г. 6.5.

Прямой способ получения некоррелированных наблюдений, когда входной шум не является белым, состоит в выборе достаточно продолжительных пауз между последовательными измерениями. Однако в этом случае время наблюдения может стать чрезмерно большим.

Другие подходы к решению этой задачи можно найти в [12, 28] и гл. 7. Работа Острэма (см. гл. 11) посвящена одновременному оцениванию параметров объекта и характеристик шума.

До сих пор на модель не накладывалось никаких ограничений, кроме требования линейности ее выхода по определяемым параметрам. В большинстве случаев u_i выбирается как сдвинутый во времени входной сигнал u . При использовании ЦВМ такой выбор очень удобен. В этом случае параметр β_i представляет собой оценку одной точки $h(i)$ весовой функции объекта. На фиг. 6.5 изображена такая модель, использующая линию задержки. Здесь предполагается, что выходной сигнал объекта $y(j)$ может быть представлен в виде суммы

$$y(j) = \sum_{i=0}^j u(i) h(j-i) + n(j), \quad j = 1, \dots, k,$$

или

$$\mathbf{y} = \mathbf{U}\mathbf{h} + \mathbf{n} = [\mathbf{u}_0, \mathbf{u}_1, \dots, \mathbf{u}_m] \mathbf{h} + \mathbf{n}.$$

В этом случае уравнение (6.57) можно переписать в виде

$$\hat{\beta} = \begin{bmatrix} \hat{\beta}_0 \\ \vdots \\ \hat{\beta}_m \end{bmatrix} = \begin{bmatrix} \tilde{\Psi}_{uu}(0, k) & \dots & \tilde{\Psi}_{uu}(m, k) \\ \vdots & & \vdots \\ \tilde{\Psi}_{uu}(m, k) & \dots & \tilde{\Psi}_{uu}(0, k) \end{bmatrix}^{-1} \begin{bmatrix} \tilde{\Psi}_{uy}(0, k) \\ \vdots \\ \tilde{\Psi}_{uy}(m, k) \end{bmatrix}, \quad (6.60)$$

где каждая $\tilde{\Psi}$ есть сумма k произведений. Если $k \rightarrow \infty$ и сигнал u эргодический, то $\tilde{\Psi}_{uu}(j, k)$ стремится к автокорреляционной функции $\Psi_{uu}(j)$. Для входной последовательности типа белого шума

$$\Psi_{uu}(j) = \begin{cases} \Psi_{uu}(0) = \sigma_u^2 & \text{при } j = 0, \\ 0 & \text{при } j \neq 0 \end{cases}$$

параметры оцениваются по формуле

$$\hat{\beta}_j = \frac{\Psi_{uy}(j)}{\Psi_{uu}(0)}.$$

Пример. Этот простой пример приведен для пояснения изложенных идей.

Объект описывается уравнением

$$y(k) = a + bu(k) + n(k), \quad \text{или} \quad \mathbf{y} = a\mathbf{1} + b\mathbf{u} + \mathbf{n}^1).$$

Выбрана модель вида

$$w(k) = \alpha + \beta u(k), \quad \text{или} \quad \mathbf{w} = \alpha\mathbf{1} + \beta\mathbf{u}.$$

Ошибка запишется как

$$e(k) = y(k) - w(k), \quad \text{или} \quad \mathbf{e} = \mathbf{y} - \mathbf{w}.$$

В качестве минимизируемой функции ошибок выбрана

$$\sum_{j=1}^k e^2(j), \quad \text{или} \quad \mathbf{e}'\mathbf{e}.$$

Истинные значения оцениваемых параметров $[a; b] = [4; 2]$; оценки обозначаются через $[\alpha; \beta]$. При соответствующей трактовке можно воспользоваться уравнением (6.52). Однако для лучшего усвоения материала снова

¹⁾ Символом $\mathbf{1}$ обозначен вектор, состоящий из единиц.

рассмотрим всю процедуру минимизации:

$$\frac{\partial}{\partial \alpha} (\mathbf{e}' \mathbf{e}) = -\mathbf{1}' [\mathbf{y} - \hat{\alpha} \mathbf{1} - \hat{\beta} \mathbf{u}] = 0,$$

$$\frac{\partial}{\partial \beta} (\mathbf{e}' \mathbf{e}) = -\mathbf{u}' [\mathbf{y} - \hat{\alpha} \mathbf{1} - \hat{\beta} \mathbf{u}] = 0,$$

или

$$\begin{bmatrix} \mathbf{1}' & \mathbf{y} \\ \mathbf{u}' & \mathbf{y} \end{bmatrix} = \begin{bmatrix} \mathbf{1}' \mathbf{1} & \mathbf{1}' \mathbf{u} \\ \mathbf{u}' \mathbf{1} & \mathbf{u}' \mathbf{u} \end{bmatrix} \begin{bmatrix} \hat{\alpha} \\ \hat{\beta} \end{bmatrix},$$

или

$$\begin{bmatrix} \sum_{j=1}^k y(j) \\ \sum_{j=1}^k u(j) y(j) \end{bmatrix} = \begin{bmatrix} k & \sum_{j=1}^k u(j) \\ \sum_{j=1}^k u(j) & \sum_{j=1}^k u^2(j) \end{bmatrix} \begin{bmatrix} \hat{\alpha} \\ \hat{\beta} \end{bmatrix}.$$

Это согласуется с полученными ранее формулами. Читателю предлагается написать простую программу для генерирования на ЦВМ последовательностей $u(k)$ и $n(k)$ (например, как последовательностей случайных чисел) с целью моделирования объекта и оценивания параметров (оценки можно изобразить графически при $k = 3, 10, 30, 100$) или проделать эти вычисления вручную, используя таблицы псевдослучайных чисел для генерирования $u(k)$ и $n(k)$.

В процессе машинного моделирования полезно повторить эксперимент несколько раз, каждый раз задавая новые начальные условия для генератора случайных чисел. Изображая графически результаты эксперимента, можно получить наглядное представление о скорости сходимости. Можно также представить наблюдения $\{u(k), y(k)\}$ на плоскости (u, y) , изобразить результаты моделирования объекта при отсутствии помех для случая $y = 4 + 2u$ и провести линию регрессии по оценкам $\hat{\alpha}$ и $\hat{\beta}$.

6.3. ТОЧНОСТЬ, НЕКОТОРЫЕ ИСТОЧНИКИ ОШИБОК [23, 26, 29, 37]

Для практики прежде всего важно выяснить источники ошибок. Различают следующие случаи:

- 1) ошибки, вызванные помехами;

- 2) ошибки усечения;
- 3) ошибки из-за неправильного определения состояния;
- 4) ошибки, связанные с упрощением при реализации;
- 5) ошибки выборочной аппроксимации.

Ошибки, вызванные помехами

Эти ошибки возникают из-за случайного шума n , который добавляется к выходному сигналу объекта. Допустим, что модель адекватна объекту (фиг. 6.5):

$$y = Ub + n,$$

$$w = U\beta,$$

где b и β — векторы одной и той же размерности. Для этого случая уже получено уравнение метода наименьших квадратов (6.53):

$$\text{cov} [\hat{\beta}] = \mathcal{E} [(\hat{\beta} - b)(\hat{\beta} - b)'] = [U'U]^{-1} U'NU [U'U]^{-1}, \quad (6.61)$$

где

$$N = \mathcal{E}[nn'].$$

Отметим, что уравнение (6.61) дает полное описание ошибок из-за помех, основанное на *априорной* информации о свойствах шума и *апостериорной* информации (измерениях) о входном сигнале объекта u . По априорной и апостериорной информации экспериментатор может получить ясное представление о точности оценок и корреляции между ошибками разных компонент вектора оценок $\hat{\beta}$. Если потребуется, можно построить машинную реализацию уравнения (6.61)¹⁾. Для пояснения основных идей рассмотрим несколько примеров при различных упрощающих предположениях.

Пример. Если помеха представляет собой белый шум и $\mathcal{E}[n] = 0$, то

$$N = \Psi_{nn}(0) I = \sigma_n^2 I$$

¹⁾ Конечно, если бы априори была известна матрица N , можно было бы использовать марковские оценки.

и

$$\text{cov} [\hat{\beta}] = \sigma_n^2 [U'U]^{-1}. \quad (6.62)$$

Если к тому же шум гауссовский, то оценка метода наименьших квадратов совпадает с оценкой максимального правдоподобия (см. гл. 5). Таким образом, оценка $\hat{\beta}$ обладает всеми свойствами оценок максимального правдоподобия и, в частности, асимптотически нормальна со средним b и ковариацией, определяемой формулой (6.62).

Пример. Допустим теперь, что на вход объекта поступает стационарный белый шум с известными математическим ожиданием и ковариацией

$$\begin{aligned}\mathbb{E}[u] &= 0, \\ \mathbb{E}[U'U] &= k\Psi_{uu}(0)I = k\sigma_u^2 I.\end{aligned}$$

Здесь оператор математического ожидания относится к стохастической переменной u . Снова рассматривается выборка длины k . Это приводит к следующей оценке:

$$\text{cov} [\hat{\beta}] \approx \frac{\Psi_{nn}(0)}{k\Psi_{uu}(0)} I = \frac{\sigma_n^2}{k\sigma_u^2} I. \quad (6.63)$$

Для каждого параметра β_j стандартное отклонение σ_{β_j} (разброс) составляет

$$\sigma_{\beta_j} \approx \frac{\sigma_n}{\sqrt{k}\sigma_u}.$$

При $k \rightarrow \infty$ плотность вероятности стремится к δ -функции т. е. разброс стремится к нулю. Следует, правда, заметить, что из-за наличия операции извлечения квадратного корня сходимость достаточно медленная. Дисперсии σ_u^2 и σ_n^2 можно рассматривать как меру относительных «мощностей». Выражение для σ_{β} не зависит от числа оцениваемых параметров.

Другой способ определения ошибки, связанной с наличием аддитивных помех, состоит в следующем. Как видно из формулы (6.38) и равенства $y = Ub + n$,

$$\hat{\beta} = [U'U]^{-1}U'y = b + [U'U]^{-1}U'n$$

или

$$\Delta \beta = \hat{\beta} - b = [U'U]^{-1} U' n = P U' n, \text{ где } P = k [p_{ij}].$$

Следовательно,

$$\Delta \beta_j = p_{0j} \tilde{\Psi}_{un}(0, k) + \dots + p_{jj} \tilde{\Psi}_{un}(j, k) + \dots \\ \dots + p_{mj} \tilde{\Psi}_{un}(m, k).$$

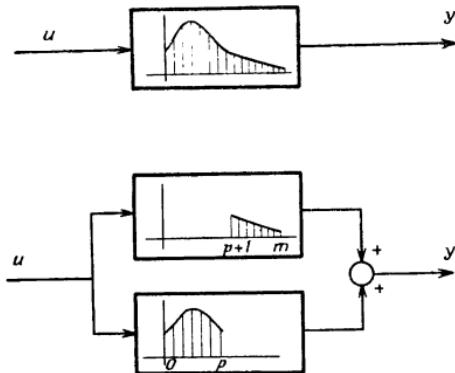
Если u — сигнал типа белого шума, то $p_{jj} = \tilde{\Psi}_{uu}^{-1}(0, k) \approx \approx \sigma_u^{-2}$ и $p_{ij} \approx 0$ для $i \neq j$. Таким образом, доминирующим членом является слагаемое, содержащее p_{jj} и

$$\Delta \beta_j \approx \frac{\tilde{\Psi}_{un}(j, k)}{\tilde{\Psi}_{uu}(0, k)} \quad (\text{для большого } k). \quad (6.64)$$

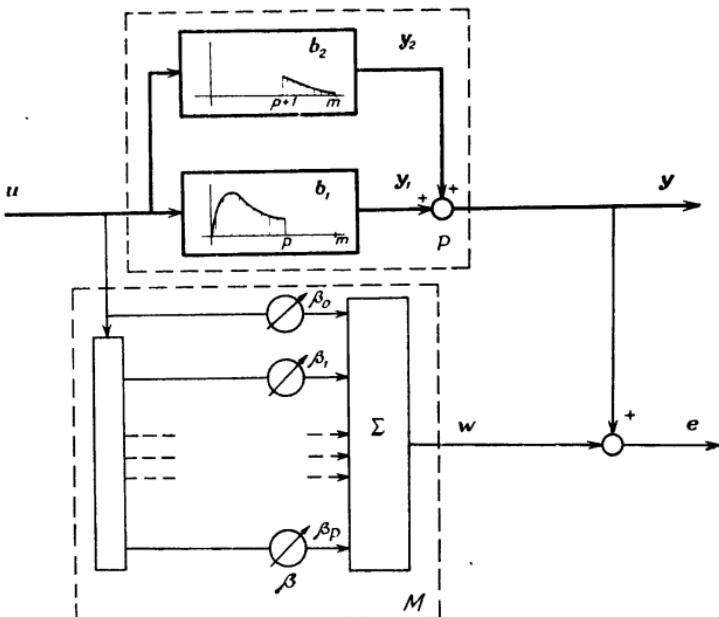
Конечно, при использовании метода наименьших квадратов в случае, когда аддитивная помеха не является белым шумом, появляется дополнительная составляющая ошибки.

Ошибки усечения

Если в концептуальной модели учитывается только часть параметров объекта, то для хорошего понимания задачи желательно рассматривать объект состоящим из двух частей (фиг. 6.6). Часть параметров (точки) весовой функции b_0, \dots, b_p может быть оценена, остальные параметры b_{p+1}, \dots, b_m не оцениваются.



Ф и г. 6.6.



Ф и г. 6.7.

Такая ситуация имеет место, если:

- 1) исходное предположение о длительности импульсной реакции системы не соответствует истине;
- 2) интерес представляет только начальная фаза (или любой заданный отрезок) отклика;
- 3) весовая функция затухает на бесконечности, т. е. объект включает чистое интегрирование.

На фиг. 6.7 параметры модели $\beta' = [\beta_0, \dots, \beta_p]$ соответствуют параметрам объекта $b'_1 = [b_0, \dots, b_p]$. Выход объекта y состоит из двух частей:

$$y = y_1(b_1) + y_2(b_2).$$

С точки зрения оценивания сигнал y_2 , соответствующий усеченной части отклика объекта, действует как аддитивное возмущение. Очевидно, что y_1 и y_2 могут быть коррелированы, так как зависят от одного входного сигнала u .

Для исследования эффекта усечения положим

$$y = Ub = [U_1; U_2] \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}, \quad \text{где } b_1 = \begin{bmatrix} b_0 \\ \vdots \\ b_p \end{bmatrix}, \quad b_2 = \begin{bmatrix} b_{p+1} \\ \vdots \\ b_m \end{bmatrix},$$

или

$$\begin{aligned} \mathbf{y} &= \mathbf{U}_1 \mathbf{b}_1 + \mathbf{U}_2 \mathbf{b}_2, \\ \mathbf{U}'_1 \mathbf{y} &= \mathbf{U}'_1 \mathbf{U}_1 \mathbf{b}_1 + \mathbf{U}'_1 \mathbf{U}_2 \mathbf{b}_2. \end{aligned} \quad (6.65)$$

Оценка параметров «усеченной» части объекта имеет вид

$$\hat{\boldsymbol{\beta}} = [\mathbf{U}'_1 \mathbf{U}_1]^{-1} \mathbf{U}'_1 \mathbf{y} \text{ или } \mathbf{U}'_1 \mathbf{y} = \mathbf{U}'_1 \mathbf{U}_1 \hat{\boldsymbol{\beta}}.$$

Сопоставляя это выражение с формулой (6.65), находим

$$\Delta \boldsymbol{\beta} = \hat{\boldsymbol{\beta}} - \mathbf{b}_1 = [\mathbf{U}'_1 \mathbf{U}_1]^{-1} \mathbf{U}'_1 \mathbf{U}_2 \mathbf{b}_2. \quad (6.66)$$

Если входной сигнал \mathbf{u} типа *белого шума*, то

$$\mathcal{E}[\mathbf{U}'_1 \mathbf{U}_2] = \mathbf{0} \text{ и } \mathcal{E}[\Delta \boldsymbol{\beta}] = \mathbf{0}.$$

Для других входных сигналов это математическое ожидание может не равняться нулю и оценки будут смешенными.

Для ковариации $\Delta \boldsymbol{\beta}$ имеем

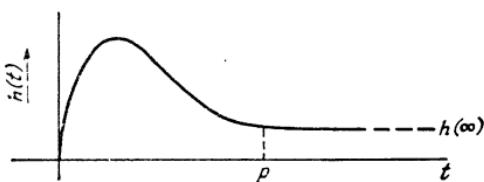
$$\text{cov}[\Delta \boldsymbol{\beta}] = \mathcal{E}[[\mathbf{U}'_1 \mathbf{U}_1]^{-1} \mathbf{U}'_1 \mathbf{U}_2 \mathbf{b}_2 \mathbf{b}'_2 \mathbf{U}'_2 \mathbf{U}_1 [\mathbf{U}'_1 \mathbf{U}_1]^{-1}]. \quad (6.67)$$

Из этих уравнений и физического смысла ясно, что $\mathcal{E}[\Delta \boldsymbol{\beta}]$ и $\text{cov}[\Delta \boldsymbol{\beta}]$ не зависят от уровня («мощности») входного сигнала.

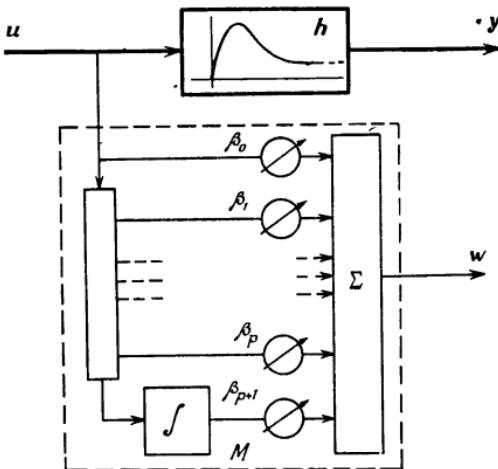
Особый интерес представляет случай, когда $h(\infty) \neq 0$ (фиг. 6.8). Параметры \mathbf{b}_1 нужно определять по формуле

$$\hat{\boldsymbol{\beta}} = [\mathbf{U}'_1 \mathbf{U}_1]^{-1} \mathbf{U}'_1 \mathbf{y}_1 = [\mathbf{U}'_1 \mathbf{U}_1]^{-1} \mathbf{U}'_1 (\mathbf{y} - \mathbf{y}_2). \quad (6.68)$$

Однако величина \mathbf{y}_2 заранее не известна, и вместо нее приходится использовать \mathbf{y} , что приводит к ошибкам. Независимо от объема выборки, используемой при оценивании, неопределенность полученных оценок не стремится к нулю. Оказывается, что для каждого параметра



Фиг. 6.8.



Ф и г. 6.9.

β_j с $j > p$ справедлива формула [29]

$$\sigma_{\beta_j} = \frac{h(\infty)}{\sqrt{2}}. \quad (6.69)$$

Если использовать y , то последний из оцениваемых параметров β_p на фиг. 6.8 может дать какую-то информацию о $h(\infty)$, хотя и искаженную ошибками усечения. Предполагалось, что $\beta_j = \beta_p$ для $j \geq p$. Таким образом, величина y_2 известна с некоторой неопределенностью. Это означает, что, используя y , частично можно компенсировать неопределенность y_2 и получить более точную оценку β_j и т. д.

Другой выход из положения указан на фиг. 6.9, где «хвост» весовой функции соответствует интегратор (сумматор). Большие постоянные времени можно учесть, включая в (концептуальную) модель вместо интегратора низкочастотный фильтр. Отметим отличие нашей схемы от схемы, рассмотренной в [25], в которой интегратор стоит перед линией задержки.

Ошибки усечения при оценке параметров сильно коррелированы и почти равны по величине и по знаку. Особенно это справедливо при больших k . Отсюда следует, что для больших k форму графика весовой функции мож-

но достаточно точно определить, однако невозможно ответить на вопрос, насколько каждая оценка отличается от соответствующего значения истинной весовой функции [29].

Ошибки из-за неправильного определения состояния

В начале наблюдений за входным и выходным сигналами объекта вектор состояния не обязательно равен нулю. Это приводит к изменению первых выборочных значений u , которые не связаны с входным сигналом u . Снова используя модель с линией задержки, находим

$$\begin{aligned} y &= U_p b + n \quad (\text{уравнение объекта}), \\ w &= U_M \beta \quad (\text{уравнение модели}), \end{aligned}$$

где

$$\left[\begin{array}{c} u(1) \\ u(2) \\ u(3) \\ \vdots \\ \vdots \\ u(m+1) \\ u(m+2) \\ \vdots \\ u(k) \\ \vdots \\ u(k-m) \\ \vdots \\ u(k) \end{array} \right] = \left[\begin{array}{c} U_p \\ U_M \end{array} \right] \left[\begin{array}{c} b \\ \beta \end{array} \right] + \left[\begin{array}{c} n \\ \vdots \end{array} \right] \quad (6.70)$$

Если при $t = 0$ объект находится в нулевом состоянии, то все элементы U_p в треугольнике A равны нулю; в противном случае они определяются значениями $y(1), \dots, y(m)$. То же самое справедливо и для конца интервала наблюдений; если входной сигнал управляем и $u(j) = 0$ при $j > k$, то треугольник B будет заполнен нулями. Если сигнал неуправляем, то элементы в B будут определяться недоступными оценкой последними выборочными значениями y . Это приводит к увеличению дисперсии оценок. Но, конечно, эта ошибка имеет значение только тогда, когда оценивание проводится по *малым* выборкам.

Этот эффект можно уменьшить. Оценивание по методу наименьших квадратов основано на минимизации

$$E = \mathbf{e}'\mathbf{e} = \sum e^2(j).$$

В процессе наблюдения мы относимся к первым и последним выборочным значениям «менее серьезно», чем к промежуточным. Это можно учесть, минимизируя следующее выражение:

$$E = \mathbf{e}'\mathbf{R}\mathbf{e}, \quad \text{где } \mathbf{R} = \begin{bmatrix} r_1 & 0 & \dots & 0 \\ 0 & r_2 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & r_k \end{bmatrix}.$$

Здесь $r_1 < r_2 < r_3 < \dots < r_i = r_j > \dots > r_{k-2} > r_{k-1} > r_k$. В результате получаем оценку вида

$$\hat{\beta} = [\mathbf{U}'\mathbf{R}\mathbf{U}]^{-1}\mathbf{U}'\mathbf{R}\mathbf{y}. \quad (6.71)$$

Другой подход к решению этой задачи состоит в одновременном оценивании параметров и состояний, начиная с $t = 0$.

Ошибки, связанные с упрощением при реализации

Как было показано, оценка

$$\hat{\beta} = [\mathbf{U}'\mathbf{U}]^{-1}\mathbf{U}'\mathbf{y} \quad (6.72)$$

является несмещенной. Операции обращения или перемножения матриц в вычислительном отношении трудоемки. Поэтому интересно было бы понять, нельзя ли

упростить формулу (6.72), приведя ее к виду

$$\beta = \frac{\mathbf{U}' \mathbf{y}}{k \tilde{\Psi}_{uu}(0, k)} \quad \text{или} \quad \beta_j = \frac{\tilde{\Psi}_{uy}(j, k)}{\tilde{\Psi}_{uu}(0, k)}, \quad (6.73)$$

что, как известно из разд. 6.11, справедливо, когда \mathbf{u} — белый шум. Из формулы (6.72)

$$\begin{bmatrix} \tilde{\Psi}_{uu}(0, k) & \dots & \dots & \tilde{\Psi}_{uu}(m, k) \\ \vdots & & & \vdots \\ \dots & \tilde{\Psi}_{uu}(1, k) & \tilde{\Psi}_{uu}(0, k) & \tilde{\Psi}_{uu}(1, k) & \dots \\ \vdots & & & \vdots & \vdots \\ \tilde{\Psi}_{uu}(m, k) & \dots & & \dots & \tilde{\Psi}_{uu}(0, k) \end{bmatrix} \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_j \\ \vdots \\ \hat{\beta}_m \end{bmatrix} = \begin{bmatrix} \tilde{\Psi}_{uy}(0, k) \\ \vdots \\ \tilde{\Psi}_{uy}(j, k) \\ \vdots \\ \tilde{\Psi}_{uy}(m, k) \end{bmatrix}$$

следует, что

$$\hat{\beta}_j = \frac{\tilde{\Psi}_{uy}(j, k)}{\tilde{\Psi}_{uu}(0, k)} - \frac{1}{\tilde{\Psi}_{uu}(0, k)} \{(\hat{\beta}_{j+1} + \hat{\beta}_{j-1}) \tilde{\Psi}_{uu}(1, k) + \dots\}. \quad (6.74)$$

Поскольку $\hat{\beta}_j$ вычисляется только по первому члену разложения без обращения матрицы, то отсюда следует, что второй член определяет смещение оценки, связанное с упрощением модели [37].

Интересно также сравнить дисперсии оценок, полученных с обращением и без обращения матрицы. Уже известно, что в случае аддитивного белого шума оценивание с обращением матрицы дает

$$\text{cov} [\hat{\beta}] = [\mathbf{U}' \mathbf{U}]^{-1} \mathbf{U}' \mathbf{N} \mathbf{U} [\mathbf{U}' \mathbf{U}]^{-1} = \sigma_n^2 [\mathbf{U}' \mathbf{U}]^{-1}. \quad (6.75)$$

Оценка, полученная без обращения матрицы, имеет вид

$$\beta = \frac{\mathbf{U}' \mathbf{y}}{k \tilde{\Psi}_{uu}(0, k)} = \frac{\mathbf{U}' \mathbf{U} \mathbf{b} + \mathbf{U}' \mathbf{n}}{k \tilde{\Psi}_{uu}(0, k)}.$$

Вообще говоря, эта оценка смещена, так как в большинстве случаев

$$\mathcal{E} [\beta] = \frac{\mathbf{U}' \mathbf{U}}{k \tilde{\Psi}_{uu}(0, k)} \mathbf{b} \neq \mathbf{I} \mathbf{b}. \quad (6.76)$$

Следовательно, ковариация имеет вид

$$\text{cov} [\beta] = \mathcal{E} \left[\left\| \beta - \frac{\mathbf{U}' \mathbf{U} \mathbf{b}}{k \tilde{\Psi}_{uu}(0, k)} \right\|^2 \right] = \frac{\mathbf{U}' \mathbf{U} \mathbf{U}' \mathbf{U}}{k^2 \tilde{\Psi}_{uu}^2(0, k)} \quad (6.77)$$

и для белого шума

$$\text{cov} [\beta] = \frac{\sigma_n^2 \mathbf{U}' \mathbf{U}}{k^2 \tilde{\Psi}_{uu}^2(0, k)}. \quad (6.78)$$

Если \mathbf{u} — белый шум, то для больших k ковариация определяется тем же выражением, что и в случае оценки, полученной с обращением матрицы [см. формулу (6.63)]. Таким образом, результаты сопоставления явных методов оценивания с обращением и без обращения матриц могут быть представлены следующей таблицей:

С обращением матрицы	Без обращения матрицы
Оценка линейная, $\hat{\beta} = \mathbf{Q} \mathbf{y}$	
Оценка несмещенная, $\mathcal{E} [\hat{\beta}] = \mathbf{b}$ $\lim_{k \rightarrow \infty} \text{cov} [\hat{\beta}] = 0 \mathbf{I}$	$\left. \begin{array}{l} \text{Оценка смещенная, если только} \\ \text{входной сигнал не является} \\ \text{белым шумом, а } k \text{ достаточно} \\ \text{большим [см. формулу (6.76)]} \end{array} \right\}$
Подходит для любых постоянно возбуждающих (возбуждающих все гармоники) сигналов	
Подходит для любых аддитивных шумов, некоррелированных с \mathbf{u}	
Требует обращения матрицы, перемножения матриц и двух умножений матрицы на вектор	$\left. \begin{array}{l} \text{Требует только одного умножения} \\ \text{матрицы на вектор, если } \mathbf{u} \text{ — белый шум, а для } \tilde{\Psi}_{uu}(0, k) \text{ заранее известно приближенное} \\ \text{выражение} \end{array} \right\}$
Если \mathbf{u} и \mathbf{n} — белые шумы, то $\text{cov} [\hat{\beta}] \approx \frac{\sigma_n^2}{k \sigma_u^2} \mathbf{I}$	

Ошибки выборочной аппроксимации

Если рассматриваемый объект непрерывен, то нужно выяснить, к чему приведет аппроксимация непрерывных сигналов их выборочными значениями при разном шаге интервала измерений (см. гл. 7).

6.4. НЕВЯЗКИ, ШУМЫ И ПОРЯДОК МОДЕЛИ

В предыдущих разделах выведены формулы для оценок параметров и их ковариационных матриц. Однако на практике в процессе исследования могут появиться вопросы, например, о свойствах помехи (возмущений) или о требуемом числе параметров модели.

Эту информацию можно получить по невязкам

$$\mathbf{e} = \mathbf{y} - \mathbf{U} \hat{\boldsymbol{\beta}} = \mathbf{U} (\mathbf{b} - \hat{\boldsymbol{\beta}}) + \mathbf{n}. \quad (6.79)$$

Невязки содержат информацию как о параметрах и состояниях, так и об аддитивной помехе.

Если заранее не ясно, сколько параметров являются существенными, то важно, чтобы процедура оценивания была *рекуррентной по числу параметров*. Это означает, что, начиная с оценки вектора параметров $\mathbf{b}'_m = [b_0, \dots, b_{m-1}]$, не нужно снова повторять все вычисления, оценивая вектор параметров $\mathbf{b}'_p = [b_0, \dots, b_{p-1}]$, $p > m$. Используя обозначение $\mathbf{b}'_p = [\mathbf{b}'_m \mid \mathbf{b}'_{p-m}]$, находим

$$\mathbf{y} = [\mathbf{U}_m \mid \mathbf{U}_{p-m}] \begin{bmatrix} \mathbf{b}'_m \\ \vdots \\ \mathbf{b}'_{p-m} \end{bmatrix} + \mathbf{n}, \quad (6.80)$$

что ведет к следующей системе нормальных уравнений:

$$\left[\begin{array}{c|c} \mathbf{U}'_m \mathbf{U}_m & \cdots \\ \hline \cdots & \cdots \end{array} \right] \left[\begin{array}{c} \hat{\boldsymbol{\beta}}_{m; m} \\ \vdots \\ 0 \end{array} \right] = \left[\begin{array}{c} \mathbf{U}'_m \mathbf{y} \\ \vdots \end{array} \right]$$

и

$$\begin{aligned} \left[\begin{array}{c|c} \mathbf{U}'_m \mathbf{U}_m & \mathbf{U}'_m \mathbf{U}_{m-p} \\ \hline \mathbf{U}'_{m-p} \mathbf{U}_m & \mathbf{U}'_{m-p} \mathbf{U}_{m-p} \end{array} \right] \left[\begin{array}{c} \hat{\boldsymbol{\beta}}_{m; p} \\ \hat{\boldsymbol{\beta}}_{p-m; p} \end{array} \right] &= \left[\begin{array}{c} \mathbf{U}'_m \mathbf{y} \\ \mathbf{U}'_{m-p} \mathbf{y} \end{array} \right] = \\ &= \left[\begin{array}{c} \mathbf{U}'_m \mathbf{U}_m \hat{\boldsymbol{\beta}}_{m; m} \\ \hline \mathbf{U}'_{p-m} \mathbf{y} \end{array} \right]. \end{aligned}$$

Здесь $\hat{\beta}_{m; m}$ и $\hat{\beta}_{m; p}$ являются соответственно старой и новой оценками вектора b_0, \dots, b_{m-1} , т. е. до и после увеличения числа параметров с m до p при одном и том же числе наблюдений. Эти уравнения можно переписать в виде

$$\begin{aligned} \mathbf{U}'_m \mathbf{U}_m [\hat{\beta}_{m; p} - \hat{\beta}_{m; m}] + \mathbf{U}'_{m-p} \mathbf{U}_{m-p} \hat{\beta}_{m-p; p} &= 0, \\ \mathbf{U}'_{m-p} \mathbf{U}_m [\hat{\beta}_{m; p} - \hat{\beta}_{m; m}] + \mathbf{U}'_{m-p} \mathbf{U}_{m-p} \hat{\beta}_{m-p; p} &= \\ &= \mathbf{U}'_{m-p} [\mathbf{y} - \mathbf{U}_m \hat{\beta}_{m; m}]. \end{aligned}$$

Решая их относительно $\hat{\beta}_{m; p}$ и $\hat{\beta}_{m-p; p}$, находим

$$\begin{aligned} \hat{\beta}_{m; p} &= \hat{\beta}_{m; m} - \mathbf{B} \mathbf{C} \mathbf{U}'_{m-p} [\mathbf{y} - \mathbf{U}_m \hat{\beta}_{m; m}], \\ \hat{\beta}_{m-p; p} &= \mathbf{C} \mathbf{U}'_{m-p} [\mathbf{y} - \mathbf{U}_m \hat{\beta}_{m; m}], \end{aligned} \quad (6.81)$$

где

$$\begin{aligned} \mathbf{A} &= [\mathbf{U}'_m \mathbf{U}_m]^{-1}, \\ \mathbf{B} &= \mathbf{A} \mathbf{U}'_{m-p} \mathbf{U}_{m-p}, \\ \mathbf{C} &= [\mathbf{U}'_{m-p} \mathbf{U}_{m-p} - \mathbf{U}'_{m-p} \mathbf{U}_m \mathbf{A} \mathbf{U}'_m \mathbf{U}_{m-p}]^{-1}. \end{aligned}$$

Член $[\mathbf{y} - \mathbf{U}_m \hat{\beta}_{m; m}]$ соответствует *невязкам* исходной модели с параметрами $\beta_0, \dots, \beta_{m-1}$. Отметим, что $\mathbf{A} = [\mathbf{U}'_m \mathbf{U}_m]^{-1}$ становится известной на предыдущем шаге и что обращаемая матрица имеет порядок только $m - p$. Таким образом, $\hat{\beta}_{m; p}$ и $\hat{\beta}_{m-p; p}$ довольно легко могут быть вычислены. На очередной итерации нужно обратить расширенную матрицу $\mathbf{U}' \mathbf{U}$. Можно показать, что эта матрица имеет следующую структуру:

$$\left[\begin{array}{c|c} \mathbf{U}'_m \mathbf{U}_m & \mathbf{U}'_{m-p} \mathbf{U}_{m-p} \\ \hline \mathbf{U}'_{m-p} \mathbf{U}_m & \mathbf{U}'_{m-p} \mathbf{U}_{m-p} \end{array} \right] = \left[\begin{array}{cc} \mathbf{A} - \mathbf{B} \mathbf{C} \mathbf{B}' & -\mathbf{B} \mathbf{C} \\ -\mathbf{C}' \mathbf{B}' & \mathbf{C} \end{array} \right]; \quad (6.82)$$

она состоит из матриц, которые уже определены.

Таким образом можно вычислять оценки для моделей с различным числом параметров. Однако до сих пор не получен ответ на вопрос, сколько параметров в действительности необходимо (каким должен быть порядок модели). Вообще говоря, увеличение числа параметров модели ведет к уменьшению функции ошибок, или функции потерь, $E = \mathbf{e}' \mathbf{e}$. Действительно E может обратиться в нуль, когда число параметров совпадает с числом k выборочных

значений входного и выходного сигналов. В этом случае помеха полностью входит в коэффициенты модели, что, конечно, нежелательно. Следовательно, нужно воспользоваться статистическим критерием для проверки того, является ли значащим уменьшение E при увеличении числа параметров.

Идея рассматривать выбор порядка модели как задачу теории принятия решений обсуждалась в работах [3, 42] и является стандартным приемом регрессионного анализа. Приводимое описание такой проверки порядка модели заимствовано у Острема [6]. Допустим, что объект описывается уравнением

$$\mathbf{y} = \mathbf{U}_m \mathbf{b}_m + \mathbf{n}, \quad (6.83)$$

где \mathbf{y} — состоит из k выборочных значений, \mathbf{b}_m — вектор параметров (например, b_0, \dots, b_{m-1}), а \mathbf{n} — вектор независимых гауссовских величин с параметрами $(0, \sigma)$. Оценка $\hat{\beta}_p$, получаемая по модели порядка p , имеет вид

$$\hat{\beta}_p = [\mathbf{U}'_p \mathbf{U}_p]^{-1} \mathbf{U}'_p \mathbf{y} = \begin{bmatrix} \mathbf{b}_m \\ \vdots \\ \mathbf{0}_{p-m} \end{bmatrix} + [\mathbf{U}'_p \mathbf{U}_p]^{-1} \mathbf{U}'_p \mathbf{n}. \quad (6.84)$$

Невязки выражаются формулой

$$\begin{aligned} \mathbf{e}_p &= \mathbf{y} - \mathbf{U}_p \hat{\beta}_p = \mathbf{U}_m \mathbf{b}_m + \mathbf{n} - [\mathbf{U}_m ; \mathbf{U}_{p-m}] \begin{bmatrix} \mathbf{b}_m \\ \vdots \\ \mathbf{0}_{p-m} \end{bmatrix} - \\ &- \mathbf{U}_p [\mathbf{U}'_p \mathbf{U}_p]^{-1} \mathbf{U}'_p \mathbf{n} = \mathbf{n} - \mathbf{U}_p [\mathbf{U}'_p \mathbf{U}_p]^{-1} \mathbf{U}'_p \mathbf{n} = [\mathbf{I} - \mathbf{A}] \mathbf{n}. \end{aligned} \quad (6.85)$$

Здесь \mathbf{A} — симметричная идемпотентная матрица, т. е.

$$\mathbf{A} = \mathbf{A}^2 = \mathbf{A}^3 = \dots;$$

то же можно сказать о матрице $[\mathbf{I} - \mathbf{A}]$ (см. приложение B). Следовательно,

$$E_p = \mathbf{e}'_p \mathbf{e}_p = \mathbf{n}' [\mathbf{I}_k - \mathbf{U}_p [\mathbf{U}'_p \mathbf{U}_p]^{-1} \mathbf{U}'_p] \mathbf{n}. \quad (6.86)$$

Теперь, если рассмотреть две модели с числом параметров p и q ($p < q$), то легко получить следующее тривиальное тождество:

$$E_q + (E_p - E_q) + (\mathbf{n}' \mathbf{I}_k \mathbf{n} - E_p) = \mathbf{n}' \mathbf{I}_k \mathbf{n}.$$

В левой части этого тождества содержатся неотрицательно определенные квадратичные формы по e_q, e_p и п. Учитывая, что соответствующие матрицы:

- симметричны, неотрицательно определены;
- ортогональным преобразованием могут быть приведены к диагональному виду с диагональю из нулей и единиц;
- имеют ранг, равный следу (инварианту) (см. приложение В), замечаем, что

$$\text{ранг } [I_k - U_q [U'_q U_q]^{-1} U'_q] = k - q,$$

$$\text{ранг } [U_q [U'_q U_q]^{-1} U'_q - U_p [U'_p U_p]^{-1} U'_p] = q - p,$$

$$\text{ранг } [U_p [U'_p U_p]^{-1} U'_p] = p.$$

По теореме Кохрана (см. [22], т. 2, стр. 360), $E_q, E_p - E_q$ и $n' I_k n - E_p$ имеют распределение χ^2 и взаимно независимы. Таким образом, если $q > p \geq m$, то E_q и $E_p - E_q$ — независимые случайные величины, имеющие распределение χ^2 с $k - q$ и $q - p$ степенями свободы соответственно. Теперь проверка гипотезы о том, существенно ли уменьшение функции ошибок при увеличении числа параметров от p до q , может быть сведена к вычислению

$$t = \frac{(E_p - E_q)/(q - p)}{\sqrt{E_q/(k - q)}}. \quad (6.87)$$

Эта величина имеет распределение F_{k-q}^{q-p} , для которого составлены многочисленные таблицы (см., например, [14]). Из этих таблиц для 5%-ного уровня значимости имеем

$$F_{120}^1 = 3,92, \quad F_{\infty}^1 = 3,84,$$

$$F_{120}^2 = 3,07, \quad F_{\infty}^2 = 3,00,$$

$$F_{120}^3 = 2,68, \quad F_{\infty}^3 = 2,60,$$

т. е.

$$P\{F_{120}^1 < 3,92\} = 0,95.$$

Таким образом, для рассматриваемого случая при $t > F$ необходима модель более высокого порядка. В технических задачах обычно интересен случай $k \gg q$, и, следовательно, важно помнить, что при $k \rightarrow \infty$ статистика F_{k-q}^{q-p}

асимптотически нормальна, т. е.

$$F_{k-q}^{q-p} \Rightarrow \frac{\chi_{q-p}^2}{q-p}.$$

Таким образом, можно пользоваться и таблицами χ^2 .

При идеальном подборе модели для объекта имело бы место равенство $e = n$. Априорные предположения о распределении помехи можно проверить по невязкам. Если предполагается, что n — белый шум, т. е. выборочные значения статистически независимы, то это можно проверить, применив корреляционные методы анализа к e ; если n считается гауссовским, то для проверки можно использовать критерии χ^2 .

Дисперсию шума σ_n^2 , которая заранее неизвестна, можно оценить по формуле

$$\sigma_n^2 = \frac{1}{k-m} e'e, \quad (6.88)$$

где k — размер выборки, а m — число определяемых параметров. Покажем это (см. [6, 34]). Из формулы (6.86) следует, что

$$\begin{aligned} \mathcal{E}[e'_m e_m] &= \mathcal{E}[n' [I_k - U_m [U'_m U_m]^{-1} U'_m] n] = \\ &= \mathcal{E}[\text{Tr}\{n' [I_k - U_m [U'_m U_m]^{-1} U'_m] n\}] = \\ &= \mathcal{E}[\text{Tr}\{I_k - U_m [U'_m U_m]^{-1} U'_m\} nn'] = \\ &= \sigma_n^2 [\text{Tr} I_k - \text{Tr}\{U_m [U'_m U_m]^{-1} U'_m\}] = \sigma_n^2 (k-m). \end{aligned}$$

Таким образом, несмещенная оценка σ_n^2 определяется формулой (6.88), т. е. получается возведением невязок в квадрат и делением на число степеней свободы.

Отметим, что в литературе по математической статистике также можно найти формулы для доверительных интервалов параметров β_i и критерии проверки гипотез $\beta_i = 0$.

Другие вопросы. В работе [18] описывается метод, в котором выборочные измерения заменяются результатами интегрирования по очень коротким промежуткам времени. Это удобно, когда на выходе объекта действует аддитивная помеха, частота которой значительно превышает частоту выборочных измерений.

Если можно управлять входом u , то возникает интересный вопрос, как выбрать U так, чтобы, насколько это возможно, уменьшить ожидаемые ошибки. Эта задача кратко рассмотрена в работе [24].

6.5. РАСПРОСТРАНЕНИЕ НА ОБОБЩЕННЫЕ МОДЕЛИ И ОБЪЕКТЫ С ОБРАТНОЙ СВЯЗЬЮ

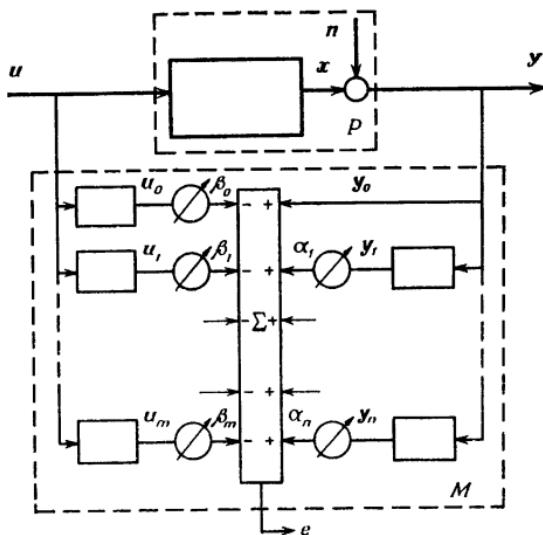
Из-за полезных свойств и удобства реализации оценок метода наименьших квадратов или марковских оценок имеет смысл попробовать распространить их на *обобщенную модель*, которая введена в гл. 2 и изображена на фиг. 6.10. В этом случае

$$e(k) = y(k) + \sum_{j=1}^n \alpha_j y_j(k) - \sum_{j=0}^m \beta_j u_j(k),$$

или

$$\mathbf{e} = \mathbf{y} + \mathbf{Y}\boldsymbol{\alpha} - \mathbf{U}\boldsymbol{\beta} = \mathbf{y} - [\mathbf{U} \mid \mathbf{Y}] \begin{bmatrix} \boldsymbol{\beta} \\ -\boldsymbol{\alpha} \end{bmatrix} = \mathbf{y} - \mathbf{\Omega}\boldsymbol{\theta}. \quad (6.89)$$

Поскольку число параметров в векторах $\boldsymbol{\alpha}$ и $\boldsymbol{\beta}$ не имеет значения, в дальнейшем предполагается, что $m = n$. Все формулы можно легко приспособить для случая $m \neq n$.



Фиг. 6.10.

Для большей наглядности параметр β_0 отброшен. Так же как и раньше при минимизации $E = \mathbf{e}'\mathbf{e}$, оценка запишется в виде

$$\hat{\theta} = [\Omega' \Omega]^{-1} \Omega' \mathbf{y}. \quad (6.90)$$

Ниже будет показано, что у этой оценки имеется серьезный недостаток — смещение, вызванное аддитивным шумом. Поэтому имеет смысл исследовать этот случай несколько подробнее, так как он охватывает одно из наиболее важных представлений линейного стационарного дискретного по времени одномерного объекта. Положим $u_i(k) = u(k-i)$ и $y_j(k) = y(k-j)$. Тогда объект будет описываться уравнениями

$$\begin{aligned} x(k) + a_1x(k-1) + \dots + a_mx(k-m) = \\ = b_1u(k-1) + \dots + b_mu(k-m), \\ y(k) = x(k) + n(k), \end{aligned} \quad (6.91)$$

а уравнение модели примет вид

$$\begin{aligned} y(k) + a_1y(k-1) + \dots + a_my(k-m) - \\ - \beta_1u(k-1) - \dots - \beta_mu(k-m) = e(k). \end{aligned} \quad (6.92)$$

Введя оператор сдвига

$$zy(k) = y(k+1),$$

можно, как и раньше, записать уравнение объекта в виде

$$\begin{aligned} A(z)x(k) &= B(z)u(k), \\ A(z) &= z^m + a_1z^{m-1} + \dots + a_m, \\ B(z) &= b_1z^{m-1} + \dots + b_m, \end{aligned}$$

или

$$\begin{aligned} A^*(z^{-1})x(k) &= B^*(z^{-1})u(k), \\ A^*(z^{-1}) &= 1 + a_1z^{-1} + \dots + a_mz^{-m}, \\ B^*(z^{-1}) &= b_1z^{-1} + \dots + b_mz^{-m}. \end{aligned}$$

Для модели получаются аналогичные выражения с операторами \mathcal{A} , \mathcal{B} , \mathcal{A}^* и \mathcal{B}^* , зависящими от параметров a и β . Используя эти обозначения, обобщенную ошибку

можно записывать разными способами:

1) обобщенная модель:

$$e(k) = \mathcal{A}^*(z^{-1}) y(k) - \mathcal{B}^*(z^{-1}) u(k), \quad (6.93)$$

которая схематично изображена на фиг. 6.11;

2) прогнозирующая модель:

$$e(k) = y(k) - w(k), \quad (6.94)$$

где

$$w(k) = \{1 - \mathcal{A}^*(z^{-1})\} y(k) + \mathcal{B}^*(z^{-1}) u(k). \quad (6.95)$$

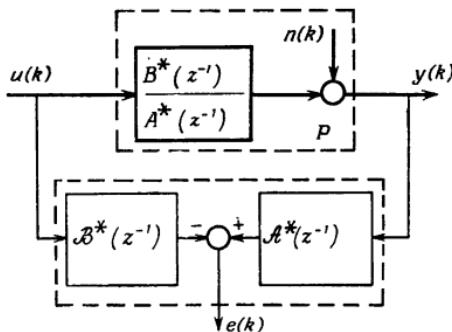
Эта модель тоже линейная по параметрам. Ее выход можно интерпретировать как основанный на предыдущих наблюдениях прогноз $y(k)$, записанный в виде $w(k) = \hat{y}(k | k-1)$ (фиг. 6.12). Таким образом, обобщенную ошибку можно рассматривать как разность между фактическим выходным сигналом и его прогнозом в соответствии с формулой (6.95). Отметим, что такое описание подходит и для случая, когда нет входных сигналов, например, при параметрическом анализе временных рядов для построения модели авторегрессии (см. [40, 43]). Приложения к ЭЭГ-анализу можно найти в [15];

3) обычная модель (и фильтрация сигнала ошибки) (фиг. 6.13):

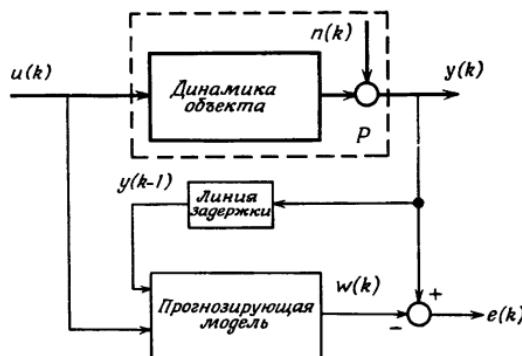
$$e(k) = \mathcal{A}^*(z^{-1}) [y(k) - y_M(k)], \quad (6.96)$$

где

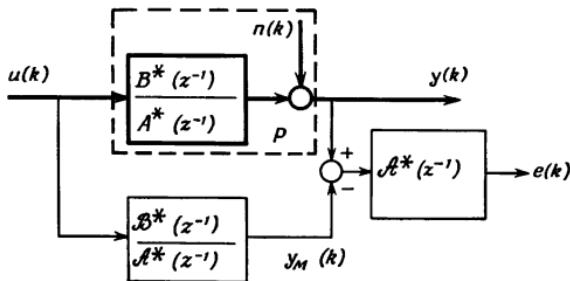
$$y_M(k) = \frac{\mathcal{B}^*(z^{-1})}{\mathcal{A}^*(z^{-1})} u(k). \quad (6.97)$$



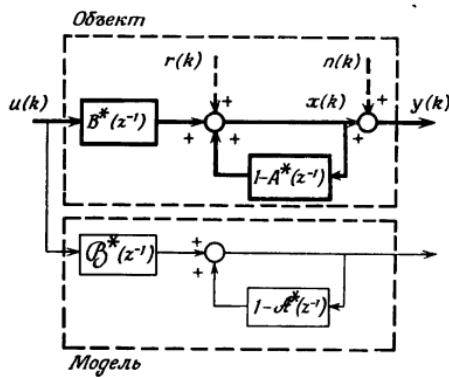
Ф и г. 6.11.



Ф и г. 6.12.



Ф и г. 6.13.



Ф и г. 6.14.

Заметим, что, как показано на фиг. 6.14, объект (модель) можно представить в виде двух частей — оператора прямого канала B^* (\mathcal{B}^*) и оператора канала обратной связи $1 - A^*$ ($1 - \mathcal{A}^*$). Отметим также, что помеха может действовать по-разному, например как $r(k)$ или как $n(k)$. Тогда объект описывается следующими уравнениями:

	$r(k) \equiv 0, n(k) \equiv 0$	$r(k) \equiv 0, n(k) \not\equiv 0$
В переменных u и x	$A^*x(k) = B^*u(k) + r(k),$ $y(k) = x(k)$	$A^*x(k) = B^*u(k),$ $y(k) = x(k) + n(k)$
В переменных u и y	$A^*y(k) = B^*u(k) + r(k)$	$A^*y(k) = B^*u(k) +$ $+ A^*n(k)$

В зависимости от типа модели и приложений можно выбрать тот или иной способ описания.

В новой интерпретации оценка

$$\hat{\theta} = [\Omega' \Omega]^{-1} \Omega' y \quad (6.98)$$

состоит из нескольких частей:

$$\hat{\theta}' = [\hat{\beta}' | -\hat{\alpha}'] = [\hat{\beta}_1, \dots, \hat{\beta}_m | -\hat{\alpha}_1, \dots, -\hat{\alpha}_m].$$

В формуле (6.98)

$$\begin{aligned} y &= \begin{bmatrix} y(1) \\ y(2) \\ \vdots \\ y(k) \end{bmatrix}, \\ \Omega &= \left[\begin{array}{cccc|cccc} u(0) & u(-1) & \dots & u(1-m) & y(0) & y(-1) & \dots & y(1-m) \\ u(1) & u(0) & \dots & u(2-m) & y(1) & y(0) & \dots & y(2-m) \\ \vdots & & & \vdots & \vdots & & & \vdots \\ u(k-1) & u(k-2) & \dots & u(k-m) & y(k-1) & y(k-2) & \dots & y(k-m) \end{array} \right]. \end{aligned}$$

Эти формулы можно запрограммировать на ЦВМ. На основе наблюдений u и y по этой программе находятся числовые значения $\hat{\theta}$, а возможно, и ковариационной матрицы. Как уже отмечалось, рассматриваемый метод

является явным или «однократным», если обработка всех наблюдений производится лишь один раз. Этот метод тесно связан с хорошо известными корреляционными методами определения точек весовой функции. Для коэффициентов корреляции получаются следующие (приближенные) выражения:

$$\begin{aligned}\widetilde{\Psi}_{uu}(j) &= \frac{1}{k-j} \sum_{i=1}^{k-j} u(i) u^*(i+j), \\ \widetilde{\Psi}_{yy}(j) &= \frac{1}{k-j} \sum_{i=1}^{k-j} y(i) y(i+j), \\ \widetilde{\Psi}_{yu}(j) &= \frac{1}{k-j} \sum_{i=1}^{k-j} y(i) u(i+j), \\ \widetilde{\Psi}_{uy}(j) &= \frac{1}{k-j} \sum_{i=1}^{k-j} u(i) y(i+j).\end{aligned}\tag{6.99}$$

Сравнивая с методом наименьших квадратов, находим, что матрицы $\Omega' \Omega$ и $\Omega' y$ состоят в основном из коэффициентов взаимной корреляции и автокорреляции, если пренебречь началом и концом серии наблюдений.

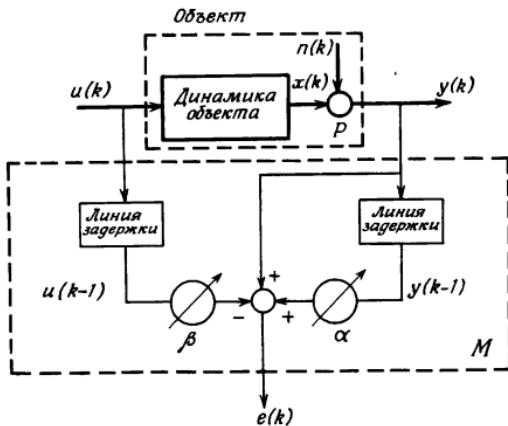
Отметим, что методы этого раздела можно без изменения применить для идентификации нелинейных объектов, являющихся линейными по параметрам, например

$$y(k) + ay(k-1) = b_1 u(k-1) + b_2 u^2(k-1).$$

Смещение оценок

Как уже упоминалось, использование метода наименьших квадратов в случае обобщенной модели приводит к смещенным оценкам. Это объясняется тем, что матрица Ω , составленная из наблюдений y , искажена шумами:

$$\Omega = [U | Y] = [U | X] + \left[\begin{array}{c|cc} 0 & n(0) & \dots & n(1-m) \\ \vdots & & & \vdots \\ n(k-1) & \dots & n(k-m) \end{array} \right],$$



Ф и г. 6.15.

или в матричной форме

$$\Omega = \Omega_{u,y} = \Omega_{u,x} + \Omega_{0,n}. \quad (6.100)$$

Наличие этих помех противоречит предположениям, при которых можно применять метод наименьших квадратов. Для того чтобы убедиться в том, что, за исключением весьма частных случаев, оценки оказываются смещеными, рассмотрим простой пример (фиг. 6.15).

Пример. Уравнения объекта:

$$x(k+1) + ax(k) = bu(k),$$

$$y(k) = x(k) + n(k).$$

Уравнение модели:

$$e(k) = y(k) + ay(k-1) - \beta u(k-1).$$

Оценка по настраиваемой модели (концептуальной) получается в результате минимизации следующей функции потерь:

$$E = \sum_{i=1}^k e^2(i).$$

Необходимые условия

$$\frac{\partial E}{\partial \alpha} = \sum_{i=1}^k e(i) y(i-1) = 0,$$

$$\frac{\partial E}{\partial \beta} = \sum_{i=1}^k e(i) u(i-1) = 0$$

можно записать в виде

$$\begin{aligned} \sum [& \{x(i) + \alpha x(i-1) - \beta u(i-1) + n(i) + \\ & + \alpha n(i-1)\} \{x(i-1) + n(i-1)\}] = 0, \\ \sum [& \{x(i) + \alpha x(i-1) - \beta u(i-1) + n(i) + \\ & + \alpha n(i-1)\} u(i-1)] = 0. \end{aligned} \quad (6.101)$$

Эти уравнения обращаются в тождества на *несмешенных оценках* $\alpha = a$ и $\beta = b$, если

- a) $n(i) = 0$, $i = 1, \dots, k$, т. е. помеха отсутствует;
- б) $\sum_i n(i) n(i-1) + a \sum_i n^2(i-1) = 0$,

что можно аппроксимировать соответствующими точками автокорреляционной функции

$$\Psi_{nn}(1) + a \Psi_{nn}(0) = 0,$$

т. е. имеется помеха, которая, пройдя часть обобщенной модели, преобразуется в белый шум невязок. Во всех остальных случаях имеет место *смещение*. Для того чтобы понять механизм этого смещения, рассмотрим простой пример. Пусть уравнение объекта имеет вид

$$y(k) - \overset{a}{0,5} y(k-1) = \overset{b}{1,0} u(k-1) + n(k) + 0,1 n(k-1).$$

Здесь $n(k)$ — последовательность независимых гауссовых $(0, 1)$ случайных величин, однако для идентификации объекта используется метод наименьших квадратов в предположении, что невязки некоррелированы. Ниже приводится типичный результат, полученный по 500 наблюдениям входа и выхода:

Параметры объекта	Оценки
$a = -0,5$	$\hat{a} = -0,643 \pm 0,029$
$b = 1,0$	$\hat{b} = 1,018 \pm 0,062$

Таким образом, имеет место очень неприятная ситуация: оценка \hat{a} неверна и к тому же вероятность того, что эта оценка ошибочна, близка к единице (истинное значение отличается от оценки на 5σ).

Теперь рассмотрим более подробно свойства оценки ¹⁾. Подстановка соотношений

$$\Omega_{u,x} = \Omega_{u,y} - \Omega_{0,n}$$

и

$$y = \Omega'_{u,y} t + n,$$

где $t' = [b' \ -a']$ — вектор параметров объекта, в выражение для оценки (6.98) дает

$$\begin{aligned} \hat{\theta} &= [\Omega'_{u,y} \Omega_{u,y}]^{-1} \Omega'_{u,y} y = \\ &= [\Omega'_{u,y} \Omega_{u,y}]^{-1} \Omega'_{u,y} [\Omega_{u,y} t - \Omega_{0,n} t + n]. \end{aligned} \quad (6.102)$$

Запишем математическое ожидание

$$\mathcal{E}[\hat{\theta}] = t - \mathcal{E}[[\Omega'_{u,y} \Omega_{u,y}]^{-1} [\Omega'_{u,y} \Omega_{0,n} t - \Omega'_{u,y} n]]. \quad (6.103)$$

Второе слагаемое в правой части представляет собой смещение оценки. Теперь исследуем асимптотические свойства этой оценки при увеличении числа наблюдений ($k \rightarrow \infty$). Рассмотрим

$$\left[\frac{\Omega'_{u,y} \Omega_{u,y}}{k} \right]^{-1} \left[\frac{\Omega'_{u,y} \Omega_{0,n}}{k} t - \frac{\Omega'_{u,y}}{k} n \right].$$

Здесь величина смещения умножена и поделена на k . Если u и n — стационарные случайные процессы, то средние по выборке являются состоятельными оценками

¹⁾ Частное сообщение Шоу (Shaw, 1970).

авто- и взаимной корреляции. Используя обозначение $p \cdot \lim_{k \rightarrow \infty}$ ¹⁾ — предел по вероятности — и теорему Слуцкого²⁾, из формулы (6.102) получим

$$p \cdot \lim_{k \rightarrow \infty} \hat{\theta} = t - \left[p \cdot \lim_{k \rightarrow \infty} \frac{\Omega'_{u,y} \Omega_{u,y}}{k} \right]^{-1} \times \\ \times \left[p \cdot \lim_{k \rightarrow \infty} \frac{\Omega'_{u,y} \Omega_{0,n}}{k} t - p \cdot \lim_{k \rightarrow \infty} \frac{\Omega'_{u,y}}{k} n \right], \quad (6.104)$$

где

$$p \cdot \lim_{k \rightarrow \infty} \frac{\Omega'_{u,y} \Omega_{u,y}}{k} = Q,$$

$$p \cdot \lim_{k \rightarrow \infty} \frac{\Omega'_{u,y} \Omega_{0,n}}{k} =$$

$$= \begin{bmatrix} 0 & & 0 \\ \vdots & \Psi_{nn}(0) & \dots & \Psi_{nn}(1-m) \\ 0 & \vdots & \vdots & \vdots \\ \Psi_{nn}(1-m) & \dots & \Psi_{nn}(0) \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & N \end{bmatrix},$$

$$p \cdot \lim_{k \rightarrow \infty} \frac{\Omega'_{u,y} n}{k} = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ \vdots \\ \Psi_{nn}(1) \\ \vdots \\ \Psi_{nn}(m) \end{bmatrix}.$$

¹⁾ Неслучайный вектор \mathbf{c} является $p \cdot \lim_{k \rightarrow \infty} \mathbf{x}(k)$ при $k \rightarrow \infty$, т. е. $p \cdot \lim_{k \rightarrow \infty} \mathbf{x}(k) = \mathbf{c}$, если

$\lim_{k \rightarrow \infty} P(|\mathbf{x}(k) - \mathbf{c}| \geq \varepsilon) = 0$ для всех $\varepsilon \geq 0$.

Это означает, что

$$\lim_{k \rightarrow \infty} E[\mathbf{x}(k)] = \mathbf{c}.$$

²⁾ Если элементы матриц $\mathbf{A}(k)$ и $\mathbf{B}(k)$ сходятся по вероятности, то

$p \cdot \lim_{k \rightarrow \infty} [\mathbf{A}^{-1}(k) \mathbf{B}(k)] = [p \cdot \lim_{k \rightarrow \infty} \mathbf{A}(k)]^{-1} p \cdot \lim_{k \rightarrow \infty} \mathbf{B}(k)$
(см. [13], стр. 118).

Считая ковариационную матрицу \mathbf{Q} положительно определенной, находим

$$\lim_{k \rightarrow \infty} \mathcal{E}[\hat{\theta}] = p \cdot \lim_{k \rightarrow \infty} \hat{\theta} =$$

$$= \mathbf{t} - \mathbf{Q}^{-1} \begin{bmatrix} \begin{bmatrix} 0 & 0 \\ 0 & N \end{bmatrix} & \begin{bmatrix} b_1 \\ \vdots \\ b_m \\ -a_1 \\ \vdots \\ -a_m \end{bmatrix} - \begin{bmatrix} 0 \\ \vdots \\ 0 \\ \Psi_{nn}(1) \\ \vdots \\ \Psi_{nn}(m) \end{bmatrix} \end{bmatrix}. \quad (6.105)$$

Если выражение, заключенное в квадратные скобки, не равно нулю, то имеется *асимптотическое смещение*. Асимптотическое смещение отсутствует, если

$$\mathbf{N}\mathbf{a} = -\begin{bmatrix} \Psi_{nn}(1) \\ \vdots \\ \Psi_{nn}(m) \end{bmatrix}, \quad (6.106)$$

т. е. если $n(k)$ порождается белым шумом $w(k)$:

$$n(k) + a_1 n(k-1) + \dots + a_m n(k-m) = w(k).$$

Это можно показать, умножая на $n(k-1)$ и переходя к математическим ожиданиям:

$$\mathcal{E}[n(k)n(k-1) + a_1 n^2(k-1) + \dots + a_m n(k-m)n(k-1)] =$$

$$= \mathcal{E}[w(k)n(k-1)],$$

.

$$\mathcal{E}[n(k)n(k-m) + a_1 n(k-1)n(k-m) + \dots + a_m n^2(k-m)] =$$

$$= \mathcal{E}[w(k)n(k-m)].$$

Здесь правые части равны нулю, так как $n(k-i)$ может зависеть только от предыдущих значений $w(k)$. Следовательно, эти соотношения совпадают с (6.106). Это условие несмещенности оценки соответствует условию «белого шума невязок», поскольку в обозначениях уравнения (6.93) для истинных значений параметров $\mathcal{A}^*(z^{-1}) = A^*(z^{-1})$

$$\text{и } \mathcal{B}^*(z^{-1}) = B^*(z^{-1})$$

— уравнение модели: $e(k) = A^*(z^{-1})x(k) + A^*(z^{-1})n(k) - B^*(z^{-1})u(k)$	
уравнение объекта: $0 = A^*(z^{-1})x(k) - B^*(z^{-1})u(k)$	

$$\begin{aligned} e(k) &= A^*(z^{-1})n(k) = \\ &= n(k) + a_1n(k-1) + \dots + a_mn(k-m) = w(k). \end{aligned}$$

Используя скользящее среднее

$$n(k) + a_1n(k-1) + \dots + a_mn(k-m),$$

можно представить помеху в виде белого шума (фиг. 6.16)

$$n(k) + a_1n(k-1) + \dots + a_mn(k-m) = w(k).$$

Условие «белого шума невязок» можно интерпретировать иначе. Рассмотрим описание объекта и внешней среды, представленное на фиг. 6.17:

$$\begin{aligned} A^*(z^{-1})x(k) &= B^*(z^{-1})u(k), \\ y(k) &= x(k) + n(k). \end{aligned}$$

Следовательно,

$$A^*(z^{-1})y(k) = B^*(z^{-1})u(k) + A^*(z^{-1})n(k). \quad (6.107)$$

Если $n(k)$ порождается белым шумом $w(k)$:

$$D^*(z^{-1})n(k) = F^*(z^{-1})w(k), \quad (6.108)$$

то, подставляя (6.108) в (6.107), находим

$$A^*(z^{-1})y(k) = B^*(z^{-1})u(k) + \frac{C^*(z^{-1})}{D^*(z^{-1})}w(k), \quad (6.109)$$

где

$$C^*(z^{-1}) = A^*(z^{-1})F^*(z^{-1}).$$

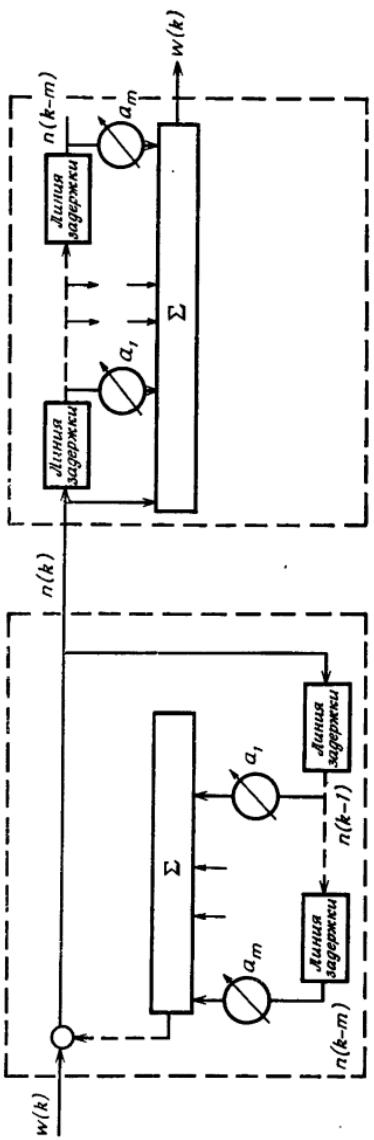
Вводя обозначения

$$\tilde{y}(k) = \frac{D^*(z^{-1})}{C^*(z^{-1})}y(k),$$

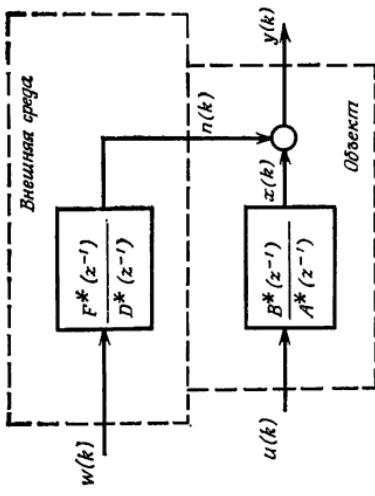
$$\tilde{u}(k) = \frac{D^*(z^{-1})}{C^*(z^{-1})}u(k),$$

можно переписать уравнение (6.107) в виде

$$A^*(z^{-1})\tilde{y}(k) = B^*(z^{-1})\tilde{u}(k) + w(k). \quad (6.110)$$

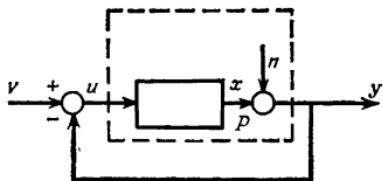


Фиг. 6.16.



Фиг. 6.17.

В результате, используя обобщенную модель, связывающую $\tilde{u}(k)$ и $\tilde{y}(k)$, приходим к невязкам типа белого шума и несмешенным оценкам.



Фиг. 6.18.

Объекты с обратной связью

Близкая задача возникает, если идентифицируемый объект охвачен обратной связью (фиг. 6.18). Заметим, что помеха действует внутри контура обратной связи. Интересно определить характеристики объекта P . В качестве входных сигналов можно рассматривать u или v . В первом случае при оценивании используются переменные u и y и определяются характеристики всей системы, охваченной обратной связью. Определение по этим характеристикам параметров объекта P не является тривиальной задачей, в частности, переход от замкнутого контура к разомкнутому может привести к накапливанию ошибок оценивания.

Во втором случае, когда при оценивании используются переменные u и y , возникает другая задача. Из-за обратной связи входной сигнал $u = v - y$ утрачивает статистическую независимость от помехи n , которая, как мы выяснили, является необходимым условием получения несмешенных оценок.

Пример. Рассмотрим задачу определения весовой функции импульсной системы для дискретного случая (выборки). Объект описывается уравнением

$$x(k) = b'_0 u(k) + b'_1 u(k-1) + \dots + b'_{m-1} u(k-m) = \\ = \sum_{i=0}^{m-1} b'_i u(k-i), \quad (6.111)$$

или

$$x(k) = \sum_{i=0}^{m-1} b'_i v(k-i) - \sum_{i=0}^{m-1} b'_i y(k-i). \quad (6.112)$$

Поскольку

$$y(k) = x(k) + n(k),$$

можно записать явное выражение для выходного сигнала

$$y(k) = \sum_{i=0}^m b_i v(k-i) - \sum_{i=1}^m b_i y(k-i) + \frac{1}{1+b'_0} n(k), \quad (6.113)$$

где $b_i = b'_i / (1 - b'_0)$. Условие «белого шума невязок» можно рассматривать, как и раньше.

Для сохранения свойства линейности по параметрам обобщенной модели и устранения вредных эффектов, связанных с аддитивной помехой, было разработано несколько методов [7]:

- а) многократный метод наименьших квадратов;
- б) метод вспомогательных переменных;
- в) принцип копирования;
- г) обобщенный метод наименьших квадратов и связанные с ним алгоритмы;
- д) метод максимального правдоподобия;
- е) метод Левина.

Многократный метод наименьших квадратов. Метод состоит в последовательном увеличении порядка модели:

$$e(k) = \mathcal{A}^*(z^{-1}) y(k) - \mathcal{B}^*(z^{-1}) u(k). \quad (6.114)$$

Степень соответствия модели объекту оценивается по величине невязок. Если увеличение порядка модели не приводит к существенному уменьшению невязок, то процедура оценивания прекращается. Острем [5] на примере простой системы первого порядка показал, что деление $\mathcal{A}^*(z^{-1})$ на $\mathcal{B}^*(z^{-1})$ приводит к хорошей модели, а остатком при делении можно пренебречь. По-видимому, строгого обоснования этого метода до сих пор не существует.

Метод вспомогательных переменных. Можно предложить простое решение задачи о величине смещения оценок вследствие действия аддитивных помех. Умножив уравнение объекта

$$\mathbf{y} = \Omega \mathbf{b} + \mathbf{n} \quad (6.115)$$

на матрицу Ω' , получим

$$\Omega' \mathbf{y} = \Omega' \Omega \mathbf{b} + \Omega' \mathbf{n}. \quad (6.116)$$

Тогда если Ω и \mathbf{n} статистически независимы, то система нормальных уравнений принимает вид

$$\hat{\mathbf{b}} = [\Omega' \Omega]^{-1} \Omega' \mathbf{y}. \quad (6.117)$$

Допустим теперь, что имеется матрица V , которая коррелирована с u , но не коррелирована с n :

$$p \cdot \lim [V' n] = 0, \quad p \cdot \lim [V' \Omega] \text{ положительно определена};$$

тогда

$$V'y = V'\Omega b + V'n = V'\Omega \hat{\beta},$$

так как n не измеряется, или

$$\hat{\beta} = [V'\Omega]^{-1} V'y, \quad (6.118)$$

что гарантирует асимптотическую несмещеннность оценки. Матрица V состоит из так называемых *вспомогательных переменных* (фиг. 6.19). Одним из частных применений этого метода является случай, когда исследуемый объект охвачен обратной связью (см. фиг. 6.18). Так как сигнал w не коррелирован с n , то w можно использовать в качестве вспомогательной переменной. Конечно, произвольный выбор V не обеспечивает оптимальности оценок, но можно найти такие вспомогательные переменные, которые придают оценке оптимальные свойства. По всем этим вопросам см. [2, 19, 21, 22, 32, 35, 41]. Схемы, предложенные в работах [19, 30], близки к методу вспомогательных переменных. В работе [38] в качестве вспомогательных переменных используются функции псевдочувствительности.

Принцип копирования. Как показано в работе [31], в этом случае оценивание начинается с канонического представления вида

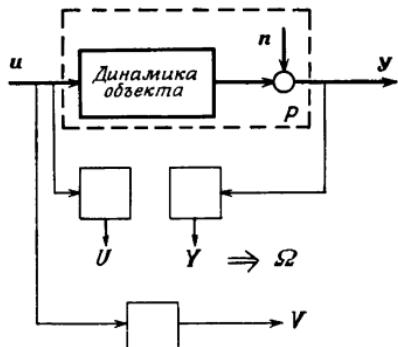
$$y(k) + a_1 y(k-1) + \dots + a_m y(k-m) = \\ = b_1 u(k-1) + \dots + b_m u(k-m) + n(k), \quad (6.119)$$

где

$$n(k) = v(k) + c_1 v(k-1) + \dots + c_m v(k-m) \quad (6.120)$$

и $\{v(k)\}$ — последовательность некоррелированных случайных величин. Отсюда следует, что

$$\begin{aligned} \mathcal{E}[n(i)u(i-j)] &= 0 \quad \text{для } j > 0, \\ \mathcal{E}[n(i)y(i-j)] &= 0 \quad \text{для } j > m. \end{aligned} \quad (6.121)$$



Ф и г. 6.19.

Для любой оценки неизвестных параметров находится последовательность $\{n^*(i)\}$. Теперь для последовательности k наблюдений имеем

$$\begin{aligned} \frac{1}{k} \sum_{i=1}^k n^*(i) u(i-j), \\ \frac{1}{k} \sum_{i=1}^k n^*(i) y(i-j). \end{aligned} \quad (6.122)$$

Эти выражения должны как можно лучше в смысле минимума среднеквадратической ошибки копировать (дублировать) соответствующие математические ожидания.

Обобщенный метод наименьших квадратов рассматривается в гл. 7.

Метод максимального правдоподобия рассматривается в гл. 11.

Метод Левина. Частный метод оценки смещения, вызванного корреляцией невязок, был предложен в работах Левина [25—27] для случая детерминированной системы с независимыми ошибками наблюдений. Результаты Левина основаны на методе Купмана и описывают смещение через собственные числа и векторы матрицы $\Omega' \Omega$. Строгое изложение метода Левина, включающее доказательство сходимости и оценку ошибок, дано в работе [4]. Этот метод также использован в работе [36].

ЛИТЕРАТУРА

1. Aitken A. C., *Statistical Mathematics*, Oliver and Boyd, Edinburgh, 1952.
2. Andeen R. E., Shipley P. P., *Digital Adaptive Flight Control Systems for Aerospace Vehicles*, *AIAA J.*, 1 (1963).
3. Anderson T. W., *The Choice of the Degree of a Polynomial Regression as a Multiple Decision Problem*, *Ann. Math. Stat.*, 33, 255—265 (1962).
4. Aoki M., Yue P. C., *On Certain Convergence Questions in System Identification*, *SIAM J. Control*, 8, 239—256 (1970).
5. Åström K. J. (см. литературу к гл. 11).
6. Åström K. J., *Lectures on the Identification Problem — the Least Squares Method*, Report Lund Inst. of Technology, Div. Automatic Control, Sweden, 1968.
7. Åström K. J., Eykhoff P. (см. литературу к гл. 2).

8. Cramer H., Mathematical Methods of Statistics, Princeton Univ. Press, Princeton, 1946; есть русский перевод: Крамер Г., Математические методы статистики, ИЛ, 1948.
9. Deutsch R. (см. литературу к гл. 5).
10. Douce J. L., Roberts P. D., Orthogonal Functions, *Control*, 8, 457—461 (1964); см. также *Control*, 8, 555 (1964).
11. Draper N. R., Smith H., Applied Regression Analysis, Wiley, N.Y., 1966.
12. Durbin J., Estimation of Parameters in Time-Series Regression Models, *Jl. R. Statist. Soc., Series B*, 22, 139—153 (1960).
13. Elgerd O. I. (см. литературу к гл. 4).
14. Fraser D. A. S., Statistics, an Introduction, Wiley, N.Y., 1958.
15. Gersch W. (см. литературу к разд. 14.7).
16. Graybill F. A., An Introduction to Linear Statistical Models, McGraw-Hill, N.Y., 1961.
17. Grenander U., Rosenblatt M. (см. литературу к гл. 4).
18. Hoppe S. G., A Least Squares Technique for the Identification of a Linear Time-Invariant Plant in a Sampled Data System (Correspondence), *IEEE Trans. Autom. Control*, AC-10, 490 (1965).
19. Hsia T. C., Landgrebe D. A., On a Method for Estimating Power Spectra, *IEEE Trans. Instrum. and Measurem.*, IM-16, 255—257 (1967).
20. Jenkins G., Watts D. (см. литературу к гл. 4).
21. Joseph P., Lewis J., Tou J., Plant Identification in the Presence of Disturbances and Application to Digital Adaptive Systems, *AIEE Transact.*, 80, Part II (Applic. and Ind.), 18—24 (1961).
22. Kendall M. G., Stuart A., The Advanced Theory of Statistics, vol. 1 (1962), 2nd ed., vol. 2, (1961), Griffin, London; есть русский перевод: Кендалл М., Стюарт А., Теория распределений, изд-во «Наука», 1966.
23. Kwakernaak H., Some Observation on the Error in the Estimation of Impulse Response (Correspondence), *IEEE Trans. Circuit Theory*, CT-10, 458—460 (1963).
24. Leonov Yu. P., On Estimate of Parameters of Linear Dynamic Systems, IFAC Symp. «Identification in Autom. Control Systems», Prague, Paper 1.2, 1967.
25. Levin M. J., Estimation of the Characteristics of Linear Systems in the Presence of Noise, Sc. D. Thesis, E. E. Dept. Columbia Univ., N.Y. 1959.
26. Levin M. J., Optimum Estimation of Impulse Response in the Presence of Noise, *IRE Trans. Circuit Theory*, CT-7, 50—56 (1960).
27. Levin M. J., Estimation of System Pulse Transfer Function in the Presence of Noise, *IEEE Trans. Autom. Control*, AC-9, 229—235 (1964); см. также AC-9, 214—215 (1964).
28. Mayne D. Q., A Method for Estimating Discrete Time Transfer Functions, IEE Convention «Advances in Computer Control», April 1967.
29. Melis J. H. A. M., Explicit Estimation of Process Parameters Using a Digital Computer (на голландском), M. Sc. Thesis, EE Dept., Univ. of Technology, Eindhoven, Netherlands, 1967.

30. Peterka V., Šmuk K., On-Line Estimation of Dynamic Model Parameters from Input-Output Data, Fourth IFAC Congress, Warsaw, Paper 26.1, 1969.
31. Peterka V., Halousková A., Tally Estimate of Åström Model for Stochastic Systems, Second IFAC Symp. «Identification and Process Parameter Estimation», Prague, Paper 2.3, 1970.
32. Reiersoll O., Confluence Analysis by Means of Lag Moments and Other Methods of Confluence Analysis, *Econometrica*, 9, 1—23 (1941).
33. Roberts P. D., Self Adjustable Orthogonal Digital Filters for System Identification and Optimization, *Proc. IEEE*, 114 (1967); см. также *Electron. Letters*, 2, 187—188 (1966).
34. Rosenberg A. E., Shen D. W. C., Regression Analysis and Its Application to the System-Identification Problem, Joint Autom. Control Conf., 1963.
35. Smets A. J., The Instrumental Variable Method and Related Identification Schemes, Rept. E.E. Dept., Univ. of Technology, Eindhoven, Netherlands, 1970.
36. Smith F. W., System Laplace-Transform Estimation from Sampled Data, *IEEE Trans. Autom. Control*, AC-13, 37—44 (1968).
37. Van den Boom A. J. W., Melis J. H. A. M., A Comparison of Some Process Parameter Estimating Schemes, Fourth IFAC Congress, Warsaw, Paper 26.5, 1969.
38. Vušković M. I., Bingulac S. P., Djorović M., Applications of the Pseudo Sensitivity Functions in Linear Dynamic Systems Identification, Second IFAC Symp. «Identification and Process Parameter Estimation», Prague, Paper 2.5, 1970.
39. Westlake J. R., A Handbook of Numerical Matrix Inversion and Solution of Linear Equations, Wiley, N.Y., 1968.
40. Wold H., A Study in the Analysis of Stationary Time Series, Almqvist and Wiksell, Uppsala, 1938.
41. Wong K. Y., Polak E., Identification of Linear Discrete Time Systems Using the Instrumental Variable Method, *IEEE Trans. Autom. Control*, AC-12, 707—718 (1967).
42. Woodside C. M., Estimation of the Order of Linear Systems, Second IFAC Symp. «Identification and Process Parameter Estimation», Prague, Paper 1.7, 1970.
43. Whittle P. (1963) (см. литературу к гл. 4).

ДОПОЛНИТЕЛЬНАЯ ЛИТЕРАТУРА

Balakrishnan A. V., Determination of Nonlinear System from Input-Output Data, Conference on Identification and Representation Problems, Princeton Univ., Princeton, N.J., 1963.

Bigelow S. C., Ruge H., An Adaptive System Using Periodic Estimation of the Pulse Transfer Function, *IRE Int. Conv. Rec.*, 9, Part 4, 25—38 (1961).

Diamessis J. E., On the Determination of the Parameters of Certain Nonlinear Systems (Correspondence), *Proc. IEEE*, 53, 319—320 (1965).

- Durbin J., Watson G. S., Testing for Serial Correlation in Least Squares Regression, *Biometrika*, 37, 409—428 (1950).
- Fagin S. L., Recursive Linear Regression Theory, Optimal Filter Theory and Error Analysis of Optimal Systems, *IEEE Convent Record*, 12, part 1 (1964).
- Fantazzzi G., The Relation Between the Sampling Time and Stochastic Error for the Impulsive Response of Linear Time-Independent Systems (Short Paper), *IEEE Trans. Autom. Control*, AC-13, 426—428 (1968).
- Girshick M. A., Savage L. J., Bayes and Minimax Estimation for Quadratic Loss Functions, Neyman J. (ed.), Second Berkeley Symp. Math. Statist. Probab., 1957, pp. 53—73.
- Graupe D., Swanick B. H., Časir G. R., Reduction and Identification of Multivariable Processes Using Regression Analysis (Short Paper), *IEEE Trans. Autom. Control*, AC-13, 564—567 (1968).
- Graybill F. A., On a Method of Correlation Analysis for Multivariate Systems, Fourth IFAC Congress, Warsaw, Paper 33.1, 1969.
- Hanuš B., Vaniček J., Evaluation of Transient Characteristics from Measured Values, IFAC Symp. «Identification in Autom. Control Systems», Prague, Paper 4.8, 1967.
- Hsia T. C., On Power Spectrum Identification Methods (Correspondence), *IEEE Trans. Autom. Control*, AC-12, 781—782 (1967).
- Kerr R. B., Surber W. H., Jr., Precision of Impulse Response Identification Based on Short Normal Operating Records, *IRE Trans. Autom. Control*, AD-6, 173—182 (1961).
- Кичатов Ю. Ф., Определение нелинейных характеристик объектов управления при гауссовых входных воздействиях, *Автоматика и телемеханика*, № 3, 470—474 (1965).
- Koopmans T. C. Statistical Inference in Dynamic Economic Models, Cowless Commission Monograph, 40, Wiley, 1950.
- Koszelnik M., Malkiewicz J., Trybuta St. (см. литературу к разд. 14.4).
- Кутиков Л. М., Обращение корреляционных матриц и некоторые задачи самонастройки, *Изв. АН СССР, Техн. кибернетика*, № 5, 42—47 (1965).
- Lindenlaub J. C., Cooper G. R., Noise Limitations of System Identification Techniques, *IEEE Trans. Autom. Control*, AC-8, 43—48 (1963).
- Линник Ю. В., Метод наименьших квадратов и основы математико-статистической теории обработки наблюдений, 2-е изд., Физматиздат, 1962.
- Loeb J., Cohen G. Derivation of the Dynamic Parameters from Recorded Measurements (на французском) *Automatisme*, 8 (1963).
- Magnes T. A., McGuire J. B., Comparison of Least Square and Minimum Variance Estimate of Regression Parameters, *Ann. of Math. Stat.*, 33, 462 (1962).
- Malinvaud E., Statistical Methods of Econometrics, Studies in Mathematical and Managerial Economics, 6, McNalley, Chicago, 1966.
- Mann H. B., Wald A. (см. литературу к гл. 4).
- Parzen E., Mathematical Considerations in the Estimation of Spectra, *Technometrics*, 3, 167—190 (1961).

Perel'man I. I., Estimation Accuracy of Impulse Response of a Linear Dynamic Plant, IFAC Symp. «Identification in Autom. Control Systems», Prague, Paper 1.9, 1967.

Placket R. L., Principles of Regression Analysis, Oxford Univ. Press, N.Y., 1960.

Попов Е. П., Лоскутов Г. М., Юсупов Р. М., О самонастраивющихся системах управления без пробных возмущающих воздействий, в книге «Дискретные и самонастраивающиеся системы», Труды II Межд. конгресса ИФАК, Базель, изд-во «Наука», 1965, стр. 338—350.

Efroymson M. A., Multiple Regression Analysis, in: Mathematical Methods for Digital Computers, Ralston A., Wilf H. (eds.), Wiley, N.Y., 1960, pp. 191—203.

Rubin O., Process Identification by Generalized Correlation (Short Paper), *IEEE Trans. Autom. Control*, AC-16, 361—363 (1971).

Shaw L., Robinson G., Invariant Estimation of Stochastic System Parameters, IFAC Symp. «Identification in Autom. Control Systems», Prague, Paper 3.16, 1967.

Tretter S. A., Steiglitz K., Power Spectrum Identification in Terms of Rational Models, *IEEE Trans. Autom. Control*, AC-12, 185—188 (1967).

Tretter S. A. (см. литературу к гл. 11).

Voronova L. I., Some Methods of Controlled Plants Dynamic Characteristics Determination and Investigation of Their Accuracy, IFAC Symp. «Identification in Autom. Control Systems», Prague, Paper 3.19, 1967.

Воронова Л. И., Крементуло Ю. В., Новый метод определения характеристик сложных динамических систем, в книге «Сложные системы управления», изд-во «Наукова думка», Киев, 1966, стр. 82—104.

Wegrzijn S., Denis G., Delisle J., The Identification of Processes by Minimization of the Distance Between Sets of Signals, Fourth IFAC Congress Warsaw, Paper 15.1, 1969.

Williams E. J., Regression Analysis, Wiley, N.Y., 1959.

ГЛАВА 7

ИМПУЛЬСНЫЕ СИГНАЛЫ (ВЫБОРОЧНЫЕ ЗНАЧЕНИЯ); НЕЯВНЫЕ МЕТОДЫ ИЛИ МЕТОДЫ С НАСТРАИВАЕМОЙ МОДЕЛЬЮ

В предыдущей главе для оценок метода наименьших квадратов и марковских оценок получена система нормальных уравнений. Эта система легко решается на ЦВМ. Имеются, однако, и недостатки:

1) обращение матриц требует больших затрат машинного времени;

2) для того чтобы получить приемлемое время счета, требуется оперативная память большого объема, поскольку в ней должна храниться вся выборка значений входного и выходного сигналов;

3) метод этот не позволяет вести вычисления в натуральном масштабе времени, т. е. обрабатывать результаты наблюдений по мере их поступления, возможно, «забывая» при этом предыдущие наблюдения.

В связи с этим третьим недостатком рассмотренный в предыдущей главе явный метод также может быть охарактеризован как *однократная*, или *одношаговая*, процедура с накоплением данных.

Во многих приложениях весьма желательно производить идентификацию объекта в процессе его функционирования. Например, можно представить себе процедуру, которая обрывается при достижении заданной точности оценки параметра. В этом случае возникает следующая задача. Допустим, что оценка $\hat{\beta}(i)$ получена по i парам наблюдений входного и выходного сигналов. Нужно ли для построения $\hat{\beta}(i+1)$ повторять процедуру идентификации, используя всю последовательность наблюдений за входом и выходом или можно использовать результаты вычисления $\hat{\beta}(i)$? Процедура идентификации, которая не

требует на каждом шаге повторной обработки всей последовательности наблюдений, называется *рекуррентной*. Такие процедуры могут использоваться в приложениях, в которых желательно обрабатывать новые наблюдения в процессе их поступления, т. е. *в натуральном масштабе времени (в замкнутом контуре)*. Этот метод, помимо того, что имеет практическую ценность, позволяет установить связь с другими разделами теории управления: нелинейной фильтрацией, стохастической аппроксимацией, обучением и адаптацией.

Если параметры объекта меняются во времени, то естественно, что не остается ничего другого, кроме отслеживания их изменений в реальном масштабе времени. Это называется *идентификацией в реальном масштабе времени*.

Существует много способов синтеза алгоритмов для явного решения задач оценивания. Однако по существу все эти методы приводят к очень похожим алгоритмам вида

$$\hat{\beta}(i+1) = \hat{\beta}(i) + \Gamma(i) e(i),$$

где $e(i)$ — рассмотренная выше (обобщенная) ошибка, а $\Gamma(i)$ — матрица коэффициентов усиления той или иной структуры.

В случае модели, линейной по параметрам, довольно несложно определить градиенты и использовать их при настройке параметров модели. Это рассматривается в разд. 7.1. При попытке применить эти методы к обобщенным моделям, так же как и в предыдущей главе, сталкиваются с проблемой смещенности оценок. Это означает необходимость разработки более тонких схем оценивания, которые обсуждаются в разд. 7.2. Для нелинейных по параметрам моделей необходимы специальные приемы определения градиента или должны использоваться прямые поисковые методы (см. разд. 7.3).

Следует отметить, что изложенные в разд. 5.2 и 5.3 сходящиеся процедуры, как детерминированные, так и стохастические, непосредственно применимы к обсуждаемой здесь задаче.

7.1. МОДЕЛИ, ЛИНЕЙНЫЕ ПО ПАРАМЕТРАМ

Задача рекуррентного, или итеративного, оценивания может быть представлена как задача улучшения старой оценки $\hat{\beta}(i)$. Наличие такой оценки означает использование некоторой модели (например, в виде программы для ЦВМ), как это показано на фиг. 7.1, а. Входной сигнал одновременно подается на объект и модель с настраиваемыми параметрами. Настройка параметров производится с помощью настраивающего устройства, на вход которого поступают выходные сигналы объекта y_i и модели w_i . В такой постановке задача идентификации в натуральном масштабе времени рассматривалась в [41].

По существу задача состоит в том, чтобы сконструировать такое настраивающее устройство, которое обеспечивало бы близость (в каком-либо смысле) параметров модели и параметров объекта. Снова отметим, что в этом случае оценка $\hat{\beta}(i)$ поступает в замкнутый контур, который обеспечивает самокорректировку различных типов ошибок (см. гл. 2).

Как отмечалось в гл. 4, линейность по параметрам приводит к относительно простым задачам, даже если объект обладает нелинейной динамикой. В этом разделе рассматривается модель следующей структуры:

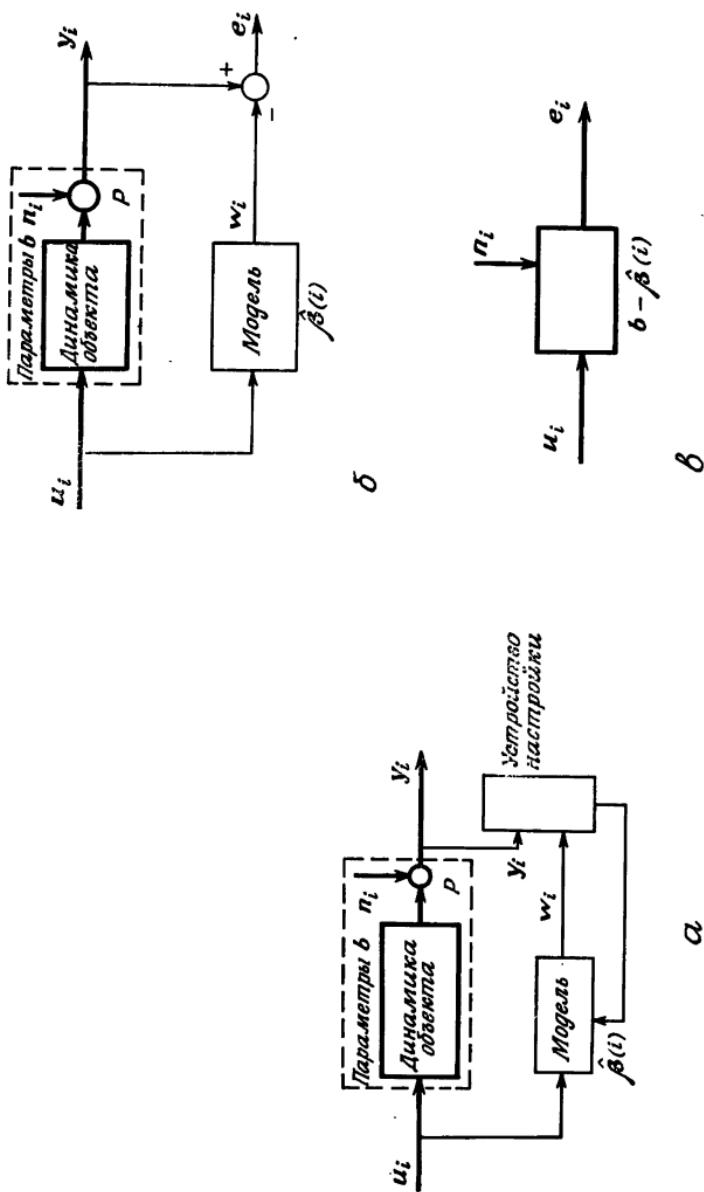
$$w(k) = \sum \beta_j u_j(k) \quad \text{и} \quad w = U\beta,$$

т.е.

$$e = y - U\beta.$$

Рекуррентные алгоритмы могут быть разной степени сложности. Здесь они обсуждаются по мере возрастания сложности в следующей последовательности:

- а) настройка, пропорциональная градиенту;
- б) метод наименьших квадратов для последовательности наблюдений;
- в) рекуррентный метод наименьших квадратов;
- г) стохастическая аппроксимация;
- д) принцип сжатых отображений.



Настройка, пропорциональная градиенту

Рассмотрим схему, изображенную на фиг. 7.1, б. Число параметров, которые нужно оценить, известно заранее; \mathbf{U} и \mathbf{n} — статистически независимы. Через $\hat{\beta}(i)$ обозначен вектор параметров модели на i -м интервале измерений. Следующее значение вектора параметров модели можно выбрать как предыдущее значение вектора, скорректированное на величину, пропорциональную градиенту функции ошибок:

$$\hat{\beta}(i+1) = \hat{\beta}(i) - \frac{1}{2} \Gamma(i) \nabla_{\beta} E_i, \quad (7.1)$$

где (см. приложение В)

$$\begin{aligned} E_i &= \mathbf{e}_i' \mathbf{R} \mathbf{e}_i, \\ \mathbf{e}_i &= \mathbf{y}_i - \mathbf{U}_i \hat{\beta}(i), \\ \nabla_{\beta} E_i &= -2 \mathbf{U}_i' \mathbf{R} \mathbf{e}_i, \end{aligned}$$

а \mathbf{y}_i и \mathbf{U}_i следует интерпретировать как наблюдения в i -м интервале измерений (фиг. 7.2). Таким образом,

$$\hat{\beta}(i+1) = \hat{\beta}(i) + \Gamma(i) \mathbf{U}_i' \mathbf{R} [\mathbf{y}_i - \mathbf{U}_i \hat{\beta}(i)]. \quad (7.2)$$

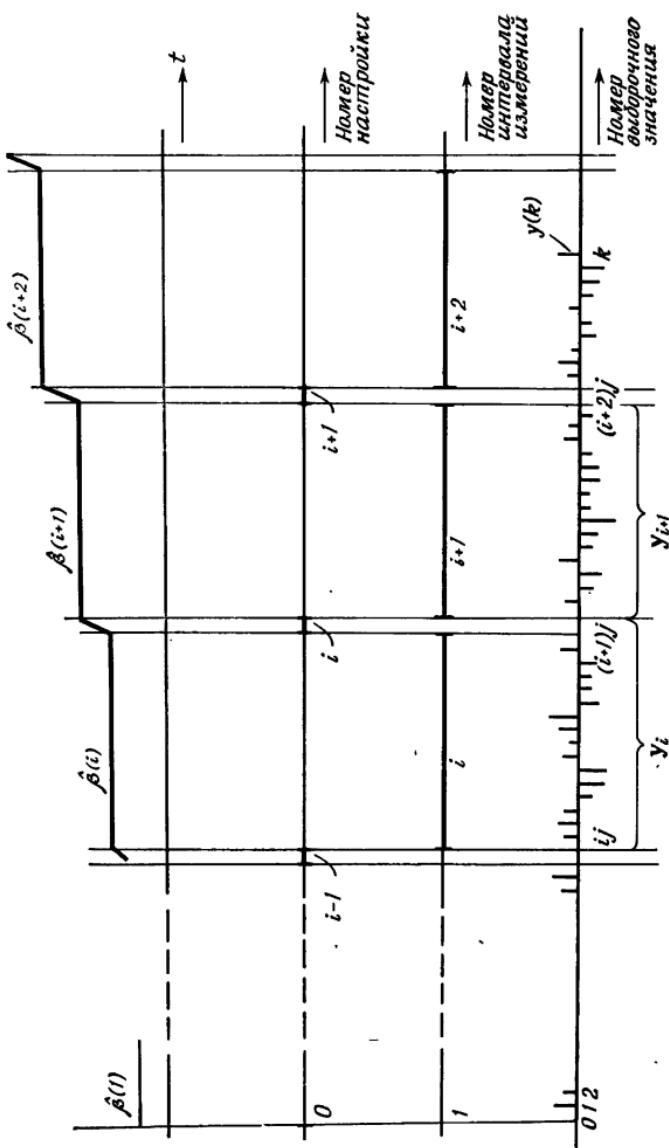
Среди возможных матриц весовых коэффициентов, как и раньше, $\mathbf{R} = \mathbf{I}$ и $\mathbf{R} = \mathbf{N}^{-1}$. Коэффициент усиления $\Gamma(i)$ может быть выбран как постоянным, так и меняющимся во времени (как функция i). Влияние этого выбора рассматривается в следующей части раздела.

Метод наименьших квадратов для последовательности наблюдений

В схеме, изображенной на фиг. 7.1, б, можно оценить вектор разности между параметрами объекта и параметрами модели

$$\Delta \hat{\beta}(i) = \mathbf{b} - \hat{\beta}(i) = [\mathbf{U}_i' \mathbf{R} \mathbf{U}_i]^{-1} \mathbf{U}_i' \mathbf{R} \mathbf{e}_i,$$

т. е. использовать уравнение (6.31) относительно \mathbf{e}_i , а не \mathbf{y} (см. фиг. 7.1, в). Это означает, что объект и модель частично компенсируют друг друга и оцениваются параметры



Фиг. 7.2.

возникающей в результате компенсации динамической системы. Следовательно,

$$\hat{\beta}(i+1) = \hat{\beta}(i) + \Gamma(i) \Delta \hat{\beta}(i).$$

Отсюда получаем, что

$$\hat{\beta}(i+1) = \hat{\beta}(i) + \Gamma(i) [U_i^T R U_i]^{-1} U_i R [y_i - U_i \hat{\beta}(i)]. \quad (7.3)$$

Сравните этот результат с уравнением (7.2). Снова можно положить $R = I$ и $R = N^{-1}$. В дальнейшем, если не оговорено противное, $R = I$. Необходимость обращения матрицы на каждом шаге процедуры относительно усложняет численную реализацию алгоритма, хотя в расчете на итерацию здесь используется меньшее число выборочных значений по сравнению с одношаговыми методами решения. Однако размер матрицы $U^T U$ не уменьшается.

Если помехи отсутствуют, то можно использовать быстро сходящийся алгоритм настройки, известный под названием «идентификация обучением». В этом случае уравнение (7.3) сводится к виду [6,23]

$$\hat{\beta}(i+1) = \hat{\beta}(i) + \Gamma \frac{u_i}{\|u_i\|^2} [y_i - U_i \hat{\beta}(i)].$$

В работе [12] рассматривается система с квантованным сигналом $\text{sgn } u_i$.

На практике задачи оценивания решаются в условиях действия аддитивного шума. Для этого случая свойства алгоритма (7.3) проанализированы в работе [39] (см. также [40]). Можно исследовать, как изменяется во времени математическое ожидание оценки. Имея в виду, что $y_i = U_i b + n_i$, из уравнения (7.3) находим

$$\mathbb{E}[\hat{\beta}(i+1)] = \Gamma(i) b + \{1 - \Gamma(i)\} \mathbb{E}[\hat{\beta}(i)], \quad i = 1, 2, \dots \quad (7.4)$$

Следовательно,

$$\mathbb{E}[\hat{\beta}(2)] = \Gamma(1) b + \{1 - \Gamma(1)\} \mathbb{E}[\hat{\beta}(1)],$$

$$\mathbb{E}[\hat{\beta}(3)] = \Gamma(2) b + \{1 - \Gamma(2)\} \Gamma(1) b + \{1 - \Gamma(2)\} \{1 - \Gamma(1)\} \mathbb{E}[\hat{\beta}(1)].$$

Поскольку $\hat{\beta}(1)$ — параметры исходной модели (известной), то можно положить $\mathbb{E}[\hat{\beta}(1)] = \hat{\beta}(1)$. Если апри-

орная информация о параметрах объекта отсутствует, естественно положить $\hat{\beta}(1) = 0$.

Интересно также рассмотреть изменение во времени ковариации. Пусть для простоты $R = I$. Тогда

$$\begin{aligned}\hat{\beta}(i+1) &= \Gamma(i)b + \{1 - \Gamma(i)\}\hat{\beta}(i) + \Gamma(i)[U_i'U_i]^{-1}U_i'n \\ \mathcal{E}[\hat{\beta}(i+1)] &= \Gamma(i)b + \{1 - \Gamma(i)\}\mathcal{E}[\hat{\beta}(i)] + 0\end{aligned}$$

$$\hat{\beta}(i+1) - \mathcal{E}[\hat{\beta}(i+1)] = \{1 - \Gamma(i)\}[\hat{\beta}(i) - \mathcal{E}[\hat{\beta}(i)]] + \Gamma(i)[U_i'U_i]^{-1}U_i'n.$$

Если n — широкополосная помеха, то в большинстве практических случаев $\hat{\beta}(i)$ и n_i можно рассматривать как независимые. Следовательно, имеем

$$\begin{aligned}\text{cov}[\hat{\beta}(i+1)] &= \{1 - \Gamma(i)\}^2 \text{cov}[\hat{\beta}(i)] + \\ &\quad + \Gamma^2(i)[U_i'U_i]^{-1}U_i'\mathcal{E}[n_i n_i'] U_i [U_i'U_i]^{-1},\end{aligned}\quad (7.5)$$

т. е. линейное уравнение с переменными коэффициентами. В случае аддитивной помехи в форме белого шума второе слагаемое сводится к

$$\Gamma^2(i)\sigma_n^2[U_i'U_i]^{-1}. \quad (7.6)$$

Если, кроме того, на вход системы подается достаточно длинный отрезок реализации белого шума (j выборочных значений), то второе слагаемое приблизительно равно

$$\Gamma^2(i)\frac{\sigma_n^2}{j\sigma_u^2}I. \quad (7.7)$$

Снова $\hat{\beta}(1)$ представляет собой исходную модель и, следовательно, является известным. Из этого вытекает, что $\text{cov}[\hat{\beta}(1)] = 0I$. Это начальное условие используется при исследовании изменения ковариации во времени.

Из формул (7.4) и (7.5) ясно, что выбор $\Gamma(i)$ решающим образом влияет на сходимость. Этот коэффициент усиления может быть выбран постоянным или зависящим от номера итерации настройки. В обоих случаях выбор $\Gamma(i) = 1$ для всех $i = l$ в соответствии с формулой (7.4) означает, что

$$\mathcal{E}[\hat{\beta}(i+1)] = b \quad (7.8)$$

для всех $i \geq l$. Таким образом, после l -й настройки оценка оказывается несмещенной независимо от коэффициентов исходной модели.

Реализация процедур оценивания в замкнутом контуре. Следуя работе [40], рассмотрим случаи

- а) $\Gamma(i) = 1/i$;
- б) $\Gamma(i) = c = \text{const.}$

а) Этот случай напоминает алгоритм стохастической аппроксимации (см. гл. 5). Из формул (7.5) и (7.7) следует, что при наличии белого шума n и белого шума u на входе системы

$$\text{cov}[\hat{\beta}(i+1)] = \frac{\sigma_n^2}{ij\sigma_u^2} \mathbf{I} \quad (7.9)$$

и

$$\lim_{i \rightarrow \infty} \text{cov}[\hat{\beta}(i+1)] = 0\mathbf{I}. \quad (7.10)$$

Это можно видеть, изображая графически последовательные значения элементов ковариационной матрицы векторов $\hat{\beta}(1), \hat{\beta}(2), \dots$.

Член ij в формуле (7.9) обозначает общий размер выборки, которая используется в течение i интервалов измерения. Сравнение с явными методами [см. формулу (6.63)] показывает, что в обоих случаях ковариация одинакова. Для достаточно малых выборок (j мало) соотношение

$$[\mathbf{U}_i^T \mathbf{U}_i]^{-1} \approx \frac{1}{j\sigma_u^2} \mathbf{I}$$

может не выполняться. В этом случае экспериментальные результаты обнаруживают сдвиг в сторону увеличения дисперсий (фиг. 7.3).

- б) Случай $\Gamma(i) = c$ приводит к формуле

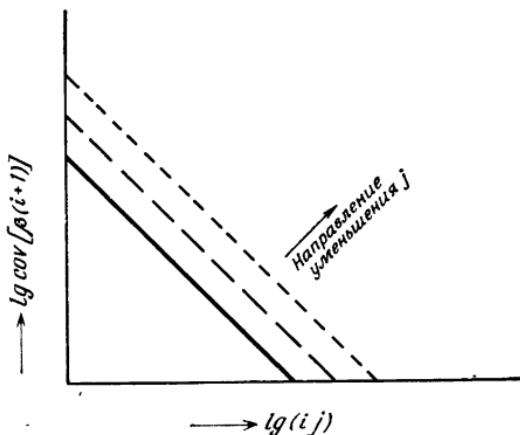
$$\mathcal{E}[\hat{\beta}(i+1)] = c\mathbf{b} + (1 - c)\mathcal{E}[\hat{\beta}(i)], \quad (7.11)$$

которую можно переписать в виде

$$\mathcal{E}[\hat{\beta}(i+1)] = \{1 - (1 - c)^i\}\mathbf{b} + (1 - c)^i \beta(1). \quad (7.12)$$

Следовательно,

$$\lim_{i \rightarrow \infty} \mathcal{E}[\hat{\beta}(i+1)] = \mathbf{b}, \quad 0 < c < 2, \quad (7.13)$$

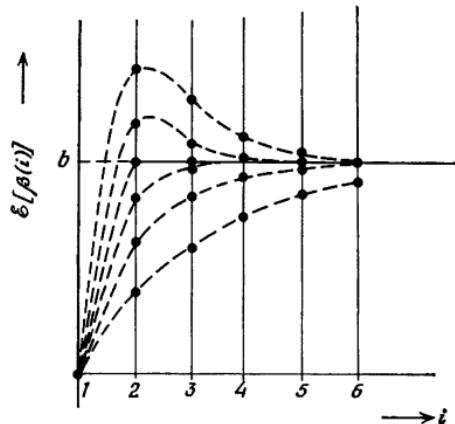


Ф и г. 7.3.

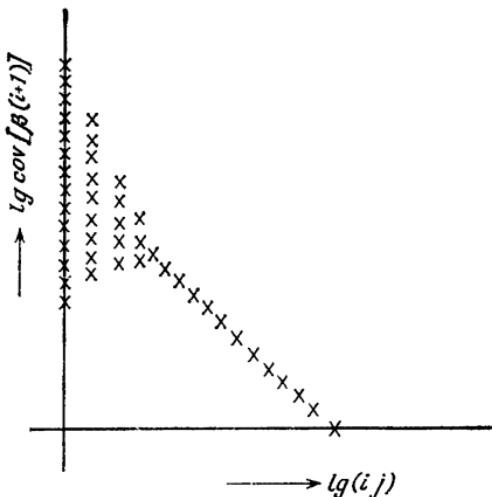
т. е. оценка является асимптотически несмещенной (фиг. 7.4). Исследуя выражения для ковариаций векторов $\hat{\beta}$ (2), $\hat{\beta}$ (3), ..., находим, что

$$\lim_{i \rightarrow \infty} \text{cov} [\hat{\beta}(i+1)] = \frac{c}{2-c} \frac{\sigma_n^2}{j\sigma_u^2} \mathbf{I}. \quad (7.14)$$

Таким образом, в случае постоянного усиления ковариация не стремится к нулю даже на бесконечном интервале наблюдения!



Ф и г. 7.4.



Ф и г. 7.5.

Очевидно, что метод «а» применим, если определяемые параметры не зависят от времени. Эти параметры могут быть найдены с любой заданной точностью, если только интервал наблюдения достаточно велик. Отметим, что в силу выбора $\Gamma(i) = 1/i$ новые наблюдения все меньше влияют на оценку по мере увеличения i .

В методе «б» и новые и старые наблюдения берутся с одним весом. В тех случаях, когда параметры меняются медленно (отслеживание параметров), необходимо обеспечить постепенное забывание предыстории. Способы организации такого забывания обсуждаются ниже.

Так же как и в гл. 6, можно исследовать, что происходит в связи с различными упрощениями, особенно если *устраняется обращение матрицы*. Оказывается, что:

1) если $\Gamma(i) = 1/i$, то ковариация такая же, как в случае использования обращения матриц с незначительными отклонениями только для небольших значений ij (фиг. 7.5);

2) если $\Gamma(i) = c$, то при больших ij ковариация асимптотически приближается к ковариации, которая получается при использовании обращения матриц в соответствии с формулой (7.14).

Рекуррентный метод наименьших квадратов

Изменим обозначение матрицы входов \mathbf{U} на \mathbf{U}_k , чтобы подчеркнуть номер последнего наблюдения [21]:

$$\mathbf{U}_k = \begin{bmatrix} u_0(1) & \dots & u_m(1) \\ \vdots & & \vdots \\ u_0(k) & \dots & u_m(k) \end{bmatrix},$$

$$\mathbf{U}_{k+1} = \begin{bmatrix} u_0(1) & \dots & u_m(1) \\ \vdots & & \vdots \\ u_0(k) & \dots & u_m(k) \\ \hline u_0(k+1) & \dots & u_m(k+1) \end{bmatrix} = \begin{bmatrix} \mathbf{U}_k \\ \hline \mathbf{u}'(k+1) \end{bmatrix}. \quad (7.15)$$

Здесь $\mathbf{u}'(k+1)$ соответствует новому наблюдению. Определим

$$[\mathbf{P}(k)]^{-1} = \mathbf{U}'_k \mathbf{U}_k \text{ невырожденная,} \quad (7.16)$$

$$[\mathbf{P}(k+1)]^{-1} = \mathbf{U}'_{k+1} \mathbf{U}_{k+1} = [\mathbf{U}'_k | \mathbf{u}'(k+1)] \begin{bmatrix} \mathbf{U}_k \\ \hline \mathbf{u}'(k+1) \end{bmatrix},$$

или

$$[\mathbf{P}(k+1)]^{-1} = [\mathbf{P}(k)]^{-1} + \mathbf{u}'(k+1) \mathbf{u}'(k+1). \quad (7.17)$$

Отметим, что размер \mathbf{P} не зависит от номера k . Используя матричные тождества, приведенные в приложении В, получим

$$\begin{aligned} \mathbf{P}(k+1) &= \mathbf{P}(k) - \mathbf{P}(k) \mathbf{u}(k+1) \times \\ &\quad \times [\mathbf{u}'(k+1) \mathbf{P}(k) \mathbf{u}(k+1) + 1]^{-1} \mathbf{u}'(k+1) \mathbf{P}(k) = \\ &= \mathbf{P}(k) - \gamma(k) \mathbf{P}(k) \mathbf{u}(k+1) \mathbf{u}'(k+1) \mathbf{P}(k), \end{aligned} \quad (7.18)$$

где

$$\gamma(k) = \{\mathbf{u}'(k+1) \mathbf{P}(k) \mathbf{u}(k+1) + 1\}^{-1}$$

является скаляром. Вводя обозначение $\mathbf{y}'_k = [y(1), \dots, y(k)]$, находим с помощью формулы (6.31) оценку \mathbf{b} :

$$\hat{\beta}(k) = \mathbf{P}(k) \mathbf{U}'_k \mathbf{y}_k, \quad (7.19)$$

$$\begin{aligned} \hat{\beta}(k+1) &= \mathbf{P}(k+1) \mathbf{U}'_{k+1} \mathbf{y}_{k+1} = \\ &= \mathbf{P}(k+1) [\mathbf{U}'_k \mathbf{y}_k + \mathbf{u}'(k+1) y(k+1)]. \end{aligned} \quad (7.20)$$

Подстановка выражения (7.18) в (7.20) дает

$$\begin{aligned} \hat{\beta}(k+1) = & \mathbf{P}(k) \mathbf{U}'_k \mathbf{y}_k + \mathbf{P}(k) \mathbf{u}(k+1) \gamma(k) \gamma^{-1}(k) y(k+1) - \\ & - \mathbf{P}(k) \mathbf{u}(k+1) \gamma(k) \mathbf{u}'(k+1) \mathbf{P}(k) [\mathbf{U}'_k \mathbf{y}_k + \mathbf{u}(k+1) y(k+1)]. \end{aligned} \quad (7.21)$$

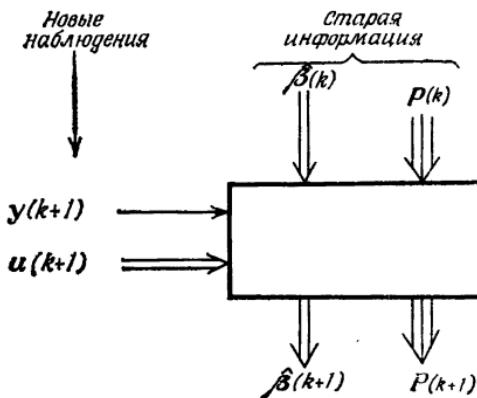
Здесь $\gamma(k) \gamma^{-1}(k) = 1$ введено для упрощения дальнейших выкладок. Используя равенство (7.19), можно преобразовать (7.21) к виду

$$\begin{aligned} \hat{\beta}(k+1) = & \hat{\beta}(k) + \gamma(k) \mathbf{P}(k) \mathbf{u}(k+1) \times \\ & \times \{y(k+1) - \mathbf{u}'(k+1) \hat{\beta}(k)\}, \end{aligned} \quad (7.22)$$

или

$$\begin{aligned} \hat{\beta}(k+1) = & \hat{\beta}(k) + \mathbf{P}(k+1) \mathbf{u}(k+1) \times \\ & \times \{y(k+1) - \mathbf{u}'(k+1) \hat{\beta}(k)\}. \end{aligned} \quad (7.23)$$

Эквивалентность двух последних соотношений можно показать непосредственной подстановкой $\mathbf{P}(k+1) \mathbf{u}(k+1)$ или используя матричные тождества из приложения В. Оценка, использующая новую пару наблюдений $\mathbf{u}(k+1)$ и $y(k+1)$, состоит из суммы старой оценки и линейного корректирующего члена. Заметим, что γ используется как нормирующий множитель. Выражение $\{y(k+1) - \mathbf{u}'(k+1) \hat{\beta}(k)\}$ соответствует той части последнего наблюдения выходного сигнала, которая не может быть объяснена наблюдением входного сигнала и последней оценкой параметров. Отметим, что в этом случае не нужно обращать матрицу, а для вычисления новой оценки используется только старая оценка и новые наблюдения. Таким образом, пара $(\hat{\beta}, \mathbf{P})$ представляет наименьшее число переменных, описывающих вход и выход объекта, которые необходимы в процессе вычислений. Это показано на фиг. 7.6. Новая итерация определяется уравнениями (7.18) и (7.22) или (7.23). На фиг. 7.7 схематично показаны операции, необходимые для вычисления новой итерации оценок. \mathbf{P}_k — симметричная матрица, и, следовательно, нужно обратить только один из ее треугольников. Отметим, что $\hat{\beta}(k+1)$, определяемая формулой (7.22) или (7.23), является оценкой



Ф и г. 7.6.

по методу наименьших квадратов; сходимость рекуррентного метода вытекает непосредственно из доказательства состоятельности оценок метода наименьших квадратов.

Формулу (7.22) можно записать также в виде

$$\hat{\beta}(k+1) = \{1 - \gamma(k)P(k)u(k+1)u'(k+1)\}\hat{\beta}(k) + \gamma(k)P(k)u(k+1)y(k+1), \quad (7.24)$$

т. е. таким образом, что новая оценка оказывается представленной в виде линейной комбинации 'старой' оценки и оценки, которая основывается только на новых наблюдениях. Весовые коэффициенты в этой линейной комбинации отражают «степень доверия» к различным ее составляющим.

Поскольку такое оценивание представляет собой просто рекуррентный вариант процедур с накоплением, рассмотренных в гл. 6, то

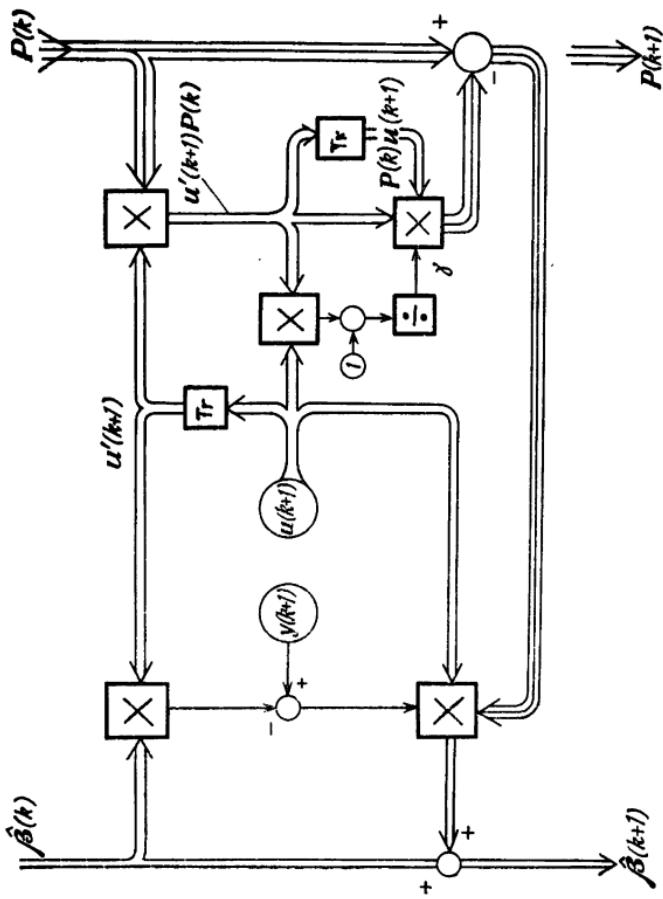
$$\mathcal{E}[\hat{\beta}(k)] = b, \quad (7.25)$$

$$\begin{aligned} \text{cov}[\hat{\beta}(k)] &= [\mathbf{U}_k' \mathbf{U}_k]^{-1} \mathbf{U}_k' \mathbf{N} \mathbf{U}_k [\mathbf{U}_k' \mathbf{U}_k]^{-1} = \\ &= P(k) \mathbf{U}_k' \mathbf{N} \mathbf{U}_k P(k). \end{aligned} \quad (7.26)$$

Если n — белый шум с ковариацией $\mathbf{N} = \sigma_n^2 \mathbf{I}$, то

$$\text{cov}[\hat{\beta}(k)] = \sigma_n^2 [\mathbf{U}_k' \mathbf{U}_k]^{-1} = \sigma_n^2 P(k). \quad (7.27)$$

→ Скаляр
 ↔ Вектор
 ⇢ Матрица



Ф и г. 7.7.

Следовательно, $\mathbf{P}(k)$ и $\mathbf{P}(k+1)$ пропорциональны ковариационным матрицам оценок в моменты k и $k+1$ соответственно.

Можно показать, что для многих типов входных сигналов матрица $\mathbf{P}(k)$ является положительно определенной и

$$\lim_{k \rightarrow \infty} \mathbf{P}(k) = 0\mathbf{I}.$$

Пример. Этот пример приводится для того, чтобы помочь читателю получить полное и ясное представление о значимости различных компонент новой оценки. Рассмотрим упрощенный вариант одномерной задачи:

$$\hat{\beta}(k+1) = \hat{\beta}(k) + P(k+1) u(k+1) \{y(k+1) - u(k+1) \hat{\beta}(k)\}, \quad (7.28)$$

$$[P(k+1)]^{-1} = [P(k)]^{-1} + u^2(k+1) = [P(0)]^{-1} + \sum_{j=1}^{k+1} u^2(j). \quad (7.29)$$

Если априорная информация о β невелика, то выбранное значение $[P(0)]^{-1}$ равно 0. Если u — стационарный белый шум с $E[u] = 0$ и $\text{var}[u] = \sigma_u^2 \mathbf{I}$, то

$$E[P(k+1)] = \frac{1}{(k+1)\sigma_u^2}. \quad (7.30)$$

Интересно сравнить полученный результат с условием сходимости по вероятности алгоритма стохастической аппроксимации (см. разд. 5.4).

В качестве упражнения рассмотрим следующий случай. Объект описывается уравнением

$$y(k) = bu(k) + n(k).$$

Выбрана модель, в которой используется значение оценки, найденное на предыдущем шаге:

$$w(k) = \beta(k-1)u(k).$$

Ошибка равна

$$e(k) = y(k) - w(k).$$

Для повторения полезно непосредственно вывести соотношения (7.28) и (7.29) (см. также разд. 5.1).

Читателю предлагается:

1) написать простую программу для ЦВМ, которая генерировала бы значения $u(k)$ и $n(k)$, например, в виде случайных чисел, моделировала бы объект и рекуррентно оценивала бы параметры

или

2) провести вычисления $u(k)$ и $n(k)$ вручную, используя таблицу случайных чисел.

Выберите $\beta(0)$, например $\beta(0) = b$. Полезно повторить вычисления несколько раз, выбирая каждый раз новые начальные условия. Графическое представление результатов экспериментов дает наглядное представление о скорости сходимости. При этом используются получающиеся значения $P(k)$ и $\mathcal{E}[P(k)]$.

Опираясь непосредственно на результаты этого раздела, можно получить следующие соотношения:

$$\hat{\beta}(k+1) = \hat{\beta}(k) + P(k+1) u(k+1) r_{k+1} \{y(k+1) - \\ - u'(k+1) \hat{\beta}(k)\}, \quad (7.31)$$

$$P(k+1) = P(k) - P(k) u(k+1) [u'(k+1) P(k) u(k+1) + \\ + r_{k+1}]^{-1} u(k+1) P(k), \quad (7.32)$$

где r_{k+1} — соответствующий элемент диагональной матрицы R из выражения для функции ошибок $e' Re$.

Более строгое изложение рекуррентного метода наименьших квадратов можно найти в работе [2].

Начальный этап процедуры оценивания. В основу начального этапа рекуррентной процедуры оценивания может быть положено несколько приемов [19]:

1) Использование всей априорной информации о β и P . Если известно, что входной сигнал u имеет вид стационарного белого шума с дисперсией σ_u^2 , то можно выбрать

$$P(l) = \mathcal{E}[U_l U_l]^{-1} = \frac{I}{l \sigma_u^2}.$$

Тогда рекуррентная процедура начинается с $(l+1)$ -го шага.

2) Использование l наблюдений для того, чтобы с помощью явного метода определить начальную оценку

$$\hat{\beta}(l) = \mathbf{P}(l) \mathbf{U}_l' \mathbf{y}_l,$$

$$\mathbf{P}(l) = [\mathbf{U}_l' \mathbf{U}_l]^{-1}.$$

После этого можно использовать рекуррентный метод. Недостатком этого способа является необходимость обращения матрицы.

3) Использование $\mathbf{P}(0) = a\mathbf{I}$, где a — очень большое действительное число, а $\hat{\beta}(0)$ выбирается произвольно. На первом шаге получаем

$$[\mathbf{P}(1)]^{-1} = [\mathbf{P}(0)]^{-1} + \mathbf{u}(1) \mathbf{u}'(1),$$

$$\begin{aligned}\hat{\beta}(1) &= \hat{\beta}(0) + \mathbf{P}(1) \mathbf{u}(1) \{y(1) - \mathbf{u}'(1) \hat{\beta}(0)\} = \\ &= \mathbf{P}(1) \{\mathbf{u}(1) y(1) + [\mathbf{P}(0)]^{-1} \hat{\beta}(0)\}\end{aligned}$$

[см. формулы (7.17) и (7.23)]. Точно так же находим

$$[\mathbf{P}(l)]^{-1} = [\mathbf{P}(0)]^{-1} + \mathbf{U}_l' \mathbf{U}_l,$$

$$\hat{\beta}(l) = \mathbf{P}(l) \{\mathbf{U}_l \mathbf{y}_l + [\mathbf{P}(0)]^{-1} \hat{\beta}(0)\}.$$

Если $[\mathbf{P}(0)]^{-1} = (1/a)\mathbf{I}$ и $a \rightarrow \infty$, то

$$[\mathbf{P}(l)]^{-1} \rightarrow \mathbf{U}_l' \mathbf{U}_l,$$

$$\hat{\beta}(l) \rightarrow \mathbf{P}(l) \mathbf{U}_l' \mathbf{y}_l,$$

т. е. имеем формулы, полученные явными методами. В этом случае можно использовать рекуррентную процедуру оценивания, начиная с первого наблюдения.

Экспоненциальное взвешивание прошлых наблюдений. Как уже отмечалось, приведенные алгоритмы представляют собой рекуррентный вариант процедур оценивания с накоплением данных (одношаговых). Это означает, что всем k выборочным значениям, которые используются при оценивании $\hat{\beta}(k)$, придаются равные веса независимо от «возраста» наблюдений. При оценивании (медленно) изменяющихся параметров это, естественно, нежелательно. Необходимо постепенно «забывать» предысторию. Математически это можно представить так:

$$\mathbf{U}_{k+1}' \mathbf{U}_{k+1} = [c \mathbf{U}_k]' [c \mathbf{U}_k] + \mathbf{u}(k+1) \mathbf{u}'(k+1),$$

или

$$c^{-2} [\mathbf{P}(k+1)]^{-1} = [\mathbf{P}(k)]^{-1} + \mathbf{u}(k+1) \times \\ \times c^{-2} \mathbf{u}'(k+1), \quad (7.33)$$

где $0 < c \leq 1$. При $\rho = c^2$ получаются следующие уравнения (см. приложение В):

$$\hat{\beta}(k+1) = \dot{\hat{\beta}}(k) + \gamma(k) \mathbf{P}(k) \mathbf{u}(k+1) \{ \mathbf{y}(k+1) - \\ - \mathbf{u}'(k+1) \hat{\beta}(k) \}, \quad (7.34)$$

$$\mathbf{P}(k+1) = \rho^{-1} [\mathbf{P}(k) - \gamma(k) \mathbf{P}(k) \mathbf{u}(k+1) \mathbf{u}'(k+1) \mathbf{P}(k)], \quad (7.35)$$

$$\gamma(k) = \{ \mathbf{u}'(k+1) \mathbf{P}(k) \mathbf{u}(k+1) + \rho \}^{-1}. \quad (7.36)$$

Замечая, что весовые коэффициенты, приписываемые прошлым наблюдениям, изменяются как функции j по закону c^{k-j} , $j \leq k$, довольно просто определить ту часть выборки, которая заметно влияет на величину оценки при каждом фиксированном c . В случае наблюдений с шумами величина c выбирается близкой к единице с тем, чтобы эффективная продолжительность наблюдений была как можно больше.

Стохастическая аппроксимация

Этот метод обсуждался в разд. 5.3. Алгоритм может иметь вид

$$\hat{\beta}(k+1) = \hat{\beta}(k) - \frac{1}{2} \Gamma(k) \nabla_{\beta} E, \quad (7.37)$$

где $\Gamma(k)$ должно удовлетворять следующим условиям:

$$\Gamma(k) \geq 0, \quad \sum_{k=0}^{\infty} \Gamma(k) = \infty, \quad \sum_{k=0}^{\infty} \Gamma^2(k) < \infty.$$

На каждом шаге выполняются соотношения

$$E = e^2(k+1), \\ e(k+1) = \mathbf{y}(k+1) - \mathbf{u}'(k+1) \hat{\beta}(k).$$

Следовательно, уравнение (7.37) можно переписать в виде

$$\hat{\beta}(k+1) = \hat{\beta}(k) + \Gamma(k) \mathbf{u}(k+1) \{ \mathbf{y}(k+1) - \\ - \mathbf{u}'(k+1) \hat{\beta}(k) \}. \quad (7.38)$$

Эта оценка сходится по вероятности к истинным значениям параметров (см. [1, 16]). Условиям на $\Gamma(k)$ удовлетворяет, например, следующая последовательность:

$$\Gamma(k) = \frac{c}{k^\alpha}, \quad c > 0, \quad 0.5 < \alpha \leq 1.$$

Сравнивая формулы (7.23) и (7.38), замечаем, что единственное отличие их состоит в том, что в формуле (7.38) вместо \mathbf{P}_{k+1} используется $\Gamma(k)$, т. е. матрица \mathbf{P} , тесно связанная с ковариационной матрицей оценки [см. формулу (7.27)], заменяется диагональной матрицей $\Gamma(k)\mathbf{I}$. Таким образом, стохастическая аппроксимация приводит к упрощенному алгоритму, который требует меньшего объема вычислений, поскольку матрицу \mathbf{P} считать не нужно. С другой стороны, если матрица \mathbf{P} отсутствует, метод стохастической аппроксимации не позволяет определить точность оценок. В смысле минимума дисперсии ошибки этот метод является субоптимальным по сравнению с алгоритмами (7.23) и (7.18). Приложениям метода стохастической аппроксимации посвящены, например, работы [13, 25, 30—32].

Принцип сжатых отражений

Метод синтеза рекуррентных алгоритмов предложен в работах [26, 27], где вместо того, чтобы для каждого k решать уравнение

$$\mathbf{U}'_k \mathbf{U}_k \hat{\beta}(k) = \mathbf{U}'_k \mathbf{y}_k$$

и доказывать, что $\hat{\beta}(k)$ удовлетворяет рекуррентному соотношению, вводится отображение

$$\mathcal{T}_k[\beta] = \beta - \gamma [\mathbf{U}'_k \mathbf{U}_k \beta - \mathbf{U}'_k \mathbf{y}_k]. \quad (7.39)$$

Здесь γ — вещественное число. Затем доказывается, что последовательность

$$\hat{\beta}(k+1) = \mathcal{T}_k[\hat{\beta}(k)] \quad (7.40)$$

при некоторых предположениях сходится при $k \rightarrow \infty$ к истинным значениям параметров. Когда алгоритм (7.39) применяется при решении задач, обычных для метода наименьших квадратов, он оказывается менее эффектив-

ным по сравнению с рекуррентным методом наименьших квадратов. Для того чтобы получить эффективный алгоритм, нужно представить γ в виде матрицы. Если

$$\gamma = [\mathbf{U}_k' \mathbf{U}_k]^{-1},$$

то алгоритм превращается в метод наименьших квадратов.

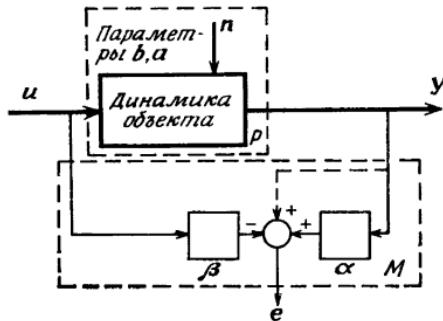
Этот метод можно применять и в более общих задачах. Он действительно был проверен при исследовании объекта, у которого как входной, так и выходной сигналы измеряются с ошибками при условии, что известна ковариационная функция ошибок измерений. Предположение о том, что известна ковариация ошибок измерений, существенно ограничивает практическую применимость метода.

7.2. РАСПРОСТРАНЕНИЕ НА ОБОБЩЕННЫЕ МОДЕЛИ И МОДЕЛИ В ПРОСТРАНСТВЕ СОСТОЯНИЙ

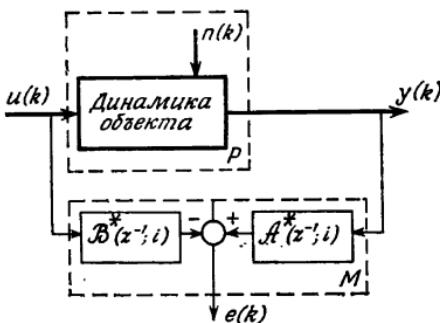
Рекуррентные схемы из предыдущего раздела можно применить и к обобщенной модели, определяемой уравнением (6.83) (фиг. 7.8):

$$\mathbf{e} = \mathbf{y} + \mathbf{Y}\alpha - \mathbf{U}\beta = \mathbf{y} - [\mathbf{U}' \mathbf{Y}] \begin{bmatrix} \beta \\ -\alpha \end{bmatrix} \triangleq \mathbf{y} - \mathbf{Q}\theta. \quad (7.41)$$

Эта модель является линейной по параметрам. Конечно, и здесь возникают те же проблемы, что и при использовании явных методов, а именно смещенность оценок параметров из-за искажения шумами \mathbf{Y} (см. разд. 6.5). Такую



Ф и г. 7.8.



Ф и г. 7.9.

обобщенную модель можно задать дискретной весовой функцией (фиг. 7.9). В этом случае

$$\begin{aligned} \mathcal{A}^*(z^{-1}; i) &= 1 + \alpha_1(i) z^{-1} + \dots + \alpha_m(i) z^{-m}, \\ \mathcal{B}^*(z^{-1}; i) &= \beta_1(i) z^{-1} + \dots + \beta_m(i) z^{-m}. \end{aligned} \quad (7.42)$$

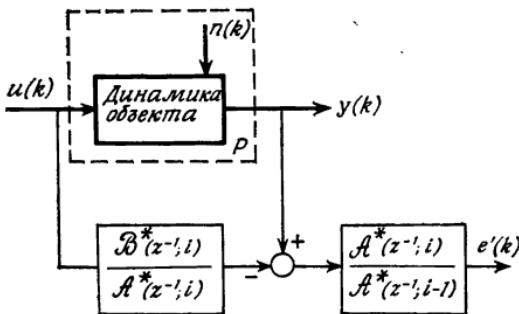
Здесь индекс i относится к значениям параметров на i -й итерации. Для этого случая Калман [18] предложил простую схему оценивания, которая ведет к асимптотически смещенным оценкам.

Способы подавления искажений, возникающих из-за аддитивных помех, перечислены в гл. 6. Обсуждение одного из этих способов, связанного с применением обобщенных и близких к ним моделей, было отложено до настоящей главы, поскольку он основан на использовании моделей с настраиваемыми параметрами. Будут рассматриваться следующие подходы:

- 1) последовательный линейный регрессионный анализ и фильтрация;
- 2) обобщенный метод наименьших квадратов;
- 3) расширенный матричный метод.

Последовательный линейный регрессионный анализ и фильтрация. Вместо того чтобы задавать обобщенную ошибку в виде (фиг. 7.9)

$$e(k) = \mathcal{A}^*(z^{-1}; i) y(k) - \mathcal{B}^*(z^{-1}; i) u(k), \quad (7.43)$$



Ф и г. 7.10.

можно, как это предлагается в [34], записать ошибку следующим образом:

$$\begin{aligned} e'(k) &= \frac{e(k)}{\mathcal{A}^*(z^{-1}; i-1)} = \frac{\mathcal{A}^*(z^{-1}; i)}{\mathcal{A}^*(z^{-1}; i-1)} y(k) - \\ &- \frac{\mathcal{B}^*(z^{-1}; i)}{\mathcal{A}^*(z^{-1}; i-1)} u(k) = \left\{ y(k) - \frac{\mathcal{B}^*(z^{-1}; i)}{\mathcal{A}^*(z^{-1}; i)} u(k) \right\} \times \\ &\times \frac{\mathcal{A}^*(z^{-1}; i)}{\mathcal{A}^*(z^{-1}; i-1)}. \end{aligned} \quad (7.44)$$

Это также обобщенная модель, линейная по параметрам $\alpha(i)$ и $\beta(i)$, однако выражение, заключенное в фигурные скобки, соответствует обычной модели (фиг. 7.10). Для дальнейшего улучшения оценок предлагается по достижении определенной сходимости произвести переключение с обобщенной модели на обычную модель:

$$e(k) = y(k) - \frac{\mathcal{B}^*(z^{-1}; i)}{\mathcal{A}^*(z^{-1}; i)} u(k). \quad (7.45)$$

Достоинство этого приема состоит в том, что аддитивная помеха не вызывает смещения оценок. В окрестности оптимальной точки нелинейность по параметрам \mathcal{A}^* едва ли скажется на сходимости. Нелинейность по параметрам также можно преодолеть с помощью квазилинеаризации [33].

Итеративная процедура имеет следующий вид. На первом шаге подсчитывается линейная регрессия

$$\hat{\theta}(1) = \left[\begin{array}{c} \hat{\beta}(1) \\ -\hat{\alpha}(1) \end{array} \right] = [\Omega' \Omega]^{-1} \Omega' y, \quad (7.46)$$

где $\alpha(1)$ и $\beta(1)$ являются параметрами $\mathcal{A}^*(z^{-1}; i)$ и $\mathcal{B}^*(z^{-1}; i)$. На следующем шаге входной и выходной сигналы подаются на фильтр $\mathcal{A}^{*-1}(z^{-1}; i)$. По отфильтрованным сигналам находится оценка $\hat{\theta}(2)$ и т. д. Доказательство сходимости этой процедуры неизвестно.

Обобщенный метод наименьших квадратов (марковские оценки). Основная идея состоит в следующем. Пусть объект описывается уравнением

$$A^*(z^{-1})y(k) = B^*(z^{-1})u(k) + n(k), \quad (7.47)$$

где A^* и B^* — полиномы, а $\{n(k)\}$ — последовательность коррелированных случайных величин. Предположим, что корреляция последовательности $\{n(k)\}$ известна. Например, $n(k)$ можно записать как

$$n(k) = G^*(z^{-1})w(k), \quad (7.48)$$

где $\{w(k)\}$ — последовательность некоррелированных случайных величин, а G^* — дискретная весовая функция. Тогда уравнение объекта можно записать в виде

$$A^*(z^{-1})y(k) = B^*(z^{-1})u(k) + G^*(z^{-1})w(k), \quad (7.49)$$

где A^* и B^* описывают объект, а G^* — свойства «внешней среды», или

$$A^*(z^{-1})\tilde{y}(k) = B^*(z^{-1})\tilde{u}(k) + w(k), \quad (7.50)$$

где

$$\tilde{y}(k) = \frac{1}{G^*(z^{-1})}y(k) \text{ и } \tilde{u}(k) = \frac{1}{G^*(z^{-1})}u(k)^1.$$

Таким образом если \tilde{u} и \tilde{y} рассматривать как входные и выходные сигналы, то имеет место обычное оценивание по методу наименьших квадратов. [Ср. с формулой (6.110).] В результате выясняется, что обобщенный метод наименьших квадратов можно интерпретировать как идентифика-

¹⁾ Заметим, что если число членов $G^*(z^{-1})$ невелико, то число существенных членов $\{G^*(z^{-1})\}^{-1}$ может быть очень большим. Отсюда следует, что для получения \tilde{u} требуется большая выборка значений u .

цию по методу наименьших квадратов с критерием вида

$$\min \sum e^2(k),$$

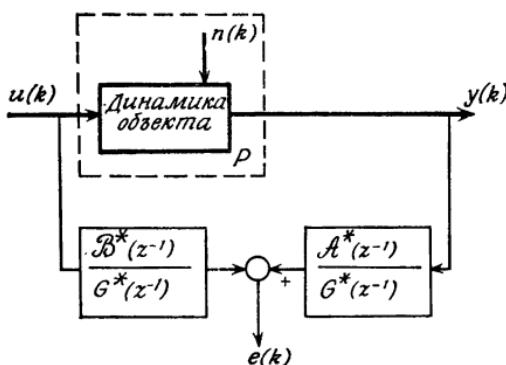
где обобщенная ошибка определяется как

$$\begin{aligned} e(k) &= \frac{\mathcal{A}^*(z^{-1})}{G^*(z^{-1})} y(k) - \frac{\mathcal{B}^*(z^{-1})}{G^*(z^{-1})} u(k) = \\ &= \mathcal{A}^*(z^{-1}) \left[\frac{y(k)}{G^*(z^{-1})} \right] - \mathcal{B}^*(z^{-1}) \left[\frac{u(k)}{G^*(z^{-1})} \right]. \end{aligned} \quad (7.51)$$

Сравните с блок-схемой, изображенной на фиг. 7.11. Это показывает, что обобщенную ошибку можно получить по входному и выходному сигналам объекта и параметрам модели α и β , которые определяются в рамках обобщенного метода наименьших квадратов.

На практике корреляция между невязками и дискретной весовой функцией G^* редко бывает известна. В работе [11] предложена рекуррентная процедура определения G^* , которая проверена как на модельных примерах, так и на практической задаче (идентификация ректификационной колонны). Эта процедура включает следующие этапы (для того чтобы подчеркнуть связь между представлениями с помощью дискретных весовых функций и с использованием матричных обозначений, они приводятся параллельно):

1) Используется обычный метод наименьших квадратов для подгонки входных и выходных последовательно-



Фиг. 7.11.

стей объекта и модели

модель $\mathcal{A}^*(z^{-1}; i) y(k) =$ $= \mathcal{B}^*(z^{-1}; i) u(k) + r(k).$	модель $\mathbf{e} = \mathbf{y} - \Omega \theta(i),$ оценка $\hat{\theta}(i) = [\Omega' \Omega]^{-1} \Omega' \mathbf{y}.$
---	--

2) Теперь последовательность значений ошибок является аппроксимацией невязок. Анализируются эти невязки и строится модель авторегрессии:

определяется $\mathcal{D}^*(z^{-1}; i) = 1 + \delta_1(i) z^{-1} + \delta_2(i) z^{-2} + \dots$ так, чтобы $\mathcal{D}^*(z^{-1}; i) \hat{r}(k) = \xi(k),$ где $\xi(k)$ — дискретный белый шум.	невязки $\mathbf{e} = \hat{\mathbf{r}} = \mathbf{y} - \hat{\Omega} \hat{\theta}(i),$ модель $\hat{\mathbf{r}} = -\hat{\mathbf{R}} \hat{\boldsymbol{\delta}}(i) + \xi,$ оценка $\hat{\boldsymbol{\delta}}(i) =$ $= -[\hat{\mathbf{R}}' \hat{\mathbf{R}}]^{-1} \hat{\mathbf{R}}' \hat{\mathbf{r}}.$ (см. разд. 6.2).
---	--

3) Фильтруется входная и выходная последовательности объекта:

$\tilde{u}(k) = \mathcal{D}^*(z^{-1}; i) u(k),$ $\tilde{y}(k) = \mathcal{D}^*(z^{-1}; i) y(k).$	$\tilde{\mathbf{u}} = \mathbf{u} + \mathbf{U} \hat{\boldsymbol{\delta}}(i),$ $\tilde{\mathbf{y}} = \mathbf{y} + \mathbf{Y} \hat{\boldsymbol{\delta}}(i).$
--	--

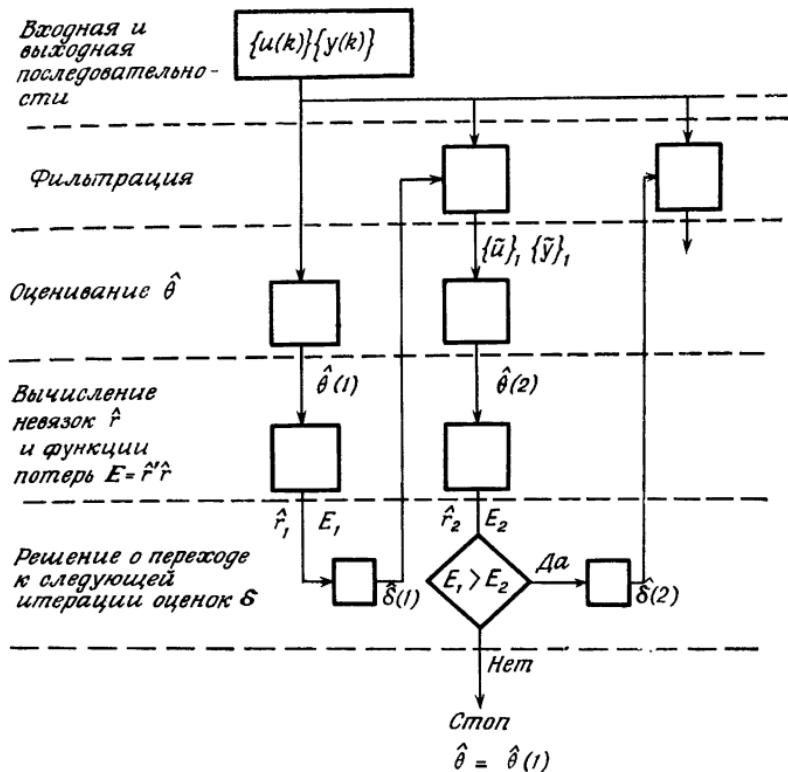
4) Снова используется метод наименьших квадратов для подгонки модели к отфильтрованным входным и выходным последовательностям объекта, т. е. повторяется первый этап.

5) Повторяется второй этап.

Эта рекуррентная процедура показана на фиг. 7.12. Правило остановки определяется сравнением относительных значений функции потерь.

Если $\mathcal{A}^*(i) \rightarrow A^*$, $\mathcal{B}^*(i) \rightarrow B^*$ и $\mathcal{D}^*(i) \rightarrow \{G^*\}^{-1}$, то получаем

$$\begin{aligned}
 A^*(z^{-1}) \{G^*(z^{-1})\}^{-1} y(k) &= \\
 &= B^*(z^{-1}) \{G^*(z^{-1})\}^{-1} u(k) + w(k),
 \end{aligned} \tag{7.52}$$



Ф и г. 7.12.

т. е. модель с некоррелированными невязками. По-видимому, к настоящему времени еще не существует каких-либо доказательств сходимости. Показано, что для простых примеров эта процедура приводит к хорошим оценкам при разумно выбранном порядке объекта. Недостаток процедуры состоит в том, что (пока еще) нет стандартных правил выбора порядка модели и порядка модели авторегрессии.

Описанная схема не является рекуррентной по данным о входных и выходных сигналах объекта; вся последовательность наблюдений обрабатывается за один шаг¹⁾. Рекуррентный вариант этой процедуры описан в работе [14]. В этом случае оценки θ и δ подстраиваются по

¹⁾ Поэтому, вообще говоря, этот метод мог бы обсуждаться в гл. 6. В эту главу он включен ради цельности изложения.

мере поступления новых наблюдений. Сначала оценки δ получаются плохими, и поэтому результаты фильтрации выборочных последовательностей $u(k)$ и $y(k)$ не являются оптимальными для оценки θ . Следовательно, для того чтобы уменьшить влияние старых наблюдений на результаты оценивания, необходимо ввести какое-то взвешивание. Поскольку в предыдущем разделе рассматривалось экспоненциальное взвешивание данных о предыстории объекта, здесь приводится лишь сводка алгоритмов для рекуррентных схем оценивания:

$$\Omega_k = [\mathbf{U}_k \mid \mathbf{Y}_k], \\ \Omega_{k+1} = \left[\begin{array}{c|c} \Omega_k & \\ \hline \omega'_{k+1} & \end{array} \right] = \left[\begin{array}{c|c} \mathbf{U}_k \mid \mathbf{Y}_k & \\ \hline \mathbf{u}'_{k+1} \mid \mathbf{y}'_{k+1} & \end{array} \right], \quad (7.53)$$

$$\tilde{y}(k+1) = y(k+1) + \mathbf{y}'(k) \hat{\delta}(k), \\ \tilde{\omega}(k+1) = \omega(k+1) + \omega'(k) \hat{\delta}(k), \\ \tilde{\omega}(k+1) = \omega(k+1) + \Omega_k \hat{\delta}(k), \quad (7.54)$$

$$\hat{r}(k+1) = y(k+1) - \omega'(k+1) \hat{\theta}(k), \\ \hat{\mathbf{r}}'(k+1) = [\hat{r}(k), \hat{r}(k-1), \dots, \hat{r}(k-p-1)], \quad (7.55)$$

$$\left\{ \begin{array}{l} \hat{\theta}(k+1) = \hat{\theta}(k) + \gamma(k) \mathbf{P}(k) \tilde{\omega}(k+1) \{ \tilde{y}(k+1) - \tilde{\omega}'(k+1) \hat{\theta}(k) \}, \\ \mathbf{P}(k+1) = \rho^{-1} [\mathbf{P}_k(k) - \gamma(k) \mathbf{P}(k) \tilde{\omega}(k+1) \tilde{\omega}'(k+1) \mathbf{P}(k)], \end{array} \right. \quad (7.56)$$

$$\left\{ \begin{array}{l} \gamma(k) = \{ \tilde{\omega}'(k+1) \mathbf{P}(k) \tilde{\omega}(k+1) + \rho \}^{-1}, \\ \hat{\delta}(k+1) = \hat{\delta}(k) + \lambda(k) \mathbf{Q}(k) \hat{\mathbf{r}}(k+1) \{ \hat{r}(k+1) - \hat{\mathbf{r}}'(k+1) \hat{\delta}(k) \}, \\ \mathbf{Q}(k+1) = \nu^{-1} [\mathbf{Q}(k) - \lambda(k) \mathbf{Q}(k) \hat{\mathbf{r}}(k+1) \hat{\mathbf{r}}'(k+1) \mathbf{Q}(k)], \\ \lambda(k) = \{ \hat{\mathbf{r}}'(k+1) \mathbf{Q}(k) \hat{\mathbf{r}}(k+1) + \nu \}^{-1}. \end{array} \right. \quad (7.57)$$

Необходимо выбрать весовые коэффициенты ρ и ν . В начале процедуры оценивания коэффициенты могут быть небольшими, например 0,975, с тем, чтобы обеспечивалась хорошая скорость сходимости; в дальнейшем ρ и ν можно

приблизить к 1,0 для того, чтобы получить хорошую точность (небольшую дисперсию). Пока еще не существует общей теории оптимального выбора ρ и v . Некоторые результаты моделирования приводятся в разд. 7.3.

Расширенный матричный метод. Предыдущий метод основывался на предположении, что шум можно представить как

$$r(k) = \frac{1}{D^*(z^{-1})} w(k), \quad D^*(z^{-1}) = 1 + d_1 z^{-1} + d_2 z^{-2} + \dots, \quad (7.58)$$

где $\{w(k)\}$ — последовательность некоррелированных случайных величин (дискретный белый шум). В зависимости от ситуации число параметров, характеризующих шум, может быть довольно большим. Поэтому минимальное число параметров при таком описании шума можно было бы получить, используя соотношения

$$\begin{aligned} r(k) &= \frac{C^*(z^{-1})}{D^*(z^{-1})} w(k), \\ C^*(z^{-1}) &= 1 + c_1 z^{-1} + c_2 z^{-2} + \dots, \\ D^*(z^{-1}) &= 1 + d_1 z^{-1} + d_2 z^{-2} + \dots. \end{aligned} \quad (7.59)$$

Таким образом, важно построить такую схему оценивания, в которой внешняя среда описывалась бы оценками параметров c и d (фиг. 7.13).

Объект и среда описываются следующими уравнениями:

$$\sum_{i=0}^n a_i x(k-i) = \sum_{i=1}^m b_i u(k-i) — динамика объекта, \quad (7.60)$$

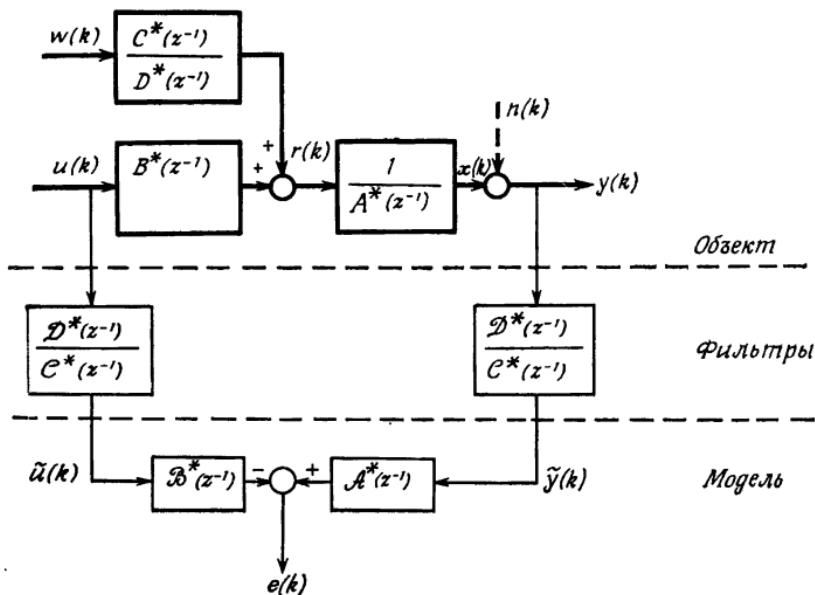
$$y(k) = x(k) + n(k) — выход объекта.$$

Отсюда

$$\sum_{i=0}^n a_i y(k-i) - \sum_{i=1}^m b_i u(k-i) = \sum_{i=0}^n a_i n(k-i) = r(k), \quad (7.61)$$

где шум на выходе описывается последовательностью $\{r(k)\}$, а $r(k)$ связан с дискретным белым шумом $w(k)$ уравнением вида

$$r(k) + \sum_{i=1}^p d_i r(k-i) = w(k) + \sum_{i=1}^q c_i w(k-i). \quad (7.62)$$



Ф и г. 7.13.

Полагая снова $a_0 = 1$, находим

$$y(k) = \sum_{i=1}^m b_i u(k-i) - \sum_{i=1}^n a_i y(k-i) + \\ + \sum_{i=1}^q c_i w(k-i) - \sum_{i=1}^p d_i r(k-i) + w(k) \quad (7.63)$$

и так же, как раньше, получаем

$$\mathbf{y} = [\mathbf{U} \mid \mathbf{Y} \mid \mathbf{W} \mid \mathbf{R}] \begin{bmatrix} \frac{b}{-a} \\ \frac{-c}{d} \end{bmatrix} + \mathbf{w}. \quad (7.64)$$

Это снова модель, линейная по параметрам объекта и шума с аддитивной помехой в виде белого шума в правой части. Как и раньше, состоятельная оценка имеет вид

$$\begin{bmatrix} \beta \\ -\alpha \\ -\gamma \\ -\delta \end{bmatrix} = [(\mathbf{U} \mid \mathbf{Y} \mid \mathbf{W} \mid \mathbf{R})' (\mathbf{U} \mid \mathbf{Y} \mid \mathbf{W} \mid \mathbf{R})]^{-1} (\mathbf{U} \mid \mathbf{Y} \mid \mathbf{W} \mid \mathbf{R})' \mathbf{y}. \quad (7.68)$$

Так как элементы матриц R и W неизвестны и не могут быть измерены непосредственно, их нужно заменить оценками \hat{r} и \hat{w} :

$$\begin{aligned}\hat{r}(k) &= y(k) + \sum_{i=1}^n \alpha_i y(k-i) - \sum_{i=1}^m \beta_i u(k-i), \\ \hat{w}(k) &= \hat{r}(k) + \sum_{i=1}^p \delta_i \hat{r}(k-i) - \sum_{i=1}^q \gamma_i \hat{w}(k-i).\end{aligned}\quad (7.66)$$

Таким образом, на каждой итерации вычисляются элементы матриц \hat{R} и \hat{W} . В начале процедуры оценивания точность невысока, поэтому для более полного учета информации о прошлых наблюдениях нужно снова ввести весовые коэффициенты. Этот алгоритм также можно записать в рекуррентной форме. Например, если определить

$$\begin{aligned}\hat{\theta}'(k+1) &= [\beta'(k+1) | -\alpha'(k+1) | \gamma'(k+1) | -\delta'(k+1)], \\ (7.67)\end{aligned}$$

$$\omega'(k+1) = [u'(k+1) | y'(k+1) | \hat{w}'(k+1) | \hat{r}'(k+1)],$$

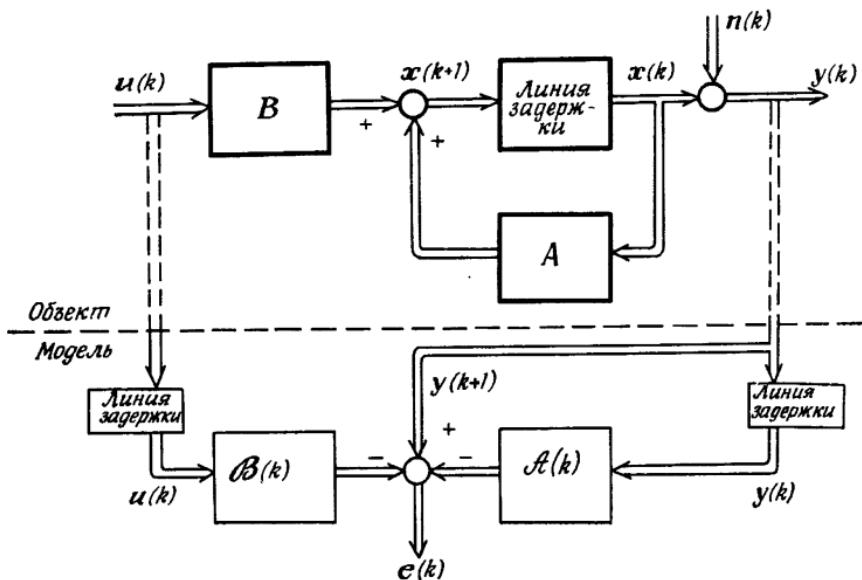
то можно применить алгоритм (7.56). Этот метод рассмотрен в работе [35] для случая $\gamma = 0$, в работе [42] — для случая $\delta = 0$, а в [36] — для общего случая. Именно из последнего источника заимствованы результаты моделирования, которые приводятся в разд. 7.3.

Другой способ борьбы со смещенностю оценок в случае обобщенных моделей сводится к использованию метода максимального правдоподобия (см. гл. 11).

Метод вспомогательных переменных (гл. 6) можно легко превратить в рекуррентный метод для работы в натуральном масштабе времени [29, 43]. Приближенный вариант метода максимального правдоподобия для работы в замкнутом контуре предложен в работе [28]. Еще одно обсуждение подобных методов можно найти в работе [20].

Модели в пространстве состояний. Изложенные методы можно применять и для оценивания параметров уравнений объекта в пространстве состояний. В работе [15] рассматривается задача оценивания матрицы A в уравнениях вида

$$\begin{aligned}x(k+1) &= Ax(k), \\ y(k) &= x(k) + n(k).\end{aligned}\quad (7.68)$$



Ф и г. 7.14.

Рассмотрим ситуацию, когда измеряется и входной сигнал (фиг. 7.14). В этом случае уравнения объекта имеют вид

$$\begin{aligned} \mathbf{x}(k+1) &= \mathbf{A}\mathbf{x}(k) + \mathbf{B}\mathbf{u}(k), \\ \mathbf{y}(k) &= \mathbf{x}(k) + \mathbf{n}(k); \end{aligned} \quad (7.69)$$

уравнение модели

$$\mathbf{e}(k) = \mathbf{y}(k+1) - \mathbf{A}(k)\mathbf{y}(k) - \mathbf{B}(k)\mathbf{u}(k). \quad (7.70)$$

Минимизация функции $E = \mathbf{e}'(k)\mathbf{e}(k)$ приводит к уравнениям вида

$$\begin{aligned} \frac{\partial E}{\partial \mathbf{A}(k)} &= -2\mathbf{e}(k)\mathbf{y}'(k), \\ \frac{\partial E}{\partial \mathbf{B}(k)} &= -2\mathbf{e}(k)\mathbf{u}'(k). \end{aligned} \quad (7.71)$$

Таким образом, итеративную процедуру можно записать как

$$\begin{aligned} \mathbf{A}(k+1) &= \mathbf{A}(k) + \Gamma\mathbf{e}(k)\mathbf{y}'(k), \\ \mathbf{B}(k+1) &= \mathbf{B}(k) + \Gamma\mathbf{e}(k)\mathbf{u}'(k). \end{aligned} \quad (7.72)$$

Коэффициент Γ снова можно выбрать:

- 1) постоянным, если A и B изменяются во времени;
- 2) равным $1/k$, если A и B постоянны; случай стохастической аппроксимации.

Из-за влияния аддитивных помех оценки будут смещены.

В работе [3] описывается процедура получения постоянных оценок элементов матрицы A , которая входит в систему уравнений вида

$$\mathbf{x}(k+1) = A\mathbf{x}(k) + \mathbf{v}(k), \quad \mathbf{y}(k) = \mathbf{x}(k) + \mathbf{n}(k), \quad (773).$$

где $\mathbf{v}(k)$ и $\mathbf{n}(k)$ — белые шумы.

7.3. МАШИННЫЕ ПРОЦЕДУРЫ И РЕЗУЛЬТАТЫ ВЫЧИСЛЕНИЙ

Так как методы, рассмотренные в разд. 7.2, в практическом плане являются многообещающими, для иллюстрации их возможностей предлагается ряд примеров. Эти примеры заимствованы из работы [36].

Объект. Во всех случаях объект описывается дискретной передаточной функцией

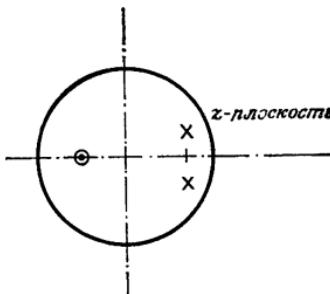
$$H(z^{-1}) = \frac{B^*(z^{-1})}{A^*(z^{-1})} = \frac{z^{-1} + 0,5z^{-2}}{1 - 1,5z^{-1} + 0,7z^{-2}}, \quad (7.74)$$

и, следовательно,

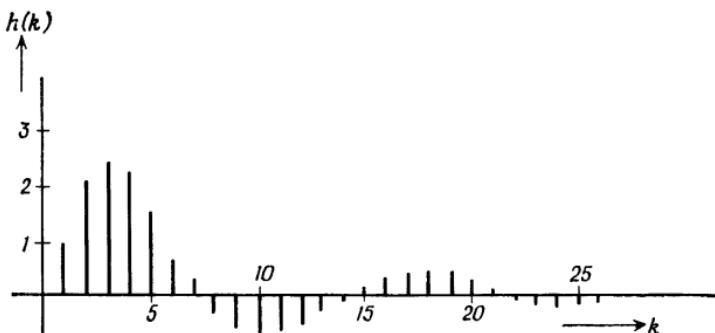
$$[b' | -a'] = [0; \ 1; \ 0,5; \ 1,5; \ -0,7].$$

$$b_0 \ b_1 \ b_2 \ -a_1 \ -a_2$$

Объект имеет простые полюсы в точках $z = 0,75 \pm 0,36j$ и нуль в точке $z = -0,5$. Диаграмма нулей и полюсов



Фиг. 7.15.



Ф и г. 7.16.

приведена на фиг. 7.15, а весовая функция — на фиг. 7.16. Во всех примерах входной сигнал является белым шумом с равномерным распределением на интервале $[-1, 1]$.

Если мощность белого шума на входе объекта σ_u^2 , то мощность выходного сигнала σ_x^2 определяется выражением [17]

$$\sigma_x^2 = \frac{\sigma_u^2}{2\pi j} \int_{|z|=1} H(z) H(z^{-1}) \frac{dz}{z}. \quad (7.75)$$

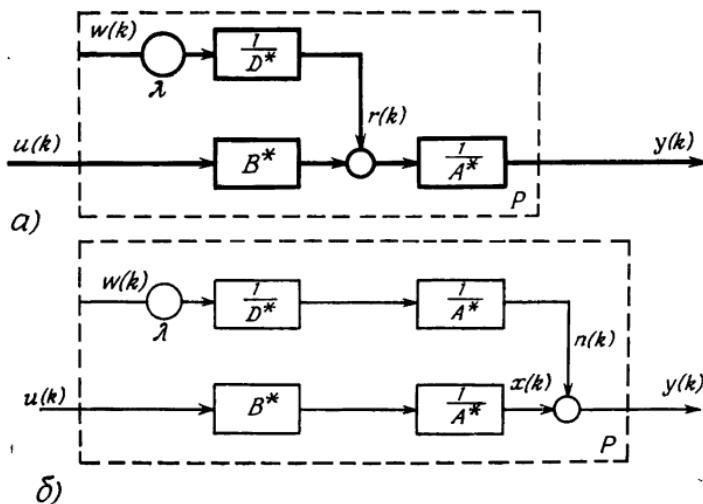
В работе [5] предлагается способ быстрого вычисления этого интеграла. Для рассматриваемого объекта отношение $\sigma_x^2/\sigma_u^2 = 18,8$. Результаты, которые приводятся в следующих частях раздела, усреднены по десяти итеративным циклам (1000 итераций в каждом цикле). Кроме того, приводятся средние значения стандартных отклонений по этим десяти циклам.

Замечания

1) Во всех таблицах, если не оговорено противное, сначала приводятся усредненные по 10 циклам значения оценок параметров. Затем даются усредненные значения стандартных отклонений.

2) Рассмотрим случайный процесс $z(k) = \mu + x(k)$, где $\mu = \text{const}$, а $x(k)$ — случайная величина с нулевым математическим ожиданием. Оценка μ определяется формулой

$$\bar{z} = \frac{1}{n} \sum_{i=1}^N z(i).$$



Ф и г. 7.17.

Доверительный интервал для оценки μ запишется как

$$\bar{z} - \frac{t_v \left(\frac{1}{2} \alpha \right) s}{\sqrt{N}} < \mu < \bar{z} + \frac{t_v \left(\frac{1}{2} \alpha \right) s}{\sqrt{N}}, \quad (7.76)$$

где s^2 — оценка дисперсии z , а v — число степеней свободы. В рассматриваемом случае $z(i)$ представляет собой оценку параметров в i -м цикле. Так как циклов 10, то $v = 9$. При построении доверительного интервала с 95%-м доверительным уровнем $t_9(1/2 \alpha) = 2,26$. Таким образом, получаем

$$\bar{z} - 0,7 \text{ s} < \mu < \bar{z} + 0,7 \text{ s}.$$

3) Программы написаны на языке АЛГОЛ-60.

Рекуррентный обобщенный метод наименьших квадратов. Используется алгоритм, задаваемый формулами (7.56) и (7.57). Разностные уравнения, описывающие объект и возмущения, имеют вид (фиг. 7.17, а и б)

$$y(k) = \sum_{i=0}^2 b_i u(k-i) - \sum_{i=1}^2 a_i y(k-i) - \sum_{i=1}^2 d_i r(k-i) + \lambda w(k), \quad (7.77)$$

$$\begin{aligned} b_0 &= 0, & b_1 &= 1,0, & b_2 &= 0,5, \\ a_1 &= -1,5, & a_2 &= 0,7, \\ d_1 &= -1,0, & d_2 &= 0,2. \end{aligned}$$

$\{w(k)\}$ — дискретный белый шум, величина которого распределена между -1 и $+1$. Этот белый шум «окрашивается» в соответствии с передаточной функцией, имеющей вид

$$\frac{1}{D^*(z^{-1})} = \frac{1}{1 - z^{-1} + 0,2z^{-2}}, \quad (7.78)$$

с полюсами в точках $z = 0,725$ и $z = 0,275$. В этом случае $\sigma_n^2/\sigma_w^2 = 98,52$.

Оцениваются пять параметров объекта и два параметра шума. После каждого оценивания d вычисляются новые значения вектора $\tilde{\omega}(k+1)$ и величины $\tilde{y}(k+1)$.

После каждого оценивания $\hat{\theta}$ вычисляются новые значения вектора $\hat{r}(k+1)$ и величины $\hat{r}(k+1)$. Экспериментальные результаты приводятся в указанных ниже таблицах.

Номер таблицы	λ	Отношение сигнал/шум на выходе объекта $\frac{S}{N} = \frac{\sigma_x^2}{\sigma_n^2}$	Весовой коэффициент $\rho = v$	Число выборочных значений, начиная с которых влияние ошибок на выходе модели снижается до 1%
7.1	0,25	3,2	+5 дБ	0,9913
7.2	0,5	0,8	-1 дБ	0,9913
7.3	1	0,2	-7 дБ	0,9913
7.4	2	0,05	-13 дБ	0,9913
7.5	4	0,0125	-19 дБ	0,9913
7.6	4	0,0125	-19 дБ	1,0
7.7	4	0,0125	-19 дБ	0,995
7.8	4	0,0125	-19 дБ	0,990
7.9	4	0,0125	-19 дБ	0,985
7.10	4	0,0125	-19 дБ	0,980
7.11	4	0,0125	-19 дБ	0,975

Таким образом, по экспериментальным данным можно судить о:

- влиянии уровня помех, табл. 7.1—7.5;
- влиянии выбора весовых коэффициентов, табл. 7.6—7.11.

Таблица 7.1

$$\frac{S}{N} = 3,2 (+5 \text{ дБ}), \rho = v = 0,9913$$

Число итераций	α_1	α_2	β_0	β_1	β_2	δ_1	δ_2
100	-1,556 0,056	0,752 0,082	0,049 0,088	1,077 0,098	0,521 0,049	-0,995 0,162	0,270 0,157
200	-1,529 0,056	0,730 0,079	0,038 0,076	1,048 0,082	0,514 0,022	-1,060 0,113	0,284 0,131
300	-1,519 0,042	0,720 0,061	0,020 0,067	1,030 0,069	0,505 0,013	-1,068 0,092	0,286 0,086
400	-1,514 0,028	0,715 0,038	0,025 0,048	1,028 0,046	0,504 0,012	-1,003 0,066	0,235 0,090
500	-1,509 0,017	0,708 0,024	0,014 0,030	1,015 0,034	0,501 0,017	-0,971 0,101	0,190 0,080
600	-1,499 0,017	0,700 0,020	-0,001 0,024	1,004 0,027	0,504 0,018	-0,985 0,076	0,201 0,061
700	-1,502 0,012	0,703 0,017	0,001 0,016	1,008 0,017	0,501 0,016	-0,950 0,077	0,172 0,058
800	-1,498 0,022	0,702 0,022	-0,011 0,024	0,990 0,031	0,489 0,020	-0,960 0,057	0,158 0,052
900	-1,498 0,017	0,708 0,019	-0,003 0,017	0,998 0,018	0,493 0,018	-0,997 0,066	0,204 0,052
1000	-1,502 0,016	0,702 0,016	0,001 0,016	1,007 -0,018	0,498 0,016	-0,985 0,090	0,196 0,087

Таблица 7.2

$$\frac{S}{N} = 0,8 (-1 \text{ дБ}), \rho = v = 0,9913$$

число итераций	α_1	α_2	β_0	β_1	β_2	δ_1	δ_2
100	-1,565 0,056	0,722 0,056	0,022 0,054	1,053 0,085	0,510 0,080	-0,925 0,132	0,244 0,152
.
1000	-1,504 0,028	0,704 0,028	0,011 0,031	1,014 0,038	0,496 0,035	-0,983 0,083	0,196 0,080

Таблица 7.3

$$\frac{S}{N} = 0,2 \text{ (-7 dB)}, \rho = v = 0,9913$$

Число итераций	α_1	α_2	β_0	β_1	β_2	δ_1	δ_2
100	-1,624 0,047	0,763 0,053	0,035 0,114	1,105 0,190	0,507 0,171	-0,849 0,102	0,232 0,158
.
1000	-1,517 0,039	0,705 0,041	0,007 0,068	1,038 0,081	0,496 0,079	-0,969 0,082	0,208 0,082

Таблица 7.4

$$\frac{S}{N} = 0,05 \text{ (-13 dB) (усредненное по 9 циклам)}, \rho = v = 0,9913$$

Число итераций	α_1	α_2	β_0	β_1	β_2	δ_1	δ_2
100	-1,672 0,049	0,806 0,042	0,043 0,225	1,106 0,226	0,456 0,258	-0,805 0,096	0,245 0,167
.
1000	-1,564 0,041	0,722 0,042	0,036 0,116	1,098 0,140	0,490 0,149	-0,926 0,069	0,243 0,091

Таблица 7.5

$$\frac{S}{N} = 0,0125 \text{ (-19 dB)}, \rho = v = 0,9913$$

Число итераций	α_1	α_2	β_0	β_1	β_2	δ_1	δ_2
100	-1,691 0,052	0,816 0,047	0,157 0,492	1,508 1,122	0,636 0,769	-0,776 0,101	0,240 0,155
.
1000	-1,587 0,041	0,731 0,043	0,023 0,243	1,142 0,260	0,470 0,272	-0,905 0,087	0,254 0,093

Таблица 7.6

$$\frac{S}{N} = 0,0125 \text{ (-19 дБ), } \rho = v = 1,0$$

Число итераций	α_1	α_2	β_0	β_1	β_2	δ_1	δ_2
100	-1,698 0,094	0,820 0,090	0,142 0,395	1,103 0,638	0,354 0,581	-0,802 0,158	0,301 0,094
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
1000	-1,640 0,070	0,764 0,053	-0,010 0,140	1,024 0,168	0,458 0,122	-0,858 0,067	0,275 0,050

Таблица 7.7

$$\frac{S}{N} = 0,0125 \text{ (-19 дБ), } \rho = v = 0,995$$

Число итераций	α_1	α_2	β_0	β_1	β_2	δ_1	δ_2
100	-1,676 0,092	0,817 0,088	0,138 0,367	1,102 0,605	0,449 0,538	-0,806 0,158	0,300 0,097
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
1000	-1,595 0,052	0,731 0,036	-0,036 0,197	0,983 0,237	0,490 0,176	-0,886 0,065	0,264 0,072

Таблица 7.8

$$\frac{S}{N} = 0,0125 \text{ (-19 дБ), } \rho = v = 0,990$$

Число итераций	α_1	α_2	β_0	β_1	β_2	δ_1	δ_2
100	-1,667 0,103	0,790 0,099	0,285 0,293	1,335 0,542	0,580 0,483	-0,865 0,173	0,334 0,071
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
1000	-1,574 0,050	0,731 0,040	-0,093 0,245	1,028 0,270	0,551 0,330	-0,926 0,071	0,272 0,064

Таблица 7.9

$$\frac{S}{N} = 0,0125 \text{ (-19 дБ)}, \rho = v = 0,985$$

Число итераций	α_1	α_2	β_0	β_1	β_2	δ_1	δ_2
100	-1,662 0,102	0,787 0,098	0,265 0,313	1,327 0,526	0,609 0,455	-0,875 0,174	0,337 0,077
:	:	:	:	:	:	:	:
1000	-1,558 0,058	0,722 0,061	0,075 0,329	1,0681 0,359	0,605 0,409	-0,933 0,082	0,266 0,080

Таблица 7.10

$$\frac{S}{N} = 0,0125 \text{ (-19 дБ)}, \rho = v = 0,980$$

Число итераций	α_1	α_2	β_0	β_1	β_2	δ_1	δ_2
100	-1,657 0,101	0,783 0,098	0,246 0,344	1,318 0,519	0,632 0,439	0,883 0,177	0,341 0,086
:	:	:	:	:	:	:	:
1000	1,549 0,073	0,715 0,081	-0,049 0,413	0,121 0,448	0,658 0,469	-0,935 0,095	0,264 0,099

Таблица 7.11

$$\frac{S}{N} = 0,0125 \text{ (-19 дБ)}, \rho = v = 0,975$$

Число итераций	α_1	α_2	β_0	β_1	β_2	δ_1	δ_2
100	-1,652 0,101	0,779 0,098	0,230 0,378	1,307 0,521	0,650 0,435	-0,891 0,181	0,345 0,097
:	:	:	:	:	:	:	:
1000	-1,543 0,089	0,711 0,098	-0,020 0,488	1,177 0,526	0,707 0,517	-0,936 0,110	0,264 0,120

С увеличением мощности шума:

1) стандартные отклонения β_0 , β_1 и β_2 растут линейно по λ , т. е. пропорционально корню квадратному из S/N ;

2) стандартные отклонения α_1 и α_2 остаются постоянными;

3) стандартные отклонения δ_1 и δ_2 остаются приблизительно постоянными (фиг. 7.18);

С увеличением мощности шума $b_i - \bar{b}_i(k)$ становится монотонно возрастающей функцией k . Кроме того, по-видимому, с увеличением мощности шума уменьшается скорость сходимости.

С уменьшением весовых коэффициентов:

1) значения оценок β_i все больше приближаются к истинным значениям параметров, хотя дисперсии оценок растут;

2) разность $a_i - \bar{a}_i$ монотонно уменьшается; стандартные отклонения снижаются, пока $\rho = v$ не станут равны 0,99, а затем снова растут;

3) значения δ_i улучшаются, хотя стандартные отклонения становятся несколько больше.

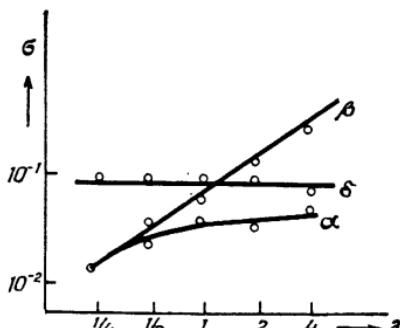
Эти эксперименты показывают, что при низком уровне шума могут быть выбраны большие значения весовых коэффициентов. В тех случаях, когда параметры постоянны, представляется разумным постепенно увеличивать весовые коэффициенты до единицы с тем, чтобы обеспечить более быструю сходимость и малые стандартные отклонения.

Расширенный матричный метод. Объект и алгоритмы описываются формулами (7.60)–(7.67). При тех же параметрах находим

$$y(k) = \sum_{i=0}^2 b_i u(k-i) - \sum_{i=1}^2 a_i y(k-i) + r(k), \quad (7.79)$$

$$b_0 = 0, \quad b_1 = 1, \quad b_2 = 0,5,$$

$$a_1 = -1,5, \quad a_2 = 0,7.$$



Ф и г. 7.18.

Теперь возмущения имеют вид

$$r(k) = \lambda \sum_i c_i w(k-i) - \sum_i d_i r(k-i) + \lambda w(k), \quad (7.80)$$

$$c_1 = 0,3, \quad c_2 = 0,0, \quad c_3 = 0,0, \quad c_4 = 0,0,$$

$$d_1 = -0,5, \quad d_2 = 0,0, \quad d_3 = 0,0.$$

Уровень шума снова определяется параметром λ . Фильтр помехи характеризуется «усилением по мощности» $\sigma_n^2/\sigma_w^2 = 43,62$.

Теперь можно исследовать, как влияет на оценивание параметров объекта использование неправильно заданного числа параметров и неправильного выбора параметров помехи. Во время эксперимента весовой коэффициент по экспоненте приближался к единице:

$$\rho(k+1) = \rho(k) + \{1 - \rho(k)\} \Delta\rho = \{1 - \Delta\rho\} \rho(k) + \Delta\rho, \quad (7.81)$$

$$\rho(0) = 0,9913.$$

Результаты экспериментов приводятся в таблицах и на рисунках, указанных ниже.

Номер таблицы	λ	Отношение сигнал/шум на выходе объекта	$\Delta\rho$	Оцениваемые параметры объекта $\alpha_1, \alpha_2, \beta_0, \beta_1, \beta_2$		Номер фигуры
				Оцениваемые параметры шума		
7.12	1	0,432	-3,7 дБ	0,001	-	7.19
7.13	1	0,432	-3,7 дБ	0,001	δ_1	7.20
7.14	1	0,432	-3,7 дБ	0,001	δ_1, δ_2	7.21
7.15	1	0,432	-3,7 дБ	0,001	$\delta_1, \delta_2, \delta_3$	7.22
7.16	1	0,432	-3,7 дБ	0,001	γ_1	7.23
7.17	1	0,432	-3,7 дБ	0,001	γ_1, γ_2	7.24
7.18	1	0,432	-3,7 дБ	0,001	$\gamma_1, \gamma_2, \gamma_3$	7.25
7.19	1	0,432	-3,7 дБ	0,001	$\gamma_1, \gamma_2, \gamma_3, \gamma_4$	7.26
7.20	0,25	6,92	+8,3 дБ	0,001	δ_1, γ_1	7.27
7.21	1	0,432	-3,7 дБ	0,001	δ_1, γ_1	7.28
7.22	4	0,0269	-15,7 дБ	0,001	δ_1, γ_1	7.29
7.23	16	0,0017	-27,7 дБ	0,001	δ_1, γ_1	7.30
7.24	1	0,432	-3,7 дБ	0	δ_1, γ_1	
7.25	1	0,432	-3,7 дБ	0,001	δ_1, γ_1	
7.26	1	0,432	-3,7 дБ	0,005	δ_1, γ_1	
7.27	1	0,432	-3,7 дБ	0,025	δ_1, γ_1	

В табл. 7.12 и на фиг. 7.19 представлен случай, когда оцениваются только параметры объекта. Как и следовало ожидать, не обнаруживается заметного смещения оценок, так что значения оценок лежат в пределах одного стандартного отклонения от истинных значений параметров.

Таблица 7.12
Оцениваются только параметры объекта

Число итераций	α_1	α_2	β_0	β_1	β_2
100	-1,709	0,859	0,024	1,042	0,386
	0,027	0,044	0,103	0,116	0,110
200	-1,701	0,857	0,012	1,029	0,386
	0,030	0,026	0,065	0,084	0,058
300	-1,690	0,850	0,014	1,026	0,394
	0,031	0,027	0,072	0,089	0,052
400	-1,690	0,853	0,059	1,057	0,398
	0,029	0,021	0,055	0,108	0,072
500	-1,693	0,851	0,042	1,035	0,392
	0,015	0,018	0,106	0,128	0,085
600	-1,687	0,849	-0,004	0,968	0,357
	0,011	0,015	0,077	0,100	0,070
700	-1,695	0,852	-0,011	0,974	0,375
	0,011	0,019	0,071	0,094	0,050
800	-1,689	0,853	-0,029	0,968	0,390
	0,016	0,014	0,083	0,090	0,044
900	-1,688	0,849	-0,011	0,974	0,380
	0,018	0,018	0,078	0,089	0,063
1000	-1,689	0,846	0,012	0,998	0,410
	0,013	0,014	0,040	0,073	0,064

Таблица 7.13

Оценивается один параметр шума; описание в «обратном» времени

Число итераций	α_1	α_2	β_0	β_1	β_2	δ_1
100	-1,558	0,742	0,005	1,038	0,482	-0,625
	0,080	0,073	0,064	0,089	0,141	0,139
·	·	·	·	·	·	·
	·	·	·	·	·	·
1000	-1,481	0,718	-0,011	0,981	0,509	-0,700
	0,041	0,029	0,040	0,051	0,065	0,051

Таблица 7.14

Оцениваются два параметра шума; описание в обратном времени

Число итераций	α_1	α_2	β_0	β_1	β_2	δ_1	δ_2
100	-1,623 0,038	0,827 0,044	-0,019 0,101	0,983 0,118	0,438 0,127	-0,661 0,166	0,217 0,119
.
.
1000	-1,512 0,034	0,696 0,033	0,005 0,049	0,998 0,055	0,507 0,062	-0,770 0,068	0,200 0,066

Таблица 7.15

Оцениваются три параметра шума; описание в «обратном» времени

Число итераций	α_1	α_2	β_0	β_1	β_2	δ_1	δ_2	δ_3
100	-1,583 0,045	0,756 0,051	0,025 0,135	1,006 0,172	0,419 0,173	-0,720 0,111	0,238 0,186	-0,104 0,127
.
.
1000	-1,511 0,037	0,710 0,034	0,010 0,052	1,008 0,057	0,503 0,060	-0,783 0,046	0,232 0,067	-0,069 0,034

Таблица 7.16

Оценивается один параметр шума; описание в прямом времени

Число итераций	α_1	α_2	β_0	β_1	β_2	γ_1
100	-1,643 0,034	0,804 0,037	-0,009 0,060	1,026 0,090	0,429 0,129	0,642 0,177
.
.
1000	-1,602 0,020	0,767 0,018	-0,014 0,033	0,995 0,040	0,465 0,053	0,677 0,053

Таблица 7.17

Оцениваются два параметра шума; описание в прямом времени

Число итераций	α_1	α_2	β_0	β_1	β_2	γ_1	γ_2
100	-1,625 0,042	0,794 0,041	-0,029 0,096	0,973 0,111	0,435 0,126	0,683 0,166	0,237 0,173
·	·	·	·	·	·	·	·
·	·	·	·	·	·	·	·
1000	-1,552 0,034	0,726 0,031	0,005 0,048	0,997 0,050	0,487 0,058	0,741 0,073	0,297 0,056

Таблица 7.18

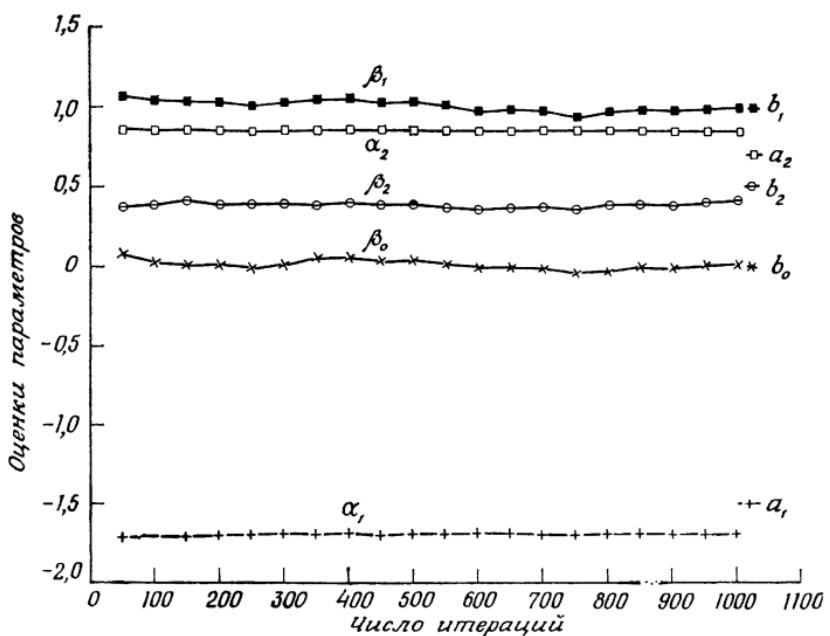
Оцениваются три параметра шума; описание в прямом времени

Число итераций	α_1	α_2	β_0	β_1	β_2	γ_1	γ_2	γ_3
100	-1,597 0,046	0,764 0,047	0,039 0,141	1,020 0,192	0,424 0,161	0,760 0,139	0,316 0,215	0,124 0,223
·	·	·	·	·	·	·	·	·
·	·	·	·	·	·	·	·	·
1000	-1,524 0,030	0,707 0,028	0,007 0,052	1,009 0,056	0,498 0,059	0,779 0,046	0,346 0,065	0,151 0,059

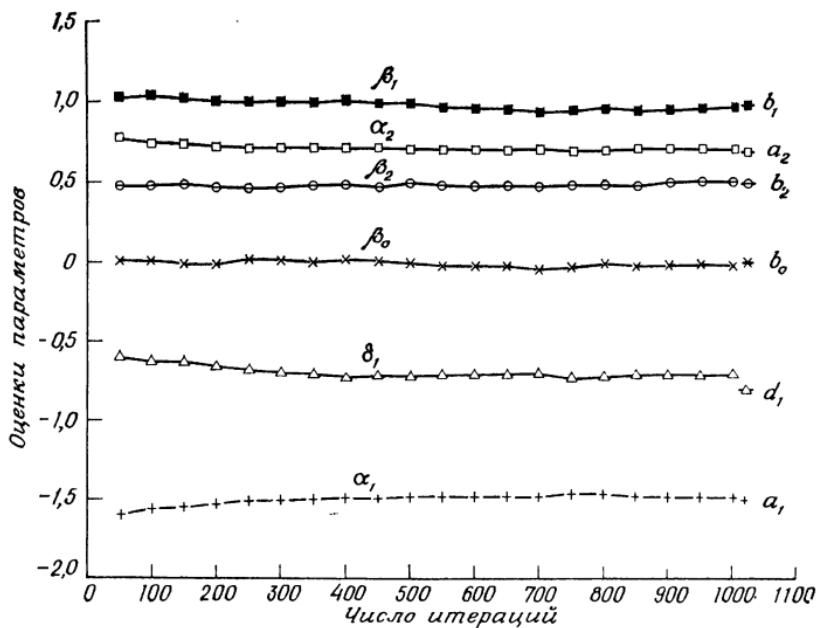
Таблица 7.19

Оцениваются четыре параметра шума; описание в прямом времени

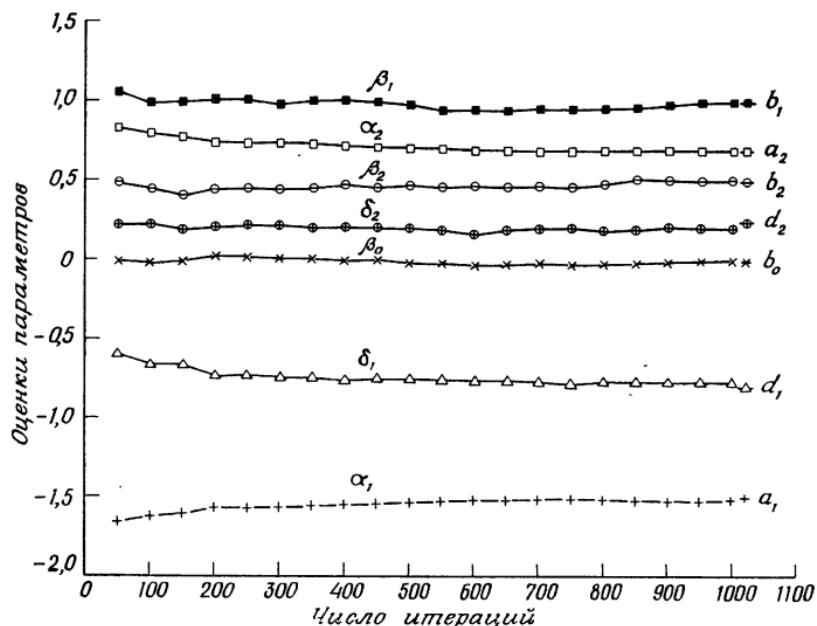
Число итераций	α_1	α_2	β_0	β_1	β_2	γ_1	γ_2	γ_3	γ_4
100	-1,621 0,046	0,788 0,046	-0,018 0,124	0,966 0,169	0,411 0,103	0,739 0,178	0,241 0,183	-0,052 0,258	0,012 0,166
·	·	·	·	·	·	·	·	·	·
·	·	·	·	·	·	·	·	·	·
1000	-1,514 0,038	0,706 0,033	0,003 0,055	1,006 0,069	0,495 0,068	0,790 0,047	0,362 0,071	0,175 0,096	0,077 0,058



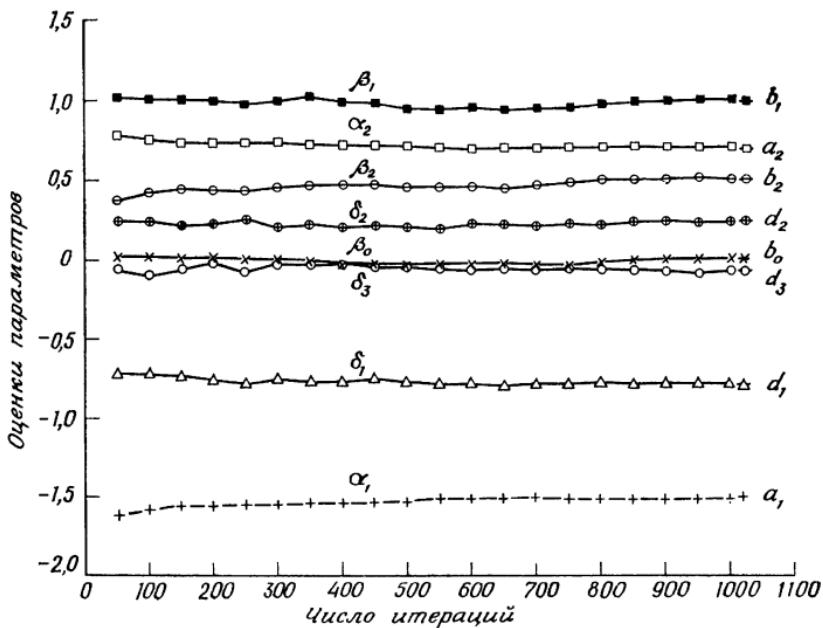
Ф и г. 7.19.



Ф и г. 7.20.



Ф и г. 7.21.



Ф и г. 7.22.

В табл. 7.13—7.15 и на фиг. 7.20—7.22 представлены случаи, когда помеха описана моделью авторегрессии (описание в «обратном» времени), т. е. помеха удовлетворяет уравнению вида

$$r'(k) = - \sum_i d_i r'(k-i) + \lambda w(k),$$

или

$$r'(k) = \frac{\lambda w(k)}{D^*(z^{-1})}. \quad (7.82)$$

Механизм образования помехи можно приближенно описать как

$$\begin{aligned} r(k) &= \frac{C^*(z^{-1})}{D^*(z^{-1})} \lambda w(k) = \frac{1+0,3z^{-1}}{1-0,5z^{-1}} \lambda w(k) = \\ &= \frac{\lambda w(k)}{1-0,8z^{-1}+0,24z^{-2}-0,072z^{-3}+\dots}. \end{aligned} \quad (7.83)$$

Увеличение числа параметров приводит к более точному описанию объекта и помехи; сравнение с фиг. 7.19 обнаруживает уменьшение смешения оценок. Заметим, что истинные значения параметров отличаются от оценок меньше чем на одно стандартное отклонение.

В табл. 7.16—7.19 и на фиг. 7.23—7.26 представлены случаи, когда помеха описывается моделью скользящего среднего (описание в прямом времени), т. е. помеха удовлетворяет уравнению вида

$$r''(k) = \lambda \{ \sum c_i w(k-i) + w(k) \},$$

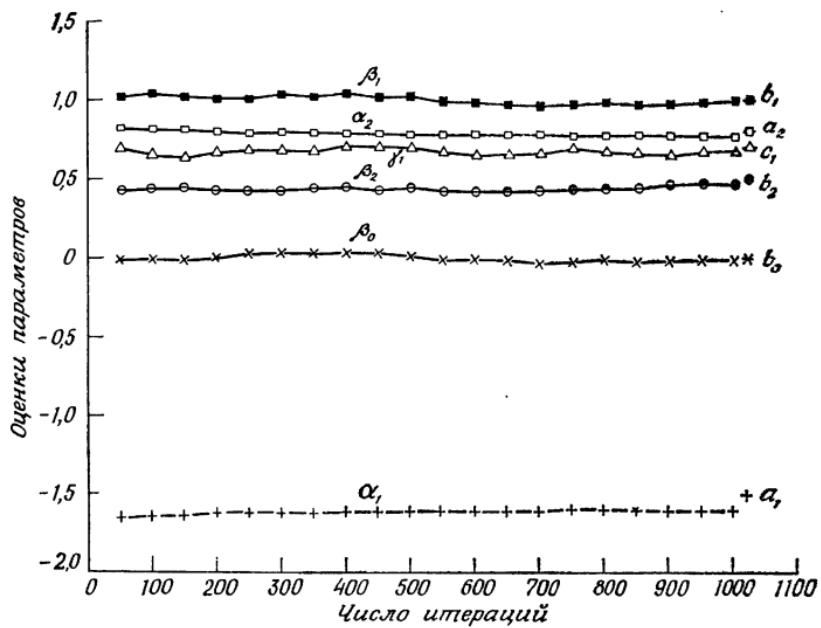
или

$$r''(k) = \lambda C^*(z^{-1}) w(k). \quad (7.84)$$

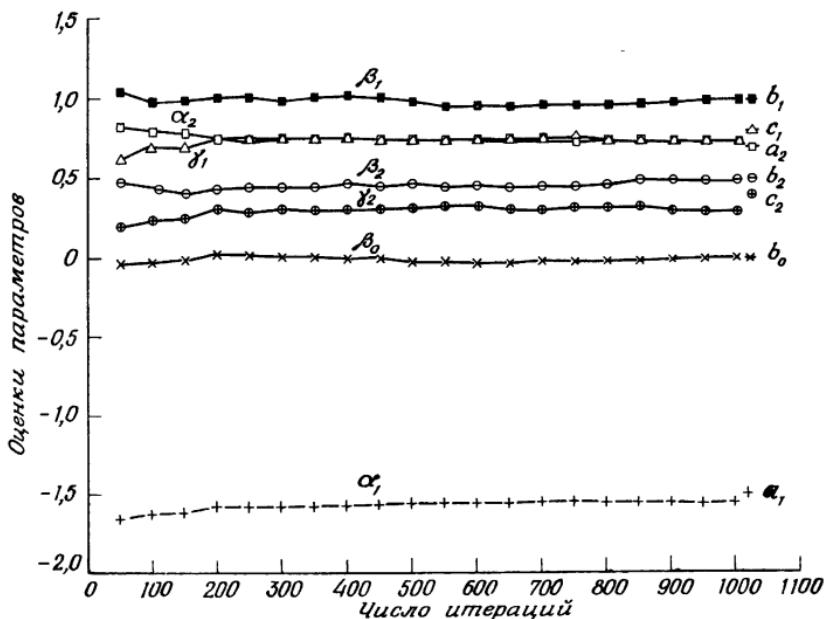
Механизм образования помехи можно приближенно описать как

$$\begin{aligned} r(k) &= \frac{C^*(z^{-1})}{D^*(z^{-1})} \lambda w(k) = \frac{1+0,3z^{-1}}{1-0,5z^{-1}} \lambda w(k) = \\ &= \{1+0,8z^{-1}+0,4z^{-2}+0,2z^{-3}+\dots\} \lambda w(k). \end{aligned} \quad (7.85)$$

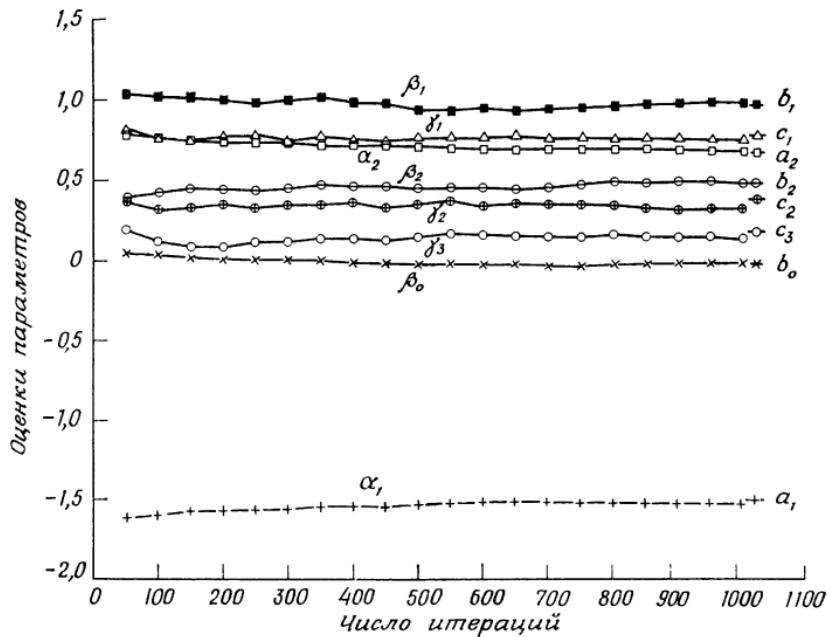
Опять увеличение числа параметров приводит к лучшему описанию помехи и объекта (см. фиг. 7.19); оказывается, что для получения сопоставимых результатов следует



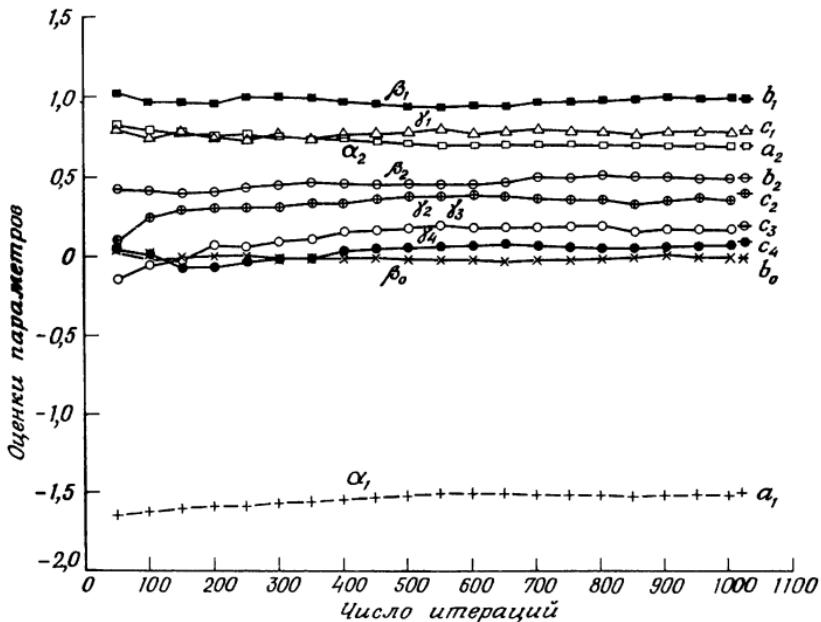
Ф и г. 7.23.



Ф и г. 7.24.



Ф и г. 7.25.



Ф и г. 7.26.

оценивать больше параметров c , чем параметров d . Это объясняется тем, что с увеличением i значения коэффициентов d_i уменьшаются быстрее, чем коэффициентов c_i .

В табл. 7.20—7.23 и на фиг. 7.27—7.30 представлен случай, когда модель задается одним параметром c и одним параметром d , т. е. параметры модели выбраны правильно. В табл. 7.21 и на фиг. 7.28 отношение S/N выбрано таким же, как и в предыдущих примерах; в этом случае получаются хорошие оценки. Далее представлены результаты экспериментов с различными отношениями сигнал/шум; на каждом шаге величина S/N уменьшается в 4 раза; так, на фиг. 7.30 $S/N = 0,0017!$

Изучение этих результатов показывает, что:

1) стандартные отклонения параметров β растут как линейная функция λ , т. е. пропорционально корню квадратному из S/N ;

2) стандартные отклонения параметров α стремятся к константам;

3) стандартные отклонения γ и δ не меняются.

В табл. 7.24—7.27 представлены случаи, когда скорость изменения весового коэффициента ρ меняется как функция k . Выбраны следующие значения $\Delta\rho$: 0, 0,001, 0,005, 0,025, такие, что разность между единицей и значением весового коэффициента равна половине аналогичной разности после следующего числа итераций: ∞ , 698, 128, 27.

Таблица 7.20

$$\frac{S}{N} = 6,92 \text{ (8,3 дБ)}$$

Число итераций	α_1	α_2	β_0	β_1	β_2	δ_1	γ_1
100	-1,519 0,019	0,714 0,025	-0,009 0,024	0,989 0,028	0,487 0,029	-0,452 0,172	0,342 0,169
.
.
1000	-1,500 0,010	0,701 0,011	0,000 0,013	0,999 0,014	0,502 0,018	-0,502 0,060	0,300 0,095

Таблица 7.21

$$\frac{S}{N} = 0,432 \text{ (-3,7 dB)}$$

Число итераций	α_1	α_2	β_0	β_1	β_2	δ_1	γ_1
100	-1,601 0,039	0,777 0,047	-0,018 0,101	0,983 0,135	0,433 0,139	-0,366 0,160	0,332 0,170
.
1000	-1,508 0,042	0,703 0,034	0,002 0,046	1,000 0,051	0,504 0,059	-0,482 0,085	0,308 0,092

Таблица 7.22

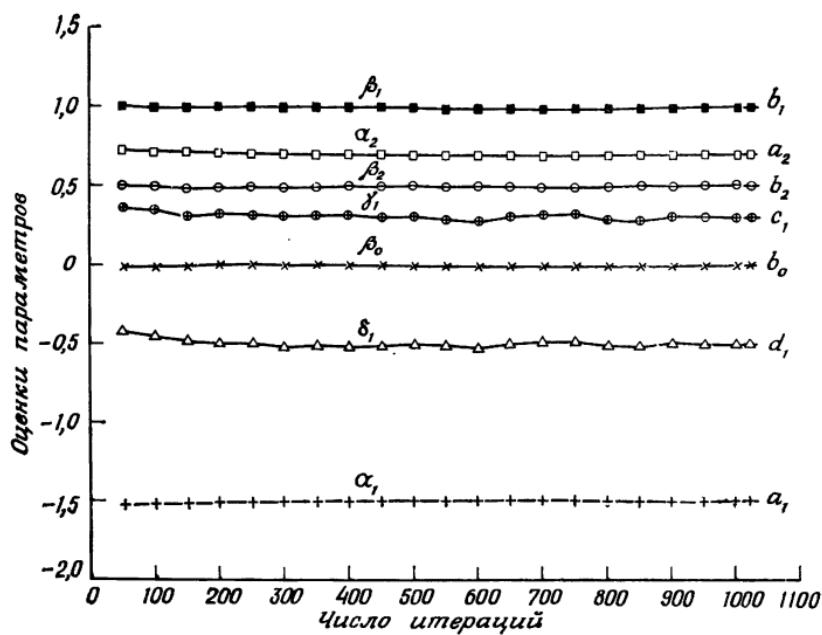
$$\frac{S}{N} = 0,0269 \text{ (-15,7 dB)}$$

Число итераций	α_1	α_2	β_0	β_1	β_2	δ_1	γ_1
100	-1,627 0,057	0,793 0,065	-0,077 0,405	0,915 0,470	0,415 0,494	-0,327 0,173	0,349 0,183
.
1000	-1,515 0,057	0,706 0,045	0,005 0,202	0,969 0,218	0,517 0,220	-0,473 0,104	0,311 0,094

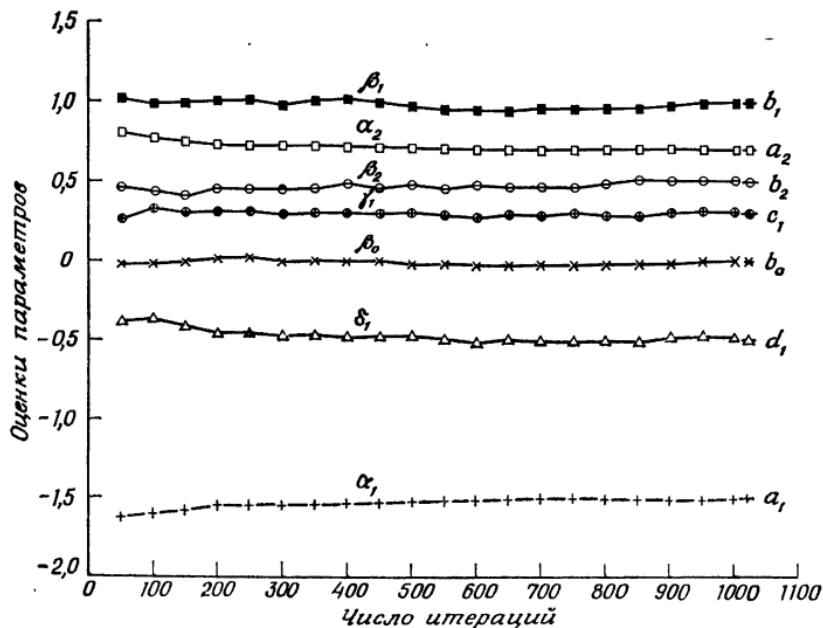
Таблица 7.23

$$\frac{S}{N} = 0,0017 \text{ (-27,7 dB)}$$

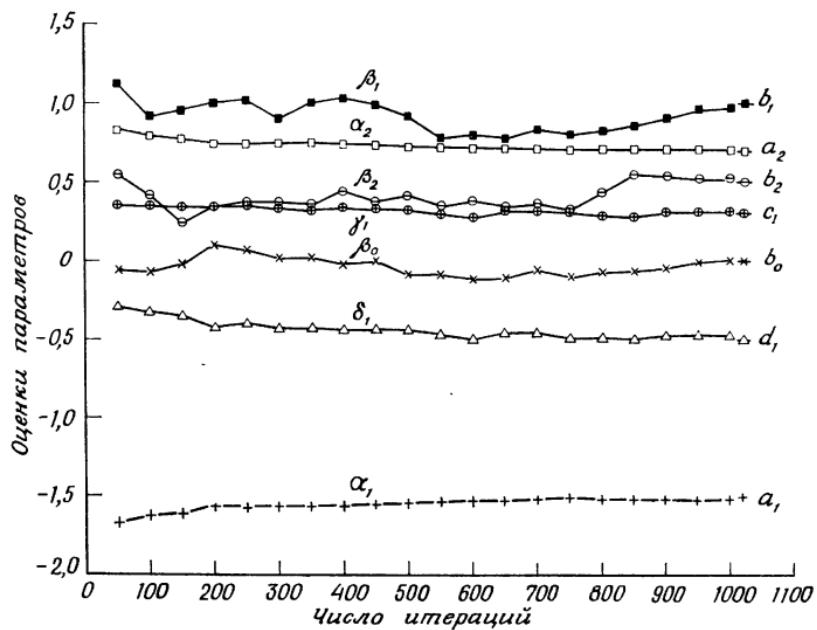
Число итераций	α_1	α_2	β_0	β_1	β_2	δ_1	γ_1
100	-1,623 0,064	0,788 0,075	-0,329 1,621	0,603 1,956	0,304 1,875	-0,342 0,155	0,330 0,180
.
1000	-1,516 0,057	0,707 0,045	-0,013 0,806	0,863 0,870	0,576 0,883	-0,471 0,106	0,312 0,097



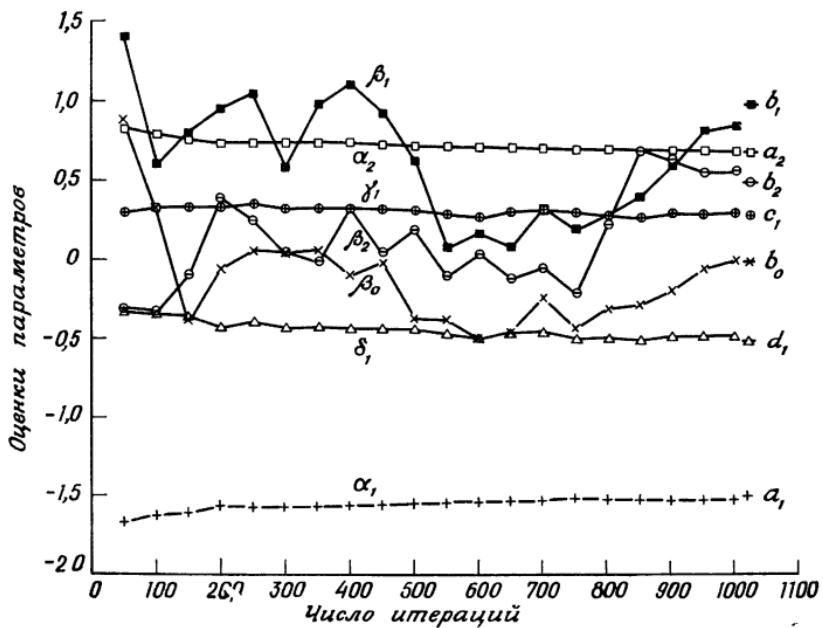
Ф и г. 7.27.



Ф и г. 7.28.



Ф и г. 7.29.



Ф и г. 7.30.

Таблица 7.24

$$\rho_0 = 0,9913, \Delta\rho = 0,0$$

Число итераций	α_1	α_2	β_0	β_1	β_2	δ_1	γ_1
100	-1,605 0,045	0,778 0,050	-0,021 0,099	0,982 0,118	0,447 0,127	-0,341 0,193	0,364 0,169
.
.
1000	-1,509 0,051	0,704 0,040	0,010 0,081	1,010 0,086	0,513 0,090	-0,474 0,106	0,333 0,111

Таблица 7.25

$$\rho_0 = 0,9913, \Delta\rho = 0,001$$

Число итераций	α_1	α_2	β_0	β_1	β_2	δ_1	γ_1
100	-1,601 0,039	0,777 0,047	-0,018 0,101	0,983 0,135	0,433 0,139	-0,366 0,160	0,332 0,170
.
.
1000	-1,508 0,042	0,703 0,034	0,002 0,046	1,000 0,051	0,504 0,059	-0,482 0,085	0,308 0,092

Таблица 7.26

$$\rho_0 = 0,9913, \Delta\rho = 0,005$$

Число итераций	α_1	α_2	β_0	β_1	β_2	δ_1	γ_1
100	-1,603 0,038	0,778 0,047	-0,018 0,098	0,984 0,133	0,434 0,138	-0,366 0,160	0,328 0,170
.
.
1000	-1,519 0,025	0,709 0,026	-0,005 0,037	0,985 0,055	0,486 0,042	-0,482 0,064	0,290 0,077

Таблица 7.27

$$\rho_0 = 0,9913, \Delta\rho = 0,025$$

Число итераций	α_1	α_2	β_0	β_1	β_2	δ_1	γ_1
100	-1,612 0,043	0,785 0,050	-0,019 0,091	0,992 0,113	0,447 0,124	-0,343 0,179	0,348 0,148
.
1000	-1,538 0,022	0,721 0,025	-0,005 0,038	0,985 0,060	0,475 0,045	-0,458 0,071	0,294 0,073

Из таблиц видно, что с увеличением ρ стандартное отклонение убывает. В общем случае при увеличении $\Delta\rho$ оценки ухудшаются, тогда как стандартное отклонение стремится к нижней границе.

С целью более глубокого изучения свойств различных схем проделана большая теоретическая и экспериментальная работа. Особенно это относится к задаче выбора весовых коэффициентов. Исключительно важной остается задача правильного подбора числа параметров α , β , γ и δ .

7.4. МОДЕЛИ, НЕЛИНЕЙНЫЕ ПО ПАРАМЕТРАМ

В гл. 4 было подчеркнуто значение линейности по параметрам. Изложенные методы относились к объектам, обладающим этим свойством. С точки зрения упрощения процедур оценивания этот тип линейности оказался более существенным, нежели линейность динамики систем. Для некоторых классов нелинейных систем, которые можно представить в виде соединения линейной динамической подсистемы и безынерционной нелинейной части, иногда удается применить методы идентификации по настраиваемой модели [10, 24].

Рассмотрим общий случай нелинейности по параметрам (модели). Выход модели можно представить в виде

$$w = f(U, \beta). \quad (7.86)$$

Функцию ошибок (или потерь) снова можно записать как

$$E = e'e, \text{ где } e = y - w = y - f(U, \beta). \quad (7.87)$$

В этом случае для минимизации E можно использовать методы, изложенные в разд. 5.2. Отметим, что для линейной по параметрам модели функция ошибок оказывалась квадратичной по параметрам. Это удобное свойство в рассматриваемом здесь случае утрачивается. В качестве примера можно привести два подхода:

а) Метод Гаусса или Гаусса — Ньютона, основанный на разложении функции f в ряд Тейлора:

$$w = f(U, \beta + \Delta\beta) \approx f(U, \beta) + \frac{\partial f}{\partial \beta} \Delta\beta,$$

или

$$w \approx f + M\Delta\beta. \quad (7.88)$$

Теперь функция w снова линейна по $\Delta\beta$, и, следовательно, $\Delta\beta$ можно найти, применяя стандартный метод наименьших квадратов. Модель настраивается по величине $G\Delta\beta$ итеративно.

б) Градиентный метод. Определяется градиент функции ошибок

$$\nabla_{\beta} E = \left[\frac{\partial E}{\partial \beta_1}, \dots, \frac{\partial E}{\partial \beta_m} \right]'. \quad (7.89)$$

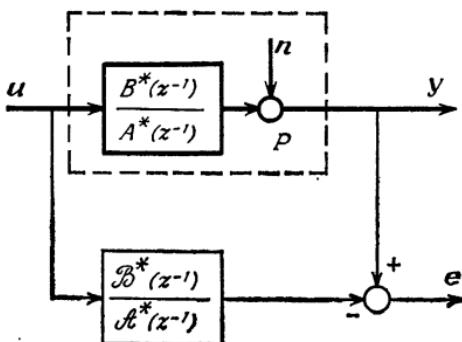
Настройка производится по величине $G\nabla_{\beta} E$.

В разд. 2.4 обсуждались вопросы вычисления частных производных и градиентов. Это можно сделать, оценивая результаты возмущения параметров модели. В этом случае может оказаться полезной гибридная вычислительная машина, состоящая из быстродействующей аналоговой части в сочетании с цифровой машиной.

Вообще говоря, алгоритм «а» может быть расходящимся, а алгоритмы метода «б» могут очень медленно сходиться в окрестности оптимума. В работе [22] используется сочетание этих методов. Для дальнейшего изучения вопросов сходимости читателю рекомендуется вернуться к гл. 5.

Причины использования нелинейных по параметрам моделей могут быть продиктованы возможными прикладными задачами или желаемыми свойствами оценок.

Если описание объекта основано на использовании дискретных весовых функций, то в качестве адекватного представления можно выбрать обобщенную модель объекта.



Фиг. 7.31.

Однако недостаток такого описания состоит в том, что наличие аддитивных помех приводит к смещению некоторых параметров. Таким образом, может понадобится рассмотреть модель, изображенную на фиг. 7.31, и минимизировать

$$E = \sum e^2(k), \quad (7.90)$$

где

$$e(k) = y(k) - \frac{\mathcal{B}^*(z^{-1})}{\mathcal{A}^*(z^{-1})} u(k). \quad (7.91)$$

Решение нелинейной задачи оптимизации было получено в работе [7] с помощью метода квазилинеаризации. Интересные применения этого метода можно найти в [8, 9, 33, 38]. Другой метод решения нелинейной задачи оптимизации предложен в работе [37], где он использовался при решении задачи определения устойчивости полета по производным.

Можно дать другую вероятностную интерпретацию функции ошибок, если предположить, что единственным источником возмущений является белый шум ошибок измерений. В работе [4] предложен метод, позволяющий рассматривать случай, когда ошибки измерений описываются стационарным случайным процессом с неизвестной спектральной плотностью. Специфичность принятых в этой работе допущений заставляет предполагать, что при их невыполнении могут возникнуть серьезные ошибки. Результатов, которые обосновывали или опровергали бы этот метод, не известно.

Задача настройки моделей весьма близка к вопросам, обсуждаемым в гл. 13, в которой содержатся ссылки на другие работы.

ЛИТЕРАТУРА

1. Albert A. E., Gardner L. A., Jr. (см. литературу к гл. 5).
2. Albert A., Sittler R. W., A Method for Computing Least Squares Estimators That Keep up with the Data, *SIAM J. Control*, 3, 384—417 (1965).
3. Anderson W. M. et. al. (см. литературу к гл. 13).
4. Åström K. J., Bohlin T., Wensmark S. (см. литературу к гл. 4).
5. Åström K. J., Jury E. I., Agniel R. C., A Numerical Method for the Evaluation of Complex Integrals (Short Paper), *IEEE Trans. Autom. Control*, AC-15, 468—471 (1970).
6. Bélanger P. R., Comments on «A Learning Method for System Identification» (Correspondence), *IEEE Trans. Autom. Control*, AC-13, 207—208 (1968).
7. Bellman R. E., Kalaba R. E. (см. литературу к гл. 13).
8. Buel J. D., Kagiwada H. H., Kalaba R. E., A Proposed Computational Method for Estimation of Orbital Elements Drag Coefficients and Potential Fields Parameters from Satellite Measurements, *Annales de Géophysique*, 23, 35—39 (1967).
9. Buel J. D., Kalaba R. E., Quasilinearization and the Fitting of Nonlinear Models of Drug Metabolism to Experimental Kinetic Data, *Mathematical Biosciences*, 5, 121—132 (1969).
10. Butler R. E., Bohn E. V., An Automatic Identification Technique for a Class of Nonlinear Systems (Short Paper), *IEEE Trans. Autom. Control*, AC-11, 292—296 (1966).
11. Clarke D. W., Generalized-Least-Squares Estimation of the Parameters of a Dynamic Model, IFAC Symp. «Identification in Autom. Control Systems», Prague, Paper 3.17, 1967.
12. Crum L. A., Wu S. H., Convergence of the Quantizing Learning Method for System Identification (Correspondence), *IEEE Trans. Autom. Control*, AC-13, 297—298 (1968).
13. Elliott D. F., Sworoder D. D. (см. литературу к гл. 5).
14. Hasting-James R., Sage M. W., Recursive Generalised-Least-Squares Procedure for On-Line Identification of Process Parameters, *Proc. IEE*, 116, 2057—2062 (1969).
15. Ho Y. C., Whalen B. W., An Approach to the Identification and Control of Linear Dynamic Systems with Unknown Parameters (Correspondence), *IEEE Trans. Autom. Control*, AC-8, 255—256 (1963).
16. Holmes J. K. (см. литературу к гл. 5).
17. Jury E. I. (см. литературу к гл. 12).
18. Kalman R. E., Design of a Self-Optimizing Control System, *Trans. ASME*, 80, 468—478 (1958).
19. Klinger A., Prior Information and Bias in Sequential Estimation (Correspondence), *IEEE Trans. Autom. Control*, AC-13, 102—103 (1968).

20. Leathrum J. F., On-Line Recursive Estimation, Joint Autom. Control Conf., 1969, pp. 173—178.
21. Lee R. C. K. (см. литературу к гл. 4).
22. Marquardt D. W., An algorithm for Least-Squares Estimation of Nonlinear Parameters, *SIAM J.*, 11, 431—441 (1963).
23. Nagumo J., Noda A., A Learning Method for System Identification, *IEEE Trans. Autom. Control*, AC-12, 282—287 (1968); см. также *AC-13*, 207—208 (1968).
24. Narendra K. S., Gallman P. G., An Iterative Method for the Identification of Nonlinear Systems Using a Hammerstein Model (Short Paper), *IEEE Trans. Autom. Control*, AC-11, 546—550 (1966).
25. Neal C. B., Bekey G. A., Estimation of the Parameters of Sampled Data Systems by Means of Stochastic Approximation, Joint Autom. Control Conf., 1969, pp. 977—978.
26. Oza K. G., Jury E. I., System Identification and the Principle of Contraction Mapping, *SIAM J. Control*, 6, 244—257 (1968).
27. Oza K. G., Jury E. I., Adaptive Algorithms for Identification Problem, Fourth IFAC Congress, Warsaw, Paper 26.4, 1969; см. также *Automatica*, 6, 795—799 (1970).
28. Panuska V., On the Maximum Likelihood Estimation of Rational Pulse Transfer Function Parameters (Correspondence), *IEEE Trans. Autom. Control*, AC-13, 304 (1968).
29. Peterka V., Šmuk K. (см. литературу к гл. 6).
30. Sakrison D. J. (см. литературу к гл. 5).
31. Saridis G. N., Stein G. (см. литературу к гл. 5).
32. Saridis G. N., Stein G. (см. литературу к гл. 4).
33. Schulz E. R., Estimation of Pulse Transfer Function Parameters by Quasilinearization (Short Paper), *IEEE Trans. Autom. Control*, AC-13, 424—426 (1968).
34. Steiglitz K., McBride L. E. (см. литературу к гл. 4).
35. Smets A. J. (см. литературу к гл. 6).
36. Talmon J. L., Approximated Gauss-Markov Estimators and Related Schemes, Report, Group Measurement and Control, E. E. Dept., Univ. of Technology, Eindhoven, Netherlands, 1971.
37. Taylor L. W., Iliff K. W., Powers B. G., A Comparison of Newton-Raphson and Other Methods for Determining Stability Derivatives from Flight Data, Paper, AIAA Third Flight Test Simulation and Support Conf., Houston, Texas, 1969.
38. Vaneček A., Fessl J., A Contribution to Parameters and State Estimation and Synthesis via Quasi-Linearization, Second IFAC Symp. «Identification and Process Parameter Estimation», Prague, Paper 2.6, 1970.
39. Van den Boom A. J. W., Implicit Estimation of Process Parameters Using a Digital Computer (на голландском), Thesis, E. E. Dept., Univ. of Technology, Eindhoven, Netherlands, 1967.
40. Van den Boom A. J. W., Melis J. H. A. M. (см. литературу к гл. 6).
41. Whitakker H. P. et al., Design of a Model Reference Adaptive Control System for Aircraft, Report R-164, Instrumentation Lab., MIT, Boston, 1958.

42. Young P. C., The Use of Linear Regression and Related Procedures for the Identification of Dynamic Processes, Proc. IEEE Symp. on Adapt. Processes, Dec. 1968.
43. Young P. C., An Instrumental Variable Method for Real-Time Identification of a Noisy Process, Fourth IFAC Congress, Warsaw, Paper 26.6; см. также *Automatica*, 6, 271—287 (1970).

ДОПОЛНИТЕЛЬНАЯ ЛИТЕРАТУРА

Bellman R. E., Kagiwada H. H., Kalaba R. E., Identification of Linear Systems via Numerical Inversion of Laplace Transforms (Correspondence), *IEEE Trans. Autom. Control*, AC-10, 111—112 (1965).

Budin M. A., A New Approach to System Identification and State Estimation, *IEEE Trans. Systems, Man. Cybernet.*, SMC-2, 396—402 (1972).

Durbin J., Estimation of Parameters in Time-Series Regression Models, *J. R. Statist. Soc., Series B*, 22B, 139—153 (1960).

Farison J. B., Graham R. E., Shelton Rey C., Identification and Control of Linear Discrete Systems, *IEEE Trans. Autom. Control*, AC-12, 438—442 (1967).

Farison J. B., Parameter Identification for a Class of Linear Discrete Systems (Correspondence), *IEEE Trans. Autom. Control*, AC-12, 109 (1967).

Genin Y., A Note on Linear Minimum Variance Estimation Problems (Correspondence), *IEEE Trans. Autom. Control*, AC-13, 103 (1968).

Holmes J. K., System Determination with Noisy Measurements, IFAC Symp. «Identification in Autom. Control Systems», Prague, Paper 3.3, 1967.

Hsieh H. C., An On-Line Identification Scheme for Multi-variable Nonlinear Systems, в сб. «Computing Methods in Optimization Problems», Neustadt L. W., Balakrishnan A. V., eds., Academic Press, N.Y., 1964, pp. 193—210.

Hsieh H. C., Least Square Estimation of Linear and Nonlinear Weighting Function Matrices, *Information and Control*, 7, 84 (1964).

Mintz M. L., Thorp J. S., A Method for Discrete System Identification Approximation, *J. Franklin Inst.*, 283, 389—403 (1966).

Müller J. A., Analysis of Control Loops by Means of Adaptive Models (на немецком), *Zeitschr. Messen, Steuern, Regeln*, 11, 78—80, 146—152 (1968).

Nishimura M., Fujii K., Suzuki Y., On-Line Estimation of the Process Parameters and Its Applications to an Adaptive Control System, Fourth IFAC Congress, Warsaw, Paper 26.3, 1969.

Ochem W., Rault A., Weingarten H., A New Identification Algorithm for Linear Discrete Time Systems, Second IFAC Symp., «Identification and Process Parameter Estimation», Prague, Paper 2.9, 1970.

Pearson A. E., A Regression Analysis in Function Space for the Identification of Multivariable Linear Systems, Third IFAC Congress, London, Paper 46 B, 1966; есть русский перевод: Пирсон Э., Применение регрессионного анализа в функциональном простран-

стве для определения характеристик многомерных линейных систем, в книге «Теория непрерывных автоматических систем и вопросы идентификации», изд-во «Наука», 1971, стр. 265—274.

Pearson J. D., Deterministic Aspects of the Identification of Linear Discrete Dynamic Systems (Short Paper), *IEEE Trans. Autom. Control*, AC-12, 764—766 (1967).

Rault A., Pouliquen R., Richalet J. (см. литературу к гл. 11).

Roy R., Sherman J., A Learning Technique for Volterra Series Representation (Short Paper), *IEEE Trans. Autom. Control*, AC-12, 761—764 (1967).

Šmuk K., Comparison of Some Algorithms for Least Squares Updating, Second IFAC Symp. «Identification and Process Parameter Estimation», Prague, Paper 2.10, 1970.

Štanković S., Velasević D., Carapić M., An Approach to On-Line Discrete-Time System Identification, Second IFAC Symp. «Identification and Process Parameter Estimation», Prague, Paper 4.2, 1970.

Wilkie D. F., Perkins W. R., Design of Model Following Systems Using the Companion Transformation, Fourth IFAC Congress, Warsaw, Paper 19.2, 1969.

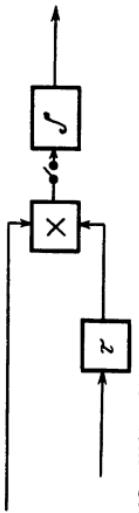
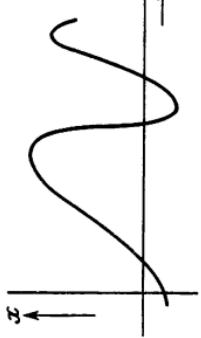
ГЛАВА 8

НЕПРЕРЫВНЫЕ СИГНАЛЫ; ЯВНЫЕ МЕТОДЫ

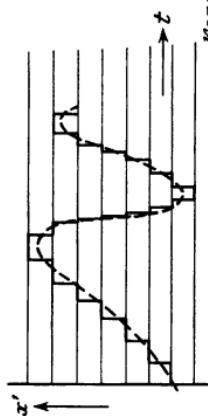
В этой главе рассматриваются корреляционные методы идентификации для сигналов, являющихся функциями непрерывного (неквантованного) времени. Ранее в гл. 2 было указано, что по способу представления амплитуды сигналов можно различать аналоговые, квантovанные и двоичные сигналы. Возможные значения амплитуды квантованного сигнала ограничены конечным числом заранее заданных уровней; для двоичных сигналов таких уровней два. Корреляционные методы для аналоговых сигналов обсуждаются в разд. 8.1. Следующий раздел посвящен описанию (нелинейной) операции квантования. В результате квантования удается упростить конструкцию коррелятора-умножителя по сравнению со случаем аналоговых сигналов, что справедливо в случае двоичных сигналов, когда для умножения можно использовать простые релейные или логические схемы. Эти вопросы рассмотрены в разд. 8.3. На фиг. 8.1 схематически представлены все упомянутые случаи.

Суть явных методов заключается в применении определенного оператора (например, умножения на функцию времени и интегрирования произведения) к входным и выходным сигналам. Корреляционные методы можно рассматривать как подкласс метода уравнений движения; это как раз случай дифференциальной аппроксимации. Этим вопросам посвящен разд. 8.4.

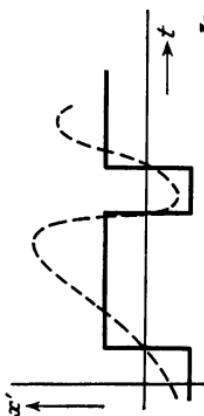
В разд. 8.5 упомянута возможность описания объектов с нелинейной динамикой корреляционными функциями более высокого порядка. Другой подход к исследованию нелинейных объектов, основанный на дисперсионных функциях, приведен в разд. 8.6.



Аналоговые сигналы



Квантованные сигналы



Двоичные сигналы

Фиг. 8.1.

8.1. КОРРЕЛЯЦИОННЫЕ МЕТОДЫ ДЛЯ АНАЛОГОВЫХ СИГНАЛОВ

Общие свойства корреляционных измерений

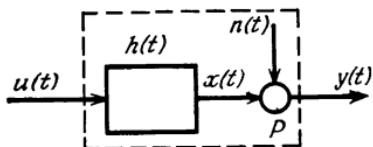
Прежде всего целесообразно напомнить, что корреляционные методы применялись для широкого круга задач, в частности для исследования случайных сигналов, идентификации систем управления, обнаружения сигналов и в таких областях, как, например, связь, радиоастрономия, радарная и ультразвуковая техника, исследование промышленных объектов, технические измерения, исследование механических колебаний, сейсмология, акустика, исследования мозга. Хотя важность корреляционных измерений признается уже давно, прошло немало времени, прежде чем необходимые приборы стали общедоступными. Это можно объяснить относительной сложностью и высокой критичностью некоторых элементов коррелятора. В частности, это относится к цепям временной задержки, широкополосным умножителям и интеграторам с большими интервалами интегрирования. Внедрение цифровой и гибридной техники облегчило решение многих из указанных задач.

Корреляционные методы позволяют оценить неизвестный параметр по критерию минимума среднеквадратической ошибки (см. гл. 2). Это легко показать для непрерывных сигналов. Рассмотрим линейный стационарный объект P , стационарные сигналы (эргоидические случайные процессы) и поставим задачу оценки весовой функции P (фиг. 8.2). Благодаря линейности выходной сигнал описывается интегралом свертки

$$y(t) = \int_0^{\infty} h(\theta) u(t - \theta) d\theta + n(t) \stackrel{\Delta}{=} h(t) * u(t) + n(t). \quad (8.1)$$

Сигналы $u(t)$ и $n(t)$ случайные; следовательно, $y(t)$ также случаен. Стационарность $u(t)$ влечет стационарность $y(t)$ для объектов с конечной «памятью» (конечным интервалом времени, на котором весовая функция $h(t)$ отлична от нуля).

Корреляционный метод. Умножая обе части формулы (8.1) на $u(t - \tau)$ и применяя операцию математического



Ф и г. 8.2.

ожидания, получаем

$$\mathcal{E}[u(t-\tau)y(t)] = \int_0^{\infty} h(\theta) \mathcal{E}[u(t-\tau)u(t-\theta)] \times d\theta + \mathcal{E}[u(t-\tau)n(t)] \quad (8.2)$$

или в обозначениях приложения Б для независимых $n(t)$ и $u(t)$ имеем

$$\Psi_{uy}(\tau) = \int_0^{\infty} h(\theta) \Psi_{uu}(\tau - \theta) d\theta + 0 = h(\tau) * \Psi_{uu}(\tau) + 0. \quad (8.3)$$

Уравнение (8.1) связывает *случайные* сигналы $u(t)$ и $y(t)$, а уравнение (8.3) — *детерминированные* функции, описывающие важные свойства этих случайных сигналов.

Минимизация среднеквадратической ошибки. Необходимо определить оценку весовой функции $h(t)$. Вновь рассмотрим усредненное по времени значение квадрата ошибки

$$E = \frac{1}{T} \int_0^T e^2 dt, \quad (8.4)$$

где

$$e(t) = y(t) - \int_0^{\infty} h(\theta) u(t - \theta) d\theta. \quad (8.5)$$

Пусть функция $\hat{h}(\theta)$ минимизирует E . Тогда произвольную функцию $h(\theta)$ можно представить в виде

$$h(\theta) = \hat{h}(\theta) + ah_a(\theta), \quad (8.6)$$

где h_a — произвольная функция, такая, что $h_a(\theta) \equiv 0$ при $\theta < 0$. По определению при $a \neq 0$ величина E больше, чем при $a = 0$. Следовательно,

$$\frac{dE}{da} \Big|_{a=0} = 0,$$

или

$$\frac{2}{T} \int_0^T \left\{ y(t) - \int_0^\infty \hat{h}(\theta) u(t-\theta) d\theta \right\} \times \\ \times \left\{ \int_0^\infty h_a(\tau) u(t-\tau) d\tau \right\} dt = 0.$$

Последнее условие можно переписать в виде

$$\int_0^\infty h_a(\tau) \left\{ \frac{1}{T} \int_0^T y(t) u(t-\tau) dt - \right. \\ \left. - \int_0^\infty \hat{h}(\theta) d\theta \frac{1}{T} \int_0^T u(t-\theta) u(t-\tau) dt \right\} d\tau = 0. \quad (8.7)$$

Ввиду произвольности $h_a(\tau)$ заключаем, что

$$\frac{1}{T} \int_0^T y(t) u(t-\tau) dt = \int_0^\infty \hat{h}(\theta) d\theta \frac{1}{T} \int_0^T u(t-\theta) u(t-\tau) dt, \quad (8.8)$$

или

$$\tilde{\Psi}_{uy}(\tau, T) = \int_0^\infty \hat{h}(\theta) \tilde{\Psi}_{uu}(\tau-\theta, T) d\theta. \quad (8.9)$$

На практике корреляционные измерения проводятся на конечном интервале времени, и в результате получаются приближенные оценки (отмеченные тильдой) истинных корреляционных функций. Используя условие эргодичности при $T \rightarrow \infty$, получаем уравнение Винера — Хопфа

$$\Psi_{uy}(\tau) = \int_0^\infty h(\theta) \Psi_{uu}(\tau-\theta) d\theta. \quad (8.10)$$

Сравните этот результат с формулой (8.3).

Исследуя $\partial^2 E / \partial \alpha^2$, можно убедиться, что при этом E действительно минимизируется. Отметим, что на $\hat{h}(\theta)$ не накладывается никаких ограничений, кроме условия физической реализуемости объекта $\hat{h}(\theta) \equiv 0$ при $\theta < 0$.

Определение $h(\theta)$ по корреляционным функциям Ψ_{uy} и Ψ_{uu} (*решение уравнения свертки*) можно выполнить на аналоговых или цифровых вычислительных машинах как во временной, так и в частотной областях. Во временной области задача решается просто, если полоса частот $u(t)$ велика по сравнению с полосой пропускания объекта. В тех случаях, когда входной сигнал является практически белым, можно использовать следующую аппроксимацию:

$$\Psi_{uu}(\tau - \theta) \approx c\delta(\tau - \theta), \quad c = \int_{-\infty}^{\infty} \Psi_{uu}(\tau) d\tau,$$

откуда с учетом свойств δ -функции Дирака имеем

$$\Psi_{uy}(\tau) \approx ch(\tau), \text{ или } h(\tau) \approx \frac{1}{c} \Psi_{uy}(\tau).$$

При этом взаимная корреляционная функция оказывается хорошим приближением для весовой функции объекта (фиг. 8.3).

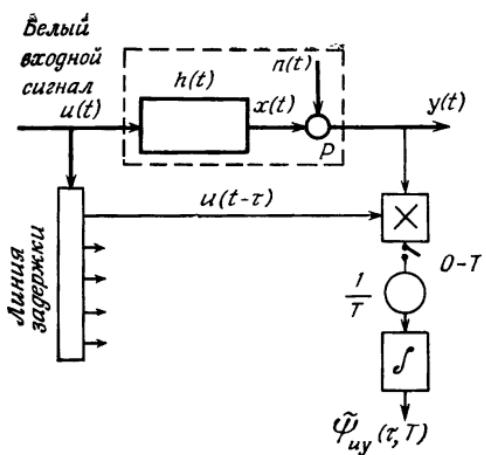
Если входной сигнал нельзя считать белым, необходимо решать задачу определения $h(\theta)$ из формулы (8.10). Один путь решения состоит в применении настраиваемой модели таким образом, чтобы $\Psi_{uy}(t)$ была ее выходным сигналом при входе $\Psi_{uu}(t)$ (см. [6, 11, 12, 35]).

Другой подход основан на аппроксимации интеграла в (8.10) суммой:

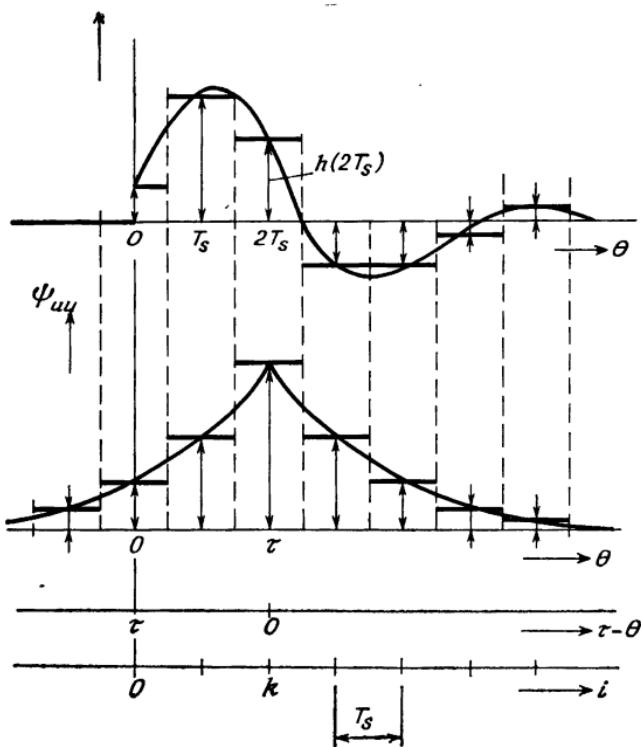
$$\begin{aligned} \Psi_{uy}(\tau) &= \int_0^{\infty} h(\theta) \Psi_{uu}(\tau - \theta) d\theta = \\ &= \int_0^{T_s/2} h(\theta) \Psi_{uu}(\tau - \theta) d\theta + \sum_{i=1}^{\infty} \int_{(i-\frac{1}{2})T_s}^{(i+\frac{1}{2})T_s} h(\theta) \Psi_{uu}(\tau - \theta) d\theta, \end{aligned}$$

или, как показано на фиг. 8.4,

$$\begin{aligned} \Psi_{uy}(kT_s) &\approx \frac{T_s}{2} h(0) \Psi_{uu}(kT_s) + \\ &+ T_s \sum_{i=1}^{\infty} h(iT_s) \Psi_{uu}((k-i)T_s). \end{aligned} \quad (8.11)$$



Ф и г. 8.3.



Ф и г. 8.4.

В качестве *примера* перепишем уравнение (8.11) в матричном виде для трех выборочных значений весовой функции

$$\begin{bmatrix} \Psi_{uy}(0) \\ \Psi_{uy}(T_s) \\ \Psi_{uy}(2T_s) \end{bmatrix} \approx T_s \begin{bmatrix} \Psi_{uu}(0) & \Psi_{uu}(-T_s) & \Psi_{uu}(-2T_s) \\ \Psi_{uu}(T_s) & \Psi_{uu}(0) & \Psi_{uu}(-T_s) \\ \Psi_{uu}(2T_s) & \Psi_{uu}(T_s) & \Psi_{uu}(0) \end{bmatrix} \times \begin{bmatrix} \frac{1}{2}h(0) \\ h(T_s) \\ h(2T_s) \end{bmatrix}.$$

Поскольку $\Psi_{uu}(\tau) = \Psi_{uu}(-\tau)$, матрица симметрична.

Эта формулировка задачи близка к приведенной в гл. 6. Разумеется, нетрудно предложить другие лучшие приближения для интеграла свертки, используя, например, линейную интерполяцию в промежутках между выборочными значениями функций h и Ψ .

Третий способ определения оценки для $h(\theta)$ состоит в переходе в частотную область. Применение преобразования Фурье к равенству (8.10) позволяет получить формулу

$$\Phi_{uy}(j\omega) = H(j\omega) \Phi_{uu}(j\omega), \quad (8.12)$$

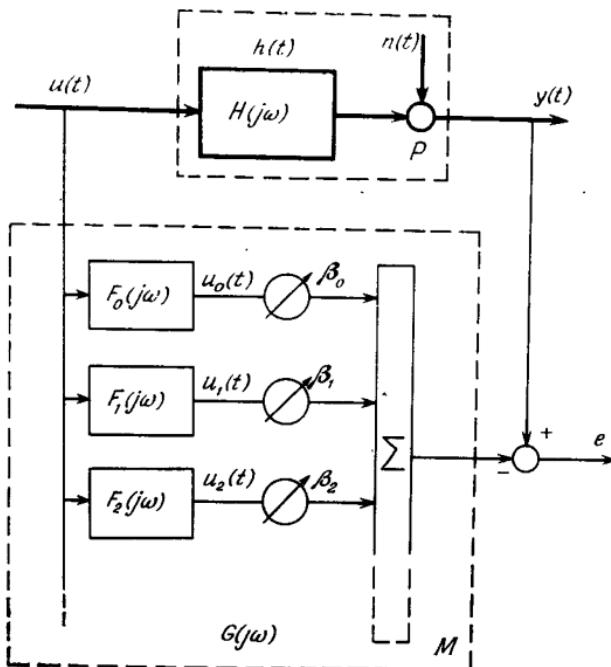
откуда

$$H(j\omega) = \frac{\Phi_{uy}(j\omega)}{\Phi_{uu}(j\omega)}.$$

По оценкам спектральных плотностей можно получить и оценку $H(j\omega)$. Практически вычисление изображений Фурье можно производить, например, известным численным методом быстрого преобразования Фурье (см. разд. 3.5).

Применение ортогональных фильтров

До сих пор обсуждение корреляционных методов велось в рамках практической реализации с использованием линий задержки. В этом, однако, нет необходимости. Другой интересный подход к определению операторов линейных динамических систем основан на использовании ортогональных фильтров [16, 17]. При этом удается обойтись без аналоговых линий задержки, остающихся



Ф и г. 8.5.

по-прежнему дорогими и критичными компонентами коррелятора. Тем не менее оказывается возможным оценить весовую функцию. Исследование свойств ортогональности можно проводить и во временной и в частотной областях. Первый подход можно сравнить с рассмотренным в разд. 3.2. Здесь мы ограничимся рассмотрением в частотной области.

Принципиальная постановка задачи иллюстрируется на фиг. 8.5. Среднеквадратическая ошибка равна

$$\overline{e^2(t)} = \Psi_{ee}(0) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \{H^*(j\omega) - G^*(j\omega)\} \times \\ \times \{H(j\omega) - G(j\omega)\} \Phi_{uu}(j\omega) d\omega + \Psi_{nn}(0), \quad (8.13)$$

где интеграл определяет обратное преобразование Фурье, а H и G характеризуют динамические свойства объекта P и модели M . Если не учитывать составляющей, вносимой шумом, правая часть последнего равенства представляет

собой квадрат разности H и G , взвешенный с весом Φ_{uu} . Модель состоит из ряда фильтров $F_i(j\omega)$, т. е.

$$G(j\omega) = \sum_{i=0}^m \beta_i F_i(j\omega). \quad (8.14)$$

Коэффициенты β_0, \dots, β_m должны «подстраиваться» таким образом, чтобы минимизировать $\overline{e^2(t)}$. Подставляя выражение (8.14) в (8.13), дифференцируя по β_k и приравнивая результат нулю при $\beta_k = \hat{\beta}_k$, получаем

$$\begin{aligned} \frac{1}{2\pi} \sum_{i=0}^m \hat{\beta}_i \int_{-\infty}^{\infty} \{F_k^*(j\omega) F_i(j\omega) + F_k(j\omega) F_i^*(j\omega)\} \Phi_{uu}(j\omega) d\omega = \\ = \frac{1}{2\pi} \int_{-\infty}^{\infty} \{F_k^*(j\omega) H(j\omega) + F_k(j\omega) H^*(j\omega)\} \Phi_{uu}(j\omega) d\omega, \\ k = 0, 1, \dots, m, \end{aligned} \quad (8.15)$$

или, заменяя интегралы символом I , имеем

$$\sum_{i=0}^m \hat{\beta}_i I_{ki} = I_{kH}. \quad (8.16)$$

Отсюда вытекает, что

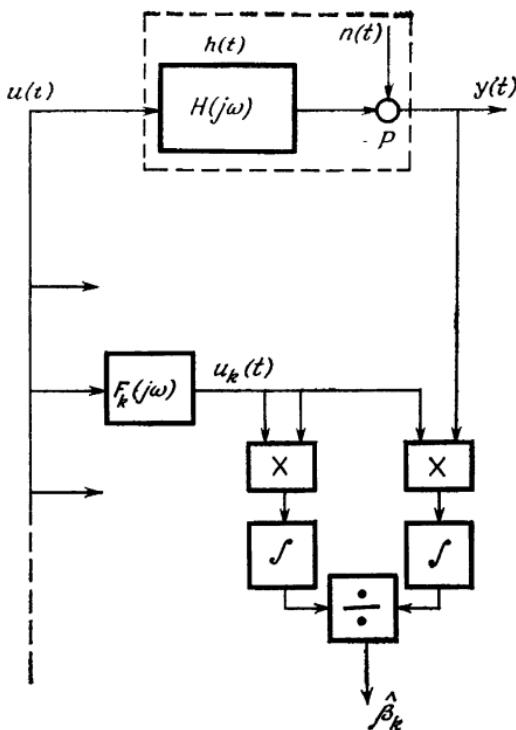
$$\hat{\beta}_k = \frac{1}{I_{kH}} \left\{ I_{kH} - \sum_{i \neq k} \hat{\beta}_i I_{ki} \right\}. \quad (8.17)$$

Следовательно, $\hat{\beta}_k$ зависит от всех других коэффициентов; после настройки каких-либо коэффициентов β_k нужно корректировать снова. Это неудобство устраняется выбором

$$\begin{aligned} I_{ki} = \frac{1}{2\pi} \int_{-\infty}^{\infty} \{F_k^*(j\omega) F_i(j\omega) + F_k(j\omega) F_i^*(j\omega)\} \Phi_{uu}(j\omega) d\omega = \\ = \delta_{ki} = \begin{cases} 1, & k=i, \\ 0, & k \neq i, \end{cases} \end{aligned} \quad (8.18)$$

т. е. F_k образуют ортонормальную систему относительно веса $\Phi_{uu}(j\omega)$. Тогда

$$\hat{\beta}_k = \frac{\int_{-\infty}^{\infty} \{F_k^*(j\omega) H(j\omega) + F_k(j\omega) H^*(j\omega)\} \Phi_{uu}(j\omega) d\omega}{2 \int_{-\infty}^{\infty} F_k^*(j\omega) F_k(j\omega) \Phi_{uu}(j\omega) d\omega}, \quad (8.19)$$



Ф и г. 8.6.

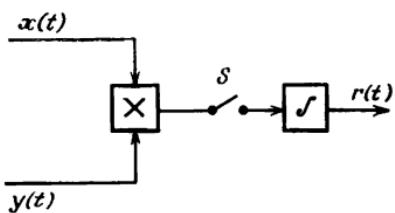
т. е. $\hat{\beta}_k$ определяется явным выражением, и зависимость между значениями разных коэффициентов отсутствует. Фильтры F_k должны быть ортонормальными с весом Φ_{uu} . Если u — белый шум, то функция $\Phi_{uu}(j\omega)$ постоянна и в качестве весовых функций фильтров F_k можно выбрать хорошо известные функции Лагерра (см. гл. 3 и фиг. 4.17).

Отметим, что числитель и знаменатель (8.19) не изменятся, если под интегралом добавить множитель $e^{j\omega t}$ при $t = 0$. Следовательно, эти интегралы можно интерпретировать как обратные преобразования Фурье, т. е. как корреляционные функции (при значении аргумента $\tau = 0$):

$$\hat{\beta}_k = \frac{\Psi_{uh} (0) + \Psi_{yu} (0)}{2\Psi_{uk} u_k (0)} = \frac{\Psi_{uh} (0)}{\Psi_{uk} u_k (0)},$$

$$k = 0, 1, \dots, m. \quad (8.20)$$

Эта процедура может быть реализована в соответствии с фиг. 8.6.



Фиг. 8.7.

Статистические ошибки корреляционных измерений. Операция усреднения по бесконечному интервалу времени является, конечно, чрезвычайно непрактичной. В реальных приложениях приходится ограничивать время корреляции конечным интервалом $[0, T]$. Следовательно, мы

вынуждены пользоваться оценками корреляционных функций $\tilde{\Psi}_{uu}(\tau, T)$ и $\tilde{\Psi}_{uy}(\tau, T)$. Эти экспериментальные корреляционные функции характеризуются отклонением (неопределенностью) относительно теоретических корреляционных функций.

При обсуждении этого вопроса воспользуемся фиг. 8.7, где x и y — эргодические стационарные случайные сигналы и ключ S при $t = 0$ закрыт. Начальное условие $r(0) = 0$. В приложении Г показано, что в этом случае

$$\mu_r(t) \stackrel{\Delta}{=} \mathbb{E}[r(t)] = t\Psi_{xy}(0) \quad (8.21)$$

и если совместное распределение x и y гауссовское, то

$$\sigma_r^2(t) \stackrel{\Delta}{=} \mathbb{E}[\{r(t) - \mu_r(t)\}^2] = 2 \int_0^t (t-v) R_i^2(v) dv, \quad (8.22)$$

где

$$R(v) = \Psi_{xx}(v) \Psi_{yy}(v) + \Psi_{xy}(v) \Psi_{yx}(v). \quad (8.23)$$

Сравнивая фиг. 8.3 и 8.7, можно заметить следующее соответствие:

$$x(t) \Rightarrow u(t-\tau),$$

$$y(t) \Rightarrow y(t) = x(t) + n(t).$$

Отсюда

$$\Psi_{xy}(0) \Rightarrow \Psi_{uy}(\tau) = \Psi_{ux}(v+\tau),$$

$$\Psi_{xx}(v) \Rightarrow \Psi_{uu}(v),$$

$$\Psi_{yy}(v) \Rightarrow \Psi_{yy}(v) = \Psi_{xx}(v) + \Psi_{nn}(v),$$

$$\Psi_{xy}(v) \Rightarrow \Psi_{uy}(v+\tau) = \Psi_{ux}(v+\tau),$$

$$\Psi_{yx}(v) \Rightarrow \Psi_{yu}(v-\tau) = \Psi_{xu}(v-\tau),$$

причем некоторые члены в этих выражениях исчезают, поскольку входной сигнал и шум предполагаются независимыми. В результате имеем

$$\mu_r(t) = t \Psi_{ux}(\tau). \quad (8.24)$$

Аддитивный независимый шум не дает вклада в математическое ожидание. Выражение для $R(v)$ можно записать в виде

$$R(v) = R_1(v) + R_2(v),$$

причем

$$R_1(v) = \Psi_{uu}(v) \Psi_{xx}(v) + \Psi_{ux}(v + \tau) \Psi_{xu}(v - \tau), \quad (8.25)$$

$$R_2(v) = \Psi_{uu}(v) \Psi_{nn}(v). \quad (8.26)$$

Составляющая R_1 присутствует, даже если аддитивного шума нет вообще; ее можно назвать *внутренней статистической неопределенностью*. Компонента R_2 представляет *неопределенность, связанную с аддитивным шумом*.

Это исследование статистических ошибок можно применить к экспериментально определяемым значениям корреляционной функции $\Psi_{ux}(\tau)$, оцениваемым на конечном интервале времени $[0, T]$:

$$\tilde{\Psi}_{ux}(\tau, T) = \frac{1}{T} \int_0^T u(t - \tau) x(t) dt. \quad (8.27)$$

Внутреннюю статистическую неопределенность можно охарактеризовать отношением сигнала к шуму S/N_1 на выходе коррелятора

$$S = \mathcal{E}[\tilde{\Psi}_{ux}(\tau, T)] = \frac{1}{T} \mu_r(T) = \Psi_{ux}(\tau), \quad (8.28)$$

$$\begin{aligned} N_1^2 &= \mathcal{E}[\tilde{\Psi}_{ux}^2(\tau, T)] - \Psi_{ux}^2(\tau) = \frac{1}{T^2} \sigma_r^2(t) = \\ &= \frac{2}{T^2} \int_0^T (T - v) R_1(v) dv. \end{aligned} \quad (8.29)$$

Таким образом, математическое ожидание выхода интегратора совпадает с истинным значением корреляционной функции независимо от длины интервала наблюдения T .

Конечно, точность (дисперсия или стандартное отклонение) найденной оценки зависит от длины этого интервала.

Выражение

$$N_1^2 = \frac{2}{T^2} \left\{ T \int_0^T R_1(v) dv - \int_0^T v R_1(v) dv \right\} \quad (8.30)$$

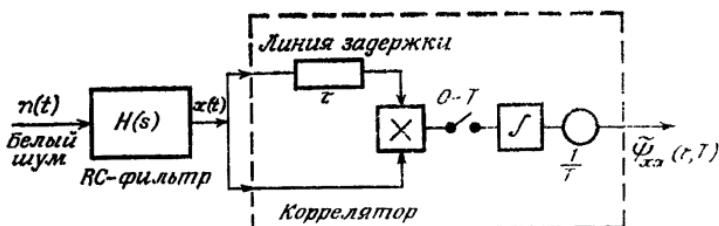
может быть упрощено при больших T , когда вторым членом в правой части можно пренебречь, т. е. когда $R_1(v)$ быстро стремится к нулю. При этом

$$\begin{aligned} N_1^2 &\approx \frac{2}{T} \int_0^T R_1(v) dv = \\ &= \frac{2}{T} \int_0^T \{ \Psi_{uu}(v) \Psi_{xx}(v) + \Psi_{ux}(v + \tau) \Psi_{xu}(v - \tau) \} dv. \end{aligned} \quad (8.31)$$

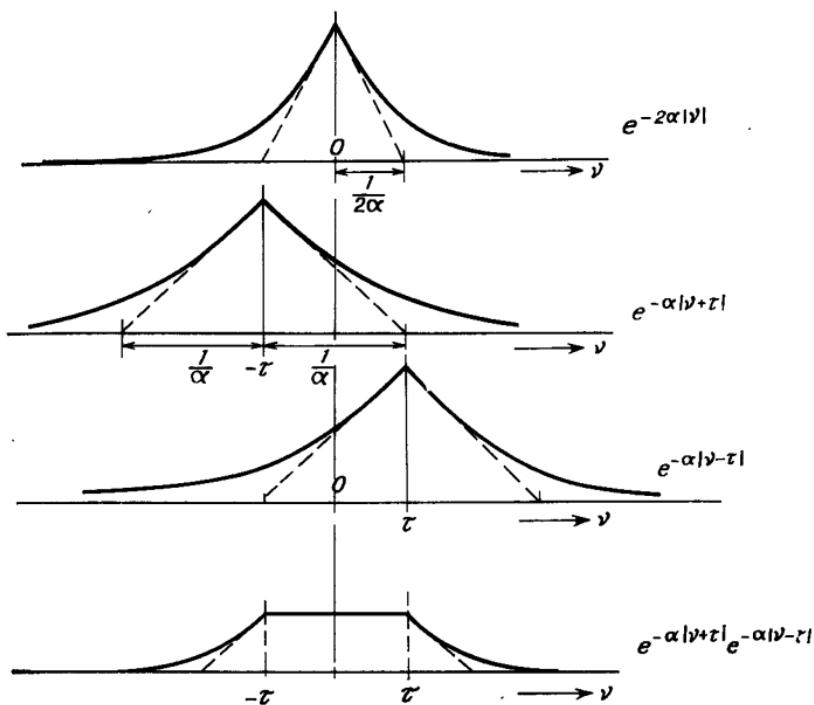
Если $\Psi_{ux}(\tau) \rightarrow 0$ при больших τ , то

$$N_1^2 \approx \frac{2}{T} \int_0^T \Psi_{uu}(v) \Psi_{xx}(v) dv. \quad (8.32)$$

Пример. Рассмотрим измерение автокорреляционной функции в соответствии с фиг. 8.8 при $u(t) = x(t)$. Белый шум $n(t)$ имеет спектральную плотность k^2 , передаточная функция RC -фильтра $H(s) = a/(s + a)$, спектральная плотность сигнала $x(t)$ есть $\Phi_{xx}(s) = k^2 a^2 / (a^2 - s^2)$, корреляционная функция $x(t)$ равна $\Psi_{xx}(\tau) = \Psi_{xx}(0)e^{-\alpha|\tau|}$,



Фиг. 8.8



Фиг. 8.9.

где $\Psi_{xx}(0) = ak^2/2$. Следовательно,

$$S = \Psi_{xx}(\tau) = \Psi_{xx}(0) e^{-\alpha|\tau|},$$

$$N_1^2 \approx \frac{2}{T^2} \int_0^T (T - \nu) R_1(\nu) d\nu,$$

где

$$\begin{aligned} R_1(\nu) &= \Psi_{xx}^2(\nu) + \Psi_{xx}(\nu + \tau) \Psi_{xx}(\nu - \tau) = \\ &= \Psi_{xx}^2(0) \{e^{-2\alpha|\nu|} + e^{-\alpha|\nu+\tau|} e^{-\alpha|\nu-\tau|}\}. \end{aligned}$$

Обращаясь к фиг. 8.9, видим, что

$$\begin{aligned} N_1^2 &= \Psi_{xx}^2(0) \times \\ &\times \frac{2\alpha T - 1 + 2e^{-2\alpha T} + [(2\alpha\tau + 1)(2\alpha T - 1) - 2(\alpha\tau)^2] e^{-2\alpha\tau}}{2(\alpha T)^2}. \quad (8.33) \end{aligned}$$

Введем безразмерные величины $\alpha\tau = \beta$ и $\alpha T = \gamma$; тогда при $\beta \geq 0$

$$S = \Psi_{xx}(0) e^{-\beta},$$

$$N_1^2 = \Psi_{xx}^2(0) \frac{2\gamma - 1 + 2e^{-2\gamma} + [(2\beta + 1)(2\gamma - 1) - 2\beta^2] e^{-2\beta}}{2\gamma^2}. \quad (8.34)$$

Рассмотрим частные случаи:

1) $\beta = 0$

$$N_1^2 = \Psi_{xx}^2(0) \frac{2\gamma - 1 + e^{-2\gamma}}{\gamma^2}, \quad (8.35)$$

$$N_1^2 \rightarrow \Psi_{xx}^2(0) \frac{2}{\gamma}, \quad \gamma \rightarrow \infty; \quad (8.36)$$

2) $\beta \rightarrow \infty$

$$N_1^2 \rightarrow \Psi_{xx}^2(0) \frac{1}{\gamma}, \quad \gamma \rightarrow \infty. \quad (8.37)$$

Следовательно, дисперсия измерений корреляционной функции обратно пропорциональна γ . Так как $\gamma = \alpha T$, это означает, что дисперсия уменьшается, если α (полоса пропускания фильтра) и (или) T (интервал наблюдения) возрастают.

В приведенном выше примере отношение сигнал/шум при длинных интервалах корреляции равно

$$\frac{S}{N_1} \approx e^{-\beta} \sqrt{\frac{\gamma}{2}}. \quad (8.38)$$

Можно задать, например, $N_1 = 0,05 S$, т. е. принять стандартное отклонение равным 5% от измеряемой величины. Отсюда при $\beta = 0$ следует, что $\gamma = \alpha T = 800$. Другими словами, корреляционный интервал должен в 800 раз превосходить постоянную времени рассматриваемого RC -фильтра. Однако этот достаточно длинный интервал приводит к небольшой точности даже при наиболее благоприятных значениях корреляционной функции ($\beta = 0$). При больших значениях $\beta = \alpha t$, т. е. при значениях корреляционной функции, соответствующих большим аргументам t , отношение S/N оказывается значительно меньше. Необходимость в длинных корреляционных интервалах возникает только из-за внутренней статистической неопределенности. Заметим, что отношение S/N не зависит от уровня входного сигнала коррелятора.

Неопределенность, связанная с возможным аддитивным шумом, накладывает еще более жесткие требования на время корреляции. Эта неопределенность имеет более простую природу и может быть выражена дополнительной «шумовой» компонентой в выходном сигнале коррелятора, а именно

$$N_2^2 = \frac{2}{T^2} \int_0^T (T - v) R_2(v) dv, \quad (8.39)$$

где, согласно формуле (8.26),*

$$R_2(v) = \Psi_{uu}(v) \Psi_{nn}(v). \quad (8.40)$$

Конечно, для этого вида ошибок отношение сигнала к шуму коррелятора S/N_2 линейно растет с увеличением уровня входного сигнала u . В качестве примера обратимся вновь к схеме на фиг. 8.8, предполагая теперь, что в один канал добавляется аддитивный шум. Пусть шум $n(t)$ и сигнал $x(t)$ имеют одинаковую ширину спектра, т.е.

$$\Psi_{nn}(\tau) = \Psi_{nn}(0) e^{-\alpha|\tau|}. \quad (8.41)$$

Отсюда

$$N_2^2 = \Psi_{xx}(0) \Psi_{nn}(0) \frac{2\gamma - 1 + e^{-2\gamma}}{2\gamma^2}, \quad (8.42)$$

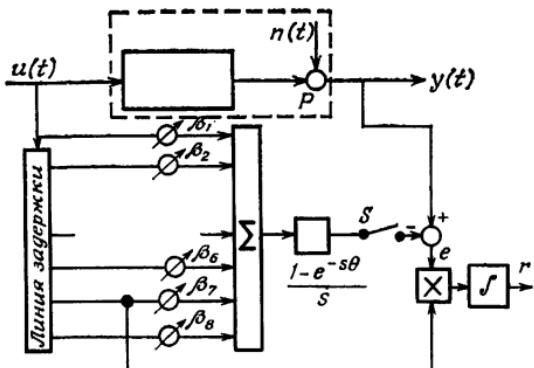
и при больших γ

$$N_2^2 \rightarrow \Psi_{xx}(0) \Psi_{nn}(0) \frac{1}{\gamma}. \quad (8.43)$$

Теперь отношение сигнала к шуму при больших интервалах корреляции оказывается равным

$$\frac{S}{N} = \frac{S}{\sqrt{N_1^2 + N_2^2}} \approx e^{-\beta} \sqrt{\frac{\gamma}{2 + [\Psi_{nn}(0)/\Psi_{xx}(0)]}}. \quad (8.44)$$

До сих пор рассматривалось определение лишь одного параметра, а именно одного значения весовой функции. Новые вопросы возникают при определении нескольких параметров. После оценивания одной величины ее оценка может считаться *aприорно известной* в следующем цикле оценивания и т. д. Интуитивно ясно, что использование априорной информации позволяет строить более эффективные процедуры оценивания. Однако решающим становится вопрос о том, как такая информация может быть



Ф и г. 8.10.

использована практически. Это осуществимо, если применять модель объекта, хорошо приспособленную для его параметрического описания, например, при исследовании весовой функции можно использовать линию задержки с отводами и потенциометрами. Фиг. 8.10 иллюстрирует этот случай [9].

Пример. Чтобы продемонстрировать особенности модели, соотношения (8.21) — (8.25) будут применены к достаточно простому примеру, а именно к объекту с весовой функцией

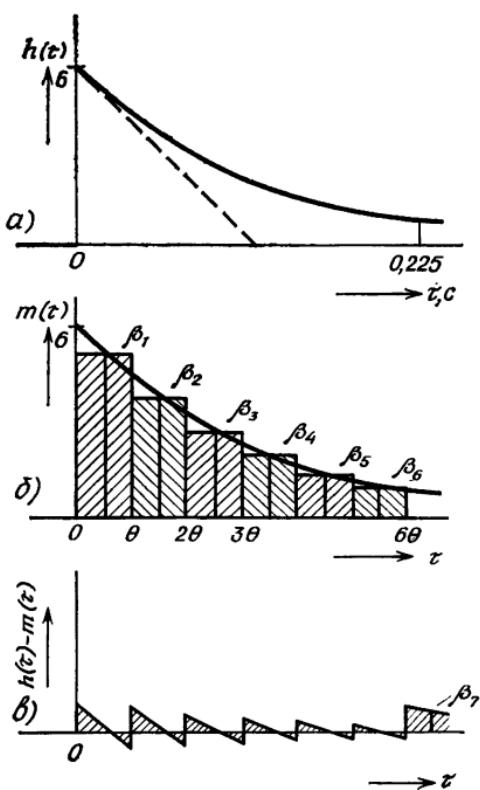
$$h(\tau) = \begin{cases} \sigma e^{-\sigma \tau} & \text{при } \tau \geq 0, \\ 0 & \text{при } \tau < 0. \end{cases}$$

Рассмотрим определение одного значения этой функции, например при $\tau = 0,225$ (фиг. 8.11, a), используя белый шум в качестве входного сигнала.

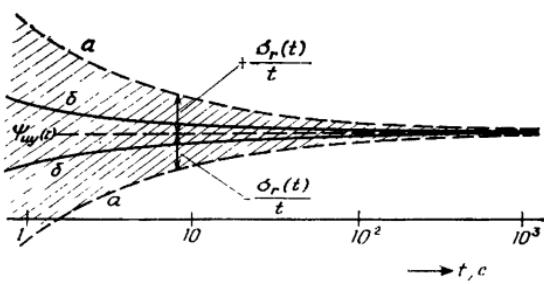
Если ключ S на фиг. 8.10 открыт, этот метод совпадает с корреляционной схемой на фиг. 8.3. В этом случае для стандартного отклонения относительно ожидаемого значения

$$\frac{1}{t} \mu_r = \Psi_{uy}(\tau) \quad \text{при } \tau = 0,225$$

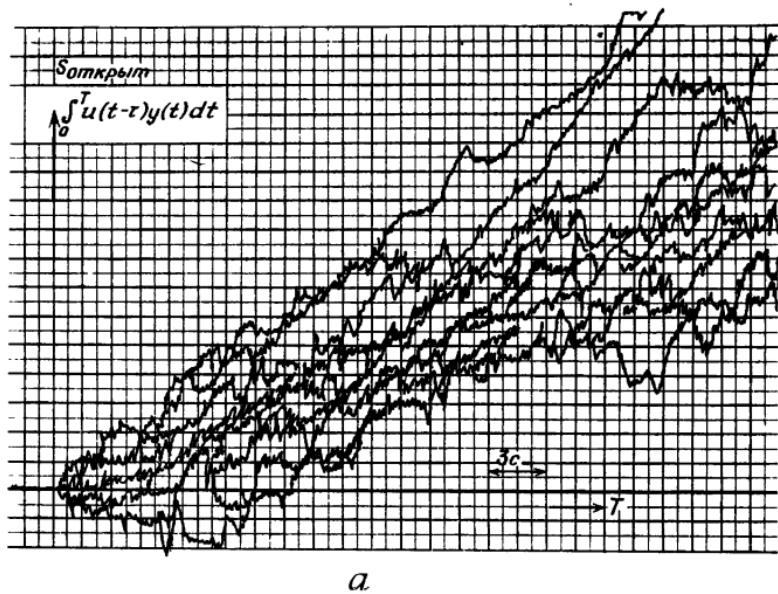
получим соотношение, представленное на фиг. 8.12 кривой a. Из этого графика можно определить корреляционный интервал, необходимый для достижения заданной точности (стандартного отклонения) оценки параметра.



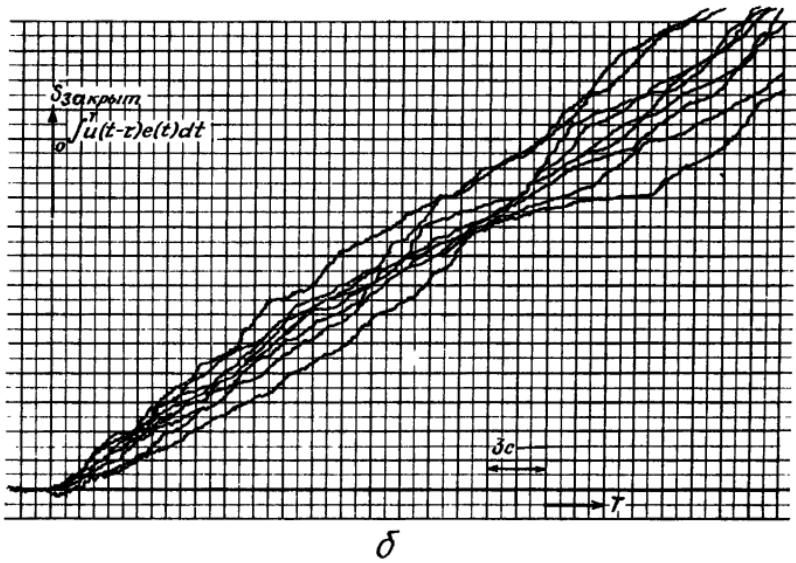
Ф и г. 8.11.



Ф и г. 8.12.



a



δ

Ф и г. 8.13.

Пусть теперь на фиг. 8.10 параметры от β_1 до β_6 уже оценены. Найденные значения используются для настройки соответствующих потенциометров; остальные потенциометры¹ устанавливаются на нуль, и ключ S закрыт. Настраиваемая часть модели, включая цепь фильтрации, имеет весовую функцию $m(\tau)$, изображенную на фиг. 8.11. Следовательно, сигнал ошибки e определяется разностью между $h(\tau)$ и $m(\tau)$ (фиг. 8.11, в); при этом в весовой функции $h(\tau) - m(\tau)$ подлежащий оценке параметр имеет свое исходное значение. Обращаясь теперь к формуле (8.25), можно заметить, что в данном случае индекс x должен быть заменен индексом e . В результате соответствующие члены уменьшаются по сравнению с первоначальными значениями, что приводит к уменьшению дисперсии (или сокращению корреляционного интервала, необходимого для достижения прежнего значения дисперсии). Для рассматриваемого примера сокращение корреляционного интервала представлено кривой b на фиг. 8.12; теперь то же самое значение стандартного отклонения достигается для корреляционного интервала, составляющего приблизительно одну десятую прежней величины.

На фиг. 8.13, а, б изображен ряд практических реализаций выходного сигнала интегратора $r(t)$ для указанных выше случаев; ключ S открыт либо закрыт. Флуктуации характеризуют внутреннюю статистическую неопределенность коррелятора при отсутствии и наличии модели. Влиянием аддитивного шума пренебрегают; его можно учесть с помощью формулы (8.26).

Дальнейшее обсуждение вопросов точности корреляционных методов можно найти в работах [3, 8, 13, 19, 20, 30]. Обширная литература посвящена применением корреляционных методов; только весьма малую ее часть можно привести в списке цитируемой (например, [1, 5, 14, 18, 27, 28, 37]) или дополнительной литературы. В настоящее время промышленность выпускает всего несколько типов корреляторов; некоторые из них оборудованы дополнительными устройствами, такими, как измерители плотности вероятностей. Другой распространенный тип дополнительных приспособлений — устройство для применения преобразования Фурье к корреляционным функциям

и получения спектральных плотностей.[†] Нередко корреляторы позволяют одновременно определять ряд значений корреляционной функции (например, 50 или 100) с помощью параллельных вычислений. Разрабатываются также корреляторы с непрерывно меняющейся временной задержкой τ и соответствующими корректирующими фильтрами [2].

8.2. КОРРЕЛЯЦИОННЫЕ МЕТОДЫ ДЛЯ КВАНТОВАННЫХ СИГНАЛОВ

Квантование амплитуды

Как указывалось ранее, операция квантования Q [7] аналоговых сигналов x и y является нелинейной, т. е. если

$$x_q = Q[x] \text{ и } y_q = Q[y], \quad (8.45)$$

то величины

$$\alpha x_q + \beta y_q \text{ и } Q[\alpha x + \beta y] \quad (8.46)$$

не обязательно равны между собой при всех x и y . Следовательно, мы неизбежно столкнемся с проблемами, характерными для нелинейных систем и связанными с тем, что для этих систем неприменим принцип суперпозиции.

Исследование операции квантования и ее применения в корреляционном анализе можно найти в статьях [36, 38]. Следующие результаты заимствованы из работы [36].

Аналоговый входной сигнал x преобразуется квантователем (аналого-цифровым преобразователем) с постоянным шагом в квантованный сигнал x_q . Характеристики зависимости входа и выхода для этого квантователя могут быть, например, такими, как на фиг. 8.14, а. Такой квантователь Q можно образовать из единичных квантователей Q_1 с характеристиками, показанными на фиг. 8.14, б, по схеме фиг. 8.15. Соответствующие соотношения имеют вид

$$\text{усиление: } x_1 = \frac{x}{q}, \quad (8.47)$$

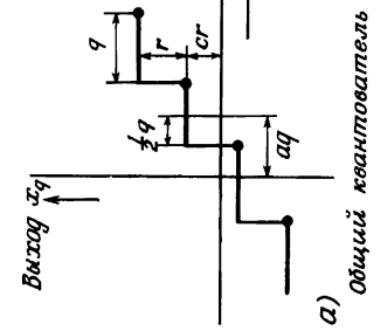
$$\text{сдвиг: } x_2 = x_1 - a, \quad (8.48)$$

$$\text{единичный квантователь: } x_3 = i \text{ при } \left(i - \frac{1}{2} \right) <$$

$$< x_2 \leqslant \left(i + \frac{1}{2} \right), \quad (8.49)$$

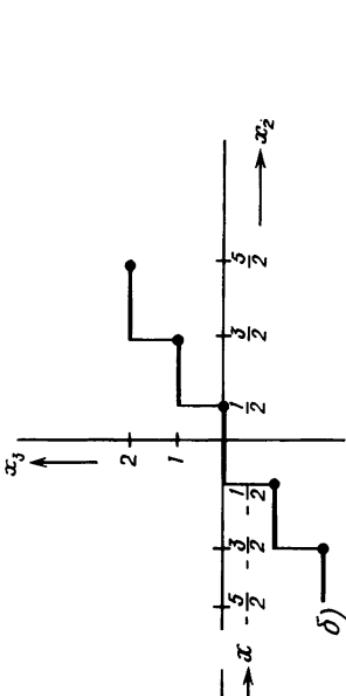
$$\text{сдвиг: } x_4 = x_3 + i, \quad (8.50)$$

$$\text{усиление: } x_q = rx_4. \quad (8.51)$$



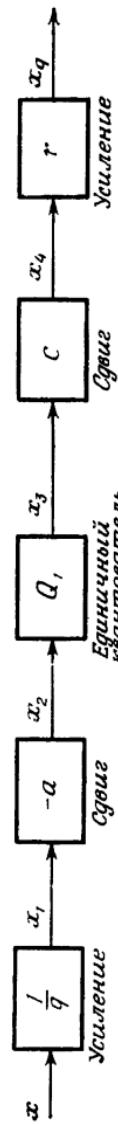
a) Общий квантователь Q_1

Ф и г. 8.14.



b) Единичный квантователь Q_1

Ф и г. 8.14.



Единичный квантователь

Умножение

Сумма

Сумма

Умножение

Ф и г. 8.15.

В работе [36] показано, что связь между плотностями распределения входа $p(x)$ и выхода $p_q(x_q)$ задается формулой

$$p_q(x_q) = \frac{1}{r} \sum_{i=-\infty}^{\infty} \left[\int_{-\infty}^{\infty} p(x) f_1 \left\{ \frac{x_q}{r} - c - \left(\frac{x}{q} - a \right) \right\} dx \right] \times \\ \times \delta \left(\frac{x_q}{r} - c - i \right), \quad (8.52)$$

где $\delta(x)$ — δ -функция Дирака и

$$f_1(x) = \begin{cases} 1 & \text{при } -\frac{1}{2} < x \leq \frac{1}{2}, \\ 0 & \text{в противном случае.} \end{cases}$$

При $a = c = 0$, $q = r = 1$ формула (8.52) сводится к

$$p_q(x_q) = \sum_{i=-\infty}^{\infty} \left[\int_{-\infty}^{\infty} p(x) f_1(x_q - x) dx \right] \delta(x_q - i), \quad (8.53)$$

т. е. к «дискретизированной» в соответствии с областью значений x_q свертке плотности распределения с функцией $f_1(x)$.

Для дальнейшего анализа и облегчения выкладок имеет смысл применить преобразование Фурье, при котором операция свертки заменяется умножением; другими словами, вводится характеристическая функция

$$P_{x_q}(s) = \mathcal{E}[\exp(jsx_q)] = \int p_q(x_q) \exp(jsx_q) dx_q, \quad (8.54)$$

так что

$$P_{x_q y_q}(s_1, s_2) = \mathcal{E}[\exp\{j(s_1 x_q + s_2 y_q)\}] = \\ = \int p_q(x_q y_q) \exp\{j(s_1 x_q + s_2 y_q)\} dx_q dy_q. \quad (8.55)$$

Дифференцируя характеристическую функцию по s_1 и s_2 , можно определить момент

$$\mathcal{E}[x_q y_q] = (j)^2 \frac{\partial^2 P_{x_q y_q}(s_1, s_2)}{\partial s_1 \partial s_2} \Big|_{s_1=s_2=0}, \quad (8.56)$$

представляющий коэффициент корреляции соответствующих сигналов. Таким образом можно определить коэф-

фициент корреляции для коррелятора-умножителя. Полученные результаты можно интерпретировать, рассматривая квантованный сигнал как суммы исходного сигнала с аддитивным шумом квантования:

$$x_q = x + n_x, \quad y_q = y + n_y, \quad (8.57)$$

$$\mathcal{E}[x_q y_q] = \mathcal{E}[xy] + \mathcal{E}[n_x y] + \mathcal{E}[x n_y] + \mathcal{E}[n_x n_y]. \quad (8.58)$$

Частные случаи этих соотношений описывают разные типы корреляторов, в том числе использующие экстремальное квантование (двоичные сигналы).

Типы корреляторов

Из предыдущего изложения ясно, что операцию квантования можно охарактеризовать следующими параметрами: a — сдвиг по оси x , c — сдвиг по оси x_q , q — величина шага по оси x , r — величина шага по оси x_q . Частные случаи:

$a = c = 0, \quad q = r = 1$ — единичный квантователь,

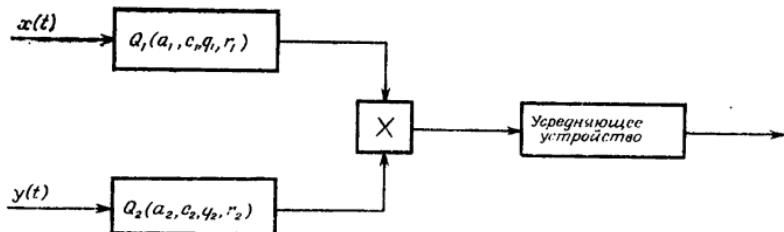
$a = c = 0, \quad q = r \rightarrow 0$ — аналоговый канал (квантование отсутствует),

$a = c = 0, \quad q = r$ — квантователь (симметричный),

$a = \frac{1}{2}, \quad c = \frac{1}{2}, \quad q \rightarrow \infty, \quad r = 2$ — двоичный квантователь.

В зависимости от типа квантования, применяемого в одном или обоих каналах коррелятора, различают (фиг. 8.16):

- 1) аналоговый коррелятор;
- 2) цифровой коррелятор;



Ф и г. 8.16.

- 3) коррелятор совпадения знаков;
 4) релейный (с круговой модуляцией) коррелятор;
 5) коррелятор Стильеса.

Таблица 8.1

		Канал y		
		двоичный	квантованный	аналоговый
Канал x	аналоговый	4)	5)	1)
	квантованный		2)	
	двоичный	3)		

Сравнение преимуществ и недостатков этих корреляторов проведено Уоттсом [36]. Здесь уместны следующие замечания.

1) **Аналоговый коррелятор.** Для этого типа коррелятора $x_q = x$ и $y_q = y$, т. е. шум квантования отсутствует и

$$\mathcal{E}[x_q y_q] = \mathcal{E}[xy] = \mathcal{E}[\tilde{\Psi}_{xy}(0, T)] = \Psi_{xy}(0), \quad (8.59)$$

где, как и раньше,

$$\Psi_{xy}(0, T) = \frac{1}{T} \int_0^T x(t) y(t) dt. \quad (8.60)$$

Подробно этот случай рассмотрен в разд. 8.1.

2) **Цифровой коррелятор.** Сигналы в обоих каналах квантованы. Оказывается, что даже при сравнительно грубом квантовании существует приемлемое соответствие между $\mathcal{E}[x_q y_q]$ и $\mathcal{E}[xy]$. Оценку погрешности можно получить из уравнения (8.58). Наихудший случай имеет место при измерении автокорреляционной функции. Тогда

$$\mathcal{E}[n_x n_y] = \mathcal{E}[n_x^2].$$

Пример. Пусть распределение вероятностей $p(n_x) = q^{-1}$ на интервале $(-q/2, q/2)$ около каждого уровня квантования равномерное. Тогда рассматриваемый второй момент равен

$$\mathcal{E}[n_x^2] = \int_{-q/2}^{q/2} n_x^2 p(n_x) dn_x = \frac{1}{3q} n_x^3 |_{-q/2}^{q/2} = q^2/12.$$

Он быстро убывает с уменьшением q . Если диапазон изменения амплитуды сигнала $(-3\sigma, 3\sigma)$ квантуется с постоянным шагом на 16 интервалов, то $q = 6\sigma/16 = 3\sigma/8$ и

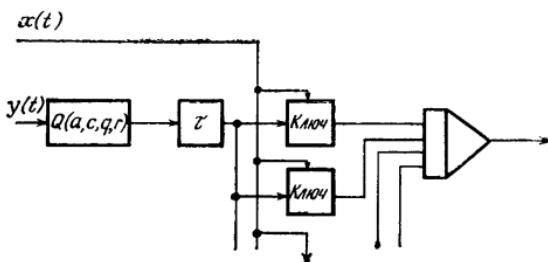
$$\mathcal{E}[n_x^2] = \frac{1}{12} \frac{9\sigma^2}{64} \approx 0,012\sigma^2.$$

Значит, для цифровых корреляторов иногда достаточно длины слова в 3 или 4 бита. В работе [7] исследовано использование дополнительного случайного сигнала для устранения смещения выхода коррелятора этого типа, т. е. для достижения относительно высокой точности при малом числе уровней квантования (см. также разд. 8.3).

3) **Коррелятор совпадения знаков.** См. разд. 8.3.

4) **Релейный коррелятор, или коррелятор с круговой модуляцией.** В том случае умножение осуществляется очень просто, а именно как $x \operatorname{sgn} y$, где $\operatorname{sgn} y = 1$ при $y \geq 0$ и $\operatorname{sgn} y = -1$ при $y < 0$; эта операция реализуется механическими или электронными реле. Разумеется, при этом теряется информация об амплитуде y . Для интерпретации результатов измерений нужно знать плотности вероятностей входных сигналов.

5) **Коррелятор Стильеса.** Этот метод предложил Уоттс [36]. Само название объясняется тем, что в корреляционном интеграле для аналогового сигнала x и квантованного сигнала y легко узнать интеграл Стильеса. Этот коррелятор представляет интерес как для теории, так и с точки зрения его практической реализации. Несмотря на довольно грубое квантование, скажем 4 или 5 уровней, ошибка (ее ожидаемое значение) измеренной корреляционной функции оказывается достаточно малой. Фиг. 8.17 иллюстрирует способ реализации коррелятора на этом



Ф и г. 8.17.

принципе. Этот принцип заслуживает внимания при вычислениях корреляционных функций на цифровых машинах.

8.3. КОРРЕЛЯЦИОННЫЕ МЕТОДЫ ДЛЯ ДВОИЧНЫХ СИГНАЛОВ

Ранее было установлено [10], что корреляционные измерения для сигналов с гауссовским совместным распределением можно производить, используя их двузначный (двоичный) вариант. При этом для построения коррелятора можно использовать цифровые средства. Когда учитываются только полярности сигналов, этот принцип называют *корреляцией совпадения знаков*.

Пусть сигналы $x(t - \tau)$ и $y(t)$ имеют гауссовское совместное распределение вероятностей со средними значениями $E[x] = E[y] = 0$, т. е.

$$p(x, y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp \left[\frac{-1}{2(1-\rho^2)} \left\{ \left(\frac{x}{\sigma_x} \right)^2 - 2\rho \left(\frac{x}{\sigma_x} \right) \left(\frac{y}{\sigma_y} \right) + \left(\frac{y}{\sigma_y} \right)^2 \right\} \right], \quad (8.61)$$

где $\rho = \rho(\tau)$. Корреляционная функция определяется формулой

$$\Psi_{xy}(\tau) = E[xy] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy p(x, y) dx dy. \quad (8.62)$$

Для плотности распределения (8.61) она сводится к

$$\Psi_{xy}(\tau) = \sigma_x \sigma_y \rho(\tau). \quad (8.63)$$

Используя двоичные сигналы $\operatorname{sgn} x$ и $\operatorname{sgn} y$ вместо исходных сигналов, получаем

$$\begin{aligned}\Psi'_{xy}(\tau) &= \mathcal{E} [\operatorname{sgn} x \operatorname{sgn} y] = \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \operatorname{sgn} x \operatorname{sgn} y p(x, y) dx dy = \\ &= \underbrace{\int_0^{\infty} \int_0^{\infty} \dots}_{\geq 0} + \underbrace{\int_{-\infty}^0 \int_{-\infty}^0 \dots}_{\leq 0} + \underbrace{\int_0^{\infty} \int_{-\infty}^0 \dots}_{\leq 0} + \underbrace{\int_{-\infty}^0 \int_0^{\infty} \dots}_{\geq 0} \dots . \quad (8.64)\end{aligned}$$

Подставляя сюда выражение (8.61), находим

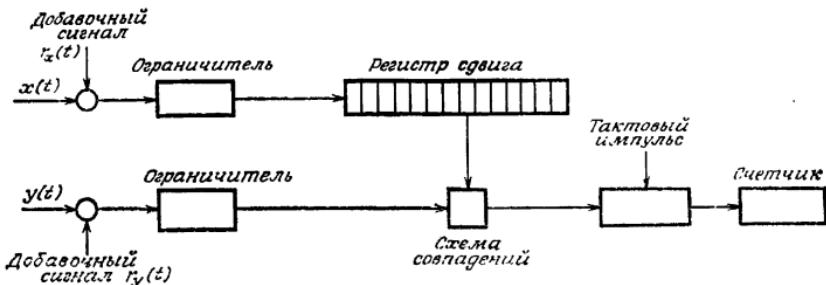
$$\Psi'_{xy}(\tau) = \frac{2}{\pi} \arcsin \rho(\tau). \quad (8.65)$$

Сравнивая Ψ и Ψ' из формул (8.63) и (8.65), замечаем, что:

1) как и следовало ожидать, в результате операции усечения из корреляционной функции исчезают мощности сигналов;

2) Ψ' монотонно зависит от $\rho(\tau)$, поскольку $-1 \leq \rho \leq 1$. Можно показать, что дисперсия этой корреляционной функции ненамного хуже, чем для нормированной корреляционной функции, вычисленной аналоговым коррелятором [34]. Интересно, что выражение (8.65) можно получить для более широкого класса плотностей распределения, например для плотностей с эллиптическими линиями уровня (распределения с эллиптической симметрией; доказательство и обсуждения см. в [21]).

Использование дополнительных сигналов. Приведенные выше рассуждения справедливы для сигналов с гауссовским распределением. На практике распределение амплитуд может быть неизвестно. В этих случаях соотношение между Ψ' и Ψ неизвестно. Был сделан ряд попыток (линеаризация) избавиться от этого недостатка корреляции совпадения знаков. Все они основаны на добавлении вспомогательных сигналов к входным сигналам коррелятора (фиг. 8.18). Анализ коррелятора этого типа обнаруживает хорошее соответствие между истинной и измеренной корреляционными функциями, несмотря на то что осуществляется перемножение знаков. В качестве



Ф и г. 8.18.

вспомогательных сигналов используются:

а) Случайные числа, которые являются реализациями двух независимых случайных величин с равномерным распределением [34]; применяются и другие распределения [7].

б) Пилообразные случайные сигналы, которые представляют собой реализации двух независимых процессов с равномерным распределением [15].

в) Детерминированные функции, в некотором смысле независимые на конечном интервале времени и сохраняющие это свойство при любых временных сдвигах (инвариантно относительно сдвига независимые функции) [22].

Отличие последнего метода от упомянутых в пунктах «а» и «б» состоит в использовании регулярных, а не случайных дополнительных сигналов. Рассматриваемое здесь свойство независимости определяется так: функции $\{f_i(t)\}$, $i = 1, 2, \dots, n$, являются инвариантно относительно сдвига независимыми в интервале $(0, T)$ тогда и только тогда, когда для любых τ_1, \dots, τ_n и любых положительных целых k_1, \dots, k_n

$$\frac{1}{T} \int_0^T f_1^{k_1}(t + \tau_1) \dots f_n^{k_n}(t + \tau_n) dt = \\ = \frac{1}{T} \prod_{s=1}^n \int_0^T f_s^{k_s}(t + \tau_s) dt. \quad (8.66)$$

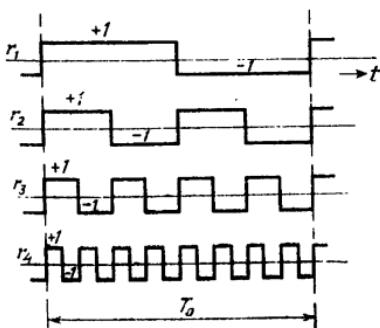
Практический интерес представляет обладающее этим свойством независимости множество двоичных функций

Радемахера

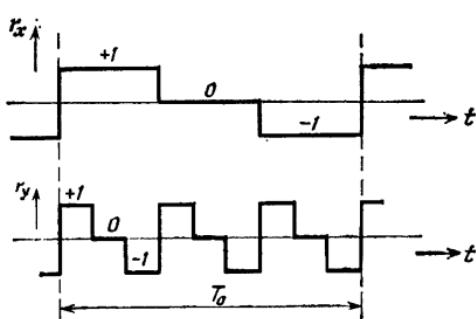
$$r_i(t) = \operatorname{sgn} \left\{ \sin \frac{2^i \pi t}{T_0} \right\}, \quad i = 1, 2, 3, \dots . \quad (8.67)$$

Они образуют ортонормальную систему. Множество этих функций полно, что, впрочем, несущественно для рассматриваемой задачи.

На фиг. 8.19 показаны первые четыре функции Радемахера. Их может генерировать двоичный счетчик, возбуждаемый периодическими импульсами. Линейные комбинации этих функций также независимы инвариантно к сдвигу, если они не содержат общей функции Радемахера. Таким способом можно формировать сигналы более чем с двумя уровнями. На фиг. 8.20 приведены графики двух таких трехуровневых сигналов, обладающих необходимым свойством независимости. Сигналы подобного типа добавляются к входным сигналам коррелятора, как показано на фиг. 8.18. Суммарные сигналы подаются на ограничители. Можно показать, что корреляционная функция этих двоичных сигналов совпадает с корреляционной функцией квантованных входных сигналов (см. разд. 8.2). Следовательно, применим и весь анализ погрешностей оценки корреляционной функции. Благодаря периодичности этих вспомогательных сигналов они обладают определенными преимуществами перед случайными вспомогательными сигналами, так как их вклад в выходной шум коррелятора (если его измерять на целом числе периодов) меньше. В работе [22] приведены значения отношения сигнала к шуму для нескольких типов корреляторов.



Фиг. 8.19.



Фиг. 8.20.

8.4. ДИФФЕРЕНЦИАЛЬНАЯ АППРОКСИМАЦИЯ И СВЯЗАННЫЕ С НЕЙ МЕТОДЫ

Явные методы идентификации по существу определяются видом операторов, применяемых к входным и выходным сигналам, в результате чего получается система уравнений для неизвестных параметров. Иллюстрацией могут служить уравнения (8.11) и (8.16). Нетрудно указать целый ряд таких операторов.

1) В предыдущих разделах исследовались операторы следующих типов:

$$\mathcal{O}_1 [y(t)] = \frac{1}{T} \int_0^T y(t) u(t-\tau) dt,$$

$$\mathcal{O}_2 [y(t)] = \frac{1}{T} \int_0^T y(t) u_q(t-\tau) dt,$$

$$\mathcal{O}_3 [y(t)] = \frac{1}{T} \int_0^T y(t) \operatorname{sgn} u(t-\tau) dt,$$

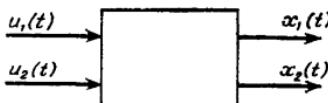
причем возможно также и квантование функции времени $y(t)$. Полученные в результате оценки корреляционной функции позволяют определить весовую функцию объекта.

2) **Метод производных.** Этот метод можно рассматривать как прямое обобщение корреляционных методов. Он известен также под названием дифференциальной аппроксимации. Рассмотрим ситуацию, схематически изображенную на фиг. 8.21. Предполагается, что все переменные состояния $x(t)$ и все входные переменные $u(t)$ измеримы и шум отсутствует. Динамика объекта описывается уравнением

$$\dot{x} = f(x, u, b). \quad (8.68)$$

Для оценки вектора неизвестных параметров b получаем

$$\dot{x} - f(x, u, \beta) = e. \quad (8.69)$$



Ф и г. 8.21.

Критерием опять является условие минимума $E(\beta)$, квадратичного по е функционала, например

$$E(\beta) = \int_0^T (e_1^2 + e_2^2) dt,$$

или в общем виде

$$E(\beta) = \int_0^T \| \dot{x} - f(x, u, \beta) \|^2 dt. \quad (8.70)$$

Необходимое условие минимума

$$\nabla_{\beta} E(\beta) |_{\beta=\hat{\beta}} = 0 \quad (8.71)$$

принимает в данном случае вид

$$\int_0^T \frac{\partial f'}{\partial \beta} \dot{x} dt = \int_0^T \frac{\partial f'}{\partial \beta} f(x, u, \hat{\beta}) dt, \quad (8.72)$$

где

$$\frac{\partial f'}{\partial \beta} = \left[\frac{\partial f_i}{\partial \beta_j} \right] = \begin{bmatrix} \frac{\partial f'}{\partial \beta_0} \\ \vdots \\ \frac{\partial f'}{\partial \beta_m} \end{bmatrix} = \begin{bmatrix} \frac{\partial f_1}{\partial \beta_0} & \cdots & \frac{\partial f_n}{\partial \beta_0} \\ \vdots & & \vdots \\ \frac{\partial f_1}{\partial \beta_m} & \cdots & \frac{\partial f_n}{\partial \beta_m} \end{bmatrix}. \quad (8.73)$$

Из системы уравнений (8.72) нужно определить неизвестные параметры $\hat{\beta}$. Если f линейна по β , сделать это совсем несложно. Подробности читатель может найти в дополнительной литературе по этому вопросу, приведенной в конце главы. Отметим, что использование описания системы в пространстве состояний вовсе не является неизбежным; кроме того, нелинейность объекта не обязательно

приводит к дополнительным осложнениям. (Как указывалось в гл. 4, принципиальное значение имеет линейность относительно параметров.)

Пример. Иллюстрацией может служить следующий простой случай. Объект описывается уравнением

$$\ddot{x} + a_1 \dot{x} + a_0 x^3 = bu, \quad (8.74)$$

а модель задается уравнением

$$\ddot{x} + a_1 \dot{x} + a_0 x^3 - \beta u = e. \quad (8.75)$$

Следовательно,

$$\frac{\partial e}{\partial a_1} = \dot{x}, \quad \frac{\partial e}{\partial a_0} = x^3, \quad \frac{\partial e}{\partial \beta} = -u. \quad (8.76)$$

Использование критерия $\min \int_0^T e^2 dt$ приводит к системе уравнений, линейных относительно неизвестных параметров:

$$\int_0^T e \dot{x} dt = 0, \quad \int_0^T e x^3 dt = 0, \quad \int_0^T e u dt = 0, \quad (8.77)$$

или

$$\begin{aligned} \alpha_1 \int_0^T \dot{x}^2 dt + \alpha_0 \int_0^T x^3 \dot{x} dt - \beta \int_0^T u \dot{x} dt &= - \int_0^T \ddot{x} \dot{x} dt, \\ \alpha_1 \int_0^T \dot{x} x^3 dt + \alpha_0 \int_0^T x^6 dt - \beta \int_0^T u x^3 dt &= - \int_0^T \ddot{x} x^3 dt, \\ \alpha_1 \int_0^T \dot{x} u dt + \alpha_0 \int_0^T x^3 u dt - \beta \int_0^T u^2 dt &= - \int_0^T \ddot{x} u dt, \end{aligned} \quad (8.78)$$

откуда можно определить оценки параметров.

До сих пор предполагалось, что x измеряется без помех. При наблюдениях зашумленного выхода, например $y(t) = x(t) + n(t)$, можно использовать уравнения (8.78),

заменяя x и \dot{x} на y и \dot{y} . Читатель должен заметить тесную связь с концепцией обобщенной модели. Благодаря этой связи ясно, что и в этом случае аддитивный шум приводит к смещенным оценкам параметров. Это смещение можно предотвратить, используя так называемый метод вспомогательных переменных (см. разд. 9.4). При этом получаются уравнения

$$\int_0^T eg_1 dt = 0, \quad \int_0^T eg_2 dt = 0, \quad \int_0^T eg_3 dt = 0, \quad (8.79)$$

в которых g_1, g_2, g_3 — вспомогательные переменные, т. е. функции времени, зависящие от $u(t)$, но не искажаемы шумом.

3) Метод преобразования Лапласа. В этом случае оператор определяется как

$$\mathcal{O}_L [y(t)] = \int_0^\infty y(t) e^{-st} dt.$$

Вычисления могут быть проведены для ряда значений s_i переменной s . При этом снова получается система уравнений относительно неизвестных параметров.

Пример. Снова рассмотрим модель (8.75):

$$\mathcal{O}_L [\ddot{x} + \alpha_1 \dot{x} + \alpha_0 x^3 - \beta u] = 0. \quad (8.80)$$

Для простоты примем нулевые начальные условия. Учитывая свойства оператора Лапласа \mathcal{L} , получим

$$\begin{aligned} s_i^2 \int_0^\infty x(t) e^{-s_i t} dt + \alpha_1 s_i \int_0^\infty x(t) e^{-s_i t} dt + \\ + \alpha_0 \int_0^\infty x^3(t) e^{-s_i t} dt = \beta \int_0^\infty u e^{-s_i t} dt. \end{aligned} \quad (8.81)$$

Используя три разных значения s_1, s_2, s_3 , можно получить три уравнения для $\alpha_1, \alpha_0, \beta$, которые решаются обращением соответствующей матрицы. Конечно, можно взять

более чем три неравных значения s и найти из получающейся системы уравнений оценки параметров по методу наименьших квадратов.

Отметим достоинство этого метода: функции времени не дифференцируются. При необходимости можно оценить и начальные условия.

4) **Метод преобразования Фурье.** Этот метод аналогичен предыдущему; оператор определяется как

$$[\mathcal{O}_F[y(t)]] = \int_0^{\infty} y(t) e^{-j\omega t} dt. \quad (8.82)$$

Отметим, что здесь возникают вопросы сходимости или ошибки усечения из-за того, что $|\exp[-j\omega t]| = 1$ для всех ω, t .

Шинброт [29] предложил название *функции метода* для используемых в предыдущих трех методах функций времени, т. е.

$$[x, x^3, u], \exp[-s_i t], \exp[-j\omega_i t],$$

и название *метод уравнений движения* для следующей процедуры оценивания:

а) умножение уравнения модели на функции, соответствующие выбранному методу;

б) интегрирование;

в) решение полученных алгебраических уравнений для неизвестных параметров.

Функции метода могут быть:

а) зависящими от входного и выходного сигналов объекта (например, дифференциальная аппроксимация);

б) зависящими только от входного сигнала объекта (например, корреляционные методы, метод вспомогательных переменных);

в) не зависящими от входных и выходных сигналов объекта (например, метод преобразований Фурье и Лапласа).

5) **Метод уравнений движения.** Рассмотрим в общем виде функции метода $g_i(t)$, не зависящие от входного и выходного сигналов объекта. При наличии подобной

независимости будем называть эти функции *модулирующими*. С помощью интегрирования по частям получаем

$$\begin{aligned} \int_0^\infty \dot{x} g_i dt &= -x(0) g_i(0) - \int_0^\infty x \dot{g}_i dt, \\ \int_0^\infty \ddot{x} g_i dt &= -\dot{x}(0) g_i(0) + x(0) \dot{g}_i(0) + \int_0^\infty x \ddot{g}_i dt. \end{aligned} \quad (8.83)$$

Эти соотношения можно подставить в уравнение модели. Приятно, что при этом не нужно дифференцировать экспериментальные реализации. Недостатком является то, что числовые значения $x(0)$ и $\dot{x}(0)$ приобретают слишком большое значение. Этот нежелательный эффект предотвращается выбором таких модулирующих функций, у которых $g_i(0) = \dot{g}_i(0) = 0$.

На практике ограничиваются конечными интервалами наблюдений. Те же рассуждения применимы и к правому концу такого конечного интервала $0 \leq t \leq T$. Следовательно, модулирующие функции должны выбираться такими, чтобы $g_i(T) = \dot{g}_i(T) = 0$. К функциям, которые (приближенно) удовлетворяют этим ограничениям, принадлежит и функция Гаусса.

Шинброт [29] показал, что этот подход применим и к нелинейным дифференциальным уравнениям вида

$$\ddot{x} + f(x) \dot{x} + g(x) x = 0, \quad (8.84)$$

где для $f(x)$ и $g(x)$ можно записать приближенные соотношения

$$\begin{aligned} f(x) &= a_{10} + a_{11}x + a_{12}x^2 + \dots, \\ g(x) &= a_{00} + a_{01}x + a_{02}x^2 + \dots \end{aligned}$$

Числовые значения этих коэффициентов можно оценить с помощью рассмотренных схем идентификации.

6) **Метод Стрейца.** Близкий к описанным выше методам подход развит Стрейцем [33], который рассмотрел задачу идентификации объекта по реализациям входных и выходных сигналов при отсутствии шумов. Пусть дифферен-

циальное уравнение объекта имеет вид

$$a_2 \ddot{y} + a_1 \dot{y} + a_0 y = b_2 \ddot{u} + b_1 \dot{u} + u, \quad (8.85)$$

где u и y — соответственно входной и выходной сигналы. Известны реализации сигналов на интервале (t_1, t_2) . Метод основан на повторном интегрировании соотношения (8.85) по интервалу (t_1, t_2) . После первого интегрирования получается

$$\begin{aligned} a_2 [\dot{y}(t_2) - \dot{y}(t)] + a_1 [y(t_2) - y(t)] + a_0 \int_t^{t_2} y(t) dt = \\ = b_2 [\dot{u}(t_2) - \dot{u}(t)] + b_1 [u(t_2) - u(t)] + \int_t^{t_2} u(t) dt. \end{aligned} \quad (8.86)$$

Повторное интегрирование дает

$$\begin{aligned} a_2 [y(t_2) \{t_2 - t\} - y(t_2) + y(t)] + a_1 [\dots] + a_0 [\dots] = \\ = b_2 [\dots] + b_1 [\dots] + [\dots] \end{aligned} \quad (8.87)$$

и, наконец,

$$\begin{aligned} a_2 [y(t_2) \{t_2(t_2 - t_1) - \frac{1}{2}(t_2^2 - t_1^2)\} - y(t_2) \{t_2 - t\}] + \\ + \int_t^{t_2} y(t) dt + a_1 [\dots] + a_0 [\dots] = \dots . \end{aligned} \quad (8.88)$$

Полагая $t = t_1$ в каждом из этих уравнений, можно оценить выражения в квадратных скобках по реализациям сигналов на (t_1, t_2) . Это можно сделать численными или графическими методами. Таким образом, опять получается система линейных уравнений для неизвестных коэффициентов:

$$\sum_{i=0}^2 a_i Y_{ij} = \sum_{i=0}^2 b_i U_{ij}, \quad b_0 = 1 \quad (8.89)$$

(индекс j обозначает номер уравнения).

Необходимо иметь столько независимых уравнений, сколько неизвестных, в данном случае — пять. Это число уравнений можно получить дальнейшим интегрированием.

Из соображений точности, однако, лучше иметь больше реализаций сигналов и интегрировать не более двух раз.

Удобнее всего определять $\dot{u}(t_1)$, $\dot{u}(t_2)$, $\dot{y}(t_1)$, $\dot{y}(t_2)$ в точках экстремума или перегиба реализаций, где некоторые из этих производных равны нулю. Эти соображения могут служить основой для выбора интервала (интервалов) (t_1, t_2) .

Сравнительно простой случай [31, 32] имеет место, когда объект описывается уравнением

$$a_n \dot{y}^{(n)} + a_{n-1} \dot{y}^{(n-1)} + \dots + a_1 \dot{y} + a_0 y = u, \quad (8.90)$$

а вход и выход постоянны при $t < 0$ и $t \rightarrow \infty$, так что

$$\dot{y}(0) = \ddot{y}(0) = \dots = y^{(n-1)}(0) = 0,$$

$$\dot{y}(\infty) = \ddot{y}(\infty) = \dots = y^{(n-1)}(\infty) = 0.$$

Пример. Если, кроме того, $y(0) = y(\infty)$ и $u(0) = u(\infty)$, то с помощью повторного интегрирования можно показать, что

$$a_0 = c \int_0^\infty u(t) dt,$$

$$a_1 = -c \int_0^\infty \int_t^\infty \{u(t) - a_0 y(t)\} dt^2, \quad (8.91)$$

$$a_2 = c \int_0^\infty \int_t^\infty \left[\int_t^\infty \{u(t) - a_0 y(t)\} dt + a_1 y(t) \right] dt^2$$

и так далее,

$$\text{где } c^{-1} = \int_0^\infty y(t) dt.$$

Из-за неизбежных ошибок измерений применимость этого метода ограничивается дифференциальными уравнениями, порядок которых не превышает 3. Однако знания коэффициентов a_0 , a_1 , a_2 дифференциального уравнения n -го порядка достаточно для аппроксимации динамики соответствующего объекта с помощью n равных постоян-

ных времени. В данном случае в операционных обозначениях Лапласа

$$a_0 (sT + 1)^n X = a_0 \left\{ s^n T^n + \binom{n}{1} s^{n-1} T^{n-1} + \dots + \right. \\ \left. + \binom{n}{2} s^2 T^2 + \binom{n}{1} sT + 1 \right\} X. \quad (8.92)$$

Следовательно,

$$nT = \frac{a_1}{a_0} \quad \text{и} \quad \binom{n}{2} T^2 = \frac{a_2}{a_0},$$

откуда определяются n и T ; n должно быть целым.

8.5. КОРРЕЛЯЦИОННЫЕ ФУНКЦИИ ВЫСШИХ ПОРЯДКОВ

В разд. 8.1—8.3 рассмотрение было ограничено корреляционными функциями двух функций времени, которые можно назвать корреляционными функциями второго порядка. В некоторых случаях интересно рассмотреть корреляционные функции порядка n , определяемые для эргодических сигналов равенством

$$\Psi_{x_1 \dots x_n}(\tau_1, \dots, \tau_n) = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T x_1(t - \tau_1) \dots x_n(t - \tau_n) dt. \quad (8.93)$$

Они используются для описания связей между нелинейными сигналами и объектами, например в винеровском методе идентификации нелинейных объектов. Эти методы пока еще нашли весьма мало практических приложений; для детального ознакомления с этим вопросом следует обратиться к дополнительной литературе.

8.6. ДИСПЕРСИОННЫЕ ФУНКЦИИ; ОПРЕДЕЛЕНИЯ И СВОЙСТВА¹⁾

Из предыдущих разделов видно, что корреляционные методы хорошо применимы для идентификации линейных объектов. Это объясняется следующими причинами: для

¹⁾ В основу этого раздела положены конспекты И. Крегтинг-Янсен из технологического университета Эйндховена (Голландия).

линейных объектов гауссовскому входу отвечает гауссовский выход; статистические свойства (эргодических) гауссовых случайных процессов исчерпывающе характеризуются авто- и взаимными корреляционными функциями (второго порядка).

С нелинейными объектами дело обстоит намного сложнее. Дифференциальная аппроксимация и связанные с ней методы применимы для оценки параметров, если модель можно сделать линейной по параметрам и схема идентификации не искажается шумом. В соответствии с изложенным в разд. 8.5 при винеровском подходе к идентификации используются корреляционные функции высших порядков.

Пример. Для иллюстрации трудностей, возникающих при попытках идентифицировать нелинейный объект, рассмотрим следующую форму связи между входом $u(t)$ и выходом $y(t)$:

$$y(t) = u^2(t). \quad (8.94)$$

Если, например, $u(t)$ — эргодический гауссовский случайный сигнал с $E[u(t)] = 0$, то взаимная корреляционная функция, как легко видеть, равна нулю:

$$\Psi_{yu}(\tau) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y u p(y, u; \tau) dy du = \int_{-\infty}^{\infty} u^3 p(u; \tau) du = 0, \quad (8.95)$$

поскольку под интегралом стоит произведение четной и нечетной функций. Несмотря на функциональную зависимость между u и y , взаимная корреляционная функция никак не указывает на ее наличие.

Избежать некоторых из этих трудностей позволяют *дисперсионные методы*, использующие математическое ожидание, корреляционные и дисперсионные функции. Изложение метода и его применений можно найти в работе [27].

Определения дисперсионных функций [25, 26]. *Взаимной дисперсионной функцией* действительных слу-

чайных процессов называется функция двух переменных

$$\Theta_{yu}(t_1, t_2) = \mathcal{E}_{u(t_2)} [\{\mathcal{E}_{y(t_1)} [y(t_1) | u(t_2)] - \mathcal{E}_{y(t_1)} [y(t_1)]\}^2], \quad (8.96)$$

где $\mathcal{E}_{y(t_1)} [y(t_1) | u(t_2)]$ — условное математическое ожидание процесса $y(t_1)$ относительно $u(t_2)$.

Если известны соответствующие плотности распределения, дисперсионная функция определяется формулой

$$\Theta_{yu}(t_1, t_2) = \int_{-\infty}^{\infty} \left\{ \int_{-\infty}^{\infty} y(t_1) p(y(t_1) | u(t_2)) dy(t_1) - \int_{-\infty}^{\infty} y(t_1) p(y(t_1)) dy(t_1) \right\}^2 p(u(t_2)) du(t_2), \quad (8.97)$$

где $p(u(t_2))$ и $p(y(t_1))$ — плотности распределения $u(t_2)$ и $y(t_1)$ соответственно, а $p(y(t_1) | u(t_2))$ — условная плотность $y(t_1)$ относительно $u(t_2)$.

Автодисперсионная функция действительных случайных сигналов (процессов) $u(t_1)$ и $u(t_2)$ определяется как

$$\Theta_{uu}(t_1, t_2) = \mathcal{E}_{u(t_2)} [\{\mathcal{E}_{u(t_1)} [u(t_1) | u(t_2)] - \mathcal{E}_{u(t_1)} [u(t_1)]\}^2]. \quad (8.98)$$

При известных плотностях распределения автодисперсионная функция определяется соотношениями, аналогичными (8.97). Автодисперсионная функция $\Theta_{uu}(t_1, t_2)$ является количественной характеристикой случайного процесса, подобной автокорреляционной функции $\Psi_{uu}(t_1, t_2)$, и характеризует внутреннюю структуру процесса $u(t)$. Взаимная дисперсионная функция $\theta_{yu}(t_1, t_2)$, подобно взаимной корреляционной функции $\Psi_{yu}(t_1, t_2)$, является мерой связи случайных процессов $u(t)$ и $y(t)$. В практических задачах нормированные значения дисперсионных функций используются для сравнения с нормированными корреляционными функциями. *Нормированная автодисперсионная функция* определяется как

$$\eta_{uu}(t_1, t_2) = \sqrt{\frac{\Theta_{uu}(t_1, t_2)}{\sigma_u^2(t_1)}}, \quad (8.99)$$

где $\sigma_u^2(t_1) = \mathcal{E}_{u(t_1)} [\{u(t_1) - \mathcal{E}[u(t_1)]\}^2]$ — дисперсия.

Нормированной взаимной дисперсионной функцией называется функция

$$\eta_{yu}(t_1, t_2) = \sqrt{\frac{\Theta_{yu}(t_1, t_2)}{\sigma_y^2(t_1)}}. \quad (8.100)$$

В случае, когда $y(t_1)$ и $u(t_2)$ независимы, функция $\Theta_{yu}(t_1, t_2)$ равна нулю, и если $\sigma_y^2(t_1) \neq 0$, то и $\eta_{yu}(t_1, t_2) = 0$. Если $u(t_1)$ и $u(t_2)$ независимы, то $\Theta_{uu}(t_1, t_2) = 0$ и $\eta_{uu}(t_1, t_2) = 0$ при $\sigma_u^2(t_1) \neq 0$. При (детерминированной) функциональной зависимости между $y(t_1)$ и $u(t_2)$ справедливы равенства $\Theta_{yu}(t_1, t_2) = \sigma_y^2(t_1)$ и $\eta_{yu}(t_1, t_2) = 1$. Если взаимная дисперсионная функция или нормированная взаимная дисперсионная функция случайных процессов $y(t_1)$ и $u(t_2)$ равна нулю для определенных значений аргументов t_1 и t_2 , то случайные величины $y(t_1)$ и $u(t_2)$ называются *недисперсированными*. Этот термин соответствует понятию некоррелированности в линейном случае. Если случайные величины не дисперсированы, то они также и не коррелированы. Однако обратное не верно. Укажем некоторые свойства дисперсионных функций:

1) Дисперсионные функции неотрицательны:

$$\Theta_{uu}(t_1, t_2) \geq 0, \quad \eta_{uu}(t_1, t_2) \geq 0,$$

$$\Theta_{yu}(t_1, t_2) \geq 0, \quad \eta_{yu}(t_1, t_2) \geq 0.$$

2) С помощью неравенства Шварца можно показать, что

$$\sigma_y^2(t_1) \geq \Theta_{yu}(t_1, t_2).$$

3) Из свойств 1) и 2) следует, что

$$0 \leq \eta_{uu}(t_1, t_2) \leq 1,$$

$$0 \leq \eta_{yu}(t_1, t_2) \leq 1.$$

4) В общем случае взаимная дисперсионная функция и нормированная взаимная дисперсионная функция несимметричны:

$$\Theta_{yu}(t_1, t_2) \neq \Theta_{uy}(t_2, t_1), \quad \eta_{yu}(t_1, t_2) \neq \eta_{uy}(t_2, t_1).$$

5) Нормированная взаимная дисперсионная функция не меньше абсолютной величины нормированной взаим-

ной корреляционной функции Y_{yu} :

$$\eta_{yu}(t_1, t_2) \geq \left| \frac{\Psi_{yu}(t_1, t_2)}{\sigma_y(t_1) \sigma_u(t_2)} \right| = |Y_{yu}(t_1, t_2)|.$$

Равенство имеет место только в линейном случае.

Во многих задачах требуется найти регрессию величины u_1 в произвольный момент t_1 на значения u_2, \dots, u_n для произвольных t_2, \dots, t_n . Связь между $u_1(t_1)$ и $u_2(t_2), \dots, u_n(t_n)$ характеризует множественная дисперсионная функция

$$\Theta_u(t_1, \dots, t_n) = \mathcal{E}_{u_2(t_2) \dots u_n(t_n)} [\{\mathcal{E}[u_1(t_1) | u_2(t_2), \dots, u_n(t_n)] - \mathcal{E}[u_1(t_1)]\}^2]. \quad (8.101)$$

Аналогично определяется нормированная множественная дисперсионная функция

$$\eta_u(t_1, \dots, t_n) = \sqrt{\frac{\Theta_u(t_1, \dots, t_n)}{\sigma_{u_1}^2(t_1)}}. \quad (8.102)$$

Примем в качестве меры нелинейности линии регрессии $\mathcal{E}[y(t_1) | u(t_2)]$ среднеквадратическое отклонение этой линии от прямой, минимизирующей величину этого отклонения. Определим степень нелинейности регрессии $y(t_1)$ на $x(t_2)$ как

$$v_{yu}(t_1, t_2) = \sqrt{\min_{a, b} \mathcal{E}_{u(t_2)} [\{\mathcal{E}_{y(t_1)}[y(t_1) | u(t_2)] - (a + bu(t_2))\}^2]} \quad (8.103)$$

Минимум в этой формуле достигается при

$$b = b(t_1, t_2) = Y_{yu}(t_1, t_2) \frac{\sigma_y(t_1)}{\sigma_u(t_2)} \quad (8.104)$$

и

$$a = a(t_1, t_2) = \mathcal{E}[y(t_1)] - Y_{yu}(t_1, t_2) \frac{\sigma_y(t_1)}{\sigma_u(t_2)} \mathcal{E}[u(t_2)]. \quad (8.105)$$

Подставляя $b(t_1, t_2)$ и $a(t_1, t_2)$ в формулу (8.103), можно получить следующее выражение для квадрата степени нелинейности:

$$v_{yu}^2(t_1, t_2) = \eta_{yu}^2(t_1, t_2) - Y_{yu}(t_1, t_2). \quad (8.106)$$

Примеры. Для иллюстрации этих идей полезно рассмотреть несколько примеров. Для упрощения обозначений аргументы t_1 и t_2 в примерах опущены.

а) y и u независимы (фиг. 8.22, а).

$$\mathcal{E}[y|u] = \mathcal{E}[y] = \text{const},$$

$$\Theta_{yu} = \int \{\mathcal{E}[y] - \mathcal{E}[y]\}^2 p(u) du = 0,$$

$$\Psi_{yu} = \int \int u y p(u, y) du dy = \mathcal{E}[u] \mathcal{E}[y].$$

б) y и u связаны линейной статистической зависимостью (фиг. 8.22, б).

$$|\quad \mathcal{E}[y|u] = au, \quad a = \frac{\Psi_{yu}}{\sigma_u^2}, \quad \mathcal{E}[u] = \mathcal{E}[y] = 0,$$

$$\Theta_{yu} = \int \left\{ \frac{\Psi_{yu}}{\sigma_u^2} \right\}^2 u^2 p(u) du = \frac{\Psi_{yu}^2}{\sigma_u^2},$$

$$\Psi_{yu} = \int \int u y p(u, y) du dy,$$

$$\nu_{yu}^2 = \eta_{yu}^2 - Y_{yu}^2 = \frac{\Psi_{yu}^2}{\sigma_u^2 \sigma_y^2} - \left(\frac{\Psi_{yu}}{\sigma_u \sigma_y} \right)^2 = 0.$$

в) y и u связаны линейной функциональной зависимостью (фиг. 8.22, в).

$$y = au, \quad \mathcal{E}[y|u] = au, \quad \mathcal{E}[u] = \mathcal{E}[y] = 0,$$

$$\Theta_{yu} = \int \{au - \mathcal{E}[u]\}^2 p(u) du = a^2 \sigma_u^2,$$

$$\Psi_{yu} = \int \int u y p(u, y) du dy = a \int u^2 p(u) du = a \sigma_u^2,$$

$$\nu_{yu}^2 = \eta_{yu}^2 - Y_{yu}^2 = a^2 \frac{\sigma_u^2}{\sigma_y^2} - a^2 \frac{\sigma_u^4}{\sigma_u^2 \sigma_y^2} = 0.$$

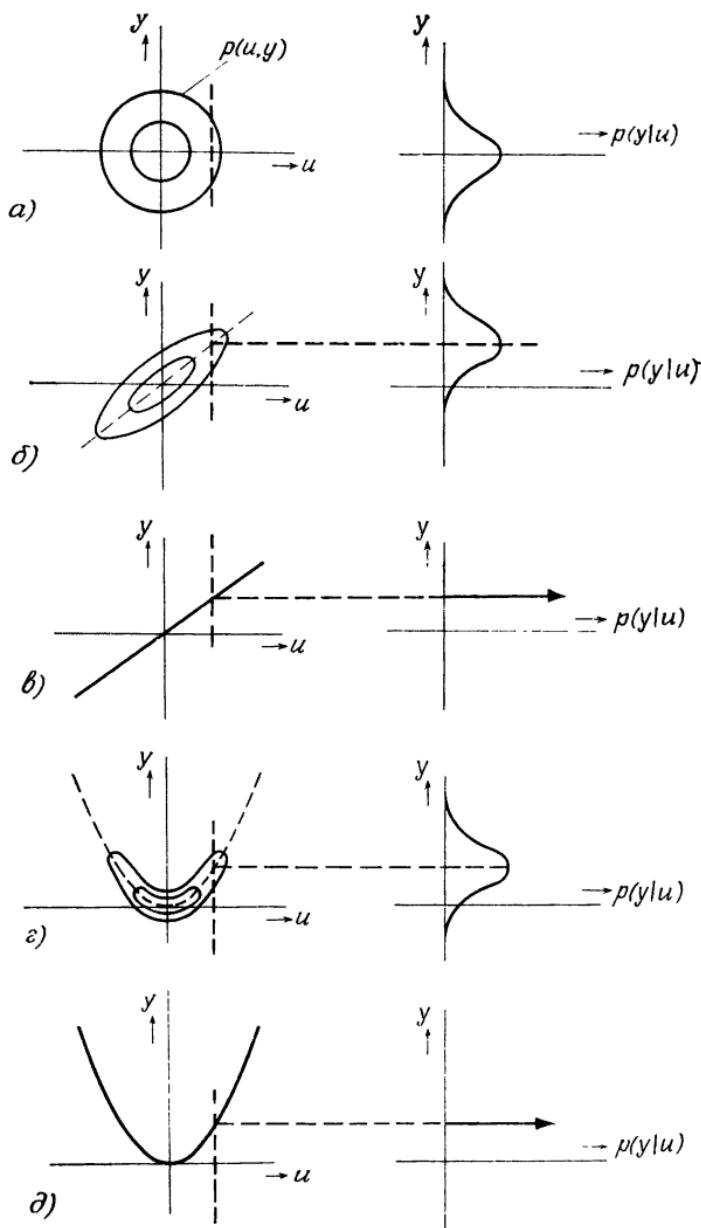
г) y и u связаны квадратичной статистической зависимостью (фиг. 8.22, г).

$$\mathcal{E}[y|u] = u^2, \quad \mathcal{E}[u] = 0, \quad p(u) = p(-u),$$

$$\Theta_{yu} = \int \{u^2 - \mathcal{E}[y]\}^2 p(u) du \geq 0,$$

$$\begin{aligned} \Psi_{yu} &= \int \int u y p(u, y) du dy = \int u \left\{ \int y p(y|u) dy \right\} p(u) du = \\ &= \int u^3 p(u) du = 0, \end{aligned}$$

$$\nu_{yu}^2 = \eta_{yu}^2 - Y_{yu}^2 = \frac{\Theta_{yu}}{\sigma_y^2} - 0 \geq 0.$$



Ф и г. 8.22.

д) y и u связаны квадратичной функциональной зависимостью (фиг. 8.22, д).

$$y = u^2, \quad \mathcal{E}[u] = 0, \quad p(u) = p(-u),$$

$$\Theta_{uy} = \int \{u^2 - \mathcal{E}[u^2]\}^2 p(u) du = \sigma_{u^2}^2 = \sigma_y^2 \neq 0,$$

$$\Psi_{yu} = \iint u y p(u, y) du dy = \int u^3 p(u) du = 0,$$

$$\nu_{yu}^2 = \eta_{yu}^2 - Y_{yu}^2 = \frac{\sigma_y^2}{\sigma_u^2} - 0 = 1.$$

Пусть вход $u(t)$ и выход $y(t)$ объекта — случайные функции. Свойства объекта можно охарактеризовать оператором \mathcal{O}_p :

$$y(t_1) = \mathcal{O}_p[u(t_2)]. \quad (8.107)$$

Если $u(t)$ и $y(t)$ доступны измерениям, задача идентификации состоит в отыскании оператора \mathcal{O}_M , являющегося приближением к истинному оператору \mathcal{O}_p в соответствии с некоторым критерием. Возможным критерием является условие минимума математического ожидания функции потерь

$$\mathcal{E}[C\{y(t_1), w(t_1)\}]. \quad (8.108)$$

Для квадратичной функции потерь $C\{y(t), w(t)\} = \{y(t) - w(t)\}^2$ минимум ее математического ожидания достигается при

$$w(t_1) = \mathcal{E}_{y(t_1)}[y(t_1) | u(t_2)]. \quad (8.109)$$

Этот случай показан на фиг. 8.23. Используя обозначения $u(t_2) = u$ и $y(t_1) = y$, можно записать дисперсию выхода в виде

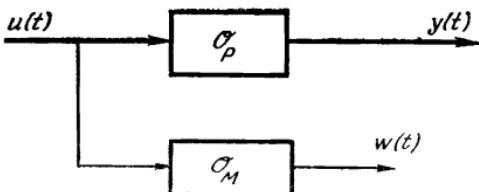
$$\sigma_y^2 = \mathcal{D}[y] = \mathcal{D}[\mathcal{E}_y[y|u]] + \mathcal{E}_{yu}[\mathcal{D}[y|u]], \quad (8.110)$$

где

$$\mathcal{D}[y] \stackrel{\Delta}{=} \int \{y - \mathcal{E}[y]\}^2 p(y) dy, \quad (8.111)$$

$$\mathcal{D}[\mathcal{E}_y[y|u]] = \int \{\mathcal{E}_y[y|u] - \mathcal{E}_y[y]\}^2 p(u) du, \quad (8.112)$$

$$\mathcal{E}_{yu}[\mathcal{D}[y|u]] = \iint \{y - \mathcal{E}_y[y|u]\}^2 p(u, y) du dy. \quad (8.113)$$



Ф и г. 8.23.

При квадратичной функции потерь выражение для дисперсии выхода $y(t)$ имеет вид

$$\mathcal{D}[y] = \mathcal{E}_u [\{w - \mathcal{E}[y]\}^2] + \mathcal{E}[\{y - w\}^2], \quad (8.114)$$

причем в первом слагаемом в правой части этого равенства легко узнать взаимную дисперсионную функцию. Дальнейшие сведения по этим вопросам читатель может почерпнуть из работ [4, 23, 24, 26].

ЛИТЕРАТУРА

1. Balchen J. G., Blandhol E., On the Experimental Determination of Statistical Properties of Signals and Disturbances in Automatic Control Systems, First IFAC Congress, Moscow, 1960, pp. 788—796; есть русский перевод: в книге «Статистические методы исследования. Теория структур, моделирование, терминология, образование», Труды I Межд. конгресса ИФАК, Изд-во АН СССР, 1961.
2. Балл Г. А., Об одном способе ускорения аппаратурного корреляционного анализа, *Автоматика и телемеханика*, № 9 (1963).
3. Bendat J. S., Piersol A. G., Measurements and Analysis of Random Data, Wiley, N.Y., 1966; есть русский перевод: Бендат Дж., Пирсол А., *Измерение и анализ случайных процессов*, изд-во «Мир», 1971.
4. Bunich A. L., Rajbman N. S., A Dispersion Equation of Non-linear Plant Identification, *Problems of Control and Inform. Theory*, 1, 29—36 (1972).
5. Buchta H., Method of Estimation of Random Errors in the Determination of Correlation Functions of Random Infralow Frequency Signals in Control System, Fourth IFAC Congress, Warsaw, Paper 33.4, 1969.
6. Chang C. M., Goodman T. P., Reswick J. B., Use of Correlation Functions to Determine System Characteristics without Applying Artificial Disturbances, in: *Regelungstechnik*, Müller G., ed., Oldenburg, München, 1956, pp. 251—256.

7. Chang K. Y., Moore A. D., Modified Digital Correlator and its Estimation Errors, *IEEE Trans. Inform. Theory*, IT-16, 699—706 (1970).
8. Davenport W. B. et al., Statistical Errors in Measurement on Random Time Functions, *J. Appl. Phys.*, 23, 377—388 (1952).
9. Eykhoff P. (см. литературу к гл. 2).
10. Faran J. J., Hills R., Correlators for Signal Reception, Techn. Memo № 27, Acoustics Research Laboratory, Harvard Univ., Mass., 1952.
11. Goodman T. P., Determination of the Characteristics of Multi-Input and Non-Linear Systems from Normal Operating Records, *Trans. ASME*, 79, 567—575 (1957).
12. Goodman T. P., Reswick J. B., Determination of Systems Characteristics from Normal Operating Records, *Trans. ASME*, 78, 259—271 (1956).
13. Green P. E., Jr., The Output Signal-to-Noise Ratio of Correlation Detectors, *IRE Trans. Inform. Theory*, IT-3, 10—18 (1957).
14. Hayashi N., On a Method of Correlation Analysis for Multivariate Systems, Fourth IFAC Congress, Warsaw, Paper 33.1, 1969.
15. Jespers P., Chu P. T., Fettweis A., A New Method for Computing Correlation Functions, *IRE Trans. Inform. Theory*, IT-8, 106—107 (1962).
16. Kitamori T. (см. литературу к гл. 3).
17. Lampard D. G. (см. литературу к гл. 3).
18. Lange F. H., Correlation Techniques, Van Nostrand, Princeton, N.J., 1967.
19. Laning J. H., Battin R. H., Random Processes in Automatic Control, McGraw-Hill, N.Y., 1956; есть русский перевод: Лэ-нинг Дж., Бэттин Р., Случайные процессы в задачах автоматического управления, ИЛ, 1958.
20. Lee Y. W., Statistical Theory of Communication, N.Y., Wiley, 1960.
21. McGraw D. K., Wagner J. F., Elliptically Symmetric Distributions, *IEEE Trans. Inform. Theory*, IT-14, 110—120 (1968).
22. Peek J. B. H., The Measurement of Correlation Functions in Correlators Using «Shift-Invariant Independent» Functions, Doctoral Dissertation, Technological Univ., Eindhoven, Netherlands, 1967.
23. Пугачев В. С., Теория случайных функций и ее применение к задачам автоматического управления, Физматгиз, 1962.
24. Райбман Н. С., Терехин А. Т., Дисперсионные функции и их применение для исследования нелинейных объектов, *Автоматика и телемеханика*, № 3 (1965).
25. Райбман Н. С., Идентификация нелинейных объектов с помощью дисперсионных функций, в книге «Теория непрерывных автоматических систем и вопросы идентификации», Труды III Межд. конгресса ИФАК, изд-во «Наука», 1971.
26. Rajbman N. S., Identification of Nonlinear Controlled Plants with Dispersion Functions, IFAC Symp. «Identification in Autom. Control Systems», Prague, Paper 3.14, 1967.
27. Rajbman N. S. (см. литературу к гл. 2).

28. Rake H., The Determination of Frequency Responses from Stochastic Signals by Means of an Electronic Analog Computer (на немецком), *Regelungstechnik*, 16, 258—260 (1968).
29. Shinbrot M., On the Analysis of Linear and Nonlinear Systems, *Trans. ASME*, 79, 547—552 (1957).
30. Соловьевников В. В. (см. литературу к гл. 12).
31. Strejc V., New Methods for Approximating Differential Equations of Control Loops using a General Input Signal (на немецком), *Acta Technica*, 3, 241—261 (1958).
32. Strejc V., Determination of Dynamical Properties of Control Loops Using Measured Input and Output Signals of a General Nature (на немецком), Tagungsbericht IV. Int. Kolloq. d. Hochschule für Elektrotechnik. Ilmenau, 1959, pp. 397—402; см. также *Zeitschr. Messen, Steuern, Regeln*, 7—11 (1960).
33. Strejc V., Evaluation of General Signals with Non-zero Initial Conditions, *Acta Technica*, 6, 378—391 (1961).
34. Veltman B. P. Th., Kwakernaak H., Theory and Application of Polarity Correlation for Analyzing Low Frequency Signals and Systems (на немецком), *Regelungstechnik*, 9, 337—364 (1961).
35. Wallman H., Electronic Integral-Transform Computer and Practical Solution of Integral Equations, *J. Franklin Inst.*, 250, 45—61 (1950).
36. Watts D. G., A general Theory of Amplitude Quantization with Application to Correlation Determination, *Proc. IEE*, 109C, 209—218 (1962).
37. Welfonder E., Stochastic Signals; Statistical Averaging under Real Conditions (на немецком), *Fortschritt Bericht VDI-Z*, 8, 165 (1969).
38. Widrow B., A Study of Rough Amplitude Quantization by Means of Nyquist Sampling Theory, *IRE Trans. Circuit Theory*, CT-3, 266—276 (1956).

ДОПОЛНИТЕЛЬНАЯ ЛИТЕРАТУРА

8.1. Действия над аналоговыми сигналами

Albert G. E., Statistical Methods in Prediction, Filtering and Detection Problems, *SIAM J.*, 8, 640—653 (1960).

Alper P., An Open-Loop Procedure for Process Parameter Estimation Using a Hybrid Computer, IFAC Symp. on «The Theory of Self-Adaptive Control Systems», Teddington, Paper 3.1, 1965.

Anderson G. W., Buland R. N., Cooper G. R., Use of Crosscorrelation in an Adaptive Control System, *Proc. Nat. Electron Conf.*, 15, 34—35 (1959).

Bär D., Practical Method of Determining the Dynamic Characteristic Values of Linear Controlled Plants with Inherent Control by Means of the Correlation Analysis (на немецком), *Zeitschr. Messen, Steuern, Regeln*, 5, 251—256 (1962).

Bartels E., Practical System Analysis Using Correlation Techniques (на немецком), *Regelungstechnik*, 14, 49—55 (1966).

Bastl W. (см. литературу к разд. 14.3).

Cooper D. C., The Probability Density Function for the Output of a Correlator with Band-Pass Input Waveforms, *IEEE Trans. Inform. Theory*, IT-11, 190—195 (1965).

Cowley P. E. A., Fano R. M., Basore B. L., A Short-Time Correlator for Speech Waves, Techn. Rept. 174, Research Lab, Electron. MIT, Cambridge, Mass., 1951.

Csáki R., Gyürki J., Some Questions of the System Identification by Stochastic Signals, IFAC Symp. «Identification in Autom. Control Systems», Prague, Paper 3.21, 1967.

Елистратов М. Р., Определение динамических характеристик нестационарных систем в процессе нормальной работы, *Автоматика и телемеханика*, № 7 (1963).

Elkind J. I., Green D. M., Starr E. A., Application of Multiple Regression Analysis to Identification of Time-Varying Linear Dynamic Systems, Joint Autom. Control Conf., 1962.

Fife D. W., Simrar: Simulated Receiver and Recorder for Statistical Measurements, *Annls. Ass. Int. Calcul Analog.*, 5, 156—159 (1963).

Goodman T. P., Hillsley R. H., Continuous Measurement of Characteristics of Systems with Random Inputs: a Step Toward Self-optimizing Control, *Trans. ASME*, 80, 1839—1848 (1958).

Goodman N. R. et al., Frequency Response from Stationary Noise: Two Case Histories, *Technometrics*, 3, 245—268 (1961).

Henderson J. G., Pengilley, Experimental Determination of System Transfer Functions from Normal Operating Data, *Jnl. Br. IRE*, 18, 149 (1958).

Hofstetter M., Some Results on Stochastic Signal Parameter Estimation Problem, *IEEE Trans. Inform. Theory*, IT-11, 422—429 (1965).

Hughes M. T. G., Noton A. R. M., The Measurement of Control System Characteristics by Means of a Cross Correlator, *Proc. IEE*, 109B, 77—83 (1962).

Янч К., Решение вероятностных задач при помощи моделирования, *Автоматика и телемеханика*, № 1 (1964).

Jenkins G. M., Cross Spectral Analysis and the Estimation of Linear Open Loop Transfer Functions, in: *Time Series Analysis*, Rosenblatt, Ed., Wiley, N.Y., 1963.

Kryze J., MUSA-6, a Universal Statistical Analyser (на немецком), *Zeitschr. Messen, Steuern, Regeln*, 6, 286, 386 (1963).

Kwakernaak H. (см. литературу к гл. 6).

Lange F. H., Zecha M., Stochastic Error Theory, Fourth IFAC Congress, Warsaw, Paper 33.3, 1969.

Леонов Ю. П., Липатов Л. Н., Прибор для статистического исследования динамических характеристик промышленных объектов при наличии шумов, *Приборостроение*, № 11 (1960).

Levey E. C., Complex Curve Fitting, *IRE Trans. Autom. Control*, AC-4, 37—43 (1959).

Lindenlaub J. C., Cooper G. R., Noise Limitations of System Identification Techniques, *IEEE Trans. Autom. Control*, AC-8, 43—48 (1963).

Mann K. H., Otte G., Werner G. W., Correlation Measuring Installation for Scientific Research and Automatic Control Teaching (на немецком), *Wiss. Z. Elektrotechn.*, 2, 53—68 (1963).

Mesch F., Non-Ideal Time Delay Circuits for Correlation Measurements, *Regelungstechnik*, 14, 70—74 (1966).

Otterman J., The Properties and Methods for Computation of Exponentially-Mapped-Past Statistical Variables, *IRE Trans. Autom. Control*, AC-5, 11—18 (1960).

Parks P. C., Clarkson B. L., Some Theory and Practice of Random Process Analysis, *Proc. Instr. Mech. Engrs.* 179, part 34 (1965).

Perlis H. J., Justification and Limitation of a 2-Parameter 2-Adjustment Self-Adaptive System, *IEEE Trans. Applic. and Ind.*, 82, 220—229 (1963).

Perlis H. J., The Minimization of Measurement Error in a General Perturbation-Correlation Process Identification System, *IEEE Trans. Autom. Control*, AC-9, 339—345 (1964).

Peschel M., Investigation of Linear Systems by Means of Appropriate Noise Processes (на немецком), *Zeitschr. Messen, Steuern, Regeln*, 5, 257—260 (1962).

Qvarnström B., Transfer Function-Determination in the Presence of Noise for a Set of Significant Input Functions, Proc. of the Fifth Inst. Instruments and Measurements Conf., Stockholm, 1960.

Райбман Н. С., Корреляционные методы определения приближенных характеристик многомерных процессов автоматических линий, *Изв. АН СССР, Энергетика и автоматика*, № 2 (1962).

Resa F. M., An Introduction to Information Theory, McGraw-Hill, N.Y., 1961.

Reswick J. B., Determine System Dynamics without Upset, *Control Engng*, 2, 50 (1955).

Schetzen M., Determination of Optimum Nonlinear Systems for Generalized Error Criteria Based on the Use Gate Functions, *IEEE Trans. Inform. Theory*, IT-11, 117—125 (1965).

Simmons G. J., A Factorization Technique for Binary Autocorrelation Functions, *Proc. IEEE*, 54, 794—795 (1966).

Соловьевников В. В., Усков А. С., Статистический анализ объектов регулирования, Машгиз, 1960.

Sydow A., Statistical Analysis by Means of the Electronic Analog Computer (на немецком), *Zeitschr. Messen, Steuern, Regeln*, 6, 501—504 (1963).

Van der Grinten P. M. E. M., Krijger W., Processing of the Auto and Cross-Correlation Functions to Step Response, *Annls. Ass. Int. Calcul. Analog.*, 5, 160—161 (1963).

Van der Grinten P. M. E. M., Stochastic Processes in Measurement and Control (на голландском, реферат на английском), Doctoral Thesis, Univ. of Technology, Eindhoven, Netherlands, 1962.

Werner G. W., Contribution to Correlation Techniques in Control Engineering, (на немецком), *Zeitschr. Messen, Steuern, Regeln*, 5, 347—354 (1963).

Werner G. W., Seide W., The Determination of Pole-Zero Distribution of an Industrial Controlled Plant by Means of Correlation

Measurements (на немецком), *Wiss. Z. Hochsch. Elektrotech.*, Ilmenau, DDR., 8, 353—359 (1962).

Westcott J. H., The Determination of Process-Dynamics from Normal Disturbance Records of a Controlled Process, in *à Regelungstechnik*, Muller G., Ed., München, Oldenbourg, 1956, pp. 256—261.

Westcott J. H., The Parameter-Estimation Problem, First IFAC Congress, Moscow, 1960, pp. 779—787; есть русский перевод: Весткотт Дж., Проблемы оценки параметров, в книге «Статистические методы исследования, теория структур, моделирование, терминология, образование», Труды I Межд. конгресса ИФАК, Изд-во АН СССР, 1961.

Wolf A. A., On the Significance of Instantaneous and Short Term Correlation Functions for a Class of Stochastic Processes, *Proc. IRE*, 50, 1983 (1962).

Wolf A. A., Dietz J. H., A statistical Theory for Parameter Identification in Physical Systems, *J. Franklin Inst.*, 274, 369—400 (1962).

Woodrow, Closed Loop Dynamics from Normal Operating Records, *Trans. Soc. Instrum. Technol.*, 10 (1958).

8.2. Действия над квантованными сигналами

Bennett W. R., Spectra of Quantized Signals, *Bell. Syst. Techn.*, 27, 446—472 (1948).

Price R., A Useful Theorem for Non-Linear Devices Having Gaussian Inputs, *IRE Trans. on Inform. Theory*, IT-4, 69—72 (1958).

Scanlon W. W., Lieberman C., Naval Ordnance and Electronic Research, *Proc. IRE*, 47, 910 (1959).

The G., Theorem on Stieltjes Correlation, *Control and Science Record*, 113, 1091 (1966).

Diamessis J. E., A Method for Determining the Parameters of Certain Time-Varying Systems (Correspondence), *Proc. IEEE*, 53, 396—397 (1965).

Ekre H., Polarity Coincidence Detection of a Weak Noise Source, *IEEE Trans. Inform. Theory*, IT-9, 18—23 (1963).

Heins W., Bos A. v. d., Veltman B. P. Th., On the Moments of Shifted Gaussian Probability Density Functions, Rept., Coop. Centre, Univ. of Technology, Delft, Netherlands, 1965.

Кузьмичев В. Н., Нестерук В. Ф., Методика и устройство для определения корреляционных функций нормальных и импульсных случайных процессов, *Автоматика и телемеханика*, № 5 (1964).

Liewers P., Buttler E. (см. литературу в разд. 14.3).

Poortvliet D. C. J. (см. литературу к гл. 10).

Ruchkin D. S., Error of Correlation Coefficient Estimates from Polarity Coincidences (Correspondence), *IEEE Trans. Inform. Theory*, IT-11, 296—297 (1965).

Соловьевников В. В., Усков А. С., Частотный метод определения динамических характеристик объектов автоматического регулирования по данным их нормальной эксплуатации, *Автоматика и телемеханика*, № 12 (1959).

Stott S. I., Canning T. I., Method of Correlating Binary Signals, *Proc. IEE*, **112**, 1835 (1965).

Van den Bos A., The Variance of Auto-Correlation Estimates of Clipped Gaussian Noise, Rept., Coop. Centre, Univ. of Technology, Delft, Netherlands, 1965.

Van Vleck J., Middleton D., The Spectrum of Clipped Noise, *Proc. IEEE*, **54**, 2—19 (1966).

Veltman B. P. Th., Van den Bos A., The Applicability of the Relay-Correlator and the Polarity-Coincidence Correlator in Automatic Control, Second IFAC Congress, Basel; есть русский перевод: Вельтман Б. П., Ван ден Бос А., Применение линейного коррелятора и коррелятора совпадения знаков в автоматическом регулировании, в книге «Теория непрерывных автоматических систем», Труды II Межд. конгресса ИФАК, изд-во «Наука», 1965.

Veltman B. P. Th., Quantizing, Sampling Frequency and Statistical Dispersion of Correlation Measurements (на немецком), *Rege lungstechnik*, **14**, 151—158 (1966).

Wolff S. S., Thomas J. B., Williams T. R., The Polarity-Coincidence Correlator: a Nonparametric Detection Device, *IRE Trans. Inform. Theory*, **IT-8**, 5—9 (1962).

Womack B. F., Watt J. T., Jr., A New Method for Determining System Parameters (Correspondence), *Proc. IEEE*, **53**, 729—730 (1965).

8.4. Метод дифференциальной аппроксимации

Bellman R. E., Kalaba R. E., Kotkin B., Differential Approximation Applied to the Solution of Convolution Equations, Rand Corp., RM-3601, 1963.

Bellman R. E., Gluss B., Roth R., Identification of Differential Systems with Time-Varying Coefficients, Rand Corp., RM-4288-PR, 1964.

Bellman R. E., Kalaba R. E., Sridhar R., Adaptive Control via Quasilinearization and Differential Approximation, *Computing*, **1**, 8—17 (1966).

Sage A. P., Eisenberg B. R., Experiments in Nonlinear and Non-stationary System Identification via Quasilinearization and Differential Approximation, Joint Autom. Control Conf., 1966, pp. 522—530.

8.5. Корреляционные функции высших порядков (см. также литературу к разд. 8.4)

Booton R. C., The Measurement and Representation of Nonlinear Systems, *IRE Trans. Circuit Theory*, **CT-1**, 32—34 (1954).

Elkind J. I., Green D. M., A Technique for Measuring the Characteristics of Nonlinear Systems, Rept. 849, Bolt Beranek and Newman, Cambridge, Mass., 1961.

Lee Y. M., Schetzen M., Some Aspects of the Wiener Theory of Nonlinear Systems, *Proc. Nat. Electron. Conf.*, **21**, 759—764 (1965).

Lee Y. W., Schetzen M., Measurement of the Kernels of a Non-linear System by Cross-Correlation, *Q. Prog. Rept. 60, Res. Lab. Electron.*, MIT, 1961, pp. 118—130; см. также *Int. Jnl. Control* (1965).

Wiener N. (см. литературу к гл. 4).

ГЛАВА 9

НЕПРЕРЫВНЫЕ СИГНАЛЫ; НЕЯВНЫЕ МЕТОДЫ (НАСТРАИВАЕМЫЕ МОДЕЛИ)

Так же как в гл. 7, методы, связанные с настройкой модели, можно интерпретировать либо как схему управления параметрами физической (аналоговой) модели, либо как вычислительную схему, в которой легко обнаружить математическую модель. В гл. 2 были введены понятия обобщенной модели и ошибки. Преимуществом такого подхода является линейная связь ошибки и подлежащих настройке параметров. Некоторые соображения общего характера приводятся в разд. 9.1. Обсуждается ряд свойств таких схем настройки моделей. Проводится различие между критериями мгновенной ошибки $E_1 = e^2(\beta, t)$ (разд. 9.2) и среднеинтегральной ошибки $E_2 = 1/T \int_{t-T}^t e^2(\beta, t) dt$

(разд. 9.3). В последнее равенство входит величина интервала усреднения T . На практике это время можно выбрать достаточно большим, чтобы удовлетворить требованиям к точности оценки. Критерий мгновенной ошибки можно применять только при достаточно пологой кривой настройки. В этом случае контур настройки, имеющий характер низкочастотного фильтра, обеспечит необходимый для достижения заданной точности эффект сглаживания.

Разд. 9.4 посвящен функциям чувствительности параметров и их применением в непрерывных и прерывистых схемах настройки. Приближенные выражения функций чувствительности, получаемые при использовании двух моделей с параметрами β и $\beta + \Delta\beta$ соответственно, рассматриваются в разд. 9.5. Заключительный раздел 9.6

посвящен применению соотношения

$$\frac{dE}{dt} = \frac{\partial E}{\partial \beta} \frac{d\beta}{dt} + \frac{\partial E}{\partial t},$$

где β меняется, например, по синусоидальному закону, а dE/dt доступно измерениям. Если величиной $\partial E/\partial t$ можно пренебречь, то удается найти $\partial E/\partial \beta$. Таким образом определяется целесообразное направление настройки параметров. При чтении этой главы и особенно при изучении методов настройки моделей необходимо помнить, что детерминированные и стохастические итерационные схемы, обсуждавшиеся в разд. 5.2 и 5.3, имеют непосредственное отношение к рассматриваемой задаче.

9.1. МОДЕЛИ, ЛИНЕЙНЫЕ ОТНОСИТЕЛЬНО ПАРАМЕТРОВ

Обобщенная ошибка

В гл. 2 была определена обобщенная ошибка

$$e = \sum_{i=0}^n \alpha_i y_i + \sum_{j=0}^m \beta_j u_j, \quad (9.1)$$

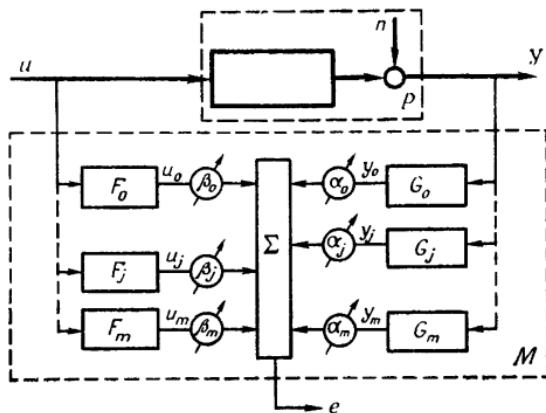
где u_j и y_i — выходы динамических операторов.

Назначением устройства на фиг. 9.1 является настройка потенциометров $\alpha_0, \dots, \alpha_n$ и β_0, \dots, β_m таким образом, чтобы достигал минимума некоторый функционал от обобщенной ошибки. Некоторое положение потенциометра можно произвольно принять за эталонное, например $\beta_0 F_0 = -1$. При таком выборе

$$e = \sum_{i=0}^n \alpha_i y_i + \sum_{j=1}^m \beta_j u_j - u. \quad (9.2)$$

Как уже указывалось, существенной особенностью таких обобщенных моделей является линейная связь между e и параметрами α_i и β_j . Поэтому частные производные (функции чувствительности параметров, см. разд. 9.4) определяются очень просто:

$$\frac{\partial e}{\partial \alpha_i} = y_i, \quad \frac{\partial e}{\partial \beta_j} = u_j. \quad (9.3)$$



Ф и г. 9.1.

Следовательно, не нужно никаких дополнительных моделей для реализации уравнений чувствительности, и все функции чувствительности параметров можно получить из одной (обобщенной) модели.

Выберем пока в качестве критерия условие минимума квадрата ошибки на интервале длины T . Для устранения эффекта случайности применим операцию математического ожидания

$$E = \mathcal{E} \left[\int_{t-T}^t e^2 dt \right]. \quad (9.4)$$

Дифференцируя по параметрам, подлежащим определению, приходим к системе из $m + n + 1$ уравнений

$$\frac{1}{2} \zeta_i = \frac{1}{2} \frac{\partial E}{\partial \alpha_i} = \mathcal{E} \left[\int_{t-T}^t ey_i dt \right] \stackrel{\Delta}{=} \langle e, y_i \rangle, \quad i = 0, 1, \dots, n, \quad (9.5a)$$

$$\frac{1}{2} \eta_j = \frac{1}{2} \frac{\partial E}{\partial \beta_j} = \mathcal{E} \left[\int_{t-T}^t eu_j dt \right] \stackrel{\Delta}{=} \langle e, u_j \rangle, \quad j = 1, \dots, m. \quad (9.5b)$$

Символ $\langle \cdot \rangle$ введен для сокращения записи. В следующем разделе ему будет дана более глубокая интерпретация. В оптимальной точке все ζ_0, \dots, ζ_n и η_1, \dots, η_m должны

быть равны нулю. Допустим теперь, что $\alpha_0, \dots, \alpha_n, \beta_1, \dots, \beta_m$ постоянны. Подставляя выражение для обобщенной ошибки (9.2) в уравнения (9.5а) и (9.5б), можно получить

$$\frac{1}{2} \begin{bmatrix} \zeta_0 \\ \vdots \\ \zeta_n \\ \eta_1 \\ \vdots \\ \eta_m \end{bmatrix} = \begin{bmatrix} \langle y_0, y_0 \rangle & \dots & \langle y_n, y_0 \rangle & \langle u_1, y_0 \rangle & \dots & \langle u_m, y_0 \rangle \\ \vdots & & \vdots & \vdots & & \vdots \\ \langle y_0, y_n \rangle & \dots & \langle y_n, y_n \rangle & \langle u_1, y_n \rangle & \dots & \langle u_m, y_n \rangle \\ \langle y_0, u_1 \rangle & \dots & \langle y_n, u_1 \rangle & \langle u_1, u_1 \rangle & \dots & \langle u_m, u_1 \rangle \\ \vdots & & \vdots & \vdots & & \vdots \\ \langle y_0, u_m \rangle & \dots & \langle y_n, u_m \rangle & \langle u_1, u_m \rangle & \dots & \langle u_m, u_m \rangle \end{bmatrix} \times$$

$$\times \begin{bmatrix} \alpha_0 \\ \vdots \\ \alpha_n \\ \beta_1 \\ \vdots \\ \beta_m \end{bmatrix} - \begin{bmatrix} \langle u, y_0 \rangle \\ \vdots \\ \langle u, y_n \rangle \\ \langle u, u_1 \rangle \\ \vdots \\ \langle u, u_m \rangle \end{bmatrix},$$

или в матричных обозначениях

$$\frac{1}{2} \nabla E = M\theta - \psi. \quad (9.6)$$

Вектор ∇E с элементами ζ_i и η_j будем называть *вектором градиента*; вектор θ с элементами α_i и β_j — *вектором параметров* и вектор ψ с элементами $\langle u, y_i \rangle$ и $\langle u, u_j \rangle$ — *эталонным вектором*.

Подобно тому как это делалось в гл. 6, можно установить следующие свойства матрицы M :

а) Из определения M следует, что $\langle u_j, y_i \rangle = \langle y_i, u_j \rangle$, а, значит, M симметрична. Кроме того, элементы главной диагонали неотрицательны.

б) Элементы матрицы зависят лишь от входного сигнала объекта u и линейных операторов модели; они не зависят от вектора коэффициентов, если (как предполагалось) этот вектор постоянен.

в) Каждая компонента вектора градиента зависит от всех параметров, что ведет к взаимной зависимости параметров.

г) Используя свойства ортогональности для определенных входных сигналов, можно добиться обращения в нуль некоторых элементов матрицы (см. гл. 3 и 4).

д) Если матрица M невырожденная, то, умножая обе части равенства (9.6) на M^{-1} , получим в правой части $M^{-1}M = I$ (I — единичная матрица). Таким образом, если вычисление M^{-1} можно реализовать в схеме оценивания, то коэффициенты становятся взаимно независимыми.

В общем виде уравнение (9.6) записывается довольно громоздко. Следует, однако, помнить, что с помощью этого уравнения определяется не менее $m + n + 1$ неизвестных коэффициентов. Более того, нетрудно понять, что при решении задачи нет необходимости определять все элементы матрицы M явным образом. Использовать уравнение (9.6) можно двумя способами:

а) Положим $\nabla E = 0$. Тогда формула (9.6) сводится к

$$M\theta = \psi, \quad \theta = M^{-1}\psi. \quad (9.7)$$

Если матрица M невырожденная, это уравнение разрешается относительно неизвестных элементов вектора θ , так как все элементы M и ψ можно определить с помощью соответствующих процедур. В результате приходим к классу методов, связанных с *применением явных математических соотношений* (см. гл. 2) и подробно обсуждаемых в гл. 8.

б) Выберем оператор, связывающий вектор градиента ∇E с вектором параметров θ :

$$\theta = \mathcal{P} \left[\frac{1}{2} \nabla E \right], \quad (9.8)$$

и будем называть \mathcal{P} *оператором политики* или просто *политикой*. Подстановка выражения (9.8) в (9.6) дает

$$\theta = \mathcal{P} [M\theta - \psi]. \quad (9.9)$$

Правильный выбор \mathcal{P} позволяет получить соотношение

$$[\zeta_0, \dots, \zeta_n, \eta_1, \dots, \eta_m] \rightarrow [0, \dots, 0, 0, \dots, 0].$$

За исключением патологических случаев, это означает, что

$$[a_0, \dots, a_n, b_1, \dots, b_m] \rightarrow [a_0, \dots, a_n, b_1, \dots, b_m],$$

где $[a_i, b_j]$ — параметры объекта.

Вследствие принятого условия постоянства вектора коэффициентов оператор политики должен иметь прерывистое строение, т. е. предусматривать проведение измерений на интервале с постоянными коэффициентами, корректировку коэффициентов модели по истечении этого интервала в соответствии с результатами измерений, затем вновь измерения, вновь корректировку и т. д. Следовательно, обозначая через i номер очередного цикла измерений и настройки, можно написать для вектора ошибки в конце интервала измерений соотношение

$$\frac{1}{2} \nabla E(i) = M\theta(i) - \Psi(i). \quad (9.10)$$

В следующем интервале настройки вектор коэффициентов изменяется в соответствии с формулой

$$\theta(i+1) = \theta(i) - \Gamma(i) \nabla E(i), \quad (9.11)$$

где $\Gamma(i)$ — матрица с действительными элементами, характеризующая коэффициент усиления контура настройки. Для этого нового значения θ на $(i+1)$ -м интервале измерений вектор ошибки определяется как

$$\frac{1}{2} \nabla E(i+1) = M\theta(i+1) - \Psi(i+1).$$

Эта процедура приводит к системе разностных уравнений. Выбрав оператор политики в виде

$$\dot{\theta} = -\frac{1}{2} \Gamma \nabla E, \quad (9.12)$$

приходим к дифференциальному уравнению

$$\dot{\theta} + \Gamma M \theta = \Gamma \Psi. \quad (9.13)$$

Строго говоря, вывод этого уравнения некорректен, так как теперь θ — переменная величина, тогда как равенство (9.6) было получено в предположении постоянства θ . При достаточно медленной сходимости уравнение (9.13) все же позволяет приближенно оценить динамические свойства объекта. Практическая реализация этих соотношений представляет собой частный случай упомянутого в гл. 2 метода настраиваемой модели; настраиваемая модель осуществляет аппроксимацию характеристик объекта.

Если модель может достаточно полно описывать динамику объекта и шум отсутствует, то существует множество параметров $[a_0, \dots, a_n, b_1, \dots, b_m]$, для которых справедливо соотношение

$$\sum_{i=0}^n a_i y_i + \sum_{j=1}^m b_j u_j - u = 0. \quad (9.14)$$

Вычитание (9.14) из (9.2) дает

$$e = \sum_{i=0}^n (\alpha_i - a_i) y_i + \sum_{j=1}^m (\beta_j - b_j) u_j. \quad (9.15)$$

Следовательно, ошибка линейно зависит от величин отклонений параметров модели от их оптимальных значений. Выбирая вновь в качестве критерия ошибки условие минимума (9.4), дифференцируя и используя обозначения (9.5а) и (9.5б), получаем

$$\frac{1}{2} \begin{bmatrix} \zeta_1 \\ \vdots \\ \zeta_n \\ \eta_1 \\ \vdots \\ \eta_m \end{bmatrix} = \begin{bmatrix} \langle y_0, y_0 \rangle \dots \langle y_n, y_0 \rangle \langle u_1, y_0 \rangle \dots \langle u_m, y_0 \rangle \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \ddots \end{bmatrix} \begin{bmatrix} \alpha_0 - a_0 \\ \vdots \\ \alpha_n - a_n \\ \beta_1 - b_1 \\ \vdots \\ \beta_m - b_m \end{bmatrix},$$

или в матричной записи

$$\frac{1}{2} \nabla E = \mathbf{M} \lambda, \quad (9.16)$$

где ∇E и \mathbf{M} имеют тот же смысл, что и в (9.8), а вектор λ с элементами $\alpha_i - a_i$ и $\beta_j - b_j$ — вектор ошибок оценок параметров.

Как и раньше, можно использовать разные политики, например

$$\Delta \lambda = -\frac{1}{2} \Gamma \nabla E, \quad (9.17)$$

т. е.

$$\Delta \lambda = -\Gamma \mathbf{M} \lambda.$$

Отсюда следует векторное разностное уравнение

$$\Delta \lambda = \lambda(i+1) - \lambda(i) = -\Gamma \mathbf{M} \lambda(i),$$

или

$$\lambda(t+1) = [\mathbf{I} - \Gamma \mathbf{M}] \lambda^*(t), \quad (9.18)$$

$$\lambda(t) = [\mathbf{I} - \Gamma \mathbf{M}]^t \lambda(0). \quad (9.18a)$$

По-прежнему коэффициент усиления Γ может зависеть от номера итерации t .

Можно положить

$$\dot{\lambda} = -\frac{1}{2} \Gamma \nabla E,$$

откуда

$$\lambda = -\Gamma \mathbf{M} \lambda, \quad (9.19)$$

или

$$\lambda(t) = \lambda(0) \exp[-\Gamma \mathbf{M} t], \quad (9.19a)$$

где $\lambda(0)$ — вектор ошибок при $t = 0$. Поскольку матрица \mathbf{M} вещественная и симметричная, ее собственные значения вещественные. Поэтому, как следует из теории функций от матриц, $\lambda(t) \rightarrow 0$ при $t \rightarrow \infty$, если все собственные значения \mathbf{M} положительные.

Выбор критерия ошибки и способы обработки информации

Допустим, что параметр $b(t)$ объекта P изменяется как произвольная неизвестная функция времени, а $\beta(t)$ — параметр модели M , изменения которого должны следовать за изменениями $b(t)$ (отслеживать параметр объекта). Величины b и β могут быть коэффициентами дифференциальных или разностных уравнений, передаточных функций (непрерывных или дискретных). Они могут также задавать расположение нулей и полюсов передаточной функции, координаты или моменты весовой функции, координаты амплитудной и фазовой характеристик.

Ошибку, характеризующую расхождение динамических свойств объекта P и модели M , можно определить многими способами. В качестве критерия можно использовать условие минимума

$$\int_{t-T}^t \{b(t) - \beta(t)\}^2 dt \quad (9.20)$$

или какого-либо аналогичного выражения. Явное выражение для $\beta(t)$ удается получить лишь в отдельных наиболее простых случаях. Кроме того, в реальных задачах функция $b(t)$ не известна. Поэтому пользоваться критерием (9.20) удается только в модельных задачах. Вследствие этого используется другое определение ошибки: $e = y - w$. В качестве критерия можно рассматривать среднее значение квадрата ошибки по тому или иному интервалу времени. Впрочем, имеется достаточная свобода в выборе критерия ошибки, например

$$E_1 = \mathcal{E} \left[\int_{t-T}^t e^2 dt \right] \stackrel{\Delta}{=} \langle e, e \rangle, \quad (9.21a)$$

$$E_2 = \mathcal{E} \left[\int_{t-T}^t |e| dt \right] \stackrel{\Delta}{=} \langle |e|, 1 \rangle, \quad (9.21b)$$

$$E_3 = \mathcal{E} \left[\int_0^\infty e^2 (t-\theta) h(\theta) d\theta \right] \stackrel{\Delta}{=} \langle e, e \rangle_h, \quad (9.21v)$$

$$E_4 = \mathcal{E} \left[\int_0^\infty |e(t-\theta)| h(\theta) d\theta \right] \stackrel{\Delta}{=} \langle |e|, 1 \rangle_h. \quad (9.21g)$$

Эти критерии имеют следующие названия: E_1 — среднеинтегральная квадратическая ошибка, E_2 — среднеинтегральная абсолютная ошибка, E_3 — средневзвешенная квадратическая ошибка, E_4 — средневзвешенная абсолютная ошибка. Интервал интегрирования в E_1 и E_2 «скользящий» вдоль оси t . Другая возможность состоит в использовании заранее заданных на оси времени интервалов (прерывистые схемы). В E_3 и E_4 производится взвешивание ошибки с помощью функций, придающих относительно меньший вес более старой информации, т. е. здесь «прошлое постепенно забывается». Этого можно достичь, применяя вместо интегратора низкочастотный фильтр с весовой функцией $h(t) \rightarrow 0$ при $t \rightarrow \infty$.

Будем, как и прежде, для сокращения записи использовать угловые скобки, как в формулах (9.21). Дифференцирование (9.21) по α_i и β_j позволяет по аналогии с (9.5)

получить

$$\frac{1}{2} \zeta_i = \frac{1}{2} \frac{\partial E_1}{\partial \alpha_i} = \langle e, y_i \rangle, \quad \frac{1}{2} \eta_j = \frac{1}{2} \frac{\partial E_1}{\partial \beta_j} = \langle e, u_j \rangle, \quad (9.22a)$$

$$\zeta_i = \frac{\partial E_2}{\partial \alpha_i} = \left\langle \frac{e}{|e|}, y_i \right\rangle, \quad \eta_j = \frac{\partial E_2}{\partial \beta_j} = \left\langle \frac{e}{|e|}, u_j \right\rangle, \quad (9.22b)$$

$$\frac{1}{2} \zeta_i = \frac{1}{2} \frac{\partial E_3}{\partial \alpha_i} = \langle e, y_i \rangle_h, \quad \frac{1}{2} \eta_j = \frac{1}{2} \frac{\partial E_3}{\partial \beta_j} = \langle e, u_j \rangle_h, \quad (9.22b)$$

$$\zeta_i = \frac{\partial E_4}{\partial \alpha_i} = \left\langle \frac{e}{|e|}, y_i \right\rangle_h, \quad \eta_j = \frac{\partial E_4}{\partial \beta_j} = \left\langle \frac{e}{|e|}, u_j \right\rangle_h. \quad (9.22c)$$

Разумеется, используемый критерий определяет форму практической реализации. Член $e / |e| = \text{sgn} e = \pm 1$ (сигнум-функция) в формулах (9.22) приводит к двоичным операциям, например, умножение $x \text{sgn} e$ можно осуществить устройством типа реле, управляемым сигналом e и переключающимся от $+x$ к $-x$.

В гл. 3 приводилась классификация используемых сигналов. Названия *непрерывные* (на наш взгляд, лучше было бы говорить *недискретные*) и *дискретные* (или *импульсные*) относятся к способу представления аргумента рассматриваемых сигналов, т. е. к непрерывному или дискретному времени. По отношению к амплитуде функций времени можно различать *аналоговые*, *квантованные* и *двоичные* сигналы. При изучении изложенного в следующих разделах нужно помнить, что для каждого из упомянутых типов сигналов можно предложить ряд схем практической реализации.

Выбор динамических операторов

В предыдущих разделах ничего не было сказано об операторах F_0, \dots, F_m и G_0, \dots, G_n , являющихся частью обобщенной модели на фиг. 9.1. Очевидно, что от выбора этих операторов зависят свойства схемы оценивания параметров, например вид получаемой информации, скорость сходимости к решению, возможное смещение решения,

помехоустойчивость, возможности упрощения практической реализации с помощью соотношений ортогональности.

Ниже указаны некоторые возможные *линейные* операторы:

F_j, G_i — дифференциаторы;

F_j, G_i — высокочастотные фильтры;

F_j, G_i — интеграторы;

F_j, G_i — низкочастотные фильтры;

F_j, G_i — полосовые фильтры;

F_j — элементы чистой временной задержки; $a_0 = 1$,
 $G_0 = 1$ и $G_i = 0$ при $i > 0$;

F_j — ортогональные фильтры, $a_0 = 1$, $G_0 = 1$ и $G_i = 0$ при $i > 0$.

Использование последних двух операторов приводит к схеме фиг. 2.3, где модель уже не является обобщенной. В таких случаях сигнал $e = y - w$ можно назвать *ошибкой, записанной в зависимых переменных*, поскольку как y , так и w зависят от входной переменной u .

Необходимо отметить, что, конечно, можно использовать и «смешанные» наборы операторов. Линейный оператор может включать также умножитель сигнала u (или y), подаваемого на один вход умножителя, на произвольную *независимую* функцию времени, подаваемую на другой вход. Это устройство может потребоваться при преобразовании компонент сигнала u (или y) к ортогональной форме.

Выбор линейных операторов зависит от полноты и простоты представления объекта. Кроме того, он определяется типом информации об объекте, которую требуется получить (например, информацией во временной или частотной области). Важно помнить, что достижение наилучшего в некотором смысле согласия с наблюдаемыми данными зависит от предположений, используемых при каждом конкретном способе описания модели. Часто также с трудом поддается анализу преобразование ошибок (неопределенности) при переходе от одного способа описания к другому.

В следующей таблице приводятся некоторые свойства различных типов моделей (см. также фиг. 2.13).

Модель	Возможности использования априорной информации	Выход линеен относительно параметров	Взаимозависимость настроек параметров	Шум вызывает смещение
С временной задержкой	Незначительные ¹⁾	Да	Незначительная	Нет
С ортогональными фильтрами	Незначительные ¹⁾	Да	Незначительная	Нет
Обобщенная Топологически идентичная	Большие ²⁾ Большие ²⁾	Да Нет	Средняя Большая	Да Нет

¹⁾ Объект (приближенно) линеен; известна приблизительная длина интервала, где весовая функция отлична от нуля.

²⁾ Известны порядок и вид дифференциальных уравнений, порядок величины оцениваемых параметров, значения отдельных параметров.

Аналогичные соображения справедливы и для нелинейных моделей. Условие, согласно которому ошибка должна быть линейна относительно параметров, ограничивает выбор таких нелинейных динамических операторов некоторым специальным классом (см. гл. 5).

Помимо изложенного выше, представляется достаточно важным рассмотреть взаимозависимость настроек модели. Для минимизации взаимозависимости параметров можно выбрать ортогональные фильтры в соответствии с разд. 3.2 и 3.3. При этом

$$\int_0^T u_i u_j dt = \langle u_i, u_j \rangle = c \delta_{ij}. \quad (9.23)$$

Учет выражений для критерия ошибки и ошибки

$$E = \int_0^T e^2 dt = \langle e, e \rangle, \quad e = y - \sum_{j=0}^m \beta_j u_j$$

приводит к следующему упрощенному соотношению:

$$-\frac{1}{2} \nabla E = \begin{bmatrix} \langle u_0, u_0 \rangle \dots \langle u_m, u_0 \rangle \\ \vdots \\ \langle u_0, u_m \rangle \dots \langle u_m, u_m \rangle \end{bmatrix} \begin{bmatrix} \beta_0 \\ \vdots \\ \beta_m \end{bmatrix} - \begin{bmatrix} \langle y, u_0 \rangle \\ \vdots \\ \langle y, u_m \rangle \end{bmatrix} = c \begin{bmatrix} \beta_0 \\ \vdots \\ \beta_m \end{bmatrix} - \begin{bmatrix} \langle y, u_0 \rangle \\ \vdots \\ \langle y, u_m \rangle \end{bmatrix},$$

или

$$-\frac{1}{2} \frac{\partial E}{\partial \beta_j} = c \beta_j - \langle y, u_j \rangle, \quad (9.24)$$

т.е. теперь производная критерия по любому из параметров не зависит от остальных параметров. Оптимальное значение параметра, определяемое из условия

$$\frac{\partial E}{\partial \beta_j} = 0,$$

равно

$$\hat{\beta}_j = \frac{\langle y, u_j \rangle}{c}.$$

Из формулы

$$E = \int_0^T e^2 dt = \int_0^T \left\{ y - \sum \beta_j u_j \right\}^2 dt$$

следует, что остаточное значение критерия равно

$$E_{\text{res}} = \int_0^T y^2 dt - c \sum_{j=0}^m \hat{\beta}_j^2. \quad (9.25)$$

Отсюда виден вклад каждого параметра модели β_j в остаточное значение критерия. Это выражение можно использовать для оценки необходимого числа параметров модели. Увеличение числа параметров модели до $m + 1, \dots$ не влияет на оптимальные значения параметров β_0, \dots, β_m .

Пока остается открытым вопрос о том, как нужно выбирать весовые функции динамических систем, чтобы их выходные сигналы $u_j(t)$ были ортогональны. Пусть желательно выполнение равенства

$$\frac{1}{T} \int_0^T u_i(t) u_j(t) dt = c \delta_{ij},$$

где

$$u_i(t) = \int_0^\infty u(t - \xi) g_i(\xi) d\xi,$$

$$u_j(t) = \int_0^\infty u(t - \xi) g_j(\xi) d\xi.$$

Отсюда следует условие

$$\begin{aligned} c\delta_{ij} &= \frac{1}{T} \int_0^T \int_0^\infty \int_0^\infty u(t - \xi) u(t - \nu) g_i(\xi) g_j(\nu) dt d\xi d\nu = \\ &= \int_0^\infty \int_0^\infty \tilde{\Psi}_{uu}(\nu - \xi, T) g_i(\xi) g_j(\nu) d\xi d\nu. \end{aligned} \quad (9.26)$$

Итак, для преобразования $u(t)$ в ортогональные сигналы $u_j(t)$ весовые функции динамических систем должны удовлетворять условию (9.26). Поэтому эти весовые функции зависят от характеристик входного сигнала. Для белого шума

$$\mathcal{E}[\tilde{\Psi}_{uu}(\nu - \xi, T)] = \sigma_u^2 \delta(\nu - \xi),$$

так что взятие математических ожиданий приводит к простому условию

$$\int_0^\infty g_i(\xi) g_j(\xi) d\xi = \frac{c}{\sigma_u^2} \delta_{ij}.$$

При произвольных входных сигналах определение весовых функций оказывается непростой задачей (см. разд. 3.2).

Очевидно, что при конкретных применениях одна ортогональная система сигналов может оказаться «лучше» другой. Под этим «лучше» может подразумеваться уменьшение числа членов, необходимых для аппроксимации (с заданной точностью) динамики объекта. Отсюда вытекает возможность настраивать отдельно некоторые параметры ортогональных фильтров таким образом, чтобы при заданном числе фильтров минимизировалась остаточная ошибка. Об ортогональных операторах для нелинейных моделей и моделей со многими входами и выходами см. гл. 4.

9.2. МИНИМИЗАЦИЯ МГНОВЕННОГО ЗНАЧЕНИЯ ОШИБКИ

В предыдущем разделе было упомянуто несколько возможных критериев ошибок. Для устранения элемента случайности использовалась операция математического ожидания. При практической реализации системы с настраиваемой моделью можно пользоваться только одним элементом статистического ансамбля. В данном разделе рассматривается минимизация мгновенного значения ошибки; минимизация интегральной ошибки является предметом следующего раздела.

Основные уравнения

Обобщенная модель показана на фиг. 9.1. Будем считать шум пренебрежимо малым. Ранее обобщенная ошибка была определена как

$$e = \sum_{i=0}^n \alpha_i y_i + \sum_{j=1}^m \beta_j u_j - u. \quad (9.27)$$

Критерием ошибки является условие минимума

$$E = e^2. \quad (9.28)$$

Отсюда следует, что

$$\begin{aligned} \frac{1}{2} \zeta_i &= \frac{1}{2} \frac{\partial E}{\partial \alpha_i} = e \frac{\partial e}{\partial \alpha_i} = e y_i, \\ \frac{1}{2} \eta_j &= \frac{1}{2} \frac{\partial E}{\partial \beta_j} = e \frac{\partial e}{\partial \beta_j} = e u_j, \end{aligned} \quad (9.29)$$

или в векторной форме

$$\frac{1}{2} \nabla E = M\theta - \psi, \quad (9.30)$$

где

$$\begin{aligned} \nabla E &= [\zeta_0, \dots, \zeta_n, \eta_1, \dots, \eta_m]', \\ \theta &= [\alpha_0, \dots, \alpha_n, \beta_1, \dots, \beta_m]', \\ \psi &= [uy_0, \dots, uy_n, uu_1, \dots, uu_m]', \end{aligned}$$

$$M = \begin{bmatrix} y_0 y_0 \dots y_n y_0 & u_1 y_0 \dots u_m y_0 \\ \vdots & \vdots & \vdots & \vdots \\ y_0 y_n \dots y_n y_n & u_1 y_n \dots u_m y_n \\ y_0 u_1 \dots y_n u_1 & u_1 u_1 \dots u_m u_1 \\ \vdots & \vdots & \vdots & \vdots \end{bmatrix}.$$

Те же рассуждения справедливы и для следующего критерия:

$$E = |e| \quad (9.28a)$$

В этом случае

$$\begin{aligned}\zeta_i &= \frac{\partial E}{\partial \alpha_i} = \frac{e}{|e|} \frac{\partial e}{\partial \alpha_i} = \frac{e}{|e|} y_i, \\ \eta_j &= \frac{\partial E}{\partial \beta_j} = \frac{e}{|e|} \frac{\partial e}{\partial \beta_j} = \frac{e}{|e|} u_j,\end{aligned}\quad (9.29a)$$

откуда получается векторное уравнение

$$\nabla E = \frac{M\theta - \Psi}{|e|}, \quad (9.31)$$

где ∇E , θ , Ψ , M имеют тот же смысл, что и раньше.

Интересно, что при таком виде критерия нет необходимости требовать, чтобы вектор θ был постоянным, как это делалось в разд. 9.1; уравнения (9.30) и (9.31) остаются справедливыми и при переменном θ . Поэтому можно выбрать оператор политики

$$\theta = \mathcal{P} \left[\frac{1}{2} \nabla E \right], \quad (9.32)$$

например, как в (9.12):

$$\dot{\theta} = -\frac{1}{2} \Gamma \nabla E. \quad (9.33)$$

С учетом равенства (9.30) отсюда получается матричное дифференциальное уравнение

$$\dot{\theta} + \Gamma M \theta = \Gamma \Psi. \quad (9.34)$$

Если вход u случаен, элементы матрицы M оказываются также случайными. Это уравнение можно еще упростить, предполагая, что модель может идеально описывать объект [18]. Это означает, что из уравнений

$$\sum_{i=0}^n a_i y_i + \sum_{j=1}^m b_j u_j - u = 0 \quad (9.35)$$

и

$$e = \sum_{i=0}^n (\alpha_i - a_i) y_i + \sum_{j=1}^m (\beta_j - b_j) u_j \quad (9.36)$$

следует

$$\dot{\lambda} = -\Gamma M \lambda \quad (9.37)$$

[см. формулы (9.14), (9.15) и (9.19)]. Для стационарного случайного входного сигнала элементы матрицы M оказываются стационарными случайными функциями.

Векторные уравнения (9.30) и (9.32) описывают *поведение* системы настройки. Для *практической реализации* достаточно уравнений (9.27) и (9.33):

$$\dot{\theta} = -\frac{1}{2} \Gamma \nabla E, \quad \begin{bmatrix} \dot{\alpha}_i \\ \dot{\beta}_j \end{bmatrix} = -\Gamma \begin{bmatrix} ey_i \\ eu_j \end{bmatrix}. \quad (9.38)$$

Используя простой матричный коэффициент усиления $\Gamma = \gamma I$, получаем

$$\begin{aligned} \dot{\alpha}_i &= -\gamma ey_i, \\ \dot{\beta}_j &= -\gamma eu_j. \end{aligned} \quad (9.38a)$$

Пример. Рассмотрим объект с простой передаточной функцией

$$H(s) = \frac{b}{1+as},$$

в которой нужно оценить a и b . В обозначениях преобразования Лапласа при нулевых начальных условиях объект описывается уравнением

$$bU(s) - asY(s) - Y(s) = 0,$$

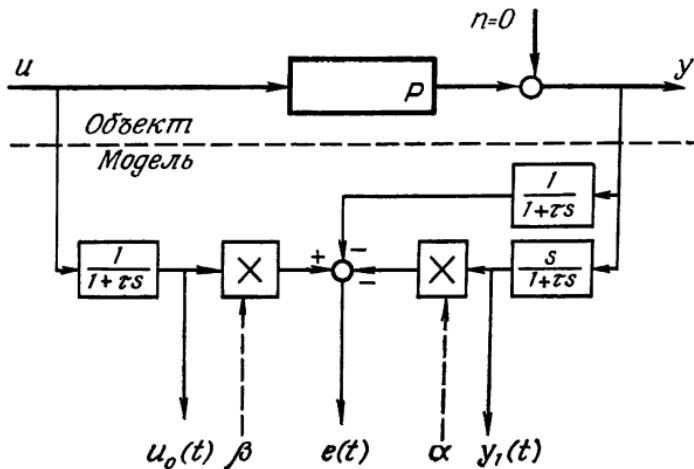
или

$$\begin{aligned} W(s) \{bU(s) - asY(s) - Y(s)\} &= 0, \\ bU_0(s) - aY_1(s) - Y_0(s) &= 0, \end{aligned}$$

где $W(s)$ — произвольная передаточная функция (фильтр), $U_0(s) = W(s)U(s)$, $Y_0(s) = W(s)Y(s)$, $Y_1(s) = W(s)sY(s)$.

Соответствующим выбором $W(s)$ можно добиться простой практической реализации передаточной функции $W(s)s$ по сравнению со случаем отсутствия вспомогательного фильтра, например

$$W(s) = \frac{1}{1+\tau s}, \quad W(s)s = \frac{s}{1+\tau s}.$$



Ф и г. 9.2.

Уравнение модели теперь записывается в виде

$$\beta U_0(s) - \alpha Y_1(s) - Y_0(s) = \mathcal{L}[e(t)]. \quad (9.39)$$

На фиг. 9.2 изображена модель для данного конкретного выбора $W(s)$.

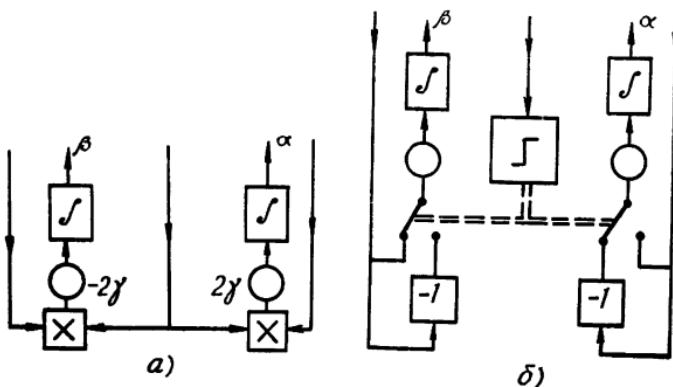
Этот пример иллюстрирует общий принцип. С применением дополнительного фильтра появляется возможность, например, дифференцирования сигналов. Этот результат можно считать прямым следствием концепции обобщенной модели. Примеры конкретного выбора моделей можно найти в [11, 25].

Теперь у нас есть возможность выбора между несколькими критериями. При $E = e^2(t)$ получаем

$$\begin{aligned} \frac{\partial E}{\partial \alpha} &= \frac{\partial e^2}{\partial \alpha} = 2e \frac{\partial e}{\partial \alpha} = 2e(-y_1), \\ \frac{\partial E}{\partial \beta} &= \dots = 2eu_0 \end{aligned} \quad (9.40)$$

и градиентные уравнения вида

$$\begin{aligned} \frac{d\alpha}{dt} &= -\gamma \frac{\partial E}{\partial \alpha} = -2\gamma e(-y_1), \\ \frac{d\beta}{dt} &= \dots = -2\gamma eu_0. \end{aligned} \quad (9.41)$$



Ф и г. 9.3.

Отсюда следует, что фиг. 9.2 нужно дополнить с помощью схемы фиг. 9.3, а. Выбирая критерий $E = |e|$, приходим к схеме фиг. 9.3, б, где вместо пары аналоговых умножителей появляются простые реле.

Свойства сходимости

Свойства сходимости и устойчивости схем оценивания параметров, приведенных в предыдущем разделе, часто оказывается довольно трудно установить. Поэтому здесь приводится другой подход, когда основное требование к оценке задается не в виде критерия ошибки, а в виде условия монотонной сходимости модели [6].

Объект задается уравнением

$$g(t, b_1, \dots, b_m) = 0, \quad (9.42)$$

где b_1, \dots, b_m — неизвестные параметры. Это уравнение справедливо в промежутке времени, когда зависимость от времени определяется несколькими функциями $u_i(t)$ и $y_i(t)$, а также их производными. Модель имеет вид

$$g(t, \beta_1, \dots, \beta_m) = e, \quad (9.43)$$

где e — ошибка, подлежащая минимизации. Теперь, однако, критерий, в соответствии с которым производится минимизация, не определяется заранее, а вытекает из рассуждений следующего типа. В общем случае справедливо соот-

ношение

$$\frac{de}{dt} = \sum_{j=1}^m \frac{\partial e}{\partial \beta_j} \frac{d\beta_j}{dt} + \frac{\partial e}{\partial t}. \quad (9.44)$$

Почленно умножим это уравнение на некоторую нечетную функцию $-f$ ошибки e , $f(-e) = -f(e)$. Допустим, что выполняется условие

$$-f(e) \sum_{j=1}^m \frac{\partial e}{\partial \beta_j} \frac{d\beta_j}{dt} > \left| -f(e) \frac{\partial e}{\partial t} \right| \text{ при } 0 \leq e_{\min} < |e|. \quad (9.45)$$

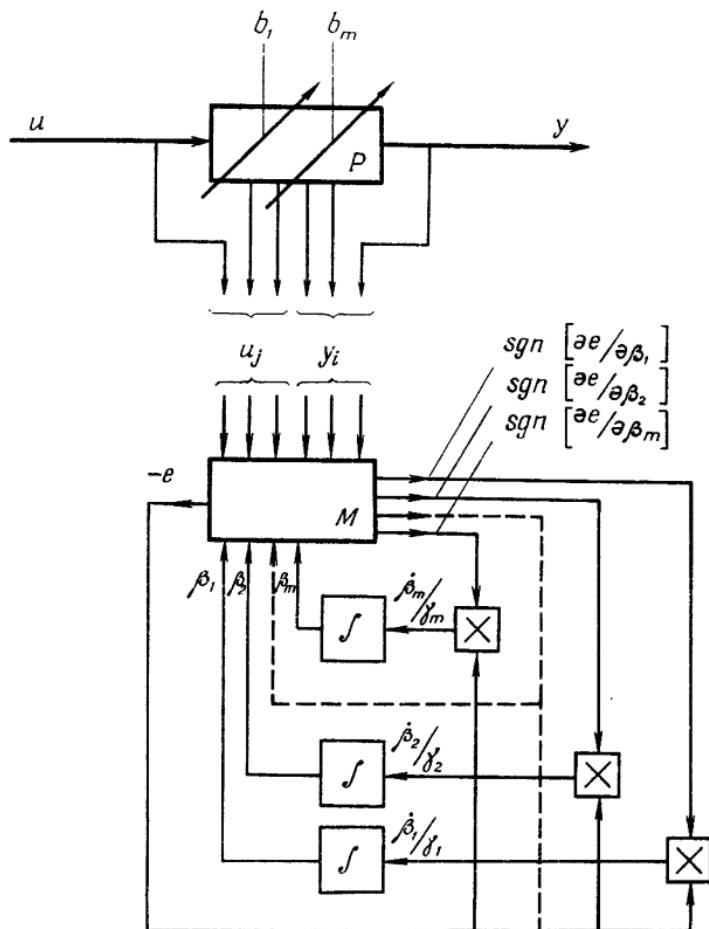
Тогда

$$-f(e) \frac{de}{dt} > 0 \text{ при } 0 \leq e_{\min} < |e|. \quad (9.46)$$

Последнее неравенство обеспечивает монотонную сходимость ошибки к нулю, пока выполняется условие (9.45). Как только ошибка уменьшится настолько, что это условие окажется нарушенным, величина ошибки начнет колебаться в интервале $-e_{\min} < e < e_{\min}$. При конкретном выборе $f(e)$ существует несколько возможных способов выбора $d\beta/dt$, обеспечивающих выполнение неравенства (9.45). Несколько примеров для случая настройки одного параметра приведено ниже. Пусть $f(e) = -e^{-1}$. Тогда

$\frac{d\beta}{dt} =$	$-\frac{1}{e} \frac{\partial e}{\partial \beta} \frac{d\beta}{dt} =$
$-\gamma e \frac{\partial e}{\partial \beta}$	$\gamma \left(\frac{\partial e}{\partial \beta} \right)^2 \geq 0$
$-\gamma e \operatorname{sgn} \left[\frac{\partial e}{\partial \beta} \right]$	$\gamma \left \frac{\partial e}{\partial \beta} \right \geq 0$
$-\gamma \operatorname{sgn} [e] \frac{\partial e}{\partial \beta}$	$\gamma \left(\frac{\partial e}{\partial \beta} \right)^2 \frac{1}{ e } \geq 0$
$-\gamma \operatorname{sgn} [e] \operatorname{sgn} \left[\frac{\partial e}{\partial \beta} \right]$	$\gamma \left \frac{\partial e}{\partial \beta} \right \frac{1}{ e } \geq 0$

при $\gamma > 0 \quad (9.47)$



Ф и г. 9.4.

На фиг. 9.4 схематически показаны реализация одного из этих случаев, а также случай одновременной настройки m параметров. Коэффициенты усиления интеграторов γ_k выбираются на основе компромисса между помехоустойчивостью (малые γ_k) и скоростью сходимости (большие γ_k).

Пример. Используем, например, метод наискорейшего спуска для настройки каждого параметра:

$$\frac{d\beta_i}{dt} = -\gamma \frac{\partial E}{\partial \beta_i}. \quad (9.48)$$

Тогда

$$E = e^2 \Rightarrow \frac{\partial E}{\partial \beta_i} = 2e \frac{\partial e}{\partial \beta_i} \text{ и } \frac{d\beta_i}{dt} = -2\gamma e \frac{\partial e}{\partial \beta_i}, \quad (9.49a)$$

$$E = |e| \Rightarrow \frac{\partial E}{\partial \beta_i} = \operatorname{sgn}[e] \frac{\partial e}{\partial \beta_i} \text{ и } \frac{d\beta_i}{dt} = -\gamma \operatorname{sgn}[e] \frac{\partial e}{\partial \beta_i}. \quad (9.49b)$$

Таким образом, соотношения (9.47) гарантируют монотонную сходимость процедуры наискорейшего спуска при использовании указанных выше критериев для настройки одного параметра.

Пользуясь формулой для (обобщенной) ошибки (9.23), можно в случае $E = e^2$ записать неравенство (9.45) в виде

$$2\gamma \left\{ \sum_{j=1}^m u_j^2 + \sum_{i=0}^n y_i^2 \right\} > \left| -\frac{1}{e} \frac{\partial e}{\partial t} \right| \quad \text{при } 0 \leq e_{\min} < |e|. \quad (9.50)$$

Для случая критерия $E = |e|$ из формулы (9.45) получаем

$$\frac{\gamma}{|e|} \left\{ \sum u_j^2 + \sum y_i^2 \right\} > \left| -\frac{1}{e} \frac{\partial e}{\partial t} \right|,$$

или

$$\gamma \left\{ \sum u_j^2 + \sum y_i^2 \right\} > \left| \frac{\partial e}{\partial t} \right| \quad \text{при } 0 \leq e_{\min} < |e|. \quad (9.51)$$

Если входной сигнал стационарен, то u_j и y_i также стационарны. От коэффициента γ зависит усиление контура.

Подобные интуитивные рассуждения о сходимости для систем рассматриваемого типа приведены в статье [22], откуда взят также и следующий пример.

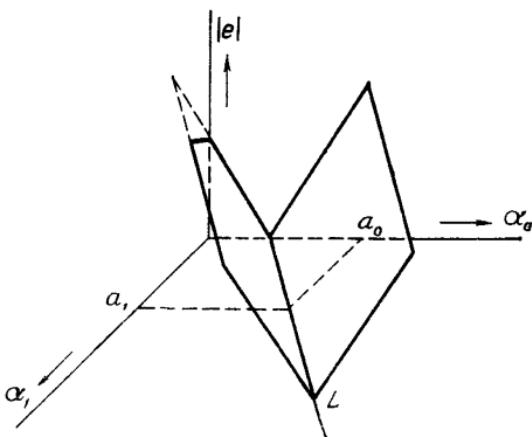
Пример. Уравнение динамики объекта имеет вид

$$a_1 \dot{y} + a_0 y = u,$$

причем коэффициенты a_1 и a_0 должны быть определены.

Для сравнения с предыдущим материалом положим $\dot{y} = y_1$, $y = y_0$, $u = u_0$. Обобщенная ошибка равна

$$e = a_1 \dot{y} + a_0 y - u = a_1 y_1 + a_0 y_0 - u_0,$$



Ф и г. 9.5.

где a_1 и a_0 — вычисленные значения коэффициентов. Абсолютная величина ошибки

$$|e| = |a_1 y_1 + a_0 y_0 - u_0|$$

представляется кусочно-линейной поверхностью в пространстве $|e(a_1, a_0)|$ (фиг. 9.5). Эта поверхность касается плоскости (a_1, a_0) по линии L , определяемой формулой

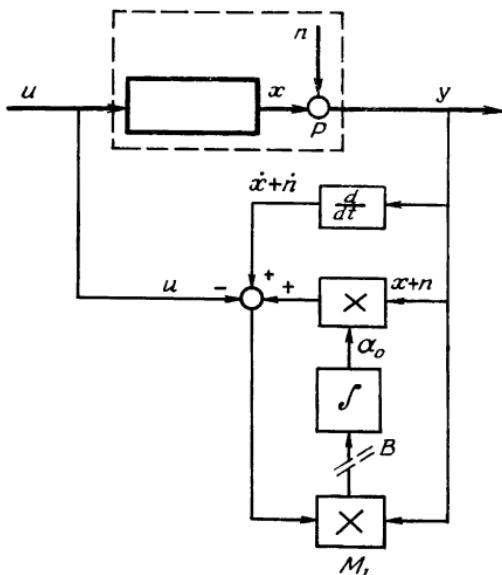
$$a_1 y_1 + a_0 y_0 - u_0 = 0.$$

Наклоны поверхности вдоль осей a_1 и a_0 вычисляются как

$$\frac{\partial |e|}{\partial a_1} = \frac{e}{|e|} y_1,$$

$$\frac{\partial |e|}{\partial a_0} = \frac{e}{|e|} y_0.$$

Справа стоят компоненты вектора градиента. Если координаты текущей точки изменяются пропорционально этим наклонам, взятым с обратным знаком, то эта точка будет двигаться к линии L по нормальной к ней траектории. Так как значения y_1 , y_0 и u_0 изменяются во времени, то линия L будет вращаться вокруг точки (a_1, a_0) . Текущая точка при этом будет равномерно стремиться к точке (a_1, a_0) . Реализация этого случая была показана на фиг. 9.3.



Ф и г. 9.6.

Исследования свойств сходимости можно найти в статье [18], где используется второй метод Ляпунова. Определим $R^2(t)$ как

$$R^2(t) = \sum_{i=0}^n (\alpha_i - a_i)^2 + \sum_{j=1}^m (\beta_j - b_j)^2 \geq 0. \quad (9.52)$$

Дифференцируя (9.52) по времени, получим

$$\frac{dR^2}{dt} = -2\gamma e \left\{ \sum (\alpha_i - a_i) y_i + \sum (\beta_j - b_j) u_j \right\} = -2\gamma e^2$$

с учетом уравнений (9.38а) и (9.36). Таким образом, $R^2(t)$ оказывается функцией Ляпунова.

Влияние аддитивного шума. До сих пор воздействие аддитивного шума не рассматривалось. Остановимся теперь на простом примере, схема которого показана на фиг. 9.6. Пусть уравнение объекта имеет вид

$$\dot{x} + a_0 x = u. \quad (9.53)$$

При отсутствии аддитивного шума n обобщенная ошибка равна

$$e = \dot{x} + a_0 x - u. \quad (9.54)$$

Теперь допустим, что $a_0 = \dot{a}_0$ и контур настройки разомкнут в точке B . Сигнал и шум независимы: $\mathcal{E}[un] = 0$ и $\mathcal{E}[\dot{un}] = 0$. При $n \neq 0$ сигнал ошибки имеет две компоненты, обусловленные шумом, а именно $a_0 n$ и \dot{n} . Выход умножителя M_1 (с обратным знаком) имеет вид

$$(x + n) \underbrace{(u - \dot{x} - a_0 x - \dot{n} - a_0 n)}_0 = -(x + n)(\dot{n} + a_0 n).$$

Поскольку сигнал и шум независимы,

$$\mathcal{E}[xn] = 0, \quad \mathcal{E}[\dot{x}\dot{n}] = 0.$$

Далее можно показать, что

$$\mathcal{E}[nn] = 0.$$

Четвертый член дает

$$a_0 \mathcal{E}[nn] > 0 \text{ при } n \neq 0. \quad (9.55)$$

В результате появляется постоянная составляющая в выходном сигнале умножителя. Поэтому при замыкании контура в точке B система выходит из равновесия. Другими словами, следует ожидать расстройки коэффициента a_0 на величину

$$a_0 \mathcal{E}[nn] = -(a_0 - \dot{a}_0) \mathcal{E}[xx]. \quad (9.56)$$

Поскольку xx — единственная комбинация входных величин M_1 , имеющая ненулевое математическое ожидание, отсюда следует, что появляется смещение коэффициента, зависящее от отношения сигнала к шуму.

Комбинация входных сигналов M_1 , не дающая вклада в математическое ожидание выхода, может, однако, увеличивать дисперсию этого выхода, повышая таким образом неопределенность a_0 .

Примеры моделирования, практической реализации и применений

Примеры подобных схем настраиваемых моделей с минимизацией мгновенного значения ошибки можно найти в ряде работ. К числу первых работ относятся статьи [3, 4], где выдвигается идея синтеза системы путем неявных

динамических вычислений. Этот прием связан с методом неявных схем в аналоговых вычислениях, примером которых может служить реализация делительного устройства с помощью умножителя и усилителя с большим коэффициентом усиления. «Неявный синтез», таким образом, представляет собой особый способ использования вычислительных устройств. Он отличается от обычных вариантов их использования, когда известны дифференциальные уравнения вместе со всеми их параметрами и возбуждающими воздействиями и требуется найти решение уравнения, являющееся зависимой переменной, как функции независимой переменной. Напротив, при неявном синтезе известны решение и возмущение и необходимо определить один или несколько неизвестных параметров дифференциального уравнения».

В этих статьях приводятся некоторые результаты цифровых вычислений для простых систем, а также иллюстрации, основанные на уравнениях типа уравнений скорости химической реакции. Дополнительную информацию о развитии этого подхода можно найти в работе [22], где рассмотрено применение этих методов для определения аэродинамических коэффициентов; см. также [23].

9.3. МИНИМИЗАЦИЯ УСРЕДНЕННОЙ ПО ВРЕМЕНИ ОШИБКИ

Основные уравнения

Здесь используются следующие обозначения разд. 9.1: обобщенная ошибка (9.2)

$$e = \sum_{i=0}^n \alpha_i y_i + \sum_{j=1}^m \beta_j u_j - u, \quad (9.57)$$

критерии ошибок, соответствующие критериям, перечисленным в (9.21):

$$\begin{aligned} \min \int_0^T e^2 dt, & \quad \mathcal{E} \left[\int_{t-T}^t e^2 dt \right] = \langle e, e \rangle, \\ \min \int_0^T |e| dt, & \quad \mathcal{E} \left[\int_0^T |e| dt \right] = \langle |e|, 1 \rangle, \end{aligned}$$

$$\min \int_0^{\infty} e^2(t-\theta) h(\theta) d\theta, \quad \mathcal{E}\left[\int_0^{\infty} e^2(t-\theta) h(\theta) d\theta\right] = \langle e, e \rangle_h,$$

$$\min \int_0^{\infty} |e(t-\theta)| h(\theta) d\theta, \quad \mathcal{E}\left[\int_0^{\infty} |e(t-\theta)| h(\theta) d\theta\right] = \langle |e|, 1 \rangle_h, \quad (9.58)$$

а также соответствующие выражения для ζ_i и η_j — компонент вектора градиента ∇E . В дальнейшем будем опускать множитель 2, появляющийся при дифференцировании квадратичного критерия. Векторное уравнение

$$\nabla E = M\theta - \psi \quad (9.59)$$

было получено в предположении, что вектор коэффициентов

$$\theta' = [a_0, \dots, a_n, \beta_1, \dots, \beta_m]$$

постоянен. Как уже отмечалось в разд. 9.1, последнее условие означает, что, строго говоря, уравнение политики

$$\dot{\theta} = -\Gamma \nabla E, \quad (9.60)$$

ведущее к дифференциальному уравнению

$$\dot{\theta} + \Gamma M \theta = \Gamma \psi, \quad (9.61)$$

не верно, поскольку вектор θ не постоянен. Впрочем, при малом усилении в контуре Γ и соответственно при медленной сходимости уравнением (9.61) можно пользоваться для приближенного описания динамического поведения системы.

Пример. Рассмотрим простой случай, когда объект описывается передаточной функцией

$$H(s) = \frac{1}{a_1 s + a_0}$$

и $\Gamma = \gamma I$. Тогда дифференциальное уравнение (9.61) имеет вид

$$-\begin{bmatrix} \dot{\alpha}_0 \\ \dot{\alpha}_1 \end{bmatrix} = \gamma \begin{bmatrix} \langle y_0, y_0 \rangle \langle y_1, y_0 \rangle \\ \langle y_0, y_1 \rangle \langle y_1, y_1 \rangle \end{bmatrix} \begin{bmatrix} \alpha_0 \\ \alpha_1 \end{bmatrix} - \gamma \begin{bmatrix} \langle u, y_0 \rangle \\ \langle u, y_1 \rangle \end{bmatrix} \quad (9.62)$$

и его решение есть

$$\begin{bmatrix} \alpha_0 \\ \alpha_1 \end{bmatrix} = \begin{bmatrix} \alpha_{00} \\ \alpha_{10} \end{bmatrix} e^{-\gamma M t} + M^{-1} \begin{bmatrix} \langle u, y_0 \rangle \\ \langle u, y_1 \rangle \end{bmatrix}, \quad (9.63)$$

где α_{00} и α_{10} — компоненты вектора расстройки при $t = 0$.

На фиг. 9.7 приведены практический пример и некоторые результаты его решения. Объектом исследования является схема автоматического измерения импеданса с помощью настраиваемой модели. Она уже приводилась на фиг. 1.8. На этот раз измеряются три параметра схемы¹⁾ C , R и L . Соответствующие элементы включаются и отключаются, чтобы продемонстрировать быстроту отклика и взаимодействие между разными каналами. Вход представляет собой треугольный периодический сигнал с основной частотой 500 Гц.

В разд. 9.1 обсуждалось использование оператора политики прерывистого типа: в i -м интервале измерений определяется вектор

$$\nabla E(i) = M\theta(i) - \Psi(i); \quad (9.64)$$

в следующем за ним интервале настройки изменяется значение вектора коэффициентов

$$\theta(i+1) = \theta(i) - \Gamma(i) \nabla E(i). \quad (9.65)$$

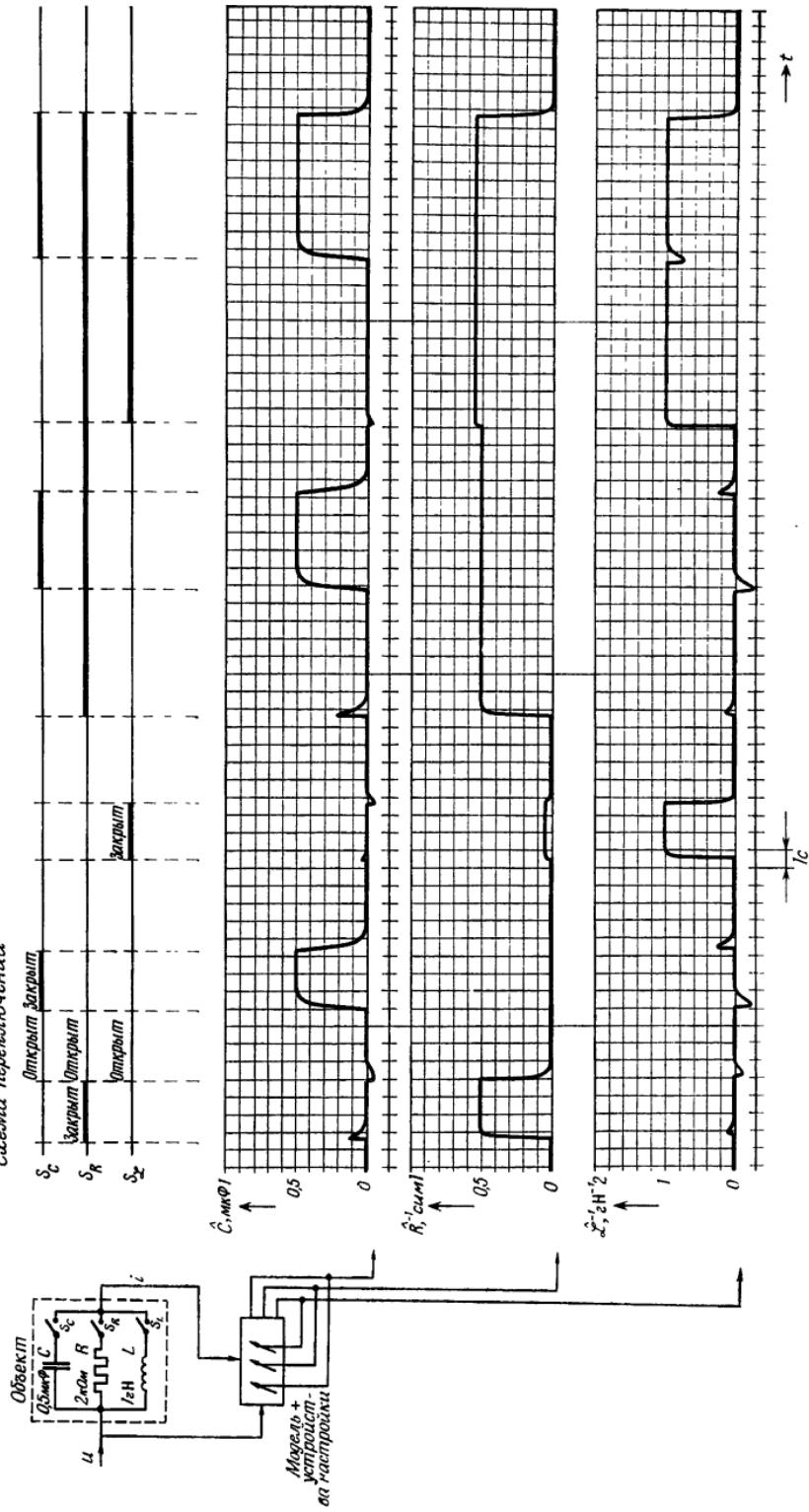
Этот вектор коэффициентов используется в $(i+1)$ -м интервале измерений и т. д. Теперь уже условие постоянства θ выполняется в течение интервала измерений. Для систем такого типа хорошо подходят критерии ошибок вида

$$\int_{iT}^{(i+1)T} e^2 dt \quad \text{и} \quad \int_{iT}^{(i+1)T} |e| dt, \quad (9.66)$$

где T — длительность интервала измерений. В работе [4] дан первый пример динамического метода наименьших квадратов.

¹⁾ Эти результаты заимствованы из проекта, выполненного в Технологическом университете Эйндховена (Голландия) Х. Х. ван де Веном, А. П. Гельсингом и др.

Система переключений



Фиг. 9.7.

Применение к управляющему устройству с переменными параметрами для аэрофотосъемки описано в [10], где указывается, что схема настройки модели устойчива, если только входной сигнал не остается постоянным в течение слишком большого интервала времени.

Вопросы сходимости

Требование устойчивости замкнутого контура настройки является необходимым условием при проектировании систем рассматриваемого типа. Поскольку система, состоящая из настраиваемой модели, объекта и механизма настройки, существенно нелинейна, проблема устойчивости оказывается нетривиальной. В работах [16, 19, 20, 24] с помощью метода Ляпунова построены различные устойчивые системы. Кушнер [14] обобщил результаты работы [16] на стохастические системы. Мощные критерии устойчивости Попова [21] и Зеймса [26] дали новое эффективное средство для конструирования механизмов настройки, обеспечивающих устойчивость систем. Первые усилия в этом направлении предпринял Ландау [15], предложивший на основе критерия Попова устойчивые системы с эталонной моделью.

9.4. ФУНКЦИИ ЧУВСТВИТЕЛЬНОСТИ ПАРАМЕТРОВ

В разд. 2.4 были указаны основные положения этого вычислительного метода, позволяющего получить частные производные (коэффициенты влияния параметров) по соответствующим параметрам системы. Эти производные можно определить одновременно с решением исходного дифференциального уравнения.

Диапазон приложения метода, основанного на изучении чувствительности (влияния) параметров, шире, чем методов оценивания параметров. Мейссингер приводит следующий список возможных применений:

- а) Предсказание решений в окрестности известного решения путем линейной экстраполяции.
- б) Определение допусков для параметров с помощью линейного прогнозирования, выделение критических параметров.

в) Приложения к статистическим исследованиям: оценивание влияния случайных параметров системы или начальных условий, экстраполяция результатов, полученных при случайных входных сигналах.

г) Оптимизация параметров системы градиентными методами в соответствии с определенным критерием качества.

д) Анализ чувствительности решения к ошибкам ЭВМ.

е) Определение границ области устойчивости системы.

ж) Изменение постоянных времени различных процессов; изменение времени нарастания, времени оседания.

з) Решение краевой задачи для обыкновенных дифференциальных уравнений.

Мы ограничимся обсуждением применения этого метода к оцениванию параметров объекта.

Методы, основанные на изучении влияния (чувствительности) параметров

Выделим теперь основные положения метода, использующего функции влияния параметров. Рассмотрим следующее неоднородное линейное дифференциальное уравнение относительно w :

$$\frac{d^2w}{dt^2} + \alpha_1 \frac{dw}{dt} + \alpha_0 w = \beta_1 \frac{du}{dt} + \beta_0 u \quad (9.67)$$

с начальными условиями

$$w(0) = \gamma_0, \quad \dot{w}(0) = \gamma_1.$$

Требуется получить решение w при конкретных значениях параметров $\alpha_1 = \alpha_{1,0}$, $\alpha_0 = \alpha_{0,0}$, $\beta_1 = \beta_{1,0}$, $\beta_0 = \beta_{0,0}$. Рассмотрим пока для наглядности только один параметр; тогда w будет функцией двух переменных, например $w = w(t; \alpha_1)$. По кривой решения, полученной при значении параметра $\alpha_{1,0}$ путем экстраполяции по $\Delta\alpha_1$, можно найти близкую кривую, соответствующую $\alpha_1 = \alpha_{1,0} + \Delta\alpha_1$,

$$w(t; \alpha_1) = w(t; \alpha_{1,0}) + \left. \frac{\partial w}{\partial \alpha_1} \right|_{\alpha_{1,0}} \Delta\alpha_1 + \\ + \frac{1}{2!} \left. \frac{\partial^2 w}{\partial \alpha_1^2} \right|_{\alpha_{1,0}} (\Delta\alpha_1)^2 + \dots \quad (9.68)$$

Необходимое для удовлетворительной аппроксимации $w(t; \alpha_{1,0} + \Delta\alpha_1)$ число членов в этом разложении зависит от величины $\Delta\alpha_1$ и поведения решения w и его частных производных по α_1 в интересующей нас области. Здесь будет рассматриваться только аппроксимация с точностью до членов первого порядка.

Частная производная $\partial w / \partial \alpha_1$, являющаяся функцией t и α_1 , называется *коэффициентом влияния* или *функцией чувствительности параметра* первого порядка. Другими коэффициентами влияния, относящимися к уравнению (9.67), являются

$$\frac{\partial w}{\partial \alpha_0}, \quad \frac{\partial w}{\partial \beta_1}, \quad \frac{\partial w}{\partial \beta_0}, \quad \frac{\partial w}{\partial \gamma_1}, \quad \frac{\partial w}{\partial \gamma_0}.$$

Два последних члена характеризуют чувствительность к изменениям начальных условий. Дифференцируя (9.67) по α_1 и учитывая, что w и \dot{w} не зависят от α_1 , получаем

$$\frac{\partial^3 w}{\partial \alpha_1 \partial t^2} + \alpha_1 \frac{\partial^2 w}{\partial \alpha_1 \partial t} + \alpha_0 \frac{\partial w}{\partial \alpha_1} = -\frac{dw}{dt}. \quad (9.69)$$

Меняя порядок дифференцирования и используя обозначение $v_1 = \partial w / \partial \alpha_1$, приходим к дифференциальному уравнению для v_1 :

$$\frac{\partial^2 v_1}{\partial t^2} + \alpha_1 \frac{\partial v_1}{\partial t} + \alpha_0 v_1 = -\frac{dw}{dt} \quad (9.70)$$

с начальными условиями

$$v_1(0) = \dot{v}_1(0) = 0,$$

следующими из того, что начальные значения w и \dot{w} постоянны и не зависят от α_1 . Уравнение (9.70) известно как *уравнение чувствительности* системы относительно параметра α_1 . При небольших изменениях α_1 из этого уравнения можно получить информацию о приближенном значении градиента v_1 . Это уравнение нетрудно промоделировать, заменяя частные производные полными:

$$\frac{d^2 v_1}{dt^2} + \alpha_1 \frac{dv_1}{dt} + \alpha_0 v_1 = -\frac{dw}{dt} \quad (9.71)$$

(*приближенное уравнение чувствительности*). Причина того, что это уравнение является всего лишь приближени-

ем, состоит в том, что соотношение между частной и полной производными имеет вид

$$\frac{dv_1}{dt} = \frac{\partial v_1}{\partial t} + \sum_{i=0}^1 \frac{\partial v_1}{\partial \alpha_i} \frac{d\alpha_i}{dt} + \sum_{j=0}^1 \frac{\partial v_1}{\partial \beta_j} \frac{d\beta_j}{dt}. \quad (9.72)$$

Следовательно, уравнение (9.71) является хорошим приближением, если изменения параметров во времени достаточно малы.

Аналогичным образом можно вывести приближенные уравнения чувствительности относительно α_0 , β_1 и β_0 . Для четырех рассматриваемых параметров получаем

$$\ddot{v}_1 + \alpha_1 \dot{v}_1 + \alpha_0 v_1 = -\dot{w}, \quad v_1 = \frac{\partial w}{\partial \alpha_1}, \quad \dot{v}_1(0) = v_1(0) = 0, \quad (9.73a)$$

$$\ddot{v}_0 + \alpha_1 \dot{v}_0 + \alpha_0 v_0 = -w, \quad v_0 = \frac{\partial w}{\partial \alpha_0}, \quad \dot{v}_0(0) = v_0(0) = 0, \quad (9.73b)$$

$$\ddot{u}_1 + \alpha_1 \dot{u}_1 + \alpha_0 u_1 = \dot{u}, \quad u_1 = \frac{\partial w}{\partial \beta_1}, \quad \dot{u}_1(0) = u_1(0) = 0, \quad (9.73c)$$

$$\ddot{u}_0 + \alpha_1 \dot{u}_0 + \alpha_0 u_0 = u, \quad u_0 = \frac{\partial w}{\partial \beta_0}, \quad \dot{u}_0(0) = u_0(0) = 0. \quad (9.73d)$$

Каждое из этих уравнений можно промоделировать с помощью отдельной модели чувствительности (см. блок-схему на фиг. 9.8). В рассматриваемом линейном случае все приближенные уравнения чувствительности оказываются одинаковыми, если не считать различий в правых частях. Это значит, что функции чувствительности параметров можно последовательно определять на одной и той же модели, используя соответствующий «связывающий член» $-\dot{w}$, $-w$, \dot{u} или u . Дальнейшие упрощения получаются, если учесть, что, согласно формулам (9.73a), (9.73b),

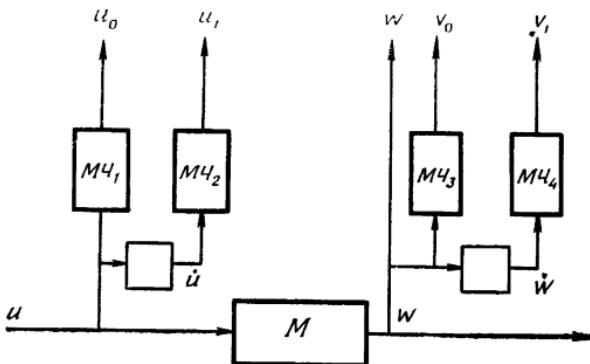
$$v_1 = \dot{v}_0, \quad (9.74)$$

согласно формулам (9.73c), (9.73d),

$$u_1 = \dot{u}_0, \quad (9.75)$$

а сравнение формулы (9.67) с (9.73c) и (9.73d) дает

$$w = \beta_1 u_1 + \beta_0 u_0. \quad (9.76)$$



Ф и г. 9.8.

Таким образом, достаточно промоделировать уравнение (9.73б) и воспользоваться соотношениями (9.74)–(9.76) для одновременного получения функций чувствительности всех четырех параметров (фиг. 9.9, б). Такая схема практической реализации требует существенно меньших затрат, чем схема, соответствующая фиг. 9.8.

Если начальные условия γ_1 и γ_0 также являются параметрами, представляющими интерес, то легко видеть, что в соответствующих уравнениях чувствительности «связывающий член» вообще отсутствует. При $\partial w/\partial \gamma_0 = v$ получаем однородное дифференциальное уравнение

$$\frac{\partial^2 v}{\partial t^2} + \alpha_1 \frac{\partial v}{\partial t} + \alpha_0 = 0 \quad (9.77)$$

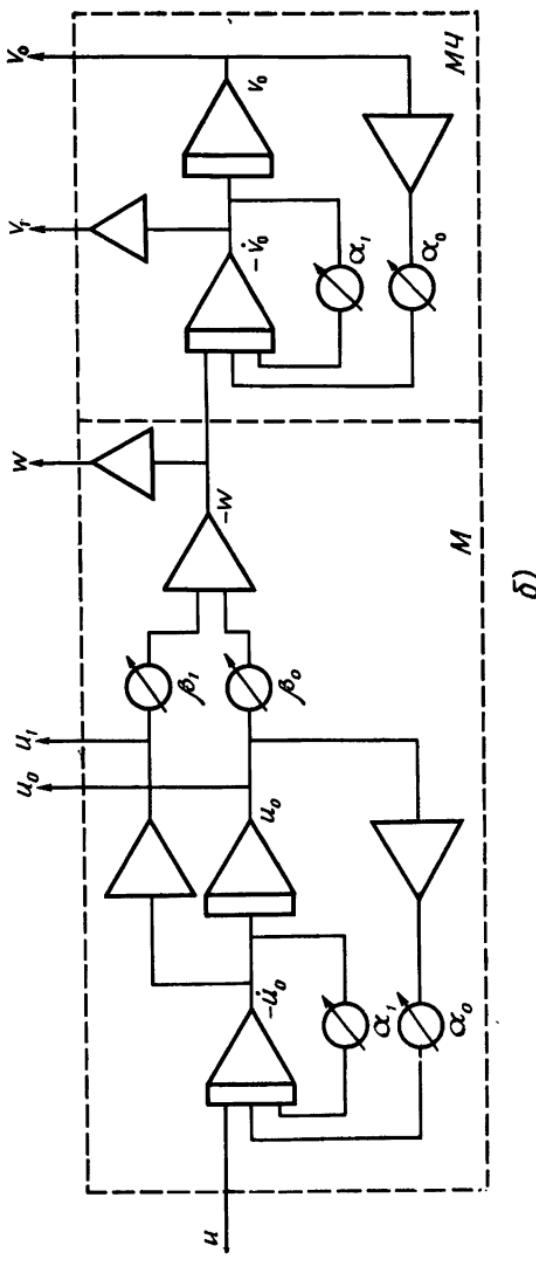
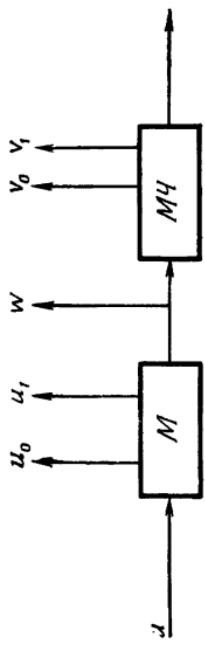
с начальными условиями

$$v(0) = 1, \quad \dot{v}(0) = 0.$$

Это уравнение решается просто путем повторного использования основной модели при тождественно равной нулю управляющей функции u и соответственно измененных начальных условиях.

Применения метода влияния параметров не ограничены линейными системами. В качестве примера нелинейной системы рассмотрим уравнение

$$\ddot{w} + \alpha_1 \dot{w} + \alpha_0 w + \mu w^3 = u. \quad (9.78)$$



Ф и г. 9.9.

Уравнения чувствительности имеют вид

$$\ddot{v}_1 + \alpha_1 \dot{v}_1 + \alpha_0 v_1 + 3\mu w^2 v_1 = -\dot{w}, \quad v_1 = \frac{\partial w}{\partial \alpha_1}, \quad (9.79a)$$

$$\ddot{v}_0 + \alpha_1 \dot{v}_0 + \alpha_0 v_0 + 3\mu w^2 v_0 = -w, \quad v_0 = \frac{\partial w}{\partial \alpha_0}, \quad (9.79b)$$

$$\ddot{v}_\mu + \alpha_1 \dot{v}_\mu + \alpha_0 v_\mu + 3\mu w^2 v_\mu = -w^3, \quad v_\mu = \frac{\partial w}{\partial \mu}. \quad (9.79v)$$

Опять уравнения различаются только «связывающими членами». Следовательно, можно последовательно использовать одну и ту же модель с управляющими функциями $-\dot{w}$, $-w$, $-w^3$. Рассмотренную задачу можно обобщить на систему n дифференциальных уравнений с m параметрами $\alpha_1, \alpha_2, \dots, \alpha_m$

$$\frac{dw_i}{dt} = f_i(w_1, w_2, \dots, w_n, t; \alpha_1, \alpha_2, \dots, \alpha_m), \quad i = 1, \dots, n. \quad (9.80)$$

Уравнения чувствительности относительно α_k , из которых определяются производные $u_{ik} = \partial w_i / \partial \alpha_k$, записываются в виде

$$\frac{\partial}{\partial \alpha_k} \left(\frac{dw_i}{dt} \right) = \frac{\partial u_{ik}}{\partial t} = \frac{\partial f_i}{\partial w_1} u_{1k} + \dots + \frac{\partial f_i}{\partial w_n} u_{nk} + \frac{\partial f_i}{\partial \alpha_k}. \quad (9.81)$$

Начальные условия нулевые, $u_{ik}(0) = 0$, если только начальные условия исходного дифференциального уравнения не рассматриваются как параметры. Приведенная формулировка справедлива как для линейных, так и для нелинейных систем. Для изучения влияния отдельного параметра α_k приходится моделировать (или программировать) всю систему уравнений чувствительности (9.81), даже если этот параметр явно входит лишь в одно уравнение исходной системы (9.80). Если, например, α_k входит только в член f_j , то в j -м уравнении чувствительности появляется «связывающий член» $\partial f_j / \partial \alpha_k$, тогда как $\partial f_i / \partial \alpha_k = 0$ при $i \neq j$. Тем не менее все остальные уравнения чувствительности содержат α_k в неявной форме в виде членов u_{1k}, u_{2k}, \dots и оказываются связанными с j -м уравнением.

Еще одна область применений обнаруживается при исследовании эффекта исключения производных более

высокого порядка из дифференциального уравнения. Допустим, что изучается уравнение

$$\ddot{w} + a_1 \dot{w} + a_0 w = u. \quad (9.82)$$

Нужно выяснить влияние члена третьего порядка

$$a_3 \ddot{\dot{w}} + \ddot{w} + a_1 \dot{w} + a_0 w = u. \quad (9.83)$$

Уравнения чувствительности относительно a_3 и a_1 имеют вид

$$a_3 \ddot{\dot{v}}_3 + \ddot{v}_3 + a_1 \dot{v}_3 + a_0 v_3 = -\ddot{w}, \quad v_3 = \frac{\partial w}{\partial a_3}, \quad (9.84a)$$

$$a_3 \ddot{\dot{v}}_1 + \ddot{v}_1 + a_1 \dot{v}_1 + a_0 v_1 = -\dot{w}, \quad v_1 = \frac{\partial w}{\partial a_1}. \quad (9.84b)$$

Следовательно, $v_3 = \dot{v}_1$, и из модели чувствительности можно получить значение коэффициента влияния этого параметра в окрестности $a_3 = 0$.

До сих пор в этом разделе рассматривались абсолютные функции чувствительности параметров, например $v_1 = \partial w / \partial a_1$. Иногда удается использовать *относительные функции чувствительности*, например $\partial w / \partial \ln a_1 = a_1 \partial w / \partial a_1$.

Метод с использованием точек чувствительности

В предыдущем разделе было установлено, что для одновременного определения нескольких функций чувствительности, помимо модели объекта, необходим еще ряд дополнительных моделей чувствительности. Это связано с усложнением аналоговой вычислительной схемы или с увеличением машинного времени, необходимого для решения подобных задач.

С другой стороны, в разд. 9.1 было показано, что при использовании обобщенной модели дополнительных моделей чувствительности не нужно — функции чувствительности могут быть измерены непосредственно. Это объясняется линейностью обобщенной модели относительно параметров.

Учитывая желательность максимально возможного упрощения схемы моделирования и сокращения машинного

времени, имеет смысл изучить типы моделей, позволяющих находить наибольшее число функций чувствительности (из числа подлежащих определению). Для этой цели используется так называемый *метод точек чувствительности* [12].

Основную его идею можно пояснить следующим образом. Рассмотрим линейный объект с передаточной функцией $H(s, a_1, \dots, a_m)$, зависящей от параметров a_i . Преобразование Лапласа от входного сигнала есть $U(s)$; тогда выходной сигнал определяется формулой

$$Y(s, a_1, \dots, a_m) = H(s, a_1, \dots, a_m) U(s). \quad (9.85)$$

Выход соответствующей модели имеет вид

$$W(s, a_1, \dots, a_m) = H(s, a_1, \dots, a_m) U(s). \quad (9.86)$$

Учитывая дифференцируемость \mathcal{L} -преобразования по параметрам, получим
(абсолютные) функции чувствительности параметра

$$\frac{\partial W(s, a_1, \dots, a_m)}{\partial a_i} = V_i(s), \quad (9.87)$$

относительные функции чувствительности параметра

$$\frac{\partial W(s, a_1, \dots, a_m)}{\partial \ln a_i} = \alpha_i V_i(s). \quad (9.88)$$

Следующий пример помогает проиллюстрировать эту идею (фиг. 9.10, а, б). Для модели справедливы соотношения

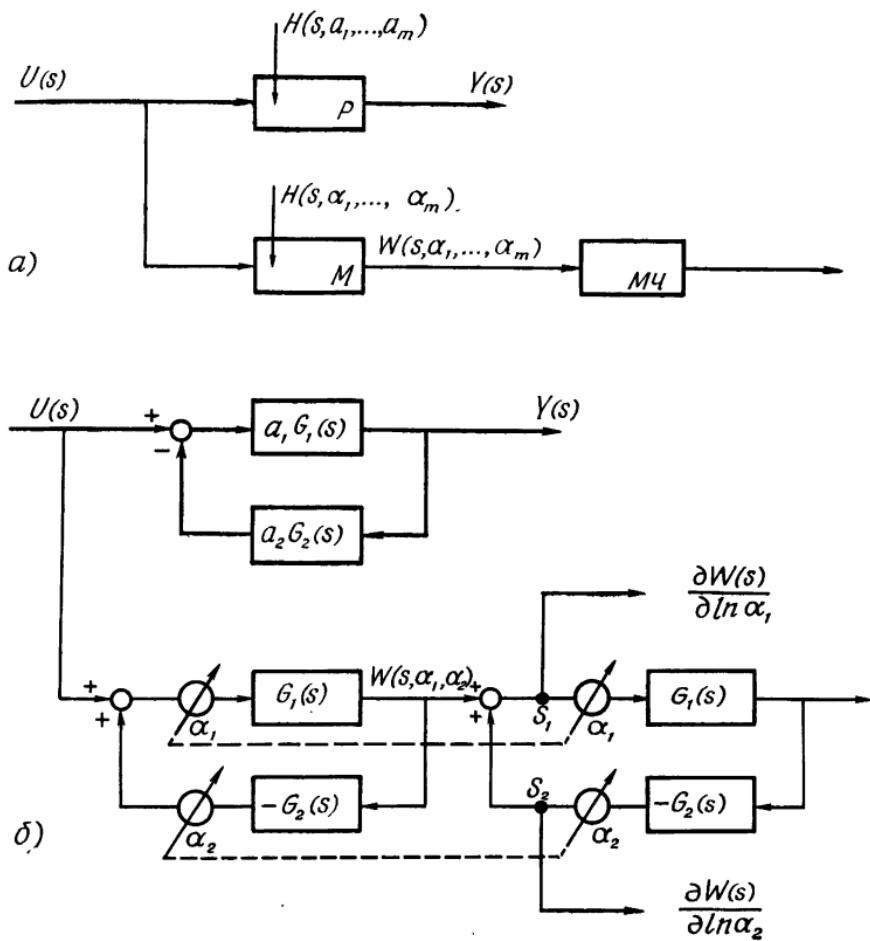
$$W(s) = H(s, \alpha_1, \alpha_2) U(s) = \frac{\alpha_1 G_1(s)}{1 + \alpha_1 G_1(s) \alpha_2 G_2(s)} U(s). \quad (9.89)$$

Отсюда для относительных функций чувствительности получаем

$$\begin{aligned} \frac{\partial W(s)}{\partial \ln \alpha_1} &= \alpha_1 \frac{\partial W(s)}{\partial \alpha_1} = \alpha_1 \frac{G_1(s)}{(1 + \alpha_1 G_1(s) \alpha_2 G_2(s))^2} U(s) = \\ &= \frac{W(s)}{1 + \alpha_1 G_1(s) \alpha_2 G_2(s)}, \end{aligned} \quad (9.90a)$$

$$\frac{\partial W(s)}{\partial \ln \alpha_2} = \alpha_2 \frac{\partial W(s)}{\partial \alpha_2} = \dots = \frac{\alpha_1 G_1(s) \alpha_2 G_2(s)}{1 + \alpha_1 G_1(s) \alpha_2 G_2(s)} W(s). \quad (9.90b)$$

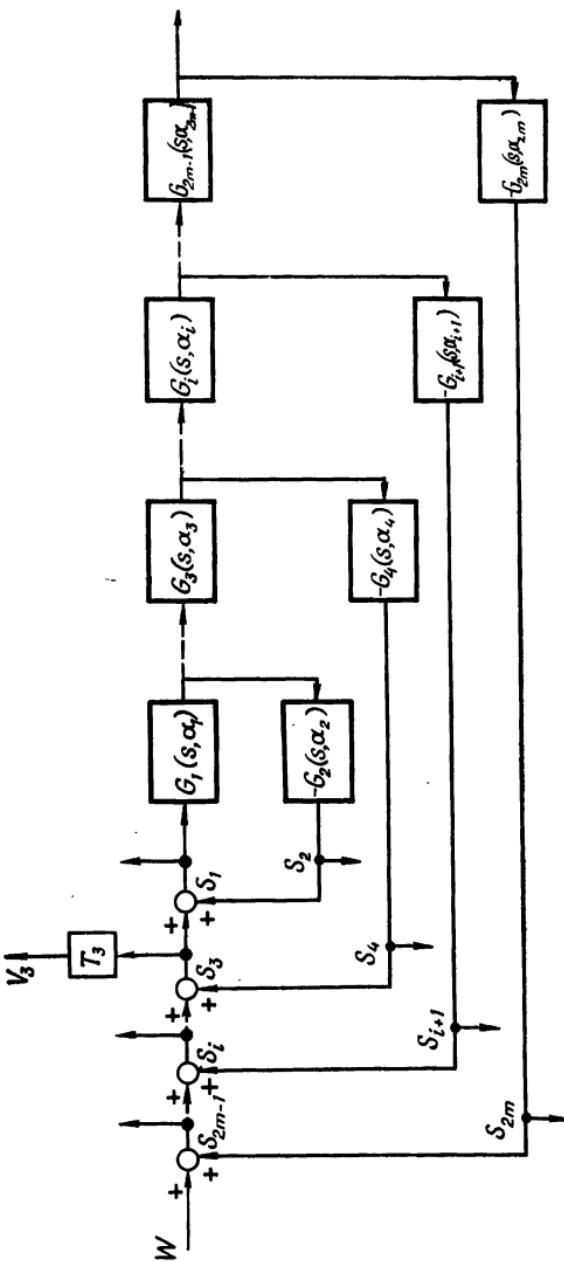
В результате приходим к схеме фиг. 9.10, б. S_1 и S_2 называются *точками чувствительности*. При аналоговом моде-



Ф и г. 9.10.

лировании обе функции чувствительности можно измерять одновременно, при цифровых вычислениях обе функции определяются по одной и той же программе.

Эту идею можно распространить на многоконтурные системы с обратной связью (фиг. 9.11). Здесь предполагается, что в каждом из $2m$ элементарных блоков имеется лишь один параметр, для которого нужно вычислить функцию чувствительности. Так же как и раньше, нетрудно показать, что S_j является точкой чувствительности для параметра из блока G_j . Остается рассмотреть вопрос



Ф и г. 9.41.

о том, каким образом параметр α_j входит в передаточную функцию $G_j(s, \alpha_j)$. Он решается введением дополнительной передаточной функции

$$T_j(s) = \frac{\partial \ln G_j(s, \alpha_j)}{\partial \ln \alpha_j}. \quad (9.91)$$

Это логарифмическая передаточная функция чувствительности, введенная ранее Боде [1]. Входом $T_j(s)$ служит сигнал, снимаемый с точки чувствительности S_j , выходом —

$$V_j(s) = \frac{\partial W(s, \dots, \alpha_j, \dots)}{\partial \ln \alpha_j}. \quad (9.92)$$

Некоторые частные случаи:

а) Если $G_j(s, \alpha_j) = \alpha_j F(s)$, то

$$T_j(s) = \frac{\alpha_j}{\alpha_j F(s)} \frac{\partial \alpha_j F(s)}{\partial \alpha_j} = 1. \quad (9.93)$$

В этом случае сигнал с S_j есть функция чувствительности и нет необходимости в добавлении каких-либо элементов в модель чувствительности (фиг. 9.9, б и 9.10, б).

б) Если $G_j(s, \alpha_j) = D(s, \alpha_j) F(s)$, т. е. передаточная функция является произведением двух передаточных функций, из которых лишь одна содержит представляющий для нас интерес параметр, то

$$T_j(s) = \frac{\alpha_j}{D(s, \alpha_j)} \frac{\partial D(s, \alpha_j)}{\partial \alpha_j}, \quad (9.94)$$

т. е. совпадает с передаточной функцией той части модели, которая содержит α_j .

Эти идеи можно также распространить на функции чувствительности высших порядков, например

$$V_{12}(s) = \frac{\partial^2 W(s, \alpha_1, \alpha_2)}{\partial \ln \alpha_1 \partial \ln \alpha_2} \text{ и т. д.,} \quad (9.95)$$

которые получаются очевидным образом из функций чувствительности первого порядка. Оказывается, что в этом случае необходима еще одна модель чувствительности.

Разумеется, анализ чувствительности использовался также и для описания объектов во временной области. Обзор соответствующей литературы можно найти в работе [13]. Много интересных статей содержат два сборника докладов симпозиумов ИФАК по чувствительности [7, 8].

Непрерывные настраиваемые модели

Рассматриваемая здесь схема приведена на фиг. 9.12. Ошибка определяется как

$$\begin{aligned} e &= y - w, \\ w &= G[u; \alpha], \end{aligned} \quad (9.96)$$

где G — некоторый функционал. Необходимо минимизировать критерий E , который можно записать как функционал L от четной функции $q(e)$:

$$E = L[q(e)]. \quad (9.97)$$

Настройка модели G осуществляется изменением параметров $\alpha' = (\alpha_1, \dots, \alpha_m)$ в соответствии со значением градиента E

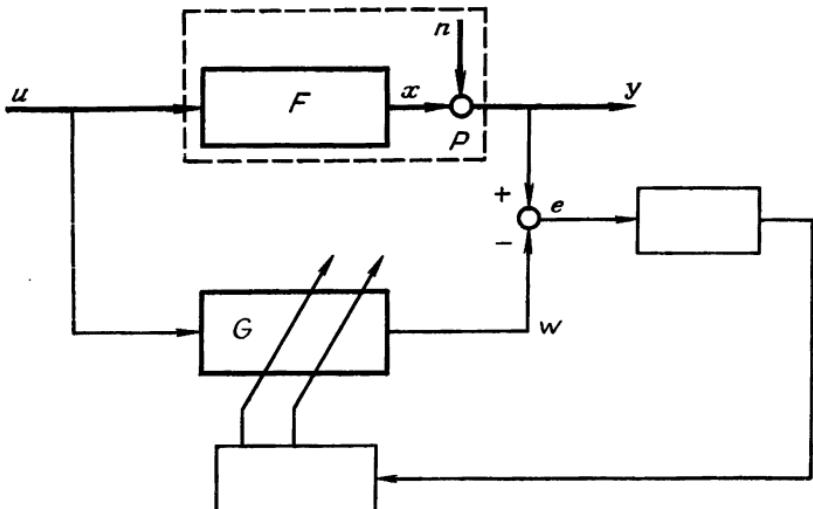
$$\dot{\alpha} = -\gamma \nabla E. \quad (9.98)$$

Компоненты вектора градиента определяются дифференцированием:

$$\frac{\partial E}{\partial \alpha_j} = L \left[\frac{\partial q(e)}{\partial e} \frac{\partial e}{\partial \alpha_j} \right], \quad (9.99)$$

$$\frac{\partial e}{\partial \alpha_j} = -\frac{\partial w}{\partial \alpha_j} = -\frac{\partial G[u; \alpha]}{\partial \alpha_j}, \quad (9.100)$$

причем $\partial w / \partial \alpha_j$ представляет собой коэффициент влияния параметра. Теперь можно определить следующий



Фиг. 9.12.

оператор:

$$H_J[u; \alpha] = \frac{\partial G[u; \alpha]}{\partial \alpha_j}, \quad (9.101)$$

откуда получаем

$$\nabla E = -L \left[\frac{\partial q(e)}{\partial e} H[u; \alpha] \right]. \quad (9.102)$$

Как указывалось в предыдущем разделе, множество операторов H , зависящих от параметра α и действующих на сигнал u , позволяет получить все функции чувствительности параметров.

Пример. Воспользуемся результатами работы [17]. Объект и модель описываются соответственно уравнениями

$$\frac{dy}{dt} + ay = u,$$

$$\frac{dw}{dt} + aw = u.$$

Уравнение чувствительности получается в результате дифференцирования уравнения модели:

$$\frac{dv}{dt} + av = -w,$$

где $v = dw/d\alpha$, а α считается постоянной. Применим в качестве критерия условие минимума

$$E = e^2 = (y - w)^2$$

и будем использовать для настройки метод наискорейшего спуска

$$\frac{d\alpha}{dt} = -\gamma \frac{\partial E}{\partial \alpha} = -2\gamma v(y - w),$$

поскольку от α зависит только w .

Поведение схемы настройки модели описывается формулами (9.98)–(9.102). Из-за ограничения, требующего постоянства α в (9.102), эти формулы позволяют лишь приближенно описать изменения α , когда эти изменения происходят достаточно медленно. В работе [17] исследованы вопросы сходимости для случаев, когда вход u является ступенчатым или синусоидальным сигналом. В первом случае можно доказать устойчивость точки равновесия

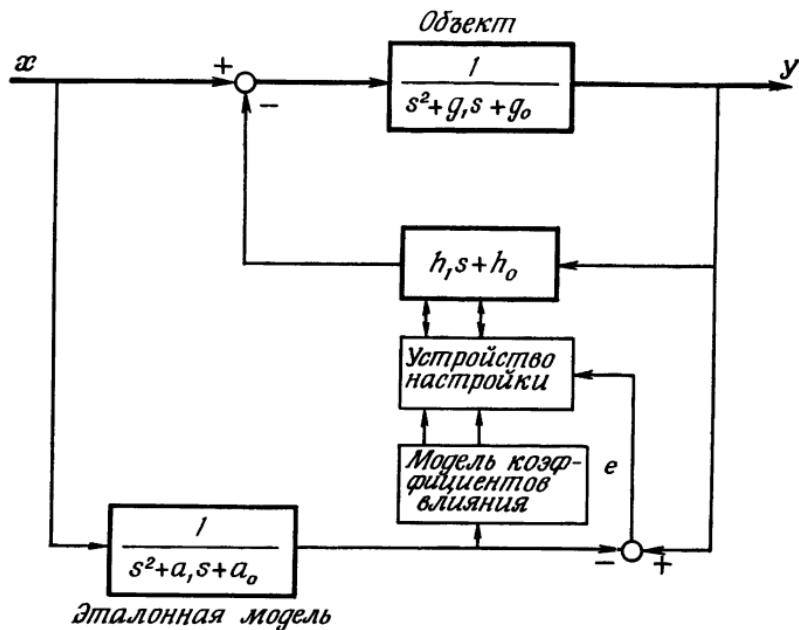
сия $a = a$. Второй случай приводит к уравнениям Матье, которые могут иметь как (асимптотически) устойчивые, так и периодические и неустойчивые решения.

При изучении устойчивости применялся второй метод Ляпунова: см. [9], а также работы, цитировавшиеся в предыдущем разделе.

Отметим, что функции чувствительности параметров играют роль *вспомогательных переменных* по аналогии с изложенным в гл. 6 и 7 для случая дискретных сигналов.

Примеры моделирования, практической реализации и применений

Хотя работа [5] и не имеет прямого отношения к оценке параметров, ее можно упомянуть как еще один пример использования коэффициентов влияния параметров. Исследуемая система изображена на фиг. 9.13. Параметры объекта g_1 и g_0 (например, изменение угловой скорости самолета по оси танглаша от отклонения управляемых поверхностей) изменяются. Эти изменения компенсируют-



Ф и г. 9.13.

ся настройкой параметров h_1 и h_0 в контуре обратной связи. Желаемые показатели системы «объект + цепь обратной связи» устанавливаются эталонной моделью, представляющей собой фиксированную аналоговую схему. Целью настройки является минимизация некоторого четного функционала от ошибки e . Это означает, что

$$\begin{aligned} g_1 + h_1 &= a_1 \rightarrow a_1, \\ g_0 + h_0 &= a_0 \rightarrow a_0. \end{aligned}$$

Такой результат получается генерированием коэффициентов влияния параметров de/da_1 и de/da_0 эталонной модели вместо соответствующих коэффициентов охваченного обратной связью объекта. Если a_1 и a_0 фиксированы, такой подход имеет то преимущество, что генерируемые коэффициенты влияния параметров представляют собой требуемые частные производные. (Это не верно для рассматривавшейся выше схемы настройки модели.)

Прерывистая настройка моделей

Как отмечалось в разд. 9.2, для непрерывных схем настройки трудно выявить свойства сходимости. Это объясняется прежде всего сложностью определения градиента при изменении (настройке) параметров модели. Рассмотрим теперь схемы, в которых параметры модели остаются постоянными при определении градиента. После интервала измерений производится настройка параметров модели, затем вновь начинается период измерений и т. д.

Если нужно настраивать два или большее число параметров, используются прерывистые схемы одного из двух типов:

а) все (или несколько) параметров настраиваются одновременно; б) в каждый момент времени может настраиваться лишь один параметр.

Пусть критерием является условие минимума

$$E = E(a_1, \dots, a_n),$$

где a_j — настраиваемые параметры. Для одновременной настройки нескольких параметров необходимо определить частные производные по этим параметрам. Выбор пути, ведущего к минимуму, обсуждался в гл. 5. Рас-

смотрим случай, когда одновременно настраивается не более одного параметра, следуя Бруннеру [2], описавшему итерационную процедуру настройки модели типа наискорейшего спуска. Критерием является условие минимума

$$E = \sum_{k=1}^K g_k e_k^2,$$

где g_k представляют веса различных составляющих ошибки. Дифференцируя по α_j , получим

$$\frac{\partial E}{\partial \alpha_j} = 2 \sum_{k=1}^K g_k e_k \frac{\partial e_k}{\partial \alpha_j}, \quad j = 1, 2, \dots, n. \quad (9.103)$$

Теперь рассмотрим итерационную процедуру настройки параметра α_j . В случае одношаговой процедуры осуществляется настройка лишь одного параметра. Ошибку можно записать в векторной форме. Пусть

$$\mathbf{e}(\alpha_j) = \{e_1(\alpha_j), \dots, e_K(\alpha_j)\}'$$

— вектор ошибки до подстройки α_j и

$$\mathbf{e}(\alpha_j + \Delta \alpha_j)$$

— тот же вектор после этой подстройки. Для получения траектории наискорейшего спуска нужно стараться сделать значения вектора ошибки $\mathbf{e}(\alpha_j + \Delta \alpha_j)$ ортогональными в метрике g_k к производной вектора $\mathbf{e}(\alpha_j)$ по параметру α_j . Это условие ортогональности можно записать в форме скалярного произведения:

$$\sum_{k=1}^K g_k \frac{\partial e_k}{\partial \alpha_j} e_k(\alpha_j + \Delta \alpha_j) = 0. \quad (9.104)$$

Это равенство представляет не что иное, как требование равенства нулю производной $\partial E / \partial \alpha_j$. Если теперь новые значения компонент вектора ошибки разложить в ряд Тейлора, сохраняя лишь члены первого порядка, то уравнение (9.104) примет вид

$$\sum_{k=1}^K g_k \frac{\partial e_k}{\partial \alpha_j} \left\{ e_k(\alpha_j) + \frac{\partial e_k}{\partial \alpha_j} \Delta \alpha_j \right\} = 0. \quad (9.105)$$

Разрешая это уравнение относительно $\Delta\alpha_j$, получим

$$\Delta\alpha_j = - \frac{\sum_{k=1}^K g_k e_k \frac{\partial e_k}{\partial \alpha_j}}{\sum_{k=1}^K g_k \left(\frac{\partial e_k}{\partial \alpha_j} \right)^2}. \quad (9.106)$$

В случае, когда e_k линейно зависят от параметров α_j , уравнения (9.104) и (9.105) полностью эквивалентны. Это как раз тот случай, когда вводится обобщенная модель (см. гл. 2 и 7).

Теперь рассмотрим одновременную настройку всех параметров после каждого периода измерений. Пусть объект и модель задаются уравнениями

$$\ddot{y} + a_1 \dot{y} + a_0 y = u, \quad (9.107)$$

$$\ddot{w} + a_1 \dot{w} + a_0 w = u. \quad (9.108)$$

Критерий имеет вид

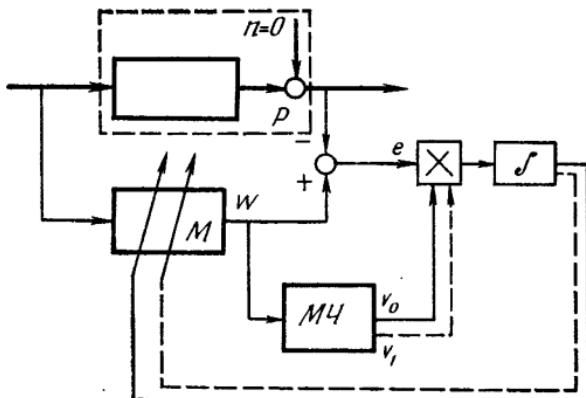
$$E(\alpha_1, \alpha_0) = \int_0^T e^2 dt = \int_0^T \{w(t; \alpha_1, \alpha_0) - y(t)\}^2 dt. \quad (9.109)$$

Для минимизации этой функции воспользуемся градиентным методом, при котором значения α_0 и α_1 изменяются пропорционально соответствующим компонентам вектора градиента ∇E . Этот метод гарантирует наискорейший спуск к минимальному значению E по траектории, ортогональной к линиям уровня $E(\alpha_0, \alpha_1) = \text{const}$. Для вычисления компонент градиента

$$\frac{\partial E}{\partial \alpha_0} = 2 \int_0^T e \frac{\partial w}{\partial \alpha_0} dt, \quad (9.110)$$

$$\frac{\partial E}{\partial \alpha_1} = 2 \int_0^T e \frac{\partial w}{\partial \alpha_1} dt \quad (9.111)$$

необходимы значения коэффициентов влияния параметров $v_0 = \partial w / \partial \alpha_0$ и $v_1 = \partial w / \partial \alpha_1$. Эти коэффициенты можно



Ф и г. 9.14.

получить как непрерывные функции времени одновременно с решением $w(t)$ (фиг. 9.14). Вычислительное устройство одновременно оценивает интегралы (9.110) и (9.111). Процедура вычислений сводится к многошаговому процессу систематических изменений значений параметров a_0 и a_1 в соответствии с направлением градиента, определяемым в конце каждого цикла вычислений. Эта последовательность сходится к желаемому минимуму E . На фиг. 9.15 показаны типичные линии уровня и направления градиента, полученные при решении задачи оценивания.

Несколько более изящным является метод с использованием квадратичного предсказания оптимума. Предположим, что ошибки вычислений отсутствуют. Тогда ошибка определяется лишь невязками

$$\Delta a_1 = a_1 - \alpha_1 = \lambda_1, \quad (9.112a)$$

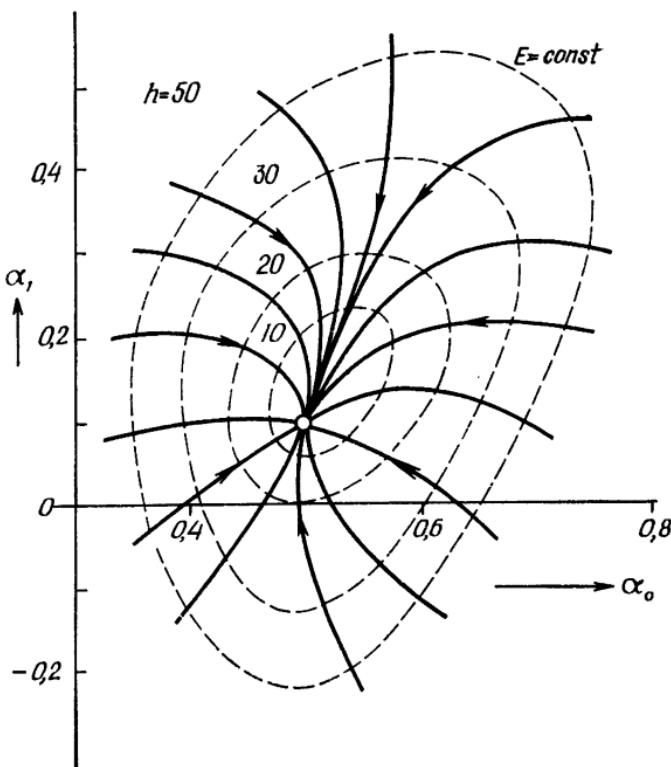
$$\Delta a_0 = a_0 - \alpha_0 = \lambda_0. \quad (9.112b)$$

Разложение ошибки в ряд в окрестности (a_0, a_1) дает

$$e(a_0, a_1) \approx e(\alpha_0, \alpha_1) + \frac{\partial e}{\partial a_0} \lambda_0 + \frac{\partial e}{\partial a_1} \lambda_1 + \dots. \quad (9.113)$$

Снова нужно из уравнений чувствительности для

$$v_0 = \partial e / \partial a_0 \text{ и } v_1 = \partial e / \partial a_1$$



Ф и г. 9.15.

определить коэффициенты влияния параметров. Так как $e(a_0, a_1) = 0$, то

$$e(a_0, a_1) + v_0\lambda_0 + v_1\lambda_1 \approx 0. \quad (9.114)$$

Поскольку предположение об отсутствии шумов и ошибок практически не может быть выполнено, следует ввести величину остаточной ошибки

$$e(a_0, a_1) + v_0\lambda_0 + v_1\lambda_1 = e_{\text{res}}. \quad (9.115)$$

Теперь можно записать критерий в виде

$$\min_{\alpha_0, \alpha_1} \int e_{\text{res}}^2 dt. \quad (9.116)$$

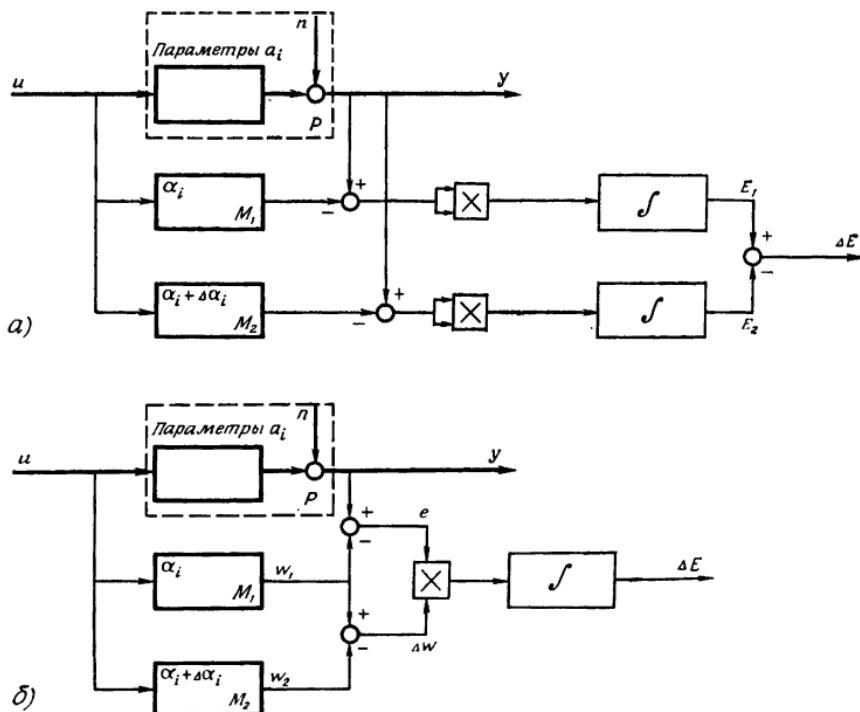
Уравнения наискорейшего спуска в данном случае имеют вид

$$\Delta\alpha_0 = -2\gamma \int_0^T v_0 e_{\text{res}} dt, \quad (9.117\text{a})$$

$$\Delta\alpha_1 = -2\gamma \int_0^T v_1 e_{\text{res}} dt. \quad (9.117\text{б})$$

9.5. ОДНОВРЕМЕННОЕ ИСПОЛЬЗОВАНИЕ ДВУХ МОДЕЛЕЙ ИЛИ ПОВТОРНОЕ ИСПОЛЬЗОВАНИЕ ОДНОЙ МОДЕЛИ

Применяя две модели с различными значениями параметров, можно получить информацию о направлении оптимальной подстройки. Таким способом удается найти (грубую) приближенную оценку $\Delta E/\Delta\alpha$ производ-



Ф и г. 9.16.

ной $\partial E/\partial a$. Можно использовать два несколько отличных метода. Связь между этими методами характеризуется следующим уравнением, получающимся при дифференцировании критерия ошибки по настраиваемому параметру:

$$\frac{\partial E}{\partial \alpha_i} = \frac{2}{T} \int_0^T e(\alpha_1, \alpha_2, \dots, t) \frac{\partial e}{\partial \alpha_i} dt. \quad (9.118)$$

При использовании одного метода получается величина $\Delta E/\Delta \alpha_i$ как оценка $\partial E/\partial \alpha_i$. Во втором методе $\Delta w/\Delta \alpha_i$ используется как приближенное значение величины

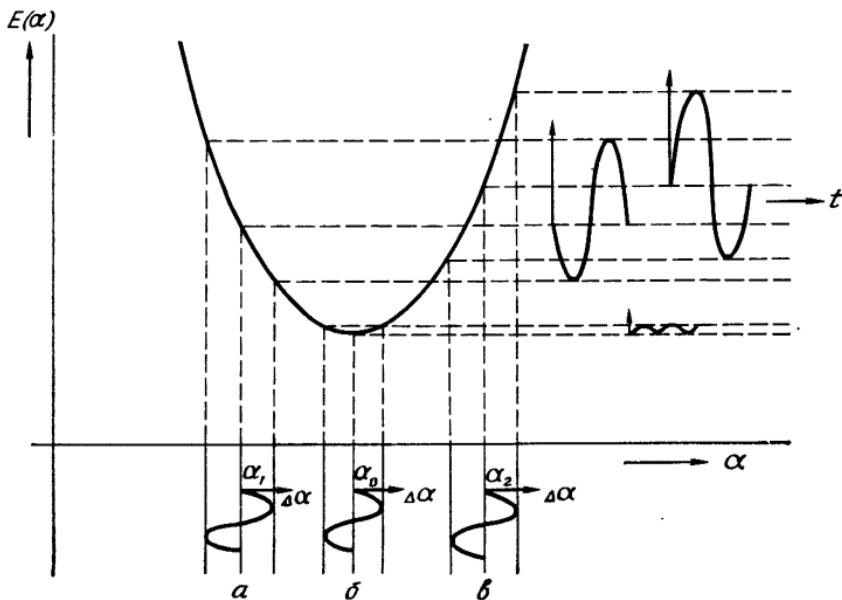
$$\frac{\partial w}{\partial \alpha_i} = -\frac{\partial e}{\partial \alpha_i} \text{ при } e = y - w(a).$$

Различие этих подходов иллюстрируется фиг. 9.16, а, б. Если входной сигнал *и* стационарен, то вместо двух моделей на фиг. 9.16, а можно ограничиться одной; при этом найденное значение критерия E_1 запоминается, затем значение параметра α_i изменяется, определяется E_2 , полученное значение E_2 сравнивается с E_1 и т. д. Этот подход описан в работах Балхена и других, приведенных в списке дополнительной литературы.

9.6. МОДЕЛЬ С ВОЗМУЩЕНИЕМ ПАРАМЕТРОВ

Метод возмущения параметров. Этот метод определения частных производных основывается на модуляции параметров модели и детектировании соответствующих изменений критерия.

На фиг. 9.17 приведены три различных состояния модели. В случае *a* значение параметра α_1 не оптимально. При изменении этого параметра относительно фиксированного значения в соответствии с синусоидальным модулирующим сигналом квадратичный критерий (если используется мгновенное значение ошибки) также изменяется. При отсутствии шумов эта реакция критерия также будет синусоидальной с определенным сдвигом по фазе. Умножая эту реакцию на синусоидальный сигнал с той же частотой и фазой, что и модулирующий сигнал параметра, и усредняя, получаем оценку $\partial E/\partial \alpha_1$. То же самое справедливо и для случая *b* с той лишь разницей, что здесь



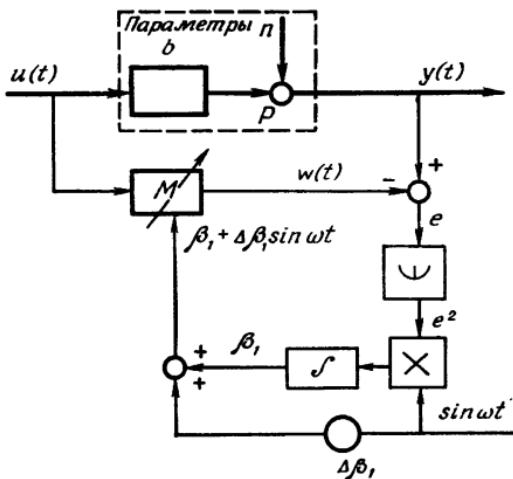
Ф и г. 9.17.

реакция критерия противоположна по фазе. В случае *б*, когда значение α_0 предполагается оптимальным, реакция квадратичного критерия содержит только высшие гармоники модулирующего сигнала. В результате умножения на сигнал с частотой модуляции и усреднения должен получиться нуль.

Возможная система настройки для одного параметра приведена на фиг. 9.18. Входные сигналы объекта и модели одинаковы. Ошибка e получается как разность выходного сигнала объекта, включая аддитивный шум, и выходного сигнала модели. Устройство возведения в квадрат образует величину e^2 , которая умножается на сигнал, возмущающий параметр, и подается на интегратор. На выходе интегратора получается значение параметра модели.

Этот процесс можно описать следующими уравнениями. Пренебрегая шумом, уравнение объекта можно записать в виде

$$a_2 \ddot{y} + a_1 \dot{y} + y - b_1 \dot{u} - b_0 u = 0. \quad (9.119)$$



Ф и г. 9.18.

Уравнение модели в этом случае должно иметь форму

$$a_2 \ddot{w} + a_1 \dot{w} + w - \beta'_1 \dot{u} - \beta_0 u = 0. \quad (9.120)$$

Здесь β'_1 — модулированный параметр

$$\beta'_1 = \beta_1 + \Delta \beta_1 \sin \omega t. \quad (9.121)$$

Вычитая уравнение модели из уравнения объекта, получим уравнение для ошибки

$$\begin{aligned} e &= y - w, \\ e &= -(a_2 \ddot{y} - a_2 \ddot{w}) - (a_1 \dot{y} - a_1 \dot{w}) + \\ &+ (b_1 - \beta_1 - \Delta \beta_1 \sin \omega t) \dot{u} + (b_0 - \beta_0) u. \end{aligned} \quad (9.122)$$

Возводя обе части этого уравнения в квадрат, умножая на $\sin \omega t$ и усредняя, получаем

$$\begin{aligned} \overline{e^2 \sin \omega t} &= -\Delta \beta_1 (b_1 - \beta_1) \overline{\dot{u}^2} + \Delta \beta_1 \overline{(a_2 \ddot{y} - a_2 \ddot{w}) \dot{u}} + \\ &+ \Delta \beta_1 \overline{(a_1 \dot{y} - a_1 \dot{w}) \dot{u}}, \end{aligned} \quad (9.123)$$

поскольку $\overline{uu} = 0$.

Таким образом, найден приближенный градиент

$$\frac{\partial E}{\partial \beta_1} \approx \overline{\frac{e^2 \sin \omega t}{\Delta \beta_1}} = - (b_1 - \beta_1) \overline{\dot{u}^2} + \\ + \overline{(a_2 \ddot{y} - \alpha_2 \dot{w}) \dot{u}} + \overline{(a_1 \dot{y} - \alpha_1 \dot{w}) \dot{u}}. \quad (9.124)$$

Первый член в правой части этого уравнения определяет знак градиента. Другие члены справа характеризуют взаимодействие между α_1 и α_2 . При $\alpha_2 = a_2$ и $\alpha_1 = a_1$ эти два члена становятся очень малыми и стремятся к нулю при стремлении (β_1, β_0) к (b_1, b_0) . Следовательно, оптимум достигается при

$$(\alpha_2, \alpha_1, \beta_1, \beta_0) = (a_2, a_1, b_1, b_0).$$

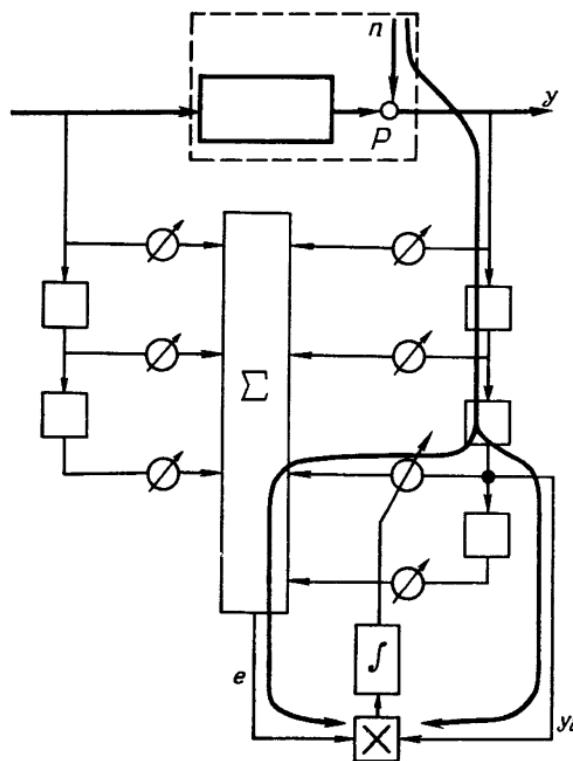
Принципиальным в инженерном отношении вопросом реализации этого метода является выбор длины интервала усреднения. При слишком длинном интервале система получается излишне инерционной, при слишком коротком — недостаточно помехоустойчивой.

Приведенное уравнение не показывает, как распространить этот метод на многопараметрические системы. Учитывая, что при умножении периодических сигналов легко осуществляется селекция частот, нетрудно понять, что при использовании различных модулирующих частот можно одновременно настраивать соответствующее число параметров.

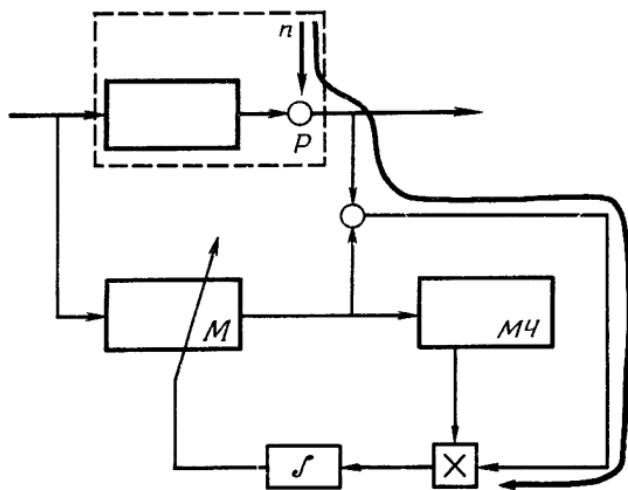
Применяя метод наискорейшего спуска, приходим к следующим уравнениям:

$$\begin{aligned} \dot{\beta}_1 &= -\gamma \frac{\partial E}{\partial \beta_1}, & \dot{\alpha}_2 &= -\gamma \frac{\partial E}{\partial \alpha_2}, \\ \dot{\beta}_0 &= -\gamma \frac{\partial E}{\partial \beta_0}, & \dot{\alpha}_1 &= -\gamma \frac{\partial E}{\partial \alpha_1}. \end{aligned} \quad (9.125)$$

Таким образом, получается четыре взаимодействующих контура настройки. В зависимости от коэффициента усиления контура γ имеет место низкая устойчивость или неустойчивость. Определение условий устойчивости для задач такого типа сводится к известному уравнению Маттье.



Ф и г. 9.19.



Ф и г. 9.20.

9.7. ЗАКЛЮЧЕНИЕ

Суть изложенного в этой главе можно кратко сформулировать следующим образом. Для методов, использующих настраиваемые модели, характерна необходимость определения частных производных по настраиваемым параметрам. Был указан ряд методов; очевидна необходимость в дополнительном оборудовании (модель чувствительности) или в добавочных наблюдениях (модель со значениями параметров β_j и $\beta_j + \Delta\beta_j$). В этом отношении весьма привлекательна обобщенная модель, для которой благодаря линейности по параметрам функции чувствительности определяются непосредственно без каких-либо вспомогательных приспособлений. Возможность такого изящного и экономного подхода достигается ценой смещения оценок, связанного с наличием любого аддитивного шума. При использовании моделей чувствительности подобная проблема не возникает. Это иллюстрируется фиг. 9.19 и 9.20, на которых показано влияние аддитивного шума соответственно при использовании обобщенной модели и модели чувствительности. Проблема смещения возникает из-за имеющего место в обобщенной модели перемножения составляющих, связанных с шумом, которые входят как в сигнал ошибки e , так и в функции чувствительности y_i .

ЛИТЕРАТУРА

1. Bode H. W., Network Analysis and Feedback Amplifier Design; есть русский перевод: Боде Г., Теория цепей и проектирование усилителей с обратной связью, ИЛ, 1948.
2. Brunner W., An Iteration Procedure for Parametric Model Building and Boundary Value Problems, *Proc. WJCC*, 517—533 (1961).
3. Clymer A. B., System Synthesis by Dynamic Implicit Computation, Meeting of the Midwestern and Central Simulation Councils, St. Louis (Mo.), 1958.
4. Clymer A. B., Direct System Synthesis by Means of Computers, *Trans. AIEE*, 77, Part I (Communication and Electronics) 798—806 (1959).
5. Donalsen D. D., Leondes C. T., A Model Referenced Parameter Tracking Technique for Adaptive Control Systems, Part I — The Principle of Adaption, Part II — Stability Analysis by the Second Method of Lyapunov, *IEEE Trans. Applic. and Ind.*, 82, № 68, 241—262 (1963).
6. Graupe K. K., The Analog Solution of Some Functional Ana-

- lysis Problems, *Trans. AIEE*, 79, part I (Communication and Electronics), 793—799 (1961).
7. IFAC Sensitivity, Sensitivity Methods in Control Theory, Proc. of an IFAC Symposium, Dubrovnik, Yugoslavia, 1964.
 8. IFAC Sensitivity, System Sensitivity and Adaptivity, Preprints of the Second IFAC Symposium, Dubrovnik, Yugoslavia, 1968.
 9. Johnson C. W., Rang E. R., The Learning Model Approach to Self-Evaluation and the Method of Lyapunov, AIEE Paper DP 61-603 (Great Lakes District Meeting, Minneapolis), 1961.
 10. Kaya Y., Yamamura S., A Self-Adaptive System with a Variable-Parameter PID Controller, *AIEE Trans.* 80, Part II (Appl. and Ind., № 58), 378—386 (1962).
 11. Kohr R. H., A Method for the Determination of a Differential Equation Model for Simple Nonlinear Systems, *IEEE Trans. Electron. Computers*, EC-12, 394—400 (1963).
 12. Кокотович П., Метод точек чувствительности в исследовании и оптимизации линейных систем управления, *Автоматика и телемеханика*, № 12 (1964).
 13. Кокотович П., Рутман Р. С., Чувствительность систем автоматического управления (обзор), *Автоматика и телемеханика*, № 4 (1965).
 14. Kushner H. J., On the Convergence of Lion's Identification Method with Random Inputs (Short Paper), *IEEE Trans. Autom. Control*, AC-15, 652—654 (1969).
 15. Landau I. D., An Approach to Synthesis of Model Reference Adaptive Control Systems, *IEEE Trans. Autom. Control* (1970).
 16. Lion P. M., Rapid Identification of Linear and Non-Linear Systems, Joint Autom. Control Conf., 1966.
 17. Margolis M., Leondes C. T., A Parameter Tracking Servo for Adaptive Control System, *IRE Trans. Autom. Control*, AC-4, 100—111 and *IRE Wescon Conv. Rec.*, Part 4, 104—115 (1959).
 18. Miller B. J., A General Method of Computing System Parameters with an Application to Adaptive Control, Conference Paper AIEE, CP 62—96, 1962.
 19. Parks P. C., Liapunov Redesign of Model Reference Adaptive Control Systems, *IEEE Trans. Autom. Control*, AC-11, 362—367 (1966).
 20. Pazdera J. S., Pottinger H. J., Linear System Identification via Liapunov Design Techniques, Joint Autom. Control Conf., 1969, 795—801.
 21. Попов В. М., Абсолютная устойчивость нелинейных систем автоматического управления, *Автоматика и телемеханика*, № 8 (1961).
 22. Potts T. F., Ornstein G. N., Clymer A. B., The Automatic Determination of Human and other System Parameters, *Proc. of the Western Joint Computer Conf.*, Los Angeles, 19, 645—660 (1961).
 23. Schoenemann P. T., An Identification Machine for Dynamic Biological Data, Ph. D. Thesis, E. E. Dept., Univ. of California, Berkeley, Calif., 1965.
 24. Shackcloth B., Butchart R. L., Synthesis of Model Reference Adaptive Systems by Liapunov's Second Method, IFAC Symp.

- Theory of Selfadaptive Control Systems, Teddington, 1965, pp. 145—152.
25. Young P. C. (см. литературу к разд. 14.1).
26. Zames G., On the Input-Output Stability of Time-Varying Non-Linear Feedback Systems, *IEEE Trans. Autom. Control*, AC-11, 228—238 (1966).

ДОПОЛНИТЕЛЬНАЯ ЛИТЕРАТУРА

9.2. Минимизация мгновенной ошибки

Butler R. E., Bohn E. V., An Automatic Identification Technique for a Class of Nonlinear Systems, *IEEE Trans. Autom. Control*, AC-11, 292—296 (1966).

Goodman T. P., Reswick J. B., Determination of System Characteristics from Normal Operating Records, *Trans. ASME*, 78, 259—271 (1956).

Кованиц П., Обобщенный дискретный аналог задачи Заде-Рагаззини, *Автоматика и телемеханика*, № 2 (1966).

Kovanic P., Identification of Operators via Static Programming, IFAC Symp. «Identification in Autom. Control Systems», Prague, Paper 1.3, 1967.

Narendra K. S., McBride L. E., Multiparameter Self-Optimizing Systems Using Correlation Techniques, *IEEE Trans. Autom. Control*, AC-9, 31—38 (1964).

Roberts J. D., A Method of Optimizing Adjustable Parameters in a Control System, *Proc. IEE*, 109B, 519—528 (1962).

Young P. C., The Determination of the Parameters of a Dynamic Process, *Radio and Electronic Engineer*, 29, 345—361 (1965).

9.3. Минимизация средней по времени ошибки

Balakrishnan A. V., A General Theory of Nonlinear Estimation Problems in Control Systems, Dept. Eng., Univ. of Calif., Los Angeles, Calif., 1961.

Hoberock L. L., Kohr R. H., An Experimental Determination of Differential Equations to Describe Simple Nonlinear Systems, *J. Basic Engr. (Trans. ASME)*, 393—398 (1967).

Lekoff K., On the Problem of Equivalence of Model and Object with Regard the Control of the Object, IFAC Symp. «Identification in Autom. Control Systems», Prague, Paper 5.2, 1967.

Maer S., On the Uniqueness of the Adjustment in Defining the Dynamic Characteristics with the Help of Selfadjusting Model, IFAC Symp. «Identification in Autom. Control Systems», Prague, Paper 5.6, 1967.

Maslov E. P., An Optimal Statistical Adaptive Model, IFAC Symp. «Identification in Autom. Control Systems», Prague, Paper 3.4, 1967.

Richalet J., Lecamus F., Hummel P., New Trends in Identification: Structural Distance and Weak Topology, IFAC Symp. «Identification and Process Parameter Estimation», Prague, Paper 1.2, 1970.

Robinson W. R., Soudack A. C., A Method for the Identifica-

tion of Timedelays in Linear Systems (Short Paper), *IEEE Trans. Autom. Control*, AC-15, 97—101 (1970).

Rosen J. B., The Gradient Projection Method for Nonlinear Programming, Part I, Linear Constraints, *SIAM J.*, 8, 181—217 (1960).

Rosen J. B., The Gradient Projection Method for Nonlinear Programming, Part II, Nonlinear Constraints, *SIAM J.*, 9, 514—532 (1961).

Roy R., Sherman J. (см. литературу к гл. 4).

Skala K., Experimental Identification of Systems by Adapting of the Parallel Model, IFAC Symp. «Identification in Autom. Control Systems», Prague, Paper 5.3, 1967.

9.4. Функции чувствительности параметров

Ariaratnam S. T., Graefe P. W. U., Linear Systems with Stochastic Coefficients II, *Int. J. Control*, 2, 161—169 (1965).

Åström K. J., On a First Order Stochastic Differential Equation, *Int. J. Control*, 1, 301—326 (1965).

Bekey G. A., McGhee R. B. (см. литературу к гл. 5).

Brainin S. M., The Identification of Random Parameters, IFAC Symp. «Identification in Autom. Control Systems», Prague, Paper 3.13, 1967.

Bogdanoff J. L., Kozin F., Moments of the Output of Linear Random Systems, *J. Acoust. Soc. Am.*, 34, 1063—1066 (1962).

Butchart R. L., Shackcloth B., Synthesis of Model Reference Adaptive Systems by Liapunov's Second Method, IFAC Symp. «Theory of Self-Adaptive Control Systems», Teddington, Paper 3.4, 1965.

Chadeev V. M., Adaptive Models of Plants, IFAC Symp. «Identification in Autom. Control Systems», Prague, Paper 5.8, 1967.

Dymock A. J., Helps K. A., Meredith J. F., The Identification of Three Parameters Using a Linear Plant Model, IFAC Symp. «Identification in Autom. Control Systems», Prague, Paper 5.7, 1967.

Fishwick W., Davies W. D. T., The Synthesis of Dynamical Models of Plants and Processes, *Proc. Instn. Mech. Engrs.*, 179, Part 3H, 29—36 (1965).

Gabor D., Wilby W. P. L., Woodcock R., A Universal Non-linear Filter, Predictor and Simulator which Optimizes Itself by a Learning Process, *Proc. IEE*, 108B, 422—438 (1961).

Graupe D., Cassir G. R., A Predictive Identification and Mode-Recognition Technique for Adaptive Control and Tracking Systems, IFAC Symp. «Identification in Autom. Control Systems», Prague, Paper 4.1, 1967.

Han K. W., Thaler G. J., Multiparameter Self-Adaptation Using Auxiliary Models, *IEEE Trans. Applic. and Ind.*, 82, 197—202 (1963).

Hughes M. T. G., May R., Estimation of Frequency Response from Random Data, IFAC Symp. «Identification in Autom. Control Systems», Prague, Paper 6.4, 1967.

Казаков И. Е., Евланов Л. Г., К теории самонастраивающихся систем с поиском градиента методом вспомогательного оператора, в книге «Дискретные и самонастраивающиеся системы», Труды II Межд. конгресса ИФАК, изд-во «Наука», 1965.

Kokotović P., et al., Sensitivity Method in the Experimental Design of Adaptive Control Systems, Third IFAC Congress, London, Paper 45B, 1966.

Kulick V. T., Sbitnev A. I., Interpolating and Extrapolating Methods of the Statistical Analysis of Dynamic Systems, IFAC Symp. «Identification in Autom. Control Systems», Prague, Paper 3.18, 1967.

Kushner H. J., A Simple Iterative Procedure for the Identification of the Unknown Parameters of a Linear Time Varying Discrete System, *Journal of Basic Engineering*, 85D, 227—235 (1963).

Lubbock J. K., Barker H. A., A Solution of the Identification Problem, *IEEE Trans. Applic. and Ind.*, 83, 166—173 (1964).

Margolis M., Leondes C. T., On the Philosophy of Adaptive Control for Process Adaptive Systems, Proc. Nat. Electron. Conf., 15, 27—34 (1959).

Margolis M., Leondes C. T., On the Theory of Adaptive Control Systems; the Learning Model Approach, First IFAC Congress, Moscow, 2, 1960, pp. 556—563; есть русский перевод: Марголис М., Леондес К., О теории самонастраивающихся систем регулирования; метод обучающейся модели, в книге «Теория дискретных оптимальных и самонастраивающихся систем», Труды I Межд. конгресса ИФАК, Изд-во АН СССР, 1961.

Marsik J., Experiments with a Self-Adjusting Model for Automatic Parameter Estimation of Control Loops (на немецком), *Zeitschr. Messen, Steuern, Regeln*, 9, 210—213 (1966).

Marsik J., Quick-Response Adaptive Identification, IFAC Symp. «Identification in Autom. Control Systems», Prague, Paper 5.5, 1967.

Meissinger H. F. (см. литературу к гл. 2).

Meissinger H. F., Parameter Influence Coefficients and Weighting Functions Applied to Perturbation Analysis of Dynamic Systems, Proc. Third Internat. Congress on Analog Computation, Opatija, Yugoslavia, 1961.

Meissinger H. F., Parameter Optimization by an Automatic Open-Loop Computing Method, Proc. Fourth Internat. Congress on Analog Computation, Brighton, England, 1964, pp. 268—274.

Meissinger H. F., Bekey G. A., Computation of Sensitivity Functions and Their Use in Parameter Optimization Problems, IFAC Conference on Sensitivity Analysis, Dubrovnik, 1964.

Meissinger H. F., Bekey G. A., An Analysis of Continuous Parameter Estimation Methods, *Simulation*, 6, 94—102 (1966).

Meredith J. F., Dymock A. J., A Self-Adaptive System Employing High Speed Parameter Identification, IFAC Symp. «Theory of Self-Adaptive Control Systems», Teddington, Paper 4.2, 1965.

Mesch F., A Self-Adjusting Controller with Self-Adjusting Reference Model (на немецком), *Zeitschr. Messen, Steuern, Regeln*, 5, 320—322 (1962).

Miller K. S., Murray F. J., A Mathematical Basis for an Error Analysis of Differential Analysers, *J. Math. Physics*, 32, № 2, 3 (1958).

Norkin K. B., Process Identification with the Use of Search Controlled Adaptive Model, IFAC Symp. «Identification in Autom. Control Systems», Prague, Paper 5.1, 1967.

Parks P. C., Stability Problems of Model-Reference and Iden-

tification Systems, IFAC Symp. «Identification in Autom. Control Systems», Prague, Paper 5.4, 1967.

Plander I., Automatic Iteration Method of Process Identification with Cyclic Adjustment of Parameters Using Analog and Hybrid Computers, IFAC Symp. «Identification in Autom. Control Systems», Prague, Paper 5.9, 1967.

Puri N. N., Weygandt C. N., Transfer Function Tracking of a Linear Time-Varying System by Means of Orthogonal Networks, *IEEE Trans. Applic. and Ind.*, 82, № 67, 234—238 (1963).

Puri N. N., Weygandt C. N., Transfer Function Tracking of a Linear Time-Varying Systems by Means of Auxiliary Simple Lag Network, Joint Autom. Control Conf., 1963, pp. 200—203.

Rubin A. I., Driban S., Meissner W. M., Regression Analysis and Parameter Identification, *Simulation*, 7, 39—47 (1967).

Sakrison D. J., Application of Stochastic Approximation Methods to System Optimization, Techn. Rept. 391, Res. Lab. Electron. MIT, Cambridge (Mass.), 1962.

Shapiro I. I., The Prediction of Ballistic Missile Trajectories from Radar Observations, McGraw-Hill, N.Y., 1958.

Thal-Larsen H., Takahashi Y., Identification of Process Parameters by Means of Models, ASME Paper 60-WA-21 (Winter Annual Meeting, N.Y.), 1960.

Tomović R., Sensitivity Analysis of Dynamic Systems, N.Y., 1964.

Wang M. C., Uhlenbeck G. E., On the Theory of the Brownian Motion II, *Rev. Mod. Phys.*, 17, 323—342 (1945).

Write A. J., The Analysis and Design of Model Reference Adaptive Control Systems, *Proc. IEE*, 113, 175—184 (1966).

Young P. C., Regression Analysis and Process Parameter Estimation: a Cautionary Message, *Simulation*, 8, 125—128 (1968).

Zhivoglyadov V. P., Kaipov V. K., Identification on Distributed Parameter Plants in the Presence of Noises, IFAC Symp. «Identification in Autom. Control Systems», Prague, Paper 3, 5, 1967.

9.5. Две модели

Balchen J. G., Hsien C., Adjustable Models for Determination of the Statistical and Dynamical Behaviour of Processes (на немецком), *Regelungstechnik*, 14, 145—150 (1966).

Blandhol E., On the Use of Adjustable Models for Determination of System Dynamics, Techn. Rep. Division of Automatic Control, Techn. Univ., Trondheim, Norway, 1962.

Blandhol E., Balchen J. G., Determination of System Dynamics by the use of adjustable Models, Second IFAC Congress, Basle, 1963, 620; есть русский перевод: Блэнхол Э., Балкен И., Определение динамических характеристик при помощи самонастраивающихся моделей, в книге «Теория непрерывных автоматических систем», Труды II Межд. конгресса ИФАК, изд-во «Наука», 1965.

9.6. Возмущения параметров

Алимов Ю. И., Приближенный расчет одного класса автоматических систем с форсированной оптимизацией параметров, в книге

«Оптимальные системы. Статистические методы», Труды II Межд. конгресса ИФАК, изд-во «Наука», 1965.

Anderson G. W., et al., A Self-Adjusting System for Optimum Dynamic Performance, *IRE Conv. Rec.*, Part 4, 182—190 (1958).

Bogirno J. J., Jr., Stability and Convergence Properties of Model-Reference Adaptive Control Systems, *IRE Trans. Autom. Control*, AC-7, 30—41 (1962).

Dorf R. C., Byars W. A., A Modulation Method for Model Reference Adaptive Control Systems, *IEEE Internat. Conv.* (1963).

Douce J. L., Ng K. C., A Six Channel Adaptive Computer, *Proc. IEE*, 111, 1757 (1964).

Douce J. L., Gupta M. M., Dynamics of the Parameter-Perturbation Process, *Control and Science Record*, 113, 1077 (1966); *Proc. IEE*, 113, 1077 (1966).

Eveleigh V. W., General Stability Analysis of Sinusoidal Perturbation Extrema Searching Adaptive Systems, Second IFAC Congress, Basle, 1963, 472; есть русский перевод: Эвелей В. У., Общий анализ устойчивости самонастраивающихся систем поиска экстремума, в книге «Дискретные и самонастраивающиеся системы», Труды II Межд. конгресса ИФАК, изд-во «Наука», 1965.

Eykhoff P., Process-Parameter Estimation Using an Analog Model, Proceed. Third Internat. Analogue Computation Meetings, Opatija, Yugoslavia, 1961, pp. 276—290.

Eykhoff P., Smith O. J. M., Optimizing Control with Process Dynamics Identification, *IRE Trans. Autom. Control*, AC-7, 140—155 (1962).

Gupta M. M., Simultaneous Generation of First and Higher-Order Gradient Functions of a Cost Function (Correspondence), *IEEE Trans. Autom. Control*, AC-13, 115—116 (1968).

Isobe T., Totani T., Analysis and Design of a Parameter-Perturbation Adaptive System for Application to Process Control, Second IFAC Congress, Basle, 1963, p. 317.

Krasovski A. A. (см. литературу к разд. 14.1).

Lackey R. B., An Adaptive Control System with Sinusoidal Parameter Perturbation, *AIEE Trans.*, 81, Part II (Applic. and Ind., № 61), 157—159 (1962).

McGrath R. J., Rideout V. C., A Simulator Study of a Two-Parameter Adaptive System, *IRE Trans. Autom. Control*, AC-6, 35—42 (1961).

McGrath R. J., Rajaraman V., Rideout V. C., A Parameter-Perturbation Adaptive Control System, *IRE Trans. Autom. Control*, AC-6, 154—162 (1961).

Narendra K. S., Baker T. S., Gradient Generation Using a Single Perturbation Signal (Correspondence), *IEEE Trans. Autom. Control*, AC-13, 298—299 (1968).

Rajamaran V., Theory of a Two-Parameter Adaptive Control System, *IRE Trans. Autom. Control*, AC-7, 20—26 (1962).

Smith K. C., Adaptive Control Through Sinusoidal Response, *IRE Trans. Autom. Control*, AC-7, 129—139 (1962).

Weygandt C. N., Puri N. N., Transfer Function Tracking and Adaptive Control Systems, *IRE Trans. Autom. Control*, AC-6, (1961).

ГЛАВА 10

ТЕСТОВЫЕ СИГНАЛЫ: ПЕРИОДИЧЕСКИЕ И БОЛЬШИЕ

Принципиальным моментом задачи оценивания является необходимость возбуждения объекта. Часто в тех случаях, когда это возможно, оказывается выгодным использовать специальные тестовые сигналы, подаваемые на объект извне; наблюдение только спонтанных флюктуаций дает менее точные результаты. Использование возмущений, содержащихся в самом объекте, затруднительно по следующим причинам:

а) Энергетический спектр возмущений может быть слишком узким или содержать доминирующие частоты, что требует предварительного приведения его к широкополосному виду («отбелыванию») либо периодической отсортировки для приближения спектра входного сигнала к спектру белого шума.

б) Из-за того что доступны лишь конечные реализации входных и выходных сигналов, точность оценивания корреляционных функций ограничена. Возникает необходимость выбора временных «окон», используемых для оценок. С расширением окна увеличивается ширина спектра, но уменьшается точность оценок.

Априорную информацию о системе можно использовать для нахождения оптимального тестового сигнала, под которым понимается, например, сигнал с заранее определенной максимальной амплитудой (допустимое возмущение), обеспечивающий получение требуемой информации с заданной точностью за минимальное время. Имеется большой выбор тестовых сигналов, однако при малой априорной информации о системе этот выбор редко бывает существенным. Исчерпывающее обсуждение тестовых сигналов можно найти в отличной монографии [33].

В этой главе будут рассмотрены некоторые типы тест-сигналов. Вообще говоря, типом входного сигнала определяется вид информации, которую с его помощью можно получить. Пусть поведение линейного объекта описывается уравнением

$$a_2 \ddot{y}(t) + a_1 \dot{y}(t) + a_0 y(t) = b_1 \dot{u}(t) + b_0 u(t),$$

или

$$\begin{aligned} a_2 \ddot{y}(t) + a_1 \dot{y}(t) + a_0 y(t) &= \\ &= b_1 \dot{u}(t - \tau) + b_0 u(t - \tau), \end{aligned} \quad (10.1)$$

где $u(t)$ — входной сигнал, а $y(t)$ — (неискаженный) выходной сигнал. Описание объекта во временной области приводит к интегралам свертки

$$y(t) = \int_{-\infty}^{\infty} h(\theta) u(t - \theta) d\theta, \quad h(t) = 0 \text{ при } t < 0, \quad (10.2)$$

$$\begin{aligned} y(t) &= \int_{-\infty}^{\infty} h(\theta) u(t - \tau - \theta) d\theta = \int_{-\infty}^{\infty} h'(\theta) u(t - \theta) d\theta, \\ h'(t) &= 0 \text{ при } t < \tau, \end{aligned}$$

представление объекта в частотной области имеет вид

$$Y(s) = H(s) U(s),$$

или

$$Y(s) = H'(s) U(s),$$

где

$$\begin{aligned} H(s) &= \frac{b_1 s + b_0}{a_2 s^2 + a_1 s + a_0}, \\ H'(s) &= \frac{b_1 s + b_0}{a_2 s^2 + a_1 s + a_0} e^{-s\tau} \end{aligned} \quad (10.3)$$

и $s = j\omega$ для синусоидальных сигналов. Вообще говоря, измерение переходных характеристик наиболее полезно при первом способе описания (разд. 10.1), тогда как при втором способе описания большей частью используются периодические сигналы (разд. 10.2).

Связь между представлениями во временной и частотной областях рассматривалась в гл. 3 и 4. Для практичес-

ских целей необходимо знать, как переходить от представления во временной области к представлению в частотной области с помощью методов преобразования Фурье (см. гл. 3). Рядом интересных свойств обладают двоичные последовательности при их использовании в качестве тестсигналов (разд. 10.3). Влияние характеристик тестсигналов на точность оценок обсуждается в гл. 11.

10.1. ИМПУЛЬСНЫЕ И СТУПЕНЧАТЫЕ СИГНАЛЫ

Для представления информации во временной области очевидна целесообразность использования импульсных и ступенчатых тестовых сигналов. Теоретический интерес представляет входной сигнал, близкий по свойствам к импульсной функции Дирака

$$u(t) = c\delta(t) \begin{cases} \delta(t) = 0, & t \neq 0, \\ \int_{-\infty}^{\infty} \delta(t) dt = 1. \end{cases} \quad (10.4)$$

После подстановки в интеграл свертки (10.2) получаем $y(t) = ch(t)$. Имеющий большее практическое значение входной тестсигнал служит приближением к функции единичного скачка

$$u(t) = cr(t) \begin{cases} r(t) = 0, & t < 0, \\ r(t) = 1, & t \geq 0. \end{cases} \quad (10.5)$$

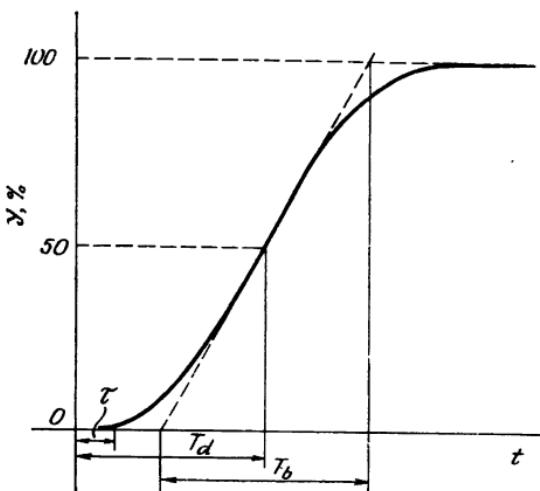
Интеграл свертки принимает вид

$$y(t) = c \int_0^t h(\theta) d\theta, \quad (10.6)$$

откуда, дифференцируя, получаем

$$\frac{dy}{dt} = ch(t). \quad (10.7)$$

Спектральная плотность ступенчатых сигналов убывает на высоких частотах, где большинство объектов обладают заметным ослаблением. Импульсные и ступенчатые сигналы не должны, конечно, наносить вред объекту либо выводить его в область нелинейности, если по усло-



Ф и г. 10.1.

виям задачи необходимо линейное представление объекта. Наконец, при исследовании объектов применяются «пологие» сигналы.

По реакции объекта на ступенчатый вход можно сразу определить ряд практически важных параметров, например (фиг. 10.1) τ — время задержки, T_d — время нарастания выходного сигнала до 50%-ного уровня, T_b — постоянную времени.

Более полный обзор и перечисление подобных определений, особенно при наличии перерегулирования, можно найти в стандартах ASA [3]. Достаточно общим является случай, когда можно ожидать, что передаточная функция имеет полюсы только на действительной оси. Следовательно,

$$H(s) = \frac{ke^{-st}}{\prod_{i=1}^n (st_i + 1)}.$$

Если можно допустить, что все постоянные времени равны между собой, то по кривой реакции на ступенчатый сигнал непосредственно определяются их количество и числовое значение каждой. Подобные правила определения постоянных времени для более общих случаев приведены, например, в [30, 33].

Влияние аддитивного шума

Рассмотрим теперь влияние аддитивного шума $n(t)$. Вместо невозмущенного выхода $x(t)$ наблюдается только $y(t) = x(t) + n(t)$. Характеристиками шума служат

$$\mathcal{E}[n(t)] = 0,$$

$$\mathcal{E}[n^2(t)] = \sigma^2.$$

Неопределенность измерений, вносимая шумом, определяется стандартным отклонением σ . Эту неопределенность можно уменьшить, повторяя испытание несколько (скажем, k) раз в предположении воспроизводимости детерминированной части сигналов. Складывая k случайных функций времени и деля сумму на k (фиг. 10.2), получим функцию с меньшим стандартным отклонением. Это нетрудно показать.

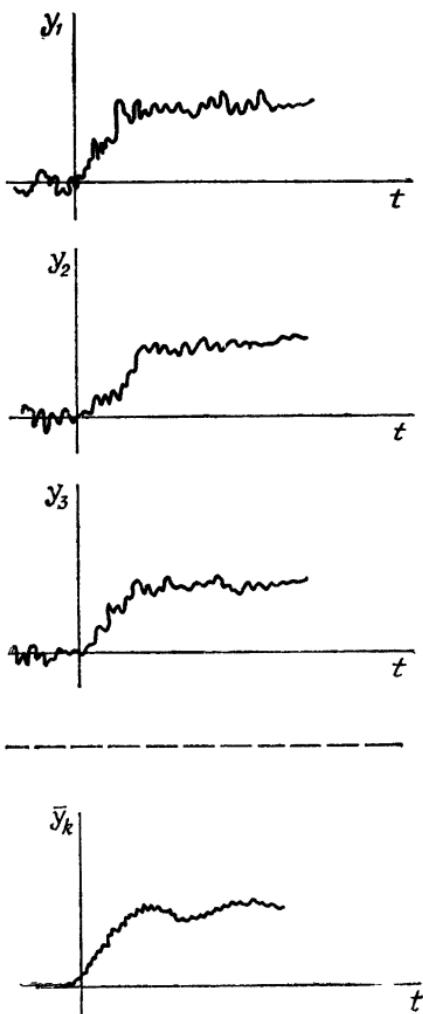
Пусть последовательные испытания начинаются в моменты $t = t_1, t_2, \dots, t_k$.

Рассмотрим значения выходного сигнала, зафиксированные через время θ после начала каждого испытания:

$$y(t_i + \theta) = x(t_i + \theta) + n(t_i + \theta),$$

или

$$y_i = x_i + n_i.$$



Фиг. 10.2.

Среднее значение по k испытаниям равно.

$$\bar{y}_k = \frac{1}{k} \sum_{i=1}^k y_i = x + \frac{1}{k} \sum_{i=1}^k n_i, \quad (10.8)$$

поскольку детерминированная часть сигнала предполагается воспроизводимой, т. е. $x_i = x$ для всех i . Следовательно,

$$\mathcal{E}[\bar{y}_k] = x, \quad \mathcal{E}[(\bar{y}_k - x)^2] = \frac{\sigma^2}{k}.$$

Последнее соотношение имеет место лишь для некоррелированных n_i и n_j , $i \neq j$. Это соотношение выполняется с удовлетворительной точностью, если для корреляционной функции шума справедливо приближенное равенство

$$\Psi_{nn}(t_i - t_j) \approx 0,$$

т. е. при достаточно разделенных во времени последовательных экспериментах. При таком условии стандартное отклонение уменьшается обратно пропорционально корню квадратному из числа экспериментов [8—10, 24]. Этот подход, называемый *методом усреднения отклика*, реализуется в серийно выпускаемом оборудовании и находит применение при изучении потенциалов мозговой активности, нервных потенциалов, кардиограмм и т. д. Кроме того, для этой цели используются небольшие цифровые вычислительные машины с аналого-цифровыми преобразователями.

Другой величиной, представляющей интерес, является дисперсия σ^2 . По k наблюдениям она оценивается следующим образом:

$$\overline{(y - \bar{y}_k)_k^2} = \frac{1}{k} \sum_{i=1}^k (y_i - \bar{y}_k)^2. \quad (10.9)$$

Математическое ожидание этой величины равно

$$\begin{aligned} \mathcal{E}[\overline{(y - \bar{y}_k)_k^2}] &= \frac{1}{k} \mathcal{E}\left[\sum_{i=1}^k (y_i - x + x - \bar{y}_k)^2\right] = \\ &= \frac{1}{k} \mathcal{E}\left[\sum (y_i - x)^2 + 2 \sum (y_i - x)(x - \bar{y}_k) + \sum (x - \bar{y}_k)^2\right] = \\ &= \sigma^2 - \frac{2}{k} \sigma^2 + \frac{1}{k} \sigma^2 = \frac{k-1}{k} \sigma^2. \end{aligned}$$

Следовательно, в качестве несмещенной оценки σ^2 следует выбирать

$$s^2 = \frac{k}{k-1} \overline{(y - \bar{y}_k)^2} = \frac{1}{k-1} \sum_{i=1}^k (y_i - \bar{y}_k)^2. \quad (10.10)$$

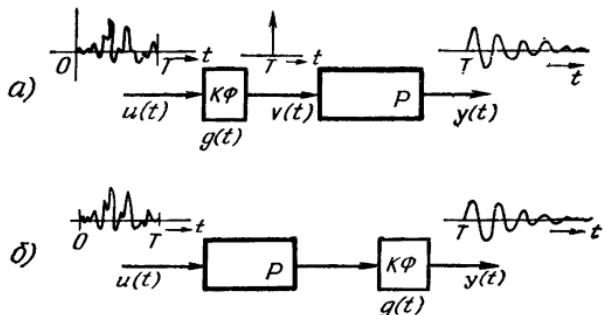
(Обратите внимание на случай $k = 1$.)

Применение корреляционных фильтров

Используемые импульсные и ступенчатые тестовые сигналы должны иметь значительную амплитуду. Это означает, что объект может оказаться в нелинейной области, где весовая функция уже не обеспечивает адекватного описания. Характеристики объекта, определенные в таких условиях, могут отличаться от характеристик в нормальных условиях. Это затруднение можно обойти, распространяя энергию тестсигнала на больший интервал времени, так чтобы амплитуда была малой. В каждом конкретном случае это можно осуществить с помощью так называемых корреляционных фильтров (КФ) и соответствующих тестсигналов (фиг. 10.3, a); см. также [22]. Тестсигнал $u(t)$ и весовая функция $g(t)$ связаны соотношением $g(t) = u(T-t)$. Следовательно, выход фильтра при таком входном сигнале есть

$$v(t) = \int_0^\infty u(t-\theta) g(\theta) d\theta = \int_0^\infty u(t-\theta) u(T-\theta) d\theta. \quad (10.11)$$

Это соотношение определяет корреляционную функцию сигнала u . При соответствующем выборе $u(t)$ и $g(t)$ сигнал



Ф и г. 10.3.

$v(t)$ может сколь угодно точно аппроксимировать импульсную функцию $cb(t)$. При этом выход $y(t)$ объекта P является аппроксимацией его весовой функции. Если объект P линеен, КФ и P можно расположить в соответствии с фиг. 10.3, б без изменения $y(t)$, причем теперь энергия сигнала $v(t) \approx cb(t)$ распределена на интервале времени T . Влияние аддитивного шума по-прежнему можно ослабить повторением эксперимента и усреднением откликов.

Свободу выбора u , определяемую условием $g(t) = u(T - t)$, можно использовать для выполнения ограничения, которое состоит в том, что максимальная амплитуда u должна оставаться в допустимых пределах. Эта задача аналогична задаче о выборе формы радиолокационных сигналов для повышения разрешающей способности [37].

10.2. СИНУСОИДАЛЬНЫЕ ТЕСТОВЫЕ СИГНАЛЫ

Информацию в частотной форме представления проще всего получить, используя синусоидальные или другие периодические тестовые сигналы. Этот метод исследования объектов имеет длинную историю [2]. Существует немало публикаций по этому вопросу [11, 16]. Синусоидальные сигналы обладают многими преимуществами. Можно ограничиться измерениями только на представляющих интерес частотах; свойства ортогональности синусоидальных сигналов предоставляют определенные возможности индексации; легко проверяются предположения о линейности; длительность эксперимента можно выбирать произвольно; генерирование и обработка таких сигналов изучены весьма глубоко; в классической теории систем с обратной связью исследование устойчивости проводится с использованием синусоидальных сигналов. Недостатком является опасность возбуждения объекта на резонансных частотах.

Принципы этого метода можно показать на следующем примере. Пусть объект описывается дифференциальным уравнением

$$\begin{aligned} a_n y^{(n)} + \dots + a_1 \dot{y} + a_0 y &= \\ &= b_m u^{(m)} + \dots + b_1 \dot{u} + u. \end{aligned} \quad (10.12)$$

Входной сигнал выбирается в виде

$$u(t) = u_0 \sin \omega t = \frac{u_0}{2j} \{e^{j\omega t} - e^{-j\omega t}\}. \quad (10.13)$$

Благодаря линейности объекта его выход после окончания переходного процесса состоит из следующей суммы откликов на каждую из двух компонент входа:

$$y(t) = \frac{u_0}{2j} \{H(j\omega) e^{j\omega t} - H(-j\omega) e^{-j\omega t}\}, \quad (10.14)$$

где

$$H(j\omega) = \frac{b_m(j\omega)^m + \dots + b_1(j\omega) + 1}{a_n(j\omega)^n + \dots + a_1(j\omega) + a_0}, \quad (10.15)$$

$$H(j\omega) = |H(j\omega)| e^{j\varphi(\omega)}, \quad (10.16)$$

$$H(-j\omega) = |H(j\omega)| e^{-j\varphi(\omega)}.$$

Следовательно,

$$\begin{aligned} y(t) &= u_0 |H(j\omega)| \sin \{\omega t + \varphi(\omega)\} = \\ &= u_0 \{c(\omega) \sin \omega t + d(\omega) \cos \omega t\}, \end{aligned} \quad (10.17)$$

где

$$c(\omega) = |H(j\omega)| \cos \varphi(\omega), \quad (10.18)$$

$$d(\omega) = |H(j\omega)| \sin \varphi(\omega),$$

$$|H(j\omega)| = \{c^2(\omega) + d^2(\omega)\}^{1/2}, \quad (10.19)$$

$$\varphi(\omega) = \arctg \left\{ \frac{d(\omega)}{c(\omega)} \right\}.$$

Подлежащими определению параметрами объекта являются $c(\omega_i)$ и $d(\omega_i)$ для ряда частот ω_i . Их можно найти, возбуждая объект сигналом $u(t) = u_0 \sin \omega_i t$ и применяя фильтр Фурье к выходному сигналу $y(t)$. Такая процедура фильтрации сводится к умножению $y(t)$ соответственно на $\sin \omega_i t$ и $\cos \omega_i t$ и усреднению по целому числу периодов:

$$\begin{aligned} c(\omega_i) &= \frac{2}{u_0 T} \int_0^T y(t) \sin \omega_i t dt, \\ d(\omega_i) &= \frac{2}{u_0 T} \int_0^T y(t) \cos \omega_i t dt, \end{aligned} \quad (10.20)$$

где

$$T = kT_i, \quad T_i = \frac{2\pi}{\omega_i}.$$

Этот подход тесно связан с корреляционными методами (см. гл. 8). Аналогично можно измерить гармонические составляющие $y(t)$, если $u(t) = u_0 \sin \omega_i t$ и выходные сигналы умножаются на синусоиды с частотами $l\omega_i$, $l = 2, 3, 4 \dots$. Другой задачей является определение из графиков Найквиста параметров a_i и b_j передаточной функции $H(j\omega)$, определяемой формулой (10.15) [33].

Влияние аддитивного шума [34]

До сих пор возмущения не рассматривались. Если $y(t)$ искажается шумом $n(t)$, результатом фильтрации Фурье будет

$$\begin{aligned} \tilde{c}(\omega_i, T) &= c(\omega_i) + \frac{2}{u_0 T} \int_0^T n(t) \sin \omega_i t dt, \\ \tilde{d}(\omega_i, T) &= d(\omega_i) + \frac{2}{u_0 T} \int_0^T n(t) \cos \omega_i t dt. \end{aligned} \quad (10.21)$$

Теперь \tilde{c} и \tilde{d} оказываются приближенными значениями, зависящими от интервала интегрирования $T = kT_i$. Поскольку сигнал $n(t)$ случаен, c и d также являются случайными величинами. В дальнейшем предполагается, что $E[n(t)] = 0$.

Применение операции математического ожидания позволяет получить

$$\begin{aligned} \mu_c(T) &= E[\tilde{c}(\omega_i, T)] = \\ &= E[c(\omega_i)] + \frac{2}{u_0 T} \int_0^T E[n(t)] \sin \omega_i t dt = c(\omega_i), \quad (10.22) \\ \mu_d(T) &= E[\tilde{d}(\omega_i, T)] = d(\omega_i). \end{aligned}$$

Следовательно, результаты измерений несмещенные. Представляет интерес дисперсия измерений

$$\begin{aligned}\sigma_c^2(T) &= \mathcal{E} [\{\tilde{c}(\omega_i, T) - c(\omega_i)\}^2] = \\ &= \mathcal{E} \left[\left\{ \frac{2}{u_0 T} \int_0^T n(t) \sin \omega_i t \, dt \right\}^2 \right] = \\ &= \frac{4}{u_0^2 T^2} \int_0^T \int_0^T \mathcal{E} [n(u) n(v)] \sin \omega_i u \sin \omega_i v \, du \, dv,\end{aligned}\quad (10.23)$$

где

$$\mathcal{E} [n(u) n(v)] = \Psi_{nn}(v - u) = \Psi_{nn}(\tau).$$

В приложении Г показано, что

$$\begin{aligned}\sigma_c^2(T) &= \frac{4}{u_0^2 T} \int_0^T \Psi_{nn}(\tau) \left\{ \left(1 - \frac{\tau}{T} \right) \cos \omega_i \tau \pm \right. \\ &\quad \left. \pm \frac{1}{\omega_i T} \sin \omega_i \tau \right\} d\tau.\end{aligned}\quad (10.24)$$

Если аддитивный шум белый, то

$$\Psi_{nn}(\tau) = c\delta(\tau)$$

и, так как $\Psi_{uu}(0) = u_0^2/2$ при $u = u_0 \sin \omega_i t$, в соответствии с равенством

$$\int_0^a \delta(\tau) d\tau = \frac{1}{2} \int_{-a}^a \delta(\tau) d\tau = \frac{1}{2}$$

можно получить

$$\sigma_c^2(T) = \sigma_d^2(T) = \frac{2c}{u_0^2 T} = \frac{c}{\Psi_{uu}(0) T}.\quad (10.25)$$

Как и в рассмотренных в предыдущих главах случаях, стандартное отклонение σ обратно пропорционально \sqrt{T} или \sqrt{k} .

Если аддитивный шум можно рассматривать как белый шум, пропущенный через низкочастотный фильтр с передаточной функцией

$$H(s) = \frac{\alpha}{s + \alpha},$$

то

$$\Phi_{nn}(s) = c \frac{\alpha^2}{\alpha^2 - s^2}, \quad \Psi_{nn}(\tau) = \frac{c}{2} \alpha e^{-\alpha|\tau|}. \quad (10.26)$$

Подставив эту корреляционную функцию в формулу (10.24), получим

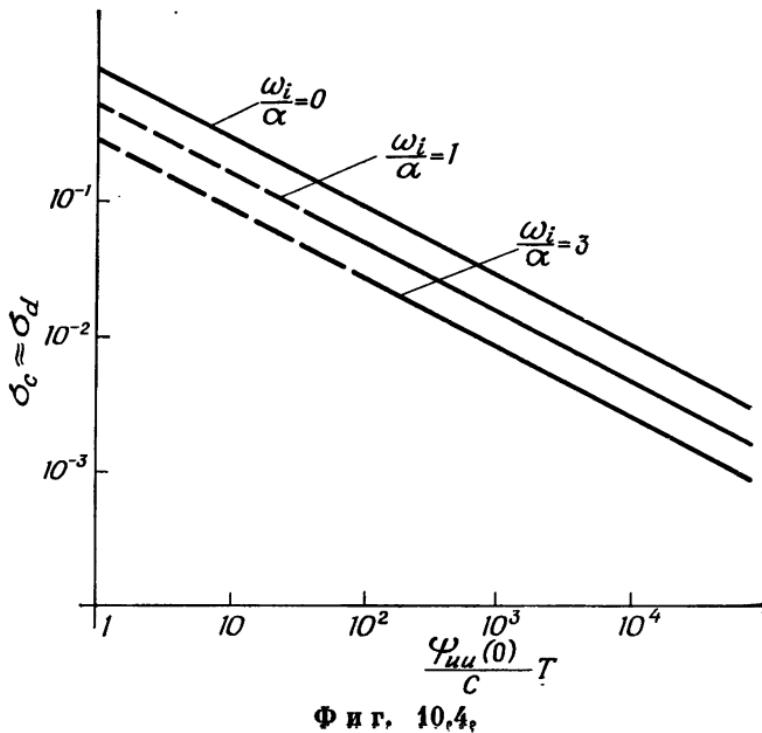
$$\sigma_c^2(T) = \frac{c}{\Psi_{uu}(0)} \frac{\alpha}{T} \left\{ \frac{\alpha}{\omega_i^2 + \alpha^2} + \frac{2}{T} \frac{\omega_i^2}{(\omega_i^2 + \alpha^2)^2} (1 - e^{-\alpha T}) \right\}, \quad (10.27a)$$

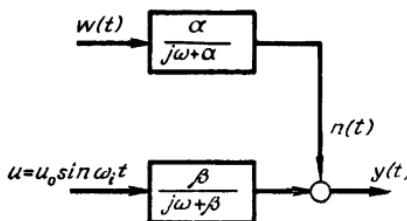
$$\sigma_d^2(T) = \frac{c}{\Psi_{uu}(0)} \frac{\alpha}{T} \left\{ \frac{\alpha}{\omega_i^2 + \alpha^2} - \frac{2}{T} \frac{\alpha^2}{(\omega_i^2 + \alpha^2)^2} (1 - e^{-\alpha T}) \right\}. \quad (10.27b)$$

При $T \rightarrow \infty$

$$\sigma_c^2(T) \approx \sigma_d^2(T) \approx \frac{c}{\Psi_{uu}(0) T} \frac{\alpha^2}{\omega_i^2 + \alpha^2}; \quad (10.28)$$

при $\alpha \rightarrow \infty$ (белый шум) эта формула сводится к (10.25). Для белого шума на достаточно большом интервале дисперсия не зависит от ω_i ; для «красного» шума она растет с увеличением ω_i (фиг. 10.4). В работе [34] показано, что





Ф и г. 10.5.

различие между уравнениями (10.27а) и (10.27б) связано с конкретным выбором интервала интегрирования $(0, T) = (0, kT_i)$. Можно использовать интервал $\{(\phi/2\pi) T_i, [k + (\phi/2\pi)] T_i\}$. Если ϕ — случайная величина, равномерно распределенная в интервале $0 \leq \phi < 2\pi$, то $\mathcal{E}_\phi [\sigma_c^2 (T)] = \mathcal{E}_\phi [\sigma_d^2 (T)]$. Помимо σ_c и σ_d , представляет интерес корреляция ошибок в c и d . Ее можно охарактеризовать величиной

$$\sigma_{cd}^2 (T) = \mathcal{E} [\{\tilde{c}(\omega_i, T) - c(\omega_i)\} \{\tilde{d}(\omega_i, T) - d(\omega_i)\}]. \quad (10.29)$$

Нетрудно показать, что $\sigma_{cd}^2 (T) \equiv 0$ независимо от T , ω_i и корреляционной функции шума $\Psi_{nn}(\tau)$.

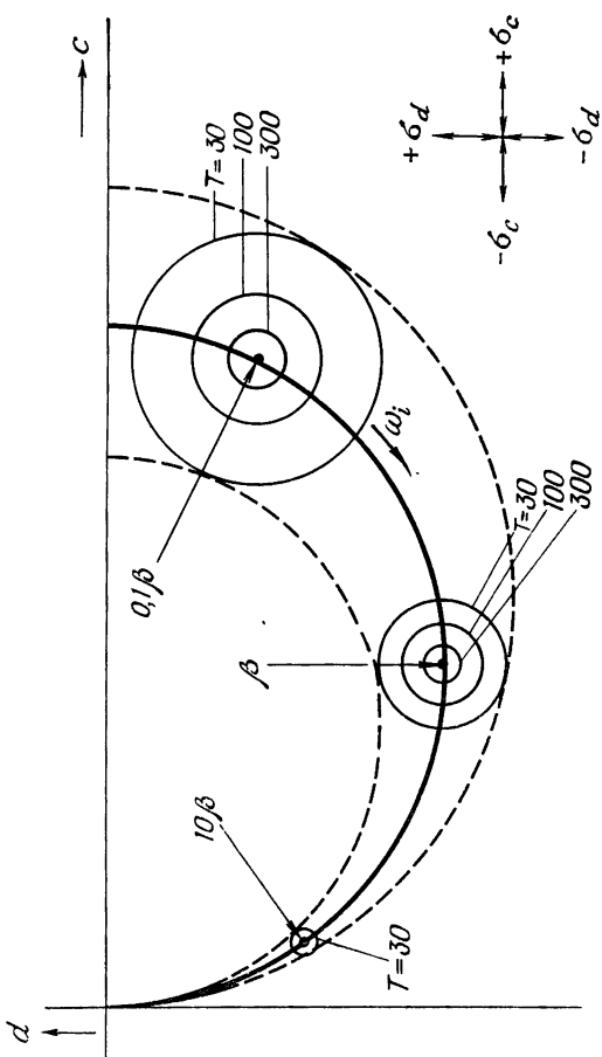
В качестве примера рассмотрим построение графика Найквиста для простого объекта первого порядка

$$H(j\omega) = \frac{\beta}{j\omega + \beta}$$

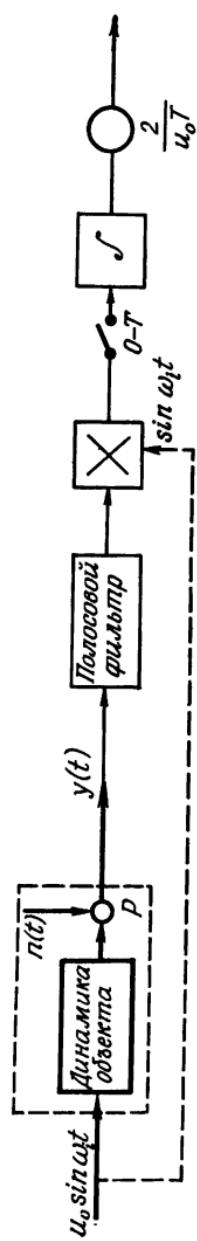
при наличии аддитивного шума вида (10.26) (фиг. 10.5). Пусть $w(t) = n(t)$ и $\beta = \alpha$. Тогда σ_c и σ_d можно вычислить при разных значениях ω_i , что с учетом равенства $\sigma_{cd} \equiv 0$ приводит к ситуации, показанной на фиг. 10.6, где окружности вокруг точек задают интервалы шириной в одно стандартное отклонение при разных длинах интервала интегрирования. Другие аспекты фильтрации Фурье рассмотрены в работах [4, 5, 14, 18, 20, 21, 23, 31, 35].

Другие родственные процедуры [34]

Исследуя нежелательное влияние аддитивного шума, можно задать вопрос, не существует ли для этого случая лучших методов. Под лучшим здесь понимается метод, дающий при том же интервале наблюдений или интегри-



Ф и г. 10.6.

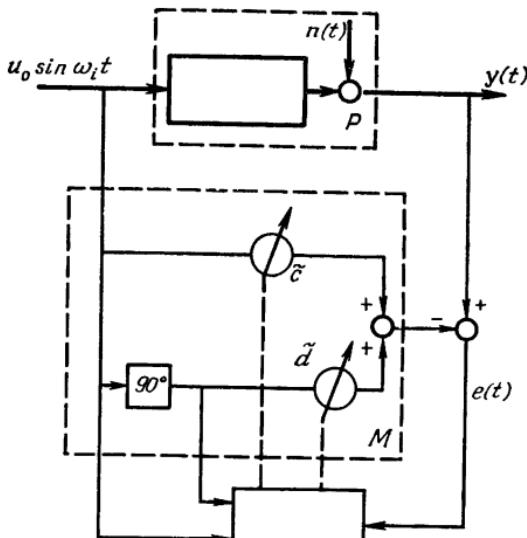


Ф и г. 10.7.

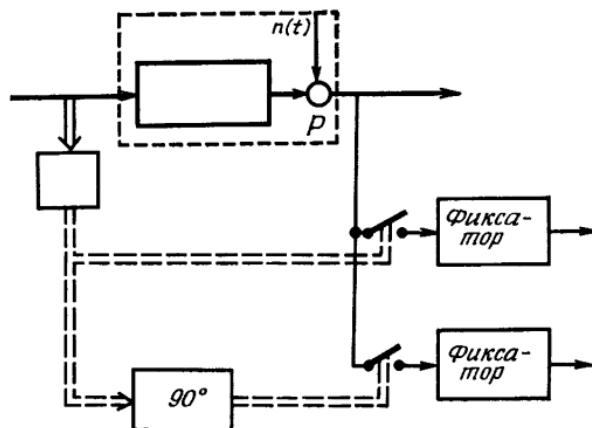
рования меньшую дисперсию, чем фильтр Фурье. В частности, можно предложить следующие методы.

а) Использование дополнительных полосовых фильтров совместно с фильтром Фурье для ослабления шума, создающего помехи наблюдениям (фиг. 10.7). В самом деле, при таком подходе удается воздействовать на корреляционную функцию $\Psi_{nn}(\tau)$ из (10.24) в благоприятном направлении. Этот метод, однако, имеет некоторые недостатки: средняя частота полосового фильтра должна быть жестко связана с частотой ω_i , на которой проводятся измерения; точность настройки полосовых фильтров относительно ω_i весьма существенна, так как незначительная расстройка проявляется в отличном от нуля фазовом сдвиге, приводящем к ошибкам в $c(\omega_i)$ и $d(\omega_i)$; при подаче синусоидального сигнала на полосовой фильтр начинаются переходные процессы, во время которых фильтр Фурье не может использоваться. При сравнении с характеристиками немодифицированного фильтра Фурье длительность этих переходных процессов нужно добавить к интервалу интегрирования фильтра Фурье. Получается, что эти недостатки перевешивают возможные достоинства.

б) Использование компенсации (настройки модели) (фиг. 10.8). Этот метод можно сравнить с методами, изло-



Ф и г. 10.8.



Ф и г. 10.9.

женными в гл. 7 и 9. В данном случае модель очень проста, поскольку имеются всего два параметра c и d . По-прежнему нужно минимизировать некоторый критерий, зависящий от ошибки $e(t)$. Можно показать, что интегральный критерий среднеквадратической ошибки тесно связан с обсуждавшимся выше фильтром Фурье. Отметим здесь работы [28, 29].

в) Использование *дискретизации времени*. Этот метод можно пояснить следующим образом. Пусть

$$\begin{aligned} y(t) &= u_0 |H(j\omega)| \sin \{\omega t + \varphi(\omega)\} = \\ &= u_0 |H(j\omega)| \sin \omega t \cos \varphi(\omega) + \\ &\quad + u_0 |H(j\omega)| \cos \omega t \sin \varphi(\omega). \end{aligned} \quad (10.30)$$

Если $\omega t_i = \pi/2 + i2\pi$, то

$$y(t_i) = u_0 |H(j\omega)| \cos \varphi(\omega) = u_0 c(\omega). \quad (10.31)$$

При $\omega t_j = j2\pi$

$$y(t_j) = u_0 |H(j\omega)| \sin \varphi(\omega) = u_0 d(\omega). \quad (10.32)$$

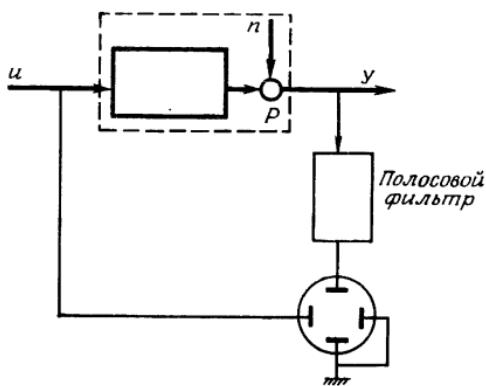
Следовательно, подлежащие определению параметры можно найти по выборочным значениям процессов в моменты t_i и t_j . Поэтому метод оказывается очень простым (фиг. 10.9). Ясно, однако, что этот метод очень чувствителен к наличию высших гармоник. Влияние аддитив-

ногого шума можно снова уменьшить усреднением по дискретным выборкам в нескольких точках t_i , $i = 0, 1, 2, \dots, k$ и t_j , $j = 0, 1, 2, \dots, k$. Сравнение дисперсий оценок, получаемых по этому методу и с помощью фильтра Фурье, обнаруживает (существенное) преимущество последнего. Можно несколько улучшить предлагаемый метод, производя выборку два раза за период и меняя знак у соответствующих значений на обратный. Несмотря на возникающие усложнения, приведенный выше вывод остается в силе.

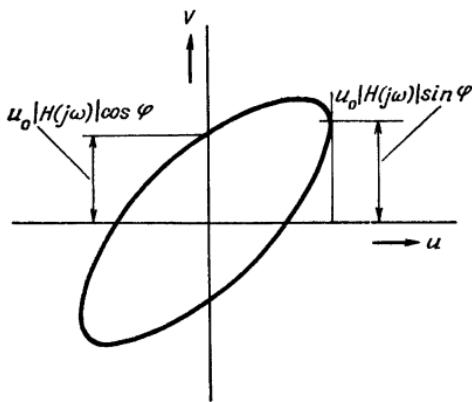
г) Использование *полосовых фильтров*. Для подавления высших гармоник и шума выходной сигнал $y(t)$ пропускается через полосовой фильтр и затем используется для получения кривых Лиссажу на осциллографе или граffопостроителе (фиг. 10.10). По этим кривым можно определить оба параметра в соответствии с фиг. 10.11. В принципе метод прост, но, как и при ранее упоминавшемся применении полосовых фильтров, здесь возникает ряд проблем. Теоретическое рассмотрение показывает, что фильтр Фурье предпочтительнее [23].

д) Использование *оптимальных фильтров*. Рассматриваемую задачу можно интерпретировать как задачу оценивания неизвестных амплитуды и фазы (или действительной и мнимой составляющих) сигнала известной частоты в присутствии помех. Имеется множество публикаций, касающихся такой постановки задачи оптимальной фильтрации, которая может быть решена с помощью линейных фильтров. Фиг. 10.12 иллюстрирует задачу с этой точки зрения; здесь $g_r(t)$ и $g_i(t)$ — весовые функции оптимальных фильтров соответственно для действительной и мнимой составляющих. Оказывается, что при аддитивном белом шуме оптимальные фильтры идентичны фильтру Фурье.

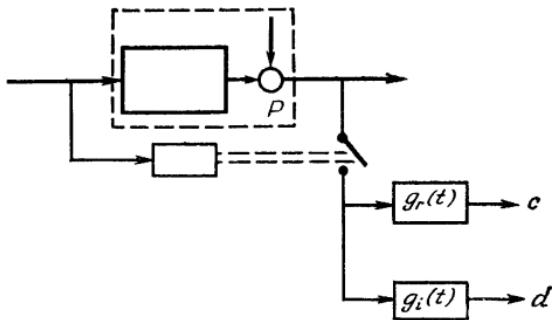
Недостатком синусоидальных сигналов является малое количество информации, приходящейся на единицу частоты тестсигнала. Увеличить эту информацию можно добавлением нескольких гармоник к периодическому тестовому сигналу. Этим способом при надлежащем выборе фазовых соотношений синусоид периодический сигнал можно превратить в двоичный (т. е. принимающий только два значения $+B$ и $-B$) без существенного иска-



Ф и г. 10.10.



Ф и г. 10.11.



Ф и г. 10.12.

жения исходного энергетического спектра [19]. Такие двоичные сигналы легко использовать при исследовании многих линейных объектов, если это допускается условиями эксплуатации. Из выходного сигнала объекта отфильтровываются гармоники, составляющие входной сигнал. При этом удается получать одновременно информацию об амплитудной передаточной функции объекта на нескольких частотах. Другой подход к определению нестационарных переходных характеристик (человек-оператор) с использованием суммы ряда синусоид описан в [32]. Пока еще нет теории, позволяющей синтезировать релейные сигналы с заранее заданным спектром. Метод проб и ошибок, а также оптимизационные процедуры с релейными сигналами могут дать хорошие приближения к определенным желаемым спектрам и привести к результатам, лучшим по сравнению с получаемыми с помощью псевдослучайных последовательностей.

При отсутствии априорной информации с помощью псевдослучайных последовательностей (разд. 10.3) легко получается общая картина процесса, на основе которой можно подобрать методы для более точного количественного исследования. Теоретической основой оптимального выбора тестовых сигналов является теория статистических решений. По мере увеличения количества информации, извлекаемой из измерений, можно применять все более точные методы обработки результатов измерений. Если это осуществляется системой автоматического управления, используется термин «дualное управление» [15].

10.3. ДВОИЧНЫЕ ПОСЛЕДОВАТЕЛЬНОСТИ

В разд. 8.1 было показано, что белый шум обладает свойствами, желательными с точки зрения определения весовой функции линейного объекта. Его корреляционная функция, аппроксимирующая функцию Дирака, позволяет обойтись без решения уравнения свертки; взаимная корреляционная функция входного и выходного сигналов объекта дает непосредственно весовую функцию объекта. Ту же идею можно использовать для выборочных сигналов; в случае белого шума матрица

[U'U] в системе нормальных уравнений (разд. 6.1) сводится к диагональной матрице при достаточной длине интервала наблюдений. При этом исчезает необходимость обращения матрицы — процедуры, аналогичной решению уравнения свертки в случае непрерывных сигналов. Поэтому не удивительно, что ряд исследователей посвятили немало времени и усилий поиску сигналов, корреляционные свойства которых аналогичны свойствам белого шума. Несмотря на то что требования того же типа имеют место в задачах модуляции радио- и ультразвуковых сигналов [37], существует, по-видимому, очень слабая связь этих задач с вопросами оценивания параметров.

Сигнал (или временная последовательность, которую предполагается построить) может быть как периодическим, так и непериодическим. В классе периодических сигналов рядом интересных практических свойств обладает подкласс двоичных тестсигналов. Действительно, эти сигналы можно генерировать с помощью релейных или цифровых устройств; вместо умножения аналоговых сигналов возникают операции релейного типа; из всех сигналов с амплитудой $-a \leq u \leq a$ двоичный сигнал $u = \pm a$ передает исследуемому объекту максимальную мощность.

На следующей диаграмме показаны основные соотношения:



Сигнал со случайными интервалами (телеграфный) имеет пуассоновское распределение числа переключений амплитуды с одного уровня на другой в заданном интервале времени. Его можно реализовать с помощью радиоактивного источника и детектора радиации, осуществляющего изменение полярности при поступлении каждой частицы. Его корреляционная функция имеет вид

$$\Psi_{uu}(\tau) = \Psi_{uu}(0) \exp[-2\lambda|\tau|], \quad (10.33)$$

где λ — действительная постоянная.

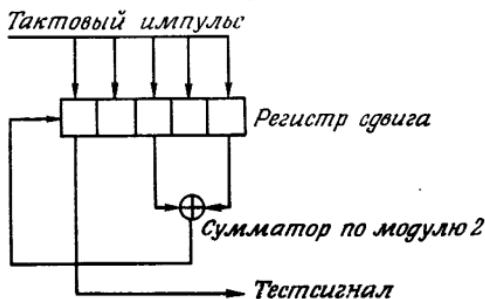
Интервалы между переключениями легко дискретизировать, используя тактовые импульсы с периодом θ . При этом изменение полярности сигнала может произойти только в конце целого числа периодов θ . Корреляционная функция этого сигнала

$$\Psi_{uu}(\tau) = \begin{cases} \Psi_{uu}(0) \left\{ 1 - \left| \frac{\tau}{\theta} \right| \right\}, & |\tau| \leq \theta, \\ 0, & |\tau| > \theta. \end{cases} \quad (10.34)$$

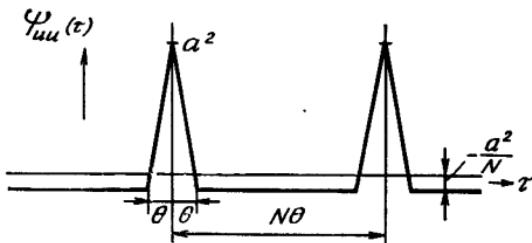
Эта корреляционная функция может сколь угодно точно аппроксимировать δ -функцию Дирака. Такие сигналы обладают еще и тем преимуществом, что легко осуществлять их задержку, используя регистр сдвига, запускаемый теми же тактовыми импульсами.

m-последовательности

Из-за случайной природы упоминавшихся сигналов приходится сталкиваться с внутренней статистической неопределенностью, которая, как объяснено в разд. 6.1, присуща корреляционным операциям. Поэтому определенными преимуществами обладают *детерминированные* сигналы, имеющие заданную корреляционную функцию — псевдослучайные двоичные последовательности (ПСДП). Одно из первых сообщений об оценке динамики объекта таким методом содержится в статье [1]. Здесь генерирование и умножение двоичных тестсигналов осуществляется с помощью кодирующего диска. Вскоре было замечено [27], что так называемые *последовательности максимальной длины* (*m*-последовательности), генерируемые регистрами сдвига с обратной связью, обладают желаемыми



Ф и г. 10.13.



Ф и г. 10.14.

корреляционными характеристиками. Они обладают следующими свойствами:

а) Корреляционная функция этих сигналов, определяемая по целому числу периодов, не содержит элемента случайности (неопределенности).

б) Их легко генерировать с помощью цифровых устройств (см., например, фиг. 10.13).

в) Более того, легко генерируется сигнал с задержкой на любое целое число тактовых импульсов θ , задающих масштаб времени.

г) Период последовательности T легко изменять в соответствии с формулой $T = (2^n - 1)\theta$, где n — число разрядов используемого регистра сдвига. В работе [26] приведены схемы обратных связей для таких регистров сдвига вплоть до $n = 33$.

д) Мощность последовательности равна $\Psi_{ii}(0) = a^2$.

е) Амплитуда постоянного тока последовательности равна только a/N ; следовательно, при больших N вряд ли возникнут отклонения от нормальных условий эксплуатации объекта.

ж) Корреляционная функция показана на фиг. 10.14; соответствующим выбором N и θ можно достичь хорошего приближения к функции Дирака.

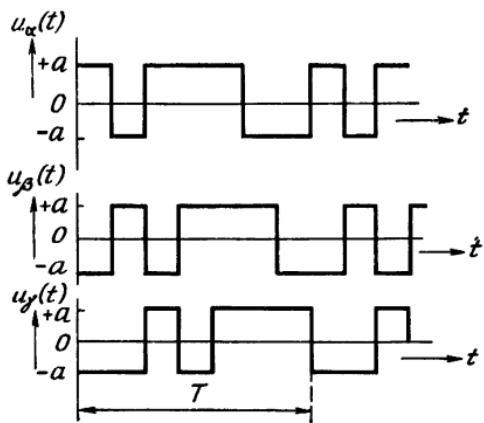
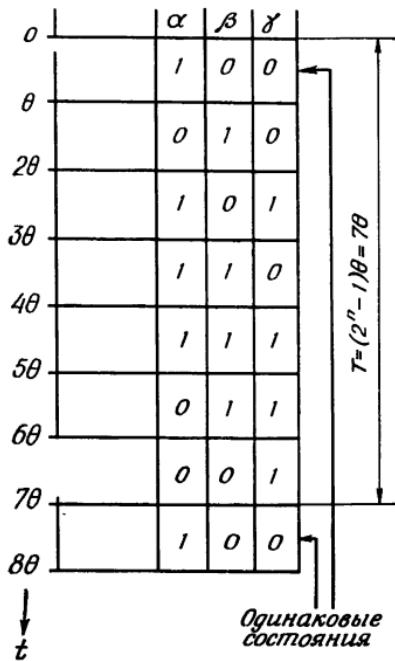
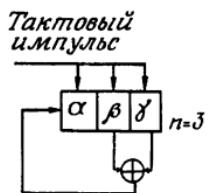
з) Сигнал, принимающий значения $+a$ и $-a$, можно подавать на самые разнообразные объекты. Такие сигналы обладают наибольшей мощностью среди сигналов, ограниченных по амплитуде.

Свойства этих сигналов подробно обсуждаются в [13]. Недостатком, связанным с периодическим характером этих сигналов, является то, что их корреляционная функция с необходимостью оказывается периодической.

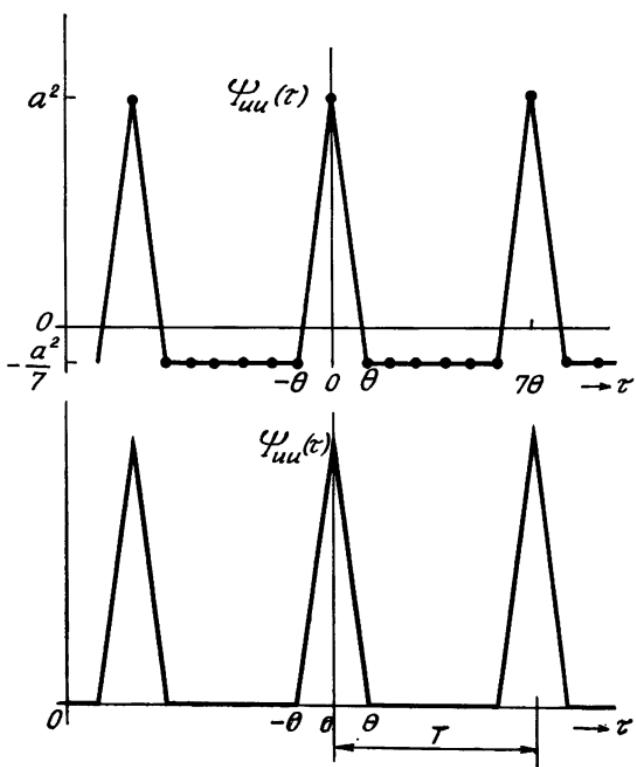
Проиллюстрировать все эти особенности удобно на простом примере (фиг. 10.15). В таблице приведены состояния регистра сдвига, соответствующие моментам времени $t = k\theta$, задаваемым тактовыми импульсами. Если состояние каждого элемента a регистра преобразуется в тест-сигнал в соответствии с соотношениями $1 \rightarrow +a$ и $0 \rightarrow -a$, то в результате получается функция $u_a(t)$. Сигналы u_b и u_y отличаются от u_a сдвигом на один и два такта соответственно. Непосредственным умножением легко вычислить значения корреляционной функции в точках $\tau = k\theta$. Можно показать, что в каждом таком интервале корреляционная функция получается линейной интерполяцией между значениями в крайних его точках. Малые отрицательные значения корреляционной функции обусловлены тем, что среди 2^n возможных состояний есть одно неиспользуемое, а именно $(0, 0, 0)$. Следовательно, при усреднении по $N = 2^n - 1$ членам последовательности остается один нескомпенсированный положительный элемент. Выходом может служить выбор немного отличающихся амплитуд положительного и отрицательного импульсов, что приводит к корреляционной функции, показанной на фиг. 10.16.

Другим аспектом, который также можно отнести к недостаткам, является отсутствие свободы при распределении мощности сигнала по частотному диапазону. Благодаря треугольной форме корреляционной функции спектр имеет огибающую вида (фиг. 10.17)

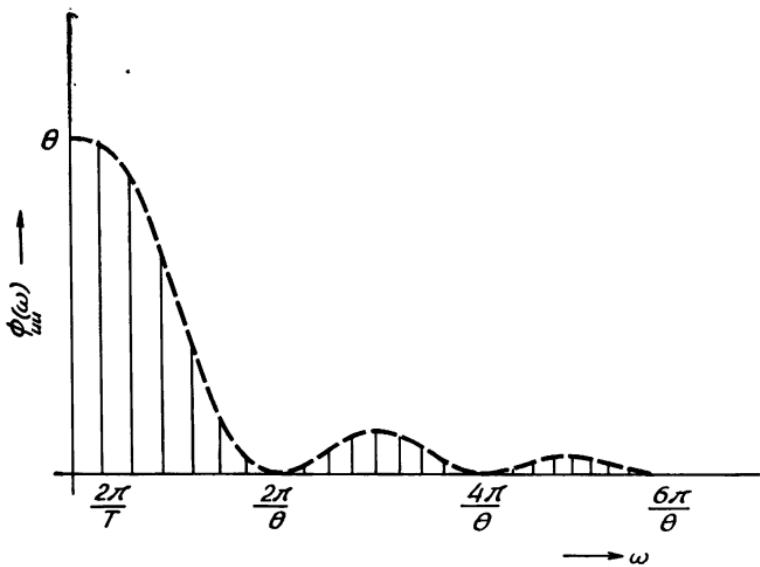
$$\theta \left(\frac{\sin \omega \theta / 2}{\omega \theta / 2} \right)^2. \quad (10.35)$$



Ф и г. 10.15.



Ф и г. 10.16.



Ф и г. 10.17.

Из-за периодичности корреляционной функции с периодом $T = N\theta$ спектр состоит из линий, отстоящих друг от друга на расстоянии $2\pi/T$. В качестве упражнения читателю предлагается определить $\Phi_{uu}(\omega)$ по $\Psi_{uu}(\tau)$ либо вычислением коэффициентов ряда Фурье периодической функции, либо исходя из формулы

$$\Psi_{uu}(\tau) = \frac{1}{\theta} f(\tau) * f(\tau) * \sum_{k=-\infty}^{\infty} \delta(\tau - kT),$$

$$f(\tau) = \begin{cases} 1, & |\tau| \leq \frac{\theta}{2}, \\ 0, & |\tau| > \frac{\theta}{2}, \end{cases}$$

где $*$ означает операцию свертки, которой соответствует умножение в частотной области. Следовательно (фиг. 10.18),

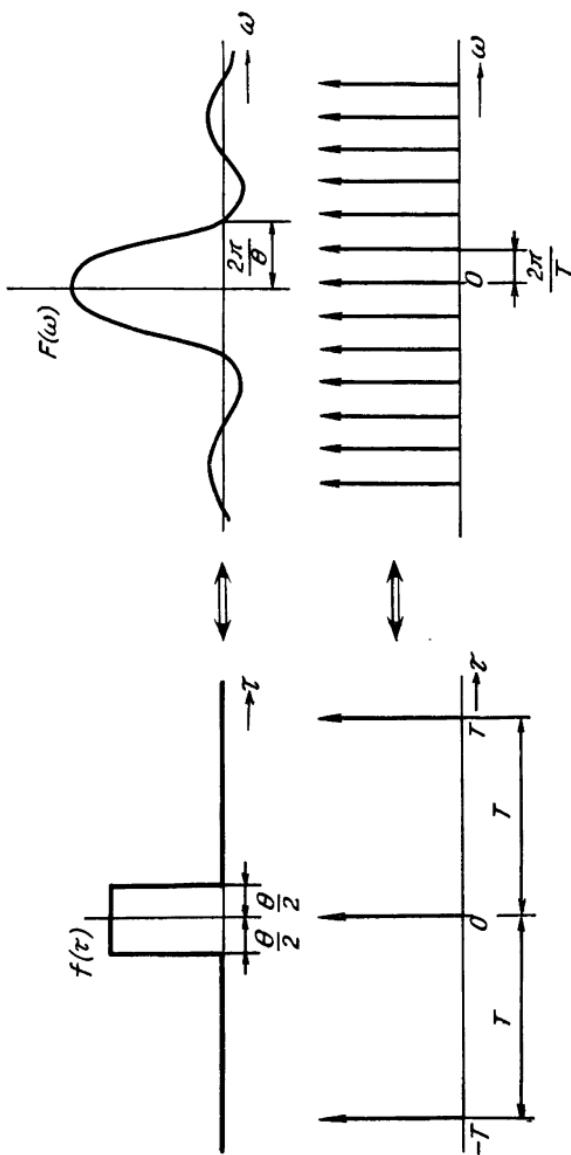
$$\Phi_{uu}(\omega) = \frac{1}{\theta} F(\omega) F(\omega) \sum_{k=-\infty}^{\infty} \delta\left(\omega - k \frac{2\pi}{T}\right). \quad (10.36)$$

Поэтому имеется большая свобода в определении полосы частот тестсигнала (выбором частоты тактовых импульсов регистра сдвига генератора m -последовательности) и частотного разрешения спектра (выбором $T = N\theta$, т. е. длины регистра сдвига и структуры обратной связи).

Использование m -последовательностей в качестве тестсигналов аналогично использованию большого числа синусоидальных компонент и измерениям на всех частотах одновременно.

Виды ошибок при использовании m -последовательностей в качестве тестсигналов

Важным аргументом в пользу выбора m -последовательностей является форма их корреляционной функции. Она может аппроксимировать δ -функцию Дирака (корреляционную функцию белого шума), что позволяет отказаться от решения уравнения свертки при отыскании значений весовой функции объекта. Будем исходить из точ-



Фиг. 10.18.

ногого соотношения между корреляционными функциями

$$\Psi_{uy}(\tau) = \int_0^{\infty} h(\theta) \Psi_{uu}(\tau - \theta) d\theta \quad (10.37)$$

и сравним результаты получающихся аппроксимаций. Можно различать следующие разновидности источников ошибок [36]:

- а) периодичность сигналов;
- б) постоянную составляющую сигналов;
- в) применение сигналов, отличных от белого шума (с ограниченным спектром);
- г) аддитивный шум;
- д) ошибки измерений и устройств сопряжения.

а) Периодичность сигналов. Поскольку m -последовательности периодичны, их корреляционные функции также обладают свойством периодичности. Поэтому при определении весовой функции объекта нужно быть уверенным, что $h(\tau) \approx 0$ при $\tau \geq T$, где T — период последовательности. В противном случае оказывается влияние дополнительных составляющих, вызванных другими пиками $\Psi_{uu}(\tau)$. Следовательно, этот метод нельзя использовать для объектов, включающих операцию интегрирования.

При использовании m -последовательностей в качестве входных сигналов объекта для оценивания по методу наименьших квадратов следует помнить, что, несмотря на их периодичность, матрица $[U'U]$ должна быть невырожденной, т. е. ее столбцы не должны быть зависимыми.

б) Постоянная составляющая сигналов. Постоянная составляющая тестсигналов приводит к появлению постоянной составляющей в корреляционной функции $\Psi_{uu}(\tau)$, а именно $-a^2/N$, которая влияет на взаимную корреляционную функцию. Для исследования этого влияния представим $\Psi_{uu}(\tau)$ в виде суммы двух частей: треугольных функций с площадью $[(N+1)/N] a^2 \theta$ и постоянного члена $-a^2/N$. Последний член дает вклад во взаимную корреляционную функцию

$$-\frac{a^2}{N} \int_0^{N\theta} h(\tau) d\tau. \quad (10.38)$$

Если $h(\tau) \rightarrow 0$ при $\tau \rightarrow T$, то эту составляющую можно выделить из $\Psi_{uy}(\tau)$ при больших положительных ($\tau \approx T$) и малых отрицательных значениях τ . Это указывает способ устранения рассматриваемой составляющей ошибки.

в) **Применение сигналов, отличных от белого шума (с ограниченным спектром).** Использование таких сигналов стимулировалось видом их корреляционной функции, которая может аппроксимировать функцию Дирака. Насколько хорошим оказывается приближение, зависит от конкретной ситуации. Поскольку «тонкая структура» передаточной функции обусловливается наименьшими постоянными времени τ_{min} исследуемого объекта, можно считать приемлемым выбор $\theta \approx \tau_{min}/5$.

Величина ошибки, конечно, зависит от особенностей реальной задачи, в частности от вида весовой функции $h(\tau)$ исследуемого объекта.

г) **Аддитивный шум.** Благодаря детерминированной природе рассматриваемых тестсигналов внутренняя статистическая неопределенность отсутствует. Исследование же неопределенности, связанной с аддитивным шумом, проводится так же, как и в предыдущих случаях. Можно показать, что дисперсия ошибки оценки, возникающей из-за аддитивного шума, равна

$$\sigma^2 = \frac{a^2 \sigma_n^2}{iN\theta}, \quad (10.39)$$

где σ_n^2 — дисперсия шума, а i — (целое) число периодов последовательности, в которых осуществляются корреляционные измерения [ср. с формулой (6.63)]. Этот случай обсуждается в [7].

д) **Ошибки измерений и устройств сопряжения.** В работе [17] исследуются ошибки, вносимые конечностью передаточных функций устройств сопряжения, используемых для подачи m -последовательностей на объекты.

Основываясь на изложенных соображениях, нужно выбрать параметры сигнала θ , N и a . Поскольку сигнал двоичный, операция умножения в таких корреляторах реализуется просто. Кроме приборов, созданных с этой целью во многих лабораториях, и небольших ЭВМ, можно использовать по крайней мере один серийный прибор такого типа (автоматический сервотестер).

Важным обобщением является использование таких тестсигналов для систем с несколькими входами и выходами. Если имеется p входов и q выходов, объект можно охарактеризовать pq весовыми функциями.

Можно показать, что корреляционные методы, описанные выше, применимы и в этом более общем случае и что весовая функция h_{ij} , соответствующая входу i и выходу j , находится из уравнения

$$\Psi_{u_i u_j}(\tau) = \int_0^\tau h_{ij}(\theta) \Psi_{u_i u_i}(\tau - \theta) d\theta \quad (10.40)$$

при условии, что u_i и остальные $p - 1$ входных сигналов статистически независимы. Если это условие не выполняется, получается система интегральных уравнений, достаточно трудная для решения. Условие независимости функций u_i можно заменить более слабым условием ортогональности на интересующем нас интервале τ (который для объектов, включающих операцию интегрирования, может оказаться бесконечным).

Можно показать, что взаимная корреляционная функция двух t -последовательностей с одинаковым θ и взаимно простыми периодами M и N может быть пренебрежимо малой для всех значений времени, если только корреляция вычисляется по интервалу $MN\theta$. Более подробная информация по использованию этих тестсигналов содержится в работах [13, 17, 25].

ЛИТЕРАТУРА

1. Anderson G. W., Buland R. N., Cooper G. R., The Aeronutronic Self-Optimizing Control System, in: Proc. of the Self-Adaptive Flight Control Systems Symp., Gregory P. C., Ed., WADC Techn. Rept. 59-49, ASTIA Document AD 209389: Wright-Patterson Air Force Base, Ohio, 1959, pp. 349-406.
2. Angström A. J., New Methods to Determine the Heat Conductivity of Bodies (на немецком), *Ann. Physik*, 114, 513-530 (1961).
3. ASA C85. 1-1963. American Standard Terminology for Automatic Control N.Y., The American Society of Mechanical Engineers, 1963.
4. Balchen J. G., A Simple Instrument for Experimental Frequency Response Tests of Control Systems, *Regelungstechnik*, 33, 200-205 (1962).
5. Balchen J. G., Blandhol E. (см. литературу к гл. 8).

6. Blackman R. B., Tukey J. W. (см. литературу к гл. 3).
7. Briggs P. A. N. et al., Correlation Analysis of Process Dynamics Using Pseudo-Random Binary Test Perturbations, *Proc. Inst. of Mech. Engrs.*, 179, 3H, 37 (1964).
8. Clark W. A., Average Response Computer (ARC-1), Quart. Progr. Rept., Research Lab. Elect., MIT, 1958, pp. 114—117.
9. Clark W. A., The Average Response Computer (ARC); a Digital Device for Computing Averages and Amplitude and Time Histograms of Electrophysiological Response, *IRE Trans. Biomed. Electron.*, BME-8, 46—51 (1961).
10. Clynes M., Kohn M., Portable 4-Channel Online Digital Average Response Computer, CAT, Proceed 4th Int'l. Conf. on Med. Electronics, N.Y., 1961.
11. Cowley P. E. A., The Application of an Analog Computer to the Measurement of Process Dynamics, *ASME Trans.*, 79, 823—832 (1957).
12. Dang Van Mien H., A Study of Some Methods of Direct Digital Control of Industrial Processes, Preprints IFAC/IFIP Symp. Digital Control of Large Industrial Systems, Toronto, June 1968.
13. Davies W. D. T. (см. литературу к гл. 2).
14. Dwyer B., A Feedback Transfer Function Analyser, *Control*, 9, 676—678 (1965).
15. Фельдбаум А. А. (см. литературу к гл. 1).
16. Fuchs A. M., A Bibliography of the Frequency Response Method as Applied to Automatic Feedback Control Systems, *ASME Trans.*, 76, 1185—1194 (1954).
17. Godfrey K. R., The Application of Pseudo Random Sequences to Industrial Processes and Nuclear Power Plant, Second IFAC Symp. «Identification and Process Parameter Estimation», Prague, Paper 7.1, 1970.
18. Hennig T., Testing for Plant Transfer Functions in Presence of Noise and Nonlinearity, *Control Engng.*, 10, 67—70, 119—124 (1963).
19. Jensen J. R., Notes on the Measurement of Dynamic Characteristics of Linear Systems, Part III, Rept. Servoteknisk Forskningslaboratorium, Denmarks Tekniske, Højscole, Copenhagen, 1959.
20. Lange F. H., Correlation Electronics (на немецком), Berlin, VEB Verlag Technik, 1962.
21. Леонов Ю. П., Липатов Л. Н., Применение статистических методов для определения характеристик объектов, *Автоматика и телемеханика*, № 9 (1959).
22. Lichtenberger W. W., A Technique of Linear System Identification Using Correlating Filters, *IRE Trans. Autom. Control*, AC-6, 183—199 (1961).
23. Mesch F., A Comparison of Frequency Response Measurement under Stochastic Disturbance (на немецком), *Zeitschr. Messen, Steuern, Regeln*, 7, 162—166 (1964).
24. MIT Communication Biophysics Group, Siebert W. M. Processing Neuroelectric Data, Tech. Rept. 351, Research Lab. Electr., MIT, Cambridge (Mass.), 1959.
25. Nikiforuk P. N., Gupta M. M., A Bibliography of the Proper-

- ties, Generation and Control Systems Application of Shift-Register Sequences, *Int. J. Control.*, 9, 217—234 (1969).
26. Peterson W. W., Error Correcting Codes, Wiley, N.Y., 1961; есть русский перевод: Питерсон У., Коды, исправляющие ошибки, изд-во «Мир», 1964.
 27. Poortvliet D. C. J., The Measurement of System-Impulse Response by Means of Cross-Correlation with Binary Signals, Techn. Rept. Electronics Lab., Univ. of Technology, Delft, Netherlands, 1962; см. также *Tijdschr. Nederl. Radio Genootschap*, 28, 253—270 (1963).
 28. Qvarnström B., On the Relative Merits of Methods to Measure Process Dynamics, *Acta Imeko*, Budapest, 14-SW-126, 1961.
 29. Qvarnström B., An Uncertainty Relation for Linear Mathematical Models, Second IFAC Congress, Basle; есть русский перевод: Кварнстрём Б., Неопределенность в линейных математических моделях, в книге «Теория непрерывных автоматических систем», Труды II Межд. конгресса ИФАК, изд-во «Наука», 1965.
 30. Schwarze G., Algorithmic Determination of the Order and Time Constants of P-, I- and D-Action Elements with two Different Time Constants and Delays up to the 6th Order (на немецком), *Zeitschr. Messen, Steuern, Regeln*, 7, 10—18 (1964).
 31. Seifert W., Commercial Frequency-Response Measuring Devices (на немецком), *Regelungstechnik*, 10, 350—353 (1962).
 32. Sheridan T. B., Experimental Analysis of Time-Variation of the Human Operator's Transfer Function, First IFAC Congress, Moscow, 1960, pp. 629—635; есть русский перевод: Шеридан Т.Б., Экспериментальное исследование изменений передаточной функции человека-оператора во времени, в книге «Теория дискретных, оптимальных и самонастраивающихся систем», Труды I Межд. конгресса ИФАК, Изд-во АН СССР, 1961.
 33. Strobel H., System Analysis Using Deterministic Test Signals (на немецком), VEB Verlag Technik, Berlin, 1968.
 34. Sins A. W., The Determination of a System Transfer Function in Presence of Output Noise (на голландском), Thesis, E.E. Dept., Univ. of Technology, Eindhoven, Netherlands, 1967.
 35. Werner G. W., Frequency Response Measurement with Low Experimental Expenditure in Disturbed Industrial Control Circuits (на немецком), *Zeitschr. Messen, Steuern, Regeln*, 5, 545—546 (1962).
 36. While C., System Identification, M. Sc. Thesis, Dept. of Electronic and Electrical Engng., Univ. of Birmingham, B'ham, England, 1968.
 37. Woodward P. M., Probability and Information Theory with Application to Radar, Pergamon, London, 1964.

ДОПОЛНИТЕЛЬНАЯ ЛИТЕРАТУРА

10.1. Импульсные и ступенчатые сигналы

Adams R. K., Digital Computer Analysis of Closed-Loop Systems Using Number Series Approach, Joint Autom. Control Conf., 1961.

- Billingsley D. S., Rekoff M. G., Evaluation of Transient Coefficients, *IRE Trans Autom. Control*, AC-6, 80—82 (1961).
- Bux D., Isermann R., A Comparison of Aperiodical Test Signals for the Measurement of the Dynamic Behaviour of Control Loops (на немецком), *Fortschritt-Bericht VDI-Zeitschrift*, 1967.
- Cuenod M., Durling A. E., A Discrete-Time Approach for System-Analysis, Academic Press, N.Y., 1969.
- Dittman H., Some Extensions in Determining Characteristic Values from the Transient Response (на немецком), *Zeitschr. Messen. Steuern. Regeln*, 6, 139—141 (1963).
- Dreike G. L., Hougen J. O., Experimental Determination of System Dynamics by Pulse Methods, *Joint Autom. Control Conf.* 1963, pp. 608—633.
- Ellington J. P., McCallion H., The Determination of Control System Characteristics from a Transient Response, *Proc. IEE*, 105C, 370—373 (1958).
- Fuchs H., On the Evaluation of Characteristic Values of PID-Units (на немецком), *Zeitschr. Messen. Steuern. Regeln*, 6, 385—386 (1963).
- Gehrman E., Schusta J., Method for Determining the Time Constants of Thermocouples (на немецком), *Wiss. Zeitschr. Hochsch. Elektrotech., Ilmenau*, DDR, 8, 419—421 (1962).
- Gerhardt G., Experimental Parameter Estimation in the Time Domain Using a-Periodic Test Signals (на немецком), *Zeitschr. Messen. Steuern. Regeln*, 9, 204—208 (1966).
- Гиршов Л. А., Гельман И. В., Добрин Л. А., Некоторые инженерные методы анализа объектов регулирования с монотонными переходными функциями. *Автоматика и телемеханика*, № 9 (1962).
- Gupta S. C., Performance of Process Models, Third IFAC Congress, London, Paper 36D, 1966; есть русский перевод: Гупта С., Качество моделей объекта, в книге «Теория непрерывных автоматических систем и вопросы идентификации», Труды III Межд. конгресса ИФАК, изд-во «Наука», 1971.
- Hilschner K., Parameter Estimation of Instruments and System from the Transient Response that can be Approximated by n Equal Time-Constants and a Dead Time (на немецком), *Zeitschr. Messen. Steuern. Regeln*, 7, 272—273 (1964).
- Hougen J. O., Walsh R. A., Pulse Testing Method, *Chemical Engineering Progress*, 57, 69—79 (1961).
- John R. M. et al., System for Processing and Recording Excitation Function Data, *Review of Scientific Instruments*, 33, 1089—1094 (1962).
- Lendaris G. G., The Identification of Linear Systems, *AIEE Trans.*, 81, Part II (Applic. and Ind., № 62), 231—242 (1962).
- Mishkin E., Haddad R. A., Identification and Command Problems in Adaptive Systems, *IRE Wescon Conv. Rec.*, Part 4, 125—135 (1959); *IRE Trans. Autom. Control*, AC-4, 121—131 (1959).
- Nyquist J. K., Schindler R. N., Gilbert R. E., Determination of Frequency Response from Step Response, *Chemical Engineering Progress, Symposium Series*, 59, 46 (1963).

Ormans G., On the Processing of Experimentally Found Transformations of Control Blocks (на немецком), Paper «Moderne Theorien und Ihre Verwendbarkeit», Heidelberg, 1956, pp. 130—133.

Radtke M., Seidel H. P., On Estimation of Static and Dynamic Parameters of Nonlinear Transfer Elements from Step Responses, IFAC Symp. «Identification in Autom. Control Systems», Prague, Paper 4.3, 1967.

Reinisch K., Use of Control Loop Simulator in Order to Obtain Simple Dimensioning Rules for Controlled Plants to Find the Characteristic Values of Controlled Plants (на немецком), *Zeitschr. Messen, Steuern, Regeln*, 5, 245—251 (1962).

Reinisch K., Wernstedt J., Model Method of Analyzing Distributed Linear and Special Nonlinear Systems by Means of Aperiodic Test Signals, Paper IFAC Symp. «Identification in Autom. Control Systems», Prague, Paper 4.2, 1967.

Sage A. P., Choate W. G., Minimum Time Identification of Nonstationary Dynamic Processes, *Proc. Nat. Electron. Conf.*, 21, 587—592 (1965).

Schechter R. S., Wissler E. H., Frequency Response from Step Input Response, *Industrial and Engineering Chemistry*, 51, 945—948 (1959).

Stepan J., Sensitivity Studies of Some Methods for Parameter Estimation of Control Loops (на немецком), *Zeitschr. Messen, Steuern, Regeln*, 8, 35—38 (1965).

Strejc V., Approximation Process for Aperiodic Transfer Characteristics (на немецком), *Regelungstechnik*, 7, 124—128 (1959).

Strejc V., The Approximation of Aperiodic Transient Responses (на немецком), *Zeitschr. Messen. Steuern, Regeln*, 3, 115—124 (1960).

Strobel H., On the Limits Imposed by Random Noise and Measurement Errors upon System Identification in the Time Domain, IFAC Symp. «Identification in Autom. Control Systems», Prague, Paper 4.4, 1967.

Turin G. L., On the Estimation in the Presence of Noise of the Impulse Response of a Random Linear Filter, *IRE Trans. Inform. Theory*, IT-3, 5—10 (1957).

Unbehauen H., A Graphical-Analytical Computation Scheme for the Determination of the Frequency Response from the Impulse Response (на немецком), *Regelungstechnik*, 11, 551—555 (1963).

Unbehauen H., Parameter Estimation of Control System Using the Measured Impulse Response (на немецком), *Zeitschr. Messen, Steuern, Regeln*, 9, 188—191 (1966).

Verheyden J., Measurement of Controller Characteristics, *Revue A*, 4, 94—99 (1962).

Weller W., The Methods of Determining Control Characteristics from the Measured Transient Response (на немецком), *Zeitschr. Messen, Steuern, Regeln*, 5, 355—363 (1962).

Werner G. W., Application of the Impulse Test Procedure for Parameter Estimation (на немецком), *Zeitschr. Messen, Steuern, Regeln*, 9, 197—201 (1966).

Zadeh L. A., The Determination of Impulse Response of Variable Networks, *J. Appl. Phys.*, 21, 642 (1950).

10.2. Синусоидальные тестсигналы

Атаманенко В. Г., Анализатор передаточной функции для экспериментального исследования динамических характеристик промышленных САР, *Приборостроение*, № 5 (1962).

Atherton D. P., Jain N. K., Identification of Double-valued Nonlinearities from Their Harmonic Response, *Electronics Letters*, 3, 485—486 (1967).

Bard M., Implementation for Automatic Registration of Frequency Plots with the Analog Computer, *Zeitschr. Steuern, Regeln, Automatisieren*, 9, 431—437 (1965).

Chen C. F., Philip B. L., Accurate Determination of Complex-Root Transfer Functions from Frequency Response Data (Short Paper), *IEEE Trans. Autom. Control*, AC-10, 356—358 (1965).

Даровских Л. Н., Экспериментальное определение передаточных функций звеньев систем автоматического регулирования при помощи типовых электронных моделей, *Автоматика и телемеханика*, № 9, (1959).

Даровских Л. Н., Определение частотных характеристик при помощи решающих блоков электронной модели, *Автоматика и телемеханика*, № 2 (1962).

Dudnikov E. E., On a Method for the Evaluation of Transfer Function Coefficients of Linear Systems from the Measured Frequency Response (на немецком), *Zeitschr. Messen, Steuern, Regeln*, 6, 453—456 (1963).

Ehret R. J., An Automatic Transfer Function Measuring and Recording System, *AIEE Trans.*, 72, Part I, 664—669 (1953).

Fraser H. J., Reece W. V. P., An Automatic Electronic Nyquist Plotter, *Proc. IEE*, 108B, 535—538 (1961).

Gochet M., Device for Measuring Isochrone Transmittance of Servosystems (на французском), *Revue A*, 3, 32—35 (1961).

Huber E. A., A Technique for the Adaptive Control of High Order Systems, *IRE Trans. Autom. Control*, AC-7, 22—29 (1962).

Hughes M. T. G., May R., Estimation of Frequency Response from Random Data, IFAC Symp. «Identification in Autom. Control Systems», Prague, Paper 6.4, 1967.

Кардашов А. А., Карюшин Л. В., Экспериментальное определение частотных характеристик автоматических систем, *Автоматика и телемеханика*, № 4 (1958).

Key K. A., The Analogue Computer as a (One Cycle) Transfer Function Analyser, Proc. Third Intern. Analogue Computation Meetings, Brussels, Presses Académiques Européennes, 1962, pp. 569—572.

Klein V., The Identification of Linear System with a Given Structure by Means of Frequency Responses, Third IFAC Congress, London, Paper 46A, 1966; есть русский перевод: Клейн В., Иден-

тификация линейной системы с заданной структурой с помощью частотных характеристик, в книге «Теория непрерывных автоматических систем и вопросы идентификации», Труды III Межд. конгресса ИФАК, изд-во «Наука», 1971.

Lees S. et al., Experimental Verification of Design Basis for Positional Servomechanisms, First IFAC Congress, Moscow, 2, 998—1008, 1961; есть русский перевод: Лиз С., Блошке Т., Браун Б., Остертаг К., Экспериментальное исследование принципиальных структурных схем позиционных следящих систем, в книге «Технические средства автоматики», Труды I Межд. конгресса ИФАК, Изд-во АН СССР, 1961.

Manton F. E. S., Analog Computer Setup Displays Frequency Response, *Control Engng.*, 9, 83—85 (1962).

Martin E., Instrument for Automatic Recording of the Logarithmic Amplitude and Phase Frequency Response (Bode Diagram) (на немецком), *Zeitschr. Messen, Steuern, Regeln*, 5, 1—6 (1962).

Minkwitz G., Zur Anwendung der Digitalen Technik bei der Phasenwinkelmessung, *Nachrichtentechnik*, 16, 97 (1966).

Nomoto A. et al., Automatisation of Dynamic Testing, First IFAC Congress, Moscow, 3, 304—309 (1960); есть русский перевод: Номото А. и др., Автоматизация исследования динамических характеристик, в книге «Технические средства автоматики», Труды I Межд. конгресса ИФАК, Изд-во АН СССР, 1961.

Ogar G. W., Obtaining the Frequency Response of Physical Systems by Analog Computer Techniques, IRE Intern. Conv. Rec., Part 2: Audio, Electronic Computers, 1961, pp. 196—210.

Oppelt W., Kleines Handbuch technischer Regelvorgänge, Verlag Chemie, Weinheim, 1972.

Parezanovic N., Frequency Characteristic Determination from the System Pulse Response by the Use of Repetitive Differential Analyser, Proc. Third Intern. Analogue Computation Meeting, Brussels, Presses Académiques Européennes, 1962, 573—575.

Reinisch K., The Use of a Controlloop Model for Obtaining Simple Measurement Rates for Linear Control Loops and for Parameter Estimation on Control Systems (на немецком), *Zeitschr. Messen, Steuern, Regeln*, 5, 245—251 (1962).

Reinisch K., A Method for the Measurement of Continuous and Sampled Control Loops Including Dead Time and Some Nonlinearities (на немецком), Habilitationsschrift, Technische Univ. Dresden, 1965.

Sanathanan C. K., Koerner J., Transfer Function Synthesis as a Ratio of Two Complex Polynomials, *IEEE Trans Autom. Control*, 8, 56—58 (1963).

Schäfer O., Feissel W., An Improved Method for Frequency Response Analysis of Industrial Control Loops (на немецком), *Regelungstechnik*, 3, 225—229 (1955).

Senf B., Strobel H., Method for Determining Transfer Functions of Linear Systems from Measured Values of Frequency Response (на немецком), *Zeitschr. Messen, Steuern, Regeln*, 4, 411—420 (1961).

Sourisse C., Wagner R., Pneumatic Transferometer (на французском), *Automatisme*, 8, 487—494 (1963).

Staffin H. K., Staffin R., Approximating Transfer Functions from Frequency Response Data, *Instruments and Control Systems*, 38, 137—144 (1965).

Strobel H., On the Problem of the Essential Frequency Range when Approximating a Measured Frequency Response (на немецком), *Zeitschr. Messen, Steuern, Regeln*, 7, 19—23 (1964).

Strobel H., On a New Method of Determining the Transfer Function by Simultaneous Evaluation of the Real and Imaginary Parts of the Measured Frequency Response, Third IFAC Congress, London, Paper 1F, 1966; есть русский перевод: в книге «Теория непрерывных автоматических систем и вопросы идентификации», Труды III Межд. конгресса ИФАК, изд-во «Наука», 1971.

Unbehauen R., The Determination of Rational Frequency Responses from Measurement Data (на немецком), *Regelungstechnik*, 14, 268—273 (1966).

Вавилов А. А., Солодовников А. И., Экспериментальное определение частотных характеристик автоматических систем, Госэнергоиздат, М.—Л., 1963.

Vichnevetsky R., Analogue Computer Simulation of a Time-Dependent Delay Using the Concept of Generalized Transfer Function, *Annls. Ass. Int. Calcul. Analog.*, 6, 105—109 (1964).

Webb E. K., Bacon N. E., A Mechanical Harmonic Analyzer, *J. of Scientific Instruments*, 39, 500—503 (1962).

Welfonder E., Parameter Estimation on Disturbed Control Systems by Means of Correlation and Periodic Testsignals (на немецком), *Fortschritt-Bericht VDI-Z.*, 8, VDI-Verlag, Düsseldorf (1966).

Welfonder E., Correlation Method for Identification on Disturbed Control Systems by Periodical Test Signal, IFAC Symp. «Identification in Autom. Control System», Prague, Paper 3.8, 1967.

Weygandt C. N., Puri N. N., Multivariable Adaptive Control System, *IEEE Trans. Applic. and Ind.*, 82, 89—94 (1963).

Weygandt C. N., Puri N. N., Transfer Function Tracking and Adaptive Control Systems, *IRE Trans. Autom. Control*, AC-6, 162—166 (1961).

Wierwille W. W., Determination by Frequency Response of Transfer Function, *AIEE. Trans*, Part 2 (Applic. and Ind., № 56), 183—189 (1961).

Wilfert H. H., Signal- and Frequency Response Analysis on Strongly-Disturbed Systems (на немецком), VEB Verlag Technik, Berlin, 1969.

Zoss L. M., Wilson H. S., Frequency Response Analysis, *Instrument Society of America Journal*, 2 (1962).

10.3. Двоичные тестсигналы

Barber D. L. A., Hammond P. H., The Use of an One-Line Digital Computer to Evaluate the Dynamic Response of a Pilot Scale Distillation Column, *Trans. Soc. Instrum. Technol.*, 17 (3), 59 (1965).

Bell D. A., Walsh Functions and Hadamard Matrices, *Electronics Letters*, 2, 340 (1966).

Boardman K. D., The Measurement of the Dynamic Behaviour of a Superheater by Random Signal Testing, IEE Conference on Automatic Control in Electricity Supply, Manchester, 1966, p. 198.

Briggs P. A. N., Godfrey K. R., Pseudo-Random Signals for the Dynamic Analysis of Multivariable Systems, *Proc. IEE*, **113**, 1259—1267 (1966).

Briggs P. A. N., Godfrey K. R., Hammond P. H., Estimation of Process Dynamic Characteristics by Correlation Methods Using Pseudo-Random Signals, IFAC Symp. «Identification in Autom. Control Systems», Prague, Paper 3.10, 1967.

Carter G. A., Shepley J. A., Roust C. B., A Cross-Correlation Experiment on an Experimental Reactor Using the Method of Binary Input and Output Sequences, IEE Colloquium on «Pseudo-Random Sequences Applied to Control Systems», London, 1967.

Chow P. E. K., Davies A. C., The Synthesis of Cyclic Code Generators, *Electronic Technology*, **41** (1969).

Clarke D. W., Godfrey K. R., Simultaneous Estimation of the First and Second Derivatives of a Cost Function, *Electronics Letters*, **2**, 338 (1966).

Cumming I. G., Frequency of Input Signal in Identification, Second IFAC Symp. «Identification and Process Parameter Estimation», Prague, Paper 7.8, 1970.

Darnell M., Synthesis of Pseudo-Random Signals Derived from p-Level m-Sequences, *Electronics Letters*, **2**, 428 (1966).

Davies A. C., Delayed Versions of Maximal-Length Linear Binary Sequences, *Electronics Letters*, **1**, 61 (1965); Further Notes on Delayed Versions of Linear Binary Sequences, *Electronics Letters*, **1**, 190 (1965).

Davies W. A., Generation of Delayed Replicas of Maximal Length Linear Binary Sequences, *Proc. IEEE*, **113**, 295—296 (1966).

Davies W. D. T., Generation and Properties of Maximum-Length Sequences 1, 2, 3, *Control and Automatic Progress*, **10**, 302 (1966).

Davies W. D. T., Identification of a System in the Presence of Low-Frequency Drift, *Electronics Letters*, **2**, 327 (1966).

Davies W. D. T., Douce J. L., On-Line System Identification in the Presence of Drift, IFAC Symp. «Identification in Autom. Control Systems», Prague, Paper 3.12, 1967.

Douce J. L., Ng K. C., The Use of Pseudo-Random Binary Test Signals in Adaptive Control, IFAC Symp. «Theory of Self-Adaptive Control Systems», Plenum Press, Teddington, 1966.

Douce J. L., Ng K. C., Walker A. E. G., System Identification in the Presence of a Ramp Disturbance, *Electronics Letters*, **2**, 243 (1966).

Elispas B., The Theory of Autonomous Linear Sequential Networks, *IRE Trans. Circuit Theory*, CT-6, 45—60 (1959).

Everett D., Periodic Digital Sequences with Pseudo-Noise Properties, *GEC Journal of Science and Technology*, England, **33**, 115 (1966).

Gardiner A. B., Logic Pseudo-Random Binary Sequence Delay Calculator and Delayed Version Generator with Automatic Delay Changing Facility, *Electronics Letters*, **1**, 123 (1965).

Gardiner A. B., Elimination of the Effect of Non-Linearities on Process Crosscorrelations, *Electronics Letters*, **2**, 164 (1966).

- Godfrey K. R., Three-Level m-Sequences, *Electronics Letters*, 2, 241 (1966).
- Godfrey K. R., Input-Transducer Errors in Binary Crosscorrelation Experiments — 3, *Proc. IEE*, 113, 1095—1102 (1966).
- Godfrey K. R., Murgatroyd M., Input-Transducer Errors in Binary Crosscorrelation Experiments — 1, *Proc. IEE*, 112, 565—573 (1965).
- Godfrey K. R., Theory and Application of Pseudo-Random Sequences, *Control and Automation Progress*, 10, 305 (1966).
- Godfrey K. R., Everett B., Bryant P. R., Input-Transducer Errors in Binary Crosscorrelation Experiments — 2, *Proc. IEE*, 113, 185—189 (1966).
- Havel J., Janáč K., Equipment for the Generation of Random and Pseudo-Random Processes, IFAC Symp., «Identification in Autom. Control Systems», Prague, Paper 6.3, 1967.
- Hazlerigg A. D. G., Noton A. R. M., The Application of Cross-correlating Equipment to Linear-System Identification, *Proc. IEE*, 112, 2385—2400 (1965).
- Hammond P. H., Barber D. L. A., The Use of an On-Line Digital Computer to Evaluate the Dynamic Response of a Pilot Scale Distillation Column, *Trans. Soc. Inst. Techn.*, 59 (1965).
- Hampton R. L., A Hybrid Analog-Digital, Pseudo-Random Noise Generator, *Simulation*, 4, 179—190 (1965).
- Henderson K. W., Some Notes on the Walsh Functions, *Trans. IEEE Electronic Computers*, EC-13, 50 (1964).
- Hill J. D., McMurtry G. J., An Application of Digital Computers to Linear System Identification, *IEEE Trans. Autom. Control*, AC-9, 536—538 (1964).
- Isobe T., Idogawa T., Kashiwagi H., Impulse Response Determination with an On-line Cross-Correlator, Third IFAC Congress, London, Paper 15D, 1966; есть русский перевод: Изобе Т. и др., Определение импульсных характеристик с помощью коррелятора, подключаемого непосредственно к объекту, в книге «Теория непрерывных автоматических систем и вопросы идентификации», Труды III Межд. конгресса ИФАК, изд-во «Наука», 1971.
- Izawa K., Furuta K., Apparatus to Measure Dynamic Characteristics, Proc. Third IFAC Congress, London, Paper 15C, 1966; есть русский перевод: Изава К., Фурута К., Аппаратура для измерения динамических характеристик, в книге «Теория непрерывных автоматических систем и вопросы идентификации», Труды III Межд. конгресса ИФАК, изд-во «Наука», 1971.
- Izawa K., Furuta K., Measurement of Dynamic Characteristics of Heat Exchanger Using Binary Random Signal, *Proc. Soc. of Instrument and Control Engng.*, 2, 101—112 (1966).
- Izawa K., Furuta K. (см. литературу к гл. 4).
- Kramer C., A Low Frequency Pseudo-Random Generator, *Electronic Engineering*, 37, 465—467 (1965).
- Krtolica R., Some Applications of Pseudo-Random Maximum Length Binary Sequences in the Statistical Identification, IFAC Symp. «Identification in Autom. Control Systems», Prague, Paper 3.11, 1967.

- Pitterman F., Schweizer G., Generation and Use of Binary Noise in Flight Tests (на немецком), *Regelungstechnik*, 14, 63—70 (1966).
- Rhihaczek A. W., Radar Signal Design for Target Resolution, *Proc. IEEE*, 53, 116—128 (1965).
- Roberts P. D., Davis R. H., Statistical Properties of Smoothed Maximal-Length Linear Binary Sequences, *Proc. IEE*, 113, 190—196 (1966).
- Sawaragi Y., Ogino K., Sensitivity Approach to Optimal Input Synthesis for Parameter Identification of Bilinear Systems. Fifth IFAC Congress, Paris, Paper 31.2, 1972.
- Scholefield, Shift Registers Generating Maximum Length Sequences, *Electronic Technology*, 37, 389 (1960).
- Schwarze G., Error Discussions for the Time-Percentage Parameter Method when Using Non-Ideal Unit Testsignals (на немецком), *Zeitschr. Messen, Steuern, Regeln*, 9, 171—173 (1966).
- Simpson H. R., Statistical Properties of a Class of Pseudo-Random Sequences, *Proc. IEE*, 113, 2075—2080 (1966).
- Smith B. M., An Asymmetrical Property of Binary Pseudo-Random Noise Generators (Correspondence), *Proc. IEEE*, 54, 793—794 (1966).
- Smith K. D., Hamilton J. C., The Logical Design of a Digital Pseudo-Random Noise Generator, *IEE Trans. Nuclear Science*, 13, 371—378 (1966).
- Stalder J. E., Cahn C. R., Bounds for Correlation Peaks of Periodic Digital Sequences (Correspondence), *Proc. IEEE*, 52, 1262—1263 (1964).
- Tausworthe R. C., Random Numbers Generated by Linear Recurrence Modulo Two, *Math. of Computation*, 19, 201—209 (1965).
- Turin G. L., An Introduction to Matched Filters, *IRE Trans. Inform. Theory*, IT-6, 311 (1960).
- Unbehauen H., Schlegel G., Estimation of the Accuracy in the Identification of Control Systems Using Deterministic Test Signals, IFAC Symp. «Identification in Autom. Control Systems», Prague, Paper 1.12, 1967.
- Van Blerkom R., Sears R. E., Freeman D. G., Analysis and Simulation of a Digital Matched Filter Receiver of Pseudo-Noise Signals, *IBM J. of Research and Development*, 9, 264—273 (1965).
- Van den Bos A., Construction of Binary Multifrequency Test Signals, IFAC Symp. «Identification in Autom. Control System», Prague, Paper 4.6, 1967.
- Van den Bos A., Estimation of Linear System Coefficients from Noisy Responses to Binary Multi-Frequency Test Signals, Second IFAC Symp. «Identification and Process Parameter Estimation», Paper 7.2, 1970.
- Van der Grinten P. M., The Application of Random Test Signals in Process Optimization, Second IFAC Congress, Basle, 1963; есть русский перевод: Ван-дер-Гринтен П., Применение случайных тестсигналов в задачах оптимизации, в книге «Оптимальные системы. Статистические методы», Труды II Межд. конгресса ИФАК, изд-во «Наука», 1965.
- Zierler N., Linear Recurring Sequences, *SIAM J.*, 7, 31 (1960).

ГЛАВА 11

БАЙЕСОВСКИЕ ОЦЕНКИ И ОЦЕНКИ ПО МЕТОДУ МАКСИМАЛЬНОГО ПРАВДОПОДОБИЯ

В гл. 5 отмечалась связь между разными типами оценок. Было показано, что марковские оценки и оценки по методу наименьших квадратов обладают тем преимуществом, что не требуют или требуют мало априорной информации. Было также установлено, что в случае гауссовского аддитивного шума эти оценки эквивалентны оценкам по методу максимума правдоподобия. Однако в зависимости от точки зрения достоинства метода наименьших квадратов могут превратиться в недостатки, если имеется априорная информация. Интуитивно ясно, что использование подходящей априорной информации позволяет улучшить оценку.

При использовании метода наименьших квадратов в настраиваемых моделях можно устанавливать начальные значения параметров в соответствии с нашими априорными знаниями или предположениями. При этом, однако, не удается достаточно просто ввести в схему оценивания степень достоверности этих начальных значений. Именно здесь оказывается полезным байесовский подход к оцениванию. Эти вопросы рассмотрены в разд. 11.1.

Разд. 11.2 посвящен оцениванию по методу максимального правдоподобия, теоретико-статистические основы которого хорошо развиты. Весьма важной является проблема максимальной достижимой точности. Она также рассматривается в разд. 11.2. Некоторые схемы практической реализации оценок этих типов обсуждаются в разд. 11.3. Требования к входным сигналам объекта рассмотрены в разд. 11.4,

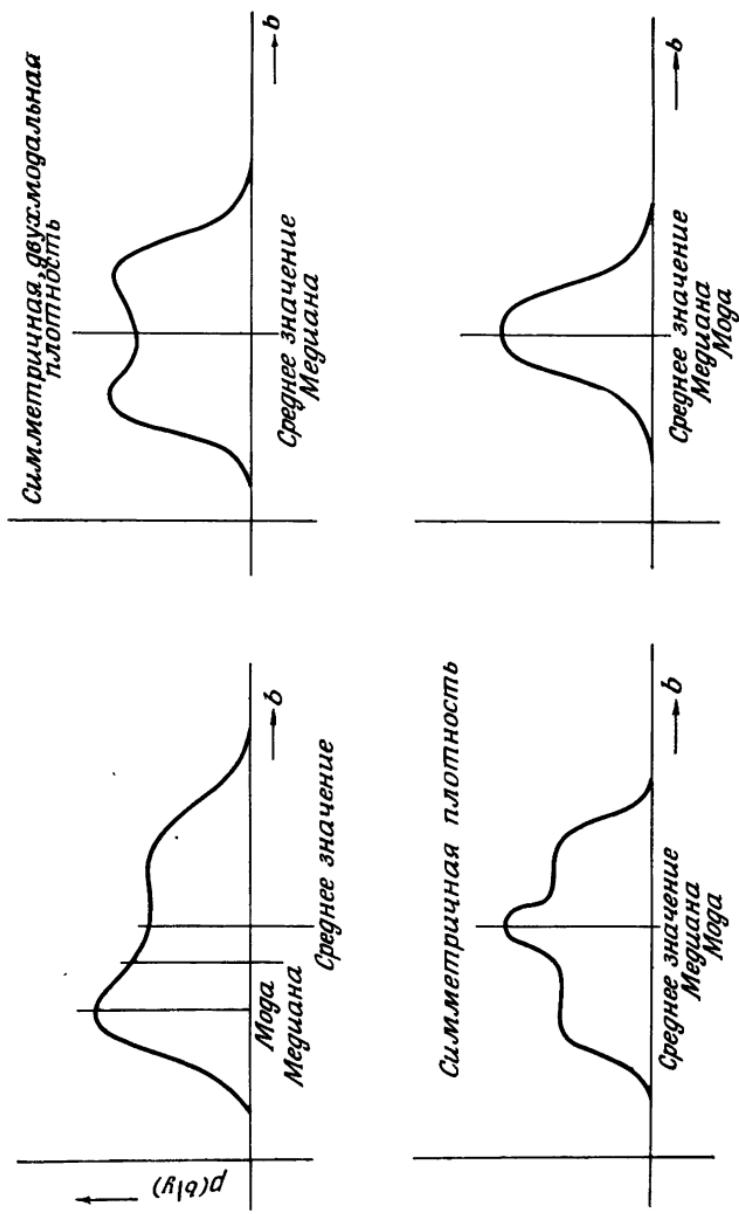
11.1. БАЙЕСОВСКИЕ ОЦЕНКИ.

Основные аспекты байесовского подхода обсуждались в разд. 5.1. Было показано, что существенным моментом этой процедуры оценивания является определение апостериорной плотности условного распределения вероятностей $p(\mathbf{b} | \mathbf{y})$ параметров относительно наблюдений \mathbf{y} . Для каждой конкретной последовательности наблюдений выходной переменной объекта $\mathbf{y}_k = \{y(1), \dots, y(k)\}$ в этой плотности заключена вся информация, представляющая интерес для экспериментатора. Из выражения для $p(\mathbf{b} | \mathbf{y}_k)$ при возрастающем k можно определить такие характеристики оценки, как смещение, состоятельность, асимптотическая нормальность (см. фиг. 5.2). Определив $p(\mathbf{b} | \mathbf{y})$, можно на основе этой информации предоставить экспериментатору возможность решать, какую оценку он считает наилучшей оценкой $\hat{\beta}$ параметра \mathbf{b} . Ряд попыток определения зависимости плотности условного распределения вероятностей от времени описан в литературе. Возможными подходами являются решение дифференциальных уравнений в частных производных типа Фоккера — Планка [3, 10, 25] и разложения выходного сигнала объекта по собственным функциям [24].

Возражения против подобной процедуры первоначального определения $p(\mathbf{b} | \mathbf{y})$ и предоставления затем возможности экспериментатору решать, какую оценку считать наилучшей, могут основываться на следующих соображениях. В ситуации, когда β является m -вектором (при $m > 2$), задача представления $p(\mathbf{b} | \mathbf{y})$ в виде, удобном для принятия решений, оказывается совсем не простой. Далее, понятие наилучшей оценки должно определяться тем или иным способом, опирающимся на выбор функции штрафа или потерь. Этот выбор остается более или менее произвольным; только в редких случаях он диктуется уже самой постановкой задачи. В литературе рассматриваются, в частности, следующие функции штрафа, связанные с плотностями условного распределения:

1) Квадратичная функция $C(\beta, b) = (\beta - b)^2 c$, приводящая к условному математическому ожиданию $p(\mathbf{b} | \mathbf{y})$;

Фиг. 11.1.



например, в одномерном случае из условия

$$\min \int c(\beta - b)^2 p(b | y) db,$$

приравнивая нулю производную по β , получаем

$$\hat{\beta} = \int bp(b | y) db,$$

т. е. среднее значение плотности условного распределения.

2) Абсолютная величина отклонения $C(\beta, b) = |\beta - b| c$; соответствующей оценкой является медиана $p(b | y)$.

3) Наиболее вероятное значение $\beta = \max_b p(b | y)$,

т. е. мода условного распределения.

Фиг. 11.1 иллюстрирует эти понятия для разных типов плотностей распределения. На фиг. 11.2 приведены некоторые типы функций штрафа. Для симметричных унимодальных распределений среднее значение, медиана и мода совпадают. Дальнейшее обсуждение вопроса выбора функций штрафа можно найти в работе [14].

В разд. 5.1 было показано, как зависит байесовская оценка $\hat{\beta}$ от плотности условного распределения вероятностей выходной переменной $p(y | b)$, априорной плотности вероятности параметра $q(b)$ и функции штрафа (потерь) $C(\beta, b)$ через подлежащий минимизации средний риск

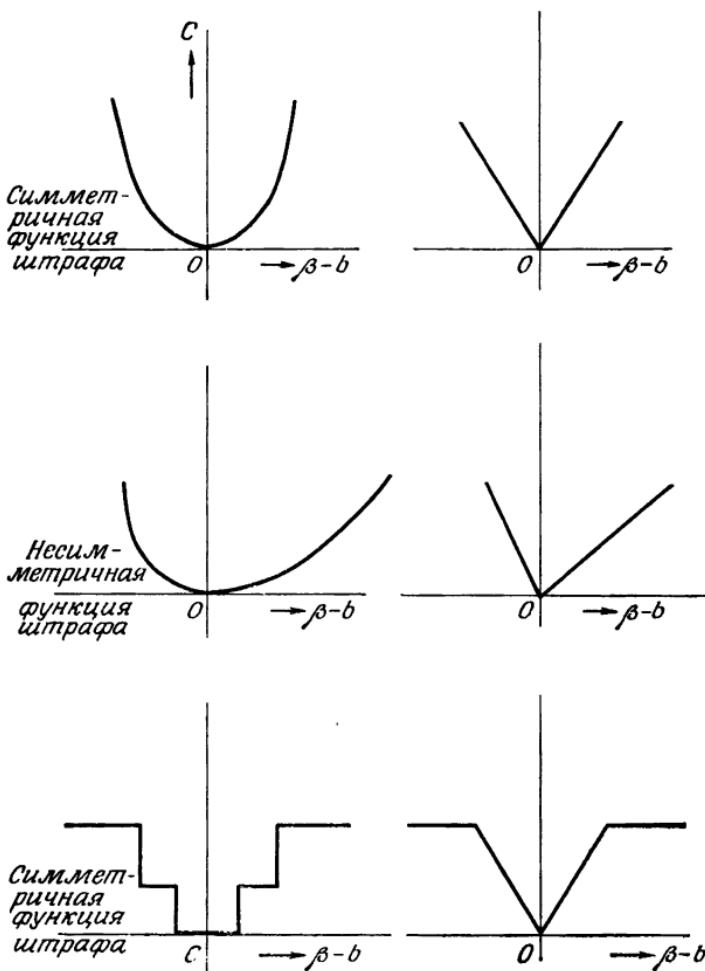
$$R(\beta) = \mathbb{E}_b [\mathbb{E}_y [C(\beta, b) | b]] = \\ = \int_m \int_k C(\beta, b) p(y, b) d^k y d^m b. \quad (11.1)$$

При заданном наблюдении y задача сводится к определению

$$\min_{\beta} \int_m C(\beta, b) p(b | y) d^m b, \quad (11.2)$$

необходимым условием которого является равенство

$$\frac{\partial}{\partial \beta} \int_m C(\beta, b) p(b | y) d^m b = 0. \quad (11.3)$$



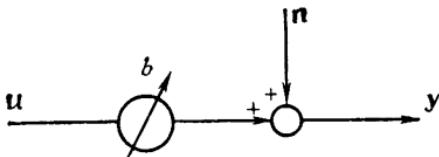
Ф и г. 11.2.

В случае скалярного параметра $\beta = \beta$ и $C(\beta, b) = (\beta - b)^2$ с получаем

$$\hat{\beta} = \int b p(b | \mathbf{y}) db. \quad (11.4)$$

Для пояснения изложенных идей рассмотрим задачу оценивания одного параметра в простой задаче, иллюстрированной на фиг. 11.3:

$$\mathbf{y} = \mathbf{u}\beta + \mathbf{u},$$



Ф и г. 11.3.

где плотности $p_b(b)$ и $p_n(n)$ — гауссовские, $E[b] = \mu_b$, $\text{var}[b] = \sigma_b^2$, $E[n] = 0$, $\text{cov}[n] = \sigma_n^2 I$, а b и n независимы. На основе этой информации легко вычислить плотность вероятности y , а именно $p_y(y)$ — гауссовская плотность, $E[y] = \mu_y$, $\text{cov}[y] = \sigma_y^2 uu' + \sigma_n^2 I$, причем формула для ковариации следует из соотношения

$$\begin{aligned} \text{cov}[y] &= E[\{y - \mu_y\}\{y - \mu_y\}'] = \\ &= E[\{u(b - \mu_b) + n\}\{u(b - \mu_b) + n\}'] = \\ &= E[(b - \mu_b)^2]uu' + E[nn']. \end{aligned} \quad (11.5)$$

Основной интерес для нас представляет плотность

$$p(b|y) = \frac{p(y, b)}{p_y(y)} = \frac{p_b(b)p_n(y - ub)}{p_y(y)}, \quad (11.6)$$

поскольку при $n = y - ub$

$$p(y, b) = p(y - ub, b) \underbrace{\left| \frac{\partial(y - ub)}{\partial y} \right|}_1 = p_b(b)p_n(y - ub).$$

Следовательно,

$$\begin{aligned} p(b|y) &= p_b p_n(y - ub) p_y^{-1}(y) = \\ &= C_1 \exp \left[-\frac{1}{2} \left\{ \frac{(b - \mu_b)^2}{\sigma_b^2} + \frac{(y - ub)'(y - ub)}{\sigma_n^2} - \right. \right. \\ &\quad \left. \left. -(y - ub)'(\sigma_b^2 uu' + \sigma_n^2 I)^{-1}(y - ub) \right\} \right], \end{aligned} \quad (11.7)$$

где C_1 — нормировочная постоянная. Отправляясь от этой формулы, читатель может потратить довольно много времени на вычисление интеграла (11.4) с учетом всех составляющих экспоненты и нормировочной постоянной. Вычисление существенно облегчается, если вспомнить некоторые свойства гауссовой плотности распреде-

лений

$$p(x) = \frac{1}{\sqrt{2\pi} \sigma_x} \exp \left[-\frac{1}{2} \left\{ \frac{x - \mu_x}{\sigma_x} \right\}^2 \right],$$

а именно

$$\int p(x) dx = 1, \quad \int (x - \mu_x) p(x) dx = 0.$$

Поэтому при интегрировании по b важно лишь выделить полный квадрат относительно b :

$$\begin{aligned} & -\frac{1}{2} \left\{ b^2 \left(\frac{1}{\sigma_b^2} + \frac{\mathbf{u}' \mathbf{u}}{\sigma_n^2} \right) - 2b \left(\frac{\mu_b}{\sigma_b^2} + \frac{\mathbf{u}' \mathbf{y}}{\sigma_n^2} \right) + \dots \right\} = \\ & = -\frac{1}{2} \left(\frac{1}{\sigma_b^2} + \frac{\mathbf{u}' \mathbf{u}}{\sigma_n^2} \right) \{(b - \alpha)^2 + \dots + \dots \}, \end{aligned} \quad (11.8)$$

где

$$\alpha = \frac{\mu_b/\sigma_b^2 + \mathbf{u}' \mathbf{y}/\sigma_n^2}{1/\sigma_b^2 + \mathbf{u}' \mathbf{u}/\sigma_n^2}. \quad (11.9)$$

Следовательно,

$$\begin{aligned} \hat{\beta} &= \int bp(b | \mathbf{y}) db = \int (b - \alpha + \alpha) p(b | \mathbf{y}) db = \\ &= \int (b - \alpha) p(b | \mathbf{y}) db + \alpha \int p(b | \mathbf{y}) db = 0 + \alpha, \end{aligned}$$

т. е.

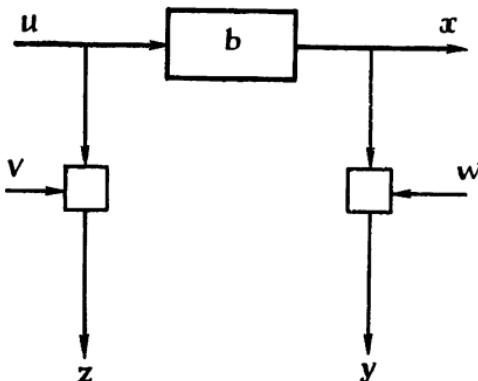
$$\hat{\beta} = \alpha = \frac{\sigma_n^2 \mu_b + \sigma_b^2 \mathbf{u}' \mathbf{y}}{\sigma_n^2 + \sigma_b^2 \mathbf{u}' \mathbf{u}}. \quad (11.10)$$

Рассмотрим два предельных случая:

$$\sigma_n^2 \neq 0, \quad \sigma_b^2 = 0, \quad \text{тогда} \quad \hat{\beta} = \mu_b;$$

$$\sigma_n^2 = 0, \quad \sigma_b^2 \neq 0, \quad \text{тогда} \quad \hat{\beta} = \frac{\mathbf{u}' \mathbf{y}}{\mathbf{u}' \mathbf{u}}.$$

В первом случае априори нет никакой неопределенности относительно b и производимые с ошибкой измерения, естественно, не могут уточнить эту информацию. Во втором случае при точных измерениях удается получить точную оценку. В общем случае $\sigma_n^2 \neq 0$, $\sigma_b^2 \neq 0$ и для построения оценки используется как априорная информация, так и данные измерений. Дисперсиями σ_n^2 и σ_b^2 определяются относительные веса, придаваемые каждой



Ф и г. 11.4.

из этих составляющих информации. Отметим, что

$$\mathcal{E}[u'y] = k\Psi_{uy}, \quad \mathcal{E}[u'u] = k\Psi_{uu}, \quad (11.11)$$

откуда следует, что вес, придаваемый измерениям, возрастает с увеличением числа выборочных значений сигналов. Если $\mu_b \neq \mathcal{E}[b]$, то при $\sigma_u^2 \neq 0, \sigma_b^2 \neq 0$ оценка смещена, но является асимптотически несмещенной.

В качестве упражнения полезно рассмотреть ту же задачу в предположении, что шум v не является белым и имеет ковариационную матрицу N . Кроме того, мы рекомендуем читателю вывести аналог формулы (11.10) для случая векторного параметра $\hat{\beta}$.

Интересная попытка применения теории статистических решений к задаче оценивания параметров объекта для еще более общего случая предпринята Масловым [28]. Рассмотренная задача иллюстрируется фиг. 11.4; предполагается, что входные и выходные сигналы импульсные; вход и выход искажаются шумами v и w соответственно; плотности распределения вероятностей u, v, w, b известны, эти случайные процессы стационарны и независимы; оператор объекта имеет вид $x(k) = F(u, b, k)$; известно, каким образом шумы добавляются к сигналам, причем в этих каналах отсутствует память и могут быть определены $p\langle z(k) | u(k) \rangle$ и $p\langle y(k) | x(k) \rangle$. Необходимо найти наилучшую оценку b :

$$\hat{\beta} = \hat{\beta}(z, y). \quad (11.12)$$

Лучшее, что можно сделать в статистическом смысле, — это определить для всех β апостериорную плотность

$$p(\beta | z, y). \quad (11.13)$$

Исходя из функции риска, нужно построить решающее правило

$$\beta = \Gamma(z, y), \quad (11.14)$$

удовлетворяющее критерию минимума риска. Это правило может быть как детерминированным, так и стохастическим. Необходимость в стохастическом правиле появляется в случае игры с сознательно противодействующим противником. В задаче оценивания, когда «противником» является природа, следует пользоваться детерминированным правилом.

Маслов рассматривал случаи, когда для оценивания используется как z , так и y и когда используется только y . Ограничимся здесь рассмотрением только первого случая. При *даных наблюдениях* z и y условный риск определяется как

$$\mathcal{E}[C(\beta, b) | z, y] = \int_{\Omega(\beta, b)} C(\beta, b) p(\beta, b | z, y) d\Omega(\beta, b). \quad (11.15)$$

Средний риск для разных экспериментов имеет вид

$$R(\beta) = \int_{\Omega(\beta, b, z, y)} C(\beta, b) p(\beta, b | z, y) p(z, y) d\Omega(\beta, b, z, y). \quad (11.16)$$

Используя правило Байеса, независимость шумов и физическую реализуемость объекта, Маслов получил несколько более общее выражение для среднего риска. Он показал, что оптимальная оценка $\hat{\beta}$ находится с помощью детерминированного решающего правила. Детали метода и некоторые примеры читатель найдет в цитированной статье.

Пока имеется сравнительно небольшое число работ, посвященных использованию байесовского подхода в задачах оценивания параметров [16, 23]. Возможно, это связано с вычислительными сложностями при оценивании условных математических ожиданий. В работе [29] предложено решение методом Монте-Карло.

11.2. ОЦЕНИВАНИЕ ПО МЕТОДУ МАКСИМАЛЬНОГО ПРАВДОПОДОБИЯ

Основные аспекты оценивания по методу максимального правдоподобия упоминались в разд. 2.2 и 5.2. Условием применения этого способа оценивания является наличие априорной информации о плотности вероятности дискретной реализации y . Функциональные связи будут записываться как $p\{y; b\}$, если b известно и интерес представляет $p(y)$, и как $L\{c; \beta\}$, если y известно (наблюдается), $y = c$ и нужно получить оценку β .

Оценкой максимального правдоподобия $\hat{\beta}$ величины b называется величина, максимизирующая функцию L , т. е. удовлетворяющая необходимому условию

$$\frac{\partial}{\partial b} \ln L\{y; b\}|_{b=\hat{b}} = 0. \quad (11.17)$$

Нужно найти абсолютный максимум L .

Проиллюстрируем этот метод простым примером. Пусть b — подлежащий оцениванию параметр, а $n(i)$ — независимый дискретный аддитивный шум с характеристиками

$$p(n) = \frac{1}{(2\pi)^{1/2} \sigma} \exp\left[-\frac{1}{2} \frac{n^2}{\sigma^2}\right], \quad (11.18)$$

$$\mathcal{E}[n] = 0, \quad \mathcal{E}[n^2] = \sigma^2.$$

Наблюдается последовательность

$$y(i) = b + n(i).$$

Отсюда для функции правдоподобия получаем

$$L\{y; b\} = \prod_{i=1}^k p\{y(i); b\} =$$

$$= \frac{1}{(2\pi)^{k/2} \sigma^k} \prod_{i=1}^k \exp\left[-\frac{1}{2} \frac{(y(i)-b)^2}{\sigma^2}\right], \quad (11.19)$$

или, логарифмируя, имеем

$$\ln L\{y; b\} = C - \frac{1}{2\sigma^2} \sum_{i=1}^k \{y(i)-b\}^2. \quad (11.20)$$

Приравнивание производной к нулю дает

$$\frac{\partial}{\partial b} [\ln L \{y; b\}]_{b=\hat{b}} = -\frac{1}{\sigma^2} \sum_{i=1}^k \{y(i) - \hat{b}\} = 0, \quad (11.21)$$

откуда, как легко было предвидеть для этого простого случая,

$$\hat{b} = \frac{1}{k} \sum_{i=1}^k y(i). \quad (11.22)$$

В виде упражнения и для сравнения с методом Байеса стоит вернуться к примеру из предыдущего раздела (фиг. 11.3) и вычислить для него оценку максимального правдоподобия.

В разд. 5.2 указывалась связь между оценками максимального правдоподобия, марковскими оценками и оценками по методу наименьших квадратов.

Достижимая точность

Выражение для точности оценивания по методу максимального правдоподобия в терминах дисперсий и ковариаций можно получить несколькими способами [22]. Рассмотрим функцию правдоподобия для k наблюдений в случае одномерного параметра b

$$L \{y(1), \dots, y(k); b\}, \quad (11.23)$$

причем

$$\int_k L dy(1) \dots dy(k) = \int_k L d^k y = 1. \quad (11.24)$$

Дифференцированием по b получаем уравнение¹⁾

$$\int_k \frac{\partial L}{\partial b} d^k y = 0, \quad (11.25)$$

которое можно записать как

$$\int_k \left(\frac{1}{L} \frac{\partial L}{\partial b} \right) L d^k y = \int_k \left(\frac{\partial \ln L}{\partial b} \right) L d^k y = \mathcal{E} \left[\frac{\partial \ln L}{\partial b} \right] = 0 \quad (11.25a)$$

¹⁾ Здесь и далее предполагается существование всех рассматриваемых производных.

Дифференцируя еще раз, получаем

$$\int_k \left\{ \left(\frac{\partial \ln L}{\partial b} \right)^2 + \frac{\partial^2 \ln L}{\partial b^2} \right\} L d^k y = 0, \quad (11.26)$$

или

$$\mathcal{E} \left[\left(\frac{\partial \ln L}{\partial b} \right)^2 \right] = - \mathcal{E} \left[\frac{\partial^2 \ln L}{\partial b^2} \right]. \quad (11.27)$$

Теперь рассмотрим $\hat{\beta} = \hat{\beta}(y)$ как оценку b :

$$\mathcal{E} [\hat{\beta}] = \int_k \hat{\beta} L d^k y = b + \lambda(b), \quad (11.28)$$

где $\lambda(b)$ — возможное смещение оценки.

Дифференцируя, получаем

$$\int_k \hat{\beta} \frac{\partial \ln L}{\partial b} L d^k y = 1 + \frac{d\lambda(b)}{db} \quad (11.29)$$

или с учетом формул (11.25a) и (11.28) имеем

$$\int_k (\hat{\beta} - f(b)) \frac{\partial \ln L}{\partial b} L d^k y = 1 + \frac{d\lambda(b)}{db}. \quad (11.30)$$

Из неравенства Коши — Шварца следует, что

$$\begin{aligned} \left\{ \int_k \{ \hat{\beta} - f(b) \}^2 L d^k y \right\} \left\{ \int_k \left(\frac{\partial \ln L}{\partial b} \right)^2 L d^k y \right\} &\geq \\ &\geq \left(1 + \frac{d\lambda(b)}{db} \right)^2. \end{aligned} \quad (11.31)$$

С помощью формул (11.28) и (11.27) отсюда получаем

$$\text{var} [\hat{\beta}] = \mathcal{E} [(\hat{\beta} - f(b))^2] \geq \frac{\left(1 + \frac{d\lambda(b)}{db} \right)^2}{\mathcal{E} \left[\left(\frac{\partial \ln L}{\partial b} \right)^2 \right]} = - \frac{\left(1 + \frac{d\lambda(b)}{db} \right)^2}{\mathcal{E} \left[\frac{\partial^2 \ln L}{\partial b^2} \right]}. \quad (11.32)$$

Это неравенство Крамера — Рао [12].

Для несмешенных оценок $\lambda(b) = 0$ и

$$\mathcal{E} [\hat{\beta}] = f(b) = b, \quad \text{var} [\hat{\beta}] \geq - \frac{1}{\mathcal{E} \left[\frac{\partial^2 \ln L}{\partial b^2} \right]}, \quad (11.32a)$$

для смещенной оценки $\lambda(b) \neq 0$ и

$$\mathcal{E}[\hat{\beta}] = b + \lambda(b), \quad \text{var}[\hat{\beta}] \geq -\frac{\left(1 + \frac{d\lambda}{db}\right)^2}{\mathcal{E}\left[\frac{\partial^2 \ln L}{\partial b^2}\right]}. \quad (11.32b)$$

В многомерном случае для несмещенной оценки можно получить (см. приложение В)

$$\text{cov}[\hat{\beta}] = \mathcal{E}[(\hat{\beta} - b)(\hat{\beta} - b)'] \geq J^{-1}, \quad (11.32b)$$

где

$$J = \mathcal{E}\left[\left(\frac{\partial \ln L}{\partial b}\right)\left(\frac{\partial \ln L}{\partial b}\right)'\right] = -\mathcal{E}\left[\frac{\partial^2 \ln L}{\partial b \partial b'}\right]. \quad (11.33)$$

Матрица J называется *информационной* (Фишер).

Для применения неравенства Крамера — Рао рассмотрим следующую модель:

$$y = Ub + n,$$

$$p(n) = \frac{1}{(2\pi)^{k/2} |N|^{1/2}} \exp\left[-\frac{1}{2} n' N^{-1} n\right],$$

откуда для функции правдоподобия получается

$$L\{y; b\} = \frac{1}{(2\pi)^{k/2} |N|^{1/2}} \exp\left[-\frac{1}{2} (y - Ub)' N^{-1} (y - Ub)\right], \quad (11.34)$$

$$\ln L\{y; b\} = c - \frac{1}{2} (y - Ub)' N^{-1} (y - Ub),$$

$$\frac{\partial \ln L}{\partial b} = U' N^{-1} (y - Ub),$$

$$\frac{\partial}{\partial b'} \left\{ \frac{\partial \ln L}{\partial b} \right\} = -U' N^{-1} U,$$

$$J = \mathcal{E}[U' N^{-1} U] = U' N^{-1} U. \quad (11.35)$$

Следовательно, ковариация ограничена величиной

$$\text{cov}[\hat{\beta}] \geq [U' N^{-1} U]^{-1}. \quad (11.36)$$

Интересно сравнить это выражение с формулой (6.47), ковариацией марковской оценки. Учитывая, что марковская оценка минимизирует дисперсию, и соответствие

марковской и максимально правдоподобной оценок при гауссовском шуме, можно не удивляться результатам этого сравнения.

Символ $\mathcal{E}[\cdot]$ в (11.35) был опущен в соответствии с предположением, что u измеряется без ошибок. Следовательно, J известна апостериори. Если желательно иметь *априорную* информацию относительно J и сигнал u (априори) является случайным, оператор математического ожидания опускать нельзя. Когда n — белый шум,

$$N^{-1} = \frac{1}{\sigma_n^2} I$$

и

$$J = \frac{1}{\sigma_n^2} \mathcal{E}[U'U]. \quad (11.37)$$

Продолжим обсуждение достижимой точности простым примером — задачей оценивания двух параметров $\beta' = \{\beta_0, \beta_1\}$ в случае, когда u_1 представляет собой ту же самую дискретную последовательность, что и u_0 , сдвинутую во времени на один шаг. В случае эргодического u_0

$$\mathcal{E}[U'U] = k \begin{bmatrix} \Psi_{uu}(0) & \Psi_{uu}(1) \\ \Psi_{uu}(1) & \Psi_{uu}(0) \end{bmatrix} = k\sigma_u^2 \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \quad (11.38)$$

и выражение для минимальной ковариации принимает вид

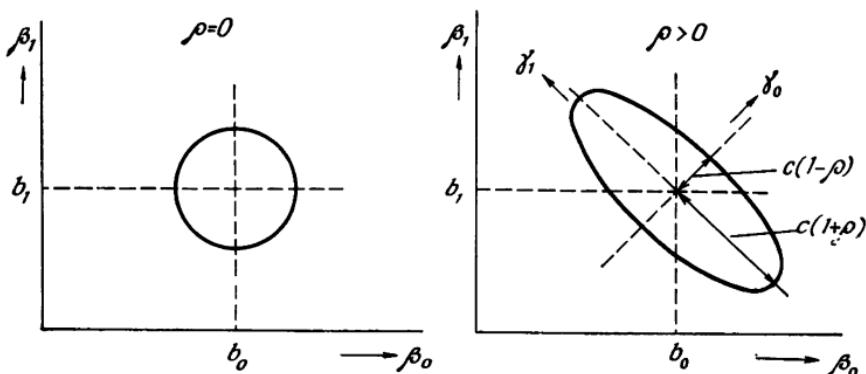
$$\text{cov}[\hat{\beta}] = \frac{\sigma_n^2}{k\sigma_u^2} \frac{1}{1-\rho^2} \begin{bmatrix} 1 & -\rho \\ -\rho & 1 \end{bmatrix} = c \begin{bmatrix} 1 & -\rho \\ -\rho & 1 \end{bmatrix}. \quad (11.39)$$

При $\rho \neq 0$ внедиагональные элементы представляют корреляцию ошибок в β_0 и β_1 . С помощью простых преобразований можно перейти к рассмотрению дисперсий компонент вектора $T\beta$, например

$$\frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} \beta_0 + \beta_1 \\ -\beta_0 + \beta_1 \end{bmatrix} = v.$$

Следовательно,

$$\begin{aligned} \text{cov}[\hat{v}] &= \frac{c}{2} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 1 & -\rho \\ -\rho & 1 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} = \\ &= c \begin{bmatrix} 1-\rho & 0 \\ 0 & 1+\rho \end{bmatrix}. \end{aligned} \quad (11.40)$$



Фиг. 11.5.

Теперь внедиагональные члены равны нулю, и компоненты $\hat{\beta}$ некоррелированы. Это положение иллюстрируется на фиг. 11.5, где приведен так называемый *эллипсоид рассеивания*. Рисунок дает наглядное представление точности, с которой могут быть определены параметры с учетом корреляции ошибок. С ростом числа выборочных значений k эллипсоид рассеивания стягивается в точку (b_0, b_1) , поскольку k входит в знаменатель (11.39).

Некоторые важные свойства

При постоянном размере выборки несмещенная оценка называется *эффективной*, если в соотношении Крамера — Рао (11.32) имеет место знак равенства. *Эффективность* определяется как

$$\eta(\beta) = \frac{\min \operatorname{var}[\hat{\beta}]}{\operatorname{var}[\beta]}. \quad (11.41)$$

Оценка $\hat{\beta}$ называется *достаточной*, если

$$p\{y(1), \dots, y(k); b\} =$$

$$= g\{\hat{\beta}; b\} H\{y(1), \dots, y(k)\}. \quad (11.42)$$

Согласно результатам Крамера [12], если эффективная оценка $\hat{\beta}$ параметра b существует, то уравнение правдоподобия имеет единственное решение $\hat{\beta}$; если существует

достаточная оценка $\hat{\beta}$ параметра b , то любое решение уравнения правдоподобия является функцией $\hat{\beta}$.

Отметим, что все оценки максимального правдоподобия либо несмещенные, либо эффективные.

При растущем размере выборки k можно изучать асимптотическое поведение $\hat{\beta} \{y(1), \dots, y(k)\}$. Если оценка $\hat{\beta}$ сходится по вероятности к b при $k \rightarrow \infty$, то она называется *состоятельной*;

$$\lim_{k \rightarrow \infty} P[|\hat{\beta} \{y(1), \dots, y(k)\} - b| \leq \varepsilon] = 1. \quad (11.43)$$

Здесь ε — сколь угодно малое положительное число. Любая максимально правдоподобная оценка состоятельна. В работе [4] предложено называть объект идентифицируемым, если оценка состоятельна. Необходимым условием является положительная определенность информационной матрицы, связанной с процессом оценивания.

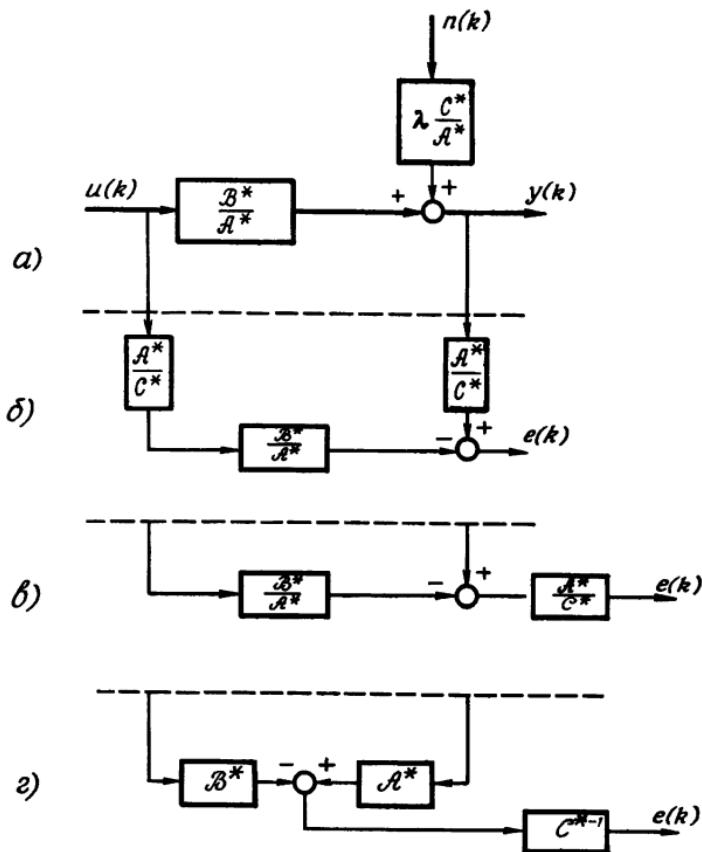
Если эффективность $\eta(\hat{\beta}) \rightarrow 1$ при $k \rightarrow \infty$, то оценка называется *асимптотически эффективной*. Можно показать, что оценки максимального правдоподобия асимптотически эффективны. Кроме того, они обладают свойством *асимптотической нормальности*, т. е. их распределение приближается к нормальному со средним b и дисперсией J^{-1} .

11.3. НЕКОТОРЫЕ СХЕМЫ ПРАКТИЧЕСКОЙ РЕАЛИЗАЦИИ

Интересное приложение процедуры оценивания по методу максимального правдоподобия к задаче оценивания параметров дано в [4]. Рассмотренный там объект описывается линейным дифференциальным уравнением с постоянными коэффициентами. Объект имеет один вход и один выход:

$$A^*(z^{-1}) y(k) = B^*(z^{-1}) u(k) + \lambda C^*(z^{-1}) n(k), \quad (11.44)$$

где z — оператор сдвига: $z^{-1}x(k) = x(k-1)$, $u(k)$ — вход объекта, $y(k)$ — выход, $n(k)$ — гауссовский аддитивный шум с независимыми значениями и дисперсией



Ф и г. 11.6.

$\sigma_n^2 = 1$, λ — уровень шума,

$$A^*(z^{-1}) = 1 + a_1 z^{-1} + \dots + a_n z^{-n},$$

$$B^*(z^{-1}) = b + b_1 z^{-1} + \dots + b_n z^{-n},$$

$$C^*(z^{-1}) = 1 + c_1 z^{-1} + \dots + c_n z^{-n}.$$

На фиг. 11.6 эти соотношения представлены блок-схемой. Многочлены A^* , B^* и C^* предполагаются взаимно простыми; некоторые из коэффициенты могут быть равны нулю; все корни A^* и C^* лежат внутри единичной окружности.

Модель описывается соотношением

$$\mathcal{E}^*(z^{-1}) e(k) = A^*(z^{-1}) y(k) - B^*(z^{-1}) u(k), \quad (11.45)$$

которое представляется блок-схемами фиг. 11.6, б, в или г. В блоках $\mathcal{A}^*/\mathcal{C}^*$ легко узнать фильтры, преобразующие случайный процесс в белый шум. Следует заметить, что уравнения (11.44) и (11.45) можно записать в векторной форме

$$\mathbf{ay} = \mathbf{bu} + \lambda \mathbf{en}, \quad (11.46)$$

$$\mathbf{ay} = \boldsymbol{\beta}\mathbf{u} + \boldsymbol{\gamma}\mathbf{e}, \quad \boldsymbol{\gamma}\mathbf{e} = \boldsymbol{\alpha}\mathbf{y} - \boldsymbol{\beta}\mathbf{u}. \quad (11.47)$$

Эта схема охватывает ряд проблем, рассмотренных в работах Острема, например:

$\mathbf{y}(k) = \mathbf{bu}(k) + \lambda \mathbf{n}(k)$ — регрессия,

$\mathbf{ay} = \lambda \mathbf{n}$ — авторегрессия,

$\mathbf{y} = \lambda \mathbf{en}$ — оценивание параметров методом скользящего среднего,

$\mathbf{ay} = \lambda \mathbf{cn}$ — оценивание дробно-рациональной спектральной плотности,

$\mathbf{ay} = \mathbf{bu} + \mathbf{n}$ — оценивание по методу наименьших квадратов

$\mathbf{ay} = \mathbf{bu} + \lambda \mathbf{en}$ — марковский подход,

$\mathbf{a}(\mathbf{y} - \lambda \mathbf{n}) = \mathbf{bu}$ — идентификация свободного от шумов объекта при наличии ошибок измерений.

Благодаря гауссовскому распределению шума n функция правдоподобия для $\mathbf{e}' = \{e(1), \dots, e(k)\}$ имеет вид

$$L(\mathbf{e}; \mathbf{u}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \lambda) = \frac{1}{(2\pi)^{k/2} \lambda^k} \exp \left[-\frac{\mathbf{e}'\mathbf{e}}{2\lambda^2} \right], \quad (11.48)$$

или иначе

$$\ln L = -\frac{k}{2} \ln 2\pi - k \ln \lambda - \frac{\mathbf{e}'\mathbf{e}}{2\lambda^2}. \quad (11.49)$$

Оценка максимального правдоподобия $\hat{\lambda}$ получается из уравнения

$$\frac{\partial \ln L}{\partial \lambda} = \left[-\frac{k}{\lambda} + \frac{\mathbf{e}'\mathbf{e}}{\lambda^3} \right]_{\lambda=\hat{\lambda}} = 0$$

в виде

$$\hat{\lambda}^2 = \frac{\mathbf{e}'\mathbf{e}}{k}. \quad (11.50)$$

Для дальнейшего удобно обозначить

$$\theta = \begin{bmatrix} \alpha \\ \beta \\ \gamma \end{bmatrix}, \quad e = e(\theta), \quad e'e = 2E(\theta),$$

где θ — вектор подлежащих оцениванию параметров. Из (11.47) и фиг. 11.6 видно, что e линейно зависит от α и что β нелинейно зависит от γ . Следовательно, поиск максимума L или $\ln L$ путем приравнивания производной к нулю не обязательно является простой процедурой. Острэм использовал метод Ньютона — Рафсона

$$\theta(i+1) = \theta(i) - \left[\left[\frac{\partial^2 E(\theta)}{\partial \theta \partial \theta} \right]^{-1} \frac{\partial E(\theta)}{\partial \theta} \right]_{\theta=\theta(i)}. \quad (11.51)$$

Подробности читатель может найти в цитированной работе. Использование метода максимального правдоподобия для оценивания параметров рассмотрено также в [3—5, 17, 36, 38].

Метод максимального правдоподобия может также рассматриваться как процедура определения коэффициентов модели предсказания

$$y_M(k) = \hat{y}(k|k-1) = \frac{\mathcal{B}^*(z^{-1})}{\mathcal{C}^*(z^{-1})} u(k) - \frac{\mathcal{A}^*(z^{-1}) - \mathcal{C}^*(z^{-1})}{\mathcal{C}^*(z^{-1})} y(k) \quad (11.52)$$

из условия минимума критерия

$$E = \sum_{i=1}^k \{y(i) - y_M(i)\}^2 = \sum_{i=1}^k e^2(i). \quad (11.53)$$

Отметим, что формулу (11.44) можно также переписать в виде

$$A^*(z^{-1}) \left[\frac{1}{\mathcal{C}^*(z^{-1})} y(k) \right] = B^*(z^{-1}) \left[\frac{1}{\mathcal{C}^*(z^{-1})} u(k) \right] + \lambda n(k). \quad (11.54)$$

Это означает, что метод максимального правдоподобия можно также рассматривать как обобщенный метод наименьших квадратов, когда характеристика фильтра $G = 1/C$ определяется из модели (11.40) [6, 8]. Метод

максимального правдоподобия широко применялся в промышленных измерениях (см., например, [3, 7, 17]). В последней из упомянутых работ проведено сравнение его с другими методами, такими, как корреляционные методы и обобщенный метод наименьших квадратов. Метод максимума правдоподобия применялся и при анализе временных рядов (если положить $B = 0$). Оценка максимального правдоподобия — существенно нелинейная функция параметров. Поскольку при анализе временных рядов преимущественно рассматриваются квадратичные функции, такие, как ковариации и спектральные плотности, можно ожидать, что оценки будут выражаться как нелинейные функции ковариаций выборки. Оценки такого рода, асимптотически эквивалентные оценкам максимального правдоподобия, для случая анализа параметрических временных рядов приводятся в [39]. Обобщения метода максимума правдоподобия на многомерный случай сделаны в [11] и [38].

11.4. ТРЕБОВАНИЯ К ВХОДНЫМ СИГНАЛАМ ОБЪЕКТА

В разд. 11.2 было показано, что информационная матрица является важной мерой точности, получаемой с помощью процедуры оценивания [см. (11.33)]. Из формул (11.37) и (11.38) ясно, каким образом при определенном представлении объекта входной сигнал u влияет на точность. Можно поставить вопрос, каким условиям должен удовлетворять сигнал u , чтобы получались состоятельные оценки [1, 4, 8]. При этих условиях должны существовать пределы

$$\bar{u} = \lim_{k \rightarrow \infty} \frac{1}{k} \sum_{i=1}^k u(i),$$

$$R_{uu}(j) = \lim_{k \rightarrow \infty} \frac{1}{k} \sum_{i=1}^k \{u(i) - \bar{u}\} \{u(i+j) - \bar{u}\} \quad (11.55)$$

и матрица

$$A_n = [a_{jl} = R_{uu}(j-l)], \quad j, l = 1, \dots, n, \quad (11.56)$$

должна быть положительно определена. Эти условия называют условиями *постоянного возбуждения* (порядка n).

Отметим, что если существует среднеквадратическое преобразование Фурье входного сигнала, т. е.

$$f_u(\omega) = \lim_{k \rightarrow \infty} \frac{1}{\sqrt{k}} \sum_{i=1}^k \{u(i) - \bar{u}\} e^{j\omega i}, \quad (11.57)$$

то

$$R_{uu}(\tau) = \int |f_u(\omega)|^2 e^{j\omega\tau} d\omega \quad (11.58)$$

и матрица A_n , определяемая формулой (11.56), автоматически оказывается неотрицательно определенной при произвольном n . Более того, если

$$f_u(\omega) > 0, \quad (11.59)$$

матрица A_n положительно определена при любом ω . Условие (11.59), таким образом, достаточно для того, чтобы входной сигнал был постоянно возбуждающим произвольного порядка. Заметим еще, что если входной сигнал — эргодический стационарный случайный процесс, то его спектральная плотность имеет вид

$$\Phi_{uu}(\omega) = |f_u(\omega)|^2. \quad (11.60)$$

Условие (11.59) означает, что спектральная плотность входного сигнала не обращается в нуль ни при каких ω . Это условие хорошо известно в корреляционном анализе.

Помимо условий постоянного возбуждения для многих применений необходимо, чтобы выходной сигнал поддерживался в определенных пределах в течение всего эксперимента. Задача синтеза входных сигналов, ограниченных во времени и по энергии и которые являются оптимальными, например, в смысле минимума дисперсий оценок, рассматривается в [1, 2, 26, 31]. Эта же задача обсуждается в [33]. Она тесно связана с задачей синтеза оптимальных сигналов в теории связи (см., например, [30]).

ЛИТЕРАТУРА

1. Aoki M., Staley R. M., On Input Signal Synthesis in Parameter Identification, Fourth IFAC Congress, Warsaw, Paper 26.2, 1969; см. также *Automatica*, 6, 431—440 (1969).
2. Aoki M., Staley R. M., Some Computational Considerations in Input Signal Synthesis Problems, in: *Computing Methods in*

- Optimization Problems, Zadeh, Neustadt, Balakrishnan, eds., vol. 2, Academic Press, N.Y., 1969.
3. Åström K. J. (см. литературу к гл. 4).
 4. Åström K. J., Bohlin T., Numerical Identification of Linear Dynamic Systems from Normal Operating Records, IFAC Symp. «Theory of Self-Adaptive Control Systems», Teddington, Engl., in: Theory of Self-Adaptive Control Systems, Hammond P. H., ed., Plenum Press, N.Y., 1966.
 5. Åström K. J., Bohlin T., Wensmark S. (см. литературу к гл. 4).
 6. Åström K. J., On the Achievable Accuracy in Identification Problems, IFAC Symp. «Identification in Automatic Control Systems», Prague, Paper 1.8, 1967.
 7. Åström K. J., Computer Control of a Paper Machine — an Application of Linear Stochastic Control Theory, IBM J. Res. and Develop., 11, 389—405 (1967).
 8. Åström K. J. (см. литературу к гл. 6).
 9. Åström K. J. (см. литературу к гл. 4).
 10. Bucy R. S., Nonlinear Filtering Theory, *IEEE Trans. Autom. Control*, AC-12 (1965).
 11. Caines P., Maximum Likelihood Identification of Multivariable Systems, Ph. D. Thesis, Imperial College, London, 1970.
 12. Cramér H. (см. литературу к гл. 6).
 13. Cramér H., Leadbetter M., Stationary and Related Stochastic Processes, Wiley, N.Y., 1967; есть русский перевод: Крамер Г., Лидбеттер М., Стационарные случайные процессы, изд-во «Мир», 1969.
 14. Deutsch R. (см. литературу к гл. 5).
 15. Florentin J. J., Optimal, Probing, Adaptive Control of a Simple Bayesian System, *J. Electron. Control*, Ser. I, 13 (1962).
 16. Galtieri C. A., The Problem of Identification of Discrete Time Processes, Report, Electronics Research Lab., Univ. of Calif., Berkeley, 1963.
 17. Gustavsson I., Maximum Likelihood Identification of Dynamics of the Ågesta Reactor and Comparison with Results of Spectral Analysis, Rept. 6903, Division of Automatic Control, Inst. of Technology, Lund, Sweden, 1969.
 18. Gustavsson I., Comparison of Different Methods for Identification of Linear Models for Industrial Processes, Second IFAC Symp. «Identification and Process Parameter Estimation», Prague, Paper 11.4, 1970; см. также *Automatica*, 8, 127—142.
 19. Helstrom C. W., Statistical Theory of Signal Detection, Pergamon Press, N.Y., 1960.
 20. Ho Y. C., Lee R. C. K., A Bayesian Approach to the Problems in Stochastic Estimation and Control, *IEEE Trans. Joint Autom. Control*, AC-9, 333—339 (1964); см. также *Joint Autom. Control Conf.*, 382—387 (1964).
 21. Kale B. K., On the Solution of Likelihood Equations by Iteration on Processes — the Multiparametric Case, *Biometrika*, 49, 479 (1962).

22. Kendall M. G., Stuart A. (см. литературу к гл. 6).
23. Kerr R. B., Bayesian Identification of System Parameters, *J. Engineering Mathematics*, 4, 273—281 (1970).
24. Kroy W. H., Stubberud A. R., Identification via Nonlinear Filtering, Joint Automatic Control Conf., 1968, pp. 394—412, см. также *Intern. J. Control.*, 6, 499—522 (1968).
25. Kushner H. J., On the Differential Equations Satisfied by Conditional Probability Densities of Markov Processes with Applications, *SIAM J. Control, Ser. A*, 2, 106—119 (1964).
26. Levadi V. S., Design of Input Signals for Parameter Estimation, *IEEE Trans. Autom. Control*, AC-11, 205—211 (1966).
27. Levin M. J., Estimation of a System Pulse Transfer Function in the Presence of Noise, *IEEE Trans. Autom. Control*, AC-9, 229—235 (1964); см. также AC-9, 214—215 (1964).
28. Маслов Е. П. (см. литературу к гл. 5).
29. McGhee R. B., Walford R. B., A Monte Carlo Approach to the Evaluation of Conditional Expectation Parameter Estimates for Non-Linear Dynamic Systems, *IEEE Trans. Autom. Control* AC-13, 29—37 (1968).
30. Middleton D., An Introduction to Statistical Communication Theory, McGraw-Hill, N.Y., 1960.
31. Nahi N. E., Wallis D. E., Jr., Optimal Inputs for Parameter Estimation in Dynamic Systems with White Observation Noise, Joint Autom. Control Conf., 1969, pp. 506—513.
32. Rao C. R., Information and Accuracy Attainable in the Estimation of Statistical Parameters, *Bull. Calcutta Math. Soc.*, 37 (1945).
33. Rault A., Pouliquen R., Richalet J., Sensibilizing Input and Identification, Proc. Fourth IFAC Congress, Warsaw, Paper 5.2, 1969.
34. Стратонович Р. Л., Условные марковские процессы, *Теория вероятностей и ее применение*, № 2 (1960).
35. Swerling P., Parameter Estimation Accuracy Formulas, *IEEE Trans. Inform. Theory*, IT-10, 302—314 (1964).
36. Tretter S. A., Steiglitz K., Power Spectrum Identification in Terms of Rational Models, *IEEE Trans. Autom. Control*, AC-12, 185—188 (1967).
37. Wald A., Statistical Decision Functions, Wiley, N.Y., 1950.
38. Woo K. T., Maximum Likelihood Identification of Noisy Systems, Second IFAC Symp. «Identification and Process Parameter Estimation», Prague, Paper 3.1, 1970.
39. Zetterberg L. H., Estimation of Parameters for a Linear Difference Equation with Application to EEG Analysis, *Math. Biosci.*, 5, 227—275 (1969).

ДОПОЛНИТЕЛЬНАЯ ЛИТЕРАТУРА

- Alspach D. L., Sorenson H. W., Nonlinear Bayesian Estimation Using Gaussian Sum Approximations, *IEEE Trans. Autom. Control*, AC-17, 439—448 (1972).
- Aoki M., On Some Convergence Questions in Bayesian Optimization Problems, *IEEE Trans. Autom. Control*, AC-10 (1965).

- Bohlin T., On the Maximum Likelihood Method of Identification, *IBM J. Res. and Develop.*, 14, 41—51 (1970).
- Davies W. D. T., Douce J. L., On-Line System Identification in the Presence of Drift, IFAC Symp. «Identification in Autom. Control Systems», Prague, Paper 3.12, 1967.
- Lainiotis D. G., A Nonlinear Adaptive Estimation Recursive Algorithm (Correspondence), *IEEE Trans. Autom. Control*, AC-13, 197—198 (1968).
- Liporace L. A., Variance of Bayes Estimates, *IEEE Trans. Inform. Theory*, IT-17, 665—669 (1971).
- Miller R. W., Roy R., Nonlinear Process Identification Using Decision Theory (Short Paper), *IEEE Trans. Autom. Control*, AC-9, 538—540 (1964).
- Qvarnström B., An Uncertainty Relation for Linear Mathematical Models, Second IFAC Congress, Basle, 1963; есть русский перевод: Квантстрем Б., Неопределенность в линейных математических моделях, в книге «Теория непрерывных автоматических систем», Труды II Межд. конгресса ИФАК, Базель, 1963, изд-во «Наука», 1965.
- Rajbman N. S., Anisimov S. A., Hovsepian F. A., Some Problems of Control Plants Identification, Fourth IFAC Congress, Warsaw, Paper 5.1, 1969.
- Rogers A. E., Steiglitz K., Maximum Likelihood Estimation of Rational Transfer Function Parameters (Short Paper), *IEEE Trans. Autom. Control*, AC-12, 594—597 (1967); см. также AC-13, 304 (1967).
- Segerstähl B., A Theory of Parameter Tracking Applied to Slowly Varying Nonlinear Systems, Fourth IFAC Congress, Warsaw, Paper 5.4, 1969.
- Seidman L. P., Performance Limitations and Error Calculations for Parameter Estimation, *Proc. IEEE*, 58, 644—652 (1970).
- Štěphán J., The Uncertainty Problem in System Identification, Second IFAC Symp. «Identification and Process Parameter Estimation», Prague, Paper 1.8, 1970.
- Ziv J., Zakai M., Some Lower Bounds on Signal Parameter Estimation, *IEEE Trans. Inform. Theory*, IT-15, 386—391 (1969).

ОДНОВРЕМЕННОЕ ОЦЕНИВАНИЕ ПАРАМЕТРОВ И СОСТОЯНИЯ

ГЛАВА 12

ОЦЕНИВАНИЕ СОСТОЯНИЯ ОБЪЕКТА (ОБЗОР)

Как отмечалось в гл. 1, задача оценивания параметров тесно связана с задачами фильтрации и оценивания состояния. В следующей главе рассматриваются методы, позволяющие получать оценки и параметров и состояния. В этой главе кратко излагаются некоторые методы оценивания состояния (фильтр Калмана), которые облегчат понимание следующей главы. Схемы обработки сигналов, которые позволяют отделить полезные сигналы от нежелательных возмущений (шумов), называются фильтрами.

Фильтр Калмана можно описать как метод (аппаратуру, программу вычислений) для обработки результатов измерений, искаженных помехами; эта обработка должна осуществляться таким образом, чтобы получалась оптимальная оценка некоторой переменной. Этот фильтр может применяться в целом ряде задач:

а) фильтрация, интерполяция, сглаживание, экстраполяция (предсказание) временных рядов и непрерывных сигналов (разд. 12.1);

б) оценивание состояния, необходимое для реализации алгоритмов оптимального управления в случаях, когда некоторые или даже все компоненты вектора состояния объекта не поддаются непосредственному измерению;

в) объединение результатов измерений некоторой переменной несколькими измерительными приборами, различающимися типом ошибок (например, в задачах навига-

ции — акселерометры, гироскопы, радиолокаторы с измерением эффекта Допплера и т. д.);

г) оценивание параметров путем перехода от вектора состояния к вектору параметров.

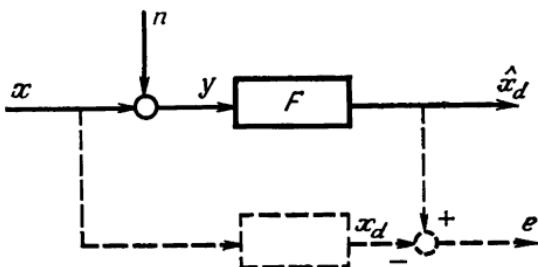
Очень интересный обзор этих методов можно найти в работе [23], где прослеживается история их развития, начиная от первых применений метода наименьших квадратов, выполненных К. Ф. Гауссом в 1795 г., и демонстрируется тесная связь этого метода с калмановской фильтрацией.

12.1. ФИЛЬТР ВИНЕРА

Во время второй мировой войны Винер выполнил свой фундаментальный, ставшие теперь классическими исследования по интерполяции, экстраполяции и сглаживанию временных рядов [25]. Для наших целей достаточно ограничиться постановкой задачи, схематически показанной на фиг. 12.1.

	Непрерывный случай	Дискретный случай
x — сигнал, несущий информацию	$x(t)$	$x(k)$
n — возмущающий сигнал (шум наблюдений)	$n(t)$	$n(k)$
x_d — желаемый сигнал: интерполированный отфильтрованный экстраполированный, предсказанный	$x(t - \tau)$ $x(t)$ $x(t + \tau),$ $\tau > 0$	$x(k - m)$ $x(k)$ $x(k + m),$ $m > 0$
\hat{x}_d — оценка x_d , полученная из сигнала $x + n$ с помощью фильтра F	$\hat{x}_d(t)$	$\hat{x}_d(k)$
$e = x_d - \hat{x}_d$ — сигнал ошибки, подлежащий минимизации выбором F в смысле наименьших квадратов	$e(t)$	$e(k)$

Конечно, выбор F должен основываться на априорной информации о процессах x и n . Задача может решаться как во временной, так и в частотной областях (см., например, [17, 22]).



Ф и г. 12.1.

Винер предположил, что сигналы $x(t)$ и $n(t)$ стационарны, т. е. их статистические свойства не зависят от времени, и их спектральные плотности известны. Фильтр F предполагается линейным, стационарным и физически реализуемым. Следовательно, его можно охарактеризовать весовой функцией $h(t)$. Если предположить, что $x(t)$ и $n(t)$ являются реализациями эргодического случайного процесса, то в качестве подходящего критерия оптимальности можно принять $\min \mathcal{E}[e^2(t)]$. При $x_d(t) = x(t)$ и $\hat{x}_d(t) = \hat{x}(t)$ этот критерий сводится к минимизации

$$\mathcal{E}[e^2(t)] = \mathcal{E}[\{\hat{x}(t) - x(t)\}^2],$$

где

$$\hat{x}(t) = \int_{-\infty}^{\infty} h(\theta) y(t-\theta) d\theta. \quad (12.1)$$

Минимизация осуществляется подобно тому, как это делалось в разд. 8.1, в результате чего приходим к необходимому условию, имеющему форму интегрального уравнения Фредгольма:

$$\Psi_{xy}(\tau) = \int_{-\infty}^{\infty} h(\theta) \Psi_{yy}(\tau-\theta) d\theta. \quad (12.2)$$

Для того чтобы фильтр был физически реализуем, нужно наложить на весовую функцию ограничение $h(t) = 0$ при $t < 0$. При этом условии вновь получается уравнение Винера — Хопфа. Решение этого уравнения является, вообще говоря, непростой задачей. Обычно с этой целью переходят в частотную область; условие физической реали-

заемости при этом вносит серьезные осложнения. Полученную в результате передаточную функцию можно записать в виде

$$H(s) = \frac{1}{\Phi_{yy}^+(s)} \left[\frac{\Phi_{xx}(s)}{\Phi_{yy}^-(s)} \right]_+, \quad (12.3)$$

где $\Phi_{yy}(s) = \Phi_{yy}^+(s) \Phi_{yy}^-(s)$, $\Phi_{yy}^+(s)$ имеет полюсы и нули в левой полуплоскости комплексной переменной s , а $\Phi_{yy}^-(s)$ — в правой полуплоскости, $[G(s)] = [G(s)]_+ + [G(s)]_-$, а $[]_+$ означает, что рассматриваются только полюсы в левой полуплоскости переменной s .

Полученный таким образом фильтр Винера оптимален для сигналов с гауссовским распределением. Аналогичные фильтры можно получить и для дискретных сигналов [6, 14]. Некоторые особенности винеровского метода фильтрации:

а) Предполагается, что априорная информация о сигнале и шуме задана в терминах корреляционных функций и спектральных плотностей. При этом возникает задача факторизации спектральной плотности, т. е. определения условий, при которых спектральную плотность можно представить в виде

$$\Phi(j\omega) = F(j\omega) F(-j\omega)$$

(см., например, [2]).

б) Определение оптимального фильтра недостаточно хорошо подходит для цифровых вычислений.

в) Поиск оптимального фильтра заканчивается определением весовой функции $h(t)$ или передаточной функции $H(s)$; синтез реального фильтра при этом может оказаться совсем не простым.

г) Метод не распространяется на нестационарные сигналы.

д) Его можно использовать для объектов со многими входами и выходами, хотя это и усложняет вычисления.

12.2. ФИЛЬТР КАЛМАНА — БЬЮСИ

Фильтр Калмана позволяет получить решение рассмотренной выше задачи, сформулированной, однако, в пространстве состояний. В отличие от метода Винера фильтр Калмана легко распространяется на случай нестационарных сигналов [8, 9].

Обычно фильтр является нестационарным. Кроме того, фильтр Калмана имеет рекуррентную форму, благодаря чему он удобен для программирования на цифровых машинах. Новые оценки получаются корректировкой старых на основе новых наблюдений; нет необходимости хранить весь массив наблюдавшихся данных, и, следовательно, требования к памяти и машинному времени оказываются довольно скромными. Даже если нас интересуют случайные сигналы в «чистом» виде, полезно рассматривать эти сигналы как получающиеся из белого шума с помощью так называемых формирующих фильтров. В дальнейшем такие фильтры будут также называться объектами. В общем случае динамику объекта можно описать уравнениями

$$\begin{aligned}\dot{\mathbf{x}} &= \mathbf{f}(\mathbf{x}, \mathbf{u}, \mathbf{v}, t), \\ \mathbf{y} &= \mathbf{g}(\mathbf{x}, \mathbf{u}, \mathbf{n}, t)\end{aligned}\tag{12.4}$$

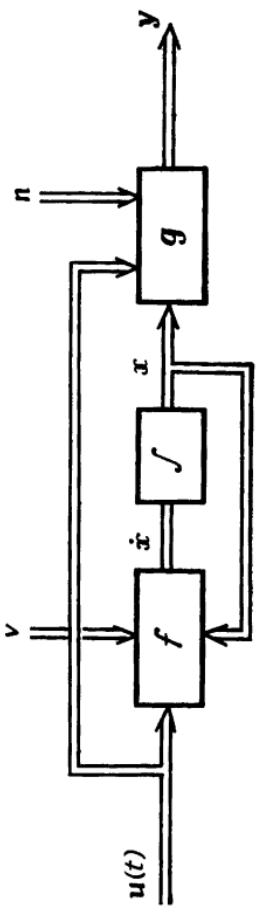
(фиг. 12.2). Для линеаризованного объекта они сводятся к

$$\begin{aligned}\dot{\mathbf{x}} &= \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{u} + \mathbf{v}, \\ \mathbf{y} &= \mathbf{C}\mathbf{x} + \mathbf{D}\mathbf{u} + \mathbf{n}\end{aligned}$$

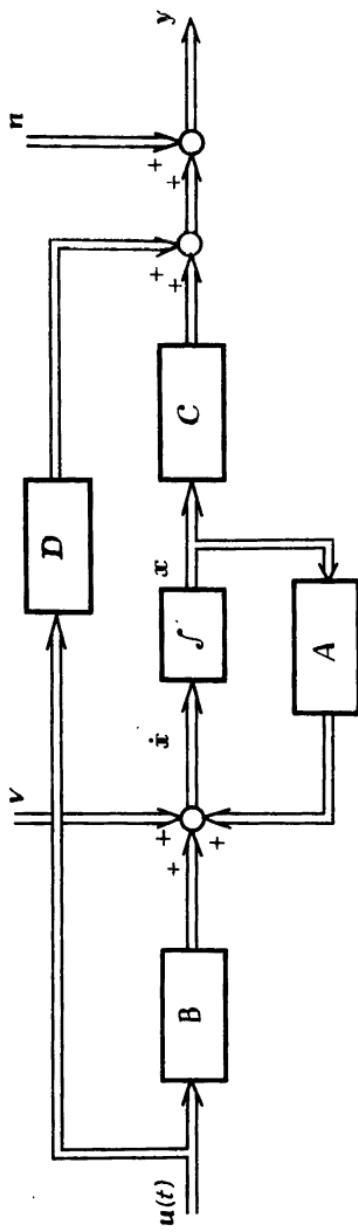
(фиг. 12.3). Функциональная связь между переменными состояниями и другими сигналами предполагается известной, включая и числовые значения параметров. Основываясь на наблюдениях \mathbf{y} в течение интервала времени $(0, T)$, нужно найти $\hat{\mathbf{x}}(t | T)$, оценку $\mathbf{x}(t)$. Как и раньше, возможны три случая: $t > T$, задача предсказания; $t = T$, задача фильтрации; $t < T$, задача сглаживания. В соответствии с изложенным в гл. 5 задачу можно решать, используя байесовский подход, оценивание по максимуму правдоподобия или метод наименьших квадратов. Здесь мы придерживаемся первого подхода как наиболее общего.

Опять исчерпывающую со статистической точки зрения информацию дает плотность условного распределения вероятностей $p(\mathbf{x} | \mathbf{y})$, где \mathbf{y} — наблюдавшиеся реализации выхода объекта. Как и в гл. 5, с помощью этой плотности вероятности можно получить разные оценки. По правилу Байеса

$$p(\mathbf{x} | \mathbf{y}) = \frac{p(\mathbf{x}, \mathbf{y})}{p(\mathbf{y})} = \frac{p(\mathbf{y} | \mathbf{x}) p(\mathbf{x})}{p(\mathbf{y})}\tag{12.5}$$



Ф и г. 12.2.



Ф и г. 12.3.

можно, используя априорную информацию об x и результаты наблюдений y , определить новую (апостериорную) плотность вероятности для x . Это естественным образом приводит к итерационной процедуре оценивания, когда результат предыдущего шага вычислений служит априорной информацией на следующем шаге, где используются новые наблюдения. Рассмотрим сначала простой одншаговый фильтр, а затем перейдем к его многошаговому варианту.

Одншаговый фильтр

Дана следующая априорная информация:

объект $y = Cx + n$ (фиг. 12.4);

$p(x)$ — гауссовская плотность, $E[x] = x_0$, $\text{cov}[x] = P_0$;
 $p(n)$ — гауссовская плотность, $E[n] = 0$, $\text{cov}[n] = N$; x и n независимы.

На основе этой информации легко получить плотность вероятности y , хотя в действительности она нам не нужна:

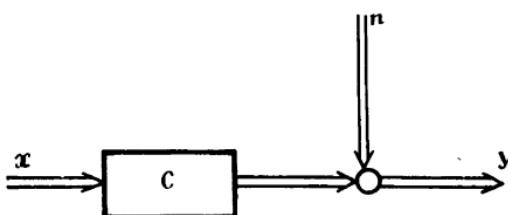
$p(y)$ — гауссовская, $E[y] = Cx_0$, $\text{cov}[y] = CP_0C' + N$, поскольку

$$\begin{aligned} \text{cov}[y] &= E[\{y - Cx_0\}\{y - Cx_0\}'] = \\ &= E[\{\{C(x - x_0) + n\}\{C(x - x_0) + n\}'\}]. \end{aligned}$$

Запишем теперь формулу для $p(x | y)$:

$$\begin{aligned} p(x | y) &= \frac{p(x, y)}{p(y)} = \frac{p(x, n) \left| \frac{\partial n}{\partial y} \right|}{p(y)} = \\ &= \frac{p(x) p(n)}{p(y)} = \frac{p(x) p_n(y - Cx)}{p(y)}. \quad (12.6) \end{aligned}$$

Сюда можно подставить выражения для всех рассматриваемых плотностей вероятности. Затем нужно выбором



Ф и г. 12.4.

функции штрафа определить вид требуемой оценки. В гл. 5 уже указывалось, что в случае гауссовского распределения использование любой четной функции штрафа и максимизация плотности вероятности приводят к одним и тем же оценкам. Мы рассмотрим максимизацию плотности распределения.

Для упрощения вычислений полезно заметить, что $p(y)$ полностью определяется априорной информацией. Поэтому ее можно рассматривать как часть нормировочной постоянной, и

$$p(x|y) = \text{const} \cdot \exp \left[-\frac{1}{2} \{ (x - x_0)' P_0^{-1} (x - x_0) + (y - Cx)' N^{-1} (y - Cx) \} \right]. \quad (12.7)$$

Это выражение нужно максимизировать по x , что можно осуществить приведением к полному квадрату выражения под знаком экспоненты [1]. Другой способ заключается в приравнивании к 0 производной $\ln p(x|y)$:

$$\{P_0^{-1}(x - x_0) - C'N^{-1}(y - Cx)\}|_{x=\hat{x}} = 0, \quad (12.8)$$

или

$$\underbrace{[P_0^{-1} + C'N^{-1}C]}_{P_1^{-1}} \hat{x} = P_0^{-1}x_0 + C'N^{-1}y, \quad (12.9)$$

откуда получаются следующие формулы:

$$\hat{x} = P_1 P_0^{-1} x_0 + P_1 C' N^{-1} y, \quad (12.10a)$$

$$\hat{x} = x_0 + P_1 C' N^{-1} \{y - Cx_0\}, \quad (12.10b)$$

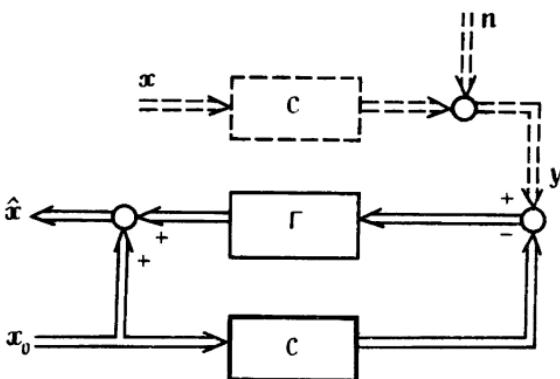
$$\hat{x} = x_0 + P_0 C' [CP_0 C' + N]^{-1} \{y - Cx_0\}, \quad (12.10b)$$

где

$$P_1 = P_0 - P_0 C' [CP_0 C' + N]^{-1} CP_0. \quad (12.11)$$

Эти выражения получаются применением матричных тождеств, приведенных в приложении В. Уравнения (12.10) представлены на фиг. 12.5. Здесь уместны следующие замечания:

а) y — единственная наблюдаемая величина, все остальные задаются заранее. Коэффициент усиления $\Gamma = P_0 C' [CP_0 C' + N]^{-1}$ не зависит от наблюдений и его можно вычислить заранее.



Ф и г. 12.5.

б) Легко показать, что $E[\hat{x}] = x_0$ и
 $E[(\hat{x} - x_0)(\hat{x} - x_0)'] = P_0 C' [CP_0 C' + N]^{-1} CP_0$. (12.12)

Это равенство поясняет смысл формулы (12.11).

в) Укажем некоторые частные случаи, облегчающие понимание сущности метода.

При $C = I$

$$\hat{x} = x_0 + P_0 [P_0 + N]^{-1} \{y - x_0\}.$$

Если существует C^{-1} , то

$$P_1 = P_0 - P_0 [P_0 + N]^{-1} P_0. \quad (12.13)$$

При $N = 0$

$$\begin{aligned} \hat{x} &= x_0 + C^{-1} CP_0 C' [CP_0 C']^{-1} \{y - Cx_0\} = \\ &= x_0 + C^{-1} \{y - Cx_0\} = C^{-1}y. \end{aligned} \quad (12.14)$$

Пример. Рассмотрим простой одномерный случай:
 $x = x$, $y = y$, P , N , C — скаляры. При $C = 1$

$$\hat{x} = x_0 + \frac{P_0}{P_0 + N} \{y - x_0\}, \quad (12.15)$$

$$P_1 = P_0 - \frac{P_0^2}{P_0 + N} = \frac{P_0 N}{P_0 + N} \quad \text{или} \quad \frac{1}{P_1} = \frac{1}{P_0} + \frac{1}{N}. \quad (12.16)$$

Члены P_0 и N напоминают мощность сигнала и шума в задачах теории связи. Если N сравнительно велико, то влияние $y - x_0$ незначительно.

Многошаговый фильтр

Дана априорная информация: уравнения объекта (фиг. 12.6, a)

$$\begin{aligned} \mathbf{x}(i+1) &= \mathbf{A}\mathbf{x}(i) + \mathbf{v}(i), \\ \mathbf{y}(i) &= \mathbf{C}\mathbf{x}(i) + \mathbf{n}(i), \end{aligned} \quad (12.17)$$

$$p\{\mathbf{x}(k) | \mathbf{y}(1), \dots, \mathbf{y}(k)\} = p\{\mathbf{x}(k) | \mathbf{Y}_k\},$$

$$p\{\mathbf{x}(k) | \mathbf{Y}_k\} \Rightarrow \mathcal{E}[\mathbf{x}(k) | \mathbf{Y}_k] = \hat{\mathbf{x}}(k | k),$$

$$\text{cov}[\mathbf{x}(k) | \mathbf{Y}_k] = \mathbf{P}(k),$$

$$p\{\mathbf{v}(i)\} \Rightarrow \mathcal{E}[\mathbf{v}(i)] = \mathbf{0}, \quad \text{cov}[\mathbf{v}(i)] = \mathbf{V},$$

$$p\{\mathbf{n}(i)\} \Rightarrow \mathcal{E}[\mathbf{n}(i)] = \mathbf{0}, \quad \text{cov}[\mathbf{n}(i)] = \mathbf{N},$$

\mathbf{v} и \mathbf{n} независимы.

Через $\hat{\mathbf{x}}(k | k)$ обозначена оценка $\mathbf{x}(k)$ по k наблюдениям $\mathbf{y}(1), \dots, \mathbf{y}(k)$. Такой же способ обозначений можно использовать и в задачах сглаживания и предсказания. В рассматриваемой задаче фильтрации достаточно ограничиться сокращенным обозначением $\hat{\mathbf{x}}(k)$. При этом получаем

$$p\{\mathbf{x}(k+1) | \mathbf{Y}_k\} \Rightarrow \mathcal{E}[\mathbf{x}(k+1) | \mathbf{Y}_k] = \hat{\mathbf{x}}(k),$$

$$\text{cov}[\mathbf{x}(k+1) | \mathbf{Y}_k] = \mathbf{A}\mathbf{P}(k)\mathbf{A}' + \mathbf{V} = \mathbf{Q}(k+1).$$

Можно также вывести выражение для плотности распределения $\mathbf{y}(k+1)$, хотя, как и прежде, оно нам не потребуется:

$$p\{\mathbf{y}(k+1) | \mathbf{Y}_k\} \Rightarrow \mathcal{E}[\mathbf{y}(k+1) | \mathbf{Y}_k] = \mathbf{C}\hat{\mathbf{x}}(k),$$

$$\text{cov}[\mathbf{y}(k+1) | \mathbf{Y}_k] = \mathbf{C}\mathbf{Q}(k+1)\mathbf{C}' + \mathbf{N}.$$

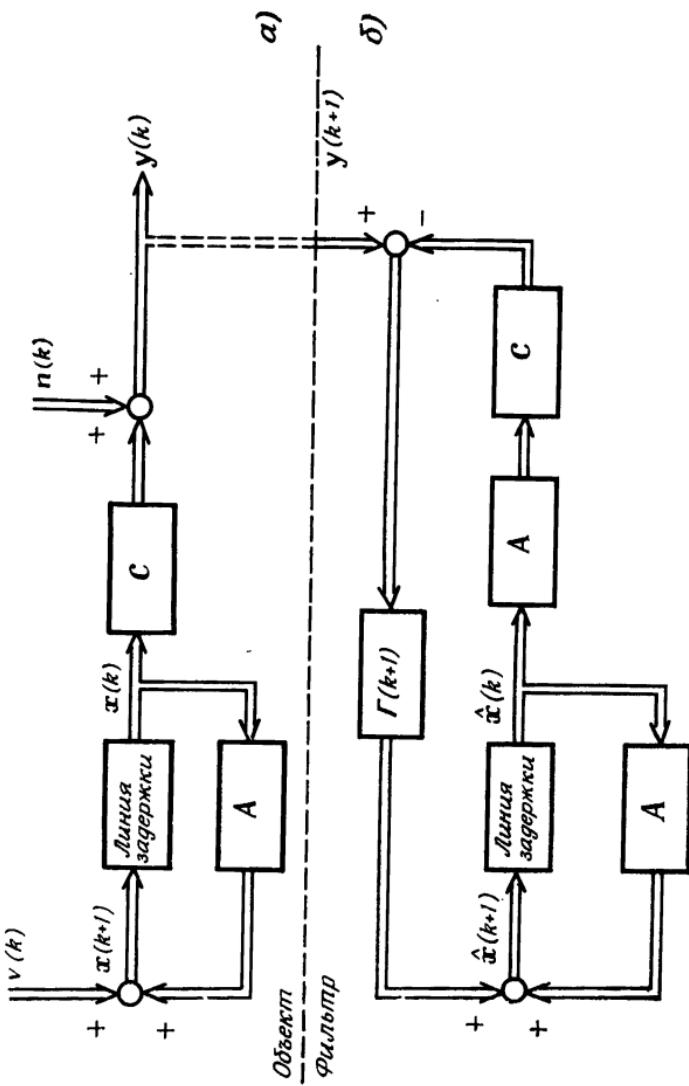
Из-за зависимости от \mathbf{Y}_k требуется несколько более общая формулировка правила Байеса:

$$p(a, b | c) p(c) = p(a, b, c) = p(a | b, c) p(b | c) p(c), \quad (12.18)$$

откуда

$$p(a | b, c) = \frac{p(a, b | c)}{p(b | c)}. \quad (12.19)$$

Фиг. 12.6.



Очевидна прямая аналогия с одношаговым случаем:
одношаговый фильтр

$$p(x|y) = \frac{p(x)p_n(y - Cx)}{p(y)},$$

многошаговый фильтр

$$\begin{aligned} p\{x(k+1)|y(k+1), Y_k\} &= \\ &= \frac{p\{x(k+1)|Y_h\} p_n\{Y(k+1) - Cx(k+1)\}}{p\{y(k+1)|Y_h\}} \end{aligned}$$

и при гауссовских распределениях

$$\begin{aligned} p\{x(k+1)|y(k+1), Y_h\} &= \\ &= \text{const} \cdot \exp \left[-\frac{1}{2} \{ \|x(k+1) - A\hat{x}(k)\|_{Q^{-1}(k+1)}^2 + \right. \\ &\quad \left. + \|y(k+1) - Cx(k+1)\|_{N^{-1}}^2 \} \right]. \quad (12.20) \end{aligned}$$

Приравнивая производную логарифма этой функции к 0, получим для оптимальной оценки $\hat{x}(k+1)$

$$\begin{aligned} Q^{-1}(k+1)[\hat{x}(k+1) - A\hat{x}(k)] - \\ - C'N^{-1}[y(k+1) - C\hat{x}(k+1)] = 0, \quad (12.21) \end{aligned}$$

откуда

$$\begin{aligned} \underbrace{[Q^{-1}(k+1) + C'N^{-1}C]}_{P^{-1}(k+1)} \hat{x}(k+1) &= \\ &= Q^{-1}(k+1) A\hat{x}(k) + C'N^{-1}y(k+1), \quad (12.22) \end{aligned}$$

$$\begin{aligned} \hat{x}(k+1) &= A\hat{x}(k) + Q(k+1)C'[CQ(k+1)C' + N]^{-1} \times \\ &\quad \times [y(k+1) - CA\hat{x}(k)]. \quad (12.23) \end{aligned}$$

Формула (12.23) соответствует (12.10в). Объект и фильтр изображены на фиг. 12.6. Обратите внимание, что часть фильтра подобна объекту.

Для сравнения с фильтром Винера укажем некоторые особенности фильтра Калмана:

- Априорная информация о сигнале (и шуме) должна задаваться в пространстве состояний.
- Уравнения фильтра имеют рекуррентную форму и хорошо подходят для цифровых вычислительных устройств.

в) Алгоритм фильтрации одновременно представляет собой непосредственное описание способа реализации фильтра.

г) Легко распространяется на нестационарные сигналы; это относится и к случаю, когда наблюдения начинаются в произвольный момент времени.

д) Распространение на многомерный случай не представляет труда. Из обширной литературы упомянем работы [13, 18].

Если в модель добавляется доступный измерению векторный входной сигнал $u(k)$:

$$\begin{aligned} \mathbf{x}(k+1) &= \mathbf{A}\mathbf{x}(k) + \mathbf{B}u(k) + \mathbf{v}(k), \\ \mathbf{y}(k) &= \mathbf{C}\mathbf{x}(k) + \mathbf{n}(k), \end{aligned} \quad (12.24)$$

то таким же способом можно получить

$$\begin{aligned} \hat{\mathbf{x}}(k+1) &= \hat{\mathbf{A}}\hat{\mathbf{x}}(k) + \mathbf{B}u(k) + \\ &+ \Gamma(k+1)[\mathbf{y}(k+1) - \mathbf{C}\hat{\mathbf{x}}(k) - \mathbf{C}\mathbf{B}u(k)], \end{aligned} \quad (12.25)$$

$$\begin{aligned} \Gamma(k+1) &= \mathbf{Q}(k+1)\mathbf{C}'[\mathbf{C}\mathbf{Q}(k+1)\mathbf{C}' + \mathbf{N}]^{-1} = \\ &= \mathbf{P}(k+1)\mathbf{C}'\mathbf{N}^{-1}, \end{aligned} \quad (12.26)$$

$$\mathbf{Q}(k+1) = \mathbf{A}\mathbf{P}(k)\mathbf{A}' + \mathbf{V}, \quad (12.27)$$

$$\begin{aligned} \mathbf{P}(k+1) &= [\mathbf{Q}^{-1}(k+1) + \mathbf{C}'\mathbf{N}^{-1}\mathbf{C}]^{-1} = \mathbf{Q}(k+1) - \\ &- \mathbf{Q}(k+1)\mathbf{C}'[\mathbf{C}\mathbf{Q}(k+1)\mathbf{C}' + \mathbf{N}]^{-1}\mathbf{C}\mathbf{Q}(k+1), \end{aligned} \quad (12.28)$$

где $\hat{\mathbf{x}}(0)$ и \mathbf{P}_0 заданы. В этой схеме

$\hat{\mathbf{A}}\hat{\mathbf{x}}(k) + \mathbf{B}u(k)$ — экстраполированное состояние;

$\mathbf{C}[\hat{\mathbf{A}}\hat{\mathbf{x}}(k) + \mathbf{B}u(k)]$ — наблюдение, соответствующее экстраполированному состоянию;

$\mathbf{y}(k+1) - \mathbf{C}\hat{\mathbf{x}}(k) - \mathbf{C}\mathbf{B}u(k)$ — невязка между наблюдавшимся и предсказанным значениями;

$\mathbf{Q}(k+1)$ — априорное значение ковариационной матрицы $\hat{\mathbf{x}}(k+1)$, основанное на k наблюдениях;

$\mathbf{P}(k+1)$ — апостериорное значение ковариационной матрицы $\hat{\mathbf{x}}(k+1)$, основанное на $k+1$ наблюдениях. Наконец, шум объекта может быть отличным от белого и иметь вид $\mathbf{G}\mathbf{v}(k)$, где $\mathbf{v}(k)$ — снова белый шум.

Можно сделать еще несколько замечаний, аналогичных приводившимся выше.

а) Коэффициент усиления $\Gamma(k+1)$ не зависит от наблюдений и может вычисляться заранее для всей процедуры оценивания сразу.

б) В одномерном случае при $C = 1$

$$P(k+1) = P(k+1) = \frac{Q(k+1)N}{Q(k+1)+N}. \quad (12.29)$$

$$\Gamma(k+1) = \Gamma(k+1) = \frac{P(k+1)}{N} = \frac{Q(k+1)}{Q(k+1)+N}. \quad (12.30)$$

в) Зависимость от времени, например, матрицы A не вносит принципиальных изменений, по крайней мере в том случае, когда эта зависимость полностью известна.

г) Сравнение рекуррентных уравнений фильтра Калмана для оценивания состояния с полученными в разд. 7.1 рекуррентными уравнениями для оценивания параметров обнаруживает их близкое соответствие. Исследование связи между этими методами можно найти в работе [3].

д) Оценка линейна относительно наблюдений. Это связано с предположением о нормальности распределения рассматриваемых случайных величин. В этом случае рассматриваемый фильтр оптимален в смысле метода наименьших квадратов. При отличных от нормального распределениях получается фильтр, *оптимальный в классе линейных фильтров*.

Уравнения фильтра можно получить не только исходя из байесовского подхода к оцениванию, как это сделано выше, но и другими способами. Рассмотрим простейший случай. Допустим, что наблюдения задаются уравнением

$$y(i) = x + n(i).$$

Нужно найти оценку

$$\hat{x} = \mathbf{q}'\mathbf{y} = q_1y(1) + \dots + q_ky(k),$$

линейную относительно наблюдений $y(1), \dots, y(k)$ и минимизирующую $E = \mathcal{E}[e^2] = \mathcal{E}[(\hat{x} - x)^2]$. Дифференцируя E до q_i , получаем необходимые условия

$$\frac{\partial E}{\partial q_i} = 2\mathcal{E}[(\hat{x} - x)y(i)] = 0, \quad i = 1, \dots, k. \quad (12.31)$$

Эти уравнения задают так называемые *условия ортогональности*, требующие, чтобы ошибка оценки $\hat{x} - x$ была

ортогональна наблюдениям. Из системы уравнений (12.31) можно определить коэффициенты q_1, \dots, q_k , задающие фильтр. Подобные условия ортогональности можно непосредственно получить и для более общих и сложных случаев. Основываясь на этих выражениях, можно вывести уравнения фильтрации для различных задач.

Уравнения фильтрации можно рассматривать также, учитывая аналогию между структурами объекта и фильтра (модели). Эта аналогия уже обсуждалась для случая винеровской фильтрации [20]. Такая точка зрения использована в теории «наблюдателя» [11, 12].

В предыдущих разделах методы оценивания состояния рассматривались для простых случаев линейного объекта и гауссовского шума. Эти методы применимы и для нелинейных объектов, если осуществить линеаризацию в окрестности известного приближенного решения уравнений объекта. В результате получаются уравнения (12.3) и (12.4), причем последнее можно использовать вместо нелинейного уравнения, если матрицы A, B, C, D составлены из частных производных, соответствующих приближенному решению.

Другой подход называют *расширенным фильтром Калмана*. В этом методе используется линеаризация относительно наилучшей оценки \hat{x} вектора состояния (и измеряемого входного вектора u при его наличии). Если, например, объект описывается уравнениями

$$\begin{aligned} x(i+1) &= f(x(i), i) + v(i), \\ y(i) &= g(x(i), i) + n(i), \end{aligned} \quad (12.32)$$

то уравнения фильтра можно записать в виде

$$\begin{aligned} \hat{x}(k+1 | k+1) &= \hat{x}(k+1 | k) + \\ &+ \Gamma(k+1) [y(k+1) - g(\hat{x}(k+1 | k), k+1)], \end{aligned} \quad (12.33)$$

где, как и раньше,

$$\hat{x}(k+1 | k) = f(\hat{x}(k | k), k) \quad (12.34)$$

является наилучшей оценкой $x(k+1)$ по k наблюдениям.

Раньше уже отмечалась тесная связь между рекуррентными уравнениями для оценивания параметров и состоя-

ния. Для линейного относительно параметров объекта эту связь легко продемонстрировать. Пусть объект описывается соотношением

$$\begin{aligned} y(k) + a_1 y(k-1) + \dots + a_n y(k-n) = \\ = b_1 u(k-1) + \dots + b_n u(k-n) + n(k). \end{aligned} \quad (12.35)$$

Следовательно, уравнения объекта и наблюдений имеют вид

$$\begin{aligned} x(k+1) &= x(k), \\ y(k) &= c'(k)x(k) + n(k), \end{aligned} \quad (12.36)$$

где

$$x(k) = \begin{bmatrix} a_1 \\ \vdots \\ a_n \\ b_1 \\ \vdots \\ b_n \end{bmatrix}, \quad c(k) = \begin{bmatrix} -y(k-1) \\ \vdots \\ -y(k-n) \\ u(k-1) \\ \vdots \\ u(k-n) \end{bmatrix}.$$

При этом задача оценивания параметров может непосредственно решаться с помощью алгоритмов фильтрации, приведенных в этой главе.

12.3. ОБСУЖДЕНИЕ, ПРИЛОЖЕНИЯ И ПРИМЕРЫ

Заслуживают внимания также следующие вопросы:

а) Возможное отсутствие сходимости процесса оценивания, т. е. появление ковариационной матрицы, которая существенно меньше истинных ошибок оценок, и возможные причины этого явления.

б) Исследование невязки $y(k+1) - CA\hat{x}(k)$, которая по своим вероятностным характеристикам должна быть близкой к (дискретному) белому шуму и с помощью которой можно проверять априорную информацию, например, об уровне шума наблюдений.

в) Случай отсутствия шума наблюдений $n(k)$, когда матрица $N = 0$ и, следовательно, не может быть обращена.

г) Случай отличного от белого (коррелированного) шума наблюдений $n(k)$; при этом целесообразно включить

в модель объекта фильтр, преобразующий белый шум в коррелированный.

д) Сигнал и шум коррелированы; здесь также оказывается полезным расширение модели.

е) Случай разных интервалов дискретизации для различных входных и выходных сигналов и (или) последовательности оценок.

ж) Влияние неточности или неполноты априорной информации; частично известные параметры объекта можно включать в схему оценивания в качестве подлежащих определению переменных состояния [5]; возможности построения адаптивных фильтров в таких случаях.

з) Наконец, случай объектов с непрерывными (по времени) сигналами, которые можно описать следующим образом.

Объект:

$$\begin{aligned}\dot{\mathbf{x}} &= \mathbf{Ax} + \mathbf{Gv}, \\ \mathbf{y} &= \mathbf{Cx} + \mathbf{n},\end{aligned}\quad (12.37)$$

$$\mathcal{E}[\mathbf{v}] = \mathbf{0}, \quad \mathcal{E}[\mathbf{v}(t) \mathbf{v}'(\theta)] = \mathbf{V}\delta(t - \theta),$$

$$\mathcal{E}[\mathbf{n}] = \mathbf{0}, \quad \mathcal{E}[\mathbf{n}(t) \mathbf{n}'(\theta)] = \mathbf{N}\delta(t - \theta),$$

$$\mathcal{E}[\mathbf{v}(t) \mathbf{n}'(\theta)] = \mathbf{0}.$$

Разумеется, использование непрерывных белых шумов \mathbf{v} и \mathbf{n} требует определенной осторожности по ряду причин математического и физического характера.

Фильтр:

$$\dot{\hat{\mathbf{x}}} = \mathbf{A}\hat{\mathbf{x}} + \mathbf{\Gamma}[\mathbf{y} - \mathbf{C}\hat{\mathbf{x}}], \quad (12.38)$$

$$\mathbf{\Gamma} = \mathbf{P}\mathbf{C}'\mathbf{N}^{-1},$$

$$\dot{\mathbf{P}} = \mathbf{AP} + \mathbf{PA}' - \mathbf{PC}'\mathbf{N}^{-1}\mathbf{CP} + \mathbf{GVG}',$$

$$\hat{\mathbf{x}}(0) = \hat{\mathbf{x}}_0, \quad \mathbf{P}(0) = \mathbf{P}_0.$$

Пример. Для иллюстрации изложенных идей рассмотрим очень простой пример. Стационарный случайный сигнал $x(k)$ с известными статистическими свойствами искажается стационарным белым шумом. Необходимо по возможности наилучшим образом оценить $x(k)$. Объект

и наблюдения определяются формулами

$$x(k+1) = ax(k) + v(k),$$

$$y(k) = x(k) + n(k),$$

$$\mathcal{E}[v(i)] = 0, \quad \mathcal{E}[v(i)v(j)] = V\delta_{ij},$$

$$\mathcal{E}[n(i)] = 0, \quad \mathcal{E}[n(i)n(j)] = N\delta_{ij},$$

$$\mathcal{E}[n(i)v(j)] = 0.$$

Фильтр описывается соотношениями

$$\hat{x}(k+1) = \hat{ax}(k) + \gamma(k+1)\{y(k+1) - \hat{ax}(k)\},$$

$$\gamma(k+1) = \frac{Q(k+1)}{Q(k+1)+N}, \quad x(0) = x_0,$$

$$Q(k+1) = a^2 P(k) + V, \quad P(0) = P_0,$$

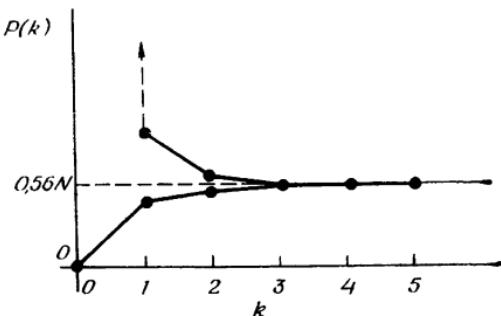
$$P(k+1) = Q(k+1) - \frac{Q^2(k+1)}{Q(k+1)+N} = \frac{Q(k+1)N}{Q(k+1)+N}.$$

Уравнения фильтра можно также записать в виде

$$\begin{aligned} \hat{x}(k+1) &= \underbrace{\frac{N}{Q(k+1)+N}}_{\alpha(k+1)} \hat{ax}(k) + \underbrace{\frac{Q(k+1)}{Q(k+1)+N}}_{\beta(k+1)} y(k+1), \\ \alpha(k+1) + \beta(k+1) &= 1. \end{aligned}$$

Величины $P(k+1)$, $Q(k+1)$ и $\gamma(k+1)$ заранее полностью определяются априорной информацией. Рассмотрим два случая:

Априорная информация отсутствует; например, $x_0 = 0$ и $P_0 = \infty$, $N = V$, $a = \sqrt{0,5} \approx 0,7$				Достоверная априорная информация; точное значение $x_0 = 0$ и $P_0 = 0$, $N = V$, $a = \sqrt{0,5} \approx 0,7$			
k	$Q(k)$	$\gamma(k)$	$P(k)$	k	$Q(k)$	$\gamma(k)$	$P(k)$
0	∞	1	∞	0	N	0,5	0
1	$1,5N$	0,6	$0,6N$	1	$1,25N$	0,555	$0,5N$
2	$1,3N$	0,565	$0,565N$	2	$1,278N$	0,56	$0,555N$
3	$1,283N$	0,56	—	3	$1,28N$	0,56	$0,56N$
4	$1,28N$	0,56	—	4	—	—	—
5	—	—	—	5	—	—	—



Ф и г. 12.7.

После нескольких итераций $\gamma(k)$ становится постоянной, и, следовательно, фильтр будет стационарным. На фиг. 12.7 показана апостериорная ковариация (неопределенность) оценки. От значений ∞ (соответственно от 0) она быстро приближается к асимптотическому значению. Скорость установления стационарного состояния зависит от параметров. Условием стационарности является

$$P(k) = P(k-1) = P.$$

Следовательно,

$$\begin{aligned} \frac{1}{P(k)} &= \frac{1}{Q(k)} + \frac{1}{N} = \frac{1}{a^2 P(k-1) + V} + \frac{1}{N}, \\ (a^2 P + V) N &= P N + P(a^2 P + V), \\ P^2 + \frac{N + V - a^2 N}{a^2} P - \frac{V N}{a^2} &= 0, \end{aligned}$$

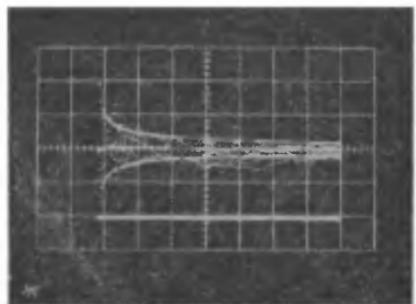
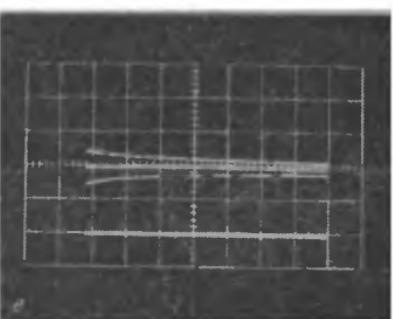
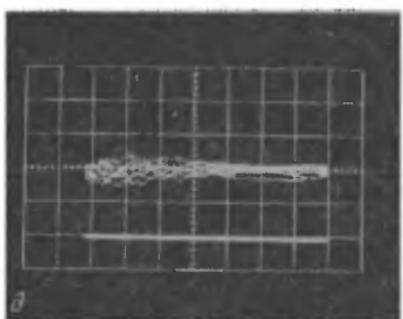
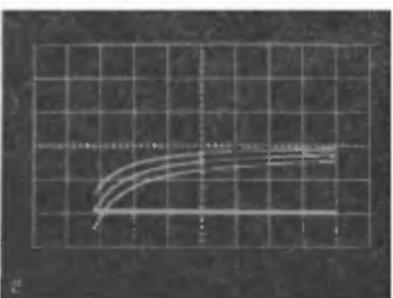
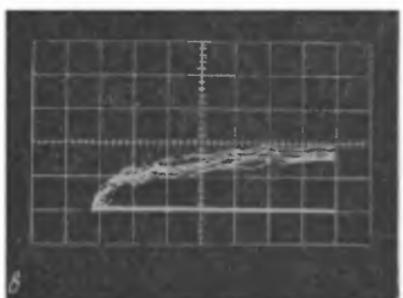
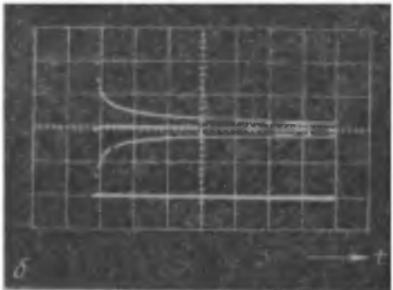
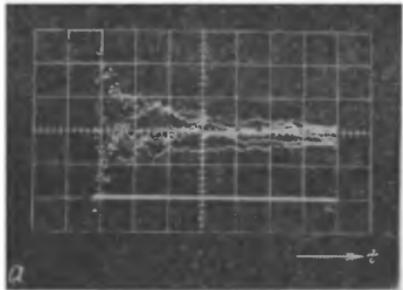
откуда можно найти P . В рассматриваемом случае

$$P^2 + 3NP - 2N^2 = 0.$$

Положительный корень $P = 0,56N$.

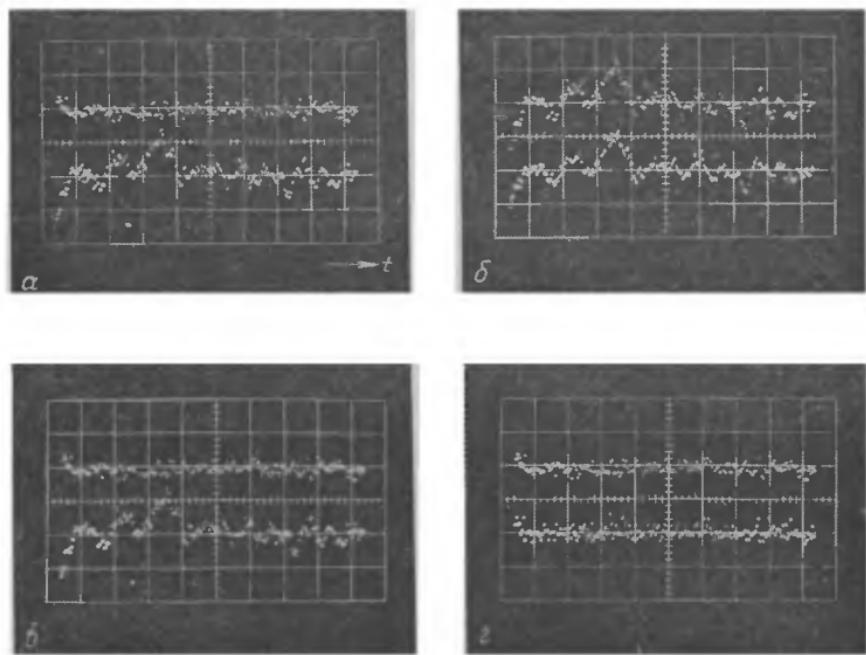
Примеры. На фиг. 12.8—12.10 приведено несколько элементарных примеров, иллюстрирующих эти алгоритмы фильтрации¹⁾. На фиг. 12.8 представлены осциллограммы для статистического случая, т. е. оценивания $\hat{x}(k)$ по дискретным результатам наблюдений, искаженным белым шумом. Осциллограммы иллюстрируют влияние принятой априорной информации относительно оцениваемых вели-

¹⁾ Эти кривые получены при моделировании на вычислительной машине PDP-8, выполненнем Дж. Дж. Шонмакером в Технологическом университете Эйндховена (Голландия).



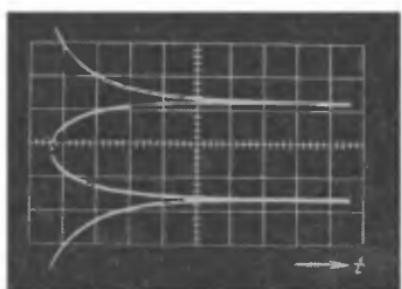
Ф и г. 12.8.

чин. Графики *а* и *б* соответствуют предположению $\hat{x}(0) = 0$, что не корректно, и большому значению $P(0)$, показывающему высокую степень недостоверности принятого значения $\hat{x}(0)$. На фиг. 12.8, *а* приведено 7 реализаций, соответствующих различным последовательностям шума, а на фиг. 12.8, *б* показано изменение стандартного отклонения относительно математического ожидания $\hat{x}(k)$. Кривые *в* и *г* соответствуют случаю, когда $\hat{x}(0) = 0$ и $P(0)$ выбрано малым, т. е. принятое значение $x(0)$ необоснованно считается достаточно достоверным. На фиг. 12.8, *в* также показаны 7 реализаций, а на фиг. 12.8, *г* — стандартное отклонение оценки. Видно, что оценка приближается к истинному значению медленно из-за неправильного выбора $\hat{x}(0)$ и $P(0)$. В случаях *д* и *е* значение $\hat{x}(0)$ было выбрано правильно и $P(0)$ мало.



Фиг. 12.9.

а — шум объекта $u(k)$ и оцениваемый сигнал $x(k)$; *б* — сигнал $x(k)$ и оценка $\hat{x}(k)$; *в* — ошибка $\hat{x}(k) - x(k)$ и оцениваемый сигнал $x(k)$; *г* — ошибка $\hat{x}(k) - x(k)$ и реализация аддитивного шума, использованная в данном примере.



Фиг. 12.10.

На фиг. 12.8, ∂ показаны 7 реализаций, на фиг. 12.8, e — стандартное отклонение. Случай \mathcal{J} соответствует объединению кривых a и b .

На фиг. 12.9 и 12.10 представлены осциллограммы для одномерного динамического случая — оценивания низкочастотного сигнала по наблюдениям, проводимым при наличии белого

шума. На фиг. 12.10 показано влияние априорной неопределенности относительно $\hat{x}(0)$. Фиг. 12.11 соответствует статическому случаю, но уже для непрерывных сигналов. Верхний график представляет результаты фильтрации обычным простым низкочастотным фильтром (первого порядка, с постоянными коэффициентами), внизу показан выход фильтра Калмана первого порядка. Входные сигналы обоих фильтров одинаковы. Объект и наблюдения задаются уравнениями

$$\dot{x}(t) = 0, \quad y(t) = x(t) + n(t),$$

где $n(t)$ — стационарный белый шум с нулевым средним и дисперсией¹⁾ $N = \sigma_n^2$. Фильтр задается уравнениями

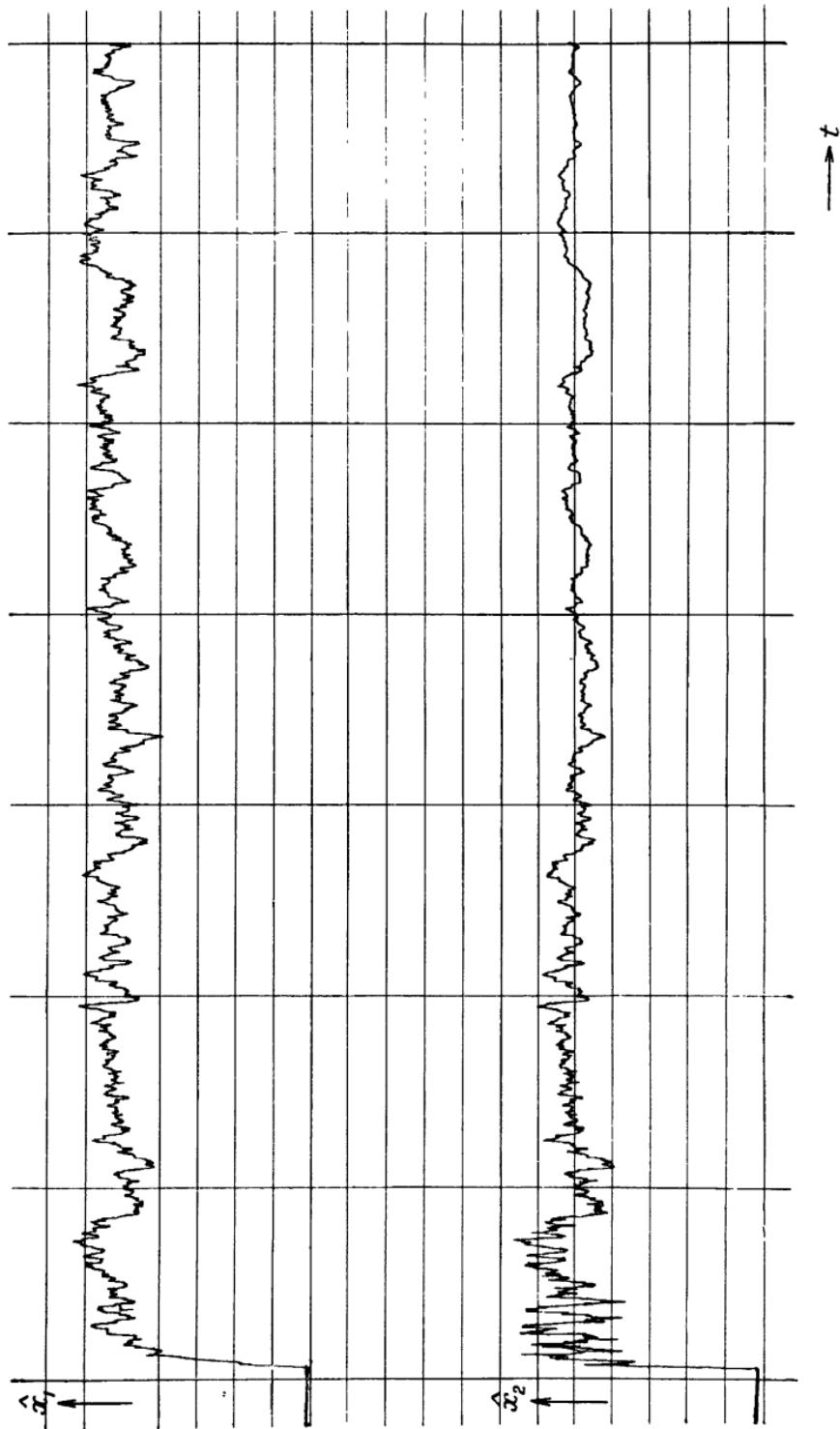
$$\begin{aligned} \dot{\hat{x}}(t) &= \frac{P(t)}{\sigma_n^2} \{y(t) - \hat{x}(t)\}, \quad \hat{x}(0) = \hat{x}_0, \\ \dot{P}(t) &= -\frac{P^2(t)}{\sigma_n^2}, \quad P(0) = P_0. \end{aligned}$$

Это простое нелинейное дифференциальное уравнение легко решается:

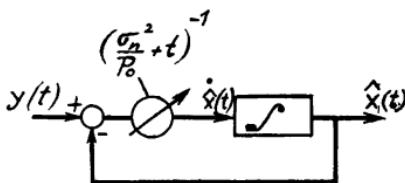
$$\frac{dP}{P^2} = -\frac{dt}{\sigma_n^2}, \quad -\frac{1}{P} = -\frac{t}{\sigma_n^2} + C = -\frac{t}{\sigma_n^2} - \frac{1}{P_0},$$

$$P(t) = \frac{\sigma_n^2}{\frac{\sigma_n^2}{P_0} + t}.$$

¹⁾ Имеется в виду интенсивность белого шума.— Прим. перев.



Ф и г. 12.11.



Ф и г. 12.12.

Итак, фильтр можно определить одним из следующих уравнений:

$$\dot{\hat{x}}(t) = \frac{1}{\frac{\sigma_n^2}{P_0} + t} \{y(t) - \hat{x}(t)\},$$

$$\left(\frac{\sigma_n^2}{P_0} + t\right) \dot{\hat{x}}(t) + \hat{x}(t) = y(t).$$

Коэффициент усиления фильтра Калмана зависит от времени (фиг. 12.12). Этот фильтр также можно рассматривать как фильтр низких частот с возрастающей постоянной времени. Отметим, что, согласно фиг. 12.10, дисперсия выхода фильтра Калмана убывает со временем, тогда как дисперсия фильтра с постоянными коэффициентами стремится к постоянному уровню.

Связь между фильтрами Винера и Калмана. Поскольку оба фильтра оптимальны в указанном выше смысле, то при соответствующих условиях они должны давать одинаковые результаты. Проиллюстрируем это на простом непрерывном случае.

Пример. Уравнения объекта и наблюдений имеют вид

$$\begin{aligned}\dot{x}(t) &= -ax(t) + v(t), \quad a > 0, \\ y(t) &= x(t) + n(t),\end{aligned}$$

где $v(t)$ и $n(t)$ — некоррелированные стационарные белые шумы с нулевыми средними и дисперсиями $V = \sigma_v^2$ и $N = \sigma_n^2$.

Фильтр Калмана задается уравнениями

$$\dot{\hat{x}}(t) = -a\hat{x}(t) + \frac{P(t)}{\sigma_n^2} \{y(t) - \hat{x}(t)\},$$

$$\dot{P}(t) = -2aP(t) - \frac{P^2(t)}{\sigma_n^2} + \sigma_v^2.$$

При $t \rightarrow \infty$ следует ожидать, что фильтр с переменными параметрами становится стационарным. Следовательно, $P(t) = P$, $\dot{P}(t) = 0$ и

$$P^2 + 2a\sigma_n^2 P + \sigma_n^2 \sigma_v^2 = 0$$

или

$$P_{1,2} = -a\sigma_n^2 \pm \sqrt{a^2\sigma_n^4 + \sigma_n^2\sigma_v^2}.$$

Величина P должна быть неотрицательной. Если положить $\sigma_v^2/\sigma_n^2 = \lambda^2$, то для стационарного фильтра можно записать

$$\dot{\hat{x}}(t) + \{a - a + \sqrt{a^2 + \lambda^2}\} \hat{x}(t) = \{-a + \sqrt{a^2 + \lambda^2}\} y(t),$$

или

$$H_K(s) = \frac{\hat{X}(s)}{Y(s)} = \frac{-a + \sqrt{a^2 + \lambda^2}}{s + \sqrt{a^2 + \lambda^2}}.$$

Фильтр Винера определяется уравнением (12.3), в котором в данном простом случае следует положить

$$\Phi_{xx}(s) = \frac{\sigma}{(a+s)(a-s)},$$

$$\Phi_{nn}(s) = \sigma_n^2,$$

$$\Phi_{yy}(s) = \sigma_n^2 + \frac{\sigma_v^2}{(a+s)(a-s)} = \sigma_n^2 \frac{(\sqrt{a^2 + \lambda^2} + s)(\sqrt{a^2 + \lambda^2} - s)}{(a+s)(a-s)}.$$

Следовательно,

$$\frac{1}{\Phi_{yy}^+(s)} = \frac{a+s}{\sigma_n(\sqrt{a^2 + \lambda^2} + s)},$$

$$G(s) = \frac{\Phi_{xx}(s)}{\Phi_{yy}^+(s)} = \frac{\sigma_v^2}{(a+s)(a-s)} \frac{a-s}{\sigma_n(\sqrt{a^2 + \lambda^2} - s)},$$

$$[G(s)]_+ = \frac{r}{a+s},$$

$$r = \lim_{s \rightarrow -a} (a+s) G(s) = \frac{\sigma_v^2}{\sigma_n} \frac{1}{a + \sqrt{a^2 + \lambda^2}},$$

где r — вычет при $s = -a$.

Отсюда

$$H_W(s) = \lambda^2 \frac{1}{a + \sqrt{a^2 + \lambda^2}} \frac{1}{s + \sqrt{a^2 + \lambda^2}} = \frac{-a + \sqrt{a^2 + \lambda^2}}{s + \sqrt{a^2 + \lambda^2}}.$$

Эквивалентность обоих фильтров следует из того, что $H_K(s) = H_W(s)$.

Эта глава является лишь вводным обзором для следующей; подробнее с упомянутыми процедурами фильтра можно познакомиться, обратившись к литературе (см., например, [15]). Среди статей по применениюм фильтра Калмана в навигации можно назвать [10, 16, 19, 24]. Приложения к задачам связи рассмотрены в [4, 21]. Ссылки на работы по приложениям к управлению в промышленности и к биологическим объектам даны в гл. 14.

ЛИТЕРАТУРА

1. Aoki M. (см. литературу к гл. 4).
2. Åström K. J. (см. литературу к гл. 4).
3. Ho V. C., *The Methods of Least Squares and Optimal Filtering*, Memo 3329-PR, RAND Corp., 1962.
4. Irwin J. D., O'Neal J. B., *The Design of Optimum dpcm Encoding Systems via the Kalman Predictor*, Joint Autom. Control Conf., 1968, pp. 130—136.
5. Jazwinsky A. H., *Stochastic Processes and Filtering Theory*, Academic Press, N.Y., 1970.
6. Jury E. I., *Theory and Application of the Z-Transform Method*, Wiley, N.Y., 1964.
7. Kailath T., Frost P. (см. литературу к гл. 4).
8. Kalman R. E., *A New Approach to Linear Filtering and Prediction Problems*, *J. Basic Eng.*, 82D, 35—45 (1960).
9. Kalman R. E., Bucy R. S., *New Results in Linear Filtering and Prediction Theory*, *J. Basic Eng.*, 83D, 95—108 (1961); есть русский перевод: Калман Р., Бьюси Р., *Новые результаты в линейной фильтрации и теории предсказания*, Труды Амер. об-ва инженеров-механиков, *Техническая механика*, 83, № 1 (1961).
10. Leondes C. T., Ed., *Theory and Applications of Kalman Filtering*, AGARDO-graph 139, 1970.
11. Luenberger D. G., *Observers for Multivariable Systems*, *IEEE Trans. Autom. Control*, AC-11, 190—197 (1966).
12. Luenberger D. G., *An Introduction to Observers*, *IEEE Trans. Autom. Control*, AC-16, 596—602 (1971).
13. Meditch J. S., *Stochastic Optimal Linear Estimation and Control*, McGraw-Hill, N.Y., 1969.
14. Ragazzini J., Franklin G., *Sampled-Data Control Systems*, McGraw-Hill, N.Y., 1958.
15. Rhodes I. B., *A Tutorial Introduction to Estimation and Filtering*, *IEEE Trans. Autom. Control*, AC-16, 688—706 (1971).
16. Richman J., Thau F., *Elements of the Kalman Filter Theory*; Part I — Theory, Part II — Applications, *Techn. News Bull.*, General Precision, Aerospace Group, 9, № 3, 4 (1966).
17. Sage A. P., *Optimum Systems Control*, Englewood Cliffs, Prentice-Hall, N.J., 1968.
18. Sage A. P., Melsa J. L., *Estimation Theory with Applications to Communications and Control*, McGraw-Hill, N.Y., 1971.

19. Schmidt S. F. et al., Case Study of Kalman Filtering in the C-5 Aircraft Navigation System, *Case Studies in System Control*, Workshop JACC, IEEE Group on Autom. Control, 1968.
20. Smith O. J. M., *Feedback Control Systems*, McGraw-Hill, N.Y., 1958.
21. Snijder D. L., *The State-Variable Approach to Continuous Estimation with Applications to Analog Communication Theory*, MIT Press, Cambridge, Mass., Research Monogr. 51, 1969.
22. Соловьев В. В., *Статистическая динамика линейных систем автоматического управления*, Физматгиз, 1960.
23. Sorenson H. W., Least-Squares Estimation: from Gauss to Kalman, *IEEE Spectrum*, 7, 63–68 (1970).
24. Toda N. F. et al., Region of Kalman Filter Convergence for Several Autonomous Navigation Modes, *AIEE J.*, 7, № 4, 1969.
25. Wiener N., *The Extrapolation, Interpolation and Smoothing of Stationary Time Series*, Wiley, N.Y., 1949.

ДОПОЛНИТЕЛЬНАЯ ЛИТЕРАТУРА

Athans M., Wishner R. P., Bertolini A., Suboptimal State Estimation for Continuous-Time Nonlinear Systems from Discrete Noisy Measurements, *IEEE Trans. Autom. Control*, AC-13, 504–514 (1968)-
Breipohl A. M., Kalman Filtering and Its Application to Reliability, *IEEE Trans. Reliability*, R-18, 127–130 (1969).

Bryson A. E., Jr., Ho Y. C., *Applied Optimal Control; Optimization, Estimation and Control*, Waltham, Mass., Blaisdell, 1969; есть русский перевод: Брайсон А., Хо Ю.-ши, *Прикладная теория оптимального управления*, изд-во «Мир», 1972.

Bryson A. E., Jr., Johansen D. E., Linear Filtering for Time-Varying Systems Using Measurements Containing Colored Noise, *IEEE Trans. Autom. Control*, AC-10, 4–10 (1965).

Bucy R. S., Linear and Nonlinear Filtering (Invited Paper), *Proc. IEEE*, 58, 854–864 (1970).

Bucy R. S., Joseph P. D., *Filtering for Stochastic Processes with Applications to Guidance*, Interscience, N.Y., 1968.

Deyst J. J., Jr., Price C. F., Conditions for Asymptotic Stability of the Discrete Minimum-Variance Linear Estimator (Short Paper), *IEEE Trans. Autom. Control*, AC-13, 702–705 (1968).

Deutsch R. (см. литературу к гл. 5).

Fischer J. R., Optimal Nonlinear Filtering, in: «*Advances in Control Systems*», Leondes C. T., ed., 1967, pp. 197–300.

Griffin R. E., Sage A. P., Large and Small Scale Sensitivity Analysis of Optimum Estimation Algorithms, *IEEE Trans. Automatic Control*, AC-13, 320–328 (1968).

Griffin R. E., Sage A. P., Sensitivity Analysis of Discrete Filtering and Smoothing Algorithms, *AIAA J.*, 7, 1890–1897 (1969).

Gura I. A., An Algebraic Solution of the State Estimation Problem, *AIAA J.*, 7, 1242–1247 (1969).

Heffer H., The Effect of Erroneous Models on the Kalman Filter Response (Short Paper), *IEEE Trans. Autom. Control*, AC-11, 541–543 (1966).

Huddle J. R., Wismer D. A., Degradation of Linear Filter Performance due to Modeling Error, *IEEE Trans. Autom. Control*, AC-13, 421—423 (1968).

Klinger A. (см. литературу к гл. 7).

Kwakernaak H., Optimal Filtering in Linear with Time Delay, *IEEE Trans. Autom. Control*, AC-12, 169—173 (1967).

Lee R. C. K. (см. литературу к гл. 4).

Leondes C. T., Novak L. M., Optimal Minimal-Order Observers for Discrete-Time Systems — a Unified Theory, Fifth IFAC Congress. Paris, Paper 38.6, 1972.

Liebel P. B., An Introduction to Optimal Estimation, Addison-Wesley, Reading, Mass., 1967.

Magne D. Q., A Solution of the Smoothing Problem for Linear Dynamic Systems, *Automatica*, 44, 73—92 (1966).

Nahi N. (см. литературу к гл. 5).

Nash R. A., Tuteur F. B., The Effect of Uncertainties in the Noise Covariance Matrices on the Maximum Likelihood Estimate of a Vector (Short Paper), *IEEE Trans. Autom. Control*, AC-13, 86—88 (1968).

Nishimura T., On the apriori Information in Sequential Estimation Problems, *IEEE Trans. Autom. Control*, AC-11, 197—204 (1966); см. также AC-12, 123 (1966).

Ohap R. F., Stubberud A. R. (см. литературу к гл. 13).

Phillipson G. A., Mitter S. K., State Identification of a Class of Linear Distributed Systems, Fourth IFAC Congress, Warsaw, Paper 12.3, 1969.

Schwartz L., Stear E. B., A Computational Comparison of Several Nonlinear Filters (Short Paper), *IEEE Trans. Autom. Control*, AC-13, 83—86 (1968).

Shaw L. G., Optimum Stochastic Control, in: Disciplines and Techniques of Systems Control, Peschon J., ed., Blaisdell, N.Y., 1965, pp. 125—185.

Sorensen H. W., Kalman Filtering Techniques, in: Advances in Control Systems, Leondes C. T., ed., Vol. 3, Academic Press, N.Y., 1966, pp. 219—292.

Tait G. R., Bélanger P. R., A Comparison of Some Parameter Identification Schemes Using First and Second Order Extended Kalman-Bucy Filters and Sensitivity Functions, Second IFAC Symp. «Identification and Process Parameter Estimation», Prague, Paper 10.3, 1970.

Wells C. H., An Approximation of the Kalman Filter Equations (Correspondence), *IEEE, Trans. Autom. Control*, AC-13, 445 (1968).

ГЛАВА 13

ОЦЕНИВАНИЕ ПАРАМЕТРОВ И СОСТОЯНИЯ

Многие вопросы современной теории управления концентрируются вокруг понятия состояния. Для применения этой теории нужно иметь возможность явным или неявным образом измерять или оценивать переменные, характеризующие состояние. Этого можно достичь методами, изложенными в предыдущей главе, если известны как структура, так и параметры объекта. Если параметры не известны, то во многих случаях возникает необходимость совместного оценивания параметров и состояния. В разд. 13.1 показано, что такие задачи совместного оценивания параметров и состояния неизбежно имеют нелинейный характер.

Допущение, что шум является аддитивным, позволяет получить функцию ошибки, как показано в разд. 13.2. В разд. 13.3 отмечается, что минимизация такой функции ошибки может привести к известной двухточечной краевой задаче, решение которой методами квазилинеаризации и инвариантного погружения демонстрируется в следующем разделе. Глава завершается примером, который может служить для читателя введением к обширному числу публикаций по (нелинейным) методам оценивания.

13.1. НЕЛИНЕЙНЫЙ ХАРАКТЕР ЗАДАЧИ

Прямой подход к задаче, рекомендуемый рядом авторов, предполагает расширение вектора состояния за счет неизвестных параметров. По-прежнему объект описывается соотношениями

$$\begin{aligned}\dot{x} &= f(x, u, a, v, t), \\ y &= g(x, u, c, n, t)\end{aligned}\tag{13.1}$$

[см. (4.64)]. Вектор параметров объекта a , если он известен или можно ожидать, что он постоянен, интерпретируется как функция времени, задаваемая уравнением

$$\dot{a} = 0. \quad (13.2)$$

Отсюда следует система уравнений вида

$$\begin{bmatrix} \dot{x} \\ \dot{a} \end{bmatrix} = \begin{bmatrix} f(x, u, a, v, t) \\ 0 \end{bmatrix}. \quad (13.3)$$

Нетрудно понять, что если даже динамика объекта линейна, его уравнение невозможно преобразовать к форме

$$\begin{bmatrix} \dot{x} \\ \dot{a} \end{bmatrix} = M \begin{bmatrix} x \\ a \end{bmatrix}, \quad (13.4)$$

где M не зависит ни от x , ни от a . Таким образом, даже для линейного объекта, линейно зависящего от параметров, задача совместного оценивания параметров и состояния *нелинейна* относительно вектора параметров и состояния $[x', a']$. Это означает, что все подходы к этой задаче должны быть типа метода настраиваемой модели, т. е. иметь итеративный характер.

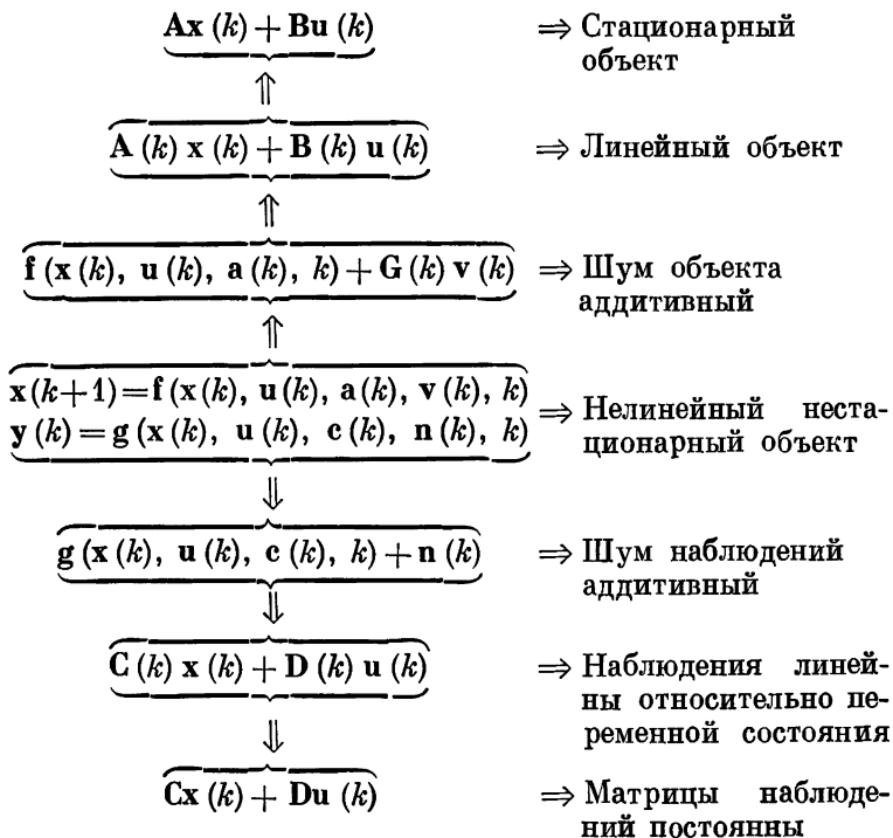
Уравнение (13.1) описывает в общем виде *нелинейный непрерывный* объект. Для выборочных процессов соответствующее представление имеет вид

$$\begin{aligned} x(k+1) &= f(x(k), u(k), a(k), v(k), k), \\ y(k) &= g(x(k), u(k), c(k), n(k), k). \end{aligned} \quad (13.5)$$

Основываясь на априорной информации об объекте и особенностях его функционирования или управления им, можно упростить это представление. Подобные упрощения часто встречаются в работах, относящихся к этому кругу задач, и служат обычно для облегчения понимания процесса оценивания. В результате можно прийти ко многим способам описания объекта ¹⁾, некоторые из которых

¹⁾ Во многих случаях эквивалентный непрерывный вариант описания объекта можно получить из дискретного варианта с помощью корректно выполненного предельного перехода с учетом математических затруднений, связанных с представлением влияния источников белого шума (Ито — Стратонович).

схематически изображены ниже и иллюстрируются фиг. 13.1.



Среди других возможных вариантов отметим следующие:

$u(k) = 0 \Rightarrow$ Отсутствует измеряемый входной сигнал.

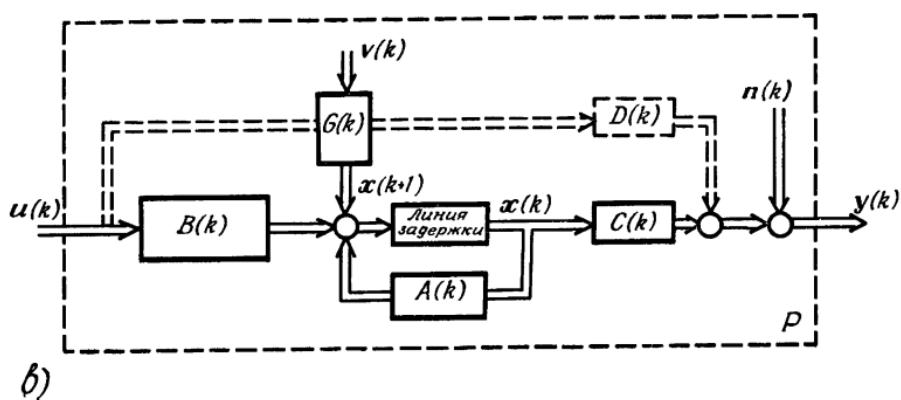
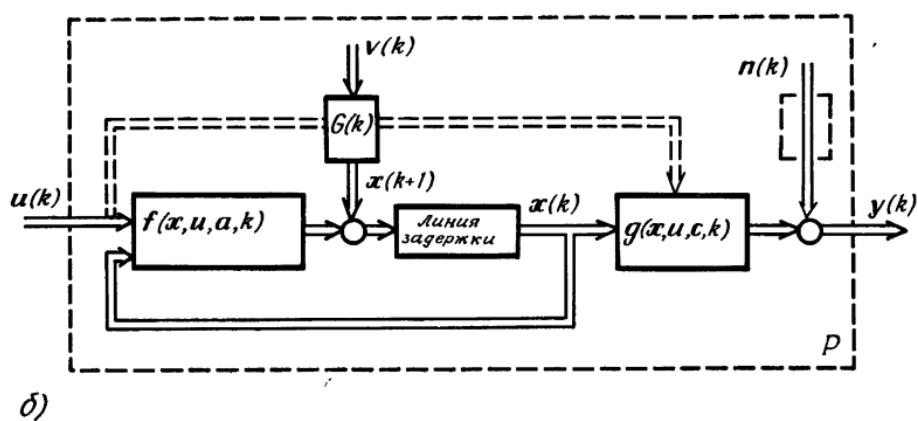
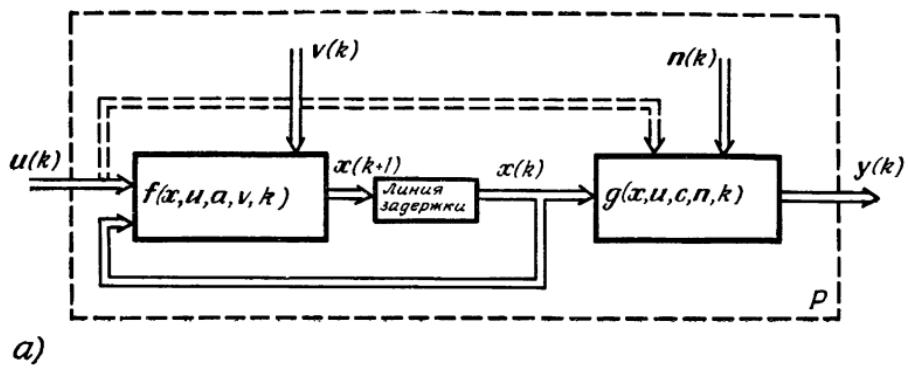
$\left. \begin{array}{l} \mathbf{a}(k) = \mathbf{a}, \\ \mathbf{c}(k) = \mathbf{c} \end{array} \right\} \Rightarrow$ Объект стационарен.

Условие, что шум является *аддитивным*, очень удобно. Например, для простого примера

$$\mathbf{y}(k) = \mathbf{g}(\mathbf{x}(k)) + \mathbf{n}(k)$$

получаем, что

$$p(\mathbf{y}(k) | \mathbf{x}(k)) = p(\mathbf{y}(k) - \mathbf{g}(\mathbf{x}(k))) = p(\mathbf{n}(k)). \quad (13.6)$$



Ф и г. 13.1.

Следовательно, плотность условного распределения вероятностей равна плотности вероятности шума (известной или предполагаемой). В частности, если эта плотность гауссовская, такой подход позволяет получить приемлемые выражения рассматриваемых функций, как это уже показывалось в предыдущей главе и вновь будет отмечено в дальнейшем.

13.2. ВЫВОД ФУНКЦИИ ОШИБКИ

Будем пока рассматривать вектор параметров и состояния x , который может включать переменные состояния и параметры, или только переменные состояния (для нелинейных объектов); или только параметры (для объектов, нелинейно зависящих от параметров). При этом все эти разновидности задачи оценивания можно рассматривать совместно. Кроме того, при таком подходе многочисленные публикации по нелинейной калмановской фильтрации, например по использованию расширенного фильтра Калмана, относятся к нашей задаче.

Как и раньше, к задаче можно подходить с разных точек зрения, а именно используя байесовский подход, оценивание по методу максимального правдоподобия или метод наименьших квадратов. По-прежнему можно изучать связь между различными оценками в частных случаях (четные функции штрафа; гауссовские плотности распределения; белый шум), как это делалось в гл. 5. Во всех случаях рассматриваемые методы сводятся к минимизации функции ошибки E относительно вектора параметров и состояния $x(i)$ на интервале $i = 0, \dots, k$. Прежде всего кратко изложим два способа вывода такой функции ошибки:

A. Объект описывается следующими уравнениями:

$$x(i+1) = f(x(i), i) + G(i)v(i), \quad (13.7)$$

$$y(i) = g(x(i), i) + n(i). \quad (13.8)$$

Если априорная информация об объекте и шуме наблюдений отсутствует, можно использовать следующий прямой подход [7]. Определим величины

$$e_1(i) = x^*(i+1) - f(x^*(i), i), \quad (13.9)$$

$$e_2(i) = y(i) - g(x^*(i), i), \quad (13.10)$$

где $\mathbf{x}^*(i)$, $i = 0, 1, \dots, k$, представляет собой оценку траектории объекта. Если бы $\mathbf{x}^*(i)$ была истинной траекторией, $\mathbf{v}(i) \equiv \mathbf{0}$ и $\mathbf{n}(i) \equiv \mathbf{0}$ при $i = 0, \dots, k$, то $\mathbf{e}_1(i) \equiv \mathbf{e}_2(i) \equiv \mathbf{0}$. Поэтому можно выбрать следующую подлежащую минимизации функцию ошибки:

$$E = \sum_{i=0}^k [\mathbf{e}'_1(i) \mathbf{Q} \mathbf{e}_1(i) + \mathbf{e}'_2(i) \mathbf{R} \mathbf{e}_2(i)]. \quad (13.11)$$

Здесь \mathbf{Q} и \mathbf{R} — неотрицательно определенные весовые матрицы; очевидно, что E представляет собой взвешенный вариант критерия наименьших квадратов. Оценка $\hat{\mathbf{x}}(i)$, обеспечивающая минимум E , является оптимальной в указанном смысле.

Б. Рассматриваемый объект по-прежнему задается уравнениями (13.7) и (13.8). Предположим теперь [5], что $\mathbf{v}(i)$ и $\mathbf{n}(i)$ — гауссовские векторные случайные величины с нулевыми средними значениями и ковариационными матрицами вида

$$\begin{aligned} \mathcal{E}[\mathbf{v}(i) \mathbf{v}'(j)] &= \mathbf{V}(i) \delta_{ij}, \\ \mathcal{E}[\mathbf{n}(i) \mathbf{n}'(j)] &= \mathbf{N}(i) \delta_{ij}, \\ \mathcal{E}[\mathbf{v}(i) \mathbf{n}'(j)] &= \mathbf{0}, \end{aligned} \quad (13.12)$$

т. е., другими словами, последовательности шума считаются белыми.

Байесовский подход. Опять используются обозначения

$$\mathbf{X}_k = [\mathbf{x}(0), \dots, \mathbf{x}(k)],$$

$$\mathbf{Y}_k = [\mathbf{y}(0), \dots, \mathbf{y}(k)].$$

Нужно снова найти плотность условного распределения $P(\mathbf{X}_k | \mathbf{Y}_k)$ величины \mathbf{X}_k относительно наблюдений \mathbf{Y}_k . По правилу Байеса

$$p(\mathbf{X}_k | \mathbf{Y}_k) = \frac{p(\mathbf{Y}_k | \mathbf{X}_k) p(\mathbf{X}_k)}{p(\mathbf{Y}_k)}. \quad (13.13)$$

Как и в гл. 5, можно использовать функцию штрафа или потерь для сведения этой апостериорной плотности распределения к простому набору числовых значений, т. е. для получения оценки по минимуму штрафа или потерь. С другой стороны, можно ограничиться определением таких числовых значений оценки, для которых плотность услов-

ногого распределения вероятностей (13.13) максимальна. При гауссовских плотностях распределения и четной функции штрафа или потерь оба метода приводят к одним и тем же оценкам. Переайдем теперь к максимизации (13.10) при условии (13.7), (13.8) и (13.12) и рассмотрим каждый член, входящий в правую часть (13.13).

Член $p(\mathbf{Y}_k)$ является наиболее простой частью максимизируемой функции. Поскольку в него входят лишь числовые значения наблюдений \mathbf{Y}_k , он играет роль нормировочной постоянной, обеспечивающей равенство единице интеграла $\int p(\mathbf{X}_k | \mathbf{Y}_k) d\mathbf{X}_k$. При максимизации этой составляющей можно пренебречь.

$p(\mathbf{Y}_k | \mathbf{X}_k)$ сводится к простому выражению по аналогии с (13.6) и с учетом того, что последовательность $\{\mathbf{n}(k)\}$ некоррелированная (белая):

$$\begin{aligned} p(\mathbf{Y}_k | \mathbf{X}_k) &= \prod_{i=0}^k p\{\mathbf{y}(i) - \mathbf{g}(\mathbf{x}(i), i)\} = \\ &= c_1 \prod_{i=0}^k \exp \left\{ -\frac{1}{2} [\mathbf{y}(i) - \mathbf{g}(\mathbf{x}(i), i)]' \times \right. \\ &\quad \left. \times \mathbf{N}^{-1}(i) [\mathbf{y}(i) - \mathbf{g}(\mathbf{x}(i), i)] \right\}, \end{aligned}$$

причем вновь нормировочная постоянная учитывается в коэффициенте c_1 . Используя сокращенное обозначение $\mathbf{z}'\mathbf{A}\mathbf{z} = \| \mathbf{z} \|^2_A$, можно записать

$$p(\mathbf{Y}_k | \mathbf{X}_k) = c_1 \exp \left\{ -\frac{1}{2} \sum_{i=0}^k \| \mathbf{y}(i) - \mathbf{g}(\mathbf{x}(i), i) \|_{\mathbf{N}^{-1}(i)}^2 \right\}.$$

$p(\mathbf{X}_k)$ можно преобразовать с учетом того, что

$$p(\mathbf{X}_i) = p(\mathbf{x}(i), \mathbf{X}_{i-1}) = p(\mathbf{x}(i) | \mathbf{X}_{i-1}) p(\mathbf{X}_{i-1}).$$

С учетом (13.7) и предположения о некоррелированности последовательности $\{\mathbf{v}(k)\}$ получим

$$\begin{aligned} p(\mathbf{x}(i) | \mathbf{X}_{i-1}) &= p(\mathbf{x}(i) | \mathbf{x}(i-1), \dots, \mathbf{x}(0)) = \\ &= p(\mathbf{x}(i) | \mathbf{x}(i-1)). \end{aligned}$$

Таким образом,

$$p(\mathbf{X}_k) = p(\mathbf{x}(0)) \prod_{i=1}^k p(\mathbf{x}(i) | \mathbf{x}(i-1)).$$

Из (13.7) также следует, что $p(\mathbf{x}(i+1) | \mathbf{x}(i))$ — гауссовская плотность со средним $\mathbf{f}(\mathbf{x}(i), i)$ и ковариационной матрицей $\mathbf{G}(i) \mathbf{V}(i) \mathbf{G}'(i)$.

Априорная плотность распределения $p(\mathbf{x}(0))$ также считается гауссовой со средним \mathbf{x}_0 и ковариационной матрицей \mathbf{P}_0 . Следовательно, рассматриваемую составляющую можно записать в виде

$$\begin{aligned} p(\mathbf{X}_k) = c_2 \exp \left\{ -\frac{1}{2} [\mathbf{x}(0) - \mathbf{x}_0]' \mathbf{P}_0^{-1} [\mathbf{x}(0) - \mathbf{x}_0] \right\} \times \\ \times \prod_{i=0}^{k-1} \exp \left\{ -\frac{1}{2} [\mathbf{x}(i+1) - \mathbf{f}(\mathbf{x}(i), i)]' \times \right. \\ \left. \times [\mathbf{G}(i) \mathbf{V}(i) \mathbf{G}'(i)]^{-1} [\mathbf{x}(i+1) - \mathbf{f}(\mathbf{x}(i), i)] \right\}, \end{aligned}$$

если только матрица $\mathbf{G} \mathbf{V} \mathbf{G}'$ обратима. Здесь c_2 также является нормировочной постоянной. В сокращенной записи

$$\begin{aligned} p(\mathbf{X}_k) = c_2 \exp \left\{ -\frac{1}{2} \| \mathbf{x}(0) - \mathbf{x}_0 \|_{\mathbf{P}_0^{-1}}^2 - \right. \\ \left. - \frac{1}{2} \sum_{i=0}^{k-1} \| \mathbf{x}(i+1) - \mathbf{f}(\mathbf{x}(i), i) \|_{[\mathbf{G} \mathbf{V} \mathbf{G}']}^{-2} \right\}. \end{aligned}$$

Объединяя выражения для всех трех составляющих, получаем

$$p(\mathbf{X}_k | \mathbf{Y}_k) = c_3 \exp \{-E_k\}, \quad (13.14)$$

где

$$\begin{aligned} E_k = \frac{1}{2} \| \mathbf{x}(0) - \mathbf{x}_0 \|_{\mathbf{P}_0^{-1}}^2 + \frac{1}{2} \sum_{i=0}^k \| \mathbf{y}(i) - \mathbf{g}(\mathbf{x}(i), i) \|_{\mathbf{N}^{-1}(i)}^2 + \\ + \frac{1}{2} \sum_{i=0}^{k-1} \| \mathbf{x}(i+1) - \mathbf{f}(\mathbf{x}(i), i) \|_{[\mathbf{G} \mathbf{V} \mathbf{G}']}^{-2}. \quad (13.15) \end{aligned}$$

Максимизация (13.14) по $\mathbf{X}_k = [\mathbf{x}(0), \dots, \mathbf{x}(k)]$ эквивалентна минимизации E_k . Последний член в (13.15) можно заменить на

$$\frac{1}{2} \sum_{i=0}^{k-1} \|\mathbf{v}(i)\|_{\mathbf{V}^{-1}(i)}^2;$$

ограничение задается формулой (13.7). Следовательно, можно минимизировать

$$\begin{aligned} E'_k = & \frac{1}{2} \|\mathbf{x}(0) - \mathbf{x}_0\|_{\mathbf{P}_0^{-1}}^2 + \frac{1}{2} \sum_{i=0}^k \|\mathbf{y}(i) - \mathbf{g}(\mathbf{x}(i), i)\|_{\mathbf{N}^{-1}(i)}^2 + \\ & + \frac{1}{2} \sum_{i=0}^{k-1} \|\mathbf{v}(i)\|_{\mathbf{V}^{-1}(i)}^2 + \\ & + \lambda'(i) [\mathbf{x}(i+1) - \mathbf{f}(\mathbf{x}(i), i) - \mathbf{G}(i)\mathbf{v}(i)], \\ & i = 0, \dots, k-1, \quad (13.16) \end{aligned}$$

где $\lambda'(i)$ — вектор множителей Лагранжа. В этой формулировке можно не требовать обратимости $\mathbf{G}\mathbf{V}\mathbf{G}'$.

Заметим, что E'_k состоит из трех частей, которые связаны с принятым допущением о значении вектора параметров и состояния при $t = 0$, с аддитивным шумом наблюдений \mathbf{n} , с шумом объекта \mathbf{v} (а также с ограничениями, накладываемыми описанием объекта).

Весовые матрицы \mathbf{P}_0^{-1} , $\mathbf{N}^{-1}(i)$ и $\mathbf{V}^{-1}(i)$ дают количественную характеристику относительной величины вклада этих составляющих в ошибку E'_k . Отметим, что $\mathbf{y}(i)$ при $i = 0, \dots, k$ представляют собой числовые значения наблюдений, проводимых над объектом. Если априорная информация об $\mathbf{x}(0)$ отсутствует, то $\mathbf{P}_0^{-1} \rightarrow 0$ и первая компонента E'_k обращается в нуль.

Заметим, наконец, что в описании объекта (13.7) предполагалось, что отсутствует измеряемая входная последовательность $\{\mathbf{u}(i)\}$, $i = 0, 1, \dots, k$. При наличии такого входа не представляет труда модифицировать это описание и все последующие выводы для учета этой дополнительной информации.

13.3. МИНИМИЗАЦИЯ ФУНКЦИИ ОШИБКИ

Функция ошибки (13.16) была получена при следующих условиях: шум наблюдений \mathbf{p} аддитивный, гауссовский, белый; шум объекта \mathbf{v} — аддитивный, гауссовский. Если распределения шумов неизвестны, выражение (13.13) можно использовать в качестве приемлемой функции ошибки, основанной на интуитивных соображениях. В этом случае в результате минимизации получаются оценки наименьших квадратов.

Минимизируя (13.16), можно найти оценку $\hat{\mathbf{x}}(j | k)$, т. е. оценку вектора параметров и состояния в момент времени j по наблюдениям $0, 1, \dots, k$. Опять можно выделить случаи интерполяции (сглаживания) при $0 \leq j < k$, оценивания текущего состояния (фильтрации) при $j = k$, экстраполяции (если используется расширенное выражение для E_k) при $j > k$.

Возможны, например, следующие подходы к поставленной задаче минимизации:

а) Метод динамического программирования, приводящий к последовательной процедуре, которая дает возможность учитывать новую поступающую информацию, но имеющий ряд свойственных ему специфических проблем, связанных с необходимым объемом памяти вычислительной машины (Беллман: проклятие размерности).

б) Дискретный принцип максимума.

в) Дискретные уравнения Эйлера — Лагранжа; последние два метода приводят к нелинейной двухточечной краевой задаче.

г) Методы «а»—«в», но с предварительной линеаризацией уравнений объекта относительно точки $\mathbf{x}^*(k | k)$, в результате чего получаются приближенные оценки. Однако такой подход позволяет получить рекуррентные схемы, в которых результаты новых наблюдений можно учитывать непосредственно по мере их поступления.

Необходимые условия минимума E'_k получаются непосредственно дифференцированием:

$$\begin{aligned} \frac{\partial E'_k}{\partial \mathbf{x}(i)} = & -[\mathbf{y}(i) - \mathbf{g}(\mathbf{x}(i), i)]' \mathbf{N}^{-1}(i) \frac{\partial \mathbf{g}(\mathbf{x}(i), i)}{\partial \mathbf{x}(i)} + \\ & + \lambda'(i-1) - \lambda'(i) \frac{\partial \mathbf{f}(\mathbf{x}(i), i)}{\partial \mathbf{x}(i)} = 0, \end{aligned} \quad (13.17)$$

$$\frac{\partial E'_k}{\partial \lambda(i)} = \mathbf{x}(i+1) - \mathbf{f}(\mathbf{x}(i), i) - \mathbf{G}(i)\mathbf{v}(i) = 0, \quad (13.18)$$

$$\frac{\partial E'_k}{\partial \mathbf{v}(i)} = \mathbf{v}'(i) \mathbf{V}^{-1}(i) - \lambda'(i) \mathbf{G}(i) = 0. \quad (13.19)$$

Рассматривая эти условия как уравнения для определения оптимальных оценок $\hat{\mathbf{x}}(i | k)$ и $\hat{\lambda}(i)$ и исключая из них $\mathbf{v}(i)$, находим

$$\begin{aligned} \hat{\mathbf{x}}(i+1 | k) &= \mathbf{f}(\hat{\mathbf{x}}(i | k), i) - \\ &- \mathbf{G}(i) \mathbf{V}(i) \mathbf{G}'(i) \hat{\lambda}(i), \quad i = 0, \dots, k-1, \end{aligned} \quad (13.20)$$

$$\begin{aligned} \hat{\lambda}(i-1) &= \frac{\partial \mathbf{f}'(\mathbf{x}(i), i)}{\partial \mathbf{x}(i)} \Big|_{\mathbf{x}(i)=\hat{\mathbf{x}}(i | k)} \hat{\lambda}(i) + \\ &+ \frac{\partial \mathbf{g}'(\mathbf{x}(i), i)}{\partial \mathbf{x}(i)} \Big|_{\mathbf{x}(i)=\hat{\mathbf{x}}(i | k)} \mathbf{N}^{-1}(i) [\mathbf{y}(i) - \mathbf{g}(\hat{\mathbf{x}}(i | k), i)], \\ &\quad i = 1, \dots, k. \end{aligned} \quad (13.21)$$

Соотношения (13.17) и (13.18) называют *каноническими уравнениями*. Аналогично выводятся граничные условия при $i = 0$ и $i = k$:

$$\begin{aligned} \hat{\mathbf{x}}(0 | k) &= \mathbf{x}_0 + \mathbf{P}_0 \frac{\partial \mathbf{g}'(\mathbf{x}(i), i)}{\partial \mathbf{x}(i)} \Big|_{\mathbf{x}(i)=\hat{\mathbf{x}}(0 | k)} \mathbf{N}^{-1}(0) \times \\ &\times [\mathbf{y}(0) - \mathbf{g}(\hat{\mathbf{x}}(0 | k), 0)] + \\ &+ \mathbf{P}_0 \frac{\partial \mathbf{f}'(\mathbf{x}(i), i)}{\partial \mathbf{x}(i)} \Big|_{\mathbf{x}(i)=\hat{\mathbf{x}}(0 | k)} \hat{\lambda}(0), \end{aligned} \quad (13.22)$$

$$\hat{\lambda}(k) = 0. \quad (13.23)$$

Формулы (13.20) — (13.23) определяют хорошо известную двухточечную краевую задачу. Граничные условия разделены: для \mathbf{x} они заданы в начальной точке рассматриваемого интервала, для λ — на его конце. Если бы оба условия относились к одному и тому же концу этого интервала, то были бы возможны непосредственные вычисления в прямом или обратном направлении. Поскольку граничные условия разделены, прямое решение двухточечной краевой задачи невозможно (см. разд. 13.4). Формулы, аналогичные (13.20) — (13.23), можно получить и при непрерывном описании объектов.

Упрощения. В предыдущем рассмотрении объект предполагался нелинейным. В разд. 13.1 уже указывалось, что даже для линейных объектов задача совместного оценивания параметров и состояния оказывается нелинейной. Тем не менее можно ожидать, что для более простых объектов будут получаться и более простые задачи оценивания. По-видимому, возможен лишь неполный анализ такой связи между сложностью объекта и сложностью задачи оценивания.

В качестве одного такого интересного случая можно рассмотреть следующий пример [1]. Пусть линейный стационарный объект описывается уравнениями

$$\begin{aligned} \mathbf{x}(i+1) &= \mathbf{A}\mathbf{x}(i) + \mathbf{v}(i), \\ \mathbf{y}(i) &= \mathbf{x}(i) + \mathbf{n}(i), \end{aligned} \quad (13.24)$$

где шумы $\mathbf{v}(i)$ и $\mathbf{n}(i)$ предполагаются белыми. Отметим, что наблюдаются все состояния. Авторы упомянутой статьи показали, что оценка \mathbf{A} вида

$$\hat{\mathbf{A}}_n = \left[\sum_{i=3}^n \mathbf{y}(i) \mathbf{y}'(i-2) \right] \left[\sum_{i=3}^n \mathbf{y}(i-1) \mathbf{y}'(i-2) \right]^{-1}$$

сходится к \mathbf{A} с вероятностью единица, а фильтр Калмана, построенный на основе этих $\hat{\mathbf{A}}_n$, дает оценку, сходящуюся с вероятностью единица к оценке, даваемой истинным фильтром Калмана. Следовательно, как параметры, так и состояние оцениваются правильно. В работе [2] можно найти другой подход к таким задачам, основанный на использовании настраиваемых моделей.

13.4. НЕКОТОРЫЕ АЛГОРИТМЫ

Из большого числа публикаций по проблемам оптимального управления значительное число работ имеет отношение к рассматриваемой задаче. Исчерпывающее обсуждение многочисленных методов, предлагаемых и рассматриваемых в этих работах (см. [4]), далеко выходит за рамки этой главы. Мы ограничимся лишь некоторыми замечаниями о квазилинеаризации и инвариантном погружении.

Квазилинеаризация

Это интересный и мощный метод. В его основе лежит метод Ньютона — Рафсона определения корня алгебраического уравнения и его обобщение на функциональные уравнения, сделанное Канторовичем. Под названием «квазилинеаризация» это направление развивается Беллманом и Калабой [3].

Для нашего изложения воспользуемся непрерывным описанием объекта, сформулированным выше, предполагая, что выход y линейно зависит от состояния x :

$$\begin{bmatrix} \dot{x} \\ \dot{a} \end{bmatrix} = \begin{bmatrix} f(x, u, a, v, t) \\ 0 \end{bmatrix}, \quad (13.25)$$

$$y(t) = Cx(t).$$

Векторы x и a имеют размерности соответственно n и m . На интервале наблюдений за объектом измеряются

$$y(j) = y_j, \quad j = 1, 2, \dots, m + n, \quad (13.26)$$

т. е. числовые значения выхода, соответствующие $m + n$ моментам времени. Выражение (13.25) дает $m + n$ уравнений, которые должны удовлетворять граничным условиям (13.26). Поэтому рассматриваемая задача называется многоточечной краевой задачей. Покажем, как можно решать эту задачу методом квазилинеаризации. Для $(i + 1)$ -х приближений функций $x(t)$ и $a(t)$ на интервале наблюдений справедливы уравнения

$$\begin{aligned} \dot{x}^{(i+1)} &= f(x^{(i)}, u, a^{(i)}) + F_1^{(i)}[x^{(i+1)} - x^{(i)}] + \\ &\quad + F_2^{(i)}[a^{(i+1)} - a^{(i)}], \end{aligned} \quad (13.27)$$

$$\dot{a}^{(i+1)} = 0$$

при условиях

$$Cx^{(i+1)} = y^{(i+1)} = y, \quad (13.28)$$

Матрицы $F_1^{(i)}$ и $F_2^{(i)}$ — якобианы:

$$F_1^{(i)} = \left[\frac{\partial f}{\partial x} \right]_{x^{(i)}, a^{(i)}}, \quad F_2^{(i)} = \left[\frac{\partial f}{\partial a} \right]_{x^{(i)}, a^{(i)}}. \quad (13.29)$$

На $(i+1)$ -й итерации задача формулируется в виде

$$\begin{aligned}\dot{\mathbf{x}}^{(i+1)} &= \mathbf{F}_1^{(i)} \mathbf{x}^{(i+1)} + \mathbf{F}_2^{(i)} \mathbf{a}^{(i+1)} + \mathbf{w}^{(i)}, \\ \mathbf{y}(j) &= \mathbf{C} \mathbf{x}^{(i+1)},\end{aligned}\quad (13.30)$$

причем $\mathbf{w}^{(i)}$ определяется на предыдущей итерации:

$$\mathbf{w}^{(i)} = \mathbf{f}(\mathbf{x}^{(i)}, \mathbf{u}, \mathbf{a}^{(i)}) - \mathbf{F}_1^{(i)} \mathbf{x}^{(i)} - \mathbf{F}_2^{(i)} \mathbf{a}^{(i)}. \quad (13.31)$$

Первая итерация начинается с произвольных начальных приближений $\mathbf{x}^{(0)}$ и $\mathbf{a}^{(0)}$. Для обеспечения сходимости итерационной процедуры начальные приближения иногда требуется выбрать достаточно хорошими. Для получения таких приближений можно использовать дифференциальную (или разностную) аппроксимацию. Итерационную процедуру можно заканчивать, когда разности между решениями, полученными на двух последовательных шагах, т. е. величины $\mathbf{x}^{(i+1)}(t) - \mathbf{x}^{(i)}(t)$, $\mathbf{a}^{(i+1)}(t) - \mathbf{a}^{(i)}(t)$, удовлетворяют заранее заданным условиям. Беллман и Калаба [3] показали, что последовательность функций $\mathbf{x}^{(0)}(t), \dots, \mathbf{x}^{(i+1)}(t)$ сходится в среднеквадратичном к истинному решению, если $\mathbf{f}(\mathbf{x}, \mathbf{u}, \mathbf{a})$ — выпуклая функция, а элементы якобиана, стоящие над и под главной диагональю, положительны. В работе [8] показано, что сходимость возможна и при невыполнении этих условий.

До сих пор предполагалось, что число измерений равно числу уравнений $n+m$. Из-за возмущений наблюдений число измерений k должно быть выбрано большим $n+m$. В этом случае можно минимизировать квадрат разности между измерениями \mathbf{y}_j и выходом модели $\mathbf{y}(j)$. В [6] рассмотрено применение этой методики при оценивании параметров и состояния. Такой же подход можно использовать и при дискретном способе описания объекта.

Инвариантное погружение

Для ясности изложения основных идей воспользуемся следующим типом описания объекта:

$$\begin{aligned}\dot{\mathbf{x}}(t) &= \mathbf{f}(\mathbf{x}(t), t) + \mathbf{G}(t) \mathbf{v}(t), \\ \mathbf{y}(t) &= \mathbf{g}(\mathbf{x}(t), t) + \mathbf{e}(t).\end{aligned}\quad (13.32)$$

Как и раньше, оцениваемые параметры включены в расширенный вектор параметров и состояния $\mathbf{x}(t)$. Так же как в разд. 13.3, приходим к соответствующей двухточечной краевой задаче на интервале $0 \leq t \leq T$:

$$\dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}(t), t) - \mathbf{G}(t) \mathbf{V}(t) \mathbf{G}'(t) \lambda(t), \quad (13.33)$$

$$\begin{aligned} \dot{\lambda}(t) = & \frac{\partial \mathbf{g}'(\mathbf{x}(t), t)}{\partial \mathbf{x}(t)} \mathbf{N}^{-1}(t) [\mathbf{y}(t) - \mathbf{g}(\mathbf{x}(t), t)] - \\ & - \frac{\partial \mathbf{f}'(\mathbf{x}(t), t)}{\partial \mathbf{x}(t)} \lambda(t) \end{aligned} \quad (13.34)$$

с граничными условиями

$$\mathbf{x}(0) = \mathbf{x}_0 - \mathbf{P}_0 \lambda(0), \quad (13.35)$$

$$\lambda(T) = 0. \quad (13.36)$$

Перепишем уравнения (13.33) и (13.34) в виде

$$\dot{\mathbf{x}}(t) = \gamma(\mathbf{x}(t), \lambda(t), t), \quad (13.37)$$

$$\dot{\lambda}(t) = \eta(\mathbf{x}(t), \lambda(t), t). \quad (13.38)$$

Обратимся теперь к граничному условию $\lambda(T) = 0$. Задача погружается (обобщается) в более общую проблему (см., например, [7]):

$$\lambda(T) = \mathbf{c}. \quad (13.39)$$

Теперь \mathbf{c} и T считаются переменными. Изменения \mathbf{c} влияют на окончательное значение λ . Изменение T до $T + \Delta T$ также оказывает влияние на величину λ . Следовательно, имеют место соотношения

$$\lambda(T) = \mathbf{c} \Rightarrow \mathbf{x}(T) = \mathbf{r}(\mathbf{c}, T),$$

$$\lambda(T + \Delta T) = \mathbf{c} + \Delta \mathbf{c} \Rightarrow \mathbf{x}(T + \Delta T) = \mathbf{r}(\mathbf{c}, T) + \Delta \mathbf{x},$$

$$\mathbf{x}(T + \Delta T) = \mathbf{r}(\mathbf{c} + \Delta \mathbf{c}, T + \Delta T).$$

Разлагая в ряд Тейлора последнее соотношение, получаем

$$\Delta \mathbf{x} = \mathbf{r}(\mathbf{c} + \Delta \mathbf{c}, T + \Delta T) - \mathbf{r}(\mathbf{c}, T) =$$

$$= \frac{\partial \mathbf{r}(\mathbf{c}, T)}{\partial \mathbf{c}} \Delta \mathbf{c} + \frac{\partial \mathbf{r}(\mathbf{c}, T)}{\partial T} \Delta T + \mathcal{O}[\Delta T^2], \quad (13.40)$$

где $\mathcal{O}[\Delta T^2]$ обозначает члены второго и более высокого порядка малости по ΔT . Из формул (13.37) и (13.38)

видно, что

$$\Delta \mathbf{x} = \gamma(\mathbf{r}(\mathbf{c}, T), \mathbf{c}, T) \Delta T + \mathcal{O}[\Delta T^2], \quad (13.41)$$

$$\Delta \mathbf{c} = \eta(\mathbf{r}(\mathbf{c}, T), \mathbf{c}, T) \Delta T + \mathcal{O}[\Delta T^2]. \quad (13.42)$$

Подставляя эти выражения в (13.40), деля на ΔT и устремляя ΔT к 0, получаем *уравнение инвариантного погружения*

$$\gamma(\mathbf{r}, \mathbf{c}, T) = \frac{\partial \mathbf{r}(\mathbf{c}, T)}{\partial \mathbf{c}} \eta(\mathbf{r}, \mathbf{c}, T) + \frac{\partial \mathbf{r}(\mathbf{c}, T)}{\partial T}. \quad (13.43)$$

В качестве приближенного решения этого уравнения в частных производных в окрестности точки $\mathbf{c} = \mathbf{0}$ выбирается

$$\mathbf{r}(\mathbf{c}, T) = \hat{\mathbf{x}}(T) - \mathbf{P}(T) \mathbf{c}. \quad (13.44)$$

Из сформулированной выше двухточечной краевой задачи при $\mathbf{c} = \mathbf{0}$ действительно получается $\hat{\mathbf{x}}(T)$, оптимальная оценка вектора параметров и состояния при $t = T$. Подстановка этого приближенного решения (13.44) в (13.43) дает

$$\begin{aligned} \gamma(\hat{\mathbf{x}}(T) - \mathbf{P}(T) \mathbf{c}, \mathbf{c}, T) &= -\mathbf{P}(T) \eta(\hat{\mathbf{x}}(T) - \mathbf{P}(T) \mathbf{c}, \mathbf{c}, T) + \\ &\quad + \frac{\partial \hat{\mathbf{x}}(T)}{\partial T} - \frac{\partial \mathbf{P}(T)}{\partial T} \mathbf{c}. \end{aligned} \quad (13.45)$$

Вернемся теперь к исходным уравнениям для γ и η именно (13.33) и (13.34), и подставим их в (13.45). В результате получим

$$\begin{aligned} \mathbf{f}(\hat{\mathbf{x}}(T) - \mathbf{P}(T) \mathbf{c}, T) - \mathbf{G}(T) \mathbf{V}(T) \mathbf{G}'(T) \mathbf{c} &= \\ &= -\mathbf{P}(T) \frac{\partial \mathbf{g}'(\hat{\mathbf{x}}(T) - \mathbf{P}(T) \mathbf{c}, T)}{\partial \hat{\mathbf{x}}} \times \\ &\quad \times \mathbf{N}^{-1}(T) [\mathbf{y}(T) - \mathbf{g}(\hat{\mathbf{x}}(T) - \mathbf{P}(T) \mathbf{c}, T)] + \\ &+ \mathbf{P}(T) \frac{\partial \mathbf{f}'(\hat{\mathbf{x}}(T) - \mathbf{P}(T) \mathbf{c}, T)}{\partial \hat{\mathbf{x}}} \mathbf{c} + \frac{\partial \hat{\mathbf{x}}(T)}{\partial T} - \frac{\partial \mathbf{P}(T)}{\partial T} \mathbf{c}. \end{aligned} \quad (13.46)$$

Поскольку $\hat{\mathbf{x}}(T) = \mathbf{r}(0, T)$, запишем разложение в ряд в окрестности этой оптимальной точки, сохраняя только

члены нулевого и первого порядка по \mathbf{c} :

$$\begin{aligned} \mathbf{f}(\hat{\mathbf{x}}(T), T) - \frac{\partial \mathbf{f}'(\hat{\mathbf{x}}(T), T)}{\partial \hat{\mathbf{x}}} \mathbf{P}(T) \mathbf{c} - \mathbf{G}(T) \mathbf{V}(T) \mathbf{G}'(T) \mathbf{c} = \\ = -\mathbf{P}(T) \frac{\partial \mathbf{g}'(\hat{\mathbf{x}}(T), T)}{\partial \hat{\mathbf{x}}} \mathbf{N}^{-1}(T) [\mathbf{y}(T) - \mathbf{g}(\hat{\mathbf{x}}, T)] + \\ + \mathbf{P}(T) \frac{\partial}{\partial \hat{\mathbf{x}}} \left[\frac{\partial \mathbf{g}'(\hat{\mathbf{x}}(T), T)}{\partial \hat{\mathbf{x}}} \mathbf{N}^{-1}(T) [\mathbf{y}(T) - \mathbf{g}(\hat{\mathbf{x}}, T)] \right] \mathbf{P}(T) \mathbf{c} + \\ + \mathbf{P}(T) \frac{\partial \mathbf{f}'(\hat{\mathbf{x}}(T), T)}{\partial \hat{\mathbf{x}}} \mathbf{c} + \frac{\partial \hat{\mathbf{x}}(T)}{\partial T} - \frac{\partial \mathbf{P}(T)}{\partial T} \cdot \mathbf{c}. \quad (13.47) \end{aligned}$$

Так как это равенство должно быть справедливо для произвольных (малых) значений \mathbf{c} , можно по отдельности приравнять члены нулевого и первого порядка по \mathbf{c} в обеих частях (13.47). В результате получаются следующие уравнения:

$$\frac{\partial \hat{\mathbf{x}}(T)}{\partial T} = \mathbf{f}(\hat{\mathbf{x}}(T), T) + \mathbf{P}(T) \frac{\partial \mathbf{g}'(\hat{\mathbf{x}}(T), T)}{\partial \hat{\mathbf{x}}} \mathbf{N}^{-1}(T) [\mathbf{y}(T) - \mathbf{g}(\hat{\mathbf{x}}(T), T)], \quad (13.48)$$

$$\begin{aligned} \frac{\partial \mathbf{P}(T)}{\partial T} = \mathbf{G}(T) \mathbf{V}(T) \mathbf{G}'(T) + \mathbf{P}(T) \frac{\partial \mathbf{f}'(\hat{\mathbf{x}}(T), T)}{\partial \hat{\mathbf{x}}} + \\ + \frac{\partial \mathbf{f}(\hat{\mathbf{x}}(T), T)}{\partial \hat{\mathbf{x}}} \mathbf{P}(T) + \mathbf{P}(T) \times \\ \times \frac{\partial}{\partial \hat{\mathbf{x}}} \left[\frac{\partial \mathbf{g}'(\hat{\mathbf{x}}(T), T)}{\partial \hat{\mathbf{x}}} \mathbf{N}^{-1}(T) [\mathbf{y}(T) - \mathbf{g}(\hat{\mathbf{x}}(T), T)] \right] \mathbf{P}(T). \quad (13.49) \end{aligned}$$

Эти уравнения задают рекуррентную оценку наименьших квадратов, так как T теперь можно рассматривать как текущую переменную t . Начальные условия вытекают из уравнения (13.35)

$$\mathbf{x}(0) = \mathbf{x}_0 - \mathbf{P}_0 \lambda(0)$$

и (13.34). Следовательно,

$$\hat{\mathbf{x}}(0) = \mathbf{x}_0, \quad \mathbf{P}(0) = \mathbf{P}_0.$$

Пример. Рассмотрим пример, приведенный в [7]. Исследуется объект второго порядка

$$\dot{x}_1 = x_2,$$

$$\dot{x}_2 = -2x_1 - a(t)x_1^3 - 3x_2 + 5 \sin t$$

с переменным параметром $a(t) = 2 \exp [-0,1t]$ и наблюдениями выходного сигнала, задаваемыми формулой

$$y = x_1 + n.$$

Если известна функциональная форма зависимости $a(t)$, но неизвестны ее начальное значение и постоянная времени, то можно записать

$$\dot{a} = -ba, \text{ или } \dot{x}_3 = -x_4 \dot{x}_3,$$

$$\dot{b} = 0, \quad \text{или } \dot{x}_4 = 0.$$

Следовательно,

$$\dot{\mathbf{x}} = \begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \\ \dot{x}_4 \end{bmatrix} = \begin{bmatrix} x_2 \\ -2x_1 - x_3 x_1^3 - 3x_2 + 5 \sin t \\ -x_4 x_3 \\ 0 \end{bmatrix} = \mathbf{f}(\mathbf{x}, t),$$

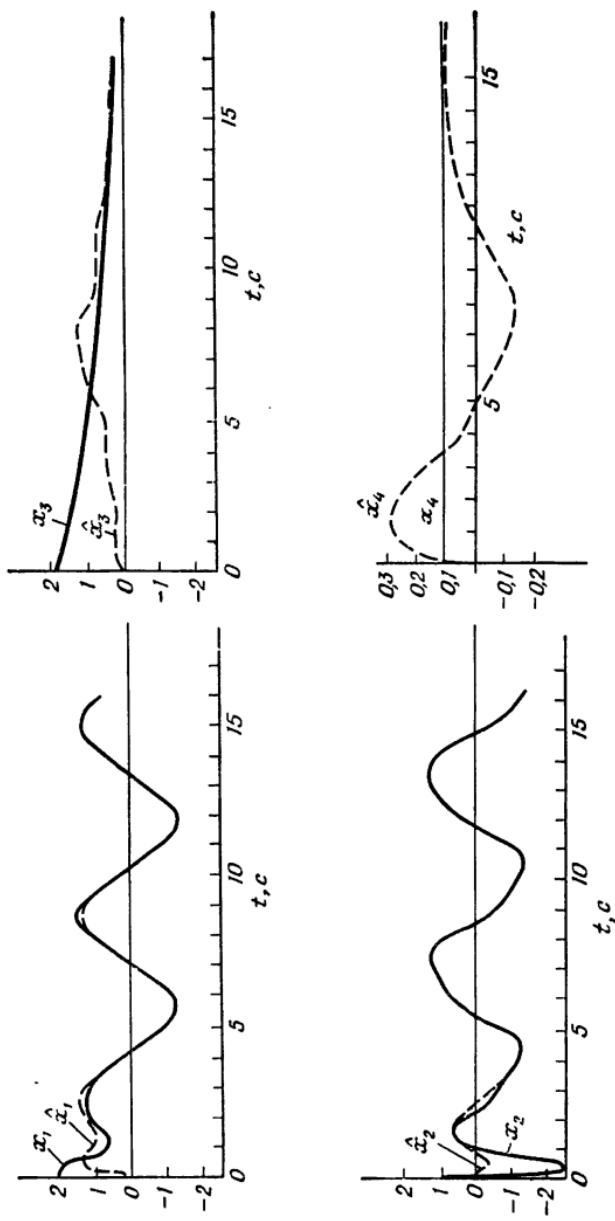
$$\mathbf{y} = \begin{bmatrix} y \\ 0 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} x_1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} n \\ 0 \\ 0 \\ 0 \end{bmatrix} = \mathbf{g}(\mathbf{x}, t) + \mathbf{n}.$$

Даны также

$$\mathbf{N}(t) = \mathbf{I} \text{ и } \mathbf{G}\mathbf{V}\mathbf{G}' = \mathbf{I}.$$

Заметим, что

$$\mathbf{P}(t) \frac{\partial \mathbf{g}'}{\partial \hat{\mathbf{x}}} = [P_{11}, P_{21}, P_{31}, P_{41}]'.$$



Ф и г. 13.2.

Итак, находим оценку в виде

$$\begin{bmatrix} \dot{\hat{x}}_1 \\ \dot{\hat{x}}_2 \\ \dot{\hat{x}}_3 \\ \dot{\hat{x}}_4 \end{bmatrix} = \begin{bmatrix} \hat{x}_2 \\ -2\hat{x}_1 - \hat{x}_3\hat{x}_1^3 - 3x_2 + 5 \sin t \\ -\hat{x}_4\hat{x}_3 \\ 0 \end{bmatrix} + \begin{bmatrix} P_{11} \\ P_{21} \\ P_{31} \\ P_{41} \end{bmatrix} (y - \hat{x}_1).$$

Уравнение для \hat{P} также можно получить непосредственно. При заданных начальных условиях $x_0 = 0$ и P_0 можно приступить к процедуре оценивания. На фиг. 13.2 показано изменение оценки во времени.

Аналогичным образом метод инвариантного погружения можно использовать при дискретном описании объектов.

13.5. ЗАКЛЮЧЕНИЕ

Поскольку невозможно здесь уделить внимание многим интересным работам, относящимся к методам минимизации функций ошибки и решения двухточечной краевой задачи, читателю следует обратиться к монографии [9] на эту тему и статьям, перечисленным в списке дополнительной литературы.

ЛИТЕРАТУРА

1. Anderson W. N., et al., Consistent Estimates of the Parameters of a Linear System, *Ann. Math. Statist.*, **40**, 2064—2075 (1969).
2. Baggerud A., Balchen J. G., An Adaptive State Estimator, Second IFAC Symp. «Identification and Process Parameter Estimation», Prague, Paper 10.1, 1970.
3. Bellman R. E., Kalaba R. E., Quasilinearization and Nonlinear Boundary-Value Problems, American Elsevier Publ., N.Y., 1965; есть русский перевод: Беллман Р., Калаба Р., Квазилинеаризация и нелинейные краевые задачи, изд-во «Мир», 1968.
4. Guenod M., Sage A. P. (см. литературу к гл. 2).
5. Cox H., On the Estimation of State Variables and Parameters for Noisy Dynamic Systems, *IEEE Trans. Autom. Control*, **AC-9**, 5—12 (1964).
6. Detchmendy D. M., Sridhar R., On the Experimental Determination of the Dynamical Characteristics of Physical Systems, *Proc. Nat. Electron. Conf.*, Chicago, Ill., **21**, 575—580 (1965).

7. Detchmendy D. M., Sridhar R., Sequential Estimation of States and Parameters in Noisy Nonlinear Dynamical Systems, *J. Basic Eng. (Trans. ASME)*, 88D, 362—368 (1966).
8. Ohap R. F., Stubberud A. R., A Technique for Estimating the State of a Nonlinear System, *IEEE Trans. Autom. Control*, AC-10, 150—155 (1965).
9. Sage A. P., Melsa J. L. (см. литературу к гл. 2).

ДОПОЛНИТЕЛЬНАЯ ЛИТЕРАТУРА

(СМ. ТАКЖЕ ДОПОЛНИТЕЛЬНУЮ ЛИТЕРАТУРУ К РАЗД. 8.4)

Aoki M. (см. литературу к гл. 4).

Balakrishnan A. V., Identification of Control Systems from Input-Output Data, IFAC Symp. «Identification in Autom. Control Systems», Prague, Paper 1.I., 1967.

Balakrishnan A. V., Neustadt L. W., eds., Computing Methods in Optimization Problems, Academic Press, N.Y., 1964.

Bar-Shalom Y., Optimal Simultaneous State Estimation and Parameter Identification in Linear Discrete-Time Systems, *IEEE Trans. Autom. Control*, AC-17, 308—319 (1972).

Bellman R. E., Kalaba R. E., Wing G. M., Invariant Imbedding and Mathematical Physics I — Particle Processes, *J. Math. Phys.*, 1, 280—308 (1960).

Bellman R. E., Kalaba R. E., Wing G. M., Invariant Imbedding and the Reduction of Two-Point Boundary Value Problems to Initial-Value Problems, *Proc. Nat. Acad. Sci.*, U.S.A., 46, 1646—1649 (1960).

Bellman R. E., Adaptive Control Processes, A Guided Tour, Princeton Univ. Press, N.J., 1961; есть русский перевод: Беллман Р., Процессы регулирования с адаптацией, изд-во «Наука», 1964.

Bellman R. E., Kalaba R. E., On the Fundamental Equations of Invariant Imbedding — I, *Proc. Nat. Acad. Sci.*, U.S.A., 47, 336—338 (1961).

Bellman R. E., Kagiwada H., Kalaba R. E., Quasilinearization, System Identification, and Prediction, Rept. Rand Corp., RM-3812-PR, 1963.

Bellman R. E., Kagiwada H., Kalaba R. E., Numerical Studies of a Two-Point Nonlinear Boundary Value Problem Using Dynamic Programming, Invariant Imbedding and Quasilinearization, Rept. Rand Corp., RM-4069-PR, 1964.

Bellman R. E., Kagiwada H. H., Kalaba R. E., Quasilinearization, Boundary-Value Problems, and Linear Programming, *IEEE Trans. Autom. Control*, AC-10, 2 (1965).

Bellman R. E., Kalaba R. E., Dynamic Programming, Invariant Imbedding and Quasilinearization: Comparisons and Interconnections, Rept. Rand Corp., RM-4038-PR, 1964.

Bellman R. E., Dynamic Programming, System Identification and Suboptimization, Rept. Rand Corp., RM-4593-PR, 1965.

Berkovec J. W., A Method for the Validation of Kalman Filter Models, *Proc. Joint Autom. Control Conf.*, 1969, pp. 488—493.

- Coasert R., On the Estimation of the Plant State and Plant Parameters and External Disturbances from Observations of the Output of a Control System, Joint Autom. Control Conf., 1964, pp. 340—348.
- Coggan G. C., Noton A. R. M., Discrete-Time Sequential State and Parameter Estimation in Chemical Engineering, *Trans. Inst. Chem. Eng.*, 40, 255—64 (1970).
- Cox H., Sequential Minimax Estimation, *IEEE Trans. Autom. Control*, AC-11, 323—324 (1966).
- Ellis T. W., Sage A. P., Application of Method for On-Line Combined Estimation and Control, Proc. Southwest IEEE Conf., 1968.
- Fan L. T., Wing C. S., The Discrete Maximum Principle, Wiley, N.Y., 1964; есть русский перевод: Фан Л. Ц., Винь Ч. С., Дискретный принцип максимума, изд-во «Мир», 1967.
- Giese C., McGhee R. B., Estimation of Nonlinear System States and Parameters by Regression Methods, Joint Autom. Control Conf., 1965, pp. 46—53.
- Habegger L. J., Bailey R. E. (см. литературу к разд. 14.4).
- Henrici P., Discrete Variable Methods in Ordinary Differential Equations, Wiley, N.Y., 1962.
- Hilborn C. G., Lainiotis D. G., Optimal Estimation in the Presence of Unknown Parameters, *IEEE Trans. Systems Sci. Cybernetics*, SSC-5, 38—43 (1969).
- Jazwinski A. H., Filtering for Nonlinear Dynamical Systems, *IEEE Trans. Autom. Control*, AC-11, 765 (1966).
- Jazwinski A. H., Adaptive Filtering, *Automatica*, 5, 475—485 (1969).
- Jazwinski A. H. (см. литературу к гл. 12).
- Kailath T., Frost P. (см. литературу к гл. 4).
- Kalaba R. E., On Nonlinear Differential Equations, the Maximum Operation, and Monotone Convergence, *J. Mathematics and Mechanics*, 8, 519—574 (1959).
- Kalaba R., Invariant Imbedding and the Analysis of Processes, Rept. Rand Corp., RM-3806-PR, 1963.
- Kalaba R., Some Aspects of Quasilinearization, in: Nonlinear Differential Equations and Nonlinear Mechanics, LaSalle J. P., Lefschetz S., ed., Academic Press, N.Y., 1963.
- Kenneth P., McGill R., Two Point Boundary Value Problem Techniques, *Advan. Contr. Syst.*, 3, 69—110 (1966).
- Kopp R. E., Orford R. J., Linear Regression Applied to System Identification for Adaptive Control Systems, *AIAA J.*, 1, 2300—2306 (1963).
- Kumar K. P. S., Sridhar R., On the Identification of Control Systems by the Quasi-Linearization Method, *IEEE Trans. Autom. Control*, AC-9, 151—154 (1964).
- Kushner H. J., Approximations to Optimal Non-Linear Filters, Joint Autom. Control Conf., 1967, pp. 613—623.
- Kushner H. J., Non-Linear Filtering: the Exact Dynamical Equations Satisfied by the Conditional Mode, *IEEE Trans. Autom. Control*, AC-12, 262—267 (1967).
- Larminat Ph. de, Tallec M. H., On-Line Identifications of Time

Varying Systems, Second IFAC Symp. «Identification and Process Parameter Estimation», Prague, Paper 3.3, 1970.

Lavi A., Strauss J. C., Parameter Identification in Continuous Dynamic Systems, IEEE Intern. Conv. Rec., Symp. on Automatic Control; Systems Science; Cybernetics; Human Factors, Part VI, 1965, pp. 49–61.

Lee R. C. K. (см. литературу к гл. 4).

Mowery V. O., Least Squares Recursive Differential-Correction Estimation in Nonlinear Problems, *IEEE Trans. Autom. Control*, AC-10, 399–407 (1965).

Neal S. R., Nonlinear Estimation Techniques, *IEEE Trans. Autom. Control*, AC-13, 705–708 (1968).

Noton A. R. M., Introduction to Variational Methods in Control Engineering, Pergamon Press, Oxford, 1965.

Pearson J. B., A Note on Nonlinear Filtering (Correspondence), *IEEE Trans. Autom. Control*, AC-13, 103–105 (1968).

Sage A. P., Masters G. W., On-Line Estimation States and Parameters for Discrete Nonlinear Dynamic Systems, *Proc. Nat. Electron. Conf.*, 22, 677–682 (1966).

Sage A. P., Masters G. W., Identification and Modeling of States and Parameters of Nuclear Reactor Systems, *IEEE Trans. Nucl. Sci.*, NS-14, 1, 279–285 (1967).

Schley C. H., Jr., Lee I., Optimal Control Computation by the Newton-Raphson Method and the Riccati Transformation, *IEEE Trans. Autom. Control*, AC-12, 139–144 (1967).

Schwartz L., Stear E. B., A Computational Comparison of Several Nonlinear Filters, *IEEE Trans. Autom. Control*, AC-13, 83–86 (1968).

Schweppe F. C., Recursive State Estimation: Unknown but Bounded Errors and System Inputs, *IEEE Trans. Autom. Control*, AC-13, 22–28 (1968).

Sylvester R. J., Meyer F., Two Point Boundary Value Problems by Quasilinearization, *SIAM J.*, 13, 586–602 (1965).

Sunahara Y., An Approximate Method of State Estimation for Non-Linear Dynamical Systems, Joint Autom. Control Conf., 1969, pp. 161–172.

Tait G. R., Bélanger P. R., A Comparison of Some Parameter Identification Schemes Using First and Second Order Extended Kalman-Bucy Filters and Sensitivity Functions, Second IFAC Symp. «Identification and Process Parameter Estimation», Prague, Paper 10.3, 1970.

Wishner R. P., Tabaczynski J. A., Athans M., A Comparison of Three Non-Linear Filters, *Automatica*, 5, 487–496 (1969).

Wonham W. M., Stochastic Problems in Optimal Control, *IEEE Conv. Rec.*, Part 2, 114–124 (1963).

Wonham W. M., Some Applications of Stochastic Differential Equations to Non-Linear Filtering, *SIAM J. Control*, 2, 347–369 (1964).

ПРИЛОЖЕНИЯ

ГЛАВА 14

ОБЗОР ПРАКТИЧЕСКИХ ПРИМЕНЕНИЙ

Как отмечалось в разд. 1.6, методы оценивания параметров и состояния могут использоваться для различных целей, например для диагностики и управления, автоматизации измерений, принятия решений и регулирования (в промышленности), распознавания образов, а также в разнообразных областях человеческой деятельности, например в связи, энергетике, машиностроении, авиации, химической промышленности, физике, геологии, экономике, биологии и медицине. В приложениях, связанных с автоматическим управлением, существенное место занимает вопрос о том, насколько процедура оценивания согласована с рассматриваемой задачей управления. Этот вопрос кратко обсуждается в разд. 14.1. В остальной части главы рассматриваются некоторые конкретные области применения, иллюстрируемые отдельными примерами: физические, механические и химические объекты (разд. 14.2), ядерные реакторы (разд. 14.3), энергетические и энергораспределительные системы (разд. 14.4), связь (разд. 14.5), авиаация и космонавтика (разд. 14.6), биологические объекты (разд. 14.7), социально-экономические системы (разд. 14.8), обучающиеся системы и распознавание образов (разд. 14.9). Этот список областей применения не представляет, конечно, единственного возможного разделения типов практических приложений, и приводимые примеры не являются наиболее важными и наиболее

характерными. Используя список литературы к данной главе, читатель может видоизменить картину приложений, приспосабливая ее к собственным интересам. Этот список литературы разделен в соответствии с тематикой разделов главы.

14.1. АВТОМАТИЧЕСКОЕ УПРАВЛЕНИЕ

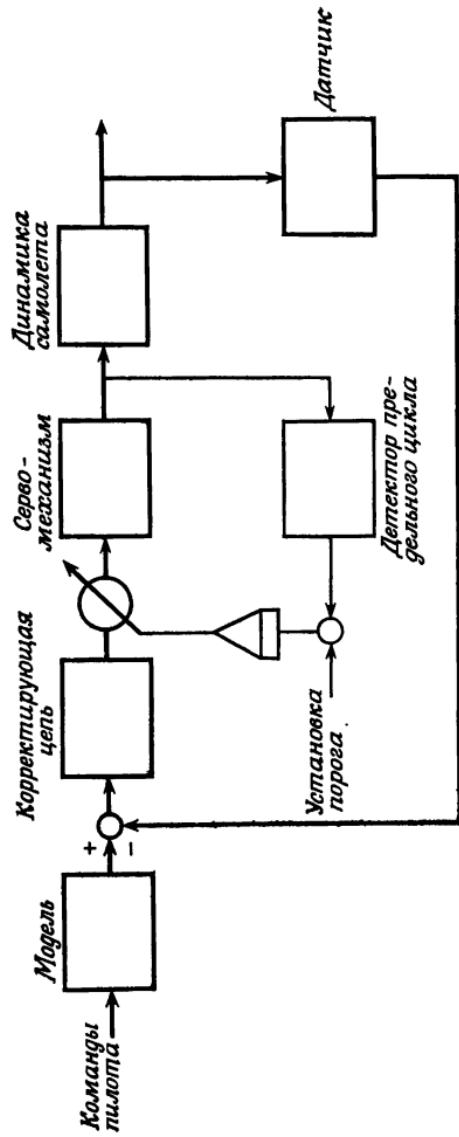
В гл. 1 подчеркивалось, что применение обратной связи оказалось эффективным средством для борьбы с неопределенностью. При этом, однако, имеются ограничения, в частности очень жесткие требования к оптимизации, предполагающие использование *оптимального* или *самооптимизирующегося* управления или значительные изменения параметров, приводящие к использованию *адаптивного* управления [7, к разд. 14.1]. В подобных случаях первостепенное значение приобретает весьма точная или постоянно обновляемая информация об объекте управления. Упоминавшийся в разд. 1.4 принцип *дуального управления* Фельдбаума предполагает одновременное осуществление двух функций: изучение или определение характеристик объекта и управление объектом для приведения его к требуемому состоянию. К сожалению, столь рациональный подход к интересующим нас задачам приводит к теоретическим результатам, практическая реализация которых даже в простейших случаях связана с чрезвычайно громоздкими вычислениями (см. [4, 11, к разд. 14.1], а также [16, к разд. 14.1]). Разумеется, делались попытки упростить эту общую задачу. С этой целью принимается априорное допущение о том, что синтез системы управления можно разделить на две части: процесс идентификации и процесс управления. По аналогии с терминологией статистической теории управления будем называть это допущение *гипотезой разделимости*. Такой подход вполне естествен, особенно если учсть множество методов, разработанных для синтеза систем управления при известных характеристиках объекта и окружающей среды. Тем не менее редко удается достичь оптимального решения, осуществляя идентификацию объекта и затем используя ее результаты в процессе синтеза системы управления, проводимого в предположении, что свойства объекта и внешней среды

точно известны. Может возникнуть необходимость изменить стратегию управления, чтобы учесть неточность результатов идентификации.

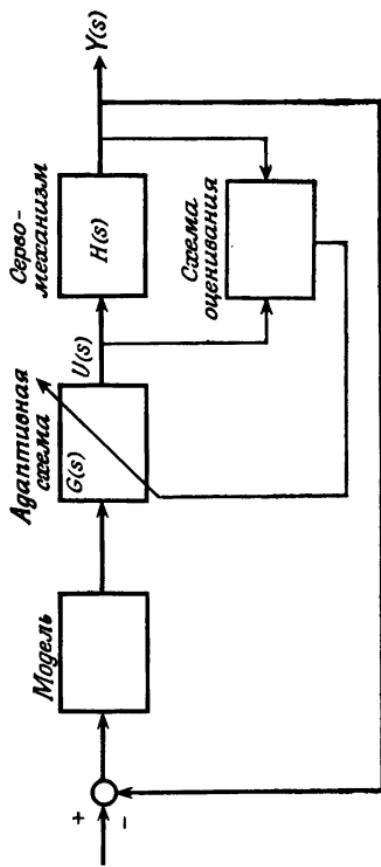
Проблемой, заслуживающей внимания, представляется строгое исследование условий, при которых гипотеза разделимости верна. Помимо того очевидного факта, что всегда желательно выбрать класс моделей \mathcal{Y} , для которого имеется достаточно разработанная теория управления, существует еще немало других интересных вопросов в области идентификации и управления, в частности следующие. Возможно ли указать рациональные способы выбора структуры модели и критерия идентификации, если заранее известно, что результаты идентификации будут использованы для решения определенной задачи управления? Какова должна быть точность решения задачи идентификации, если гипотеза разделимости должна выполняться по меньшей мере с заданной заранее ошибкой? Частичные ответы на эти вопросы для ограниченного класса задач даны в работе [2, к разд. 14.1].

Самоорганизующиеся и адаптивные системы

Начиная с ранней работы [3, к разд. 14.1] было сделано много попыток и предложений, касающихся адаптивных и самоорганизующихся систем (см., например, [1, 5, 6, 8–10, 12, 13, 15, к разд. 14.1]). Многие из предложенных методов опираются на методы оценивания. Имеется, однако, возможность реализовать принципы самоорганизации и адаптации непосредственно. Типичным примером являются системы поиска оптимума или настройки устройств управления аэрофотосъемкой, работающие в натуральном масштабе времени. В подобных случаях для определения направления движения к оптимуму может возникнуть необходимость введения тестсигналов для нахождения представляющих интерес величин (стационарных точек, параметров). Один класс адаптивных систем известен под названием *эталонных моделей*. К первым работам в этом направлении принадлежит [20, к разд. 14.1]. Этот метод можно проиллюстрировать на примере системы управления самолетом (фиг. 14.1). На разных высотах в зависимости от плотности воздуха изменяется эффектив-



Фиг. 14.1.



Фиг. 14.2.

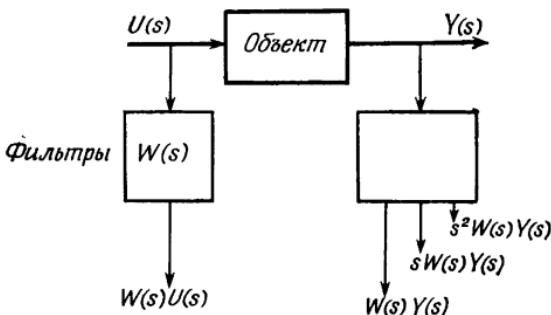
ность рулей. В результате изменяются динамика самолета и его реакция на управляющие воздействия. Команды пилота вызывают реакцию модели (второго порядка), представляющую желаемый отклик самолета. Автоматической настройкой коэффициента усиления K усиление контура управления системы делается максимально большим, так что этот контур образует жесткую систему. Настройка коэффициента усиления до максимального значения производится с помощью выделения малого предельного цикла. Если предельный цикл вообще не выделяется, коэффициент K возрастает; если амплитуда предельного цикла становится слишком большой, K убывает. Таким образом, нет необходимости в явном использовании методов оценивания.

Рассмотрим теперь два простых примера системы управления, в которых *методы оценивания* используются для адаптации. Первый пример представляет простую сервосистему, описанную в работах [21, 22, к разд. 14.1]. Схематически эта система показана на фиг. 14.2. Задачей блока «схема оценивания» является определение коэффициентов передаточной функции

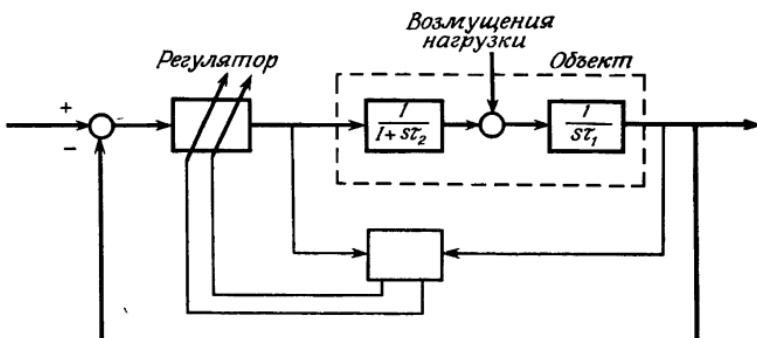
$$H(s) = \frac{1}{a_2 s^2 + a_1 s + a_0}.$$

На основании этих оценок производится настройка передаточной функции $G(s)$ адаптивной схемы так, чтобы с максимальной точностью выполнялось равенство $G(s) = H^{-1}(s)$. Если этого удается достичь, то характеристики системы управления определяются моделью, которую можно выбрать правильно. В схеме оценивания использовались аналоговые фильтры, позволяющие получать сигнал и его производные без применения дифференциаторов (см. фиг. 14.3 и разд. 9.2). С помощью дискретизации по времени и аналого-цифрового преобразования получается система уравнений для трех неизвестных параметров, откуда можно *явными* методами определить оценку α вектора $a = [a_2, a_1, a_0]$.

Второй пример системы управления, где используются методы оценивания, заимствован из [14, к разд. 14.1] (см. также [17, 18 к разд. 14.1]). В этом случае объект представляет собой электронный регулятор скорости



Ф и г. 14.3.

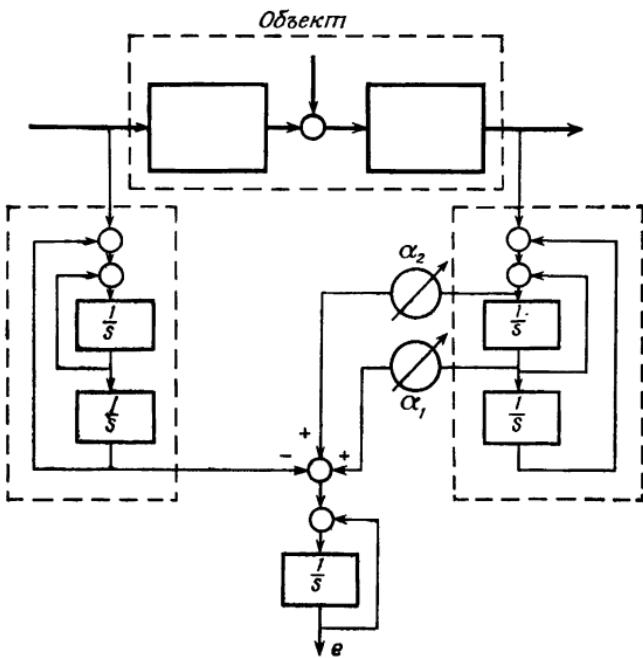


Ф и г. 14.4.

с передаточными функциями (фиг. 14.4)

$$\frac{1}{1+s\tau_2}, \quad \frac{1}{s\tau_1},$$

где τ_1 определяется моментом инерции, а τ_2 — постоянная времени управляющего элемента, передающего момент вращения. Оба параметра $a_1 = \tau_1$ и $a_2 = \tau_1\tau_2$ оцениваются *неявным* методом (*настройкой модели*). Чтобы избежать дифференцирования, снова используются аналоговые фильтры, выделяющие сигнал и его производную; см. фиг. 14.5, где a_1 и a_2 настраиваются с помощью градиентного метода по критерию интегральной квадратичной ошибки. (Для исключения постоянной составляющей ошибки e в схему включен низкочастотный фильтр.) Опираясь на полученные значения параметров, можно настраивать управляющее устройство явными методами,



Фиг. 14.5.

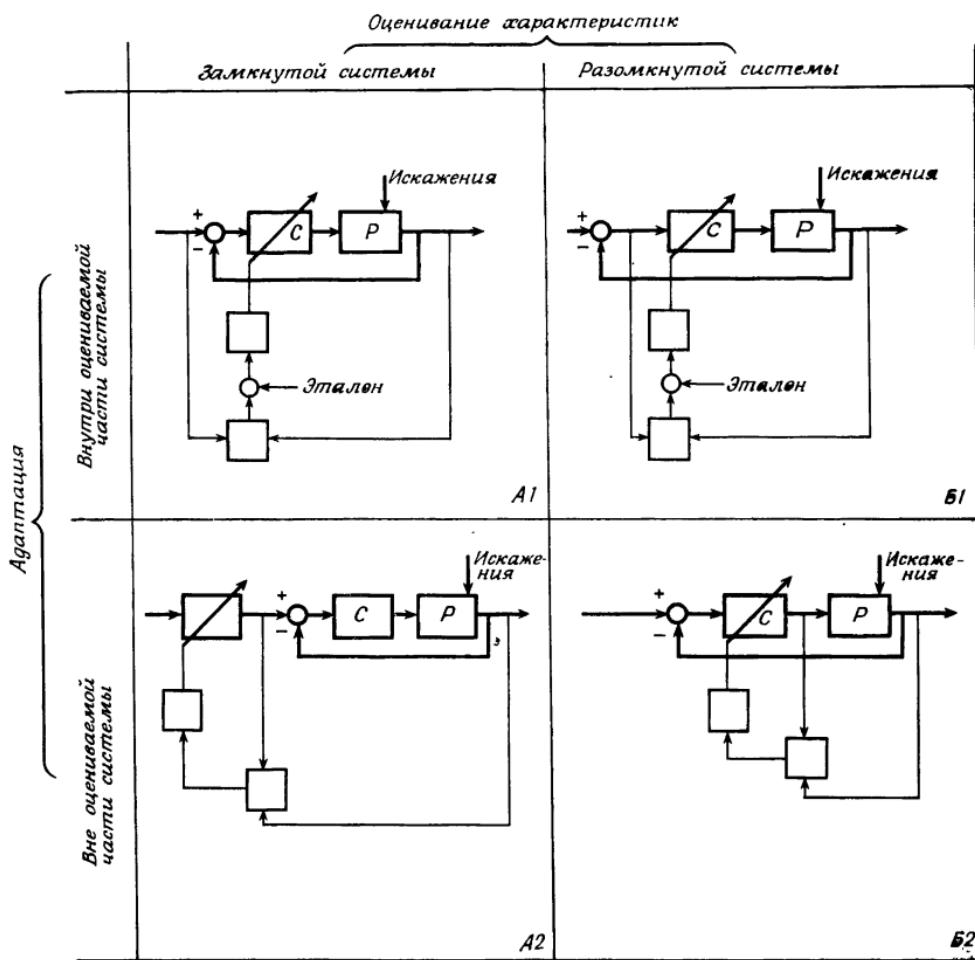
используя «обобщенный симметричный оптимум», как это делается в цитировавшихся работах. Показано, что схемы такого типа работают удовлетворительно; они нашли ряд практических применений (см. также [19, к разд. 14.1]).

На фиг. 14.6 показаны различные структуры адаптивных систем [18, к разд. 14.1]. Основные различия и свойства этих схем таковы:

1) Первый столбец (фиг. 14.6) — *оценивание характеристик замкнутой системы*. Результаты оценивания можно сделать независимыми от возмущений, если эти возмущения не связаны с входным сигналом системы.

2) Второй столбец — *оценивание разомкнутой системы*. Если настраиваемый контур управления неустойчив, то оценивание должно производиться быстро при больших амплитудах сигнала. Однако при обычных условиях один из используемых для оценивания сигналов оказывается малым, так что оценивание происходит медленно. Помехи вызывают смещение оценок.

а) Первая строка — *адаптация внутри оцениваемой*



Ф и г. 14.6.

части системы. Если управляющее устройство C принадлежит к оцениваемой части системы, то нет необходимости полностью определять неизвестные параметры; достаточно проверить, больше они или меньше желаемых величин.

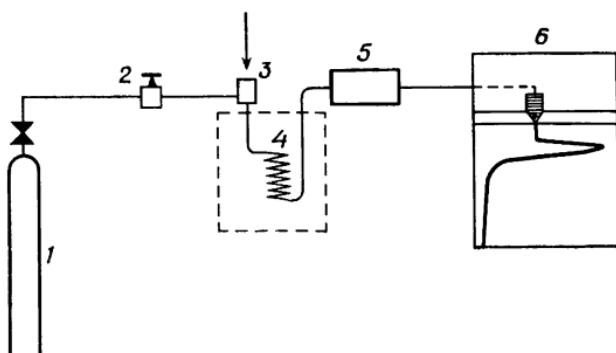
б) Вторая строка — адаптация вне оцениваемой части системы. Быстрое изменение в адаптивном элементе не влияет на процедуру оценивания.

14.2. ФИЗИЧЕСКИЕ, МЕХАНИЧЕСКИЕ И ХИМИЧЕСКИЕ ОБЪЕКТЫ

Существует обширная литература по выбору, обоснованию, использованию и оцениванию моделей для разнообразных целей в этих областях. Вводные замечания об использовании моделей в промышленности для стабилизирующего и оптимального управления, а также некоторые примеры динамических и статических моделей Ван-дер-Гринтена содержатся в обзорной статье [9, к разд. 14.2]. Много работ посвящено моделированию разнообразных технических объектов, например [15, 20, 29, к разд. 14.2]. В рамках данного раздела бесполезно пытаться нарисовать ясную и полную картину многочисленных приложений. Кроме уже упоминавшихся статей, читателю следует обратиться к дополнительной литературе. Ряд замечаний о моделях и методах испытания приводится в работе [20, к разд. 14.2].

Интересные данные об экспериментах, проведенных с использованием псевдослучайных двоичных сигналов в качестве возмущающих воздействий и корреляционных методов, сообщаются в работе [14, к разд. 14.2]. В статье [12, к разд. 14.2] содержится подробное обсуждение применения таких методов в нефтеперегонной промышленности. Сошлемся еще на некоторые работы по применению в конкретных областях: ректификационные колонны — [25, к разд. 14.2]; химические реакции — [4, 6, 21, 23, 27, 28, 30, к разд. 14.2]; парогазовые реакторы — [24, 31, к разд. 14.2]; бумагоделательные машины — [11, к разд. 14.2]; рефрижераторы — [19, 32, к разд. 14.2]; печи и топки — [1, 5, 7, 10, 18, к разд. 14.2]; теплообменники — [8, 16, к разд. 14.2]; обжиг цемента — [2, 26, к разд. 14.2]; газовая хроматография — [3, 13, 17, 22, к разд. 14.2].

Для примера рассмотрим подробнее последнюю статью. На фиг. 14.7 схематически изображен газовый хроматограф. Он состоит из цилиндра 1, заполненного газом, например водородом; редуктора 2, обеспечивающего постоянство потока газа; инжектора 3, вводящего исследуемый образец в газовый поток; колонки 4, т. е. трубки, где происходит разделение компонент (эта колонка под-

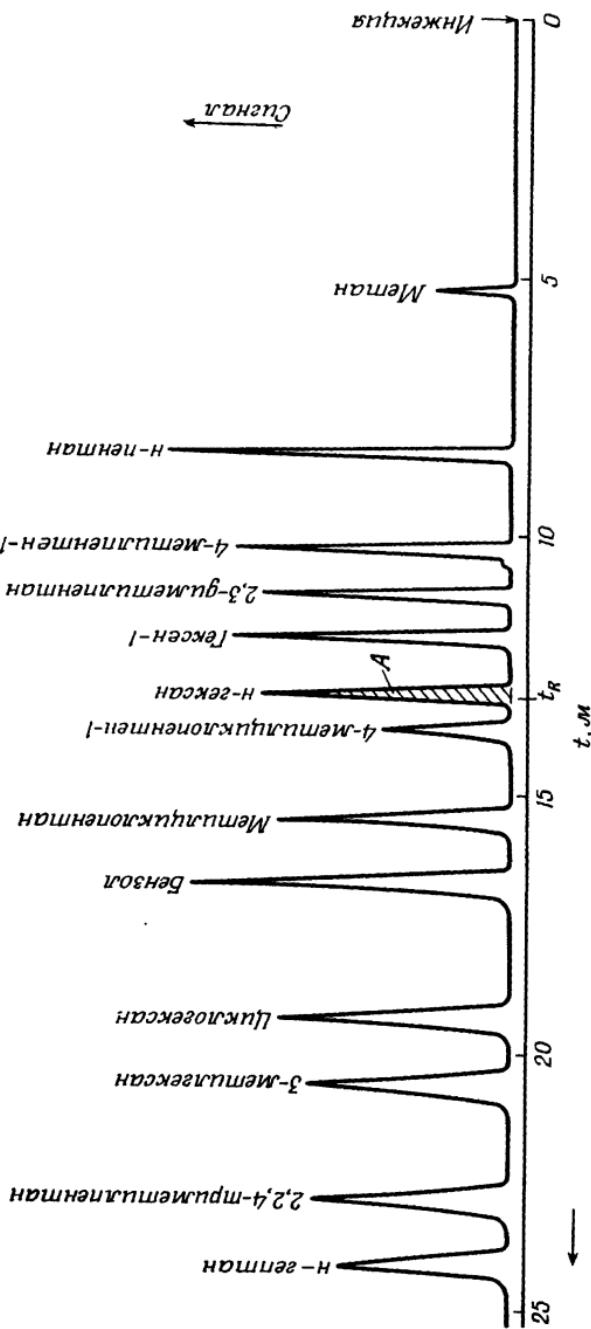


Ф и г. 14.7.

ключена к термостату); детектора 5, измеряющего некоторые физические свойства газа, например теплопроводность; записывающего устройства 6.

Небольшой образец исследуемого вещества, находящийся, как правило, в жидком виде, испаряется в газовом потоке. Благодаря физическим свойствам колонки молекулы образца отстают от газового потока, причем это отставание различно для молекул разных типов. В конце колонки образец обнаруживается в виде последовательности выбросов на записываемой хроматограмме (фиг. 14.8). Временная задержка по отношению к моменту инъекции (время запаздывания) t_R служит характеристикой каждого компонента образца, площадь под выбросом A является мерой количества этого компонента в образце. Ширина (стандартное отклонение σ) выброса также является важной характеристикой. Хотя основное внимание уделяется свойствам образца (входного «сигнала»), эти свойства можно считать также характеристиками объекта (измерительного прибора), осуществляющего разделение. Поэтому задачу можно сформулировать как задачу оценивания (для каждого выброса) вектора параметров $\beta = [t_R, A, \sigma]'$. На практике хроматограммы могутискажаться шумом, связанным, например, с загрязнением несущего газа, флуктуациями температуры, электрическими шумами в детекторе и усилителе. Методика, предложенная в цитированной статье, предусматривает (в предположении аддитивности и стационарности шума) следующие операции;

Ф и г. 14.8.



оценивание свойств шума при отсутствии полезного сигнала; пропускание сигнала через фильтр, преобразующий шум к белому; получение оценки $\hat{\beta}$.

14.3. ЯДЕРНЫЕ РЕАКТОРЫ

В этой области также предлагалось и реализовывалось немало схем оценивания. Приведем несколько новых характерных примеров:

а) Измерение подкритического коэффициента деления (нейтронов); см. [1, к разд. 14.3]. Решение достигается введением источника нейтронов, модулируемого псевдослучайной двоичной последовательностью, и определением корреляции получающегося сигнала (с задержкой) с величиной плотности нейтронного потока.

б) Определение кинетических параметров реактора по корреляционным измерениям, основанным на статистике флуктуаций нейтронных цепных реакций; см. работы [3, 8, к разд. 14.3], где используется метод корреляции знаков [4, к разд. 14.3] и применяется квазилинеаризация.

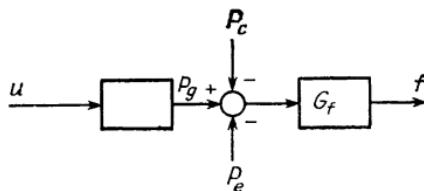
в) Оценивание параметров спаренных реакторов [2, 6, к разд. 14.3].

г) Оценивание состояния с помощью фильтра Калмана — Бьюси в условиях, когда зависящая от времени реактивность рассматривается как одна из переменных состояния [7, к разд. 14.3], а также совместное оценивание параметров и состояния посредством решения двухточечной краевой задачи [5, к разд. 14.3].

14.4. ЭНЕРГЕТИЧЕСКИЕ И ЭНЕРГОРАСПРЕДЕЛИТЕЛЬНЫЕ СИСТЕМЫ

а) Для успешного решения проблем, возникающих при частых изменениях нагрузки, требуется много информации и контроль за параметрами энергетических систем. Примеры литературы по задачам оценивания в энергетике: [3, 11, 13, к разд. 14.4].

б) В задачах диспетчеризации можно пользоваться оценками состояния (и параметров) [8, 10, 14, 15, к разд. 14.4].



Ф и г. 14.9.

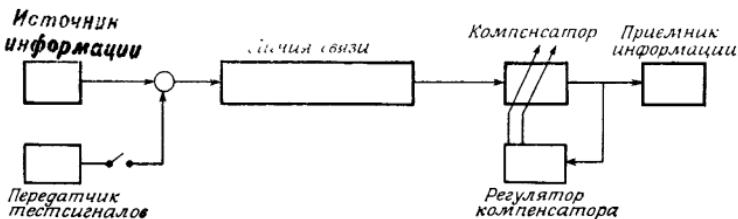
в) Для настройки параметров устройства, управляющего частотой нагрузки, нужно знать передаточную функцию G_f в соответствии с фиг. 14.9, где u — управляющий сигнал от генератора; P_g , P_c , P_e — генерируемая, потребляемая и «экспортируемая» мощности, а f — частота. См. [1, 4, 9, к разд. 14.4]. Общий интерес представляет работа [7, к разд. 14.4].

14.5. СВЯЗЬ

Существует тесная связь между многими задачами в области связи и в области оценивания состояний и параметров, в частности доминирующее значение возмущений (шумов), обусловливающих принципиальные ограничения на достижимые результаты. Подобная связь очевидна в задачах разделения сигнала и шума, оценивания параметров сигналов (например, расстояния и скорости объекта при радиолокации), оценивания синхронизации импульсов и т. д. (см. [6, к разд. 14.5]). Ниже кратко перечисляются некоторые конкретные применения:

а) Ослабление и сдвиг по фазе линий связи зависят от длины, типа кабеля, температуры, влажности и т. д. Для устранения по возможности нежелательного изменения характеристик и проявления отрицательных факторов используются специальные компенсаторы. Различные типы компенсаторов рассматриваются в работе [1, к разд. 14.5].

Автоматическая компенсация стала популярным объектом исследований и применений. Много статей посвящено адаптивным фильтрам, которые автоматически подстраиваются к изменениям в характеристиках линии связи, например, так, чтобы минимизировать взаимодействие между символами. Используемые методы можно рас-



Ф и г. 14.10.

сматривать как разновидность метода настраиваемой модели, когда модель (компенсатор) соединяется последовательно с нестационарным объектом (фиг. 14.10). Распространенным типом компенсатора является линия задержки с несколькими ответвлениями и настраиваемым ослабителем в каждом ответвлении. Обзор методов автоматической компенсации приводится в работе [3, к разд. 14.5]; много статей по этому вопросу указано в списке дополнительной литературы.

б) Использование адаптивных фильтров в качестве избирательных частотных ограничителей [7, к разд. 14.5].

в) Адаптивное изменение избыточности при передаче информации [5, к разд. 14.5].

г) Адаптивные антенные системы [8, к разд. 14.5].

д) Оценивание характеристик многоканальных систем связи [2, к разд. 14.5].

е) Оценивание или измерение электрических (электронных) характеристик самонастраивающимися мостами [4] или методами настраиваемой модели (фиг. 1.7 и 9.7).

14.6. АВИАЦИЯ И КОСМОНАВТИКА

В разд. 14.1 упоминались применения идентификации для целей управления на современных самолетах. Подобные адаптивные системы могут быть рассчитаны: 1) на компенсацию изменений динамических характеристик летательного аппарата, т. е. создание системы управления с обратной связью с хорошими показателями устойчивости и управляемости в условиях, когда характеристики разомкнутой системы могут подвергаться сильным изменениям, или 2) на компенсацию внешних возмущений,

связанных с атмосферной турбулентностью и внешними управляющими воздействиями.

Обзор подобных применений в авиации и космонавтике см. в [7, к разд. 14.6]. Ряд статей о конкретных схемах адаптивного управления можно найти в [1, к разд. 14.6]. Всего лишь несколько статей из большого числа публикаций по этим вопросам приведено в списке дополнительной литературы.

Для определения неизвестных параметров авиационной и космической техники применялся ряд методов [6, к разд. 14.6]:

оценивание динамики спутника итерационным методом наименьших взвешенных квадратов [8, к разд. 14.6];

оценивание аэродинамических коэффициентов ракет с помощью анализа чувствительности методом интегральных наименьших квадратов [3, к разд. 14.6];

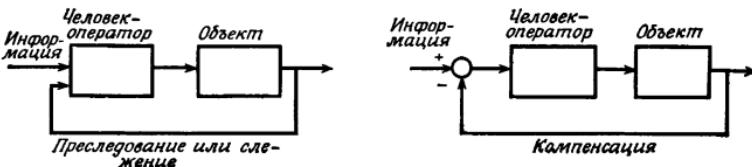
оценивание методами частотных характеристик [2, к разд. 14.6];

оценивание параметров авиационной техники с помощью квазилинеаризации [4, к разд. 4.6] и методом максимума правдоподобия [5, к разд. 4.6]. Еще Гаусс применял метод наименьших квадратов для определения переменных состояния в астрономии. И теперь эта процедура оценивания широко используется для вычисления орбит космических аппаратов. (См. литературу, цитированную в гл. 12 и 13.)

14.7. БИОЛОГИЧЕСКИЕ ОБЪЕКТЫ

Из-за большой сложности живых организмов, трудностей непосредственного измерения существенных величин, важности более точных диагнозов и лучшего лечения болезней методы оценивания следует считать, по крайней мере потенциально, заслуживающими внимания в данной области. Важная роль этих методов тесно связана с изучением кибернетических аспектов (опирающихся на понятия системы и управления) в биологических объектах; примеры из этой области содержатся в ряде книг [2, 14, 15, 20, к разд. 14.7].

Человек-оператор. В течение ряда лет весьма популярным объектом исследований оставался человек как



Ф и г. 14.11.

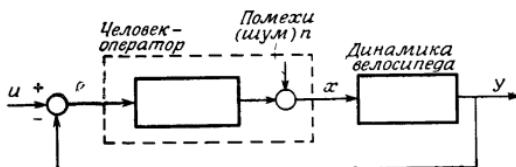
элемент системы управления; см. фиг. 1.5. Одной из первых является работа [21, к разд. 14.7]. Интерес к вопросам идентификации и оценивания человека-оператора обусловливается прежде всего задачами управления различными сухопутными, водными, воздушными, космическими и лунными средствами передвижения. Важными вопросами при этих исследованиях являются трудности, связанные с выбором способа управления; изменения динамических характеристик оператора со временем в результате обучения или из-за усталости; адаптационные способности оператора, проявляющиеся в реакции на неожиданные изменения динамики объекта; например из-за поломок и т. д. Проблемам этого типа уделялось много внимания [6, 7, 8, к разд. 14.7]. Исчерпывающий список литературы приведен в [9, к разд. 14.7]; много интересных материалов содержится в трудах конференций НАСА по управлению с участием человека [16, к разд. 14.7]. Появление значительного количества разнообразных статей обусловлено многочисленностью аспектов проблемы, среди которых можно выделить, например:

а) тип управления: преследование или компенсация (фиг. 14.11), одно или несколько направлений (степеней свободы);

б) тип требуемой информации: весовая функция; коэффициенты передаточной функции, постоянные или случайным образом меняющиеся со временем; нелинейности; размер выборки;

в) тип входных характеристик: детерминированный многочастотный сигнал, случайный сигнал с заданной плотностью вероятности и спектральной плотностью;

г) динамика объекта: неустойчивость, нелинейность, внезапные изменения (отказы);



Ф и г. 14.12.

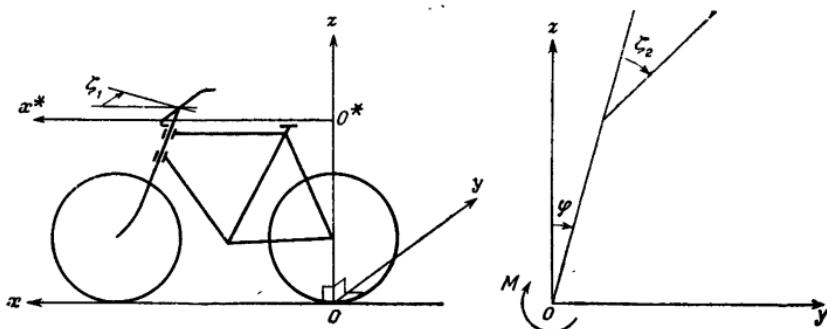
- д) тип предварительной обработки сигналов, например ускорение, предсказание;
- е) влияние внешней среды на оператора;
- ж) тип применяемых методов оценивания.

Лишь весьма ограниченное число публикаций упоминается в списке дополнительной литературы.

Тот факт, что человек-оператор включен в замкнутый контур управления (фиг. 14.12), приводит к возникновению существенной проблемы циркулирующего шума, т. е. сигналы n и e оказываются коррелированными. Следовательно, использование для оценивания только сигналов e и x может привести к смещению оценок. Один из подходов к этой проблеме заключается, например, в сведении ее к методам, в которых используются вспомогательные переменные, получаемые из сигнала u .

Большинство исследований НАСА поведения человека-оператора при моделировании экстраординарных заданий проводилось высококвалифицированными и тренированными людьми. Опишем кратко один интересный эксперимент, касающийся человека на улице, а именно велосипедиста. Это описание основано на частном сообщении¹⁾ и статье [19, к разд. 14.7]. Модель представляет собой жестко закрепленную раму велосипеда, усилия к которой прикладываются гидравлической системой управления. Задачей испытуемого является «управление» велосипедом, т. е. удержание в равновесии этой неустойчивой по своей природе системы. Гидравлическая система обеспечивает естественные усилия, связанные с гироскопическим момен-

¹⁾ Р. Дж. Бойтен, Г. Дж. Стассен, А. ван-Лантерен, годовой отчет группы по системам «человек — машина» (1969), Дельфтский технологический университет, инженерно-механический факультет, лаборатория технических измерений и управления.



Ф и г. 14.13.

том переднего колеса и центробежным моментом велосипеда при движении по криволинейному пути. Правила отсчета знаков иллюстрируются фиг. 14.13; принятые следующие обозначения:

$\zeta_1(t)$ — угол между рулем и рамой;

$\zeta_2(t)$ — угол между верхней частью тела и рамой;

$\varphi(t)$ — угол между осью Oz и рамой;

J, M — момент инерции и статический момент велосипеда и нижней части тела относительно оси Ox ;

J^*, M^* — момент инерции и статический момент верхней части тела относительно оси O^*x^* ;

z_1 — расстояние между осями O^*x^* и Ox ;

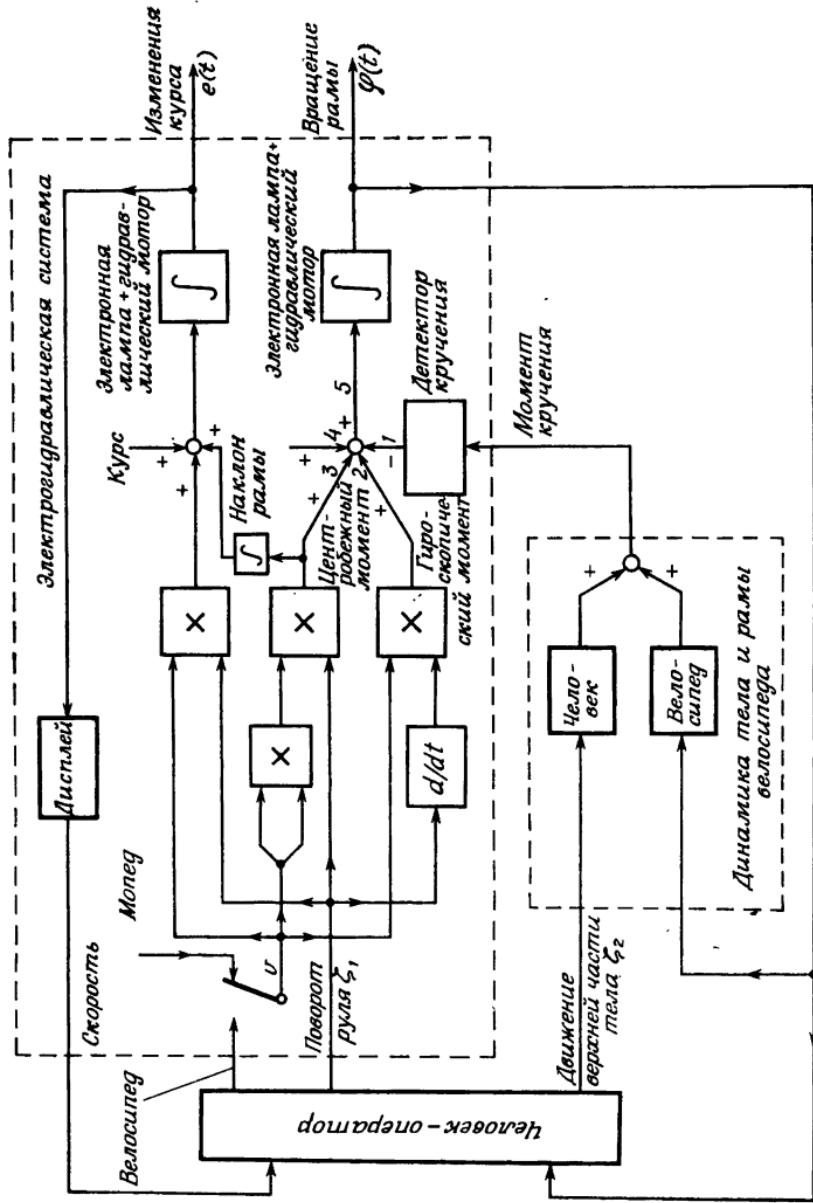
v — поступательная скорость;

A, B — постоянные.

При незначительных упрощающих предположениях моменты относительно оси Ox удовлетворяют следующему соотношению:

$$Av\dot{\zeta}_1(t) + Bv^2\dot{\zeta}_2(t) = \{J\ddot{\varphi}(t) - Mg\varphi(t)\} + \\ + \{(J^* + z_1M^*)\dot{\zeta}_2(t) - M^*g\zeta_2(t)\}.$$

Первое слагаемое слева представляет собой гирокопический момент переднего колеса, а второе слагаемое — центробежный момент при движении по криволинейному пути; первое слагаемое в фигурных скобках справа — момент, возникающий при отклонении рамы и нижней части тела от положения равновесия, а второе слагаемое — момент, возникающий при отклонении верхней части тела



Фиг. 14.14.

относительно рамы. На фиг. 14.14 приведена блок-схема экспериментальной установки. С помощью курсового тест-сигнала человеку-оператору указывается цель; отклонение от курса $e(t)$ служит сигналом ошибки. В этой экспериментальной ситуации основное внимание обращается на динамические свойства человека-оператора. Модель оператора представляется следующими выражениями:

передаточная функция от вращения рамы к вращению руля

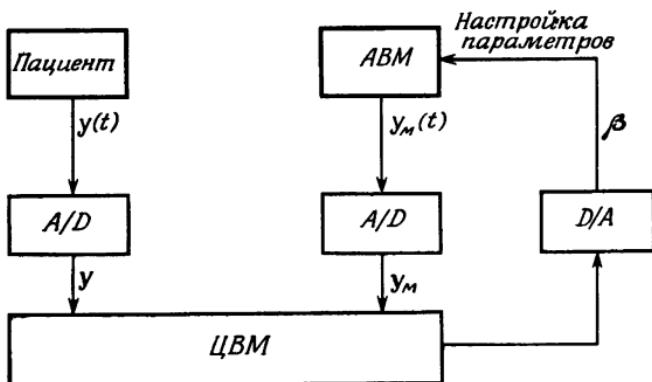
$$H_1(s) = \left(\alpha_{11} + \frac{\alpha_{12}}{s} + \alpha_{13}s \right) e^{-\tau_1 s},$$

передаточная функция от вращения рамы к движению верхней части тела

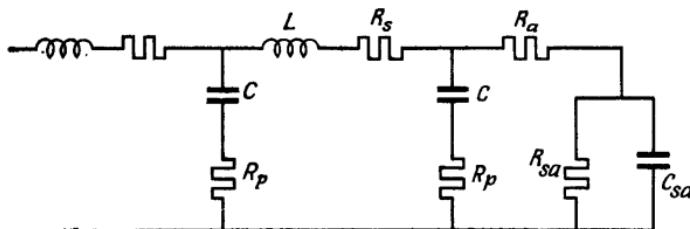
$$H_2(s) = \left(\alpha_{21} + \frac{\alpha_{22}}{s} + \alpha_{23}s \right) e^{-\tau_2 s}.$$

Использовались методы настраиваемой модели с помощью небольшой цифровой вычислительной машины. По-прежнему необходима осторожность по отношению к проблемам, связанным с циркулирующим шумом. Одно из возможных направлений исследований заключалось в изучении влияния медикаментов и алкоголя на различные параметры системы.

Кровообращение. Предпринималось немало попыток построить (электрические) модели системы кровообращения (в целом или какой-либо ее части), в частности у человека. Сердечно-сосудистый аппарат представляет систему с распределенными параметрами, динамическое поведение которой необходимо описывать уравнениями в частных производных. Исследовались разные способы аппроксимации, например путем моделирования на аналоговых вычислительных машинах [4, 11, 17, к разд. 14.7]. Одной из целей подобных исследований может быть определение параметров системы кровообращения, которые трудно измерить непосредственно, например эластичность вен или артерий. Как показано на фиг. 14.15, аналоговая модель настраивается цифровой вычислительной машиной таким образом, чтобы результаты, получаемые на модели, оказывались по возможности наиболее близкими к реальным данным, получаемым от пациентов. Пример модели



Ф и г. 14.15.

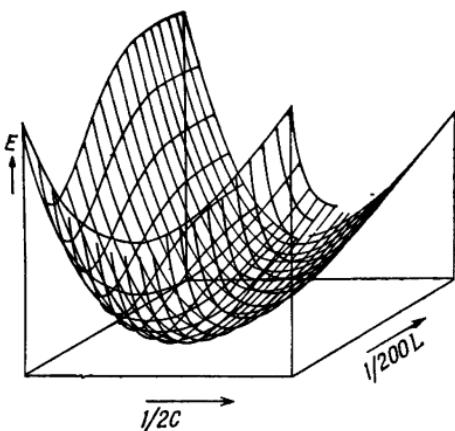


Ф и г. 14.16.

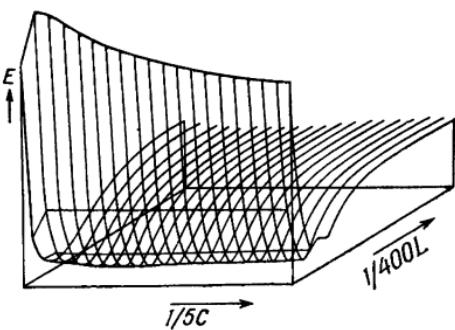
аорты человека¹⁾ показан на фиг. 14.16. Инерцию крови представляют индуктивности, эластичность стенок сосудов — емкости. Сопротивления моделируют гидравлическое сопротивление и эффекты вязкости в стенках. Очевидно, что это всего лишь один из многих возможных путей моделирования аорты.

На фиг. 14.17 показана функция ошибки, или показатель эффективности этой модели, в зависимости от изменений величин индуктивностей и емкостей относительно их номинальных значений. Интересное явление, которое может служить и предостережением, иллюстрируется на фиг. 14.18. Существенные изменения в поведении функции ошибки вызваны большими изменениями тех же самых

¹⁾ Частное сообщение К. А. М. ван-ден-Брекеля и Г. В. Тиммера, электротехнический факультет Технологического университета Эйндховена (Голландия).



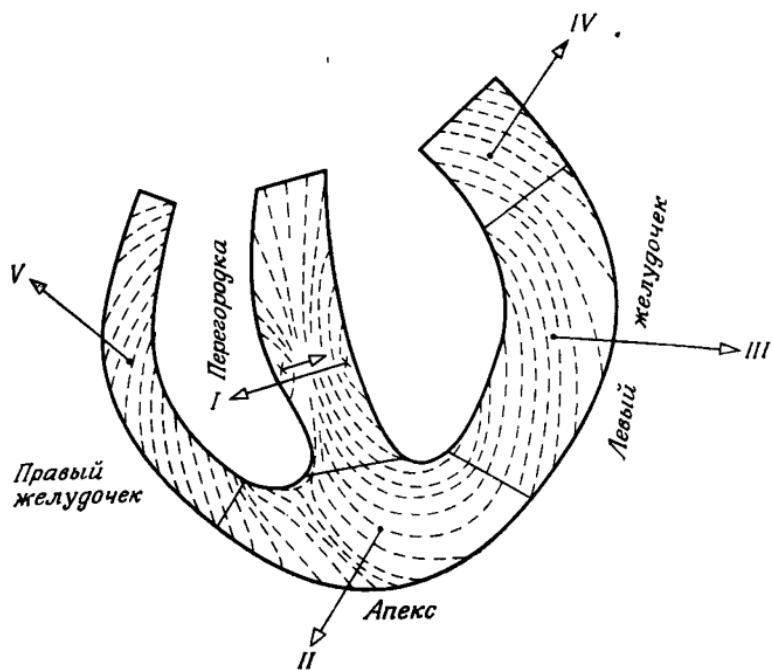
Ф и г. 14.17.



Ф и г. 14.18.

параметров относительно тех же номинальных значений, что и на фиг. 14.17. В списке дополнительной литературы упомянуто несколько работ по этому вопросу.

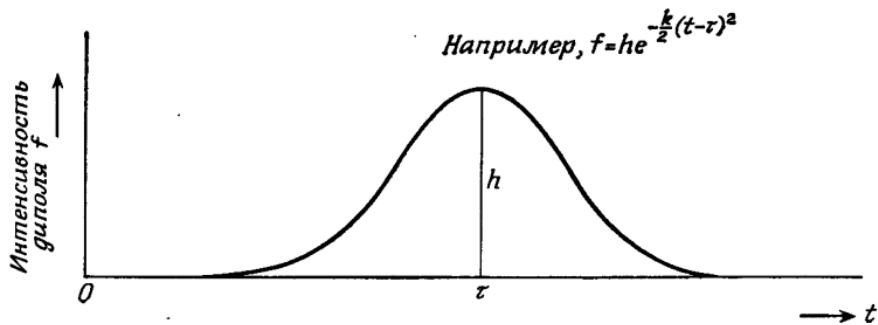
Электрическая активность сердца. Благодаря электрическим явлениям в сердечной мышце (волна деполяризации) электрокардиограмму можно снимать на поверхности тела. С помощью адекватной модели сердца и свойств окружающих частей тела можно моделировать электрокардиограмму. Такая модель может, например, состоять из m диполей, характеризующих различные сегменты сердечной мышцы (фиг. 14.19), причем каждый диполь определяется своей функцией интенсивности (фиг. 14.20).



Ф и г. 14.19.

Если эти функции можно задать с помощью n параметров, модель содержит $m \times n$ параметров.

Комбинации сигналов, представляющих потенциалы электродов, являются ортогональными распределенными в пространстве сигналами, приблизительно равными алгебраическим суммам функций интенсивностей дипо-



Ф и г. 14.20.

лей¹⁾ (фиг. 14.21). Принципиальным является вопрос о возможности определения параметров этих диполей по измеренной электрокардиограмме. Если такая возможность существует, то с помощью методов оценивания, например по настраиваемой модели, можно локализовать дефекты сердца и определить их степень. См. по этому поводу [3, 10, к разд. 14.7].

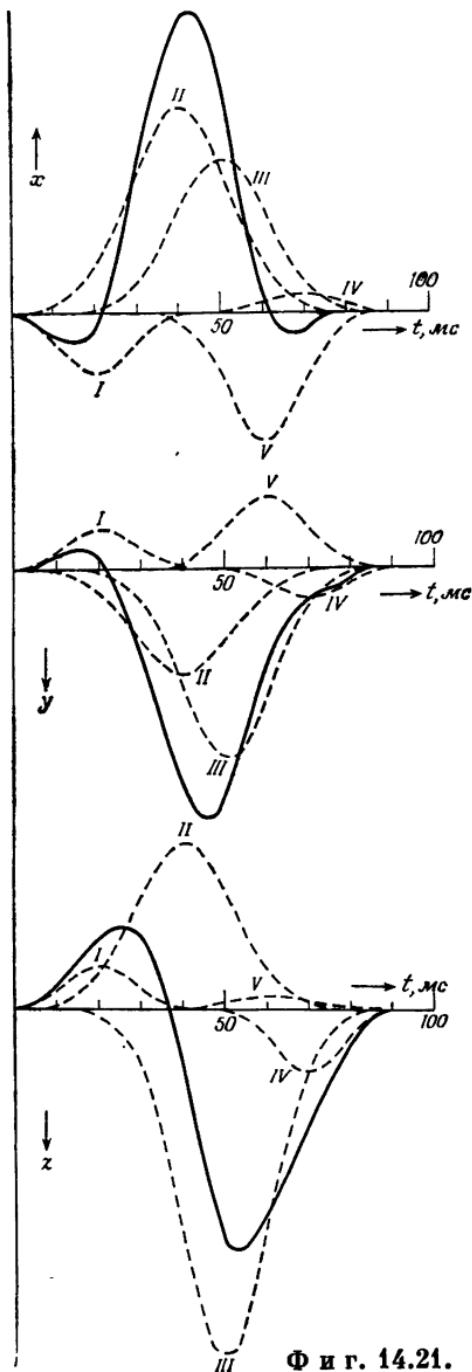
Среди других приложений к биологическим объектам отметим лишь моделирование:

- а) мышечной активности [1, 13, к разд. 14.7];
- б) с целью анализа электрокардиограммы [12, к разд. 14.7];
- в) зрительной аккомодации у человека [18, к разд. 14.7].

14.8. СОЦИАЛЬНО-ЭКОНОМИЧЕСКИЕ СИСТЕМЫ

В течение многих лет специалисты по эконо-метрике занимались

¹⁾ Частное сообщение А. А. Г. Дамена, электротехнический факультет Технологического университета Эйндховена (Голландия).



Ф и г. 14.21.

разработкой моделей основных частей национальной экономики и других областей экономической деятельности. Рассмотрение примеров таких моделей не входит в круг задач этой книги. Ограничимся несколькими ссылками на литературу общего характера: [2, 8, 9, к разд. 14.8]. Эти методы нашли развитие в так называемой индустриальной динамике, городской динамике и мировой динамике [3, 4, 5, к разд. 14.8]; см. также [10, к разд. 14.8].

14.9. ОБУЧАЮЩИЕСЯ СИСТЕМЫ И РАСПОЗНАВАНИЕ ОБРАЗОВ

Читатель должен помнить о тесной связи, существующей между процедурами оценивания параметров и задачами, относящимися к обучающимся системам и распознаванию образов. Эта связь объясняется общностью проблем выбора структуры модели и способов настройки весовых коэффициентов (параметров). Опять в списке дополнительной литературы указаны лишь некоторые работы по этим вопросам.

Природа этой главы такова, что в зависимости от интересов читателя она может показаться содержащей как слишком много, так и чрезмерно мало материала. Имелось в виду только проиллюстрировать чрезвычайное многообразие целей и средств оценивания параметров и состояния, предлагавшихся и использовавшихся к настоящему времени. Не следует забывать о возможностях, предоставляемых индексом цитируемости при поиске более важных публикаций по проблематике, затрагиваемой какой-либо определенной статьей.

Возникшая благодаря конкретным нуждам в некоторых частных областях, стимулируемая широкими потенциальными возможностями и уже продемонстрированными достижениями, поддерживаемая растущей стоимостью оборудования для обработки информации, теория оценивания параметров и состояния интенсивно развивается. В этой области науки нетрудно увидеть одно из средств улучшения жизни человека, связанного, например, с усилением экономического развития, ростом безопасности полетов благодаря применению этих методов для управления и контроля профилактического обслуживания, созданием

более совершенных моделей социально-экономических явлений, вкладом в улучшение здравоохранения и т. д. и т. п.

В данный момент еще невозможно точно предсказать ни возможные приложения, ни окончательные ограничения применимости рассматривавшихся методов. Необходимость в дальнейших теоретических обоснованиях и в накоплении опыта практических приложений очевидна как из литературы, так и, в частности, из данной книги. Хочется надеяться, что эти исследования послужат на благо человечества.

Г. У. Лонгфелло, «Псалом жизни»

Не восторги, не печали
 Рок готовит нам в конце,
 Только ближе, чем вначале,
 Будет завтра наша цель.

ЛИТЕРАТУРА К РАЗД. 14.1

1. Aseltine J. A. et al., A Survey of Adaptive Control Systems, *IRE Trans. Autom. Control*, PGAC-6, 102—108 (1958).
2. Åström K. J., Wittenmark B., Problems of Identification and Control, Report Electronic Sciences Laboratory, Univ. of South Calif., Los Angeles, Calif. J. Math. Anal. Appl., 1969.
3. Draper C. S., Li V. T., Principles of Optimalizing Control Systems and an Application to the Internal Combustion Engine, ASME Publication, N.Y., 1951.
4. Фельдбаум А. А. (см. литературу к гл. 1).
5. Gibson J. E., Mechanising the Adaptive Principle, *Control Engng.*, 7, 109—115 (1960).
6. Jacobs O. L. R., A Review of Self-Adjusting Systems in Automatic Control, *J. Electronics and Control*, 10, 311—322 (1961).
7. Jacobs O. L. R., Two Uses of the Term «Adaptive» in Automatic Control, *IEEE Trans. Autom. Control*, AC-9, 574—575 (1964).
8. Kalman R. E. (см. литературу к гл. 7).
9. Красовский А. А., Динамика непрерывных систем автоматического регулирования с экстремальной настройкой корректирующих устройств, в книге «Теория дискретных оптимальных и самонастраивающихся систем», Труды I Межд. конгресса ИФАК, Изд-во АН СССР, 1961.
10. Mathias R. A., Van Nice R. I., Adaptive Control Systems: a Survey, *Electro Technology* (1960).
11. Mendes M., An on Line Adaptive Control Method, IFAC Symp. «Identification and Process Parameter Estimation», Prague, Paper 6.5, 1970; см. также *Automatica*, 7 (1971); Dual Adaptive

- Control of Time-Varying Systems (на немецком), *Regelungstechnik*, 18, 157—164 (1970).
12. Mishkin E., Braun L., Adaptive Control Systems, McGraw-Hill, N.Y., 1961.
 13. Pun L., The Evolution of Optimal Control (на французском), *Automatisme*, 6, 47—53, 103—111, 148—152 (1961).
 14. Rumold G., Speth W., Self-Adjusting PI Controller (на немецком), *Siemens Z.*, 42, 765—768 (1968).
 15. Рутман Р. С., Самонастраивающаяся система с настройкой по динамической характеристике, *Автоматика и телемеханика*, № 9 (1962).
 16. Snegur A., Tomashov V., Dubina L., Synthesis of Chemical Reactor Adaptive Control System and Analysis of System Stability with Inaccurate a Priori Knowledge on Reactor Model, IFAC Symp. «Identification in Autom. Control», Prague, Paper 7.14, 1967.
 17. Speth W., Simple Methods for the Rapid Self-Adaptation of Automatic Controllers in Drive Applications, Fourth IFAC Congress, Warsaw, Paper 50.4, 1969.
 18. Speth W., Self-Adjusting Control Systems in Electrical-Drive Technology (на немецком), Dr.-Ing. Dissertation, Techn. Univ., Braunschweig, Germany, 1971.
 19. Spiegel P., Parameter Control in Electrical Drives (на немецком), *Zeitschr. Messen, Steuern, Regeln*, 12, 197—202 (1969).
 20. Whitakker H. P. et al., Design of a Model Reference Adaptive Control System for Aircraft, Report R-164, Instrumentation Lab., MIT, Cambridge (Mass.), 1958.
 21. Young P. C., Process Parameter Estimation and Self-Adaptive Control, in: The Theory of Self-Adaptive Control Systems, Hammond P., ed., Teddington, England, 1965, pp. 118—139.
 22. Young P. C., Applying Parameter Estimation to Dynamic Systems, *Control Engng*, 16, Oct. 119—125 (1969); Nov. 118—124 (1969).

ДОПОЛНИТЕЛЬНАЯ ЛИТЕРАТУРА

Bakke R. M., Adaptive Direct Digital Control with Multi-Parameter Adjustment, Paper IFAC-Symp. «Theory of Self-adaptive Control Systems», Teddington, 1965.

Bohn E. V., Butler R. E., Mukherji M. R., Parameter-Tracking Models for Adaptive Control Systems, *Proc. IEE*, 113, 378 (1966).

Box G. E. P., Jenkins G. M., Mathematical Models for Adaptive Control and Optimization, Univ. of Wisconsin, Department of Statistics, Techn. Rept. 45, 1965.

Braun L., Jr., On Adaptive Control Systems, *IRE Nat. Conv. Rec.*, Part 4, 32—44 (1959); *IRE Trans. Autom. Control*, AC-4, 30—42 (1959).

Braun L. E., Mishkin E., Truxal J. G., Approximate Identification of Process Dynamics in Computer Controlled Adaptive Systems, First IFAC congress, Moscow, 1960; есть русский перевод: Браун Л., Мишкін Э., Траксель Дж., О приближенном определении дина-

мических характеристик регулируемого процесса в самонастраивающейся системе, в книге «Теория дискретных оптимальных и самонастраивющихся систем», Труды I Межд. конгресса ИФАК, Изд-во АН СССР, 1961.

Бутковский А. Г., Лернер А. Я., Об оптимальном управлении системами с распределенными параметрами, *Автоматика и телемеханика*, № 6 (1960).

Chang S. S. L., On the Relative Time of Adaptive Processes, *IEEE Trans.*, AC-10, 11—17 (1965).

Choquette P., Noton A. R. M., Watson C. A. G., Remote Computer Control of an Industrial Process, *Proc. IEEE*, 58, 10—16 (1970).

Eveleigh V. W., Adaptive Control and Optimization Techniques, McGraw-Hill, N.Y., 1967.

Gibson J. E., Nonlinear Automatic Control, McGraw-Hill, N.Y., 1963.

Graupe D., Cassir G. R., Adaptive Control by Predictive Identification and Optimization (Short Paper), *IEEE Trans. Autom. Control*, AC-12, 191—199 (1967).

Haimes Y. Y., Wismer D. A., A Computational Approach to the Combined Problem of Optimization and Parameter Estimation, *Automatica*, 8, 337—347 (1972).

Hougen J. O., Experimental Techniques and the Synthesis of Control Systems, Progress in Engineering, vol. 3, Heyward Books, London, 1966.

Kershaw W. V., Adaptive Control by Plant Identification, *Control Engng.*, 12, 103 (1965).

Kondo B., Suzuki T., Statistical Consideration of an Optimal Control System Based on the Alternate Parameter-Adjustment Method, *Electrical Engineering in Japan*, 84, 88 (1964).

Kondo B., Suzuki T., Adaptive Control of Correlation-type Optimizing Control System, *Electrical Engineering in Japan*, 85, 1 (1965).

Kumar K. S. P., Sridhar R., A Note on Combined Identification and Control (Correspondence), *IEEE Trans. Autom. Control*, AC-9, 118 (1964).

Кунцевич В. М., Исследование устойчивости простейших самонастраивающихся систем, *Автоматика и телемеханика*, № 8 (1965).

Macmillan R. H., Rees N. W., Automatic Control Systems, *Process Control and Automation*, 575—580 (1960); 623—629 (1960).

Маслов Е. П., Осовский Л. М., Самонастраивающаяся система управления с моделью, *Автоматика и телемеханика*, № 6 (1966).

Moe M. L., Murphy G. J., An Approach to Self-Adaptive Control Based on the Use of Time Moments and a Model Reference, *IRE Trans. Autom. Control*, AC-7, 82—92 (1962).

Narendra K. S., McBride L. E., Multiparameter Self Optimizing Systems Using Correlation Techniques, *IEEE Trans. Autom. Control*, AC-9, 31 (1964).

Nikiforuk P. N., Lashyn E. K., Self-Zeroing Self-Adaptive Control System, *Proc. IEE*, 110 (1963).

Peterson I., Jaaksoo Y., On the Optimal Adaptive Control of Linear Multidimensional Plant, IFAC Symp. «Identification in Autom. Control Systems», Prague, Paper 5.11, 1967.

Reinisch K., The Effect of Errors in Process Analysis on the Synthesis of Control Loops and the Consequences on the Model Method for the Analysis of Control Loops (на немецком), *Zeitschr. Messen, Steuern, Regeln*, 9, 183—187 (1966).

Sage A. P. (см. литературу к гл. 12).

Sano A., Terao M., Measurement Optimization in Batch Process Optimal Control, Fourth IFAC Congress, Warsaw, Paper 12.4, 1969; см. также *Automatica*, 6 (1969).

Sarachik P. E., Identification of the Steady State Operator for Discrete Self-Optimizing Systems, *IEEE Trans.*, AC-10, 83—85 (1965).

Sawaragi Y., Sunahara Y., Nakamizo T., Statistical Decision Theory in Adaptive Control Systems, Academic Press, N.Y., 1967.

Schmidt G., A Self-Adjusting Control Loop with a Reference Model (на немецком), *Regelungstechnik*, 10, 145—151 (1962).

Sklansky J., Learning Systems for Automatic Control, *IEEE Trans. Autom. Control*, AC-11 (1966).

Staffin W., Executive-Controlled Adaptive Systems, *AIEE Trans.*, Part II (Applics. and Ind., № 46), 78, 523—530 (1960).

Štěpán J., The Identification of Controlled Systems and the Limit of Stability, IFAC-Symp. «Identification in Autom. Control Systems», Prague, Paper 1.11, 1967.

Stout T. M., ed., Computers in Industrial Process Control, *Proc. IEEE*, 58, 1—132 (1970).

Sworoder D. D., A Study of the Relationship Between Identification and Optimization in Adaptive Control Problems, *J. Franklin Inst.*, 281, 198 (1966).

Takahashi Y., Rabens M. J., Auslander D. M., Control and Dynamic Systems, Addison-Wesley, London, 1970.

Tou J. T., Digital and Sampled Data Control Systems, McGraw-Hill, N.Y., 1959.

Vasu G., Self-Adaptive System for Automatic Control of Dynamic Performance by Controlling Gain, Phase Shift, Gain Margin, Phase Margin or Slope, Paper AIEE, 60—1381 (Fall General Meeting, Chicago), 1960.

Wells C. H., Larson R. E., Application of Combined Optimum Control and Estimation Theory to Direct Digital Control, *Proc. IEEE*, 58, 16—22 (1970).

ЛИТЕРАТУРА К РАЗД. 14.2

1. Bailey A. J., Modelling of a Steel Reheating Furnace, *IEE*, London (1968).
2. Bailey M. J., On Line Dynamic Modelling for Improved Control, *IEE*, 177—184 (1969).
3. Booth A. D., Resolution of (Complex) Gas Chromatograms Using a Digital Computer, *Trans. Soc. Instrum. Technol.*, 19, 12—16 (1967).
4. Bremermann H., Identification of Rate Constants in Chemical Cycles of Known Structure, IEEE Symposium on Adaptive Processes, Decision and Control, Austin, Tex., U.S.A., 1970.

5. Brown P., The Adaptive Control of Electric Arc Furnaces, *IEE* (1969).
6. Буровой И. А., Мамедов А. А., Построение структуры модели для идентификации одного класса гетерогенных каталитических процессов, *Автоматика и телемеханика*, № 7 (1970).
7. Carlson N. R., Model for the Basic Oxygen Process in Steel Refining, *IEEE Trans. Syst. Sci. Cybernetics*, SCC-2, 41—44 (1966).
8. Corlis R. G., Luus R., Use of Residuals in the Identification and Control of Two-Input, Single-Output Systems, *Industr. Engng. Chem. Fundamentals*, 8, 246—253 (1969).
9. Eykhoff P., Van der Grinten P. M. E. M., Kwakernaak H., Veltman B. P. Th. (см. литературу к гл. 1).
10. Элькинсон Л. З., Математическая модель стеклоплавильной печи относительно управляемых параметров, *Стекло и керамика*, 25 (1969).
11. Gerdin K., Markov Modelling via Correlation Analysis Applied to Identification of a Paper Machine in the Absence of Test Signal, Second IFAC Symp. «Identification and Process Parameter Estimation», Prague, Paper 2.8, 1970.
12. Godfrey K. R., Dynamic Analysis of an Oil-refinery Unit Under Normal Operating Conditions, *Proc. IEE*, 116, 879—888 (1969).
13. Godfrey K. R., Devenish M., An Experimental Investigation of Continuous Gas Chromatography Using Pseudo-Random-Binary Sequences, *Measurement and Control*, 2, 228—233 (1969).
14. Godfrey K. R., The Application of Pseudo-Random Sequences to Industrial Processes and Nuclear Power Plant, Second IFAC Symp. «Identification and Process Parameter Estimation», Prague, Paper 7.1, 1970; см. также The Theory of the Correlation Method of Dynamic Analysis and its Application to Industrial Processes and Nuclear Power Plant, *Measurement and Control*, 2, 165—172 (1969).
15. IEE, Industrial Application of Dynamic Modelling, Conference Held at Durham, England, 16—18 Sept., 1969.
16. Isermann R., Measurement and Calculation of the Control Behaviour of a Super Heater, *Fortschritt-Bericht VDI Z.*, 6, VDI-Verlag, Dusseldorf (1965).
17. Kelly P. C., Harris W. E., Application of Method of Maximum Posterior Probability to Estimation of Gas-Chromatographic Peak Parameters, *Analytical Chemistry*, 43, 1184—1195 (1971).
18. Loosemore M. J., Bray J. W., Optimisation Model for a Lead-Zinc Blast Furnace, I. Ch. E. Symp. Nottingham, London, 1967, pp. 19—31.
19. Manczak K., Identification of an Evaporator Plant as the Exemplification of the Analytical Identification of Complex Technological Units, Paper IFAC Symp. «Identification in Autom. Control Systems», Prague, Paper 2.2, 1967.
20. Murrill P. W., Pike R. W., Smith C. L., *Chem. Engng*, 75, № 23, 117 (1968); № 25, 165 (1968); № 27, 103 (1968); 76, № 2, 167 (1969); № 4, 105 (1969); № 5, 111 (1969); № 7, 151 (1969); № 11, 195 (1969); № 13, 97 (1969); № 18, 125 (1969).

21. Nesbit R. A., Engel R. D., An Example Program for the Determination of Chemical Rate Coefficients from Experimental Data, *Simulation*, 8, 133—137 (1967).
22. Ng K. C., Moss G. C., Continuous Gas Chromatography, Proc. Conference on Laboratory Automation, London, 1970, pp. 197—208.
23. Пайзиев Э., Идентификация сложных объектов управления при помощи частных корреляционных и дисперсионных функций, *Автоматика и телемеханика*, № 10 (1968).
24. Price R. J., Rippin D. W. T., Single Variable On-Line Adaptive Optimisation of a Water Gas Shift Reactor, *Chem. Engng. Sic.*, 23, 593—617 (1968).
25. Rees N. W., Dynamic Identification and Optimum Control Studies on a Pilot Distillation Column, Third IFAC Congress, London, 1966; есть русский перевод: в книге «Теория непрерывных автоматических систем и вопросы идентификации», Труды III Межд. конгресса ИФАК, изд-во «Наука», 1971.
26. Rocco V., Building a Dynamic Model to Simulate in Real Time the Evolution of a Rotating Cement Kiln (на итальянском), *Automaz. e Strum.*, 16, 146—152 (1968).
27. Sawaragi Y., Takamatsu T., Fukwa H., Sayama H., Fundamental Approach to the Identification and Optimization of Chemical Reaction Processes, Proc. IFAC Symp. on Systems Engineering for Control System Design, Tokyo, 1965, pp. 415—423.
28. Seinfeld J. H., Nonlinear Estimation Theory, *Indust. Engng. Chem.*, 62, 32—42 (1970).
29. Stout T. M., Mathematical Models for Computer Control Systems, First IFAC Congress, Moscow, 1961, pp. 991—997; есть русский перевод: Старт Т. М., Математические модели для систем автоматического регулирования, в книге «Статистические методы исследования. Теория структур, моделирование, терминология, образование», Труды I Межд. конгресса ИФАК, Изд-во АН СССР, 1961.
30. Vanderschamp R. E., Perlmutter D. D., Constrained and Weighted Least Squares Procedures for Parameter Estimation, *Ind. and Eng. Chem. Fundamentals*, 10, 150—153 (1971).
31. Wise H. E., Price R. J., Rippin D. W. T., A Comparison of the Use of Pseudo Random Binary Sequences and Frequency Response Methods for the Determination of the Dynamics of a Pilot Scale Water Gas Shift Reactor, Fifth Intern. Congress of AICA, Lausanne, 1967, p. 171.
32. Závorka J., Sutek L., Aguado A., Delgado E., The Dynamics of Multiple-effect Evaporators in the Sugar Industry, IFAC Symp. «Identification in Autom. Control Systems», Prague, Paper 2.2, 1967.

ДОПОЛНИТЕЛЬНАЯ ЛИТЕРАТУРА

Åström K. J., Control Problems in Papermaking, Proc. IBM Scientific Computing Symp. on Control Theory and Applications, Yorktown Heights, N.Y., 1964.

Åström K. J., Computer Control of a Paper Machine — an Application of Linear Stochastic Control Theory, *IBM J. of Res. and Develop.*, **11**, 389—405 (1967).

Буяновский Л. А., Гришунин А. Д., Восстановление динамических характеристик объекта методом стохастической аппроксимации, *Изв. вузов, Нефть и газ*, № 1 (1970).

Coggan G. C., Noton A. R. M., Discrete-Time Sequential State and Parameter Estimation in Chemical Engineering, *Trans. Inst. Chem. Eng.*, **40**, 255—264 (1970).

Damert K., A Standard Program for Determining Kinetic Constants from Differential Equations by Non-Linear Regression (на немецком), *Chem. Techn.*, **21**, 458—459 (1969).

Fitzgerald T. J., Determination of the Average Impulse Response for a Fluidized Bed., *Dissertation Abstracts B*, **27**, Paper 67—7024, 1967.

Hougen J. O., Experiences and Experiments with Process Dynamics, *Chem. Eng. Progress Monograph*, Series 4, 60 (1964).

Ибрагимов И. А., Метт М. С., Исследование многостадийного нефтехимического процесса методом нелинейной корреляции, *Изв. вузов, Нефть и газ*, № 10 (1967).

Kaplick K., Investigation of Process by Regression Analysis (на немецком), *Chem. Techn.*, **19**, 466, 538—542 (1967); **20**, 30—34 (1968).

Клюев Ю. Б., Методика предварительного анализа непрерывных технологических процессов как объектов управления, *Механизация и автоматизация управления*, № 6 (1967).

Michalski B., Mathematical Model and Optimization of Industrial Manufacturing Process of Phthalic Anhydride by Fluidization Method, *Pomiary Automat. Kontrola*, **16**, 30—33 (1970).

Paul R. J. A., Al-Naimi T. T., Das-Gupta D. K., Dynamics of a Gas-Solid Fluidized bed, *Intern. J. Contr.*, **12**, 817—834 (1970).

Pell T. M., Jr., Aris R., Problems in Chemical Reactor Analysis with Stochastic Features, Control of Linearized Distributed Systems on Discrete and Corrupted Observations, *Indust. Eng. Chem. Process Design Develop.*, **9**, 15—20 (1970).

Ray W. H., Aris R., An Adaptive Control of the Batch Reactor. IV. A More Sophisticated Controller, *Automatica*, **4**, 139—161 (1967).

Reid K. N., Lateral Dynamics of an Idealized Moving Web, *Joint Autom. Control Conf.*, 1969, pp. 955—963.

Reid K. N., Lateral Dynamics of a Real Moving Web, *Joint Autom. Control Conf.*, 1969, pp. 964—976.

Sastry V. A., Vetter W. J., Caston R., The Application of Identification Methods to an Industrial Process, *Joint Autom. Control Conf.*, 1969, pp. 787—794.

Schaefer W., Mathematical Problems of Simulation and Optimisation of Processes in Chemical Engineering (на немецком), XII Intern. Sci. Colloquium, Ilmenau, 1967, pp. 7—11.

Vaclavek V., Studies on System Engineering. III. Optimal Choice of the Balance Measurements in Complicated Chemical Engineering Systems, *Chem. Engng. Sci.*, **24**, 947—956 (1969).

ЛИТЕРАТУРА К РАЗД. 14.3

1. Bastl W., Correlation Procedure Applied to Measuring in Nuclear Reactor Systems (на немецком), *Regelungstechnik*, 14, 56—63 (1966).
2. Hendrikson R. A., Murphy G., Cross Spectral Density Measurements in a Coupledcore Reactor, *Nucl. Sci. Eng.*, 31, 215—221 (1968).
3. Liewers P., Buttler E., Measurement of Correlation Functions of Reactor Noise by Means of the Polarity Correlation Method, IFAC Symp. «Identification in Automatic Control Systems», Prague, Paper 3.7, 1967.
4. Roman W. S., Hsu C., Habegger L. J., Parameter Identification in a Nonlinear Reactor System, *IEEE Trans. Nucl. Sci.*, NS-18, 426—429 (1971).
5. Sage A. P., Masters G. W., Identification and Modeling of States and Parameters of Nuclear Reactor Systems, *IEEE Trans. Nucl. Sci.*, NS-14, 279—285 (1966).
6. Schwalm D., Parameter Estimation of Coupled Reactors by Correlation Methods, *Atomkernenergie*, 17, 140—142 (1971).
7. Venerus J. C., Estimation of the Dynamic Reactivity Using Digital Kalman Filtering, *Nucl. Sci. Eng.*, 40, 199—205 (1970).
8. Wilkie J. D. F., Off-Line Statistical Identification of the Kinetics of a Nuclear Reactor (a Comparison of Digital Models), IFAC Symp. «Identification in Autom. Control Systems», Prague, Paper 3.20, 1967.

ДОПОЛНИТЕЛЬНАЯ ЛИТЕРАТУРА

Balcomb J. D., Demuth H. B., Gyftopoulos E. P., A Cross-Correlation Method for Measuring the Impulse Response of Reactor Systems, *Nucl. Sci. Eng.*, 11, 159—166 (1961).

Cohn C. E., Determination of Reactor Kinetic Parameters by Pile Noise Analysis, *Nucl. Sci. Eng.*, 5, 331 (1959).

Dang Van Ba M., Roch A., Control of a Nuclear Reactor: Measurement of Transfer Functions (на французском), *Bull Techn. Suisse Romande*, 93, 335—339 (1967).

Dragt J. B., Accurate Reactor Noise Measurements in a Low Power Critical Reactor, *Nukleonik*, 6, 188 (1966).

Kemeny L. G., Hybrid Simulation of Nuclear Reactor System Stochastic Models, Proc. Intern. Symp. on Analog and Hybrid Computation Applied to Nuclear Energy, Versailles, 1968.

Kovanic P., Identification of Operators in the Reactor Physics (на немецком), *Atomkernenergie*, 12, 404—408 (1967).

Larsen B., The Use of Statistical Methods in the Determination of Reactor Dynamics, *Neue Techn.*, 9, 141—145 (1967).

Robinson E., Correlation Experiments in Nuclear Reactors by Digital Measurement of Counting Rates (на немецком), *Atomkernenergie*, 13, 8—12 (1968).

Sumner A., Optimal Identification of Some Parameters of a Nuclear Reactor Dynamical System, Third IFAC Congress, London, Paper 26 A, 1966.

Walker P. A. W., Ward E. J., Maudsley D., Identification of a Nuclear Reactor Control System by Cross-Correlation, *J. Brit. Nucl. Energy Soc.*, 9, 85—93 (1970).

ЛИТЕРАТУРА К РАЗД. 14.4

1. Cuenod M., Quazza G., Dynamic Statistical Analysis of Electric Power System Controls, *Inst. Engrs. Austral. Elect. Eng. Trans.*, EE-5, 125—135 (1969).
2. Debs A. S., Larson R. E., A Dynamic Estimator for Tracking the State of a Power System, *IEEE Trans. Power Apparatus. and Systems*, PAS-89, 1670—1678 (1970).
3. Debelle J., Foureau A., Vazquez R., Parametric Control of Boilers and Identification of Furnace Perturbation, Third IFAC Congress, London, 1966.
4. Donati F., Fiorio G., Genesio R., Mauro V., Statistical Identification of Interconnected Power Systems, *Ricerche di Automatica*, 1, 30—56 (1970).
5. Eklund K., A Comparison of a Drum Boiler-Turbine Model to Measurements and Models Obtained by Identification, Fifth IFAC Congress, Paris, Paper 5.3, 1972.
6. Gustavsson I. (см. литературу к гл. 11).
7. Handschin E., ed., Real-Time Control of Electric Power Systems, Elsevier, Amsterdam, 1972.
8. Habegger L. J., Bailey R. E., Minimum Variance Estimation of Parameters and States in Nuclear Power System, Fourth IFAC Congress, Warsaw, Paper 12.2, 1969.
9. Koszelnik M., Malkiewicz J., Trybula St., A Method to Determine the Transfer Functions of Power Systems, Fourth IFAC Congress, Warsaw, Paper 33.2, 1969.
10. Scheppe F. C. et al., Power System Static State Estimation. Part I — Exact Model; Part II — Approximate Model; Part III — Implementation, *IEEE Trans. Power Apparatus. and Systems*, PAS-89, 120—135 (1970).
11. Stanton K. N., Estimation of Turboalternator Transfer Functions Using Normal Operating Data, *Proc. IEE*, 112, 1713—1720 (1965).
12. Сулайменова Л. К., Экспериментальное определение динамических характеристик блока котел — турбина статистическим методом, *Автоматика и телемеханика*, № 2 (1965).
13. Thompson F. T., A Dynamic Model of a Drum-Type Boiler System, *IEEE Trans. Power Apparatus. and Systems*, PAS-86, 625—635 (1967).
14. Toyoda J. et al., An Application of State Estimation to Short-Term Load Forecasting, *IEEE Trans. Power Apparatus. and Systems*, PAS-89, 1678 (1970).
15. Uemura K., State Estimation of Large-Scale Electric Power Systems by Decomposition Methods, Fifth IFAC Congress, Paris, Paper 7.5, 1972.

ЛИТЕРАТУРА К РАЗД. 14.5

1. Bode H. W., Variable Equalisers, *Bell System Techn. J.*, 17, 229—244 (1938).
2. Goldman J., Estimation of a Random Timevarying Transfer Function, *IEEE Trans. Inform. Theory*, IT-13, 598—599 (1968).
3. Kanal L., Survey of Automatic Equalization, IEEE Computers and Communic. Conf., Rome, 1969, pp. 23—33.
4. Lückers J., Mulders C. E., An Automatic A. C. Bridge with Phase-Sensitive Detectors, *Revue A*, 7, 88—92 (1965).
5. Lucky R. W., Adaptive Redundancy Removal in Data Transmission, *Bell. Syst. Techn. J.*, 47, 549—573 (1968).
6. Middleton D. (см. литературу к гл. 11).
7. Orth R. W., Frequency-Selective Limiters and Their Application, *IEEE Trans. Electromagn. Compatibility*, EMC-10, 273—283 (1968).
8. Widrow B., Mantey P. E., Griffiths L. J., Goode B. B., Adaptive Antenna Systems, *Proc. IEEE*, 55, 2143—2159 (1967).
9. Lucky R. W., Rudin H. R., Generalized Automatic Equalization for Communication Channels, *Proc. IEEE*, 54, 439—440 (1966).
10. Mohrmann K., Implementation of an Adaptive Equalizer for Fast Data Transmission Involving a Minimum of Complexity, *Nachrichtentech. Z.*, 23, 36—42 (1970).
11. Niessen C. W., Automatic Channel Equalization Algorithm, *Proc. IEEE*, 55, 700—701 (1967).
12. Potter J. B., The Adaptive Equalisation of Communication Systems, Particularly Those Using Waveform Fidelity Criteria, *Proc. Instn. Radio Electronics Engrs. Australia*, 30, 326—336 (1969).
13. Proakis J. G., Miller J. H., An Adaptive Receiver for Digital Signalling through Channels with Intersymbol Interference, *IEEE Trans. Inform. Theory*, IT-15, 484—497 (1969).
14. Proakis J. G., Adaptive Digital Filters for Equalization of Telephone Channels, *IEEE Trans. Audio Electroacoust.*, AU-18, 195—200 (1970).
15. Rudin H. R., Jr., Automatic Equalization Using Transversal Filters, *IEEE Spectrum*, 4, 53—60 (1967).
16. Rudin H. R., Jr., A Continuously Adaptive Equalizer for General-Purpose Communication Channels, *Bell Syst. Techn. J.*, 48, 1865—1884 (1969).
17. Schonfeld T. J., Rapidly Converging Second Order Tracking Algorithms for Adaptive Equalization, Proceed. Intern. Symp. on Information Theory, Noordwijk, Netherlands, 1970,

ЛИТЕРАТУРА К РАЗД. 14.6

1. AGARD, Advanced Control System Concepts, 1970.
2. Dale O. E., Cohen R., Multi-Parameter Identification in Linear Continuous Vibratory Systems, Joint Autom. Control Conf., 1970, pp. 556—564.

3. Huff C. F., Asher R. B., Sebesta H. R., Parameter Sensitivity of Aero-Dynamic Coefficients Determined from Experimental Data, *Journal of Spacecraft and Rockets*, 7, 1265—1267 (1970).
4. Larson D. B., Fleck J. T., Quasilinearization Techniques, Proceedings Nat. Electron. Conf., Chicago, Ill., 1969, pp. 95—100.
5. Mehra R. K., Maximum Likelihood Identification of Aircraft Parameters, Joint Autom. Control Conf., 1970, p. 442—444.
6. Rault A., Pouliquen R., Lefevre C., Applications of Identification Methods to Aeronautics, Second IFAC Symp. «Identification and Process Parameter Estimation», Prague, Paper 11.2, 1970.
7. Smyth R. K., Ehlers H. L., Survey of Adaptive Control Application to Aerospace Vehicles, AGARD Advanced Control System Concepts, 1970, pp. 1—21.
8. Farrel J. L., Newton J. K., Miller J. A., Connely J. J., Optimal Estimation of Rotation-Coupled Flexural Oscillations, Joint Autom. Control Conf., 1969, pp. 433—441.

ДОПОЛНИТЕЛЬНАЯ ЛИТЕРАТУРА

Andeen R. E., Shipley P. P., Digital Adaptive Flight Control Systems for Aerospace Vehicles, *AIAA J.*, 1 (1963).

Buel J. D., Kagiwada H. H., Kalaba R. E. (см. литературу к гл. 7).

Dymock A. J., Helps K. A., Meredith J. F., The Place of Adaptive Control Techniques in the Future Development of Automatic Flight Control Systems, AGARD Advanced Control System Concepts, 1970, pp. 87—100.

Farrel J. L., Newton J. K., Miller J. A., Connely J. J., Optimal Estimation of Rotationcoupled Flexural Oscillations, Joint Autom. Control Conf., 1969, pp. 433—441.

Grusas P. A., Satellite Attitude Determination from Celestial Sightings, *Journal of Spacecraft and Rockets*, 6, 1007—1012 (1969).

Hollstien R. B., Gruber W. H., A Response Regulation Technique for Self-Adaptive Control of Flexible Aircraft, AGARD Advanced Control System Concepts, 1970, pp. 121—138.

Howell G. C., Shanks G. T., Self-Adaptive Manoeuvre Demand Control Systems: Some Practical Design Considerations, AGARD Advanced Control System Concepts, 1970, pp. 149—168.

Kishi F. H., On-Line Computer Control Techniques and their Application to Re-entry Aerospace Vehicle Control, in: *Advances in Control Systems*, Leondes C. T., ed., Academic Press, N.Y., 1964.

Nash R. A., Jr., The Estimation and Control of Terrestrial Inertial Navigation System Errors due to Vertical Deflections, *IEEE Trans. Autom. Control*, AC-13, 329—338 (1968).

Nikiforuk P. N., Pronovost R. J., A Three Variable Self-Adaptive Flight Control System Using Discrete Interval Binary Noise Perturbation Signals, IFAC Symp. «Identification in Autom. Control Systems», Prague, Paper 3.2, 1967.

Taylor L. M., Iliff K. W., Powers B. G. (см. литературу к гл. 7).

ЛИТЕРАТУРА К РАЗД. 14.7 .

1. Bahler A. S., Modeling of Mammalian Skeletal Muscle, *IEEE, Trans. Biomed. Eng.*, 15, 249—257 (1968).
2. Bayliss L. E., Living Control Systems, Freeman, San Francisco, 1966.
3. Bellman R. et al., Estimation of Heart Parameters Using Skin Potential Measurements, *Communic. ACM*, 7 (1964).
4. Beneken J. E. W., A Mathematical Approach to Cardio-Vascular Function, Doctoral Dissertation, Univ. of Utrecht, Netherlands, 1965.
5. Buel J. D., Kalaba (см. литературу к гл. 7).
6. Clymer A. B., Accomplishment in Human Simulation, ASME Paper, 60-AV-26, Aviation Conf., Dallas, 1960.
7. Clymer A. B., Ax A. F., Possibilities for Human Simulation, ASME Paper, 60-AV-37, Aviation Conf., Dallas, 1960.
8. Cosgriff R. L., Briggs G. E., Accomplishments in Human Operator Simulations, ASME Paper, 60-AV-40, Aviation Conf., Dallas, 1960.
9. Costello R. G., Higgins T. J., An Inclusive Classified Bibliography Pertaining to Modeling the Human Operator as an Element in an Automatic Control System, *IEEE Trans. Human Factors Electr.*, HFE-7, 174—181 (1966).
10. Damen A. A. H., Piceni H. A. L., The Multiple Dipole Model of the Ventricular Depolarisation, Report, Group Measurement and Control, EE Dept., Univ. of Technology, Eindhoven, Netherlands, 1971.
11. De Pater L., An Electrical Analogue of the Human Circulatory System, Doctoral Dissertation, Univ. of Groningen, [Netherlands, 1966].
12. Gersch W., Spectral Analysis of e.e.g. by Autoregressive Decomposition of Time Series, *Mathematical Biosciences*, 7, 205—222 (1970).
13. Gottlieb G. L., Agarwal G. C., Stark L., Stretch Receptor Models. I. Single-efferent Single-afferent Innervation, *IEEE Trans. Man-Mach. Syst.*, MMS-10, 17—27 (1969).
14. Mesarovic M. D., ed., Systems Theory and Biology, Springer, 1968.
15. Milsum J. H., Biological Control Systems Analysis, McGraw-Hill, N.Y., 1966.
16. NASA. Annual NASA-University Conference on Manual Control, Scient. and Technical Inform. Division, Office of Technology Utilization, NASA, Washington, DC 20546.
17. Noordergraaf A. et al., Analog of the Arterial Bed, in: Pulsatile Blood Flow, Attinger E. O., ed., McGraw-Hill, N.Y., 1963.
18. O'Neill W. D., Sanathanan C. K., Brodkey J. S., A Minimum Variance, Time Optimal Control System Model of Human Lens Accommodation, *IEEE Trans. on Syst. Science and Cybernetics*, SSC-5, 290—299 (1969).
19. Stassen H. G., The Polarity Coincidence Correlation Technique — a Useful Tool in the Analysis of Human Operator Dynamics, *IEEE Trans. Man-Machine Syst.*, MMS-10, 34—39 (1969).

20. Stark L., Neurological Control Systems Studies in Bioengineering, Plenum Press, N.Y., 1968.
21. Tustin A., The Nature of the Operator's Response in Manual Control and its Implications for Controller Design, *J. IEE*, 94, Part II A, 190—207 (1947).

ДОПОЛНИТЕЛЬНАЯ ЛИТЕРАТУРА

ЧЕЛОВЕК-ОПЕРАТОР

Angel E. S., Bekey G. A., Adaptive Finite-State Models of Manual Control Systems, *IEEE Trans. Man-Machine Syst.*, MMS-9, 15—20 (1968).

Baron S., Kleinman D. W., The Human as an Optimal Control and Information Processor, *IEEE Trans. Man-Machine Syst.*, MMS-10, 9—17 (1969).

Bekey G. E., Discrete Models of the Human Operator in a Control System, Second IFAC Congress, Basle, 1963, pp. 430—434.

Bekey G. A., Meissinger H. F., Rose R. E., Mathematical Models of Human Operators in Simple Two-Axis Manual Control Systems, *IEEE Trans. Human Factors Electron.*, HFE-6, 42—52 (1965).

Bekey G. A., Neal C. B., Identification of Sampling Intervals in Sampled-Data Models of Human Operators, *IEEE Trans. Man-Machine Syst.*, MMS-9, 138—142 (1968).

Blank G. L., Schumacher P. J., Jr., The Human Operator as an Optimal Controller, Proceedings Sixth Allerton Conference on Circuit and System Theory, Monticello, Ill, 1968, pp. 680—689.

Costello R. G., The Surge Model of the Well-Trained Human Operator in Simple Manual Control, *IEEE Trans. Man-Machine Syst.*, MMS-9, 2—9 (1968).

Donaldson P. E. K., Error Decorrelation Studies on a Human Operator Performing a Balancing Task, *Med. Electronics Biol. Engng.* 2, 393—395 (1964).

Elkind J. I., Forgie C. D., Characteristics of the Human Operator in Simple Manual Control Systems, *IRE Trans. Autom. Control*, AC-4, 44—55 (1959).

Elkind J. I., Miller D. C. On the Process of Adaptation by the Human Controller, Third IFAC Congress, London, 1966; есть русский перевод: Элкинд Дж., Миллер Д., О процессах адаптации человека-оператора, в книге «Труды III Межд. конгресса ИФАК», изд-во «Наука», 1971.

Kahn A., Beare A. C., A Critical Re-evaluation of the Human Transfer Function Problem, *IEEE Trans. Aerospace Electronic Syst.*, AES-2, 711—718 (1966).

Kreifeldt J. G., A Sampled-Data Pursuit Tracking Model, *IEEE Trans. Human Factors Electronics*, HFE-6, 65—73 (1965).

Kreil W., Schweizer G., Man Acting as a Controller, A Contribution to the Dynamical Transfer Model (на немецком), *Regelungstechnik*, 16, 49—56 (1968).

Lang R. P., Vacroux A. G., Operator-Centred Adaptive Compensation of Continuous Manual Control Systems, *IEEE Trans. Man-Machine Syst.*, MMS-10, 53—59 (1969).

McRuer D. T., Graham D., Krendel E. S., Reisener W. C., Jr., System Performance and Operator Stationarity in Manual Control Systems, Third IFAC Congress, London, 1966; есть русский перевод: Мак-Руэр Д. Т., Грэхэм Д., Крендел Е. С., Рейснер В. К., Рабочая характеристика системы и стационарность оператора в системах ручного управления, в книге «Дискретные самоадаптирующиеся и обучающиеся системы», Труды III Межд. конгресса ИФАК, изд-во «Наука», 1971.

Naslin P., Raoult J. C., Continuous and Sampled-Data Models of the Human Operator in a Control Loop (на французском), Second IFAC Congress, Basle, 1963, pp. 421—426; есть русский перевод: Наслен П., Рауль Ж., Непрерывные и импульсные модели человека-оператора как звена цепи управления, в книге «Автоматизация процессов управления», Труды II Межд. конгресса ИФАК, изд-во «Наука», 1965.

Ornstein G. N., The Automatic Analog Determination of Human Transfer Function Coefficients, Medical Electronics, *Biological Engineering* (Pergamon Press), 1, 377—387 (1963).

Paskin H. M., A Discrete Stochastic Optimal Control Model of the Human Operator, Proceedings of the National Aerospace Electronics Conference, Dayton, Ohio, 1970.

Phatak A. V., Bekey G. A., Model of the Adaptive Behavior of the Human Operator in Response to a Sudden Change in the Control Situation. *IEEE Trans. Man-Machine Syst.*, MMS-10, 72—80 (1969).|

Preyss A. E., Meiry J. L., Stochastic Modeling of Human Learning Behavior, *IEEE Trans. Man-Machine Syst.*, MMS-9, 36—46 (1968).

Potts T. E., Ornstein G. N., Clymer A. B. (см. литературу к гл. 9). Sheridan T. B. (см. литературу к гл. 10).

Smith H. J., Human Describing Functions Measured in Flight and on Simulators, *IEEE Trans. Human Factors Electronics*, HFE-8, 264—268 (1967).

Snyder D. L., A Nonlinear Study of Compensatory Manual Control Systems, *IEEE Trans. Human Factors Electronics*, HFE-5, 25—28 (1964).

Taylor L. W., Balakrishnan A. V., Identification of Human Response Models in Manual Control Systems, IFAC Symp. «Identification in Autom. Control Systems», Prague, Paper 1.7, 1967.

Цибулевский И. Е., Адаптация человека-оператора к изменениям коэффициента усиления последовательно включенного усилительного звена, *Автоматика и телемеханика*, № 3 (1967).

Wieringa R. D., An Evaluation of a Pilot Model Based on Kalman Filtering and Optimal Control, *IEEE Trans. Man-Machine Syst.*, MMS-10, 108—117 (1969).

Wierwille W. W., Improvement of the Human Operators Tracking Performance by Means of Optimum Filtering and Prediction, *IEEE Trans. Human Factors Electronics*, HFE-5, 20—24 (1964).

Wierwille W. W., A Theory for Optimal Deterministic Characterization of Timevarying Human Operator Dynamics, *IEEE Trans. Human Factors Electronics*, HFE-6, 53—61 (1965).

Young L. R., Green D. M., Elkind J. I., Kelly J. A., Adaptive Dynamic Response Characteristics of the Human Operator in Simple Manual Control, *IEEE Trans. Human Factors Electronics*, HFE-5, 6—13 (1964).

КРОВООБРАЩЕНИЕ

Abel F. L., An Analysis of the Left Ventricle as a Pressure and Flow Generator in the Intact Systemic Circulation, *IEEE Trans. Biomed. Engng.*, BME-13, 182—188 (1966).

Bekey G. A., Merritt M. J., Assali N. S., Synthesis and Optimization of a Model of Cardiovascular Dynamics, Proc. 4th Intern. Analogue Computation Meetings, Brighton, 1964, pp. 383—388.

Burrus C. S., Parks T. W., Watt T. B., Jr., A Digital Parameter-Identification Technique Applied to Biological Signals, *IEEE Trans. Biomed. Engng.*, BME-18, 35—37 (1971).

Ellis M. F., Seaba P. J., Alton F. D., Levy D. M., A Hybrid Computer Simulation of the Mammalian Cardiovascular System, *19th Annual Conf. Engng. Med. and Biol.*, San Francisco, 18, 27, 1966.

Goldwyn R. M., Watt T. B., Jr., Modal Analysis of the Dicrotic Portion of the Human Blood Pressure Curve, *IEEE Intern. Convention Record*, 14, 17—26 (1966).

Goldwyn R. M., Watt T. B., Jr., Arterial Pressure Pulse Contour Analysis via a Mathematical Model for the Clinical Quantification of Human Vascular Properties, *IEEE Trans. Biomed. Engng.*, BME-14, 11—17 (1967).

Granick J. L., An Electrical Analog of the Cardiovascular System and Its Use in Simulating Pathological Cardiology, *19th Annual Conf. Engng. Med. and Biol.*, San Francisco, 8, 130 (1966).

King R. E., Blodgett R. E., A Network Realization of Aortal Haemodynamics, Symp. on Generalized Networks, N.Y., 1966, pp. 599—606.

Morehouse F., Apter J. T., Amiot L., Use of Hybrid Computers to Analyze Behavior of Detailed Models for Biological Systems. I. Model for Aortic Blood Flow, *Comput. Biol. and Med.*, 1, 29—40 (1970).

Pickering W. D., Nikiforuk P. N., Merriman J. F., Analogue Computer Model of the Human Cardiovascular Control System, *Med. Biol. Engng.*, 7, 401—410 (1969).

Pollack G. H., Reddy R. V., Noordergraaf A., Input Impedance, Wave Travel and Reflections in the Human Pulmonary Arterial Tree: Studies Using an Electrical Analog, *IEEE Trans. Biomed. Engng.*, BME-15, 151—164 (1968).

Rideout V. C., Computer Simulation Study of the Pulmonary Circulation, *Simulation*, 12, 239—245 (1969).

Rideout V. C., Sims J. B., Computer Study of the Effects of Small Nonlinearities in the Arterial System, *Mathemat. Biosciences*, 4, 411—426 (1969).

Strano J. J., Welkowitz W., Fich S., Measurement and Utilization of in Vivo Blood-Pressure Transfer Functions of Dog and Chicken Aortas, *IEEE Trans. Biomed. Engng.*, BME-19, 261—271 (1972).

Topham W. S., An Analog Model of the Control of Cardiac Output, *Simulation*, 8, 49—56 (1967).

Verdouw P. D., Noordergraaf A., A Critical Consideration of the Design of Electrical Analogs for the Pressure-Flow Relationship in Arteries, *19th Annual Conf. Engng. Med. and Biol.*, San Francisco, 8, 181 (1966).

Watt T. B., Jr., Goldwyn R. M., Clinical Estimation of Human Vascular Properties via a Mathematical Model Describing the Contour of the Arterial Pressure Pulse, *19th Annual Conf. Engng. Med. and Biol.*, San Francisco, 8, 182 (1966).

Westerhof N., Noordergraaf A., Wave Travel and Input Impedance for Various Types of Networks in an Electrical Model of the Human Systemic Arterial Tree, with and without Elastic Tapering, *19th Annual Conf. Engng. Med. and Biol.*, San Francisco, 8, 90 (1966).

Willard P. W., Beck H. R., A System for Measurement of Cardiovascular Parameters, *Med. Res. Engng.*, 8, 33—36 (1969).

ЛИТЕРАТУРА К РАЗД. 14.8

1. Bergstrom A. R., *The Construction and Use of Economic Models*, English Univ. Press, London, 1967.
2. Fisher F. M., *The Identification in Econometrics*, McGraw-Hill, N.Y., 1966.
3. Forrester J. W., *Industrial Dynamics*, MIT Press, Cambridge; есть русский перевод: Форрестер Дж., *Основы кибернетики предприятия (индустриальная динамика)*, изд-во «Прогресс», 1971.
4. Forrester J. W., *Urban Dynamics*, MIT Press, Cambridge, 1969.
5. Forrester J. W., *World Dynamics*, Wright-Allen Press, Cambridge, 1971.
6. Johnston J., *Econometric Methods*, McGraw-Hill, N.Y., 1963.
7. Koopmans T. C., *Linear Regression Analysis of Economic Time Series*, Haarlem, Netherlands, Bohn, 1937.
8. Manetsch T. J., Transfer Function Representation of the Aggregate Behaviour of a Class Economic Processes, *IEEE Trans. Autom. Control*, AC-11, 693—698 (1966).
9. Naylor T. H., *Computer Simulation Experiments with Models of Economic Systems*, Wiley, N.Y., 1971.
10. Young J. W., Arnold W. F., Brewer J. W., Parameter Identification and Dynamic Models of Socioeconomic Phenomena, *IEEE, Trans. Systems, Man. Cybernet.*, SMC-2, 460—467 (1972).

ЛИТЕРАТУРА К РАЗД. 14.9

Barát J., Muszély Gy, Pattern Recognition Algorithm Adapting with Searching Method, IFAC Symp. «Identification in Autom. Control Systems», Prague, Paper 1.4, 1967.

Biglow J. W., McVey E. S., Moore J. W., Pattern Recognition in a Learning Automatic Control System, IFAC Symp. «Identification in Autom. Control Systems», Prague, Paper 1.5, 1967.

Fu K. S., Sequential Methods in Pattern Recognition and Machine Learning, Academic Press, 1968; есть русский перевод: Фу К., Последовательные методы в распознавании образов и обучении машин, изд-во «Наука», 1971.

Gaines B. R., Techniques of Identification with the Stochastic Computer, IFAC Symp. «Identification in Autom. Control Systems», Prague, Paper 6.1, 1967.

Hawkins J., Self Organizing Systems: a Review and Commentary, *Proc. IRE*, 49, 31—48 (1961).

Ho Y. C., Agrawala A. K., On Pattern Classification Algorithms — Introduction and Survey (Expository Paper), *IEEE Trans. Autom. Control*, AC-13, 676—690; см. также *Proc. IEEE*, 56, 2101—2114 (1968).

Nilson N. J., Learning Machines, McGraw-Hill, N.Y., 1965.
Цыпкин Я. З. (см. литературу к гл. 1).

ПРИЛОЖЕНИЕ А

СПИСОК ОБОЗНАЧЕНИЙ

Приводятся только наиболее важные обозначения. Общие правила (с редкими исключениями) таковы: строчные полужирные буквы означают векторы; прописные полужирные буквы — матрицы; прописными буквами \mathcal{F} , \mathcal{L} и \mathcal{Z} обозначаются преобразования переменных, обозначаемых соответствующими строчными буквами.

A, B, C, D — передаточные функции, динамические операторы;

C — функция штрафа (потерь);

D — оператор дифференцирования;

E — четная функция (функционал) от ошибки, функция штрафа или потерь;

F — функция (функционал), $F [\]$;

F, G, H — передаточные функции, динамические операторы, $F (\)$;

J — критерий;

L — функция правдоподобия;

M — модель;

$N(\mu, \sigma^2)$ — функция нормального распределения со средним μ и дисперсией σ^2 ;

N — мощность шума, например, на выходе объекта или коррелятора;

P — объект, вероятность, $P [\]$;

R — функция Ляпунова;

S — система, мощность сигнала, например, на выходе объекта или коррелятора;

T — время, интервал наблюдения;

A — матрица системы;

B — распределительная матрица;

- C** — матрица измерений (выходная);
D — левая треугольная матрица, матрица вход — выход;
F — матрица частных производных вектор-функции f ;
I — единичная матрица;
J — информационная матрица;
M — матрица;
N — ковариация аддитивного шума;
P — ковариационная матрица оценки;
Q — оценочная матрица;
R — положительно определенная (весовая) матрица, выборочная матрица шума;
T — матрица преобразования (подобия);
U — выборочная матрица сигналов $[u_0, \dots, u_m]$;
W — выборочная матрица белого шума;
Y — выборочная матрица выходных сигналов;

a, a, b, b — коэффициенты дифференциальных и разностных уравнений, передаточной функции объекта, (переменные) параметры объекта;

c, d — коэффициенты (ортогонального) разложения;

e, e — (обобщенная) ошибка;

\mathbf{e}_i — базисные векторы;

f — частота;

f, f, g, g, h, h — функции, весовые функции;

i — ток;

j — мнимая единица;

i, j, k, l, m, n — целые числа;

n, n — аддитивный шум;

p, q — плотность распределения вероятностей;

r — весовой коэффициент;

r — вектор шума;

s — аргумент преобразования Лапласа;

t — время;

u, u — наблюдаемый вход объекта (модели), напряжение;

u, u_j, v, v_i — выход линейных операторов в (обобщенной) модели, функции чувствительности параметров;

v, v, w, w — недоступные измерениям шумы;

w, w — выход модели, весовая функция;

x, x — состояние объекта;

- y, \bar{y} — выход объекта;
 z — оператор сдвига, аргумент z -преобразования;
- $\mathcal{A}, \mathcal{A}, \mathcal{B}, \mathcal{B}, \mathcal{C}, \mathcal{D}$ — оценки величин A, A, B, B, C, D ;
 \mathcal{D} — оператор дисперсии, $\mathcal{D}[\cdot]$;
 \mathcal{E} — оператор математического ожидания, $\mathcal{E}[\cdot]$;
 $\mathcal{F}, \mathcal{F}^{-1}$ — прямое и обратное преобразования Фурье, $\mathcal{F}[\cdot]$;
 $\mathcal{F}^+(\mathcal{F}^-)$ — преобразование Фурье, действующее только на положительные (отрицательные) функции времени;
- $\mathcal{L}, \mathcal{L}^{-1}, \mathcal{L}_2, \mathcal{L}_2^{-1}$ — прямое и обратное одно- и двустороннее преобразования Лапласа, $\mathcal{L}[\cdot]$;
 \mathcal{O} — динамический оператор, $\mathcal{O}[\cdot]$;
 \mathcal{P} — оператор политики, $\mathcal{P}[\cdot]$;
 \mathcal{S} — оператор дискретизации времени, $\mathcal{S}[\cdot]$ — класс систем;
- \mathcal{T} — отображение;
- \mathcal{U} — класс входных сигналов;
- $\mathcal{Z}, \mathcal{Z}^{-1}$ — прямое и обратное z -преобразования, $\mathcal{Z}[\cdot]$;
- Γ — скалярный (матричный) коэффициент усиления;
 Δ — приращение;
 Θ — дисперсионная функция;
 Π — знак произведения;
 Π_i — ортогональные многочлены;
 \sum — знак суммы;
 \bar{Y} — нормированная корреляционная функция;
 Φ — переходная или фундаментальная матрица, $\Phi(t), \Phi(k)$;
 Φ — спектральная плотность, $\Phi(\omega)$;
 Ψ — корреляционная функция, $\Psi(\tau)$;
 Ω — выборочная матрица входа и выхода объекта;
 ∇ — градиент, ∇_{β} — градиент по β ;
- a, α, β, β — оценки a, b , коэффициенты уравнений и передаточной функции модели, настраиваемые параметры модели;
 γ — коэффициент усиления, коэффициент ортогонального разложения;

- δ — дельта-функция Дирака: $\delta(t) = 0$ при $t \neq 0$,

$$\int_{-\infty}^{\infty} f(t) \delta(t - \theta) dt = f(\theta);$$
- δ_{mn} — символ Кронекера;
- ε — малая положительная константа;
- ζ, η — компоненты вектора градиента;
- η — нормированная дисперсионная функция;
- θ — временная задержка, переменная интегрирования;
- Θ — вектор параметров $[\beta', \alpha']'$ или $[\beta', \alpha', \gamma', \delta']'$;
- λ — вектор ошибок;
- λ — собственное значение;
- μ — первый момент случайной величины;
- ν, ζ — переменные интегрирования;
- ξ — состояние модели;
- π — многочлены, отношение длины окружности к ее диаметру;
- ρ — коэффициент корреляции;
- σ — стандартное отклонение;
- σ^2 — дисперсия, $\sigma_x^2 = \mathbb{E}[(x - \mu)^2]$;
- τ — время, постоянная времени;
- $[\varphi$ — характеристическая функция, фаза;
- ψ — эталонный вектор;
- ω — круговая частота;
- ω — вектор $[u', y', w', r']'$.

ВЕРХНИЕ И НИЖНИЕ ИНДЕКСЫ И ДРУГИЕ ОБОЗНАЧЕНИЯ

- $u^{(n)}, \dot{u}$ — n -я и первая производные;
- $u(k)$ — значение в k -й выборочной точке;
- u_i — i -я компонента вектора u ;
- u_i — значение вектора в i -м интервале измерений;
- U_k — матрица наблюдений с номерами $1, \dots, k$;
- $u^{(i)}$ — i -е приближение;
- $u(k+1)$ — новое наблюдение u в момент $k+1$;
- $\beta(i)$ — вектор после i -го цикла настройки;
- $E(i)$ — величина после i -го измерения;
- u^* — дискретный вариант функции $u(t)$;
- G^* — передаточная функция импульсной (дискретной) системы;
- $\overline{G(j\omega)}$ — комплексно-сопряженная величина к $G(j\omega)$;

\bar{x} — среднее значение по счетному ансамблю;
 \bar{x}_N — среднее значение по конечному ансамблю,

$$\bar{x}_N = \frac{1}{N} \sum_{i=1}^N x_i;$$

$\langle x \rangle$ — среднеинтегральное значение функции по бесконечному интервалу:

$$\langle x \rangle = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T x(t) dt;$$

$\langle x \rangle_T$ — среднеинтегральное значение по конечному интервалу:

$$\langle x \rangle_T = \frac{1}{T} \int_0^T x(t) dt;$$

$\langle x, y \rangle$ — некоторая нелинейная процедура обработки сигналов, например $\langle x, y \rangle = \int_0^T x(t) y(t) dt$;

$x * h$ — свертка;

$\angle H(j\omega)$ — фазовый угол $H(j\omega)$;

$| \cdot |$ — абсолютная величина, определитель матрицы;

\equiv — тождественное равенство;

$\| \cdot \|$ — знак нормы;

$\| \cdot \|_{\mathbf{R}}^2$ — положительно определенная форма с весовой матрицей \mathbf{R} ;

$\stackrel{\Delta}{=}$ — равенство по определению;

sgn — знак;

$\text{sgn } e = e / |e|$;

$\text{var } [\cdot]$ — дисперсия;

$\text{cov } [\cdot]$ — ковариация;

$\mathbf{A}^{(-1)}$ — обобщенная обратная (псевдообратная) матрица;

$p(z | x)$ — условная плотность;

$\hat{\beta}$ — оптимальная оценка;

$\tilde{\Psi}$ — приближенное значение (корреляционной функции);

$\text{Re } [\cdot], \text{Im } [\cdot]$ — действительная и мнимая части [];

\mathbf{U}' — знак транспонирования;

$\int_k \dots d^k y$ — k -мерный интеграл.

ПРИЛОЖЕНИЕ Б

ОСНОВНЫЕ ПОНЯТИЯ ИЗ ТЕОРИИ ВЕРОЯТНОСТЕЙ И ТЕОРИИ СЛУЧАЙНЫХ ФУНКЦИЙ

Данная книга рассчитана на читателя, знакомого с основами теории вероятностей и теории случайных функций. Это приложение может служить справочником по наиболее употребительным обозначениям и понятиям. Оно широко используется в книге без дополнительных ссылок.

Множества. Соотношения между событиями определяются правилами теории множеств. Понятия: множество, подмножество, элемент, пустое множество. Операции (фиг. Б.1): \bar{A} — дополнение, $A \cup B$ — объединение, $A \cap B$ — пересечение. Свойства:

Коммутативность	Ассоциативность	Дистрибутивность
$A \cup B = B \cup A$ $A \cap B = B \cap A$	$(A \cup B) \cup C = A \cup (B \cup C)$ $(A \cap B) \cap C = A \cap (B \cap C)$	$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$ $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$

Вероятность. Каждому событию A из множества возможных событий приписывается число $P [A]$, называемое вероятностью.

Аксиома I: $P [A] \geq 0$.

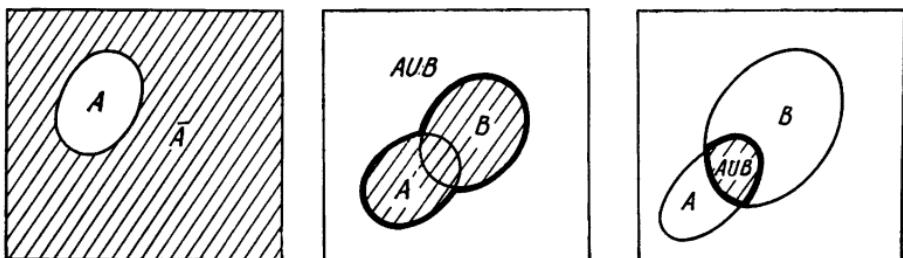
Аксиома II: $P [\Omega] = 1$, если Ω — достоверное событие.

$P [\phi] = 0$, если ϕ — невозможное событие.

Аксиома III: $P [A \cup B] = P [A] + P [B]$, если $A \cap B = \phi$, т. е. если события несовместны.

Аксиома IV: (при счетном числе возможных событий)

$P [A_1 \cup A_2 \cup A_3 \cup \dots] = \sum_{i=1}^{\infty} P [A_i]$, если $A_i \cap A_j = \phi$ для всех $i \neq j$.



Ф и г. Б.1.

Сложные события. На основе двух простых событий A_i и B_j рассматривается сложное событие $[A_i, B_j]$.

$$0 \leq P[A_i, B_j] \leq 1,$$

$$\sum_{i,j} P[A_i, B_j] = 1,$$

$$P[A_i] = \sum_j P[A_i, B_j], \text{ если } B_k \cap B_j = \emptyset \text{ для всех } k \neq j,$$

$$P[B_j] = \sum_i P[A_i, B_j], \text{ если } A_k \cap A_i = \emptyset \text{ для всех } k \neq i.$$

Условная вероятность. Определение:

$$P[A_i | B_j] = \frac{P[A_i, B_j]}{P[B_j]}, \text{ если } P[B_j] \neq 0.$$

Следовательно,

$$P[A_i | B_j] P[B_j] = P[A_i, B_j] = P[B_j | A_i] P[A_i].$$

Для независимых событий

$$P[A_i | B_j] = P[A_i],$$

$$P[A_i, B_j] = P[A_i] P[B_j].$$

Часто событие можно охарактеризовать числом или набором чисел. Действительной случайной величиной¹⁾ X называется величина, которая может принимать ряд значений, причем для каждого действительного числа x определена вероятность $P[X \leq x]$. Случайная величина может быть *дискретной* (если она имеет конечное или счетное множество различных значений) и *непрерывной*²⁾. Для этих величин можно определить следующие понятия.

¹⁾ В этом приложении случайные величины обозначаются прописными буквами в отличие от обозначений, принятых в основном тексте.

²⁾ Здесь под непрерывными величинами понимаются величины с дифференцируемой функцией распределения; существуют также величины, не являющиеся ни непрерывными, ни дискретными.—*Прим. перев.*

	Случайная величина X	
	дискретная	непрерывная
Одномерный случай. Функция распределения (фиг. Б.2).	$F(x) = P[X \leq x]$	$F(x) = P[X \leq x]$
Плотность распределения ¹⁾	$p_i = P[X = x_i]$	$p(x) = \frac{dF(x)}{dx}$
Вероятность	$\sum p_i$	$p(x) \Delta x \approx$ $\approx P[x < X \leq x + \Delta x]$
Свойства	$0 \leq p_i \leq 1$ $\sum_i p_i = 1$ $F(x) = \sum_{x_i \leq x} p_i$	$\int_a^b p(\xi) d\xi = P[a < X \leq b]$ $0 \leq p(x)$ $\int_{-\infty}^{\infty} p(\xi) d\xi = 1$ $F(x) = \int_{-\infty}^x p(\xi) d\xi$ $F(-\infty) = 0$ $F(\infty) = 1$
Примеры распределений вероятностей.	Биномиальное $P[X = x] = \binom{n}{x} p^x (1-p)^{n-x},$ $x = 0, 1, \dots, n$ Пуассоновское $P[X = x] = \frac{\lambda^x e^{-\lambda}}{x!},$ $x = 0, 1, 2, \dots$	Нормальное (фиг. Б.3) $p(x) = \frac{1}{\sigma \sqrt{2\pi}} \times$ $\times \exp \left[-\frac{1}{2} \left(\frac{x-\mu}{\sigma} \right)^2 \right],$ $-\infty \leq x \leq \infty$ Экспоненциальное $p(x) =$ $= \begin{cases} \lambda e^{-\lambda x}, & 0 \leq x \leq \infty \\ 0, & -\infty < x < 0 \end{cases}$

1) В отечественной литературе под плотностью распределения дискретной случайной величины обычно понимается линейная комбинация δ-функций в точках, соответствующих возможным значениям случайной величины с коэффициентами, равными их вероятностям. — Прим. перев.

	Случайная величина X	
	дискретная	непрерывная
Двумерный случай		
Совместная функция распределения	$F(x, y) = P[X \leq x, Y \leq y]$	$F(x, y) = P[X \leq x, Y \leq y]$
Совместная плотность распределения	$p_{ij} = P[X = x_i, Y = y_j]$	$p(x, y) = \frac{\partial^2 F(x, y)}{\partial x \partial y}$
Совместная вероятность	$\sum_{i,j} p_{ij}$	$p(x, y) \Delta x \Delta y \approx P[x < X \leq x + \Delta x, y < Y \leq y + \Delta y]$
Свойства	$0 \leq p_{ij} \leq 1$	$0 \leq p(x, y)$
	$\sum_{i,j} p_{ij} = 1$	$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p(\xi, \eta) d\xi d\eta = 1$
	$F(x, y) = \sum_{\substack{x_i \leq x \\ y_i \leq y}} p_{ij}$	$F(x, y) = \int_{-\infty}^x \int_{-\infty}^y p(\xi, \eta) d\xi d\eta$
Частные функции распределения ¹⁾	$F(x, \infty) = P[X \leq x, Y \leq \infty] = P[X \leq x],$ $F(\infty, y) = P[X \leq \infty, Y \leq y] = P[Y \leq y]$	$F(x, \infty) = P[X \leq x, Y \leq \infty] = P[X \leq x],$ $F(\infty, y) = P[X \leq \infty, Y \leq y] = P[Y \leq y]$
Частные плотности распределения	$p_i = \sum_j p_{ij}$	$p(x) = \frac{dF(x, \infty)}{dx} = \int_{-\infty}^{\infty} p(x, \eta) d\eta,$

1) Здесь и далее часто различные функции обозначаются одной и той же буквой (с разными аргументами).

	Случайная величина X	
	дискретная	непрерывная
Частные плотности распределения	$p_j = \sum_i p_{ij}$	$p(y) = \frac{dF(\infty, y)}{dy} =$ $= \int_{-\infty}^{\infty} p(\xi, y) d\xi$ $F(x y) =$ $= P[X \leq x Y = y]$ $F(y x) =$ $= P[Y \leq y X = x]$
Условная функция распределения		$p(x y) = \frac{p(x, y)}{p(y)}$ $p(y x) = \frac{p(x, y)}{p(x)}$ $p(x y)p(y) = p(x, y) =$ $= p(y x)p(x)$ $F(x, y) = F(x)F(y)$ $p(x, y) = p(x)p(y)$
Условная плотность распределения	$P[X = x_i Y = y_j] = \frac{p_{ij}}{p_j}$ $P[Y = y_j X = x_i] = \frac{p_{ij}}{p_i}$	
Независимость X, Y	$F(x, y) = F(x)F(y)$ $p_{ij} = p_i p_j$	
Пример двумерного распределения		Нормальное (фиг. Б.4) $p(x, y) =$ $= \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \times$ $\times \exp\left[-\frac{1}{2}Q\right]$ $Q = \frac{1}{1-\rho^2} \times$ $\times \left\{ \left(\frac{x-\mu_x}{\sigma_x} \right)^2 - \right.$ $- 2\rho \left(\frac{x-\mu_x}{\sigma_x} \right) \times$ $\times \left(\frac{y-\mu_y}{\sigma_y} \right) +$ $\left. + \left(\frac{y-\mu_y}{\sigma_y} \right)^2 \right\}$

Продолжение табл.

	Случайная величина X	
	дискретная	непрерывная
Многомерный случай Совместная функция распределения	$F(x, y, z, \dots) =$ $= P[X \leq x, Y \leq y,$ $Z \leq z, \dots]$ и т. д. аналогично предыдущему	$F(x, y, z, \dots) =$ $= P[X \leq x, Y \leq y,$ $Z \leq z, \dots]$

Часто работать с распределениями случайной величины оказывается очень сложно. Нередко допустимо (и намного проще) использовать *параметры* распределения вместо самих *функций*. При этом применяются следующие понятия.

	Случайная величина X	
	дискретная	непрерывная
Одномерный случай		
Математическое ожидание (среднее значение)	$E[\dots] = \sum_i \dots p_i$ $E[f(X)] = \sum_i f(x_i) p_i$	$E[\dots] = \int_{-\infty}^{\infty} \dots p(x) dx$ $E[f(X)] =$ $= \int_{-\infty}^{\infty} f(x) p(x) dx$
Свойство: линейность	$E[\alpha X + \beta Y] = \alpha E[X] + \beta E[Y]$	
n -й момент	$E[X^n] = \sum_i x_i^n p_i$	$E[X^n] = \int_{-\infty}^{\infty} x^n p(x) dx$

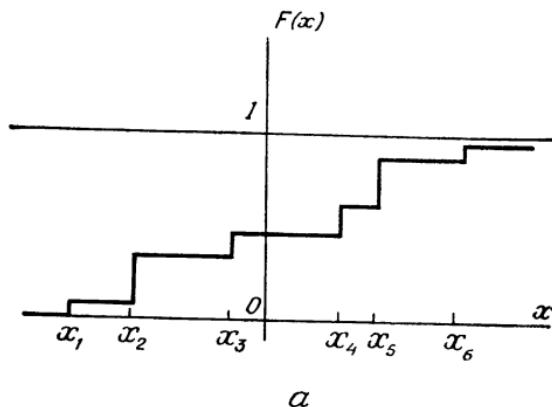
Продолжение табл.

	Случайная величина X	
	дискретная	непрерывная
Первый момент математиче- ское ожидание	$\mu = \mathbb{E}[X]$ $\mu = \sum_i x_i p_i$	$\mu = \int_{-\infty}^{\infty} x p(x) dx$
n -й центральный момент		$\mathbb{E}[(X - \mu)^n]$
Первый централь- ный момент		$\mathbb{E}[(X - \mu)] = 0$
Второй централь- ный момент дисперсия	$\sigma^2 = \mathbb{E}[(X - \mu)^2] = \mathbb{E}[X^2] - \{\mathbb{E}[X]\}^2$ $\sigma^2 = \sum_i x_i^2 p_i - \mu^2$	$\sigma^2 = \int_{-\infty}^{\infty} x^2 p(x) dx - \mu^2$
Стандартное отклонение (разброс)		σ
Двумерный случай Математическое ожидание (среднее значе- ние) _i	$\mathbb{E}[\dots] = \sum_{i, j} \dots p_{ij}$ $\mathbb{E}[f(X, Y)] = \sum_{i, j} f(x_i, y_j) p_{ij}$ $\mu_x = \mathbb{E}[X]$ $\mu_x = \sum_{i, j} x_i p_{ij}$	$\mathbb{E}[\dots] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \dots p(x, y) dx dy$ $\mathbb{E}[f(X, Y)] = \int \int \int f(x, y) p(x, y) dx dy$ $\mu_x = \int \int x p(x, y) dx dy$

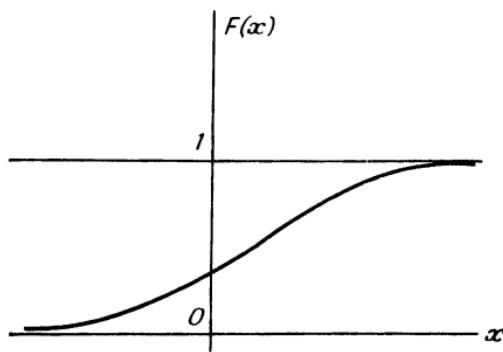
Продолжение табл.

	Случайная величина X	
	дискретная	непрерывная
Дисперсия	$\sigma_x^2 = \mathbb{E} [(X - \mu_x)^2]$ $\sigma_x^2 = \sum_{i,j} x_i^2 p_{ij} - \mu_x^2$	$\sigma_x^2 =$ $= \int \int x^2 p(x, y) dx dy - \mu_x^2$
Ковариация	$\sigma_{xy}^2 = \text{cov}[XY] = \mathbb{E} [(X - \mu_x)(Y - \mu_y)] =$ $= \mathbb{E}[XY] - \mu_x \mu_y$	
Коэффициент корреляции	$\rho_{xy} = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$	
Условное математическое ожидание (условное среднее)	$\mathbb{E}[X Y = y_j] =$ $= \frac{\sum_i x_i p_{ij}}{p_j}$	$\mathbb{E}[X Y] =$ $= \frac{\int x p(x, y) dx}{p(y)}$
Свойство Независимые X, Y	$\mathbb{E}[\mathbb{E}[X Y]] = \mathbb{E}[X]$ $\mathbb{E}[XY] = \mathbb{E}[X] \mathbb{E}[Y]$ $\rho_{xy} = 0$ $\sum_{i,j} x_i y_j p_{ij} =$ $= \left\{ \sum_i x_i p_i \right\} \left\{ \sum_j y_j p_j \right\} =$ $= \mu_x \mu_y$	$\int \int x y p(x, y) dx dy =$ $= \left\{ \int x p(x) dx \right\} \times$ $\times \left\{ \int y p(y) dy \right\} = \mu_x \mu_y$

Случайный процесс — семейство случайных величин $X(t)$, зависящих от параметра t , принимающего значения из множества T . *Реализация (выборочная функция)* случайного процесса получается, если каждому t из T приписывается определенное значение $X(t)$. В данной книге t — время, T — временной интервал наблюдений. В этом случае можно рассматривать реализацию как случайный сигнал, а случайный процесс — как множество случайных сигналов. В принципе это множество сигналов должно реализовываться с помощью большого числа физических систем с (макроскопически) идентичной структурой, находящейся в различных состояниях.



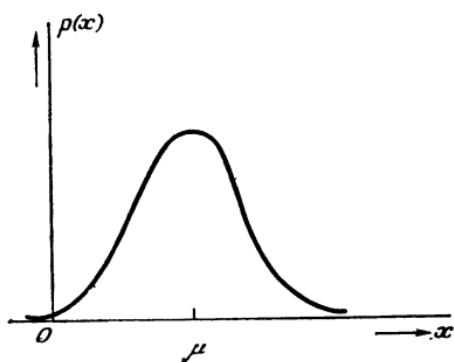
α



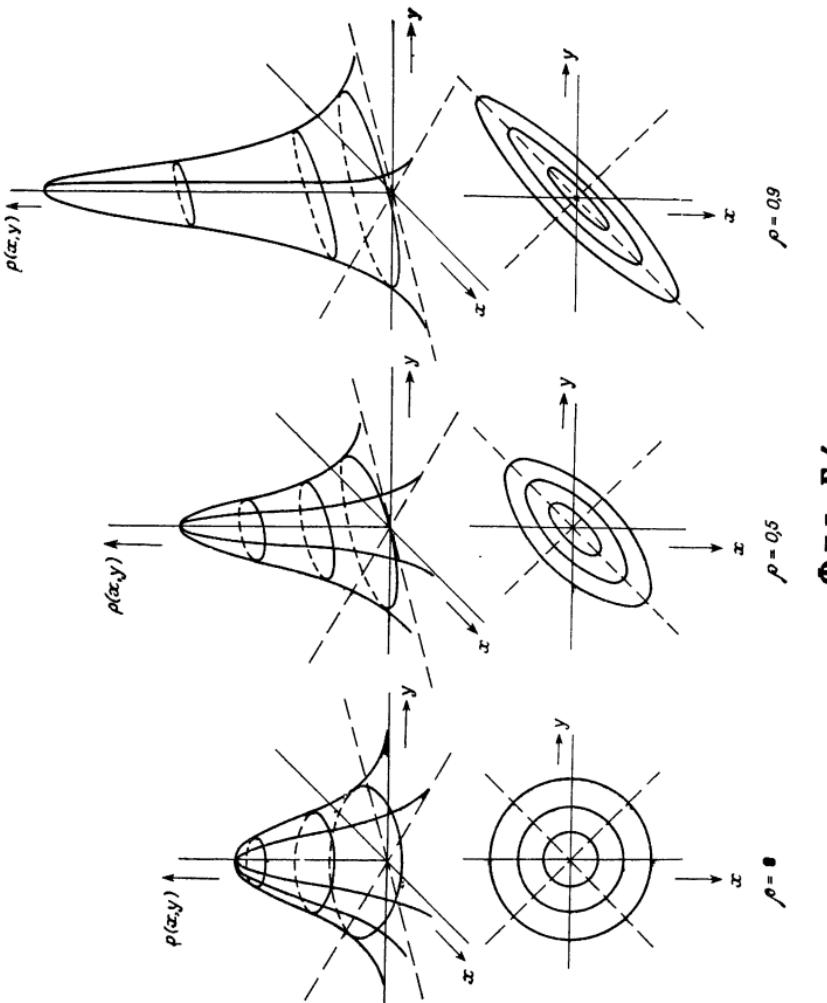
б

Ф и г. Б.2.

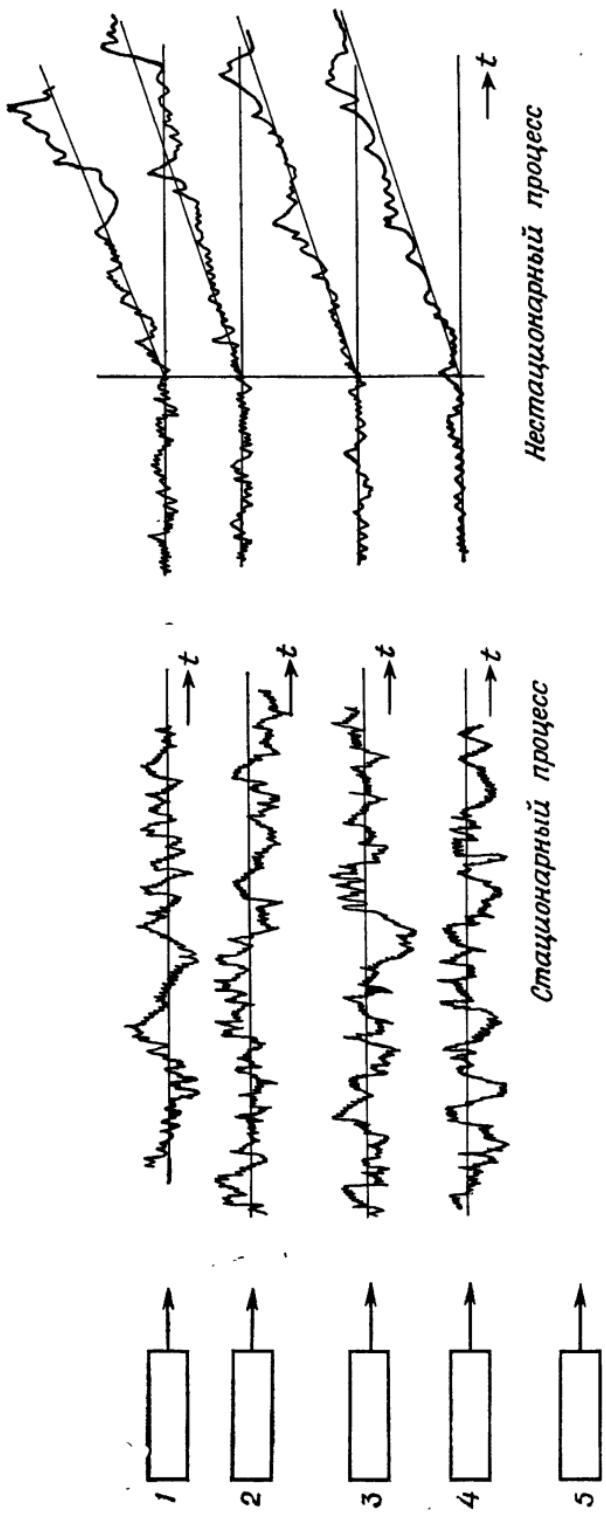
а – дискретные случайные величины; *б* – непрерывные случайные величины.



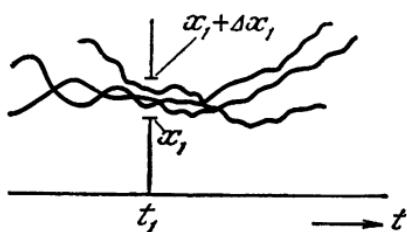
Ф и г. Б.3.



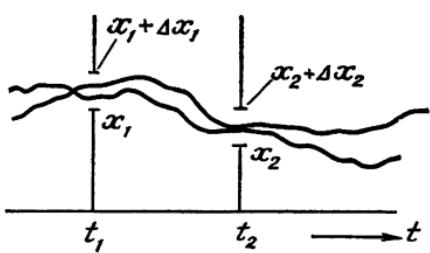
Ф и г. Б.4.



Ф и г. Б.5.



Фиг. Б.6.



Фиг. Б.7.

дящихся при идентичных условиях (фиг. Б.5). Если каждая система выдает случайный сигнал, то можно считать, что статистические свойства всех этих сигналов одинаковы. При этих условиях можно определить вероятностные характеристики амплитуды $X(t_1)$ в момент t_1 :

$$F_1(x_1, t_1) = P[X(t_1) \leq x_1],$$

$$p_1(x_1, t_1) = \frac{dF_1(x_1, t_1)}{dx_1},$$

$$p_1(x_1, t_1) \Delta x_1 \approx P[x_1 < X(t_1) \leq x_1 + \Delta x_1]$$

(если функция F_1 дифференцируема). На фиг. Б.6

показаны некоторые сигналы (элементы множества), траектории которых проходят в момент t_1 через интервал $(x_1, x_1 + \Delta x_1]$.

Более детальное описание случайного процесса достигается при помощи изучения статистического ансамбля значений процесса, соответствующих *двум* моментам времени (фиг. Б.7):

$$58 \quad F_2(x_1, t_1; x_2, t_2) = P[X(t_1) \leq x_1; X(t_2) \leq x_2],$$

$$p_2(x_1, t_1; x_2, t_2) = \frac{\partial^2}{\partial x_1 \partial x_2} F_2(\dots),$$

$$p_2(x_1, t_1; x_2, t_2) \Delta x_1 \Delta x_2 \approx$$

$$\approx P[x_1 < X(t_1) \leq x_1 + \Delta x_1; x_2 < X(t_2) \leq x_2 + \Delta x_2].$$

Аналогично можно найти плотности вероятности p_3, p_4, \dots , характеризующие распределения вероятностей амплитуд для трех, четырех и т. д. моментов времени и дающие еще более детальное описание случайного процесса. Как и для случайных величин, при описании случайных процессов можно ограничиться некоторыми параметрами вместо плотностей распределения. В качестве таких

параметров обычно используются *моменты по множеству* (т. е. полученные усреднением по статистическому ансамблю), например

$$\mu_x(t_1) = \overline{X(t_1)}^{\text{MH}} = \mathbb{E}[X(t_1)] = \int_{-\infty}^{\infty} x_1 p(x_1, t_1) dx_1,$$

$$\overline{X^2(t_1)}^{\text{MH}} = \mathbb{E}[X^2(t_1)] = \int_{-\infty}^{\infty} x_1^2 p(x_1, t_1) dx_1,$$

$$\begin{aligned} \Psi_{xx}(t_1, t_2) &= \overline{X(t_1) X(t_2)}^{\text{MH}} = \mathbb{E}[X(t_1) X(t_2)] = \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_1 x_2 p(x_1, t_1; x_2, t_2) dx_1 dx_2. \end{aligned}$$

Последнее выражение определяет *корреляционную функцию*¹⁾, имеющую важное значение для описания свойств случайного процесса. Аналогично связь между двумя случайными процессами X и Y можно охарактеризовать *взаимной корреляционной функцией* $\Psi_{xy}(t_1, t_2)$.

Эти определения справедливы для произвольных случайных процессов, в том числе и нестационарных. Чрезвычайно важный подкласс образуют *стационарные* процессы, т. е. такие, для которых все статистические свойства инвариантны к сдвигу по оси времени (фиг. Б.5):

$$\begin{aligned} F_n(x_1, t_1 + \theta; \dots; x_n, t_n + \theta) &= \\ &= F_n(x_1, t_1; \dots; x_n, t_n) \end{aligned}$$

при всех θ . Отсюда, в частности, следует, что все параметры стационарных процессов не меняются при сдвиге во времени:

$$\mu_x(t_1) = \mu_x(t) = \mu_x,$$

$$\Psi_{xx}(t_1, t_2) = \Psi_{xx}(t, t + \tau) = \Psi_{xx}(\tau),$$

где $\tau = t_2 - t_1$.

¹⁾ В отечественной литературе корреляционной функцией принято называть функцию $\Psi_{xx}(t_1, t_2) = \mu_x(t_1) \mu_x(t_2) = \mathbb{E}[(X(t_1) - \mathbb{E}[X(t_1)]) (X(t_2) - \mathbb{E}[X(t_2)])]$. — Прим. перев.

Для таких процессов имеют смысл понятия усреднения по времени и моментов по времени. Если соответствующие моменты по множеству и по времени совпадают,

$$\lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T X(t) dt = \overline{X(t)}^{\text{sp}} = \overline{X(t)}^{\text{MH}} = \int_{-\infty}^{\infty} x p(x) dx,$$

$$\lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T X^2(t) dt = \overline{X^2(t)}^{\text{sp}} = \overline{X^2(t)}^{\text{MH}} = \int_{-\infty}^{\infty} x^2 p(x) dx,$$

$$\begin{aligned} \Psi_{xy}(\tau) &= \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T X(t) Y(t + \tau) dt = \overline{X(t) Y(t + \tau)}^{\text{sp}} = \\ &= \overline{X(t) Y(t + \tau)}^{\text{MH}} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x y p(x, y, \tau) dx dy \quad \text{и т. д.,} \end{aligned}$$

то процесс называют *эргоидическим*.

Некоторые свойства корреляционных функций стационарных процессов:

$\Psi_{xx}(\tau)$ существует, если существует $E[X^2(t)]$.

$\Psi_{xx}(0)$ (средняя энергия) положительна, если

$E[X^2(t)] \neq 0$.

$\Psi_{xx}(\tau) = \Psi_{xx}(-\tau)$

$|\Psi_{xx}(\tau)| \leq \Psi_{xx}(0)$.

Если $X(t)$ содержит периодическую составляющую, то $\Psi_{xx}(\tau)$ также ее содержит.

Если $X(t)$ не содержит периодической или постоянной составляющей, то $\Psi_{xx}(\tau) \rightarrow 0$ при $\tau \rightarrow \infty$.

Для описания случайных процессов в частотной области используются *спектральные и взаимные спектральные плотности* (энергетические спектры) $\Phi_{xx}(j\omega)$ и $\Phi_{xy}(j\omega)$, которые посредством преобразования Фурье или двустороннего преобразования Лапласа выражаются через соответствующие корреляционные функции (уравнения Вине-

ра — Хинчина):

$$\Phi_{xy}(j\omega) = \mathcal{F}[\Psi_{xy}(\tau)] = \int_{-\infty}^{\infty} \Psi_{xy}(\tau) e^{-j\omega\tau} d\tau,$$

$$\Psi_{xy}(\tau) = \mathcal{F}^{-1}[\Phi_{xy}(j\omega)] = \frac{1}{2\pi} \int_{-\infty}^{\infty} \Phi_{xy}(j\omega) e^{j\omega\tau} d\omega,$$

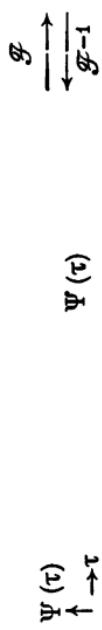
$$\Phi_{xy}(s) = \mathcal{L}_2[\Psi_{xy}(\tau)] = \int_{-\infty}^{\infty} \Psi_{xy}(\tau) e^{-s\tau} d\tau, \quad \alpha < \operatorname{Re}[s] < \beta,$$

$$\Psi_{xy}(\tau) = \mathcal{L}_2^{-1}[\Phi_{xy}(s)] = \frac{1}{2\pi j} \int_{c-j\infty}^{c+j\infty} \Phi_{xy}(s) e^{s\tau} ds, \quad \alpha < c < \beta.$$

На фиг. Б.8 показаны некоторые из таких пар функций. Обратите внимание на использование j в аргументе \mathcal{F} -преобразования, чем обеспечивается соответствие $j\omega \leftrightarrow s$ между \mathcal{F} - и \mathcal{L}_2 -преобразованиями. Следует отметить еще одну особенность преобразования \mathcal{L}_2 . При определении корреляционной функции можно выбирать $c = 0$, т. е. интегрировать по оси $j\omega$, которая делит комплексную s -плоскость на две части — левую полуплоскость $\operatorname{Re}[s] < 0$, соответствующую $\tau > 0$, и правую полуплоскость $\operatorname{Re}[s] > 0$, соответствующую $\tau < 0$ (фиг. Б.9).

В качестве применения методов описания сигналов в частотной и временных областях рассмотрим фиг. Б.10. В эргодическом случае нужно определить взаимную корреляционную функцию и взаимную спектральную плотность

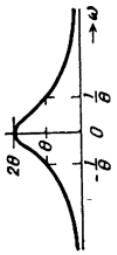
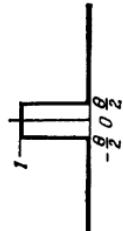
$$\begin{aligned} \Psi_{yz}(\tau) &= \overline{Y(t) Z(t + \tau)} = \\ &= \left\{ \int_{-\infty}^{\infty} g(\theta) X(t - \theta) d\theta \right\} \left\{ \int_{-\infty}^{\infty} h(\xi) X(t + \tau - \xi) d\xi \right\} = \\ &= \int_{-\infty}^{\infty} d\theta g(\theta) \int_{-\infty}^{\infty} d\xi h(\xi) \overline{X(t - \theta) X(t + \tau - \xi)} = \\ &= \int_{-\infty}^{\infty} d\theta g(\theta) \int_{-\infty}^{\infty} d\xi h(\xi) \Psi_{xx}(\tau - \xi + \theta). \end{aligned}$$



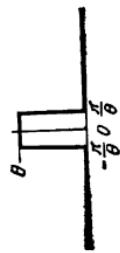
$$\Psi(\tau) = e^{-|\tau|/\theta}$$

$$\Psi(\tau) = \frac{\sin \pi \frac{\tau}{\theta}}{\pi \frac{\tau}{\theta}}$$

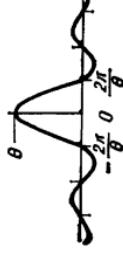
$$\Psi(\tau) = \begin{cases} 1, |\tau| < \frac{\theta}{2} \\ 0, |\tau| > \frac{\theta}{2} \end{cases}$$



$$\Phi(j\omega) = \frac{2\theta}{1 + (\omega\theta)^2}$$

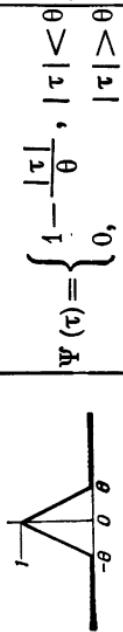


$$\Phi(j\omega) = \begin{cases} \theta, |\omega| < \frac{\pi}{\theta} \\ 0, |\omega| > \frac{\pi}{\theta} \end{cases}$$

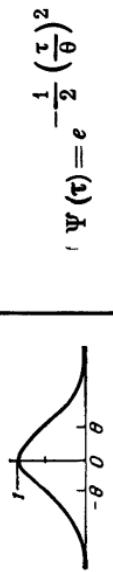


$$\Phi(j\omega) = \theta \frac{\sin \frac{\omega\theta}{2}}{\frac{\omega\theta}{2}}$$

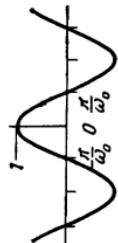
Ф и г. Б.8.



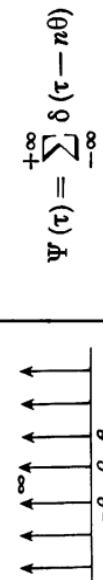
$$\Psi(\tau) = \begin{cases} 1 - \frac{|\tau|}{\theta}, & |\tau| < \theta \\ 0, & |\tau| > \theta \end{cases} \Leftrightarrow \Phi(j\omega) = \theta \left(\frac{\omega\theta}{2} \right)^2$$



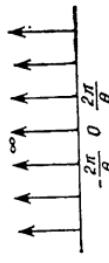
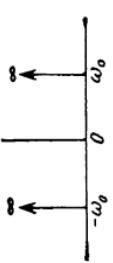
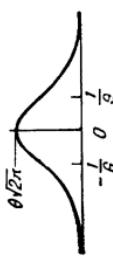
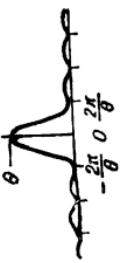
$$\Psi(\tau) = e^{-\frac{1}{2}} \left(\frac{\pi}{\theta}\right)^2 \tau^2 \Leftrightarrow \Phi(j\omega) = \theta \sqrt{2\pi} e^{-\frac{1}{2}(\theta\omega)^2}$$

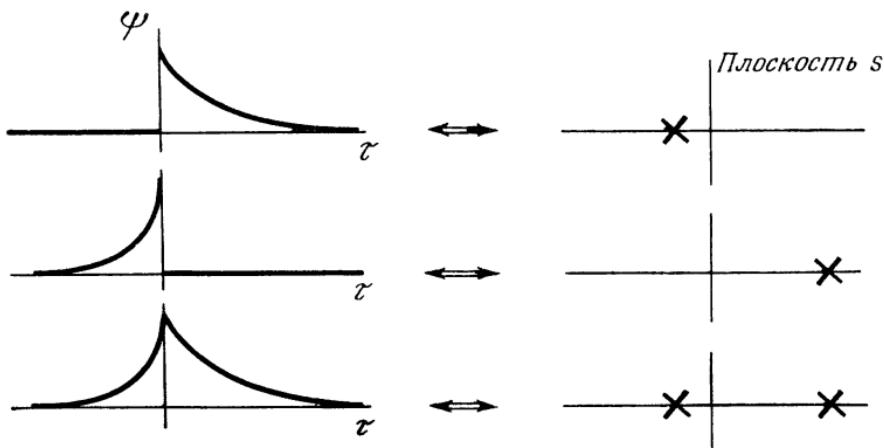


$$\Psi(\tau) = \cos \omega_0 \tau \Leftrightarrow \Phi(j\omega) = \pi \{ \delta(\omega + \omega_0) + \delta(\omega - \omega_0) \}$$



$$\Psi(\tau) = \sum_{n=-\infty}^{+\infty} \delta(\tau - n\theta) \Leftrightarrow \Phi(j\omega) = \sum_{n=-\infty}^{+\infty} \delta\left(\omega - \frac{2\pi}{\theta}\right)$$





Ф и г. Б.9.

Полученное выражение представляет собой двойную свертку. Применяя двустороннее преобразование Лапласа, получаем взаимную спектральную плотность

$$\begin{aligned} \Phi_{yz}(s) &= \mathcal{L}_2 [\Psi_{yz}(\tau)] = \int_{-\infty}^{\infty} \Psi_{yz}(\tau) e^{-s\tau} d\tau = \\ &= \int_{-\infty}^{\infty} d\theta g_i(\theta) e^{s\theta} \int_{-\infty}^{\infty} d\xi h_i(\xi) e^{-s\xi} \int_{-\infty}^{\infty} \Psi_i(\tau - \xi + \theta) e^{-s(\tau - \xi + \theta)} d\tau = \\ &\quad = G(-s) H_i(s) \Phi_{xx}(s), \end{aligned}$$

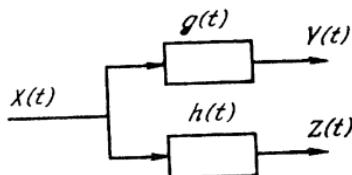
где

$$G(s) = \mathcal{L}_2 [g(t)],$$

$$H(s) = \mathcal{L}_2 [h(t)],$$

$$\Phi_{xx}(s) = \mathcal{L}_2 [\Psi_{xx}(\tau)],$$

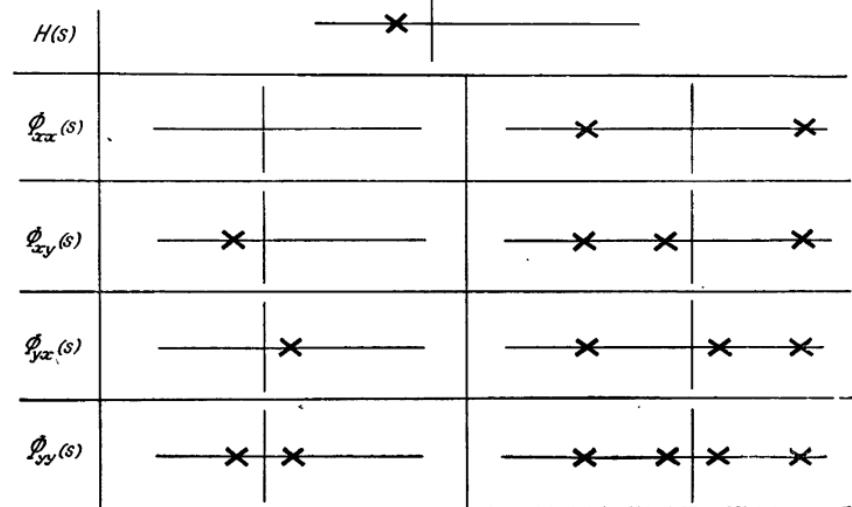
причем $g(t) = h(t) = 0$ при $t < 0$ (условия физической реализуемости). Эти преобразования осуществляются по



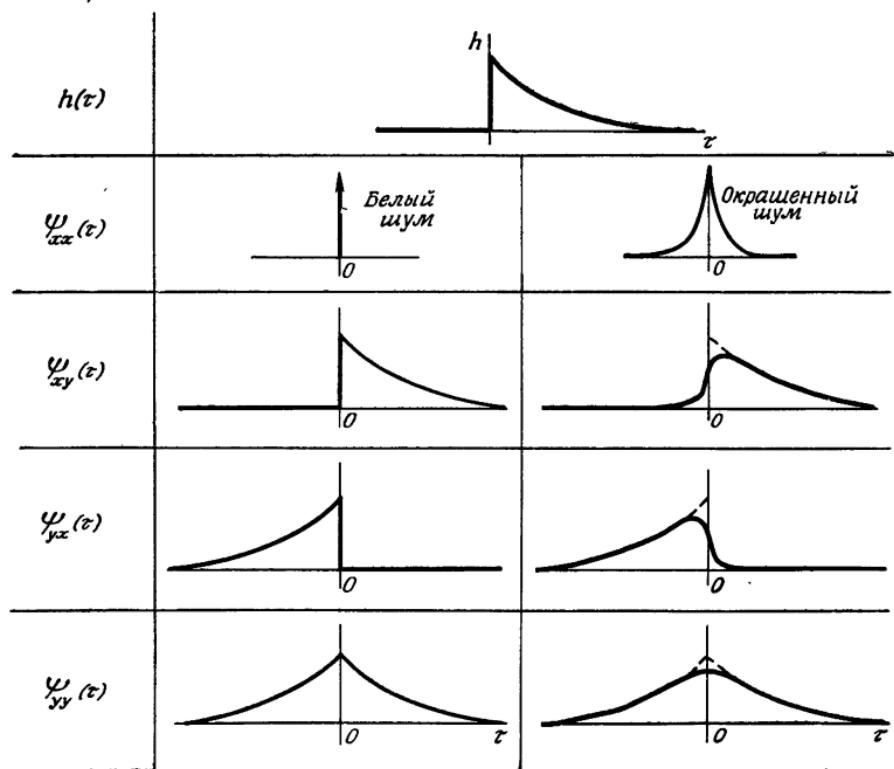
Ф и г. Б.10.

Частотная область

Плоскость s



Временная область



Ф и г. Б.11.

схеме

$$\Phi_{yz}(s) = G(-s) H(s) \Phi_{xx}(s)$$

Передаточная функция с аргументом $-s$ воздействует на первый индекс спектральной плотности, передаточная функция с аргументом $+s$ — на второй. На фиг. Б.11 эта процедура иллюстрируется для двух весьма простых примеров. Аналогичное соотношение с заменой $G(s) \rightarrow G(j\omega)$ и $G(-s) \rightarrow \bar{G}(j\omega)$ справедливо и при использовании преобразования Фурье.

ПРИЛОЖЕНИЕ В

ОСНОВНЫЕ ПОНЯТИЯ ТЕОРИИ МАТРИЦ

Обозначения.

Вектор ($k \times 1$ -матрица) $m \times n$ -матрица

$$\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_k \end{bmatrix}, \quad \mathbf{A} = [a_{ij}] = \begin{bmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \dots & a_{mn} \end{bmatrix}.$$

Термины: размерность, элементы, строка, столбец, алгебраическое дополнение, след, диагональная и единичная матрицы.

Матричная алгебра.

Транспонирование:

$$\begin{aligned} \mathbf{x}' &= [x_1, \dots, x_k], \\ \mathbf{A}' &= [a_{ji}], \\ [\mathbf{A}']' &= \mathbf{A}, \quad [\mathbf{A} + \mathbf{B}]' = \mathbf{A}' + \mathbf{B}', \\ [\mathbf{ABC}]' &= \mathbf{C}' \mathbf{B}' \mathbf{A}'. \end{aligned}$$

Квадратная матрица \mathbf{A} :

симметричная, если $\mathbf{A}' = \mathbf{A}$;

кососимметрическая, если $\mathbf{A}' = -\mathbf{A}$;

ортогональная, если $\mathbf{A}'\mathbf{A} = \mathbf{A}\mathbf{A}' = \mathbf{I}$ (\mathbf{I} — единичная матрица);

идемпотентная, если $\mathbf{A}'\mathbf{A} = \mathbf{A}$.

Определитель:

если \mathbf{A} — $n \times n$ -матрица, то $|\mathbf{A}| = \sum_{j=1}^n a_{ij} A_{ij}$, A_{ij} — алгебраическое дополнение элемента a_{ij} ;

матрица вырожденная, если $|A| = 0$;
 $|AB| = |A| \cdot |B|$.

Ранг:

наивысший порядок невырожденной квадратной подматрицы;

число линейно независимых строк (столбцов);
 ранг $C = \text{ранг } ACB$, если A и B невырожденные,
 C и ACB называются эквивалентными.

Присоединенная матрица

$$\tilde{A} = [A_{ij}]'$$

$$\tilde{A}\tilde{A} = A\tilde{A} = |A| \cdot I.$$

Сложение матриц ассоциативно и коммутативно.

Умножение:

ассоциативно $A[BC] = [AB]C$;

дистрибутивно $A[B + C] = AB + AC$;

некоммутативно, за исключением случая $CA = AC$,
 $C = cI$;

умножение на скаляр $cAB = AcB = ABc$;

скалярное умножение $x'y \rightarrow$ скаляр;

двучленное умножение $xy' \rightarrow$ матрица.

При матричном умножении

произведение, последний множитель которого — вектор-столбец, является вектор-столбцом ($ABx = y$);

произведение, первый множитель которого — вектор-строка, является вектор-строкой ($y'BC = d'$);

произведение, начинающееся с вектор-строки и заканчивающееся вектор-столбцом, представляет скаляр ($x'Ay = c$).

Квадратичные формы:

$$[x_1, x_2] \begin{bmatrix} a & b \\ b & c \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = ax_1^2 + 2bx_1x_2 + cx_2^2.$$

$x'A_1x$ — квадратичная форма

$\begin{cases} > 0 & \text{положительно} \\ < 0 & \text{отрицательно} \\ \geq 0 & \text{неотрицательно} \\ \leq 0 & \text{неположительно} \end{cases}$	$\begin{cases} \text{определенная} \\ \text{определенная} \\ \text{определенная} \\ \text{определенная} \end{cases}$
--	--

$x'A_2y$ — билинейная форма

A_1 всегда можно преобразовать к симметричному виду. Если Q — положительно определенная, то $[x - x_0]' Q^{-1} [x - x_0] = c$ — эллипсоид с центром x_0 и полуосами, квадраты которых равны корням характеристического уравнения Q .

Вырожденность:

матрица A вырожденная, если существует $x \neq 0$, такое, что $Ax = 0$ или $A'x = 0$.

Обращение:

$$A^{-1}A = I = AA^{-1},$$

$$A^{-1} = |A|^{-1}\tilde{A},$$

$$[A^{-1}]' = [A']^{-1},$$

$$[A^{-1}]^{-1} = A,$$

$$[ABC]^{-1} = C^{-1}B^{-1}A^{-1}.$$

Если $A' = A^{-1}$, то A ортогональная;

если $A' = A$, $C' = C$, то ¹⁾

$$[A + BCB']^{-1} = A^{-1} - A^{-1}B[C^{-1} + B'A^{-1}B]^{-1}B'A^{-1} = \\ = [I + A^{-1}BCB']^{-1}A^{-1},$$

$$CB'[A + BCB']^{-1} = [C^{-1} + B'A^{-1}B]^{-1}B'A^{-1}$$

или, транспонируя, имеем

$$[A + BCB']^{-1} BC = A^{-1}B[C^{-1} + B'A^{-1}B]^{-1}.$$

Производя здесь замены типа $C \rightarrow -C$, $A \rightarrow -A$, $A \rightarrow A^{-1}$, $C \rightarrow C^{-1}$, $B \rightarrow B'$, можно получить еще группу формул.

Собственные значения:

$A - \lambda I$ — характеристическая матрица (A — квадратная);

$|A - \lambda I| = f(\lambda)$ — характеристическая функция;

$|A - \lambda I| = f(\lambda) = 0$ — характеристическое уравнение;

¹⁾ Доказывается умножением обеих частей равенства на $[A + BCB']$:

$$I = [A + BCB'][A^{-1} - A^{-1}B[C^{-1} + B'A^{-1}B]^{-1}B'A^{-1}] = \\ = I - B[C^{-1} + B'A^{-1}B]^{-1}B'A^{-1} + BCB'A^{-1} - \\ - BCB'A^{-1}B[C^{-1} + B'A^{-1}B]^{-1}B'A^{-1} = \\ = I + B[C - [C^{-1} + B'A^{-1}B]^{-1}[I + CB'A^{-1}B]]B'A^{-1} = \\ = I + BC[I - I]B'A^{-1} = I.$$

$\lambda_1, \dots, \lambda_n$ — корни характеристического уравнения (собственные числа или значения);

$[A - \lambda_i I] x_i = 0 \Rightarrow x_1, \dots, x_n$ — характеристические (собственные) векторы;

$$r_i = x_i / |x_i|;$$

$$R = [r_1, r_2, \dots, r_n].$$

Если A — симметричная вещественная матрица, то существует хотя бы одна матрица R , такая, что $R'AR = D$, $A = RDR'$, где D — диагональная матрица; если A — симметричная вещественная матрица, то все λ_i вещественные;

если при этом все $\lambda_i > 0$, то A положительно определенная, а при $\lambda_i \geq 0$ — неотрицательно определенная.

Преобразование подобия:

A и B — подобные матрицы, если $B = T^{-1}AT$ ($TT^{-1} = T^{-1}T = I$);

$$A = TBT^{-1};$$

$$A^i = TB^iT^{-1}.$$

Конгруэнтное преобразование:

A и B — конгруэнтные матрицы, если $B = T'AT$ (A и B симметричные). Таким способом B можно преобразовать к диагональному виду с элементами $-1, 0, 1$, причем число единиц инвариантно и равно рангу.

Ортогональное преобразование:

$$B = T^{-1}AT,$$

$$T^{-1} = T'$$

(T — ортогональная матрица).

Симметричная идемпотентная матрица:

$$A^2 = A,$$

ее характеристический вектор x_i :

$$A^2x_i = Ax_i = \lambda_i x_i, A^2x_i = A\lambda_i x_i = \lambda_i^2 x_i, \lambda_i = 0 \text{ или } 1.$$

Матрица

$$[I - A]^2 = I^2 - 2A + A^2 = I - A \text{ также идемпотентна.}$$

Обобщенное обращение (псевдообращение):

обозначение $A^{(-1)}$;

определение $A^{(-1)}$: $AA^{(-1)}A = A$;

если A квадратная невырожденная, то $A^{(-1)} = A^{-1}$;

если $A^{(-1)}A = I$ ($AA^{(-1)} = I$), то $A^{(-1)}$ — левая (правая) обратная.

Матричный анализ.

Производная по скалярному аргументу:

$$\frac{d\mathbf{A}}{dt} = \left[\frac{da_{ij}}{dt} \right], \quad \frac{d\mathbf{x}}{dt} = \begin{bmatrix} \frac{dx_1}{dt} \\ \vdots \\ \frac{dx_k}{dt} \end{bmatrix},$$

$$\frac{d\mathbf{x}'}{dt} = \left[\frac{dx_1}{dt}, \dots, \frac{dx_k}{dt} \right],$$

$$\frac{d}{dt} (\mathbf{A} + \mathbf{B}) = \frac{d\mathbf{A}}{dt} + \frac{d\mathbf{B}}{dt}, \quad \frac{d}{dt} (\mathbf{ABC}) = \frac{d\mathbf{A}}{dt} \mathbf{BC} + \\ + \mathbf{A} \frac{d\mathbf{B}}{dt} \mathbf{C} + \mathbf{AB} \frac{d\mathbf{C}}{dt},$$

$$\frac{d}{dt} \mathbf{A}^n = \frac{d\mathbf{A}}{dt} \mathbf{A}^{n-1} + \mathbf{A} \frac{d\mathbf{A}}{dt} \mathbf{A}^{n-2} + \dots + \mathbf{A}^{n-1} \frac{d\mathbf{A}}{dt},$$

$$\frac{d}{dt} \mathbf{A}^{-1} = -\mathbf{A}^{-1} \frac{d\mathbf{A}}{dt} \mathbf{A}^{-1}, \quad \frac{d}{dt} |\mathbf{A}| = \text{Sp} \left[\frac{d\mathbf{A}}{dt} \tilde{\mathbf{A}} \right].$$

Производная по векторному аргументу:

$$\frac{\partial}{\partial \mathbf{x}} = \nabla_{\mathbf{x}} \stackrel{\Delta}{=} \begin{bmatrix} \frac{\partial}{\partial x_1} \\ \vdots \\ \frac{\partial}{\partial x_k} \end{bmatrix} \quad (\text{оператор градиента}),$$

$$\frac{\partial}{\partial \mathbf{x}'} = \nabla_{\mathbf{x}'} \stackrel{\Delta}{=} \left[\frac{\partial}{\partial x_1}, \dots, \frac{\partial}{\partial x_k} \right].$$

Скалярная функция c

$$\nabla_{\mathbf{x}} c = \frac{\partial c}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial c}{\partial x_1} \\ \vdots \\ \frac{\partial c}{\partial x_k} \end{bmatrix}, \quad \nabla_{\mathbf{x}\mathbf{x}'} c = \nabla_{\mathbf{x}} \nabla_{\mathbf{x}'} c = \\ = \begin{bmatrix} \frac{\partial^2 c}{\partial x_1 \partial x_1} & \cdots & \frac{\partial^2 c}{\partial x_1 \partial x_k} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 c}{\partial x_k \partial x_1} & \cdots & \frac{\partial^2 c}{\partial x_k \partial x_k} \end{bmatrix};$$

Вектор-функция \mathbf{y}

$$\nabla_{\mathbf{x}} \mathbf{y}' = \begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \cdots & \frac{\partial y_m}{\partial x_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial y_1}{\partial x_k} & \cdots & \frac{\partial y_m}{\partial x_k} \end{bmatrix}, \quad \nabla_{\mathbf{x}'} \mathbf{y} = \begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \cdots & \frac{\partial y_1}{\partial x_k} \\ \vdots & \ddots & \vdots \\ \frac{\partial y_m}{\partial x_1} & \cdots & \frac{\partial y_m}{\partial x_k} \end{bmatrix},$$

$$\nabla_{\mathbf{x}} \mathbf{x} = \mathbf{I}, \quad \nabla_{\mathbf{x}'} \mathbf{x}' = \mathbf{I},$$

$$\nabla_{\mathbf{x}} [\mathbf{y}' \mathbf{x}] = \mathbf{y} = \nabla_{\mathbf{x}'} [\mathbf{x}' \mathbf{y}], \quad \nabla_{\mathbf{x}} [\mathbf{x}' \mathbf{x}] = 2\mathbf{x},$$

$$\nabla_{\mathbf{x}} [\mathbf{x}' \mathbf{A} \mathbf{x}] = \mathbf{A} \mathbf{x} + \mathbf{A}' \mathbf{x}, \quad \nabla_{\mathbf{x}'} [\mathbf{x}' \mathbf{A}] = \mathbf{A}.$$

Градиент скалярного произведения

$$\nabla_{\mathbf{x}} [\mathbf{y}' \mathbf{z}] = [\nabla_{\mathbf{x}} \mathbf{y}] \mathbf{z} + [\nabla_{\mathbf{x}} \mathbf{z}'] \mathbf{y}.$$

Градиент квадратичной формы с симметричной, неотрицательно определенной, не зависящей от аргумента \mathbf{x} матрицей \mathbf{A}

$$\nabla_{\mathbf{x}} [\mathbf{y}' \mathbf{A} \mathbf{y}] = \nabla_{\mathbf{x}} [[\mathbf{y}' (\mathbf{A}')^{1/2}] [\mathbf{A}^{1/2} \mathbf{y}]] =$$

$$= 2 [\nabla_{\mathbf{x}} [\mathbf{y}' (\mathbf{A}')^{1/2}]] \mathbf{A}^{1/2} \mathbf{y} = 2 [\nabla_{\mathbf{x}} \mathbf{y}'] \mathbf{A} \mathbf{y},$$

где

$$\mathbf{A} = \mathbf{R}' \mathbf{D} \mathbf{R} = [\mathbf{R}' \mathbf{D}^{1/2}] [\mathbf{D}^{1/2} \mathbf{R}], \quad \mathbf{A}^{1/2} = \mathbf{D}^{1/2} \mathbf{R},$$

$|\nabla_{\mathbf{x}} \mathbf{y}|$ — якобиан (функциональный определитель).

Стационарные точки:

скалярная функция c

$$\frac{\partial c}{\partial \mathbf{x}} \Big|_{\mathbf{x}=\hat{\mathbf{x}}} = 0, \text{ минимум, если}$$

$\frac{\partial}{\partial \mathbf{x}'} \frac{\partial c}{\partial \mathbf{x}} \Big|_{\mathbf{x}=\hat{\mathbf{x}}}$ положительно определена;

скalarный квадрат

$$\mathbf{e}' \mathbf{e} (\mathbf{e} = \mathbf{z} - \mathbf{U} \boldsymbol{\beta}),$$

$$\frac{\partial \mathbf{e}' \mathbf{e}}{\partial \boldsymbol{\beta}} = -2 \mathbf{U}' \mathbf{e} = -2 \mathbf{U}' [\mathbf{z} - \mathbf{U} \boldsymbol{\beta}],$$

$$\frac{\partial}{\partial \boldsymbol{\beta}} \frac{\partial \mathbf{e}' \mathbf{e}}{\partial \boldsymbol{\beta}} = 2 \mathbf{U}' \mathbf{U}.$$

Матричные ряды и экспонента:

$$f(A) \stackrel{\Delta}{=} \sum_{i=0}^{\infty} \alpha_i A^i,$$

$$e^{At} \stackrel{\Delta}{=} I + At + \frac{[At]^2}{2!} + \frac{[At]^3}{3!} + \dots$$

Теорема Кэли — Гамильтона: если A — $n \times n$ -матрица, то любой полином или сходящийся степенной ряд от A может быть записан в виде линейной комбинации A^{n-1}, \dots, A, I . Например, если A — квадратная матрица порядка 3, то

$$e^{At} = a_0 I + a_1 A + a_2 A^2.$$

Неравенство Шварца (схема доказательства):

$$[Ax + By]' [Ax + By] \geqslant 0,$$

$$\underbrace{x' A' Ax}_C + \underbrace{x' A' By}_D + \underbrace{y' B' Ax}_D + \underbrace{y' B' By}_E \geqslant 0,$$

$$\underbrace{[C^{1/2}x + C^{-1/2}Dy]' [C^{1/2}x + C^{-1/2}Dy]}_{= 0, \text{ если выбрать } x = -C^{-1}Dy} + y' [E - D'C^{-1}D] y \geqslant 0$$

Следовательно,

$$y' [E - D'C^{-1}D] y \geqslant 0 \quad \text{при всех } y, \text{ т. е.}$$

$$E - D'C^{-1}D \geqslant 0,$$

$$B'B \geqslant [B'A][A'A]^{-1}[A'B].$$

ЛИТЕРАТУРА

Bellman R., Introduction to Matrix Analysis, McGraw-Hill, N.Y., 1960; есть русский перевод: Беллман Р., Введение в теорию матриц, изд-во «Наука», 1969.

Bodewig E., Matrix Calculus, Amsterdam, North-Holland Publ. Comp., 1956.

Derusso P. M., Roy R. J., Close C. M., State Variables for Engineers, 1965; есть русский перевод: Деруссо П., Рой Р., Клоуз Ч., Пространство состояний в теории управления, изд-во «Наука», 1970.

Frame J. S., Matrix Functions and Applications, IEEE Spectrum, 1, Part I: Matrix Operations and Generalized Inverses, March, 1964, pp. 208—220.

- Part II: Functions of a Matrix, April, 1964, pp. 102—108.
Part III: Application of Matrices to Systems Analysis, May, 1964,
pp. 100—109.
Part IV: Matrix Functions and Constituent Matrices, June, 1964,
pp. 123—131.
Part V: Similarity Reductions by Rational or Orthogonal Matrices,
July, 1964, pp. 103—109.
Гантмахер Ф. Р., Теория матриц, изд-во «Наука», 1967.
Householder A. S., The Theory of Matrices in Numerical Analysis, Blaisdell Publ. Comp., N.Y., 1964.
Zurmühl R., Matrizen (на немецком), Springer-Verlag, Berlin,
1964.

ОБОБЩЕННОЕ ОБРАЩЕНИЕ МАТРИЦ

- Albert A., Regression and the Moore-Penrose Pseudoinverse, Academic Press, N.Y., 1972.
Penrose R., On Best Approximate Solutions of Linear Matrix Equations, *Proc. Cambridge Phil. Soc.*, 52, 17—19 (1956).
Penrose R., A Generalized Inverse for Matrices, *Proc. Cambridge Phil. Soc.*, 51, 406—413 (1955).
Manherz R. K., Jordan B. W., Hakimi S. L., Analog Methods for Computation of the Generalized Inverse, *IEEE Trans. Autom. Control*, AC-13, 582—585 (1968).

ПРИЛОЖЕНИЕ Г

ПЕРЕХОДНЫЕ ПРОЦЕССЫ СТАТИСТИЧЕСКОГО СРЕДНЕГО И ДИСПЕРСИИ ПРИ КОРРЕЛЯЦИОННЫХ ИЗМЕРЕНИЯХ

Произведение двух случайных сигналов. Рассматриваемая система показана на фиг. Г.1. Эта система является даже несколько более общей, чем необходимо для наших целей. Стационарные случайные сигналы x и y (элементы эргодического статистического ансамбля) перемножаются между собой. Потенциометр «модулирует» произведение функцией времени $f(t)$, которая, например, может быть ступенчатой функцией, обусловленной переключениями некоторого ключа. (Если одна или обе функции времени x и y «модулированы» по мощности, функция $f(t)$ может описывать эту модуляцию.) Модулированное произведение подается на блок G с весовой функцией $g(t)$. Предполагая, что $f(t) = 0$ при $t < 0$ и принимая нулевые начальные условия, получаем

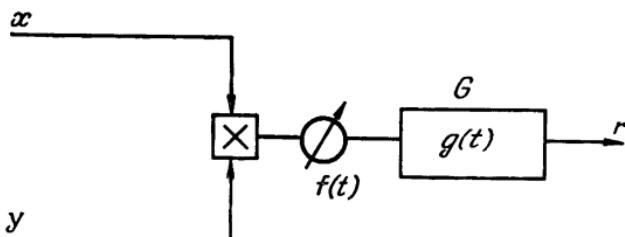
$$r(t) = \int_0^t f(t-\theta) x(t-\theta) y(t-\theta) g(\theta) d\theta. \quad (\Gamma.1)$$

Если $f(t)$ — функция единичного скачка в момент $t = 0$, то

$$r(t) = \int_0^t x(t-\theta) y(t-\theta) g(\theta) d\theta. \quad (\Gamma.2)$$

Если блок G — интегратор, то $g(t)$ также является ступенчатой функцией. Изменяя порядок интегрирования, находим

$$r(t) = \int_0^t x(\tau) y(\tau) d\tau. \quad (\Gamma.3)$$



Ф и г. Г.1.

Следовательно, оба случая — низкочастотный фильтр и интегратор можно описать одной и той же формулой (Г.2).

В этой постановке необходимо найти выражения для среднего значения и дисперсии выходного сигнала r как функции времени.

а) Среднее значение. Поскольку оба сигнала предполагаются стационарными, среднее значение r равно

$$\begin{aligned} \mu_r(t) &= E[r(t)] = E\left[\int_0^t f(t-\theta) x(t-\theta) y(t-\theta) g(\theta) d\theta\right] = \\ &= \int_0^t f(t-\theta) E[x(t-\theta) y(t-\theta)] g(\theta) d\theta = \\ &= \Psi_{xy}(0) \int_0^t f(t-\theta) g(\theta) d\theta. \quad (\Gamma.4) \end{aligned}$$

Если функция $f(t)$ ступенчатая,

$$f(t) = \begin{cases} 1 & \text{при } t \geq 0, \\ 0 & \text{при } t < 0, \end{cases}$$

то

$$\mu_r(t) = \Psi_{xy}(0) \int_0^t g(\theta) d\theta. \quad (\Gamma.5)$$

Если используется интегратор

$$g(t) = \begin{cases} 1 & \text{при } t \geq 0, \\ 0 & \text{при } t < 0, \end{cases}$$

то

$$\mu_r(t) = \Psi_{xy}(0)t. \quad (\Gamma.6)$$

б) Дисперсия.

Дисперсия r равна

$$\sigma_r^2(t) = E[\{r(t) - \mu_r(t)\}^2] = E[r^2(t)] - \mu_r^2(t). \quad (\Gamma.7)$$

Это можно переписать в виде

$$\begin{aligned} \sigma_r^2(t) &= E \left[\int_0^t f(t-\theta) x(t-\theta) y(t-\theta) g(\theta) d\theta \times \right. \\ &\quad \times \left. \int_0^t f(t-\xi) x(t-\xi) y(t-\xi) g(\xi) d\xi \right] - \\ &\quad - \left\{ \Psi_{xy}(0) \int_0^t f(t-\theta) g(\theta) d\theta \right\}^2 = \\ &= \int_0^t \int_0^t f(t-\theta) f(t-\xi) E[x(t-\theta) y(t-\theta) x(t-\xi) y(t-\xi)] \times \\ &\quad \times g(\theta) g(\xi) d\theta d\xi - \left\{ \Psi_{xy}(0) \int_0^t f(t-\theta) g(\theta) d\theta \right\}^2. \quad (\Gamma.8) \end{aligned}$$

Необходимо найти четвертый момент. Введем моментную производящую функцию

$$\begin{aligned} P(s) &= E[\exp s'x] = \\ &= \int \int \int \int \exp \{s_1 x_1 + s_2 x_2 + s_3 x_3 + s_4 x_4\} p(x) dx_1 dx_2 dx_3 dx_4, \quad (\Gamma.9) \end{aligned}$$

где

$$s = \begin{bmatrix} s_1 \\ s_2 \\ s_3 \\ s_4 \end{bmatrix}, \quad x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix},$$

$p(x)$ — четырехмерная плотность вероятности рассматриваемого случайного процесса. Интересующий нас момент определяется смешанной производной четвертого порядка

от $P(s)$:

$$\frac{\partial^4 P(s)}{\partial s_1 \partial s_2 \partial s_3 \partial s_4} \Big|_{s=0} = \mathcal{E}[x_1 x_2 x_3 x_4]. \quad (\Gamma.10)$$

Предположим, что рассматриваемые случайные величины имеют нулевое математическое ожидание и гауссовское совместное распределение вероятностей

$$p(x) = \frac{1}{(2\pi)^2 |M|^{1/2}} \exp \left\{ -\frac{1}{2} x' M^{-1} x \right\},$$

где

$$M = \begin{bmatrix} \Psi_{11} & \dots & \Psi_{1n} \\ \vdots & & \vdots \\ \vdots & & \vdots \\ \Psi_{n1} & \dots & \Psi_{nn} \end{bmatrix} \text{ и } \Psi_{ij} = \mathcal{E}[x_i x_j] = \Psi_{ji}.$$

Тогда нетрудно показать, что производящая функция равна

$$P(s) = \frac{1}{(2\pi)^2 |M|^{1/2}} \int \int \int \int \exp \left\{ -\frac{1}{2} x' M^{-1} x + \frac{1}{2} x' s + \frac{1}{2} s' x \right\} dx_1 dx_2 dx_3 dx_4.$$

Выделяя полный квадрат

$$\begin{aligned} -\frac{1}{2} \{x' M^{-1} x - x' s - s' x\} &= \\ &= -\frac{1}{2} \{(x' M^{-1/2} - s' M^{1/2}) (M^{-1/2} x - M^{1/2} s) - s' M s\} \end{aligned}$$

и учитывая свойства интегралов от плотностей распределения, получаем, что

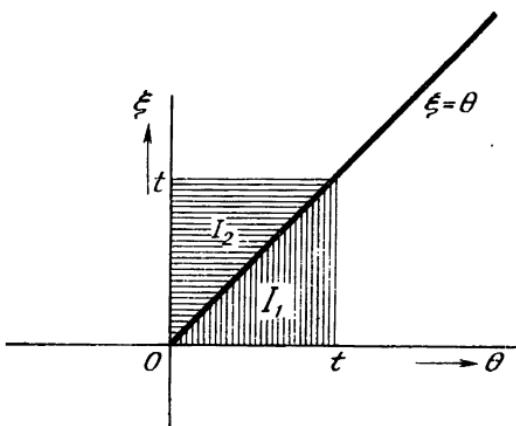
$$P(s) = \exp \left\{ \frac{1}{2} s' M s \right\}, \quad (\Gamma.11)$$

откуда в соответствии с (Г.10) находим

$$\mathcal{E}[x_1 x_2 x_3 x_4] = \Psi_{12} \Psi_{34} + \Psi_{13} \Psi_{24} + \Psi_{14} \Psi_{23}. \quad (\Gamma.12)$$

В рассматриваемом случае

$$\begin{aligned} \mathcal{E}[x(t-\theta) y(t-\theta) x(t-\xi) y(t-\xi)] &= \\ &= \Psi_{xy}^2(0) + \Psi_{xx}(\theta-\xi) \Psi_{yy}(\theta-\xi) + \\ &+ \Psi_{xy}(\theta-\xi) \Psi_{yx}(\theta-\xi) = \Psi_{xy}^2(0) + R(\theta-\xi) \end{aligned} \quad (\Gamma.13)$$



Ф и г. Г.2.

(обозначение $R(\theta - \xi)$ введено для удобства). Поскольку

$$\Psi_{xx}(-v) = \Psi_{xx}(v), \quad \Psi_{yy}(-v) = \Psi_{yy}(v),$$

$$\Psi_{xy}(-v) = \Psi_{yx}(v),$$

то $R(-v) = R(v)$. Первый член в (Г.13) $\Psi_{xy}^2(0)$ приводит к выражению

$$\begin{aligned} & \int_0^t \int_0^t f(t-\theta) f(t-\xi) \Psi_{xy}^2(0) g(\theta) g(\xi) d\theta d\xi = \\ & = \left\{ \Psi_{xy}(0) \int_0^t f(t-\theta) g(\theta) d\theta \right\}^2, \quad (\text{Г.14}) \end{aligned}$$

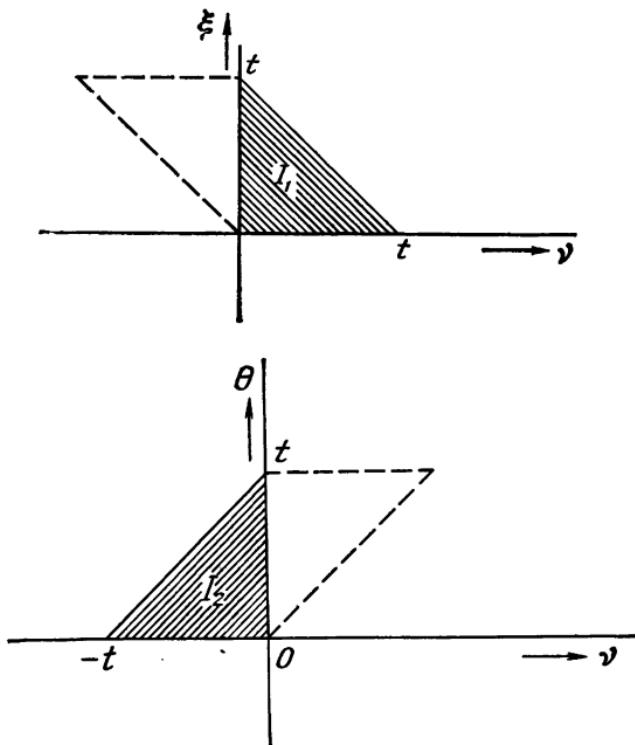
тогда как второй член порождает соответствующую составляющую в соотношении (Г.8). Следовательно, подставляя (Г.13) в (Г.8), получаем

$$\sigma_r^2(t) = \int_0^t \int_0^t f(t-\theta) f(t-\xi) R(\theta - \xi) g(\theta) g(\xi) d\theta d\xi. \quad (\text{Г.15})$$

Полагая $\theta - \xi = v$ и деля область интегрирования (фиг. Г.2) на две части (фиг. Г.3), имеем

$$\begin{aligned} & \text{для } I_1: 0 \leq \xi \leq \theta \leq t \Rightarrow 0 \leq \xi \leq (\xi + v) \leq t, \\ & \text{для } I_2: 0 \leq \theta \leq \xi \leq t \Rightarrow 0 \leq \theta \leq (\theta - v) \leq t. \end{aligned}$$

| якобиан | = 1,



Ф и г. Г.3.

Таким образом, дисперсию как функцию времени можно представить в виде суммы двух интегралов:

$$\begin{aligned} \sigma_r^2(t) = & \int_0^t dv R(v) \int_0^{t-v} f(t-\xi-v) f(t-\xi) g(\xi+v) g(\xi) d\xi + \\ & + \int_0^t dv R(v) \int_0^{t+v} f(t-\theta) f(t-\theta+v) g(\theta) g(\theta-v) d\theta. \end{aligned}$$

Поскольку $R(-v) = R(v)$, оба интеграла равны

и

$$\sigma_r^2(t) = 2 \int_0^t dv R(v) \int_0^{t-v} f(t-v-\xi) f(t-\xi) g(v+\xi) g(\xi) d\xi, \quad 0 \leq v \leq t, \quad (\Gamma.16)$$

где

$$R(v) = \Psi_{xx}(v)\Psi_{yy}(v) + \Psi_{xy}(v)\Psi_{yx}(v).$$

Если $f(t)$ — ступенчатая функция,

$$f(t) = \begin{cases} 1 & \text{при } t \geq 0, \\ 0 & \text{при } t < 0, \end{cases}$$

то

$$\sigma_r^2(t) = 2 \int_0^t dv R(v) \int_0^{t-v} g(v+\xi) g(\xi) d\xi. \quad (\Gamma.17)$$

Если используется интегратор с весовой функцией

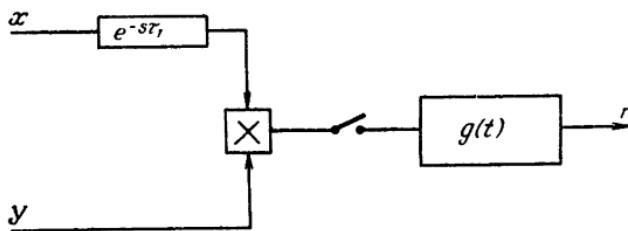
$$g(t) = \begin{cases} 1 & \text{при } t \geq 0, \\ 0 & \text{при } t < 0, \end{cases}$$

то

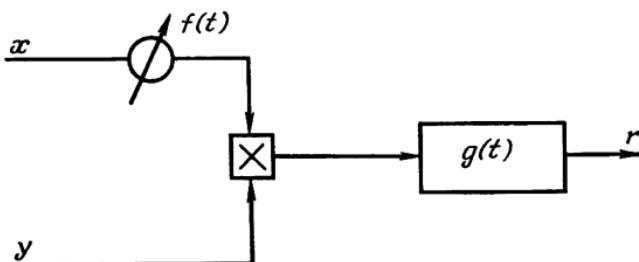
$$\sigma_r^2(t) = 2 \int_0^t (t-v) R(v) dv. \quad (\Gamma.18)$$

в) Обсуждение результатов. Во-первых, нужно отметить, что эти результаты справедливы для стационарных случайных процессов, для которых можно определить $\Psi_{xy}(0)$ и четвертый смешанный момент $E[x(t-\theta)y(t-\theta)x(t-\xi)y(t-\xi)]$. В качестве примера были взяты гауссовские процессы.

Во-вторых, следует отметить, что хотя было принято, что временная задержка τ равна нулю, на самом деле полученные результаты сохраняют силу и при отказе от этого предположения. Если нужно определить $\mu_r(t)$ и $\sigma_r^2(t)$ при другом значении аргумента, например $\tau = \tau_1$, это можно сделать введением фиктивного элемента задержки в блок-схему формирования сигнала (фиг. Г.4).



Ф и г. Г.4.

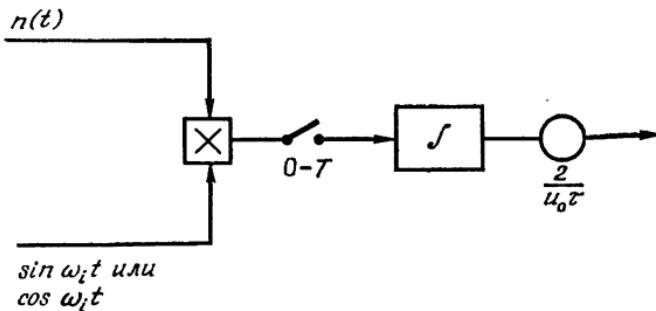


Ф и г. Г.5.

Наконец, можно показать, что выведенные формулы могут быть использованы также и для изучения переходных процессов среднего значения и дисперсии в схемах типа изображенной на фиг. Г.5, где стационарный сигнал модулируется функцией $f(t)$ и затем поступает на линейный оператор с весовой функцией $g(t)$.

Произведение синусоидального и случайного сигналов. Рассмотрим систему, изображенную на фиг. Г.6. Другими словами, ограничимся изучением влияния аддитивного шума. Как и в гл. 10, область интегрирования ограничена интервалом $(0, T)$, где $T = kT_i$, k — целое число и $T_i = 2\pi/\omega_i$.

a) **Среднее значение.** В гл. 10 отмечалось, что ввиду независимости сигнала и шума последний не влияет на математическое ожидание.



Ф и г. Г.6.

б) **Дисперсия.** В гл. 10 были получены следующие выражения для дисперсии:

$$\left. \begin{aligned} \sigma_c^2(T) \\ \sigma_d^2(T) \\ \sigma_{cd}^2(T) \end{aligned} \right\} = \frac{4}{u_0^2 T^2} \int_0^T \int_0^T \Psi_{nn}(\theta - \xi) \left\{ \begin{aligned} \sin \omega_i \xi \sin \omega_i \theta \\ \cos \omega_i \xi \cos \omega_i \theta \\ \sin \omega_i \xi \cos \omega_i \theta \end{aligned} \right\} d\xi d\theta. \quad (\Gamma.19)$$

Эти формулы можно преобразовать к виду, аналогичному (Г.15). Вновь, полагая $\theta - \xi = v$ и деля область интегрирования (фиг. Г.2) на две части (фиг. Г.3) ($t = T$), получаем

$$\begin{aligned} \sigma_c^2(T) &= \frac{4}{u_0^2 T^2} \int_0^T dv \Psi_{nn}(v) \int_0^{T-v} \sin \omega_i \xi \sin \omega_i (\xi + v) d\xi + \\ &+ \frac{4}{u_0^2 T} \int_0^T dv \Psi_{nn}(v) \int_0^{T+v} \sin \omega_i (\theta - v) \sin \omega_i \theta d\theta. \end{aligned}$$

Поскольку $\Psi_{nn}(-v) = \Psi_{nn}(v)$, оба интеграла равны и

$$\sigma_c^2(T) = \frac{8}{u_0^2 T^2} \int_0^T dv \Psi_{nn}(v) \int_0^{T-v} \sin \omega_i \xi \sin \omega_i (\xi + v) d\xi,$$

где

$$\sin \omega_i \xi \sin \omega_i (\xi + v) = \frac{1}{2} \cos \omega_i v - \frac{1}{2} \cos \omega_i (2\xi + v),$$

откуда следует выражение (10.24)

$$\begin{aligned} \sigma_c^2(T) &= \frac{4}{u_0^2 T^2} \int_0^T \Psi_{nn}(v) \left\{ (T-v) \cos \omega_i v - \right. \\ &\quad \left. - \frac{1}{2\omega_i} \sin \omega_i (2T-v) + \frac{1}{2\omega_i} \sin \omega_i v \right\} dv = \\ &= \frac{4}{u_0^2 T} \int_0^T \Psi_{nn}(v) \left\{ \left(1 - \frac{v}{T} \right) \cos \omega_i v + \frac{1}{\omega_i T} \sin \omega_i v \right\} dv. \end{aligned} \quad (\Gamma.20)$$

Аналогично можно вывести выражения для $\sigma_d^2(T)$ и $\sigma_{cd}^2(T)$.

ПРЕДМЕТНЫЙ УКАЗАТЕЛЬ

- Автоматическая корректировка 47
Автоматический контроль 45
Авторегрессия 247, 534
Адаптация 33
Аппроксимация дифференциальная 392, 582
— стохастическая 229, 317
Априорная информация 22, 54
— в байесовском оценивании 193, 523
— в корреляционных методах 377
— в неявных методах 315
— при оценивании состояния 548
- Вектор ошибок оценок параметров 421
— параметров 418
— эталонный 418
Вероятность 639
— условная 640
Взвешивание экспоненциальное 316
Возбуждение постоянное 536
Время задержки 480
- Гармоника 121
Гипотеза разделимости 593
Градиент 72, 355, 418, 456
Группировка переменных 140
- Динамическое программирование 578
Дисперсионная функция 400
Дисперсия 645
Дифференциальная аппроксимация 392
Доверительный интервал 333
Дуальное управление 33, 593
- Задача краевая двухточечная 578
— многоточечная 581
 z -преобразование 87, 279
- Идентификация 25
— в реальном времени 300
— обучением 305
— рекуррентная 300
Идентифицируемость 37, 153
Изменение избыточности адаптивное 605
Инвариантное погружение 584
Интервал доверительный 333
Информация априорная 22, 54
- Квазилинеаризация 581
Квантователь 382
Ковариационная матрица оценок 251, 306, 553
Ковариация 646
Контроль автоматический 45
Корректировка автоматическая 47
Коррелятор 385
— аналоговый 386

- релейный 387
- совпадения знаков 387
- Стилтьеса 387
- цифровой 386
- Корреляционная функция 95, 372, 651
 - высшего порядка 400
 - — двоичных сигналов 388
 - — оценивание 372
- Коэффициент влияния параметров 74, 446
 - корреляции 384, 646
 - усиления 72, 303, 531, 554
- Критерий (функционал) ошибки 58, 67, 422
 - интегральной абсолютной ошибки 423, 440
 - — квадратичной ошибки 423, 440

- Матрица вход-выход 120
 - выходная (наблюдений) 120
 - идемпотентная 275, 659
 - информационная 529
 - ковариационная 251, 306, 553
 - переходов (фундаментальных решений) 119
 - псевдообратная (обобщенная обратная) 672
 - распределительная 120
 - системы 120
- Медиана 520
- Метод вспомогательных переменных 293, 395, 458
 - Гаусса — Ньютона 218
 - групповой обработки данных 177
 - Левина 294
 - Ляпунова 213, 444, 538
 - максимального правдоподобия 65, 199, 526
 - Маркуардта 219
 - Монте-Карло 525
 - наименьших квадратов 64, 202, 303, 542
 - — — многократный 292
 - — — обобщенный 294,
- Методы градиентные 72, 211
 - итеративные 58, 211
 - корреляционные 361
 - неявные 58, 299, 415
 - поисковые 211
 - табличные 211
 - явные 58, 242, 361
- Мода 520
- Модель 17, 61
 - авторегрессии 247, 534
 - каноническая 169
 - линейная 110, 158, 169
 - — нестационарная 109, 132
 - — по параметрам 76, 158, 301
 - настраиваемая 56, 67, 299, 415
 - — нелинейная 112
 - — по параметрам 354
 - обобщенная 32, 77, 278, 470
 - — регрессионная 247
 - обратимая 32
 - прогнозирования (предсказания) 280, 541
 - топологически идентичная 426
 - эквивалентная 25
 - эталонная 458, 594
 - Моменты 129, 645
 - m*-последовательность 497

- Наблюдаемость 37, 153
 Наблюдатель 555
 Наблюдения 17, 35
 Настройка прерывистая 459
 — циклическая 213
 Невязки 274
 Независимость 643
 Неопределенность 20, 377
 Неравенство Бесселя 85
 — Коши — Шварца 528
 — Крамера — Рао 50, 528
 — Шварца 254, 665
- Обновленное представление 173
 Обучение 33, 305
 Объект 23, 25
 Овраг Розенброка 223
 Оператор политики 419, 430
 Отражение сжатое 318
 Оценивание 24, 55
 — байесовское 193, 518
 — в замкнутом контуре 58
 — марковское 65, 202, 253
 — неявное 58
 — по методу максимального правдоподобия 65, 199, 526
 — — — наименьших квадратов 64, 202, 251, 303, 535
 — порядка модели 275
- Оценка
 — асимптотически несмешенная 201
 — — — нормальная 201, 532
 — — — эффективная 201, 532
 — достаточная 193, 531
 — инвариантная 201
 — итеративная 205
 — линейная 202
 — минимального риска 66, 197
 — одноразовая 205
 — смешенная 191
 — с минимальной дисперсией 202
 — состоятельная 193, 532
 — эффективная 193, 531
- Ошибка 30, 67, 263, 269
 — входная 32
 — выборочной аппроксимации 273
- выходная 31, 62
 — мгновенная 429
 — обобщенная 32, 77, 416, 429
 — остаточная 249
 — упрощения при реализации 270
 — усредненная по времени 440
 — усечения 265
- Параметры 25, 113
 Переменные вспомогательные 293, 458
 Плотность распределения 95, 520, 641
 — — — апостериорная условная 194, 518
 — — — — при оценке состояния 548
 — спектральная 95, 537, 652
- Погружение инвариантное 584
 Полиномы Колмогорова — Габора 134, 178
 — Лагерра 86, 149, 164
 — Лежандра 86
 — Чебышева 86
 — Эрмита 86, 149
- Политика 419
 Порядок модели 275
 Последовательности двоичные псевдослучайные 496
 — максимальной длины (m -последовательности) 497
- Постоянное возбуждение 37, 536
 Правило Лейбница 117
 — решающее 525
- Пределочные теоремы 49
 Предсказание (прогнозирование) 280, 535
 Представление обновленное 173
 Преобразование конгруэнтное 672
 — Лапласа 87, 139, 395, 452
 — ортогональное 672
 — подобия 672
 — Фурье 83, 97, 396, 652
 — — быстрое 99
 — — — дискретное 98
- Принцип копирования 293
 — максимума 578

- сжатых отражений 318
- Программирование динамическое 578
- Пространство параметров 113
 - функциональное 112
- Процедура Шмидта 85
- Процесс случайный 646
 - стационарный 651
- Равенство Парсеваля 85
- Распознавание образов 41, 616
- Распределение гауссовское (нормальное) 95
 - с эллиптической симметрией 389
- Реакция на единичный импульс 114, 124
- Реализация 56
- Регистры сдвига 497
- Регрессия 242, 321
- Риск действия 35
 - изучения 35
 - средний 197, 525
 - условный 196
- Ряд Вольтерра 112, 133
 - Тейлора 208, 355
 - Фурье 83
- Свертка 116, 363
- Сигналы 39, 81
 - аналоговые 82
 - выборочные (импульсные) 81, 242, 299
 - двоичные 82, 388
 - дополнительные 389
 - квантованные 82, 382
 - непрерывные 81, 361, 415
 - псевдотелеграфные 496
 - телеграфные 496
 - тестовые (пробные) 36, 257, 477
 - большие 479
 - — периодические 484
- Система управления оптимальная 44
 - — самонастраивающаяся (адаптивная) 45, 594
 - — стабилизирующая 600
 - — структура 598
- Смещение 191, 283
- Собственные векторы 121, 672
 - числа (значения) 121, 672
- Сопряженное уравнение 132
- Среда 17
- Среднее 520
 - скользящее 534
 - условное 646
- Средний риск 197, 525
- Степень нелинейности 404
- Стохастическая аппроксимация 229, 317
- Стохастический (случайный) процесс 646
- Структура модели 24
- Суперпозиция 109
- Схема настройки дискретная (прерывистая) 73, 420
 - — непрерывная 72
- Схемы сходимости детерминированные 207
 - — непрерывные 211
 - — прерывистые (дискретные) 213
 - — стохастические 229
- Сходимость 207
 - импульсные сигналы 305
 - монотонная 436
 - непрерывные сигналы 305
- Теорема Байеса 194, 574
 - Кохрана 276
 - Кэли — Гамильтона 123, 665
 - Слуцкого 287
- Теоремы предельные 49
- Точка чувствительности 452
- Точность 51, 262, 517
- Управляемость 37, 153
- Уравнение свертки 366
 - решение 366
 - сопряженное 132
 - Фредгольма 543
 - чувствительности 446
- Уравнения канонические 579
 - нормальные 249
 - Фоккера — Планка 518
 - Эйлера — Лагранжа 578

- Условная вероятность 640
 Условное среднее 646
 Условный риск 196
- Фильтр адаптивный** 605
 — Винера 258, 542, 552
 — Калмана — Бьюси 173, 541, 552
 — — расширенный 555
 — Фурье 485
- Фильтрация линейная последовательная** 320
 — оптимальная 33
- Фильтры** 92
 — корреляционные 483
 — ортогональные 257, 368
 — полосовые 491
- Форма квадратическая** 670
- Функции Лагерра** 86, 149, 164
 — метода 396
 — модулирующие 397
 — независимые инвариантно относительно сдвига 390
 — ортонормальные 83
 — Радемахера 391
 — Эрмита 86, 149
- Функция** 53
 — весовая (импульсная переходная) 114
 — выборочная 646
 — дисперсионная 400
 — корреляционная 95, 372, 651
 — нечетная 67
 — ошибки 208, 423, 573
 — передаточная 127, 131
- — импульсная (дискретная) 127
 — переходная 114
 — потерпь (штрафа) 25, 407, 520
 — правдоподобия 66, 199, 526
 — производящая 669
 — распределения 641
 — характеристическая матрицы 671
 — — распределения вероятностей 97, 384
 — четная 67
 — чувствительности 74, 446
 — — логарифмическая 455
 — — относительная 451
- Функция-окошко** 98
- Шум** 202, 230, 267, 273, 283
 — входной 120
 — квантования 385
 — наблюдений (измерений) 120
 — объекта 120
- Экстраполяция** 541, 553
- Эллипсоид рассеивания** 531
- Ядро** 133
- Якобиан** 581

ОГЛАВЛЕНИЕ

Предисловие к русскому изданию	5
Предисловие автора к русскому изданию	10
Предисловие автора	11
Глава 1. Понятия и идеи. Построение модели, оценивание параметров и состояний	17
1.1. Модели	17
1.2. Построение модели	23
1.3. Структура, параметры и состояния	25
1.4. Описание проблемы	29
1.5. Условия нормального функционирования	35
1.6. Некоторые области применения	41
1.7. Дальнейшие перспективы	49
1.8. Заключение	50
Глава 2. Основные подходы: статистический и инженерный	53
2.1. Общие аспекты задачи	54
2.2. Типы реализации	56
2.3. Различные статистические подходы	60
2.4. Методы оценивания по настраиваемой модели	67
2.5. Оценивание параметров и состояний	78
2.6. Заключение	79
ПРЕДВАРИТЕЛЬНЫЕ СВЕДЕНИЯ	
Глава 3. Сигналы: детерминированные и случайные	81
3.1. Типы сигналов	81
3.2. Ортогональные функции	82
3.3. Описание детерминированных сигналов	87
3.4. Описание стохастических сигналов	94
3.5. Обработка сигналов; быстрое преобразование Фурье	97
Глава 4. Модели объектов: линейные с постоянными и переменными параметрами, нелинейные	108
4.1. Классы моделей	108
4.2. Линейные модели	114
4.3. Нестационарные линейные модели	132

4.4. Нелинейные модели	133
4.5. Управляемость, наблюдаемость, идентифицируемость	153
4.6. Модели	158
Глава 5. Некоторые вероятностные понятия. Теория оценивания, сходящиеся алгоритмы, стохастическая аппроксимация	189
5.1. Теория оценивания	190
5.2. Детерминированные сходящиеся алгоритмы	207
5.3. Вероятностные алгоритмы; стохастическая аппроксимация	229
ОЦЕНИВАНИЕ ПАРАМЕТРОВ	
Глава 6. Выборочные сигналы; явные методы	242
6.1. Регрессионный анализ	242
6.2. Реализация разомкнутых схем оценивания	255
6.3. Точность, некоторые источники ошибок [23, 26, 29, 37]	262
6.4. Невязки, шумы и порядок модели	273
6.5. Распространение на обобщенные модели и объекты с обратной связью	278
Глава 7. Импульсные сигналы (выборочные значения); неявные методы или методы с настраиваемой моделью	299
7.1. Модели, линейные по параметрам	301
7.2. Распространение на обобщенные модели и модели в пространстве состояний	318
7.3. Машины процедуры и результаты вычислений	331
7.4. Модели, нелинейные по параметрам	354
Глава 8. Непрерывные сигналы; явные методы	361
8.1. Корреляционные методы для аналоговых сигналов	363
8.2. Корреляционные методы для квантованных сигналов	382
8.3. Корреляционные методы для двоичных сигналов	388
8.4. Дифференциальная аппроксимация и связанные с ней методы	392
8.5. Корреляционные функции высших порядков	400
8.6. Дисперсионные функции; определения и свойства	400
Глава 9. Непрерывные сигналы; неявные методы (настраиваемые модели)	415
9.1. Модели, линейные относительно параметров	416
9.2. Минимизация мгновенного значения ошибки	429
9.3. Минимизация усредненной по времени ошибки	440
9.4. Функции чувствительности параметров	444
9.5. Одновременное использование двух моделей или повторное использование одной модели	464
9.6. Модель с возмущением параметров	465
9.7. Заключение	470

Глава 10. Тестовые сигналы: периодические и большие	477
10.1. Импульсные и ступенчатые сигналы	479
10.2. Синусоидальные тестовые сигналы	484
10.3. Двоичные последовательности	495
Глава 11. Байесовские оценки и оценки по методу максимального правдоподобия	518
11.1. Байесовские оценки	518
11.2. Оценивание по методу максимального правдоподобия	526
11.3. Некоторые схемы практической реализации	532
11.4. Требования к входным сигналам объекта	536
ОДНОВРЕМЕННОЕ ОЦЕНИВАНИЕ ПАРАМЕТРОВ И СОСТОЯНИЯ	
Глава 12. Оценивание состояния объекта (обзор)	541
12.1. Фильтр Винера	542
12.2. Фильтр Калмана — Бьюси	544
12.3. Обсуждение, приложения и примеры	556
Глава 13. Оценивание параметров и состояния	569
13.1. Нелинейный характер задачи	569
13.2. Вывод функции ошибки	573
13.3. Минимизация функции ошибки	578
13.4. Некоторые алгоритмы	580
13.5. Заключение	588
ПРИЛОЖЕНИЯ	
Глава 14. Обзор практических применений	592
14.1. Автоматическое управление	593
14.2. Физические, механические и химические объекты	600
14.3. Ядерные реакторы	603
14.4. Энергетические и энергораспределительные системы	603
14.5. Связь	604
14.6. Авиация и космонавтика	605
14.7. Биологические объекты	606
14.8. Социально-экономические системы	615
14.9. Обучающиеся системы и распознавание образов	616
Приложение А. Список обозначений	634
Приложение Б. Основные понятия из теории вероятностей и теории случайных функций	639
Приложение В. Основные понятия теории матриц	659
Приложение Г. Переходные процессы статистического среднего и дисперсии при корреляционных изменениях	667
Предметный указатель	676

УВАЖАЕМЫЙ ЧИТАТЕЛЬ!

Ваши замечания о содержании книги, ее оформлении, качестве перевода и другие просим присыпать по адресу: 129820, Москва, И-110, ГСП, 1-й Рижский пер., 2, издательство «Мир».

П. Эйхофф

**ОСНОВЫ ИДЕНТИФИКАЦИИ
СИСТЕМ УПРАВЛЕНИЯ**

Редактор Л. Якименко

Художник Ю. Урманчеев

Художественный редактор В. Бисенгалиев

Технический редактор Н. Борисова

Сдано в набор 28/X 1974 г.

Подписано к печати 25/IV 1975 г.

Бумага кн.-журн. 84×108^{1/32}=

=10,94 бум. л. Усл. печ. л. 36,75.

Уч.-изд. л. 35,19. Изд. № 20/7540

Цена 2 р. 64 к. Зак. 0565

ИЗДАТЕЛЬСТВО «МИР»

Москва, 1-й Рижский пер., 2

**Ордена Трудового Красного Знамени
Московская типография № 7 «Искра
революции» Союзполиграфпрома при**

**Государственном комитете Совета
Министров СССР по делам издательств,
полиграфии и книжной торговли.
Москва, К-1, Трехпрудный пер., 9.**