**МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РЕСПУБЛИКИ КАЗАХСТАН**

**КАЗАХСКИЙ НАЦИОНАЛЬНЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ имени К.И. САТПАЕВА**

Д.Ж. Сыздыков

# ИДЕНТИФИКАЦИЯ СИСТЕМ АВТОМАТИЧЕСКОГО УПРАВЛЕНИЯ

Рекомендовано Научно-методическим советом университета в качестве учебного пособия

Алматы 2012

УДК 618.5 (075.8)
С 95

**С 95 Сыздыков Д.Ж.** Идентификация систем автоматического управления: Учеб. пособие – Алматы: КазНТУ, 2012. – 278 с.
Ил. 62

ISBN 978–601–228–378–5

*В данном учебном пособии изложены основные понятия теории идентификации в системах управления. Описаны основные методы идентификации. Рассматриваются модели детерминированных и стохастических объектов, анализируются особенности применения временных, частотных, стохастических непараметрических и параметрических методов. Рассматриваются постановка и решение задачи идентификации линейных и нелинейных объектов в условиях неполной определенности их характеристик с применением устойчивых алгоритмов.*

*Учебное пособие предназначено для магистров специальности 050702 – «Автоматизация и управления», а также может быть применено бакалаврами и магистрами при изучении дисциплин «Моделирование систем управления», «Автоматизация технологических процессов и производств», «Адаптивные системы управления».*

УДК 618.5 (075.8)

Рецензенты:
*Бекбаев А.Б.* зав. каф. «ЭиАТК», д-р техн. наук, проф.
*Биттеев Ш.Б.* каф. «АиУ» КазАТК им. М. Тынышбаева, д-р техн. наук, проф.
*Хисаров Б.Д.* зав. каф. инженерной кибернетики Алматинского университета энергетики и связи, канд. техн. наук, проф.

Печатается по плану издания Министерства образования и науки Республики Казахстан на 2012 г.

ISBN 978–601–228–378–5

© Сыздыков Д.Ж., 2012
© КазНТУ, 2012

# Аннотация

В данном учебном пособии изложены основные понятия теории идентификации в системах управления. Описаны основные методы идентификации. Рассматриваются модели детерминированных и стохастических объектов, анализируются особенности применения временных, частотных, стохастических непараметрических и параметрических методов. Рассматривается постановка и решение задачи идентификации линейных и нелинейных объектов в условиях неполной определенности их характеристик с применением устойчивых алгоритмов.

Учебное пособие предназначено для магистров специальности 050702 – Автоматизация и управления, а так же может быть применена бакалаврами и магистрами при изучении дисциплин «Моделирование систем управления», «Автоматизация технологических процессов и производств», «Адаптивные системы управления».

# ВВЕДЕНИЕ

Основные этапы проектирования, внедрения и сопровождения автоматизированных систем управления сопряжены с решением задач построения математических моделей и моделированием реальных технических объектов или технологических процессов. Отличительной особенностью этих задач стало в настоящее время обращение к сложным объектам и процессам, повышению требований к качеству управления.

Моделирование объектов различной физической природы является одним из приемов научного познания, а основные закономерности процессов построения моделей исследуются в различных разделах теории познания. Методы моделирования лежат в основе теории управления и широко используются в прикладных разделах этой науки. Особое место данного направления занимают проблемы построения математических моделей, ориентированных для целей адаптивного управления.

Построение математической модели достаточно сложного объекта представляет собой довольно трудоемкий процесс, включающий этапы выбора структуры модели идентифицируемого объекта, выбора или разработки метода или численных алгоритмов идентификации, предварительной обработки результатов измерений, получения оценок характеристик модели, анализа этих оценок и проверки адекватности модели и реального объекта. Задача каждого из указанных этапов составляет весьма сложную проблему.

Вполне естественно, что при решении задач идентификации заданного объекта в реальных условиях приходится сталкиваться с массой новых неясных, и, возможно, нерешенных проблем. Трудности могут иметь место непосредственно при выборе вида математической модели объекта и модели процесса функционирования объекта, включающего кроме объекта модели помехи измерительных устройств. Сложным может оказаться этап выбора или разработка метода и устойчивых вычислительных алгоритмов идентификации с уче-

том ограниченных возможностей вычислительных средств и аппаратуры дистанционного измерения и передачи данных.

В связи с этим в данной работе не ставится своей целью дать исчерпывающий ответ на все вопросы, возникающие в задачах восстановления математических моделей сложных объектов в реальных условиях, их эксплуатации и испытаний. Изложенные в книге сведения по теории идентификации не могут дать полного представления обо всей проблеме построения математических моделей технических объектов. Более подробные исследования по проблемам идентификации, достаточно полно представляющих существующие методы и алгоритмы идентификации, можно найти, например, в работах [3, 9,10, 13, 17, 20–25, 35–39].

В рамках объема учебного пособия представлены основные сведения о методах идентификации технических объектов, реализация которых наиболее часто используется в инженерной практике. Проведен анализ задач, возникающих при разработке систем идентификации, которые можно представить в виде двух больших классов. К первому классу относятся задачи, которые решаются на этапе исследования и математического описания объектов управления, когда для описания точных и адекватных моделей допустимо использование сложных и трудоемких алгоритмов. Ко второму классу относятся задачи идентификации настройки параметров и моделей непосредственно в процессе функционирования систем управления, когда требования простоты, обеспечения быстрой и приближенной корректировки моделей являются определяющими.

Отдельные главы учебного пособия посвящены задачам идентификации сложных технических объектов и их особенностям. Представлены новые работы в теории идентификации по конструированию эффективных и достаточно простых систем идентификации на базе методологических принципов общего параметра. Основной идеей данного подхода является отказ от крайних позиций – одновременной раздельной настройки всех параметров модели и настройки одного общего параметра. Рассматриваются более гибкие подходы последовательного поиска и комбинированные подходы, которые позволяют реализовать преимущества обоих перечисленных выше "крайних" случаев – высокую точность и высокую скорость сходимости процессов идентификации.

# ГЛАВА 1. ПРОБЛЕМА ИДЕНТИФИКАЦИИ

## 1.1. Операторные возмущающие воздействия

В настоящее время комплекс проблем, связанный с созданием математических моделей технических объектов, технологических процессов, экономики и живой природы, формирует одно из основных направлений современной науки и техники. Это объясняется тем, что математические модели объектов широко применяются как при разработке и испытании систем управления этими объектами, так и при их эксплуатации.

Построение математической модели технического объекта (или объекта) может производиться тремя методами: аналитическим, экспериментальным и экспериментально-аналитическим.

Аналитический метод предусматривает получение математического описания объекта на основе законов физики, механики, химии и т. д., и технологических данных. Такой подход дает положительный результат, если рассматриваемый объект хорошо изучен. Если же объект изучен недостаточно или же он настолько сложен, что аналитическое составление его математической модели практически невозможно, прибегают к экспериментальным методам. При экспериментально - аналитическом методе априорная модель, полученная аналитическим путем, уточняется при соответствующих экспериментах.

Взаимодействие объекта с внешней средой можно представить с помощью простейшей схемы (рис. 1.1). Воздействия внешней среды на объект в обобщенном виде изображены стрелками, направленными к объекту и обозначенными через $u$ и $v$. Объект в свою очередь воздействует на окружающую среду. Это воздействие на рисунке показано стрелкой, направленной от объекта и обозначенной через $x$, являющейся выходным воздействием или выходной величиной объекта.

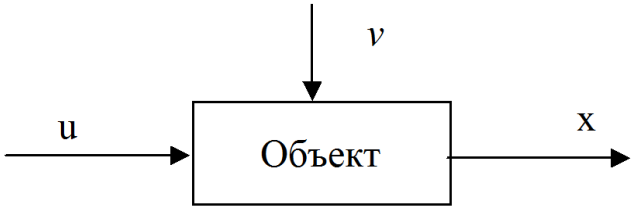



Рис. 1.1. Простейшая схема воздействия среды на объект

Всю совокупность воздействий внешней среды на объект можно разделить на две группы. Такое разделение можно произвести по различным признакам, например, по характеру влияния воздействий внешней среды на переменные объекта. В соответствии с этим в первую группу входят те воздействия, которые в точке приложения изменяют переменные состояния аддитивно. Последнее означает, что сигналы, пропорциональные этим воздействиям, суммируются с сигналами, пропорциональными соответствующим переменным состояния. Данные входные воздействия могут быть как управляющими и, так и возмущающими $\nu$ воздействиями.

Вторая группа воздействий внешней среды изменяет переменные состояния объекта косвенно, обычно неаддитивно. Эти воздействия непосредственно приводят к изменению оператора объекта, под которым понимают закон (правила) преобразования входных воздействий в выходные переменные величины объекта.

Математически операторы определяются в соответствующих пространствах, т. е. на множествах элементов, под которыми совершаются преобразования. Примерами таких пространств служат пространства: непрерывных функций; непрерывных функций, имеющие непрерывные производные до $n$-го порядка $n \geq 0$ и т. д. Множества входных и выходных сигналов ряда объектов и систем автоматического управления могут рассматриваться как те или иные метрические пространства.

Формально оператор объекта характеризуется структурой и параметрами. Внешние воздействия $\nu$ могут влиять как на структуру, так и на параметры объекта. В зависимости от этого различают структурные и параметрические воздействия.

Отметим некоторые особенности операторных возмущений.

Первая особенность связана с тем, что их действие проявляется только при наличии входных воздействий или при ненулевых начальных условиях, которые могут быть сведены к эквивалентным входным воздействиям.

В качестве второй особенности операторных возмущающих воздействий отметим тот факт, что по отношению к последним любая система является обычно нелинейной. Это касается и линейных систем по принятой в теории управления терминологии.

Третья особенность, как следствие первой, связана с тем, что эффект воздействия операторных возмущений на процесс управления

может проявляться со значительным сдвигом во времени по отношению к моменту появления возмущения.

Операторные возмущающие воздействия вносятся в объект в течение всего его жизненного цикла (проектирования, производства, хранения, эксплуатации и т. д.), но проявляются они только при испытаниях или при эксплуатации объекта.

Отметим, что операторные воздействия могут быть как полезными (управляющими), так и вредными (помехи). Полезны, например, настройки параметров системы в ходе наладки регулятора или в процессе адаптации.

Остановимся кратко на факторах, определяющих операторные воздействия среды на объект. Наиболее существенными факторами следует считать физическое окружение, степень развития науки, техники и производства и уровень организации эксплуатации объекта.

В понятие "физические окружение" включаются источники энергии, температура, влажность, вибрации и механические удары, радиоактивное излучение и т. д. Все эти факторы определяют старение элементов и блоков объекта. Под влиянием температуры, влажности и других факторов значения параметров объекта со временам "уходят" от своих номинальных значений. При этом объект подвергается, как правило, параметрическим возмущающим воздействиям. При синтезе и проектировании системы управления техническим объектом, как правило, необходимо иметь математическую модель объекта. При этом стремление учесть наиболее полно все тонкости технической системы часто приводит к такой сложной математической модели, что практическое использование ее для синтеза системы не всегда представляется возможным. Ввиду этого приходится идти на значительные упрощения модели. Упрощенные (приближенные) модели применяются и при отсутствии "полного" математического описания объекта. Все эти упрощения являются источником операторных, в основном структурных, возмущений.

Привнесенные в систему (объект) на этапах проектирования и производства операторные возмущения проявляются при ее эксплуатации. Кроме того, система может подвергаться дополнительным операторным воздействиям как при транспортировке и хранении, так и во время самой эксплуатации. Величины этих воздействий определяются как физическим окружением, так и уровнем организации транспортировки, хранения и эксплуатации.

## 1.2. Особенности решения задач идентификации

### 1.2.1. Идентификация и математическая статистика

Основная проблема теории идентификации связана с разработкой методов построения математических моделей объектов. Построение моделей по реализациям входных и выходных сигналов является наиболее характерной обратной задачей [10, 22, 24, 36, 38].

Стохастичность, динамичность и специфика решения обратных задач и определяет особенности процесса идентификации.

Стохастичность приводит к тому, что большинство методов идентификаций строится на основе математической статистики, которая в прикладном плане занимается описанием, обработкой и анализом экспериментальных данных с учетом случайных ошибок измерений и помех в каналах передачи данных.

Практически все методы математической статистики находят или могут найти применение при идентификации технических объектов.

Известно, что основными задачами математической статистики являются:

- оценивание по статистическим данным законов распределения случайных величин и функций;
- оценивание по статистическим данным неизвестных параметров распределения и числовых характеристик случайных величин и функций;
- проверка правдоподобия статистических гипотез;
- сглаживание экспериментальных зависимостей;
- оптимальная организация экспериментов.

Идентификация производится обычно на этапе полной или вторичной обработки информации. С позиции математической статистики целью идентификации является получение значений физических величин в результате ряда косвенных измерений. Так, при параметрической идентификации обычно определяются значения некоторых параметров $u_1, u_2, ..., u_n$, связанных с измеряемыми величинами $z_1, z_2, ..., z_k$, известными соотношениями

$$f_N(u_1, ..., u_n, z_1, ..., z_k), \ N = \overline{1,s} \geq n.$$

Величины $z_1, z_2, ..., z_k$ являются, как правило, результатами измерений входных и выходных сигналов объекта. Они могут измеряться

однократно или многократно различными методами в зависимости от точности метода измерений, характера измеряемой величины, возможностей измерителей и телеметрических приборов. В данной схеме задача идентификации сводится к получению наилучших оценок параметров $u_1 (i = \overline{1,n})$ и исследованию точностей найденных оценок.

В математической статистике предлагается предварительно вычислить несмещенные оценки $\hat{z}_j$ величин $z_j$, а затем разрешить соотношения

$$f_N(\theta_1,...,\theta_n, z_1,...,z_k) = 0$$

относительно искомых оценок $\hat{\theta}_1,...,\hat{\theta}_n$, т.е. $\hat{\theta}_i = \varphi_i(z_1,...,z_k), \ i = \overline{1,n}.$

Заметим, что оценки $\hat{\theta}_i$ могут зависеть и от оценок моментов второго порядка (дисперсий, корреляционных моментов, средних квадратических отклонений).

Очевидно, что в общем случае функции $f_N$ нелинейны, вследствие чего, оценки $\hat{\theta}_i$ могут оказаться смещенными. Поэтому важно после получения этих оценок провести анализ их точности, проверить их достоверность. Кроме того, анализ точности оценок позволяет исследовать надежность полученных результатов, сравнить найденные оценки параметров $\hat{\theta}_i$ с результатами, полученными другими методами. На этапах обработки и проектирования объектов и систем обработки измерений детальный анализ точности $\hat{\theta}_i$ позволяет оценить сравнительное влияние ошибок измерений, помех в каналах телеметрии, результатов первичной обработки и других факторов на окончательный результат.

Под первичной обработкой будем понимать процесс получения статистических характеристик результатов прямых измерений. В зависимости от потребности вторичной обработки такими характеристиками могут быть законы распределения или определенные числовые характеристики измеряемых сигналов. При получении этих характеристик предварительно производится отбраковка (исключение) из обработки аномальных измерений, "прореживание" принимаемых по телеметрическим каналам сигналов, если измерительные средства выдают результаты прямых измерений с большой частотой, сжатие информации и т. д. Эти операции непосредственно влияют на точность статистических характеристик результатов прямых измерений, которые в свою очередь определяют точностные характеристики результатов полной (вторичной) обработки, т. е. результатов идентификации.

Оценка точности может быть точечной и интервальной. Точечная оценка представляет собой некоторую функцию от выборочных данных, которая в определенном смысле наилучшим образом оценивает неизвестное значение параметра. Анализ точности при этом включает в себя обычно исследование оценок на состоятельность, несмещенность и эффективность.

Точечная оценка, являясь случайной величиной, практически всегда отличается от истинного значения параметра. Поэтому важно получить определенную информацию о возможных границах истинного значения оцениваемого параметра. Такую информацию дает интервальная оценка, характеризующаяся доверительными интервалами. Если для получения точечной оценки достаточно иметь только результаты измерений, то для получения интервальных оценок необходимо, кроме того, знать законы распределения измеряемых сигналов.

Кроме указанного, анализ точности полученных в ходе идентификации результатов может включать в себя этап проверки адекватности модели, т. е. пригодности. В статистике разработаны специальные критерии, используемые для проверки гипотезы об адекватности специальных (обычно статистических) моделей. Большинство методов рассматривает остаточную дисперсию или критерий Фишера.

Полное исследование точности результатов идентификации представляет подчас довольно сложную и длительную процедуру даже при использовании современных вычислительных средств. Поэтому при оперативной идентификации можно ограничиваться только получением точечных оценок параметров модели объекта.

Кроме вопросов обработки результатов измерений в экспериментальной практике, современная математическая статистика занимается проблемой оптимальной в определенном смысле организации самого эксперимента. Построение математической модели объекта производится в процессе эксперимента – активного или пассивного. Во многих случаях это эксперимент чрезвычайно сложный, длительный и дорогостоящий. На проведение многих экспериментов затрачиваются огромные материальные, временные и людские ресурсы. Поэтому важно научиться рационально проводить такие эксперименты, используя в максимально возможной степени содержащуюся в каждом наблюдении информацию, уметь извлекать наибольшее количество информации об идентифицируемом объекте при ограниченных затратах на организацию эксперимента. Указанные вопросы

оптимальной организации и управления экспериментом рассматриваются в теории планирования эксперимента. В этой теории можно выделить два основных направления: планирования экстремальных экспериментов и планирование экспериментов для построения математических моделей объектов и явлений.

Планирование экстремальных экспериментов направлено на построение локальной (точечной) модели объекта, при которой изучаемый процесс удовлетворяет некоторому критерию оптимальности. По существу методы планирования такого типа позволяют настроить процесс на оптимальный режим. В задачах идентификации находят в основном применение методы планирования эксперимента второго типа.

В задачах идентификации эксперимент имеет ту особенность, что он, как правило, пассивный, т. е. проводится в условиях нормального функционирования объекта. При этом для обработки представляется возможным получать лишь по одной реализации случайных функций, описывающих входные и выходные сигналы объекта. Специфика работы телеметрических систем приводит иногда к тому, что эти реализации могут оказаться довольно "короткими".

### 1.2.2. Идентификация и теория управления

Как прикладное научное направление теория идентификации зародилась в теории управления. Очевидно, что для "хорошего" управления необходимо знать то, чем управлять, т. е. знать описание (модель) управляемых объектов. Последние в теории и практике правления, как правило, являются динамическими и описываются дифференциальными или интегро-дифференциальными уравнениями [1–3, 18, 22–25, 28, 36].

Методы и приемы теории управления используются на таких этапах идентификации, как выбор типа модели и его предварительный анализ, выбор показателей и критериев близости оптимальности, выбор метода идентификации. Предварительный анализ может охватывать выяснение условий управляемости, наблюдаемости и идентифицируемости объекта.

В связи с указанным весь процесс идентификации условно можно разделить на два этапа. На первом этапе методами теории управления разрабатываются общая схема и методы идентификации, на втором –

производится реализация этого метода с привлечением математической статистики. Динамичность объектов принципиальных затруднений не вызывает, т. к. дифференциальные и интегральные уравнения при наличии реализаций входных и выходных переменных сводятся к алгебраическим уравнениям относительно оцениваемых параметров.

Отметим также, что возможны постановки задач идентификации непосредственно на "языке" теории управления.

Укажем две наиболее характерные постановки.

В соответствии с первой постановкой задача идентификации формулируется в виде задачи синтеза оптимального управления. В этой задаче отыскивается допустимое управление, задаваемое обычно в определенном классе функций, которое доставляет наименьшее значение критерию оптимальности. При идентификации минимизируется критерий близости за счет выбора допустимого оператора или параметров модели.

Вторая постановка позволяет свести задачу параметрической идентификации к задаче наблюдения. Это достигается тем, что неизвестные параметры модели рассматриваются как дополнительные фазовые координаты (параметры состояния). Происходит искусственное расширение фазового пространства состояния объекта. Формулируется задача восстановления (наблюдения) расширенного вектора фазовых координат по доступным для измерения выходным переменным объекта. Для решения этой задачи могут быть привлечены все методы решения задач наблюдения линейных и нелинейных систем.

### 1.2.3. Идентификация и обратные задачи

Идентификация представляет собой наиболее характерную типовую обратную задачу. Обратные задачи широко распространены в физике, технике и других отраслях знаний. В теории управления обратными являются задачи восстановления входных сигналов по измеряемым выходным сигналам и задача идентификации. Характерная особенность всех обратных задач – их некорректность. Пусть в задаче параметрической идентификации уравнение для оценивания параметров $\theta$ имеет вид

$$F\theta = s, \qquad (1.1)$$

где $F$ – оператор, осуществляющий соответствующее отображение; $s$ – величины, которые получаются путем измерений и поэтому известны приближенно.

Задача определения решения $\theta$ из пространства $C_\theta$ по исходным данным $s \in C_s$ называется корректно поставленной, если выполняются следующие условия:

1. Для всякого элемента s существует решение $\theta \in C_\theta$;
2. Решение определяется однозначно;
3. Задача устойчивая.

При этом задача нахождения решения $\theta = f(s)$ называется устойчивой, если для любого числа $\varepsilon > 0$ можно найти такое число $\delta(\varepsilon) > 0$, что из неравенства $\rho_s(s_1, s_2) \le \delta(\varepsilon)$ следует неравенство $\rho_\theta(\theta_1, \theta_2) \le \varepsilon$, где $\theta_i \in f(s_i)$, $C_s$, $\theta_i$, $C_\theta$, $i = 1,2$; $\rho_\theta, \rho_s$ – соответствующие расстояния в нормированных пространствах.

На практике для устойчивой задачи малым ошибкам в исходных данных соответствуют малые ошибки в решении $\theta$.

Задачи, не удовлетворяющие хотя бы одному из перечисленных выше условий, называются некорректно поставленными.

Рассмотрим возможные причины некорректности задач идентификации. Исходными данными при идентификации являются составляющие вектора $s$ (апостериорная информация). Последние получаются путем измерений входных и выходных сигналов объекта и содержат неизвестные измерительные погрешности. Могут иметь место погрешности и из-за различных аппроксимаций и представлений входных и выходных сигналов. В результате при некоторых значениях вектора s уравнение (1.1) относительно $\theta$ может не иметь решения (в данном пространстве $C_\theta$). При этом нарушается первое условие корректности задачи – условие существования решения.

Если уравнение (1.1) является нелинейным относительно $\theta$, то решение может оказаться неоднозначным, что приводит к невыполнению второго условия. Наконец, даже если решение существует и единственно, то оно может не обладать свойством устойчивости. Это в конкретных задачах будет выражаться в том, что малые ошибки в исходных данных (в величине $s$) будут вызывать значительные ошибки в решении $\theta$. Причина "усиления" ошибок исходных данных заклю-

чается в том, что обратный оператор $F^{-1}$ не является непрерывным. Таким образом, нарушается третье условие корректности.

Сказанное ни в коей мере не означает, что некорректность задач идентификации делает практически бесполезным построение модели. Опыт, интуиция всегда позволяли в рамках возможного конструировать или определять приближенные решения задач идентификации, устойчивые к малым изменениям исходных данных. Это достигалось, в частности, за счет максимального использования всей априорной и, может быть, апостериорной информации относительно модели объекта, за счет использования специальных входных сигналов при активной идентификации.

Многие методы содержали в себе элементы регуляризации, т. е. алгоритмы, позволяющие найти решения некорректных задач. Например, широко используемый в теории идентификации метод наименьших квадратов позволяет найти решение избыточной (возможно, несовместной) алгебраической системы. Известны модификации МНК, обеспечивающие и устойчивость решения.

В настоящее время развивается теоретическая база, позволяющая целенаправленно находить решения таких некорректных задач, какими являются задачи идентификации. Основу этой базы составляют такие методы решения некорректных задач, как метод подбора, метод квазирешений, метод регуляризации [10, 26].

### 1.2.4. Чувствительность методов идентификации к отклонению законов распределения от предполагаемых. Помехоустойчивая идентификация

Существующие методы идентификации, как, например, метод наименьших квадратов, адаптивные алгоритмы идентификации позволяют получать оценки параметров модели при отсутствии информации относительно законов распределения помех. В этом отношении эти методы являются универсальными. Но оказывается, что аналогичные методы являются порой чрезвычайно чувствительными к отклонению законов распределения помех от предполагаемых. В условиях неучитываемых помех и редких выбросов эти методы оценивания теряют работоспособность.

В связи с отмеченным в теории оценивания и теории идентификации в последние годы, начали проводиться исследования по разра-

ботке так называемых помехоустойчивых или робастных методов и алгоритмов, обладающих малой чувствительностью к отклонениям законов распределения от предполагаемых.

Решение указанной проблемы связано с возможностью наиболее полного учета в формулировке задачи и метода идентификации имеющейся априорной информации как количественного, так и качественного характера. К такой априорной информации могут относиться данные о классе законов распределения помех (например, законы распределения с ограниченной дисперсией, законы распределения, близкие в определенном смысле к нормальному и т. д.). При этом учитывается тот факт, что выбор метода идентификации определяется законом распределения помехи. Покажем это на примере идентификации объекта, описываемого уравнением

$$x = f(\theta, u) + \nu, \qquad (1.2)$$

где $u$ – входной сигнал; $x$ – выходной сигнал; $\nu$ – случайная помеха; $f$ – некоторая функция, определенная с точностью до параметра $\theta$ (в общем случае векторного).

Требуется по измеренным значениям входных и выходных сигналов $u_1,..., u_N$, $x_1,..., x_N$ получить оценку $\hat{\theta}$ параметра $\theta$. При этом предполагается, что помехи $\nu_1,...,\nu_N$ независимы и одинаково распределены. Плотность распределения $p(\nu)$ задана лишь своим классом. Величина $\hat{\theta}$ оценивается в результате минимизации показателя

$$J(\hat{\theta}) = \sum_{i=1}^{N} Q[\varepsilon_i(\hat{\theta})], \qquad (1.3)$$

где $$\varepsilon_i(\hat{\theta}) = x_i - f(\hat{\theta}, u_i).$$

Очевидно, что если Q – квадратичная функция, то мы получаем метод наименьших модулей.

Вначале допустим, что закон распределения помехи полностью известен и что для выбора метода идентификации используется принцип максимального правдоподобия. Тогда, как известно из математической статистики, в случае нормального распределения принцип максимального правдоподобия приводит к МНК, а в случае распределения Лапласа – к методу наименьших модулей. Можно также показать, что при равномерном распределении помехи оптимальный метод идентификации должен строиться на основе чебышевского приближения. Последнее означает, что оценка $\hat{\theta}$ равна

$$\hat{\theta} = arg \min_{\theta} \max_{1 \le i \le N} \left| x_i - f(\hat{\theta}, u_i) \right|.$$

Так как фактические распределения обычно отличаются от предполагаемых, то получаемые оценки могут оказаться неустойчивыми (чувствительными) к распределениям.

Основная идея построения помехоустойчивых (грубых) методов идентификации заключается в следующем.

Пусть $\hat{\theta}$ – произвольная несмещенная оценка для $\theta$, построенная по результатам N измерений. Применим к рассматриваемой задаче известное в математической статистике неравенство Рао-Крамера

$$M\left\{(\theta - \hat{\theta})(\theta - \hat{\theta})^T\right\} \ge \sum_{i=1}^{N} \left[\nabla f(\hat{\theta}, u_i) \nabla^T f(\hat{\theta}, u_i)\right]^{-1} J^{-1}(p), \quad (1.4)$$

где $\nabla f(\hat{\theta}, u) = \frac{\partial f(\hat{\theta}, u_i)}{\partial \hat{\theta}}$; $J(\rho) = \int_{-\infty}^{\infty} \frac{[\rho'(v)]^2}{\rho(v)} dv$ – фишеровская информация (информационное количество Фишера).

Пусть относительно плотности распределения $p(v)$ известно лишь то, что она принадлежит некоторому классу $C_p$. Очевидно, "наихудшим" распределением явится то, что плотность $p(v)$ которого максимизирует правую часть неравенства Рао-Крамера. Так как в правой части (1.4) от функции $\tilde{p}(v)$ зависит только величина $J(p)$, то предыдущее условие сводится к минимизации фишеровской информации. Таким образом, наихудшей плотностью распределения $\tilde{p}(v)$ является та, которая определяется соотношением

$$\tilde{p}(v) = arg \min_{p \in C_p} J(p). \quad (1.5)$$

Используем эту плотность распределения для формирования функции правдоподобия (оценочной функции) [24, 34, 37, 38]

$$L(\hat{\theta}) = ln\, p(v). \quad (1.6)$$

В результате решение задачи идентификации сводится к решению вариационной задачи (1.5) и получения оценки параметра $\theta$ путем минимизации показателя $J(\hat{\theta})$ с оценочной функцией (1.6).

В литературе показано [21], что определяемая таким образом оценка идентифицируемых параметров является асимптотически оптимальной в минимаксном смысле для данного класса распределения $C_P$.

# 1ГЛАВА 2. МАТЕМАТИЧЕСКИЕ МОДЕЛИ ОБЪЕКТОВ ИДЕНТИФИКАЦИИ

## 2.1. Общие сведения о математических моделях и их классификация

Под математической моделью реального объекта понимается совокупность соотношений (например, уравнений, неравенств, логических условий и т. д.), определяющих характеристики состояний объекта (а через них и выходные сигналы) в зависимости от параметров объекта, входных сигналов, начальных условий и времени.

Понятию "соотношение" в этом определении придается весьма широкий смысл. В некоторых случаях эти соотношения могут быть представлены в виде явных функций от параметров системы. В других случаях математическая модель представляет собой совокупность уравнений (алгебраических, дифференциальных и др.) относительно переменных состояний объекта и выходных (наблюдаемых) сигналов. При этом параметры объекта входят в коэффициенты уравнений, а входные сигналы – в правые части.

При математическом описании технических объектов их часто схематически изображают в виде обобщенной блок-схемы, представленной на рис. 2.1, с выделением различных групп входных и выходных переменных по следующим признакам.

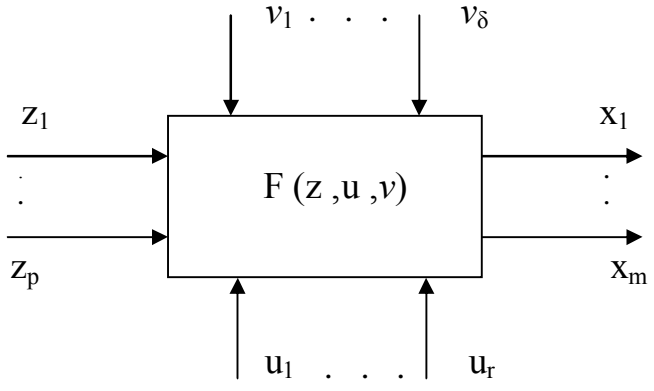



Рис. 2.1. Обобщенная блок-схема технических объектов с входными и выходными переменными

1. Контролируемые и неуправляемые переменные $z^T = (z_1,...,z_p)$, значения которых можно измерить, но нельзя произвольно изменить. Например, это могут быть параметры внешней среды (температура, давление, напряжение сети и т. п.), характеристики сырья и полупродуктов для технологического процесса и т. д.

2. Контролируемые и управляемые переменные $u^T = (u_1,...,u_r)$, значения которых в любой момент времени могут быть измерены и изменены контролируемым образом, либо человеком-оператором, либо автоматической системой из допустимого диапазона изменения. Например, управляющие воздействия на управляемых входах технического объекта (температура, скорость вращения, угол поворота и т. п.).

3. Неконтролируемые и неуправляемые переменные $\nu^T = (\nu_1,...,\nu_\delta)$, которые недоступны контролю и направленному изменению, но характеризуют множество реально существующих факторов, влияющих на текущее состояние технического объекта. Например, многочисленные примеси в сырье и реагентах, степень изношенности оборудования и т. д.

4. Контролируемые выходные переменные $x^T = (x_1,...,x_p)$, которые характеризуют реакцию объекта, результат его функционирования. Например, показатели качества и количества конечного продукта для технологической системы, параметры текущего состояния объекта и т. д.

Вид некоторого оператора $F$, согласно которого в процессе функционирования объекта осуществляется преобразование всех значений входных векторных переменных $u, z, \nu$ в значениях выходных переменных вектора $x$, определяет характер модели объекта, т. е.

$$x = F(z, u, \nu).$$

Учет всех компонент векторных переменных $x, z, u, \nu$ в реальных условиях функционирования не представляется возможным. Поэтому уже на стадии разработки моделей технического объекта как управляемого объекта число входных и выходных переменных сокращается в ряде случаев, что приводит к неадекватной (гомоморфной) модели технического объекта.

Для преодоления таких сложностей используются различные способы упрощения математического описания технического объекта. Решению такой задачи способствует то, что современные технологические процессы представляют собой некоторое число взаимо-

связанных объектов, каждый из которых описывается одно- или многомерной моделью, отражающей самые существенные связи процесса. Такое упрощение позволяет использовать в дальнейшем формальные математические модели объектов, сохраняющие в моделях существенные составляющие.

Математические модели могут быть получены либо чисто теоретическим способом, основанным на известных физических, физико-химических и других зависимостях, либо эмпирическим путем экспериментов по наблюдениям входных и выходных переменных объекта, либо разумной комбинацией этих двух подходов.

Математические модели чрезвычайно широко используются при исследовании различных объектов и процессов. Широкое применение математических моделей обусловлено следующими факторами:

1. Математическая модель безразлична к форме своего материального воплощения;

2. Реализация математической модели на современных ЭВМ и ее экспериментальное исследование связаны с гораздо меньшими материальными и временными затратами по сравнению с теми затратами, которые требуются для экспериментов над реальными объектами;

3. Математические модели позволяют широко применять при исследовании объектов формальный аппарат современной математики.

Сложность и многообразие реальных объектов не позволяют строить для них абсолютно адекватные модели. Математическая модель, будучи формализованным абстрактным описанием объекта, в состоянии охватить только основные характерные его свойства, оставляя в стороне несущественные факторы.

Математические модели могут быть представлены в различных видах в зависимости от типа моделируемого объекта, условий его функционирования, объема априорной информации и других факторов. Так, например, в зависимости от вида моделируемого объекта модели могут быть линейными и нелинейными, непрерывными и дискретными, статическими и динамическими, детерминированными и стохастическими и т. д.

По форме своего представления математические модели подразделяются на два больших класса: параметрические и непараметрические модели. Параметрические модели представляют собой совокупность математических соотношений, в которых параметры объекта фигурируют либо в явном виде, либо в виде коэффициентов уравнений.

К таким моделям относятся алгебраические и дифференциальные уравнения, полиномиальные выражения, передаточные функции.

Непараметрические модели представляют собой описание объекта в виде графиков, таблиц, графов и т. д. Таковыми являются, например, переходные и частотные характеристики объекта. Между параметрическими моделями существует связь, позволяющая по одному виду модели построить другой вид. Так, например, по передаточной функции легко получить переходную, импульсную переходную функцию и частотные характеристики [10, 18, 20, 24, 26].

Любая из указанных моделей представляет собой абстрактное информационное отражение реального объекта на формализованном языке математики, т. е. является его образом. Все эти абстрактные образцы связывают функционально входные и выходные сигналы объекта. По количеству входов и выходов объекты и соответствующие им модели классифицируют на одномерные и многомерные.

## 2.1. Модели статических объектов

Объекты, которые не могут иметь запасы энергии и связь между входными и выходными сигналами которых можно представить чисто алгебраическими выражениями, называются статическими. Математические модели статических объектов описываться как линейными, так и нелинейными математическими выражениями и в зависимости от свойств изучаемого объекта, с учетом упрощения записи, можно представить в следующих видах:

а) $$x = f(u); \qquad (2.1)$$

б) $$\varphi_i(x_1,...,x_n,u_1,...,u_m) = 0, \quad i = \overline{1,n}; \qquad (2.2)$$

в) $$x = f(u_1,...,u_m); \qquad (2.3)$$

г) $$\varphi_i(x_1,...,x_n,u) = 0, \quad i = \overline{1,n}. \qquad (2.4)$$

Системы (2.2) и (2.4) при удовлетворении условиям теоремы о существовании неявных функций могут быть разрешены относительно выходных сигналов $x_1, x_2, \ldots, x_n$

$$x_i = f_1(u_1,...,u_m), \quad i = \overline{1,n};$$
$$x_i = f_i(u), \quad i = \overline{1,n},$$

или в векторной форме

$$x = F(\bar{u}); \tag{2.5}$$
$$x = F(u). \tag{2.6}$$

Математические модели, представляемые уравнениями (2.1) – (2.6), содержат в своей структуре параметры (числовые величины). В случае, если статические объекты являются линейными (т. е. объектами, для которых справедлив принцип суперпозиции), то уравнения (2.1) – (2.4) имеют соответственно вид

$$x = ku; \quad Ax = Bu \tag{2.7}$$

или

$$\sum_{j=1}^{n} a_{ij} x_j = \sum_{j=1}^{m} b_{ij} u_{ij}, \quad i = \overline{1,n}; \tag{2.8}$$

$$x = k^T u; \quad x = ku; \; x_i = k_i u, \quad i = \overline{1,n}. \tag{2.9}$$

В приведенных соотношениях $k$, $a_{ij}$, $b_{ij}$, $k_j$ являются числовыми величинами параметров, которые могут быть функциями времени.

Если $|A| \neq 0$, уравнения (2.8) могут быть разрешены относительно входных переменных

$$x_i = \sum_{j=1}^{m} k_{ij} u_j, \quad i = \overline{1,n}, \tag{2.10}$$

или

$$x = Ku. \tag{2.11}$$

В случае, если оператор $f(\cdot)$ математической модели (2.3) может быть достаточно хорошо аппроксимирован полиномиальной структурой, например, второго порядка по $u_i$ ,..., $u_m$ вида

$$x = \sum_{i=1}^{m} a_i u_i + \sum_{i=1}^{m}\sum_{j=1}^{m} a_{ij} u_i u_j, \tag{2.12}$$

то здесь $a_i$, $a_j$ являются числовыми величинами параметров. Проведем в (2.12) замену переменных и параметров как

$u_1^2 = u_{m+1}, ..., u_m^2 = u_{2m}, \qquad u_1 u_2 = u_{2m+1}, ..., u_m u_{m-1} = u_k,$

$a_i = \theta_1, a_{11} = \theta_{m+1}, ..., a_{mm} = \theta_{2m} \qquad a_{12} = \theta_{2m+1}, ... a_{m,m-1} = \theta_k.$

Тогда любое из уравнений (2.2) можно переписать в виде уравнения

$$x = \sum_{i=1}^{N} \theta_i u_i = \theta^T u, \qquad N = m + m^2, \tag{2.13}$$

которое представляет собой модель изучаемого объекта в виде полиномиальной структуры, в которую параметры входят линейно.

Как следует, например, из выражений (2.11), (2.12) модели линейные по параметрам, могут быть линейными (2.11) и нелинейными (2.12) по физическим переменным. Такие модели часто называются линейными в смысле вхождения в их структуру параметров $\theta_1, ..., \theta_N$.

В практической проблеме идентификации статических объектов представляют интерес и нелинейные по параметрам модели. В этом случае иногда возможно осуществить их преобразование в линейную форму. Приведем некоторые примеры таких преобразований нелинейных по параметрам моделей в линейную форму.

Пусть объект с m входами $u_1(t)$ ,..., $u_m(t)$ и выходом $x(t)$ описывается уравнением

$$x(t) = k u_1^{\theta_1} u_2^{\theta_2} ... u_m^{\theta_m}. \tag{2.14}$$

Очевидным является преобразование

$$z(t) = \theta_0 + \theta_1 \varphi_1(t) + ... + \theta_m \varphi_m(t), \tag{2.15}$$

где

$$z(t) = ln\ x(t), \quad \theta_0 = ln\ k, \quad \varphi_i(t) = ln\ u_i(t), \qquad i = \overline{1,m}. \tag{2.16}$$

Исходная нелинейная относительно идентифицируемых параметров модель (2.14) за счет замены переменных (2.16) преобразована в линейную модель (2.15).

Рассмотрим экспоненциальные модели нелинейных объектов

$$x_1(t) = exp\{\sum_{i=1}^{m} \theta_i u^i(t)\}; \qquad x_2(t) = exp\{\sum_{i=1}^{m} \theta_i u_i(t)\}.$$

Обе эти модели довольно просто преобразуются в линейные модели вида

$$z_1(t) = ln\ x_1(t) = \sum_{i=1}^{m} \theta_i u^i(t); \qquad z_2(t) = ln\ x_2(t) = \sum_{i=1}^{m} \theta_i u_i(t).$$

В качестве третьего примера приведем нелинейные модели вида

$$x_3(t) = \frac{1}{\sum_{i=1}^{m} \theta_i u^i(t)}; \quad x_4(t) = \frac{1}{\sum_{i=1}^{m} \theta_i u_i(t)}.$$

Эти модели преобразуются в линейные путем замены переменной

$$z_i(t) = \frac{1}{x_i(t)}, \qquad i = 3,4;$$

$$z_3(t) = \sum_{i=1}^{m} \theta_i u^i(t); \quad z_4(t) = \sum_{i=1}^{m} \theta_i u_i(t).$$

Указанные и аналогичные преобразования наиболее эффективны при параметрической идентификации, когда априори известно уравнение нелинейной модели (с точностью до параметров).

Дальнейшее применение методов идентификации к преобразованным линейным моделям (2.7), (2.9), (2.10), (2.11), (2.12) позволяют получать линейные оценки параметров $\theta^T = (\theta_0, \theta_1, ..., \theta_k)$.

Обобщая выше рассмотренные исследования представим линейные модели изучаемых объектов теперь в виде

$$x(t) = \theta^T u(t) + \nu(t), \qquad (2.17)$$

где $x(t)$ – выходная переменная; $u(t)$ – вектор наблюдаемых входных переменных; $\theta$ – вектор неизвестных параметров; $\nu(t)$ – помеха; $t$ – момент времени наблюдения.

Рассмотренный подход можно применить и при структурной идентификации, когда известен только класс объектов (степенные, экспоненциальные или другие модели объекта). Задача идентификации решается при этом методе прямого подбора типа и структуры модели.

## 2.2. Формы записи дифференциальных уравнений

Объекты, для которых свойственно иметь запас энергии, называются динамическими. Наиболее употребительной моделью динамических объектов являются дифференциальные уравнения. Порядок системы дифференциальных уравнений, описывающей модель объекта, непосредственно не определяется количеством входов и выходов, а зависит от операторов, преобразующих входные сигналы в выходные.

Дифференциальное уравнение модели одномерного объекта имеет вид

$$\varphi(t, x, x', ..., x^{(n)}, u, u', ..., u^{(n)}) = 0. \qquad (2.18)$$

Это дифференциальное уравнение n-го порядка относительно переменной *x(t)*. Обычно его можно разрешить относительно старшей производной $x^{(n)}(t)$

$$x^{(n)}(t) = f(t, x, x', ..., x^{(n-1)}, u, u', ..., u^{(m)}). \qquad (2.19)$$

Запасенную объектами энергию определяют начальными условиями

$$x^{(i)}(t_0) = x_0^{(i)}, \qquad i = \overline{0, n-1}, \qquad (2.20)$$

которым в момент $t_0$ должны удовлетворять переменная *x(t)* и ее производные $x'(t), ..., x^{(n-1)}(t)$.

Для одномерного линейного динамического объекта уравнение (2.19) имеет вид обыкновенного дифференциального уравнения

$$x^{(n)}(t) + \sum_{i=0}^{n-1} a_i(t) x^{(i)}(t) = \sum_{j=0}^{m} b_i(t) u^{(j)}(t). \qquad (2.21)$$

### 2.2.1. Запись дифференциальных уравнений в нормальной форме Коши

Для решения задачи идентификации с использованием ЭВМ модель (2.18) целесообразно представить в виде нормальной системы. Нормальной системой или системой в нормальной форме Коши называют систему дифференциальных уравнений первого порядка, разрешенных относительно производных. Для получения нормальной системы введем обозначения

$$x = x_1, \ \frac{dx}{dt} = x_2, ..., \ \frac{dx_{n-1}}{dt} = x_n.$$

В результате получим искомую систему

$$\begin{aligned} &\frac{dx_1}{dt} = x_2, \\ &\cdots\cdots\cdots\cdots, \\ &\frac{dx_{n-1}}{dt} = x_n, \\ &\frac{dx_n}{dt} = f(t, x_1, ..., x_n) \end{aligned} \qquad (2.22)$$

с начальными условиями

$$x_i(t_0) = x_{i0}, \ i = \overline{1, n}. \qquad (2.23)$$

Для общности систему (2.22) можно записать в виде

$$\frac{dx}{dt} = f_i(t, x_1, ..., x_n), \qquad i = \overline{1, n}, \tag{2.24}$$

или в векторной форме

$$\frac{dx}{dt} = f_i(t, x), \qquad x(t_0) = x_0, \tag{2.25}$$

где $x^T = [x_1, x_2, ..., x_n]$; $F^T = [f_1, f_2, .., f_n]$.

Модель многомерного динамического объекта описывается $N$ уравнением вида

$$\begin{aligned} &\varphi_j(t, x_1, x_1', ..., x_1^{(n_1)}, ..., x_N, x_N', ..., x^{(n_N)}, \\ &u_1, u_1', ..., u_1^{(m_1)}, ..., u_M', ..., u_M^{(n_M)}) = 0, \\ &x_j^{(i)}(t_0) = x_{j0}^{(i)}, \; i = \overline{0, n_{j-1}}, \; j = \overline{1, N}. \end{aligned} \tag{2.26}$$

Из теоремы о существовании неявных функций следует, что уравнение (2.26) можно разрешить относительно старших производных

$$x_j^{(n_j)} = f(t, x_1, x_1', ..., x_1^{(n_{1-1})}, ..., x_N^{(n_{N-1})}, \; j = \overline{1, N}. \tag{2.27}$$

Введем новые переменные

$$x_1' = x_2, ..., \; x_1^{(n_1+1)} = x_{n_1}, \; x_2 = x_{n_1+1}, ..., x_2^{(n_2-1)} = x_{n_1-n_2}, ..., x_N^{(n_N-1)} = x_l,$$

$$l = \sum_{j=1}^{N} n_j.$$

Тогда уравнение (2.27) в нормальной форме записывается в виде следующего векторного уравнения:

$$\frac{dx}{dt} = f_i(t, x), \qquad x(t_0) = x_0. \tag{2.28}$$

Для одномерного линейного динамического объекта, описываемого дифференциальным уравнением (2.21), нормальная система будет иметь вид

$$\begin{aligned} &\frac{dx_1}{dt} = x_2, \\ &\text{..... ....}, \\ &\frac{dx_{n-1}}{dt} = x_n, \\ &\frac{dx_n}{dt} = -\sum_{i=1}^{n} a_{i-1}(t)\, x_i + \sum_{j=0}^{m} b_j(t)\, u_{j+1}. \end{aligned} \tag{2.29}$$

Для общности эту систему можно записать

$$\frac{dx_i}{dt} = -\sum_{j=1}^{n} a_{i,j-1}(t)\, x_j + \sum_{j=0}^{m} b_{ij}(t)\, u_{j+1}, \qquad (2.30)$$

где $u_{j+1} = u^{(j)}, i = \overline{1,n}$ или в векторно-матричной форме

$$\frac{dx}{dt} = Ax + Bu\ , \qquad (2.31)$$

где

$$A = [-a_{ij}]_n^n, B = [b_{ij}]_{m+1}^n.$$

Для многомерного линейного объекта (2.27) дифференциальное уравнение принимает вид

$$\frac{dx}{dt} = Ax + Bu\ , \qquad x(t_0) = x_0, \qquad (2.32)$$

где

$$A = [\, a_{ij}\,]\ {}_{\ell}^{\ell}, \quad B = [\, b_{ij}\,]\ {}_{s}^{\ell}, \quad \ell = \sum_{j=1}^{N} n_j, \quad s = \sum_{j=1}^{M} (m_i + 1)\ .$$

Элементы векторов x или $x_i$ во всех предыдущих уравнениях называют переменными состояния объекта или фазовыми координатами, а сам вектор *x*-вектором состояния. Множество этих векторов составляет пространство состояний (фазовое пространство). Элементы векторов *u* или $u_i$ в зависимости от их природы называются переменными управления или возмущениями.

Во многих физических объектах измеряются и передаются по соответствующим каналам не сигналы вектора состояния *x*, а другие сигналы функции составляющих вектора фазовых координат. Обозначим измеряемые величины через $y_1(t),\ldots,y_s(t),$ причем обычно $s = n$. Тогда уравнение измерения (наблюдения) примет вид

$$y(t) = F[t, x(t)]. \qquad (2.33)$$

Для линейного объекта это соотношение линейное

$$y(t) = C(t)\ x(t). \qquad (2.34)$$

Матрица *C(t)* называемся матрицей измерения или наблюдения. При измерениях, описываемых выражениями (2.33) и (2.34), вектором выходных сигналов (или просто вектором выхода) является вектор $y(t)$. Важно помнить, что между понятиями вектора входа, вектора выхода и вектора состояния существует принципиальное различие. Все составляющие вектора входа и выхода являются вполне конкретными физическими вели-

чинами (токами, напряжениями, углами и т. д.). Элементы вектора состояния могут рассматриваться как некоторые абстрактные переменные, физическая природа которых при исследованиях не всегда существенна.

Векторно-матричная запись модели линейного динамического объекта с учетом уравнения наблюдения принимает вид

$$\frac{dx}{dt} = A(t)x + B(t)u,$$
$$y(t) = C(t)\ x(t)\ , \qquad x(t_0) = x_0\ . \tag{2.35}$$

Если матрицы *A(t), B(t)* и *C(t)* постоянны, то объект называется объектом с постоянными коэффициентами или стационарным объектом. Уравнения такого объекта имеют вид

$$\frac{dx}{dt} = Ax\ +\ Bu\ ,$$
$$y(t) = Cx(t)\ . \tag{2.36}$$

При наличии погрешностей при измерении и возмущений, выходные (регулируемые) сигналы задаются линеаризованным матричным уравнением:

$$y(t) = C(t)x(t) + v(t), \tag{2.37}$$

где *y(t)* – вектор регулируемых (измеряемых) величин; *C(t)* – матрица связи вектора измерений с вектором состояний; *v(t)* – вектор ошибок измерений (помехи).

Структура линейной непрерывной системе реализующая уравнения (2.35) и (2.37) приведена на рис. 2.2

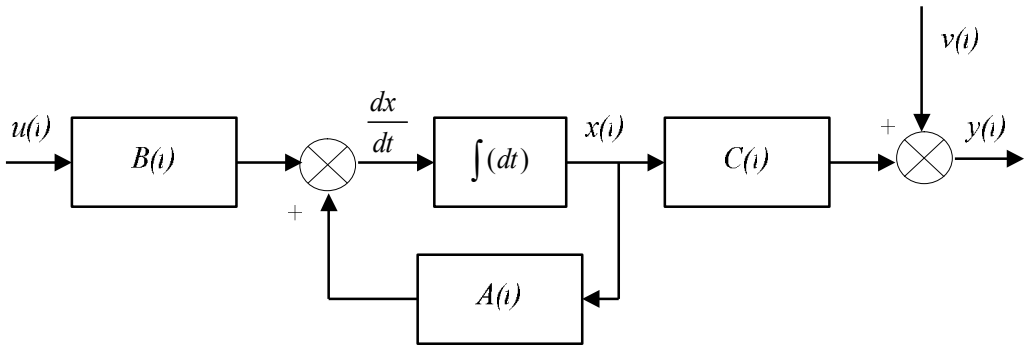



Рис. 2.2. Структура математической модели определяемая уравнениями (2.35), (2.37)

Данная структура соответствует математической модели объекта построенной в пространстве состояний его входных *u(t),* выходных *y(t)* и внутренних, или фазовых координат *x(t).*

### 2.2.3. Канонические представления системы

Пусть в векторном пространстве (пространстве состояний) $R$ задан базис определенный на координатах пространства состояний $x_1, x_2, .....x_k$. Из линейной алгебры [1, 20, 23, 26] известно, что этот базис может быть получен из другого базиса с помощью линейного преобразования $x_i = \sum_{j=1}^{k} t_{ii} y_i \;\; i = 1,2,...k$, или в матричной форме

$$x = Ty, \tag{2.38}$$

где $$T = \begin{pmatrix} t_{11} & t_{12} & \dots & t_{1k} \\ t_{21} & t_{22} & \dots & t_{2k} \\ \dots & \dots & \dots & \dots \\ t_{k1} & t_{2k} & \dots & t_{kk} \end{pmatrix}.$$

Можно, наоборот, выразить вектор $y$ через вектор $x$

$$y = T^{-1}x. \tag{2.39}$$

Уравнения (2.38) и (2.39) являются уравнениями замены базиса в пространстве состояний $R$ и, по сути, представляют уравнения перехода от одной системы координат к другой. Очевидно, что можно выбрать бесконечно большое число базисов или систем координат в пространстве состояний. При переходе к новому базису необходимо и достаточно, чтобы матрица перехода $T$ была не вырожденной, что выполняется, если определитель этой матрицы не равен нулю $T \neq 0$. Следовательно, между множеством координатных преобразований и множеством матриц $T$ существует взаимно однозначное соответствие при фиксированных базисах, соответствующих этим преобразованиям.

Используя линейные преобразования (2.38) можно отображать в пространство состояний системы и другие ее пространства (пространства управлений, возмущений и регулируемых координат), что и задается уравнениями связи (2.35).

При решении таких уравнений требуется вычисление свободных и вынужденных движений системы, что в конечном итоге приводит к необходимости решения системы однородных алгебраических уравнений вида:

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n = \lambda x_1, \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n = \lambda x_2, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_{n1}x_1 + a_{n2}x_2 + \ldots + a_{nn}x_n = \lambda x \end{cases} \tag{2.40}$$

или в векторной форме

$$Ax = \lambda x. \tag{2.41}$$

Такая система уравнений получается при подстановке в дифференциальные уравнения системы (2.35) какого-нибудь их частного решения. К такой системе сводится вычисление установившихся режимов работы системы (2.35). По сути уравнения (2.40) отображают базис $x_1, x_2, \ldots x_k$ сам в себя и поэтому характеризуют свойства такого отображения, или свойства матрицы $A$, соответствующей этому отображению.

В линейной алгебре эта задача известна как задача вычисления собственных значений и собственных векторов матрицы, задающей линейное преобразование.

Значения параметра $\lambda$, для которого существуют нетривиальные решения (2.40), называются собственными значениями матрицы $A$. Соответствующие им векторные решения (2.41) называют собственными векторами матрицы $A$.

Столбец, составленный из элементов собственного вектора, для конкретного значения $\lambda_i$, называют модальным столбцом.

Перенеся правую часть уравнения (2.41) влево, получим

$$(A - \lambda I)x = 0, \tag{2.42}$$

где I – единичная матрица.

Это уравнение обладает нетривиальным решением только тогда, когда его определитель равен нулю

$$(A - \lambda I) = \begin{vmatrix} a_{11} - \lambda & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} - \lambda & \cdots & a_{2n} \\ \ldots & \ldots & \cdots & \ldots \\ a_{n1} & a_{n2} & \cdots & a_{mn} - \lambda \end{vmatrix} = 0. \tag{2.43}$$

Разложение этого определителя дает характеристическое уравнение системы или матрицы

$$(\lambda)^n + b_{n-1}(\lambda)^{n-1} + \ldots + b_1(-\lambda) + b_0 = 0, \tag{2.44}$$

из которого могут быть найдены все значения λ. Если теперь подста-

вить найденные значения $\lambda_i$ в уравнение (2.15) и решить его, то вычисленные значения составляющих вектора $x$ для каждого значения $\lambda_i$ будут собственными векторами матрицы $A$.

Собственные векторы матрицы имеющей действительные и различные собственные значения обладают следующими важными свойствами:

1. Собственные векторы такой матрицы попарно ортогональны;

2. Собственные векторы матрицы $n$-го порядка порождают $n$-мерное векторное пространство;

3. Собственные векторы матрицы $n$-го порядка образуют ортогональный базис $n$-мерного векторного пространства.

Знание собственных значений и векторов матрицы системы позволяет осуществлять линейные преобразования (2.35) в пространстве состояний, придавая различный вид этой матрице. В задачах управления наиболее часто используются следующие канонические виды матрицы А [26]:

1. Диагональная каноническая форма,

$$\text{А}\partial = \begin{pmatrix} \lambda_{11} & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \lambda_n \end{pmatrix}, \qquad (2.45)$$

где $\lambda_i$ – собственные различные значения матрицы А.

Диагональной канонической форме соответствует следующая структурная схема рис. 2.3.

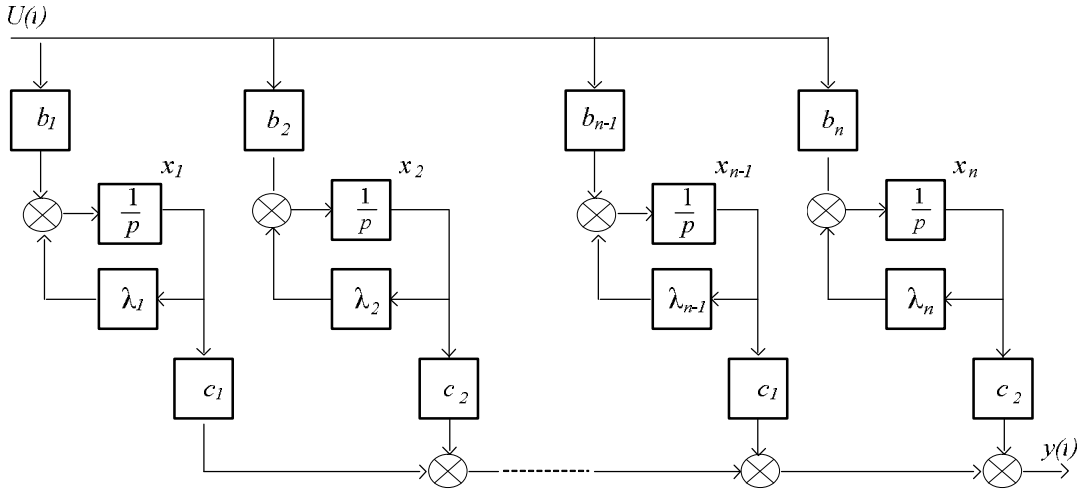



Рис. 2.3. Структурная схема диагональной канонической формы

2. Каноническая форма управляемости

$$A_y = \begin{pmatrix} 0 & 1 & \dots 0 \\ 0 & 0 & \dots 0 \\ \dots & \dots & \dots \dots \\ -d_0 & -d_1 & \cdots -d_{n-1} \end{pmatrix}, \qquad (2.46)$$

где $d_i$ – коэффициенты характеристического уравнения матрицы (2.44).

Структурная схема для канонической формы управляемости показана на рис. 2.4.

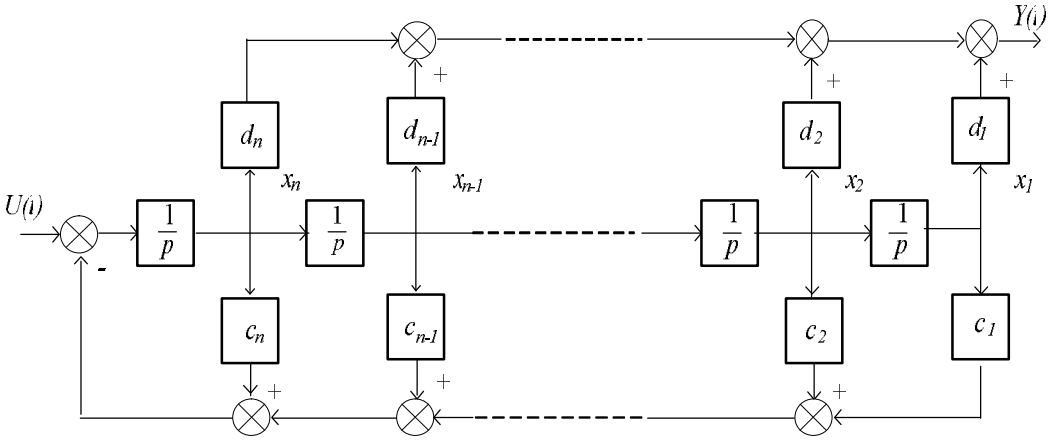



Рис. 2.4. Структурная схема канонической формы управляемости

2. Каноническая форма наблюдаемости

$$A_н = \begin{pmatrix} 0 & 0 & \dots & -b_0 \\ 1 & 0 & \dots & -b_1 \\ \dots & \dots & \cdots & \dots \\ 0 & \dots & 1 & -b_{n-1} \end{pmatrix}. \qquad (2.47)$$

Этой канонической форме соответствует структура рис. 2.4.

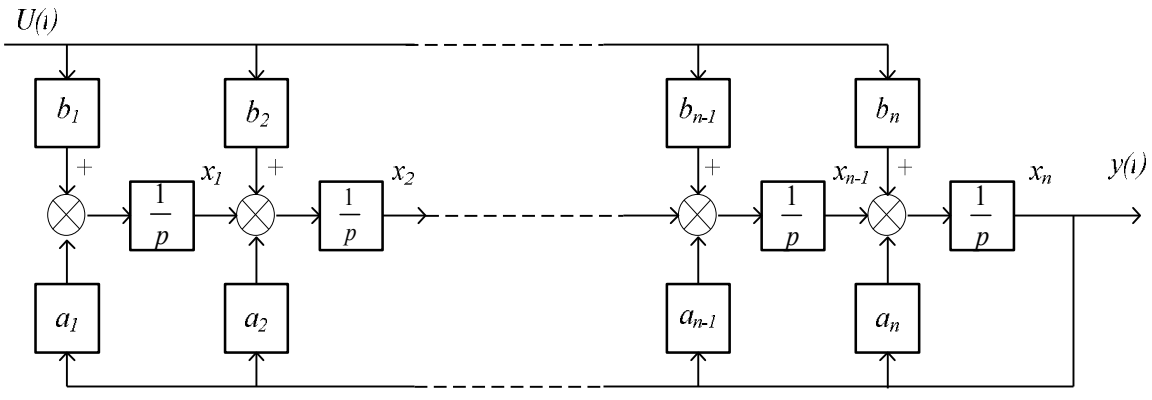



Рис. 2.5. Структурная схема канонической формы наблюдаемости

Наибольший интерес представляется получение диагональной канонической формы, т. к. все остальные формы легко получаются из нее путем линейных преобразований. Для диагонализации матрицы $A$ используется следующее выражение:

$$A\,\partial = \begin{pmatrix} \lambda_{11} & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \lambda_n \end{pmatrix} = T^{-1} AT, \qquad (2.48)$$

где $\lambda_1, \lambda_2, \dots \lambda_n$ – собственные значения матрицы $A$, а ее нормированные собственные векторы являются столбцами матрицы $T$. Условием нормирования является равенство единице модуля собственных векторов.

Собственные векторы и собственные значения матрицы $A$ однозначно определяют динамические свойства системы, задаваемые уравнением (2.36).

**Пример 2.1.** Рассмотрим приведение многомерного объекта, задаваемого уравнением

$$\begin{aligned} \frac{dx_1}{dt} &= a_{11}x_1 + a_{12}x_2 + b_{11}u_1, \\ \frac{dx_2}{dt} &= a_{21}x_1 + a_{22}x_2 + b_{22}u_2, \\ y &= c_1x_1 + c_2x_2 \end{aligned} \qquad (2.49)$$

к канонической диагональной форме.

Матрицы объекта равны

$$\mathrm{A} = \begin{pmatrix} -0{,}5 & 3 \\ -0{,}25 & -2{,}5 \end{pmatrix}; \ \mathrm{B} = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}; \ \mathrm{C} = \begin{pmatrix} 3 & 2 \end{pmatrix}.$$

Вычислим собственные значения матрицы $A$ из условия (2.43)

$$\begin{vmatrix} -0{,}5-\lambda & 3 \\ -0{,}25 & -2{,}5-\lambda \end{vmatrix} = 0. \qquad (2.50)$$

Раскрывая полученный определитель, получим характеристическое уравнение объекта

$$\lambda^2 + 3\lambda + 2 = 0. \qquad (2.51)$$

Корни этого уравнения или собственные значения матрицы $A$ равны $\lambda_1 = -1;\ \lambda_2 = -2$.

Найдем собственные векторы матрицы, подставляя в (2.50) вычисленные собственные значения. Для первого собственного значения имеем:

$$(-0,5 - \lambda_1)\, t_{11} + 3t_{21} = 0, \qquad (2.52)$$
$$-0,25t_{11} + (-2,5 - \lambda_1)\, t_{21} = 0.$$

При $\lambda_1 = -1$ получаем следующую систему уравнений для вычисления первого собственного вектора

$$0,5t_{11} + 3t_{21} = 0,$$
$$-0,25t_{11} - 1,5t_{21} = 0.$$

Откуда $t_{11} = -6t_{21}$.

Конкретные значения первого собственного вектора определяются условием нормировки

$$\sqrt{t_{11}^2 + t_{21}^2} = 1.$$

Подставляя сюда решение уравнений $t_{11} = -6t_{21}$, получим

$$t_{11} = \frac{6}{\sqrt{37}}; t_{21} = -\frac{1}{\sqrt{37}}.$$

Аналогично найдем и второй собственный вектор

$$t_{12} = -\frac{2}{\sqrt{5}}; t_{21} = \frac{1}{\sqrt{5}}.$$

Зная собственные векторы можно записать выражение для матрицы T, задающей переход в новую систему координат

$$\boldsymbol{T} = \begin{pmatrix} \frac{6}{\sqrt{37}} & -\frac{2}{\sqrt{5}} \\ -\frac{1}{\sqrt{37}} & \frac{1}{\sqrt{5}} \end{pmatrix}.$$

Найдем обратную матрицу

$$\boldsymbol{T}^{1} = \begin{pmatrix} \frac{\sqrt{37}}{4} & \frac{\sqrt{37}}{2} \\ \frac{\sqrt{5}}{4} & \frac{3\sqrt{5}}{2} \end{pmatrix}.$$

Не трудно убедиться, что

$$\boldsymbol{T} \times \boldsymbol{T}^{1} = I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Новый вектор координат q задается линейным преобразованием

$$x = Tq.$$

Подставляя его в уравнения объекта (2.32), получим

$$\boldsymbol{T}\frac{dq}{dt} = ATq + Bu.$$

Умножая обе части уравнения на $T^{-1}$ слева будем иметь

$$\boldsymbol{I}\frac{dq}{dt} = T^{-1}ATq + T^{-1}Bu.$$

Матрица $Aд = T^{-1}AT$, в соответствии с (2.48), будет иметь диагональный вид, где в главной диагонали стоят ее собственные значения. Действительно

$$Aд = \begin{pmatrix} \frac{\sqrt{37}}{4} & \frac{\sqrt{37}}{2} \\ \frac{\sqrt{5}}{4} & \frac{3\sqrt{5}}{2} \end{pmatrix} \times \begin{pmatrix} -0{,}5 & 3 \\ -0{,}25 & -2{,}5 \end{pmatrix} \times \begin{pmatrix} \frac{6}{\sqrt{37}} & -\frac{2}{\sqrt{5}} \\ -\frac{1}{\sqrt{37}} & \frac{1}{\sqrt{5}} \end{pmatrix} = \begin{pmatrix} -1 & 0 \\ 0 & -2 \end{pmatrix}.$$

Вычислим новую матрицу управления

$$Bд = \begin{pmatrix} \frac{\sqrt{37}}{4} & \frac{\sqrt{37}}{2} \\ \frac{\sqrt{5}}{4} & \frac{3\sqrt{5}}{2} \end{pmatrix} \times \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} \frac{\sqrt{37}}{4} & 37 \\ -\frac{\sqrt{5}}{4} & 3\sqrt{5} \end{pmatrix}$$

и матрицу измерения

$$Cд = \begin{pmatrix} \frac{6}{\sqrt{37}} & -\frac{2}{\sqrt{5}} \\ -\frac{1}{\sqrt{37}} & \frac{1}{\sqrt{5}} \end{pmatrix} = \begin{pmatrix} \frac{16}{\sqrt{37}} & -\frac{4}{\sqrt{5}} \end{pmatrix}.$$

Тогда в новых координатах $q_1$, $q_2$ уравнения объекта примут вид

$$\frac{dq_1}{dt} = \lambda_1 q_1 + b_{d11}u_1 + b_{d12}u_2,$$

$$\frac{dq_2}{dt} = \lambda_2 q_2 + b_{d21}u_1 + b_{d22}u_2,$$

$$y = c_{d1}q_1 + c_{d2}q_2.$$

### 2.2.4. Символическая форма записи линейного дифференциального уравнения

Пусть имеем исходное дифференциальное уравнение

$$\sum_{i=0}^{n} a_i x^{(i)}(t) = \sum_{j=0}^{m} b_j u^{(j)}(t). \qquad (2.53)$$

Уравнение (2.53) можем записать в следующем виде:

$$\sum_{i=0}^{n} a_i p^i x = \sum_{j=0}^{m} b_j p^j u \,, \qquad (2.54)$$

где $p = \frac{d}{dt}, \quad p^i = \frac{d^i}{dt^i}$.

При записи и преобразовании дифференциального уравнения оператор р можно рассматривать как алгебраический сомножитель, а выражение $p^i x \,(i = \overline{0,n})$ или $p^j u \,(j = \overline{0,m})$ как произведение, не обладающее свойством коммутативности. Учитывая это замечание, перепишем уравнение (2.54), вынеся *x* и *u* за скобки

$$(\sum_{i=0}^{n} a_i p^i)\, x = (\sum_{j=0}^{m} b_j p^j)\, u. \qquad (2.55)$$

Введем обозначения

$$A(t) = a_n p^n + a_{n-1} p^{n-1} + \ldots + a_0,$$

$$B(t) = b_m p^m + b_{m-1} p^{m-1} + \ldots + b_0,$$

с помощью которых уравнение (2.55) запишем в более компактной форме

$$A(p)\, x = B(p)\, u. \qquad (2.56)$$

В уравнении (2.56) дифференциальный оператор *A(p)* при выходной величине называют собственным оператором, а оператор *B(p)* при входной величине – оператором воздействия.

В случае многомерного объекта, имеющего *M* входов и *N* выходов можно составить систему из *N* уравнений в символической (операторной) форме. В общем виде эта система выглядит так

$$A_{11}(p)x_1 + \ldots + A_{1N}(p)x_N = B_{11}(p)u_1 + \ldots + B_{1M}(p)u_M \,,$$

$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

$$A_{N1}(p)x_1 + \ldots + A_{NN}(p)x_N = B_{N1}(p)u_1 + \ldots + B_{NM}(p)u_M$$

или, в более компактной форме

$$\sum_{j=1}^{N} A_{ij}(p)\, x_j = \sum_{j=1}^{M} B_{ij}(p)\, u_j, \quad i = \overline{1,N}.$$

Здесь $A_{ij}(p),\ B_{ij}(p)$ обозначают стационарные линейные операторы, т. е. полиномы от оператора дифференцирования с постоян-

ными коэффициентами. В матричной форме эта система записывается следующим образом:

$$A(p)\,x = B(p)u, \qquad (2.57)$$

где

$$A(p) = \begin{bmatrix} A_{11}(p)....A_{N1}(p) \\ .... \quad .... \quad .... \\ A_{N1}(p)....A_{NN}(p) \end{bmatrix}, \qquad B(p) = \begin{bmatrix} B_{11}(p)....B_{M1}(p) \\ .... \quad .... \quad .... \\ B_{N1}(p)....B_{NM}(p) \end{bmatrix},$$

$$x^T = [x_1,...x_N], \quad u^T = [u_1,...u_M].$$

### 2.2.5. Запись линейного дифференциального уравнения в форме изображений Лапласа

Если исходное дифференциальное уравнение (2.53) имеет постоянные коэффициенты и $a_i(i=\overline{0,n})$ и $b_j(j=\overline{0,m})$ и начальные условия нулевые, то его можно видоизменить с помощью преобразования Лапласа [2, 10, 11, 17].

Преобразованием Лапласа называют соотношение

$$F(s) = \int_0^\infty f(t)\,e^{-st}dt,$$

ставящее в соответствие функции *f(t)* вещественного переменного функцию *F(s)* комплексного переменного $s\,(s=\sigma+j\omega)$. При этом функцию *f(t)* называют оригиналом, а функцию *F(s)* – изображением или изображением по Лапласу. Связь между оригиналом и изображением условно записывается так

$$F(t) \div F(s) \text{ или } F(s) \div F(t).$$

Иногда пользуются символической записью $f(s) = L\{f(t)\}$, где *L* – оператор Лапласа.

Предполагается, что функция *f(t)*, которая подвергается преобразованию Лапласа, обладает следующими свойствами:

1. *f(t)* определена и кусочно-дифференцируема на всей положительной полуоси $[0,\infty]$;

2. *f(t)=0* при *t<0;*

3. существуют такие положительные числа $M$ и $C$, что $|f(t)| \le Me^{ct}$ при $0 \le t \le \infty$.

Функции, обладающие указанными тремя свойствами, часто называют функциями оригиналами.

Рассмотренное преобразование Лапласа иногда называют односторонним преобразованием Лапласа в отличие от двухстороннего преобразования

$$F(s) = \int_0^\infty f(t)\, e^{-st} dt,$$

которое имеет место, когда $f(t) \ne 0, \quad t < 0$.

Соотношение

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

определяющее по известному изображению его оригинал (в точках непрерывности последнего), называют обратным преобразованием Лапласа. В нем интеграл берется вдоль любой прямой $Re\, s = \sigma_0 > c$. Символически обратное преобразование Лапласа записывают так

$$f(t) = L^{-1}\{f(s)\},$$

где $L^{-1}$ – обратный оператор Лапласа.

Применим преобразование Лапласа к обеим частям уравнения (2.53)

$$L\{a_n x^{(n)}(t) + \ldots + a_0 x(t)\} = L\{b_m x^{(m)}(t) + \ldots + b_0 u(t)\}.$$

Используя свойства линейности и дифференцирования оригинала при нулевых начальных условиях, получим

$$(a_n s^n + a_{n-1} s^{n-1} \ldots + a_0)\, x(s) = (b_m s^m + b_{m-1} s^{m-1} + \ldots + b_0)\, u(s), \quad (2.58)$$

где

$$u(s) = L\{u(t)\}, \qquad x(s) = L\{x(t)\}.$$

Тогда вместо уравнения (2.58) получим

$$A(s)\, x(s) = B(s)\, u(s), \quad (2.59)$$

где

$$A(s) = a_n s^n + a_{n-1} s^{n-1} + \ldots + a_0,$$

$$B(s) = b_m s^m + b_{m-1} s^{m-1} + \ldots + b_0.$$

Это уравнение как уравнение (2.58), адекватно исходному дифференциальному уравнению (2.53) только при нулевых начальных условиях. Если начальные условия не нулевые, то уравнениями (2.58) и (2.59) как математическим описанием объекта пользоваться нельзя.

В случае многомерного объекта можно также составить систему уравнений в изображениях Лапласа. По аналогии с выражением (2.57) эта система в матричной форме при нулевых начальных условиях записывается так $A(s)\,x(s) = B(s)\,u(s)$, где

$$A(s) = \begin{bmatrix} A_{11}(s) \ldots . A_{N1}(s) \\ \ldots . \quad \ldots . \quad \ldots . \\ A_{N1}(s) \ldots . A_{NN}(s) \end{bmatrix}, \qquad B(s) = \begin{bmatrix} B_{11}(s) \ldots . B_{M1}(s) \\ \ldots . \quad \ldots . \quad \ldots . \\ B_{N1}(s) \ldots . B_{NM}(s) \end{bmatrix},$$

$$x^T(s) = \left[x_1(s), \ldots x_N(s)\right], \qquad u^T(s) = \left[u_1(s), \ldots u_M(s)\right].$$

## 2.3. Передаточные функции и передаточные матрицы

Для описания одномерного линейного динамического объекта широко используется другой вид параметрической модели, так называемая передаточная функция (ПФ) в следующих формах:

- ПФ в операторной форме;
- ПФ в форме изображений Лапласа.

Отношение оператора воздействия к собственному оператору называют ПФ в операторной форме, т. е.

$$W(p) = \frac{B(p)}{A(p)} = \frac{b_m p^m + b_{m-1} p^{m-1} + \ldots + b_0}{a_n p^n + a_{n-1} p^{n-1} + \ldots + a_0}. \qquad (2.60)$$

С помощью ПФ уравнение (2.53) можем записать так

$$x = W(p)\,u. \qquad (2.61)$$

Уравнения (2.55), (2.56) и (2.61) называют уравнениями в символической или операторной форме записи.

Используя запись дифференциального уравнения в символической форме (2.55) или (2.56) и рассматривая формально собственный оператор *A(p)* и оператор воздействия *B(p)* как обычные алгебраические сомножители, ПФ в операторной форме можем определить как отношение выходной величины к входной.

Наряду с ПФ в операторной форме широко используют ПФ в форме изображений Лапласа.

Передаточной функцией в форме изображений Лапласа называют отношение изображения выходной величины к изображению входной величины при нулевых начальных условиях

$$W(s) = \frac{x(s)}{u(s)} = \frac{b_m s^m + b_{m-1} s^{m-1} + \ldots + b_0}{a_n s^n + a_{n-1} s^{n-1} + \ldots + a_0}. \qquad (2.62)$$

Передаточную функцию в форме (2.62) можно получить из ПФ в форме (2.60), если в последней сделать подстановку $p = s$. В общем случае это следует из того, что дифференцированию оригинала – символическому умножению оригинала на $p$ – при нулевых начальных условиях соответствует умножение изображения на комплексное число $s$. Сходство между ПФ в форме (2.60) и в форме (2,62) чисто внешнее, и оно имеет место только в случае стационарных объектов.

Используя ПФ (2.62), уравнение (2.58) в изображениях Лапласа можем записать в более компактной форме

$$x(s) = W(s)\,u(s). \qquad (2.63)$$

Это уравнение также эквивалентно исходному дифференциальному уравнению (2.53) только при нулевых начальных условиях.

В случае многомерного объекта для его полного описания необходимо иметь NM передаточных функций. Эти передаточные функции записывают в виде

$$W(s) = \begin{bmatrix} W_{11}(s) \ldots . W_{1M}(s) \\ \ldots . \quad \ldots . \quad \ldots . \\ W_{N1}(s) \ldots . W_{NM}(s) \end{bmatrix}. \qquad (2.64)$$

Эту матрицу называют передаточной матрицей. Она дает полное описание многомерного объекта при нулевых начальных условиях. С ее помощью уравнение (2.57) многомерного объекта в изображениях Лапласа можем записать в следующем виде:

$$x(s) = W(s)\,u(s). \qquad (2.65)$$

Для вычисления передаточной матрицы используем следующий прием.

Умножим слева обе части матричного уравнения (2.57) $A(s)\,x(s) = B(s)\,u(s)$ на обратную матрицу $A^{-1}(s)$.

Получим $x(s) = A^{-1}(s)\,B(s)\,u(s)$.

Приравнивая правую часть полученного уравнения к правой части равносильного ему уравнения (2.65), получим

$$W(s) = A^{-1}(s)\ B(s). \qquad (2.66)$$

Передаточные функции формально можно рассматривать как новую форму записи дифференциальных уравнений. В соответствии с принятой классификацией обе эти модели относятся к параметрическим. К классу параметрических моделей относятся и так называемые регрессионные модели, в которых связь между входной и выходной величинами описывается следующим выражением:

$$x(u) = a_0 f_0(u) + a_1 f_1(u) + \ldots + a_n f_n(u).$$

Это выражение называют уравнением регрессий относительно искомых коэффициентов $a_i (i = \overline{0,n})$. В качестве функций могут быть использованы, например, члены разложений в ряд Фурье

$$x(u) = \frac{a_0}{2} + a_1 \sin u + a_2 \cos u + a_3 \sin 2u + a_4 \cos 2u + \ldots + a_{2n-1} \sin n\,u + a_{2n} \cos n\,u,$$

или члены разложения в ряд Тейлора

$$x(u) = a_0 + a_1\, u + \ldots + a_n u^n.$$

Искомые коэффициенты зависят от параметров идентифицируемого объекта, поэтому такую модель можно считать параметрической. Однако в отличие от дифференциальных уравнений и передаточных функций регрессионные модели используют обычно для идентификации объектов, структура и параметры которых неизвестны, т. е. при структурной идентификации.

## 2.4. Временные характеристики

В качестве непараметрических моделей могут быть названы переходная и импульсно-переходная функции и их графики – временные характеристики. Их используют при описании линейных объектов как стационарных, так и нестационарных.

Переходной функцией объекта называют функцию, описывающую изменение выходной величины объекта, когда на его вход подается единичное ступенчатое воздействие при нулевых начальных условиях. Единичное ступенчатое воздействие аналитически описывается единичной функцией, которую обозначают символом $\Delta(t)$ или $1(t)$, причем

$$\Delta(t) = 1(t) = \begin{cases} 1 & \text{при} \quad t \geq 0 \\ 0 & \text{при} \quad t < 0. \end{cases}.$$

Функция $\Delta(t) = 1(t)$ обладает производными любого порядка, интегрируема и может быть подвергнута преобразованию Лапласа, причем

$$L\{1(t)\} = \frac{1}{s}.$$

Обозначим переходную функцию через $h(t)$. Тогда она определится как решение уравнения

$$h^{(n)}(t) + \sum_{i=0}^{n-1} a_i h^{(i)}(t) = \sum_{j=0}^{m} b_j \Delta^{(j)}(t), \qquad (2.67)$$

$$h^{(i)}(t_0) = 0, \ (i = \overline{1, n-1}).$$

Импульсной переходной, или весовой функцией, называется функция, которая описывает реакцию объекта на единичное импульсное воздействие при нулевых начальных условиях. Единичное импульсное воздействие физически можно представить как очень узкий импульс, ограничивающий единичную площадь. Математически он описывается дельта – функцией $\delta(t)$, обладающей следующими свойствами:

$$\delta(t) = \begin{cases} \infty & \text{ïðè} \quad t = 0 \\ 0 & \text{ïðè} \quad t \neq 0 \end{cases},$$

$$\int_{-\infty}^{\infty} \delta(t)\, dt = \int_{-\varepsilon}^{\varepsilon} \delta(t)\, dt = 1,$$

где $\delta$ – любое положительное число.

С единичной функцией $1(t) = \Delta(t)$ дельта-функция $\delta(t)$ связана соотношением

$$\delta(t) = \frac{d\Delta(t)}{dt}.$$

Дельта-функция также интегрируется и обладает производными любого порядка, причем как от самой дельта-функции, так и от ее производных могут быть взяты преобразования Лапласа

$$L\{\delta(t)\} = 1, \qquad L\{\delta^{(1)}(t)\} = s, \ldots, \ L\{\delta^{(m)}(t)\} = s^m.$$

Импульсную переходную функцию (ИПФ) обозначим через $\omega(t)$. Ее можно определить как решение дифференциального уравнения

$$\omega^{(n)}(t)+\sum_{i=0}^{n-1}a_i\omega^{(i)}(t)=\sum_{j=0}^{m}b_j\delta^{(j)}(t), \qquad (2.68)$$

$$\omega^{(i)}(t_0)=0,\ (i=\overline{0,n-1}).$$

Установим, как связана весовая функция с передаточной функцией. Для этого рассмотрим линейное дифференциальное уравнение с постоянными коэффициентами в общем виде

$$\sum_{i=0}^{n}a_i x^{(i)}(t)=\sum_{j=0}^{m}b_j u^{(j)}(t).$$

В изображениях Лапласа это уравнение принимает уравнения (2.63)

$$x(s)=W(s)u(s),$$

где передаточная функция

$$W(s)=\frac{\sum_{j=0}^{m}b_j s^j}{\sum_{i=0}^{n}a_i s^i}.$$

Легко видеть, что уравнение (2.63) справедливо и в случае, когда $u(t)=\delta(t)$. В соответствии с определением ИПФ при $u(t)=\delta(t)$ переменная $x(t)=\omega(t)$. Поэтому вместо уравнения (2.63) можно записать

$$L\{\omega(t)\}=W(s)\,L\{\delta(t)\}.$$

Второй сомножитель правой части этого уравнения $L\{\delta(t)\}=1$, поэтому

$$W(s)=L\{\omega(t)\}, \qquad (2.69)$$

т. е. передаточная функция определяется как изображение Лапласа весовой (импульсной переходной) функции. Соответственно ИПФ получается из ПФ путем обратного преобразования Лапласа

$$\omega(t)=L^{-1}\{W(s)\}. \qquad (2.70)$$

Пусть теперь $u(t)=\delta(t)$, и тогда в соответствии с определением переходной функции $x(t)=h(t)$. В этом случае вместо уравнения (2.63) имеем

$$L\{h(t)\}=W(s)\ L\{1(t)\}.$$

Подставим в это уравнение вместо *W(s)* его значение из формулы (2.63), и, учитывая, что $L\{\Delta(t)\}=\frac{1}{s}$, получим

$$L\{h(t)\}= L\{\omega(t)\}\frac{1}{s},$$

откуда имеем

$$L\{\omega(t)\}= sL\{h(t)\}.$$

Так как при нулевых начальных условиях умножению изображения на *s* соответствует дифференцирование оригинала, то из последнего равенства получаем

$$\omega(t)=\frac{d\,h(t)}{d\,t}. \qquad (2.71)$$

Весовая и переходная функции, как и передаточная функция, являются исчерпывающими моделями линейного одномерного стационарного объекта при нулевых начальных условиях. По ним можно однозначно определить выходную величину (реакцию объекта) при произвольном входном воздействии. Действительно, исходя из уравнения (2.63) на основании теорем о свертке, получим

$$x(t)=\int_o^t \omega(t-\tau)\,u(\tau)d\,\tau. \qquad (2.72)$$

Эта формула, как и уравнение (2.63), справедлива только при нулевых начальных условиях.

Заметим, что весовая функция ω (t) удовлетворяет условию

$$\omega(t)=0 \text{ при } t=t_0=0,$$

которое принято называть условием физической осуществимости. Оно вытекает из того очевидного факта, что в любом реальном динамическом объекте переходный процесс $x(t)=\omega(t)$ не может возникнуть раньше момента приложения входного воздействия (предполагается, что входное воздействие прикладывается к объекту в момент $t_0=0$). В силу этого условия, в выражении (2.63) в качестве верхнего предела у интеграла можно принять ∞, т. е.

$$x(t)=\int_0^\infty \omega(t-\tau)\,u(\tau)d\,\tau.$$

Модели в виде переходных и весовых функций могут использоваться также для описания одномерных линейных динамических объ-

ектов с переменными параметрами, т. е. нестационарных объектов. Однако в силу особенностей нестационарных объектов эти модели более сложные. Так, вследствие того, что параметры нестационарного объекта являются функциями времени, при разных моментах приложения импульсного воздействия на такой объект его весовая функция будет различной. Пусть значение момента приложения импульса относительно момента начала отсчета $t_o$ равна τ. В этом случае функция зависит от аргументов t и τ является решением дифференциального уравнения

$$\omega^{(n)}(t)+\sum_{i=1}^{n-1}a_i(t)\omega^{(i)}(t,\tau)=\sum_{j=0}^{m}b_j(t)\delta^{(j)}(t-\tau),$$

$$\omega^{(i)}(t,\tau)\big|_{t=t_0}=0,(i=\overline{0,n-1}).$$

Условия физической осуществимости для функции $\omega(t,\tau)$ имеет вид $\omega(t,\tau)=0$ при $\tau<0$ и $\tau>t$. Для одномерного нестационарного линейного объекта вместо выражения (2.72) имеем

$$x(t)=\int_0^t \omega(t,\tau)\,u(\tau)\,d\tau\,.$$

Рассмотренные модели можно обобщить и на случай многомерных объектов. Так, по аналогии с передаточной матрицей вводится понятие весовой матрицы. Модель в виде весовой матрицы *W(t)* для многомерного стационарного линейного объекта с M входами и N выходами можно составить следующим образом. На *j*-й вход объекта подается воздействие в виде одиночного импульса $\delta(t-t_o)$ при всех остальных нулевых входах. При этом на всех выходах наблюдаются соответствующие переходные характеристики $\omega_{1j}(t),\omega_{2j}(t),\omega_{3j}(t),...,\omega_{Nj}(t)$. Эти функции составят *j*-й столбец весовой матрицы $(j=\overline{1,M})$. Возбуждая указанным образом последовательно все входы объекта, получим все элементы матрицы

$$W(t)=\begin{bmatrix}\omega_{11}(t)... & \omega_{1M}(t), \\ ... \quad ... & ... \\ \omega_{N1}(t)... & \omega_{NM}(t)\end{bmatrix}.$$

По весовой матрице нетрудно определить с помощью преобразования Лапласа передаточную матрицу

$$W(s) = \int_0^\infty W(t)\, e^{-st}\, dt .$$

Известно, что интеграл от матрицы равен матрице интегралов от ее элементов.

Весовая и передаточная матрицы являются исчерпывающими моделями многомерного объекта и позволяют определить его реакцию на произвольные воздействия при нулевых начальных условиях по формуле

$$x(t) = \int_0^\infty W(t-\tau)\, u(\tau)\, d\tau .$$

Условие физической осуществимости, обобщенное на случай многомерного объекта, принимает вид

$$W(t) = 0,\ t < 0 .$$

## 2.5. Дискретные модели

В настоящее время проблемы проектирования и широкое внедрение систем автоматизированного управления сложными техническими объектами обуславливает необходимость иметь математическое описание дискретных систем (объектов). Под дискретными системами понимают такие системы, для которых все или часть переменных состояния являются функциями, определенными только для последовательности дискретных значений независимой переменной [3, 10, 12, 26].

К таким системам можно отнести собственно дискретные системы, как ЦВМ, цифровые фильтры, системы релейной автоматики и т. д.

Дискретные системы также получаются при использовании ЦВМ для целей исследования или управления непрерывными технологическими процессами и объектами. В этом случае для их описания естественно использовать разностные уравнения.

### 2.5.1. Представление исходного дифференциального уравнения в разностной форме

Пусть дана функция *f(t)* и постоянное положительное число *h*, называемое шагом. Выражение

$$\Delta f(t) = f(t+h) - f(t) \qquad (2.73)$$

называется первой конечной разностью или конечной разностью первого порядка функции *f(t)* в точке *t*. Аналогично определяются разно-

сти высших порядков, например $\Delta^2 f(t) = \Delta[\Delta f(t)]$ . Дважды применяя формулу (2.73), получим

$$\Delta^2 f(t) = \Delta f(t+h) - \Delta f(t) = [f(t+2h) - f(t+h)] - [f(t+h) - f(t)] =$$
$$= f(t+h) - 2f(t+h) + f(t) .$$

Если определена конечная разность (n-l)-го порядка $\Delta^{n-1} f(t)$, то конечная разность *n*-го порядка есть $\Delta[\Delta^{n-1} f(t)]$. Методом полной математической индукции может быть получена

$$\Delta^n f(t) = \sum_{k=0}^{n} (-1)^{n-k} c_n^k f(t+kh) , \qquad (2.74)$$

где

$$c_n^k = \frac{n(n-1)...(n-k+1)}{1\,2...k} .$$

Формула (2.74) позволяет выразить конечную разность n-го порядка через значения функции *f(t)* в точках *t, t+h,...,t + n h*. Можно получить также формулу, в которой *f(t + nh)* выражается через конечные разности этой функции в точке *t*. Для этого примем, что $\Delta^0 f(t) = f(t)$, тогда $f(t+h) = \Delta^0 f(t) + \Delta^{-1} f(t)$. Методом полной математической индукции легко доказывается формула

$$f(t+nh) = \sum_{k=0}^{n} c_n^k \Delta^k f(t) . \qquad (2.75)$$

Линейным разностным уравнением называется уравнение вида

$$\sum_{i=0}^{n} a_i(t) \Delta^i(t) = \sum_{j=0}^{m} b_j(t) \Delta^j u(t) . \qquad (2.76)$$

Здесь x(t) – неизвестная, а $a_i(t), b_j(t)$ и $u(t)$ заданные функции $(i = \overline{0,n})$, $(j = \overline{0,m})$. Пользуясь формулой (2.74), уравнение (2.76) можем записать в виде

$$\sum_{i=0}^{n} c_i(t)\, x(t+ih) = \sum_{j=0}^{m} d_j(t)\, u(t+ih) . \qquad (2.77)$$

Если $c_n(t) \neq 0$ и $c_0(t) \neq 0$, то число n называется порядком уравнения.

В разностном уравнении вида (2.76) или (2.77) обычно без ограничения общности шаг *h* полагают равным единице, т. к. этого всегда можно добиться изменением масштаба отсчета по оси *t*. К тому же дискрет-

ный отсчет времени t заменяют номером k рассматриваемого отсчета *(k = 0, 1, 2,...),* а саму функцию *f(t)* задают дискретными значениями через равноотстоящие интервалы времени *h*, т. е. в моменты *0, h, 2h, 3,....* В промежутках между этими моментами времени значение функции равно нулю. Числа k являются целыми и положительными числами. При таком представлении рассматриваемую функцию обозначают через *f(kh),* или с учетом того, что *h –1,2,...* через *f(k).* Такую функцию обычно называют решетчатой. Решетчатая функция образуется из непрерывной функции путем придания аргументу последних значений, кратных *h*.

Тогда разность первого порядка для решетчатой функции *f(t)* описывают в виде

$$\Delta f(t) = f(t+1) - f(t). \qquad (2.78)$$

Разность первого порядка в теории решетчатых функций выполняет такую же роль, которую в теории непрерывных функции играет первая производная.

Разность *n*-го порядка будет

$$\Delta^n f(t) = \sum_{r=0}^{n} (-1)^r \frac{n!}{r!(n-r)!} f(t+n-r). \qquad (2.79)$$

С учетом введенных обозначений ранее записанные уравнения (2.76) и (2.77) принимают вид

$$\sum_{i=0}^{n} a_i \Delta^i x(t) = \sum_{j=0}^{m} b_j \Delta^j u(t), \qquad (2.80)$$

$$\sum_{i=0}^{n} c_i \Delta^i x(t-i) = \sum_{j=0}^{m} d_j \Delta^j u(t-j). \qquad (2.81)$$

Связь между коэффициентами $a_i$ и $b_j$ исходного дифференциального уравнения и коэффициентами $c_i$ *u* $d_j$ *(i =* $\overline{0,n}$, *j =*$\overline{0,m}$*)* разностного уравнения (2.81) устанавливается посредством формул (2.74) и (2.79).

Уравнение в конечных разностях в отличие от дифференциальных уравнений обладает двумя существенными особенностями:

1. Записанное в форме (2.81), оно дает возможность определять последующие значения отклика системы *x(t+n)* через значения отклика в предыдущие моменты времени, т. е. через значения *x(t+n-l), x(t+n-2),* вплоть до *x(t=0),* и через значения *u(t+m), u(t+m-l),..., u(t=0).* Аналогичную возможность дает и разностное уравнение, записанное в форме (2.77). Таким образом, разностное уравнение, записанное в форме (2.77) или в форме (2.81), представляет собой рекуррентную

формулу для последовательного определения откликов системы через отклики в предыдущие моменты времени.

2. Если $c_0 \equiv 0$, то уравнение (2.77) или (2.81) может быть сведено линейной заменой независимой переменной к уравнению более низкого порядка. Предположим, что в общем случае $c_0 = c_1 = c_2 = ... = c_{l-1} = 0,$ $a\ c_\ell,\ c_{\ell+1},...,c_n \neq 0.$ Тогда, сделав замену независимой переменной $t'=t+\ell,$ мы преобразуем уравнение (2.81) в уравнение такого же вида, но порядка $n-\ell$. В линейных дифференциальных уравнениях никакая линейная замена независимой переменной не может привести к понижению порядка уравнения. Это свойство может быть использовано для значительного сокращения вычислений при идентификации параметров объектов, описываемых линейными дифференциальными уравнениями высоких порядков. При переходе от дифференциального уравнения к его разностному аналогу следует всякий раз проанализировать исходное уравнение с целью определения возможности понижения порядка разностного уравнения за счет описанной выше линейной замены независимой переменной.

### 2.5.2. Дискретные модели стационарных систем

При анализе стохастических систем, встречающихся в самых различных областях науки и техники, исходными данными для анализа являются реализации случайного процесса генерируемого этой системой. Полученные в виде графиков, или осциллограмм, реализации случайного процесса обрабатываются и представляются в виде временного ряда. Временной ряд содержит ординаты реализации случайного процесса снятые в дискретные и равноотстоящие моменты времени. Следовательно, о свойствах исходной непрерывной системы судят по результатам цифровой обработки сигналов (временных рядов) формируемых системой. В связи с этим широкое распространение получили цифровые параметрические стохастические модели авторегрессии и скользящего среднего (АРСС-модели). Эти модели достаточно просты и включают обычно небольшое число параметров, которые необходимо оценивать по наблюдениям. АРСС-модели могут быть использованы как для изучения временных рядов, так и при определении статистических характеристик этих рядов. Широко используются такие модели в

управлении, экономике, медицине, геофизике, при обработке звуковых сигналов [7, 22, 26–28].

АРСС процессом порядка (*p, q*) называется ряд

$$x(k) = \sum_{i=1}^{p} c_i x(k-i) + \sum_{j=1}^{q} d_j e(k-j) + e(k), \qquad (2.82)$$

где *x(k)* – значения временного ряда в *k*-й момент времени; *e(k)* – последовательность независимых, одинаково распределенных случайных величин с нулевым математическим ожиданием и единичной дисперсией (белый шум); $\{c_i,\ i = 1,\ p\}$ – параметры авторегрессии; $\{d_j, j = 1, q\}$ – параметры скользящего среднего.

Частными случаями АРСС (p, q) процессов является процесс АР(p) – авторегрессии порядка *p*:

$$x(k) = \sum_{i=1}^{p} c_i x(k-i) + e(k), \qquad (2.83)$$

и процесс СС(q) – скользящего среднего порядка q:

$$x(k) = \sum_{j=1}^{q} d_j e(k-j) + e(k). \qquad (2.84)$$

Уравнения (2.82) и (2.83) описывают рекурсивные фильтры, а уравнение (2.84) –трансверсальный фильтр. Таким образом, процессы АРСС (p, q), АР(p) и СС(q) можно рассматривать как отклики соответствующих линейных фильтров на входной бело-шумный процесс *{e(k)}* с характеристиками

$$M\{e(j)e(i)\} = \sigma_e^{\,2}\delta_{ij}, \qquad M\{e(j)\} = 0,$$

где $\delta_{ij}$ – символ Кронекера; *M* – символ математического ожидания; $\sigma_e^2$ – дисперсия.

Следовательно, условиями стационарности этих процессов являются условия устойчивости соответствующих фильтров: рекурсивный фильтр устойчив, если все корни характеристического уравнения

$$q^p - c_1 q^{p-1} - \ldots - c_p = 0$$

находятся внутри окружности единичного радиуса [7,10,11]. Трансверсальный фильтр порядка *q* устойчив без ограничения на параметры.

Если в качестве стохастической системы рассматривается одномерный объект управления, то АРСС-модель объекта примет вид

$$x(k)=\sum_{i=1}^{p}a_i x(k-i)+\sum_{j=1}^{q}b_j u(k-j)+e(k), \qquad (2.85)$$

где *x(k), u(k)* выходная и входная координаты объекта.

Аналогично (2.51) АР-модель запишется как

$$x(k)=\sum_{i=1}^{p}a_i x(k-i)+bu(k)+e(k), \qquad (2.86)$$

$$x(k)=\sum_{j=1}^{p}b_i u(k-j)+e(k). \qquad (2.87)$$

Уравнения (2.85) – (2.87) являются линейными разностными уравнениями объекта управления, которые можно записать в символической форме, как:

АРСС – модель

$$A(z)x(z)=B(z)u(z)+e(z), \qquad (2.88)$$

АР – модель

$$A(z)x(z)=bu(z)+e(z), \qquad (2.89)$$

СС – модель

$$x(z)=B(z)u(z)+e(z), \qquad (2.90)$$

где x*(z), u(z)* и *e(z)* – *z* – изображения соответствующих сигналов;

$A(z)=1-\sum_{i=1}^{n}a_i z^{-i}$, $B(z)=1-\sum_{j=0}^{m}b_j z^{-j}$ – полиномы по оператору запаздывания.

Вводя дискретную передаточную функцию объекта, как отношение *z* – изображений сигнала на входе к сигналу на выходе при нулевых начальных условиях можно записать

$$W(z)=\frac{x(z)}{u(z)}=\frac{b_{01}+b_1 z^{-1}+\ldots+b_m z^{-m}}{1+a_1 z^{-1}+a_2^{-2}z^2+\ldots+a_n z^{-n}}=\frac{B(z)}{A(z)}. \qquad (2.91)$$

При наличии запаздывания в объекте равному целому число периодов дискретизации $\tau = d$ выражение для дискретной передаточной функции необходимо умножить на $z^{-d}$

$$W_d(z)=W(z)z^{-d}=\frac{B(z)}{A(z)}z^{-d}. \qquad (2.92)$$

Приводя помехи, действующие на объект управления к выходу, можно получить структурную схему объекта управления (рис. 2.6).

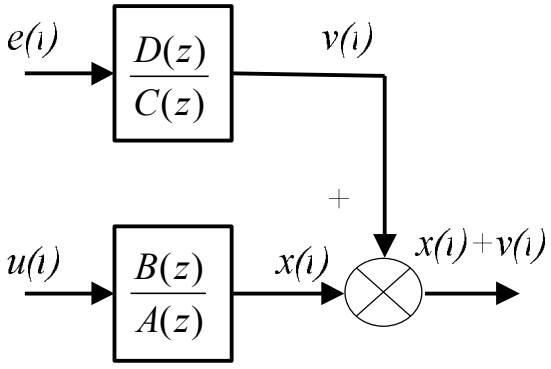



Рис. 2.6. Структурная система объекта управления

Для шума (по аналогии) передаточная функция будет иметь вид

$$W_x(z) = \frac{f(z)}{e(z)} = \frac{1 + d_1 z^{-1} + ... + d_s z^{-s}}{1 + c_1 z^{-1} + ... + c_r q^{-r}} = \frac{C(z)}{D(z)}, \qquad (2.93)$$

где *C(z),* и *D(z)* – полиномы по оператору запаздывания.

Объединив выражения (2.91) и (2.93), получим модель объекта с шумом измерений

$$x(z) = \frac{A(z)}{B(z)} z^{-d} u(z) + \frac{1}{A(z)} e(z). \qquad (2.94)$$

В зависимости от типа модели шума, при котором гарантируется сходимость оценок модели (2.94), используются модели частного вида:
– МП-модель (модель максимального правдоподобия)

$$x(z) = \frac{B(z)}{A(z)} z^{-d} u(z) + \frac{D(z)}{A(z)} e(z); \qquad (2.95)$$

– НК-модель (модель наименьших квадратов)

$$x(z) = \frac{B(z)}{A(z)} z^{-d} u(z) + \frac{1}{A(z)} e(z), \qquad (2.96)$$

$$x(t) = \frac{B(z^{-1})}{A(z^{-1})} u(t) + \frac{C(z^{-1})}{D(z^{-1})} e(t) \qquad (2.97)$$

или в виде

$$A^*(z^{-1})\, x(t) = B^*(z^{-1})\, u(t) + C^*(z^{-1})\, e(t), \qquad (2.98)$$

где

$$\begin{aligned} A^*(z^{-1}) &= A(z^{-1})\, D(z^{-1}), \\ B^*(z^{-1}) &= B(z^{-1})\, D(z^{-1}), \\ C^*(z^{-1}) &= C(z^{-1})\, A(z^{-1}). \end{aligned} \qquad (2.99)$$

Без потери общности авторегрессионный фильтр скользящего среднего $H(z^{-1})$ для белого шума $e(t)$ можно представить как $H(z^{-1}) = C(z^{-1}), A^{-1}(z^{-1})$. Тогда уравнение (2.97) также представимо в переобозначениях уравнений (2.100), (2.101)

$$A(z^{-1})\, x\,(\mathrm{t}) = \; B\,(z^{-1})\, u\,(\mathrm{t}) + \tilde{N}\,(z^{-1})\, e\,(\mathrm{t}), \qquad (2.100)$$

где $A(z^{-1})$, $B(z^{-1})$, $C(z^{-1})$ – полиномы по оператору запаздывания вида

$$\begin{aligned} A(z^{-1}) &= 1 + a_1 z^{-1} + \ldots + a_n z^{-n}, \\ B(z^{-1}) &= 1 + b_1 z^{-1} + \ldots + b_n z^{-n}, \\ C(z^{-1}) &= 1 + c_1 z^{-1} + \ldots + c_n z^{-n}. \end{aligned} \qquad (2.101)$$

В общем случае уравнение (2.98) с учетом вида полиномов (2.101) может быть представлено в виде линейного разностного уравнения как

$$x(t) + \sum_{i=0}^{n} a_i x(t-i) = \sum_{i=0}^{n} b_i u(t-i) + \sum_{i=0}^{n} c_i e(t-i). \qquad (2.102)$$

Система (2.100), описываемая разностным уравнением с двумя входами и одним выходом, имеет полиномы одного порядка, что не является принципиальным. В случае невыполнения данного условия, то некоторые коэффициенты старших степеней достаточно приравнять нулю. Кроме того, для линейных дискретных стационарных вполне управляемых и наблюдаемых динамических систем полиномы $A(z^{-1})$, $B(z^{-1})$, $C(z^{-1})$ не имеют общих множителей, а для устойчивых систем нули полинома $A(z^{-1})$ лежат вне окружности с единичным радиусом.

Для решения задач идентификации уравнения (2.98) описывают широкий класс объектов и процессов, такие как: регрессии, авторегрессии, скользящего среднего, временные ряды и т. п. В этих случаях полиномы $A(z^{-1})$, $B(z^{-1})$, $C(z^{-1})$ принимают различную форму для описания различных процессов.

Если принять, что $A(z^{-1})=1$, $C(z^{-1})=1$, то уравнения (2.98) и (2.102) соответственно принимают вид

$$x\,(\mathrm{t}) = \; B\,(z^{-1})\, u\,(\mathrm{t}) + e\,(\mathrm{t})$$

или

$$x(t) = \sum_{i=0}^{n} b_i \; u \; (t-i) + e(t). \qquad (2.103)$$

Данные уравнения соответствуют динамическому объекту.

Уравнение вида

$$x\,(\mathrm{t}) = \sum_{i=0}^{n} b_i\, u_i\,(\mathrm{t}) + e\,(\mathrm{t})\,, \qquad (2.104)$$

аналогичное уравнению (2.103), в котором вместо *u(t-i)* записано $u_i(t)$ представляет собой статический объект с *n+1* входами и одним выходом. Уравнения (2.103), (2.104) соответствуют регрессионным моделям.

Уравнения (2.98), (2.92) соответствуют регрессионно-авторегрессионным моделям.

Последние уравнения могут быть записаны в более удобной форме.

Так, например, при оценке параметров динамической системы (2.100) при $C\,(z^{-1}) = 1$ последняя может быть представлена в виде

$$x(t) = \theta^T u(t) + \nu\,(t)\,, \qquad (2.105)$$

где $\theta^T = [-a^T, -b^T] = [-a_1,...,-a_n, b_0, b_1,..., b_n] = [\theta_1,...,\theta_N]$, $N = 2n+1$;

$u^T(t) = \left[x\,(\mathrm{t}\text{-}1),...,x\,(\mathrm{t}\text{-}\mathrm{n}), u\,(\mathrm{t}), u\;(\mathrm{t}\text{-}1),...,u\,(\mathrm{t}\text{-}\mathrm{n})\right]$; $\nu(t)$ – последовательность возмущений представляет стационарный стохастический процесс (рис. 2.6), включает в себя все шумовые составляющие, которые воздействуют на систему.

Если в уравнении (2.105) вместо переменных $x(t-i)$ и $u(t-j)$ соответственно записано $u_i(t)$ и $u_{n+1+j}(t)$ $(i = \overline{1,n}; j = \overline{0,n})$, то данное уравнение описывает статический объект с одним выходом $x(t)$ и $N = 2n+1$ входами $u^T(t) = [u_1(t),...,u_N(t)]$.

Как было отмечено выше, весовая функция является исчерпывающей характеристикой линейной динамической системы, и определение ее сводится к решению интегрального уравнения, которое в случае стационарности одномерной системы имеет вид (2.72).

Если перейти от уравнения (2.72) к его дискретной форме, то в общем случае получим уравнение типа свертки

$$x(t) = \sum_{i=0}^{\infty} \omega(i) u(t-i)\,, \qquad (2.106)$$

где ω (*i*) – временная или импульсная характеристика объекта.

Для устойчивых объектов всегда можно предположить, что $\omega(i) \approx 0$ при достаточно большом $i > n$. Поэтому можно ограничиться лишь конечным числом значений весовой функции и уравнением (2.106), приближенно записать в виде

$$x(t) = \sum_{i=0}^{n} \omega(i) u(t-i)\,, \qquad (2.107)$$

которое аналогично уравнению (2.104) для скользящих средних. В этом случае поведение системы характеризуется только не параметрами $c_0$ , $c_1$,..., $c_m$, а набором величин дискретов $\omega(0)$, $\omega(1)$,..., $\omega(\mathrm{n})$ временной характеристики $\omega(i)$, которые подлежат определению.

Для стохастического случая уравнение (2.107) можно представить в виде

$$x(t) = \sum_{i=0}^{n} \omega(i) u(t-i) + e(t), \qquad (2.108)$$

которое аналогично уравнению (2.103). При исследовании технических объектов, описываемых последними уравнениями (2.107), (2.108), можно также представить в общем виде выражением (2.105).

### 2.5.3. Описание дискретной системы в пространстве состояний

Известно, что представление объекта в той или иной эквивалентной форме определяется практической последовательностью построения систем идентификации, устройств наблюдения, размещения корней (полюсов) замкнутой системы.

В ряде практических приложений иногда удачным является представление моделей в пространстве состояний. Наиболее распространенным случаем модели состояния является модель, уравнение состояния которой имеет вид

$$\begin{aligned} x(\mathrm{k}) &= A\, x(\mathrm{k}\text{-}1) + B\, u(\mathrm{k}\text{-}1)\,, \\ y(\mathrm{k}) &= C\, x(\mathrm{k}\text{-}1) + D\, u(k\text{-}1)\,, \end{aligned} \qquad (2.109)$$

где $A$ – матрица коэффициентов размера $nxn$; $B$ – матрица коэффициентов управления размера $nxm$; $C$ – матрица наблюдения размера $rxm$; $D$ – матрица входа rxm. Векторы $x$, $u$, $y$ имеют соответственно размерности $n$, $m$, $r$, что соответствуют модели с $m$-входами, $r$-выходами и n переменными состояния.

При описании систем автоматического управления обычно матрица $D=0$, а матрица $C$, определяющая связь в виде алгебраических уравнений между выходными переменными y с переменными состояния x постоянна.

Для случая стохастических моделей, например, когда в (2.98) $b_o=0$, приведением разностных уравнений к стандартной форме уравнений состояний можно показать, что модель (2.98) имеет эквивалентное представление в пространстве состояний, описываемой уравнениями вида

$$\begin{aligned} x(k) &= A\,x(k-1) + b\,u(k-1) + ce(k-1), \\ y(k) &= d^T\,x(k-1) + e(k), \end{aligned} \tag{2.110}$$

где

$$A = \begin{bmatrix} -a_1 \dots & \\ -a_2 . & I \\ . & \cdots \\ . & \cdots \\ . & \cdots \\ -a_n \cdot & 0 \end{bmatrix}; \quad b = \begin{bmatrix} b_1 \\ b_2 \\ . \\ . \\ . \\ b_n \end{bmatrix}; \quad c = \begin{bmatrix} c_1 - a_1 \\ c_2 - a_2 \\ . \\ . \\ . \\ c_n - a_n \end{bmatrix}; \quad d = \begin{bmatrix} 1 \\ 0 \\ . \\ . \\ . \\ 0 \end{bmatrix}.$$

Отметим, что метод переменных состояния и метод передаточных функций взаимосвязаны и дополняют друг друга при исследовании и проектировании автоматизированных систем управления.

## 2.6. Нелинейные модели

### 2.6.1. Нелинейные непрерывные детерминированные модели

*Дифференциальная модель*

Модель такого объекта в простейшем одномерном случае выражается нелинейными дифференциальными уравнением вида

$$x^{(n)} = f(x^{(n-1)},...,x,u^{(m)},...,u), \tag{2.111}$$

где $x^{(i)} - i$-я производная выхода $x$; $u^{(j)} - j$-я производная входа $x(t)$; $f$ – нелинейная скалярная функция $n+m+1$ аргументов, которую следует идентифицировать по наблюдениям поведения входа и выхода объекта [8,12,19, 20, 22–24, 26].

Введем обозначения:

$$x = x_1; \quad \dot{x}_1 = \dot{x}_2; \dots \quad x^{(n-1)} = x_n. \tag{2.112}$$

Тогда модель (2.111) записывается в виде системы дифференциальных уравнений:

$$\dot{x}_1 = x_2;$$

$$\dot{x}_2 = x_3; \qquad (2.113)$$

$$\dots\dots$$

$$\dot{x}_n = f(x_n, x_{n-1}, \dots, x_1, u_n, \dots, u_m).$$

Естественно обобщить эту систему следующим образом:

$$\dot{x}_1 = f_1(x_1, \dots, x_n, u, \dots, u^{(m)});$$

$$\dots\dots$$

$$\dots\dots$$

$$\dot{x}_n = f_n(x_1, \dots, x_n, u_m, \dots, u).$$

В векторной форме эта система принимает более простой вид:

$$\dot{x} = F(x, \upsilon), \qquad (2.114)$$

где

$$X = (x_1, \dots, x_n) = (x, \dot{x}, \dots, x^{(n-1)}), \qquad (2.115)$$

$$U' = (u, \dot{u}, \dots, u^{(m)}),$$

а $F$ – векторная функция двух векторных аргументов, которая и подлежит идентификации по наблюдениям входных и выходных сигналов.

При малых $n$ и $m$ можно использовать прямое численное дифференцирование (2.114), но для больших $n$ и $m$ это неприемлема, т. к. операция численного дифференцирования вносит неизбежные погрешности, которые обесценивают производные, начиная со второго и, в лучшем случае, с третьего порядка.

Поэтому для определения (2.115) целесообразно обратиться к сглаживанию исходных функций методом наименьших квадратов или к процедуре фильтрации.

*Кусочно-линейная модель*

Будем искать $F(\bullet, \bullet)$ в классе кусочно-линейных функций:

$$F(X, U) = \hat{O}(X, U, \theta),$$

где $\Phi$ – кусочно-линейная функция с параметрами $\theta = (\theta_1, \dots, \theta_k)$. Параметрами $\theta$, как обычно, следует назначать параметры, определяющие опорную последовательность функции $\Phi$:

$$< U_i^0, X_i^0 > \quad i = (1, \dots, q), \qquad (2.116)$$

где $U°=(X,U)$ – аргумент, а $X^0$ – значение функции $\Phi$.

Возможны различные формы задания $\theta$. Наиболее полной является вектор $\theta=(U_1^0,...,U_q^0,X_1^0,...,X_q^0)$, содержащий $k=q(n+2m)$ идентифицируемых параметров. Если взять

$$U_i^0=(X(t_i),U'(t_i)),\ i=1,...,q, \tag{2.117}$$

т. е. сократить опорную последовательность до

$$\theta=(t_1,...,t_q,X_1^0,...,X_q^0), \tag{2.118}$$

то $k=q(m+1)$. Но при этом необходимо еще определить моменты $t_i(i=1,...,q)$, где фиксируется опорная последовательность.

И, наконец, опорную последовательность можно всю заимствовать из (2.116), добавив к (2.117) еще

$$X_i^0=\dot{X}(t_i)\ (i=1,...,q). \tag{2.119}$$

В этом случае $\theta=(t_1,...,t_q)$ и $k=q$, т. е. будут идентифицироваться моменты времени $t_i(i=,\ldots,\ k)$.

*Функциональная модель*

Пусть $F$ является известной (в общем случае заданной алгоритмически) функцией с неизвестными параметрами $\theta$:

$$F=F(X,U,\theta).$$

В этом случае система уравнений

$$\dot{X}=F(X,U,\theta)$$

интегрируется численно (например, методом Рунге-Кутта) при заданных начальных условиях X(0) и фиксированных значениях идентифицируемых параметров $\theta=(\theta_1,...\theta_k)$. Полученное решение $X=X(t,\theta)$ естественно обуславливает получить функцию невязки, минимизация которой решает задачу идентификации динамического объекта.

*Модели линейные относительно идентифицируемых параметров*

Линейность такого рода всегда позволяет построить относительно простые и достаточно эффективные алгоритмы идентификации.

*Модель Вольтерра.* Эта модель связана с рядом Вольтерра, которым представляется выход нелинейной динамической системы [8, 22, 23]:

$$x(t)=\int_0^\infty w_1(\tau_1)u(t-\tau_1)d\tau_1+\int_0^\infty\int w_2(\tau_1,\tau_2)u(t-\tau_1)u(t-\tau_2)d\tau_1d\tau_2+...$$

$$...=\int_0^\infty...\int w_1(\tau_1,...,\tau_l)u(t-\tau_1)...u(t-\tau_1)d\tau_1...d\tau_1+...$$

Легко заметить, что первый член ряда является реакцией линейного объекта на вход ***x(t).*** Последующие члены учитывают нелинейность объекта. Запишем этот ряд в компактном виде

$$x(t)=\sum_{i=0}^{\infty}\int_{0}^{\infty}\ldots\int w_i(\tau_1,\ldots,\tau_i)\prod_{j=1}^{i}u(t-\tau_1)d\tau. \qquad (2.120)$$

Здесь $w_i(\tau_1,\ldots,\tau_i)(i=1,\ldots,l)$ – обобщенные весовые функции, которые и подлежат идентификации. Эти функции обладают следующими свойствами:

1) $w_i(t_1,\ldots,t_i)=0$, если $t_j<0$; $j[1,\ldots,i]$, т. е. если хотя бы один ее аргумент отрицателен;

2) $\lim\limits_{tj\to\infty} w_i(t_1,\ldots,t_i)=0$, где $j\in[1,\ldots,i]$, т. е. весовая функция стремится к нулю при неограниченном увеличении любого из ее аргументов.

Для идентификации естественно параметризовать задачу, т. е. представить эти функции в виде разложения по заданной системе:

$$w_i(t_1,\ldots,t_i)=\sum_{j=1}^{k_i}\theta_{ij}\varphi_{ij}(t_1,\ldots,t_i),(i=1,\ldots,l), \qquad (2.121)$$

где

$$\varphi_{ij}(t_1,\ldots,t_i),\ i=1,\ldots,l,\ \ j=1,\ldots,k_1 \qquad (2.122)$$

- выбранная система функций, содержащая $k=\sum_{i=1}^{l}k_i$ элементов. Каждый элемент этой системы должен обладать свойствами, аналогичными свойствам функций $w_i$ (•).

Такую систему функций можно легко построить, если воспользоваться одномерной системой весовых опорных функций: $w_i(t)$ $(i=1,\ldots,n)$.

С помощью функций $\omega_i(t)$ образуем функции $\varphi_{ij}(.)$ вида:

$$\varphi_{ij}(t_1,\ldots,t_i)=\prod_{q=1}^{i}w_q(t_q). \qquad (2.123)$$

Легко убедиться в том, что функции $\varphi_{ij}$ обладают указанными выше свойствами. Естественно такое представление не может быть универсальным, но оно достаточно хорошо аппроксимирует весовые функции и, что самое главное, позволяет довольно просто решать задачу идентификации нелинейных объектов.

### 2.6.2. Нелинейные непрерывные стохастические модели

*Дифференциальная модель*

Моделью объектов такого рода в простейшем случае является нелинейное дифференциальное уравнение вида [23]

$$\begin{gathered} \upsilon^{(p)} = f(\upsilon^{(p-1)},...,\upsilon;u^{(l)},...,u), \\ x = \upsilon + \varepsilon_x, \\ u = u + \varepsilon_u, \end{gathered} \tag{2.136}$$

где $f$ – идентифицируемая нелинейная функция. Этот случай отличается от (2.111) – (2.113) лишь наличием помех. Воспользуемся этим.

Аналогично (2.113) введем обозначения:

$$\upsilon = \upsilon_1; \quad \dot{\upsilon} = \upsilon_2; \quad \ldots \quad \upsilon^{(p-1)} = \upsilon_p.$$

Тогда модель (2.136) приобретает вид:

$$\begin{gathered} \dot{\upsilon}_1 = \upsilon_2, \\ \dot{\upsilon}_2 = \upsilon_3, \\ ..... \\ \dot{\upsilon}_{p-1} = \upsilon_p, \\ \dot{\upsilon}_p = f(\upsilon_p,\upsilon_{p-1},...,\upsilon_p,u^{(t)},...,u). \end{gathered} \tag{2.137}$$

Эту модель естественно обобщить:

$$\dot{\upsilon}_i = f_i(\upsilon_1,...,\upsilon_p,u^{(l)},...,u) \quad (i = 1,..., p)$$

и представить в векторной форме

$$\begin{gathered} \dot{V} = F(V,U'), \\ x = \upsilon_1 + \varepsilon_x, \\ u = u + \varepsilon_u, \end{gathered} \tag{2.138}$$

где

$$V = (\upsilon_1,...,\upsilon_p) = (\upsilon,\dot{\upsilon},...,\upsilon^{(p-1)}),$$

$$U' = (u_1,...,u_{l+1}) = (u_1,\dot{u},...,u^{(l)}),$$

а $F$ – векторная функция двух векторных аргументов.

В общем случае, когда $n \geq 1$ $(m = 1)$, модель (2.138) представляется в виде

$$\begin{aligned} \dot{V} &= F(V, U_1^{'},...,U_n^{'}), \\ x &= \upsilon_1 + \varepsilon_x(t), \\ u &= u + \varepsilon_u(t), \end{aligned} \qquad (2.139)$$

где $u_i$, – измерение $i$ – го входа и

$$U'_i = (u_i, \dot{u}_i,...,u^{(l)}{}_i) \quad (i = 1,...,n).$$

Как видно, структура задач (2.138) и (2.139) очень похожа. Поэтому можно ограничиться рассмотрением задачи (2.138). Исходная информация о работе объекта содержится в наблюдениях его входа и выхода:

$$\beta = (<u_t,\ x_t> \ 0 \le t \le T). \qquad (2.140)$$

Прежде всего, преобразуем эту информацию к виду, удобному для идентификации объекта типа (2.138). Для этого можно воспользоваться методами сглаживания функций $x_t$ и $y_t$, описанными в приложении. Эти методы позволяют определить и все необходимые производные исходной последовательности и тем самым преобразовать ее к виду

$$\beta = (<U'_t,\ V_t> \quad 0 \le t \le T).$$

Теперь процесс идентификации может производиться любым из рассмотренных в (2.111) – (2.113) уравнений. Однако помехи $\varepsilon_u(t)$ и $\varepsilon_x(t)$ естественно препятствуют эффективной оценке производных в (2.139), и эти методы можно рекомендовать лишь при малом уровне помех или при незначительном порядке системы (2.138). При значительных помехах лучше обращаться к адаптивным методам идентификации, избегая по возможности оценок производных.

*Модель Вольтерра*

Эта модель строится на основе ряда Вольтерра (2.120). Будем для простоты рассматривать два члена ($l$ = 2) в ряду (2.120). Это упрощение не принципиально, и полученные результаты легко распространяются на любое конечное число членов указанного ряда.

Для рассматриваемого случая модель Вольтерра имеет вид:

$$\upsilon(\xi) = \int_0^\infty \omega(\tau) u(\xi - \tau) d\tau + \iint_0^\infty \omega'(\tau_1 \tau_2) u(\xi - \tau_1) u(\xi - \tau_2) d\tau_1 d\tau_2 .$$

Умножим это выражение на $u(\xi - t)$ и проинтегрируем по $\xi$ в пределах от 0 до $\infty$. В результате, после очевидных преобразований, получаем:

$$R_{u\upsilon}(t)=\int_0^{\infty}\omega(\tau)R_{uu}(t-\tau)d\tau+\iint_0^{\infty}\omega'(\tau_1,\tau_2)R_{uuu}(t-\tau_1,t-\tau_2)d\tau_1 d\tau_2,$$

где к известным корреляционным функциям

$$R_{uu}(\tau)=\lim_{T\to\infty}\frac{1}{T}\int_0^{T}u(\xi)u(\xi+\tau)d\xi,$$

$$R_{u\upsilon}(\tau)=\lim_{T\to\infty}\frac{1}{T}\int_0^{T}u(\xi)\upsilon(\xi+\tau)d\xi, \qquad (2.141)$$

добавляется третья

$$R_{uuu}(\tau_1,\tau_2)=\lim_{T\to\infty}\frac{1}{T}\int_0^{T}u(\xi)u(\xi+\tau_1)u(\xi+\tau_2)d\xi.$$

Теперь, учитывая конечность исходной выборки (2.126), получаем модель в окончательном виде

$$\hat{R}_{ux}(t)=\int_0^{T_\omega}\omega(\tau)\hat{R}_{uu}(t-\tau)d\tau+\iint_0^{T_\omega}\omega'(\tau_1,\tau_2)\hat{R}_{uuu}(t-\tau_1,t-\tau_2)d\tau_1 d\tau_2, \qquad (2.142)$$

где

$$\hat{R}_{uu}(\tau)=\frac{1}{T-\tau}\int_0^{T-\tau}u_\xi u_{\xi+\tau}d\xi,$$

$$\hat{R}_{ux}(\tau)=\frac{1}{T-\tau}\int_0^{T-\tau}u_\xi u_{\xi+\tau}d\xi, \qquad (2.143)$$

$$\hat{R}_{uuu}(\tau_1,\tau_2)=\frac{1}{T-\tau}\int_0^{T-\tau}u_\xi u_{\xi+\tau_1}u_{\xi+\tau_2}d\xi(\tau=max(\tau_1,\tau_2)).$$

## 2.7. Математические модели внешних воздействий

### 2.7.1. Характеристики внешних воздействий и их оценивание

Внешние воздействия, как отмечалось в п. 2.1 могут быть полезными (управляющими сигналами $u$) и помехами (возмущающими воздействиями $\nu$). Управляющие сигналы, вырабатываемые устройством управления, являются полностью наблюдаемыми. Возмущающие воздействия, в отличие от них, как правило, ненаблюдаемые и случайные сигналы. В результате выходные переменные объекта $y(t)$ определяются не только входными сигналами $x$(t), но и ненаблюдае-

мыми и неуправляемыми воздействиями (помехами), что вызывает неконтролируемые отклонения выходных переменных от заданных значений. При повторении процессов управления, происходящих в системе, выходные переменные могут иметь различные значения при одних и тех же значениях времени отсчитываемых от начала процесса. Выходная величина объекта при каждом повторном цикле управления, в этом случае, представляет собой реализацию одного и того же случайного процесса управления.

Таким образом, под действием ненаблюдаемых, неуправляемых и случайных внешних воздействий наблюдаемые переменные объекта также становятся случайными сигналами, являющимися реализациями случайного процесса управления.

Для количественной оценки и сравнения различных случайных сигналов используют различные характеристики этих сигналов, представляющие собой абстрактные математические понятия, которые существуют объективно, но не могут быть измерены или определены в строгом смысле слова [20, 26].

К таким характеристикам относятся:

1.Функция распределения вероятностей случайного процесса или интегральная функция распределения. *F(y,t)* – функция распределения вероятностей, это вероятность того, что случайный процесс *x(t)* в момент времени *t* принимает значения меньше *y*

$$F(y) = P\{x(t) < y\}; \tag{2.144}$$

2. Плотность вероятностей или дифференциальное распределение (распределение) *w(x, t)*

$$F(y,t) = \int_{-\infty}^{y} w(x,t)dx \tag{2.145}$$

откуда

$$w(y,t) = \frac{dF(y,t)}{dy}; \tag{2.145}$$

3. Математическое ожидание случайного процесса $M[x(t)] = m(t)$

$$m(t) = \int_{-\infty}^{\infty} x(t)w(x,t)dx; \tag{2.147}$$

4. Дисперсия случайного процесса

$$D(t) = \int_{-\infty}^{\infty} [x(t) - m(t)]^2 w(x,t)dx \tag{2.148}$$

или $$D(t) = M\left[x^2(x)\right] - \left\{m(t)\right\}^2; \qquad (2.149)$$

5. Корреляционная (автокорреляционная) функция $R_{xx}(t_1,t_2)$.

Корреляционная функция – это математическое ожидание произведений двух значений одного и того же сигнала, сдвинутых по времени

$$R_{xx}(t_1,t_2) = M\left[x(t_1),x(t_2)\right]; \qquad (2.150)$$

6. Взаимная корреляционная функция $R_{xy}(t_1,t_2)$. Взаимная корреляционная функция – это математическое ожидание произведений двух сигналов один, из которых сдвинут относительно другого по времени.

$$R_{xy}(t_1,t_2) = M\left[x(t_1),y(t_2)\right]. \qquad (2.151)$$

Точное определение этих характеристик невозможно, т. к. неизвестен вид закона распределения и конечное число реализаций случайного процесса.

Поэтому в реальных условиях эти характеристики вычисляют приблизительно, оценивая их с какой-то погрешностью.

Оценка характеристик случайных процессов проводится на основе принятия гипотез о стационарности и эргодичности случайного процесса.

Случайный процесс называют стационарным, если характеризующая его функция распределения не зависит от времени. Отсюда следует, что от времени не будут зависеть и все характеристики случайного процесса. Условие стационарности значительно упрощает вычисление характеристик случайных процессов, т. к. в выражениях (2.144) – (2.151) исчезает аргумент времени.

Однако и в этом случае для вычисления характеристик необходимо достаточно большое количество независимых реализаций случайного процесса (ансамбль реализаций).

Эргодическая гипотеза позволяет заменить ансамбль реализаций одной реализацией снятой за достаточно продолжительный интервал времени.

Согласно эргодической гипотезе средние значения случайного сигнала по множеству и времени совпадают

$$M\left[x(t)\right] = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} x(t)dt. \qquad (2.152)$$

Тогда для случайных стационарных эргодических процессов оценки их характеристик (2.144) – (2.151) с учетом конечности вре-

мени наблюдения *T,* записываются в следующем виде:

1. Оценка математического ожидания

$$\hat{m} = \frac{1}{T}\int_0^T x(t)dt; \qquad (2.153)$$

2. Оценка дисперсии

$$\hat{D} = \frac{1}{T}\int_0^T [x(t) - m]^2 dt \qquad (2.154)$$

или

$$\hat{D} = \frac{1}{T}\left[\int_0^T x^2(t)dt - \int_0^T x(t)dt\right]; \qquad (2.155)$$

3. Оценка корреляционной функции

$$R_{xx}(\tau) = \frac{1}{T}\int_0^T x_0(t)x_0(t+\tau)dt, \qquad (2.156)$$

где $x_0(t) = x(t) - m$ – центрированный случайный сигнал.

Корреляционную функцию центрированного сигнала еще называют ковариационной или автоковариационной функцией;

4. Спектральная плотность мощности $S(\omega)$, связанная с корреляционной функцией преобразованием Фурье:

$$S_{xx}(\omega) = \int_{-\infty}^{\infty} R_{xx}(\tau)e^{-j\omega\tau}d\tau, \qquad (2.157)$$

$$S_{xx}(\tau) = \frac{1}{2\pi}\int_{-\infty}^{\infty} S_{xx}(\omega)e^{-j\omega\tau}d\omega.$$

Для получения приемлемой точности оценок характеристик случайных процессов длительность реализации процесса, по которой вычисляются оценки, должна превышать интервал корреляции. Интервал корреляции $\tau_{max}$ – это значение аргумента корреляционной функции, начиная с которого все ее последующие значения не превышают (0,01 – 0,05) R(0). [6, 9, 26].

Одним из наиболее распространенных подходов к моделированию случайных сигналов является метод формирующих фильтров [3, 13, 26]. Моделирование случайного сигнала этим методом осуществляется в предположении, что он является реакцией линейной системы на случайный входной сигнал, характеристики которого известны.

Пусть на вход линейной системы с передаточной функцией *W(p)* и функцией веса $\omega(t)$ действует центрированный, стационарный, эргоди-

ческий, случайный сигнал *x(t)* с известной корреляционной функцией $R_{xx}(\tau)$ и спектральной плотностью $S_{xx}(\omega)$.Тогда случайный сигнал на выходе системы вычисляется с помощью интеграла по формуле

$$y(t) = \int_{-\infty}^{\infty} \omega(\tau) x(t-\tau) d\tau . \qquad (2.158)$$

Умножив это выражение сначала на $x(t + \tau)$, а затем на $y(t + \tau)$, и проинтегрировав обе части по $\tau$ в пределах от $-T$ до $T$ (при $T \to \infty$), получим соотношения Винера-Ли:

$$R_{xy}(\tau) = \int_{-\infty}^{\infty} \omega(t) R_{xx}(t-\tau) dt ,$$

$$R_{yy}(\tau) = \int_{-\infty}^{\infty} \omega(t) R_{yx}(t-\tau) dt . \qquad (2.159)$$

Отсюда следует, что

$$R_{yy}(\tau) = \int_{-\infty}^{\infty} R_{\omega\omega}(\tau) R_{xx}(t-\tau) dt , \qquad (2.160)$$

где

$$R_{\omega\omega}(\lambda) = \int_{-\infty}^{\infty} \omega(\tau)\omega(\tau+\lambda) dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} \left|W(j\omega)\right|^2 e^{j\omega\lambda} d\omega . \qquad (2.161)$$

Эти соотношения значительно упрощаются, если представить их через спектральные плотности, используя соотношения Винера-Хинчина (2.157)

$$S_{xy}(\omega) = W(j\omega) S_{xx}(\omega),$$

$$S_{yy}(\omega) = W(-j\omega) S_{xx}(\omega), \qquad (2.162)$$

$$S_{yy}(\omega) = W(-j\omega) S_{yx}(\omega) = W(j\omega) S_{yx}(\omega) = \left|W(j\omega)\right|^2 S_{xx}(\omega).$$

При моделировании в качестве входного сигнала используют стационарный белый шум со следующими характеристиками:

$$S_{xx} = S_0,$$

$$R_{xx} = \frac{2}{2\pi} S_0 e^{j\omega\lambda} d\omega = S_0 \delta(\tau), \qquad (2.163)$$

где $S_0$ – интенсивность белого шума.

Белый шум представляет собой сумму гармонических колебаний всех частот, имеющих одну и ту же дисперсию амплитуды. При наличии на входе линейной системы белого шума на ее выходе формируется случайный сигнал характеристики (2.159) –(2.162).

$$R_{xy}(\tau) = S_0 \omega(\tau), \tag{2.164}$$

$$R_{yy}(\tau) = S_0 R_{\omega\omega}(\tau),$$

$$R_{yy}(\omega) = |W(j\omega)|^2 S_0. \tag{2.165}$$

Очевидно, что успех и точность статистического моделирования зависят в основном от качества формирования последовательности случайных чисел, имеющих свойства белого шума.

Задача получения последовательности случайных чисел обычно разбивается на две. Вначале получают последовательность случайных чисел, имеющих равномерное распределение в интервале [0,1]. Затем из нее получают последовательность случайных чисел, имеющих произвольный закон распределения. Один из способов такого преобразования состоит в использовании нелинейных преобразований. Пусть необходимо получить случайную величину *x*, функция распределения вероятностей которой

$$F(y) = P\{x \le y\}. \tag{2.166}$$

Значения искомой функция распределения вероятностей лежат в диапазоне $0 \le F(y) \le 1$. Если теперь придавать функции *F(y)* произвольные случайные значения, лежащее в диапазоне ее существования [0,1], то значения ее аргумента будут иметь заданный закон распределения. Значения функции *F(y)* задаются от датчика случайных чисел имеющих равномерное распределение в диапазоне [0,1], а значения аргумента находятся как обратная функция от *F(y)*

$$x = y\, F_{обр}(y),$$

$$0 \le y \le 1. \tag{2.167}$$

Таким образом, основная проблема моделирования случайных последовательностей заключается в получении независимых, равномерно распределенных чисел в интервале [0,1].

При аппаратном генерировании чаще всего используют шумящие электронные устройства, шум которых связан с шумами электронных ламп или транзисторов. При усилении этих шумов получается напряжение, которое является случайным процессом. Если брать его значения, достаточно далеко отстоящие друг от друга, так чтобы они были некоррелированы, то величины получаемых напряжений образуют последовательность независимых случайных величин.

Для программной генерации случайных чисел разработано большое количество специальных программ, имеющихся практически в каж-

дом языке программирования. На ЦВМ принципиально невозможно получить идеальную последовательность случайных чисел из-за конечного числа разрядов, поэтому такие последовательности являются псевдослучайными. Псевдослучайные последовательности характеризуются длиной отрезка апериодичности и длиной периода. Под длиной отрезка апериодичности понимается совокупность последовательно полученных случайных чисел, в которой ни одно число не встречается дважды. Под длиной периода последовательности случайных чисел понимается такая длительность этой последовательности, при превышении которой, числа начнут периодически повторяться. У современных программ эти характеристики имеют порядок $10^{10}$–$10^{30}$, что вполне приемлемо для задач моделирования случайных процессов.

При моделировании непрерывных систем в качестве генератора белого шума можно использовать телеграфную волну, порождаемую процессом Пуассона, который в свою очередь является частным случаем марковского процесса. Функция, задаваемая телеграфной волной, принимает только положительные и отрицательные значения равные по модулю, причем последовательность изменений знака представляет собой процесс Пуассона со средней скоростью изменения λ. Такой процесс стационарен и эргодичен, если он начинается с $t \to -\infty$ и для него

$$M[x(t)] = 0,$$
$$R_{xx}(\tau) = a^2 e^{-2\lambda|\tau|}, \qquad (2.168)$$
$$S_{xx}(\omega) = \frac{4a^2\lambda}{\omega^2 + 4a^2}.$$

Если выбрать среднюю скорость изменения знака телеграфной волны в несколько раз большей частоты генерируемого случайного сигнала, то приближенно можно считать, что телеграфная волна обладает свойствами белого шума.

### 2.7.2. Линейные модели и их применение для оценивания характеристик случайных процессов

Как известно [20], спектральная плотность процесса, полученного в результате прохождения белого шума через линейную систему, равна произведению интенсивности входного шума на квадрат модуля комплексной частотной характеристики системы. В свою очередь, комплексная частотная характеристика рекурсивного формирующего

фильтра представляет собой функцию, полученную в результате подстановки $z = e^{j\omega}$, где $j = \sqrt{-1}$.

Определим спектральную плотность процесса АРСС (p, q), если передаточная функция рекурсивного фильтра

$$G(z) = \frac{D(z)}{C(z)} = \frac{1 + \sum_{k=1}^{q} d_k z^{-k}}{1 - \sum_{i=1}^{p} c_i z^{-i}}. \qquad (2.169)$$

Делая подстановку $z = e^{j\omega}$, в передаточную функцию получим

$$G(e^{j\omega}) = \frac{1 + \sum_{k=1}^{q} d_k e^{-j\omega\omega}}{1 - \sum_{i=1}^{p} c_i e^{-j\omega\omega}}. \qquad (2.170)$$

Тогда квадрат модуля комплексной частотной характеристики

$$\left|G(e^{j\omega})\right|^2 = \frac{\left(1 + \sum_{k=1}^{q} d_k \cos k\omega\right)^2 + \left(\sum_{k=1}^{q} d_k \sin k\omega\right)^2}{\left(1 - \sum_{i=1}^{p} c_i \cos i\omega\right)^2 + \left(\sum_{i=1}^{p} c_i \sin i\omega\right)^2}, \qquad (2.171)$$

следовательно, спектральная плотность АРСС-процесса описывается передаточной функцией (2.172)

$$S(\omega) = \left|G(e^{j\omega})\right|^2 \sigma_e^2 = \sigma_e^2 = \frac{\left(1 + \sum_{k=1}^{q} d_k \cos k\omega\right)^2 + \left(\sum_{k=1}^{q} d_k \sin k\omega\right)^2}{\left(1 - \sum_{i=1}^{p} c_i \cos i\omega\right)^2 + \left(\sum_{i=1}^{p} c_i \sin i\omega\right)^2}, \qquad (2.172)$$

где $\sigma_e^2$ – дисперсия процесса.

Отметим, что оценивание спектральной плотности (2.172) сводится к оцениванию коэффициентов передаточной функции (2.169) и дисперсии ненаблюдаемого белого шума, т. е. к идентификации АРСС-процесса. Таким образом, АРСС-процессы позволяют получать оценки спектральной плотности непосредственно по наблюдениям, минуя вычисления статистических характеристик наблюдений.

Рассмотрим теперь автокорреляционные функции (АКФ) процессов АР($p$), СС($q$), АРСС($p$, $q$). АКФ процесса АР($p$) выражается разностным уравнением, которое аналогично уравнению, описывающему сам процесс

$$m = c_1R(m-1) + c_2R(m-2) + \ldots + c_pR(m\text{-}p),$$

где $R(m)$ – значение АКФ при дискретном сдвиге $m$.

Таким образом, АКФ процесса АР(р) бесконечна, и представляет собой затухающую экспоненту или экспоненциально затухающую косинусоиду.

Выражение для дисперсии процесса АР($p$) имеет вид

$$\sigma_{AP}^2 = \frac{\sigma_e^2}{1 - c_1R(1) - c_2R(2) - \ldots - c_pR(p)}. \qquad (2.173)$$

АКФ процесса СС(q) определяется следующим образом:

$$R(m) = \begin{cases} \dfrac{d_m + d_1c_{m+1} + \ldots + d_{q-m}c_p}{1 + d_1^2 + \ldots + d_q^2}, m = \overline{1,q} \\ 0, m > q \end{cases}. \qquad (2.174)$$

АКФ процесса скользящего среднего равна нулю при сдвиге, большем $q$, т. е. большем, чем порядок процесса. Дисперсия процесса СС($q$)

$$\sigma_{CC}^2 = \left(1 + d_1^2 + \ldots + d_q^2\right)\sigma_e^2. \qquad (2.175)$$

Общего выражения АКФ для АРСС-процесса для $m[0,\infty]$ не существует.

В каждом случае АКФ вычисляется по известным $\{c_i\}$ и $\{d_k\}$ и выражению для автоковариационной функции процесса АРСС($p$, $q$), которое имеет вид

$$R(m)=c_1R(m-1)+ \ldots + c_pR(m-p)+R_{xe}(m)+d_1R_{xe}(m-1)+ \ldots + d_0R_{xe}(m\text{-}q), \qquad (2.176)$$

где $R(m)$ – значение автоковариационной функции при сдвиге $m$; $R_{xe}(m)$ – значение взаимной ковариационной функции процесса и шума.

Рассмотренные модели случайных последовательностей позволяют с помощью небольшого числа параметров описать обширный класс случайных процессов. На практике часто оказывается, что адекватное описание наблюдаемых временных рядов достигается с помощью моделей авторегрессии, скользящих средней или комбинированной модели, в которых $p$ и $q$ не больше трех [56].

Рассмотрим процессы АР(1) и АРСС (1,1) более подробно. Процесс авторегрессии первого порядка (марковский процесс)

$$v(k)=c_1v(k-1)+e(k) \qquad (2.177)$$

часто используется в качестве модели случайных сигналов и возмущений [18].

Для стационарности процесса необходимо, чтобы $c_i \in [-1,1]$.

Согласно формуле (2.176) АКФ процесса АР(1) удовлетворяет разностному уравнению первого порядка:

$$R(m) = c_1 R(m-1), \quad m > 0, \tag{2.178}$$

которое при $R(0)$=1 имеет решение

$$R(m) = c_1^m, m \geq 0. \tag{2.179}$$

Согласно выражению (2.175) дисперсия процесса имеет вид

$$\sigma^2 = \frac{\sigma_e^2}{1 - R(1)c_1}, \tag{2.180}$$

или заменяя $R(1)$ на $c_1$ ,

$$\sigma^2 = \frac{\sigma_e^2}{1 - c_1^2}.$$

Используя формулу (2.172), найдем спектр процесса.

Смешанный процесс авторегрессии – скользящего среднего первого порядка АРСС (1,1) – описывается формулой

$$v(k)=c_1 v(k-1)+d_1 ek-(k-1)+e(k) \tag{2.181}$$

Этот процесс, как и процесс АР(1), стационарен, если $c_1 \in [-1,1]$. Определим АКФ процесса. Из выражения (2.176) получаем

$$\begin{aligned} &R(0) = c_1 R(1) + \sigma_e^2 + d_1 R_{xe}(-1), \\ &R(1) = c_1 R(0) + d_1 \sigma_e^2, \\ &\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ &R(m) = c_1 R(m-1), \\ &m \geq 2. \end{aligned} \tag{2.182}$$

Умножим значение (2.181) на $e(k-1)$ и, переходя к математическим ожиданиям, получим

$$R_{xe}(-1) = (c_1 + d_1)\sigma_e^2.$$

Отсюда автоковариационная функция процесса равна:

$$\begin{aligned} &R(0) = \frac{\sigma_e^2 (1 + 2c_1 d_1 + d_1)}{1 - c_1^2} \\ &R(1) = \frac{\sigma_e^2 (1 + c_1 d_1)(c_1 + d_1)}{1 - c_1^2} \\ &\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ &R(m) = c_1 R(m-1) \\ &m \geq 2 \end{aligned} \tag{2.183}$$

Следовательно,

$$R(0) = \frac{(1 + c_1 d_1 + d_1)(c_1 + d_1)}{1 + 2c_1 d_1 + d_1^2},$$

$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \qquad (2.184)$$

$$R(m) = c_1 R(m-1).$$

Следуя формуле (2.171), спектр процесса АРСС (1,1)

$$S(\omega) = \frac{1 + 2d_1 \cos\omega + d_1^2}{1 - 2c_1 \cos\omega + c_1^2}\sigma_e^2, \ 0 \le \omega < \pi. \qquad (2.185)$$

Таким образом, для описания стохастических систем можно применять модели СС*(q),* АР*(p)* или АРСС*(p, q)*. Доказательство этого следует из теоремы представления Острема [20]. Согласно этой теореме, для заданной в виде рациональной функции спектральной плотности S*(ω),* существует асимптотически устойчивая линейная динамическая система, такая, что при воздействии на ее вход белым шумом с дискретным временем, выходным сигналом будет стационарный процесс со спектральной плотностью S*(ω)*.

# ГЛАВА 3. ОБЩИЕ СВЕДЕНИЯ И ПОНЯТИЯ О ЗАДАЧАХ ИДЕНТИФИКАЦИИ

## 3.1. Постановка задачи и особенности процесса идентификации

### 3.1.1. Основные определения. Общая схема процесса идентификации

Под идентификацией будем понимать процедуру построения оптимальной (наилучшей) в определенном смысле математической модели объекта по реализациям его наблюдаемых входных и выходных сигналов [10, 22, 24, 36, 38].

Допустим, что функционирование реального объекта описывается соотношением

$$x = F(u, \nu), \tag{3.1}$$

где $x$ – вектор наблюдаемых выходных переменных (сигналов); $u$ – вектор наблюдаемых входных переменных (сигналов); $\nu$ – вектор ненаблюдаемых входных переменных (сигналов); $F$ – оператор, связывающий входные и выходные переменные объекта.

Задачей идентификации является построение модели с оператором $F_M$, который бы в определенном смысле был близок к оператору объекта $F$, т. е.

$$F_M \approx F. \tag{3.2}$$

Очевидно, что связь между входными сигналами $u_m$ и выходными сигналами $x_m$ модели при этом описывается соотношением

$$x_M \approx F_M(u_M). \tag{3.3}$$

Отметим, что "близость", согласно выражению (3.2), является весьма условной, а тождественное совпадение операторов $F$ и $F_M$, вообще практически невозможно. В реальных условиях операторы $F$ и $F_M$ могут иметь разные области задания, вследствие этого разные структуры, размерность и т. д. Задача осложняется, кроме того, тем, что во многих случаях априорная информация об операторе F сама является весьма относительной, т. е. в свою очередь модельной. Поэтому в процедурах идентификации близость операторов $F$ и $F_M$ оценивают обычно по реакциям объекта и модели на одно и то же входное воздействие u. При этом

$$\hat{x} \approx F_M(u). \tag{3.4}$$

Для оценивания этой близости вводится скалярная функция

$$Q = Q(x, \hat{x}), \tag{3.5}$$

удовлетворяющая обычно следующим трем условиям:

1. Она является неотрицательной, т. е.

$$Q(x, \hat{x}) \geq 0 \text{ при любых } x \text{ и } \hat{x}; \tag{3.6}$$

2. Равна нулю, при x $= \hat{x}$, т. е.

$$Q(x, \hat{x}) = 0; \tag{3.7}$$

3. Непрерывна и выпукла по обоим аргументам, т.е.

$$\begin{gathered} Q\left[(1-\lambda)x_1 + \lambda x_2, \hat{x}\right] \leq (1-\lambda)Q(x_1, \hat{x}) + \lambda Q(x, \hat{x}), \\ Q\left[x, (1-\lambda)\hat{x}_1 + \lambda \hat{x}_2\right] \leq (1-\lambda)Q(x, \hat{x}) + \lambda Q(x, \hat{x}_2), \\ 0 \leq \lambda \leq 1. \end{gathered} \tag{3.8}$$

Применительно к мере Q используются разные термины: функция потерь, функция невязки, показатель адекватности модели реальному объекту, показатель близости, показатель качества идентификации и т. д.

Вследствие (3.4) выражение для показателя (3.5) имеет вид

$$Q = Q\left[(x, F_M(u))\right]. \tag{3.9}$$

В силу свойств показателя близости можно предположить, что эквивалентность выходных сигналов объекта и модели будет достигнута при условии

$$Q=0. \tag{3.10}$$

Отсюда следует очевидное правило: оператор $F_M$ модели надо подбирать таким, чтобы выполнялось условие (3.10). На практике порой приходится удовлетворяться достижением минимума величины Q. Этого в общем случае можно добиться путем реализации какого-либо алгоритма минимизации функции (3.9). Формально такую процедуру можно записать следующим образом:

$$Q(F_M) \to \min_{F_M \in C_F} \to \hat{F}_M \tag{3.11}$$

или

$$\hat{F}_M = arg \min_{F_M \in C_F} Q(F_M),$$

где $C_F$ – класс (множество) допустимых операторов; $\hat{F}_M$ – оптимальная оценка оператора $F_M$.

Обобщенная схема системы идентификации показана на рис. 3.1. Система предусматривает измерения входных и выходных сигналов объекта и использование определенной априорной информации об объекте и внешних условиях.

Более подробная схема системы идентификации реализующей правило (3.11), изображена на рис. 3.2.

На последней схеме через $\nu_2$ и $\nu_3$ обозначены помехи каналов измерений (передачи) соответственно входного и выходного x сигналов, через $z$ – сигнал перестройки оператора $F_M$ модели, через $\nu_1$ – ненаблюдаемые входные переменные объекта.

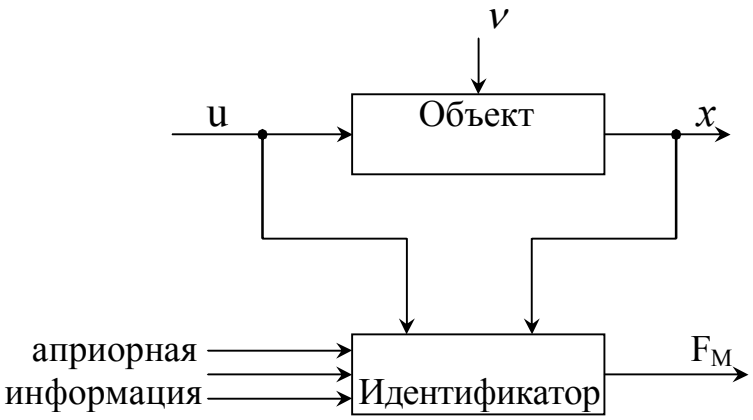



Рис. 3.1. Обобщенная схема системы идентификации

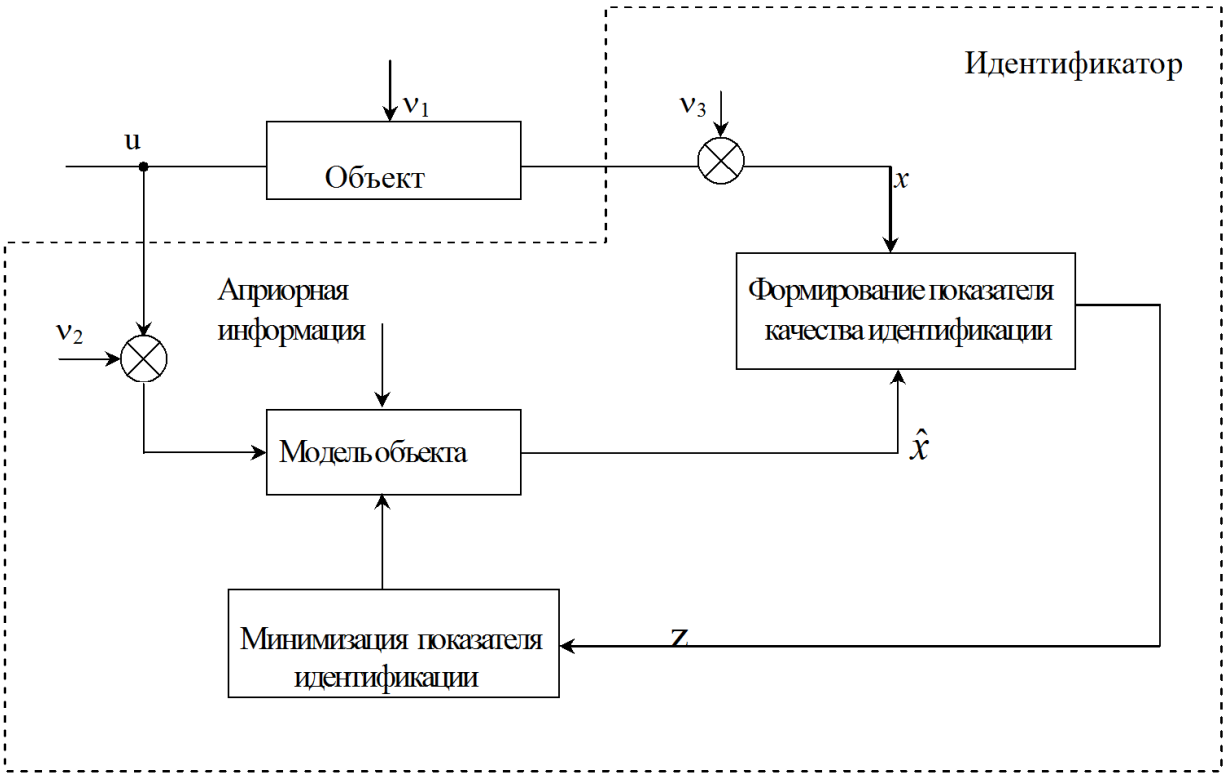



Рис. 3.2. Структурная схема системы идентификации реализующей правило (3.11)

### 3.1.2. Основные этапы идентификации

В соответствии с определением, данным в п. 3.l, структурной схемой на рис. 3.2 и содержанием процедур идентификации, можно выделить следующие основные этапы этой процедуры:

1. Выбор и обоснование класса (типа) моделей объекта идентификации;
2. Выбор показателя и критерия близости объекта и модели;
3. Выбор метода (принципа) идентификации и разработки соответствующих алгоритмов;
4. Выбор численных методов для реализации алгоритмов идентификации;
5. Получение и соответствующая обработка данных входных и выходных сигналов объекта;
6. Выбор вычислительных средств (системы) и реализация процедуры идентификации;
7. Анализ полученных результатов;
8. Использование полученных результатов по целевому назначению.

Выбор и обоснование класса математической модели объекта является важнейшим этапом процедуры идентификации, существенно влияющим на все последующие этапы. Классификация и описание наиболее типовых моделей изложены во второй главе. Здесь отметим лишь, что выбор модели объекта в каждом конкретном случае производится с учетом всей априорной информации об объекте идентификации. Эта априорная информация может содержать данные о классе объекта линейный – нелинейный, стационарный – нестационарный, непрерывный – дискретный, наблюдаемый – ненаблюдаемый и т. д.), о порядке (размерности) уравнений, описывающих объект и т. д. На выбор модели определенное влияние могут оказать и такие факторы, как целевое предназначение модели, возможности имеющейся в распоряжении исследователя вычислительной техники, квалификация специалистов, привлекаемых к решению задачи идентификации, и прочие факторы, выступающие в качестве ограничивающих условий.

Проблемы выбора показателей и критериев близости, а также методов и алгоритмов идентификации тесно связаны между собой при решении задач идентификации. Данным проблемам в научной литературе по теории идентификации уделяется основное внимание.

Последующие этапы (выбор численных методов, измерения и обработка входных и выходных сигналов объекта, выбор вычислительных средств и реализация на них метода идентификации) принципиально не отличаются от аналогичных этапов решения других задач на ЭВМ, например задач автоматизированного или автоматического управления.

На этапе анализа результатов идентификации основное внимание уделяется исследованию точности и надежности (достоверности) полученных оценок оператора или параметров объекта. При необходимости на этом этапе может быть произведена проверка пригодности построения модели к использованию по назначению.

Если окажется, что построенная модель не удовлетворяет предъявляемым к ней требованиям, то процедура идентификации может быть повторена с коррекцией всех или части этапов. Например, могут быть пересмотрены тип и структура модели, показатель близости, метод идентификации, вычислительные средства и т. д. В итоге построение математической модели сводится к многошаговой (итерационной) процедуре.

Во многих случаях проверка пригодности сводится к проверке адекватности модели объекта с использованием более сложного (чем при оценивании оператора или параметров модели) показателя близости (адекватности).

Пригодность модели определяется требованиями, которые зависят от той цели, ради которой строится данная модель. Дело в том, что в науке и технике задача идентификации является не самостоятельной, а подчиненной конкретным целям, определяемым системой, в которой используется построенная модель. Это – управление, прогнозирование, диагностирование, контроль, распознавание образов, принятие решений, например, при проведении испытаний техники, процесс познания окружающего нас мира.

В связи с указанным естественным является рассмотрение системы идентификации в качестве подсистемы основной системы (суперсистемы). Основная система характеризуется своими целями, показателями и критериями эффективности (качества). Им должны быть подчинены соответствующие характеристики подсистемы идентификации. Логично также исследование задачи идентификации производить совместно с исследованием основной задачи (например, управ-

ления или диагностирования). Это вполне соответствовало бы требованиям системного подхода, который в настоящее время начинает широко внедряться в исследовательскую практику и в практику управления. Но, к сожалению, достаточно обоснованные методы такого подхода к проблеме идентификации в настоящее время отсутствуют. К тому же выделение идентификации в самостоятельную задачу в рамках данной книги сделано из чисто методических и методологических соображений; т. к. методы построения математических моделей существенно отличаются от решения основных задач теории управления, диагностики, контроля и т. д.

### 3.1.3. Априорная и апостериорная информация

Априорная информация о модели и о входных и выходных сигналах объекта характеризует степень предварительной изученности объекта и условий его функционирования. Учитывая сложность решения задач идентификации, трудно переоценить роль указанной априорной информации при построении модели объекта. Априорная информация оказывает существенное влияние практически на все этапы процесса идентификации. Особенно велика ее роль при обосновании методов идентификации.

Обычно исследователь располагает определенными сведениями об объекте и о входных и выходных сигналах объекта. Заметим, что отсутствие каких-либо сведений (априорной информации) об объекте и сигналах делает бессмысленной, на современном уровне развития теории управления, задачу построения моделей более или менее сложных объектов. При этих условиях исследователь имеет дело с объектом типа "черный ящик". На практике обычно рассматриваются объекты типа "серый ящик", о которых априори что-то известно.

Наиболее значимой для специалистов по идентификации является информация о том, известна или неизвестна структура модели объекта, к какому типу относится модель.

Структура модели характеризуется оператором $F_M$. Поэтому под структурой будем понимать вид оператора с точностью до его параметров (коэффициентов). Оператор определяет фактически и тип модели. Признаками, характеризующими тип модели, как уже было рассмотрено, является стохастичность модели (детерминированные и стохастиче-

ские), динамичность (динамические и статические), нелинейность по параметрам (линейные и нелинейные по параметрам), нелинейность по переменным состояния или фазовым координатам, дискретность (непрерывные и дискретные), размерность (одномерные и многомерные), распределенность параметров (сосредоточенными параметрами и распределенными параметрами), переменность параметров (с постоянными параметрами и переменными параметрами) и т. д.

Даже в рамках известного типа модели чрезвычайно важной может оказаться дополнительная информация о структуре и свойствах оператора *F*. Например, только факт отнесения модели к нелинейному классу существенно не упрощает задачу идентификации, т. к. класс нелинейных систем чрезвычайно многообразен и, можно сказать, неограничен. В связи с этим любые сведения о структуре уравнений, о типе нелинейности (гладкие, разрывные, однозначные и т. д.), о характере процессов в объекте (монотонность, автоколебательность и пр.), о свойствах линейной части, и другие данные могут оказаться весьма полезными исследователю.

В зависимости от степени изученности структуры объекта различают структурную и параметрическую идентификацию.

Структурной идентификацией называется процесс построения модели в случае, когда априори отсутствует полная (исчерпывающая) информация о структуре (и, естественно, о параметрах) оператора F. Синонимом понятия "структурная идентификация" является "идентификация в широком смысле".

При идентификации в широком смысле вначале определяется структура оператора *F* (к какому классу принадлежит этот оператор), затем оцениваются неизвестные параметры этого оператора.

Заметим, что даже структурная идентификация возможна только тогда, когда об объекте (модели) что-то известно. Как уже было указано, отсутствие всякой информации об объекте практически не позволяет построить модель. Так, определение порядка линейного дифференциального уравнения, описывающего поведение объекта, возможно только в предположении о том, что объект линеен и одномерен. Сведения о линейности и одномерности могут составить априорную информацию о модели. Конечно, в данной ситуации принципиально задача идентификации может быть решена и при отсутствии, каких-бы то ни было сведений об объекте. Для этого вначале необхо-

димо определить размерность объекта, затем проверить гипотезу о линейности и после этого оценить размерность уравнения, описывающего объект с последующим оцениванием параметров. Но даже для этого простейшего случая с линейным объектом необходимо реализовать весьма трудоемкий длительный по времени процесс. Если же идентифицируется достаточно сложный нелинейный объект, то необходимость проверки огромного количества возможных описаний объекта сведет на нет оперативность решения задачи.

Если структура оператора модели априори известна, то в процессе идентификации оцениваются только неизвестные параметры. При этом говорят о параметрической "идентификации или об идентификации в узком смысле". При идентификации в узком смысле класс оператора $F$ задается с точностью до параметров.

Обозначим вектор неизвестных параметров $\theta_1, \theta_2, ..., \theta_n$ через $\theta$. Тогда формальную запись (3.11), условно описывающую в общем виде задачу идентификации, можно переписать в виде

$$Q[F_M(\theta)] \to \min_{\theta \in C_\theta} \to F_M(\hat{\theta}), \qquad (3.12)$$

где $C_\theta$ – множество допустимых значений параметров $\theta$.

В некоторых случаях процедура (3.12) приводит к линейным или нелинейным алгебраическим уравнениям, неизвестными в которых являются искомые параметры $\theta$. Аналогичные уравнения в дальнейшем будем называть уравнениями идентификации.

Рассмотренная выше априорная информация об объекте носит в основном качественный характер. Но могут иметь место и количественные предварительные данные об объекте. Пусть, например, известно, что параметры объекта с заданной структурой являются случайными величинами, которые необходимо оценить в процессе идентификации. При этом могут быть заданы априорные законы распределения в виде функции распределения $F(\theta)$ или плотности распределения $f(\theta)$ с точностью до параметров распределения. Эти законы и образуют априорную информацию об объекте.

Априорная информация о входных и выходных сигналах содержит качественные сведения о природе этих сигналов (детерминированные или стохастические), о типе сигналов (непрерывные или дискретные), об их размерности (одномерные или многомерные) и т. д. Для случайных входных и выходных сигналов могут быть известны

их качественные и количественные характеристики. Информация качественного характера касается класса случайных сигналов: случайные величины или случайные процессы, стационарные или нестационарные, эргодические или неэргадические, марковские, полумарковские или немарковские процессы и т. д. Информация количественного характера содержит априорные законы распределения или числовые характеристики входных сигналов и помех.

В отличие от априорной информации апостериорная информация обычно имеет количественное выражение. Она представляет собой результаты измерений входных и выходных сигналов объекта.

## 3.2. Критерии и показатели качества идентификации

### 3.2.1. Общие сведения

В настоящее время завершенной теории целенаправленного выбора (синтеза) показателей и критериев качества идентификации нет. Во многом выбор показателей и критериев производится субъективно, с учетом опыта исследователя, особенностей задачи, объема априорной информации, возможностей вычислительной техники и т. д. Обычно в качестве показателя качества выбирается функционал $J(\varepsilon)$, характеризующий точность идентификации (близость модели и объекта). Критерий формирует требование к показателю $J(\varepsilon)$. В задачах идентификации это требование, как правило, сводится к минимизации показателя $J(\varepsilon)$ по идентифицируемым характеристикам структуры или параметрам. Заметим, что в литературе зачастую понятия показатель и критерий неоправданно отождествляются.

Мера погрешности (невязки) в функционале $J(\varepsilon)$ определяется видом пространства оптимизации (минимизации) показателя $J(\varepsilon)$, а функционал $J(\varepsilon)$ – выбором метрики в этом пространстве. Обычно величина $\varepsilon$ может определяться в одном из трех пространств: невязок уравнений объекта, выходного сигнала, оцениваемых параметров.

Так, для объекта с уравнением

$$x^{(n)} + \sum_{i=0}^{n-1} a_i x^{(i)} = \sum_{i0}^{m} b_i u^{(i)}$$

невязка уравнения определяется выражением

$$\varepsilon(t) = x^{(n)} + \sum_{i=0}^{n-1} a_i x^{(i)} - \sum_{i0}^{m} b_i u^{(i)} .$$

Предполагаем, что выходные сигналы совпадают с переменными состояния (фазовыми координатами), т. е. $y_i = x_i$.

В пространстве выходных сигналов мера погрешности ε равна

$$\varepsilon(t) = x(t) - \hat{x}(t), \tag{3.13}$$

где $\hat{x}(t)$ – выходной сигнал модели объекта.

В пространстве идентифицируемых параметров в качестве меры погрешности естественно рассмотреть вектор

$$\Delta\theta = \theta - \hat{\theta}, \tag{3.14}$$

где

$$\theta^T = \left[a_0, a_1, ..., a_{n-1}, b_0, b_1, ..., b_m\right],$$
$$\hat{\theta}^T = \left[\hat{a}_0, \hat{a}_1, ..., \hat{a}_{n-1}, \hat{b}_0, \hat{b}_1, ..., \hat{b}_m\right].$$

Функционал $J(\varepsilon)$ формируется на основе рассмотренных невязок ε. Чаще всего выбираются среднеквадратичный, среднемодульный и чебышевский (максимального отклонения) показатели. При идентификации в пространстве невязок уравнений и выходных сигналов показатель $J(\varepsilon)$ в общем виде определяется выражениями

$$J(\varepsilon) = \left[\int_{t_0}^{t_1} |\delta(t)|^k dt\right]^{\frac{1}{k}}$$ – в непрерывном случае;

$$J(\varepsilon) = \left[\sum_{i=1}^{N} |\varepsilon(t_i)|^k dt\right]^{\frac{1}{k}}$$ – в дискретном случае.

При k=2 имеем среднеквадратичный показатель, при k=1 – среднемодульный, при $k \to \infty$ – чебышевский (показатель максимального отклонения).

В пространстве идентифицируемых параметров могут быть построены показатели типа среднеквадратичного

$$J(\varepsilon) = \sum_{i=0}^{n-1} (a_i - \hat{a}_i)^2 + \sum_{i=0}^{m} (b_i - \hat{b}_i)^2 \tag{3.15}$$

и среднемодульного

$$J(\varepsilon) = \sum_{i=0}^{n-1} |a_i - \hat{a}_i| + \sum_{i=0}^{m} |b_i - \hat{b}_i| . \tag{3.16}$$

Для решения практических задач наибольшее употребление нашли меры погрешности в пространстве выходных сигналов и среднеквадратичный показатель качества идентификации. При использовании невязок уравнений объекта возникает проблема измерения высших производных входного *u(t)* и выходного *x(t)* сигналов. В связи с этим разрабатываются методы предварительного преобразования уравнений объекта таким образом, чтобы избежать необходимости измерения указанных производных. Невязка *ε(t)* находится для преобразованного уравнения. Некоторые алгоритмы преобразования исходных уравнений объекта рассмотрены во второй главе.

В данном пункте по существу дана классификация показателей качества идентификации в зависимости от вида пространства оптимизации функционала *J(ε)*. В то же время весьма важными признаками классификации являются природа измеряемых сигналов и идентифицируемых параметров, а также объем априорной информации.

В зависимости от первого признака величины $\varepsilon$ и соответственно $J(\varepsilon)$ могут быть детерминированными (неслучайными) или стохастическими (случайными). Во втором случае при решении задачи идентификации необходимо оценивать вероятностные характеристики показателя $\hat{J}(\varepsilon)$. В простейшем случае это математическое ожидание $M\{\hat{J}(\varepsilon)\}$.

Второй признак (объем априорной информации) играет определяющую роль в выборе показателя качества идентификации. Наиболее употребительные показатели и критерии такие как: байесовские, максимума правдоподобия, наименьших квадратов полностью зависят от объема априорной информации об идентифицируемых параметрах и условиях измерений.

### 3.2.2. Ограничения

В 3.1 было введено множество допустимых операторов $C_F$. Наличие такого множества свидетельствует о том, что на выбор оператора модели накладываются определенные ограничивающие условия (ограничения). Эти ограничения могут отражать класс (тип) операторов, условия физической реализуемости модели и другие факторы.

При параметрической идентификации ограничения могут иметь вид равенств

$$g_i(\theta) = 0, \quad i = \overline{1, m_1}, \quad m_1 < n \tag{3.17}$$

или неравенств

$$g_i(\theta) \le 0, \quad i = \overline{1, m_2}. \tag{3.18}$$

В простейшем случае ограничения могут свестись к требованию неотрицательности таких идентифицируемых параметров, как постоянная времени, коэффициент усиления, вес, размеры и т. д.

В качестве ограничений при параметрической идентификации линейных систем могут выступать, например, известные в теории автоматического управления условия устойчивости в форме алгебраических неравенств Рауса-Гурвица, частотные критерии устойчивости, логарифмические частотные критерий устойчивости и т. д.

В результате задача идентификации может быть сведена к задаче отыскания условного экстремума.

Так при минимизируемом функционале близости *J(θ)* и ограничениях в виде равенств на основе метода множителей Лагранжа составляется новый функционал

$$J(\theta, \Lambda) = J(\theta) + \Lambda^T G(\theta), \tag{3.19}$$

где $\Lambda^T = [\lambda_1, ..., \lambda_{m_1}]$ – вектор множителей Лагранжа.

При этом отыскание минимума нового функционала сводится к решению алгебраической системы

$$\nabla_\theta J(\theta, \Lambda) = \nabla_\theta J(\theta) + G_\theta \Lambda = 0,$$

$$\nabla_\lambda J(\theta, \Lambda) = G(\theta) = 0,$$

где

$$G_\theta = \left[ \frac{\partial g_i}{\partial \theta_j} \right], \quad i = \overline{1, m_1}, \quad j = \overline{1, n},$$

$$\nabla_\theta = \left[ \frac{\partial}{\partial \theta_1}, ..., \frac{\partial}{\partial \theta_n} \right].$$

При наличии ограничений типа неравенств условия оптимальности задаются теоремой Куна-Таккера из математического программирования.

## 3.3. Классификация методов идентификации

### 3.3.1. Структурная и параметрическая идентификация

Под методом идентификации будем понимать способ и совокупность приемов построения математической модели объекта.

На современном этапе своего развития теория идентификации характеризуется обилием методов и алгоритмов построения моделей объектов. Единых, универсальных методов, приемлемых для всех типов объектов и их входных и выходных сигналов пока не существует. Это объясняется рядом причин, в т. ч. большим разнообразием типов идентифицируемых объектов и сложностью решаемых задач.

Выбор конкретного метода идентификации определяется многими факторами. К числу основных факторов относятся цель идентификации, тип модели объекта, конкретный вид показателей и критериев качества идентификации (близости), особенности входных и выходных сигналов и измерительных помех, объем априорной информации об объекте и сигналах, возможности имеющихся в распоряжении исследователя вычислительных средств.

По объему априорной информации об объекте различают методы структурной и параметрической идентификации. При структурной идентификации используют методы, позволяющие оценить структуру и параметры модели объекта. В некоторых задачах можно ограничиться только определением структуры модели объекта.

Заметим, что структура, объекта, вообще говоря, может и не совпадать (фактически и не совпадает) со структурой модели. Во многих случаях заведомо ставится задача аппроксимации сложного объекта более простой моделью. Особенно характерно это для построения математических моделей сложных систем и процессов.

При параметрической идентификации выбираются методы, позволяющие оценить параметры при известной (заданной) структуре. Методы параметрической идентификации в отличие от структурной идентификации разработаны более детально, особенно для построения моделей, линейных относительно параметров. Наиболее широко для параметрической идентификации применяется метод наименьших квадратов. В задачах идентификации, решаемых в системах автоматизированной обработки, наиболее типовыми являются задачи параметрической идентификации.

### 3.3.2. Активная и пассивная идентификация

По характеру входного воздействия различают методы активной и пассивной идентификации. Активная идентификация предполагает подачу на вход объекта и модели специальных испытательных (тестовых, пробных) сигналов. При пассивной идентификации измеряются естественные входные и выходные сигналы объекта, имеющие место в процессе нормальной эксплуатации.

При активной идентификации имеются широкие возможности для оптимального планирования всего процесса построения математической модели, включая синтез подходящих входных сигналов. Для этой цели широко привлекаются методы теории планирования эксперимента. Очевидно, что применяя специально синтезированные входные сигналы, можно существенно повысить оперативность процесса построения модели за счет уменьшения интервала наблюдения или же на заданном интервале получить больший объем информации о свойствах объекта.

При активной идентификации существенно упрощаются и сами методы построения моделей. В качестве пробных сигналов при активной идентификации используются сигналы вида импульсной функции, ступенчатой функции, гармонической функции, "белого шума" и т. д.

Однако реализовать процесс активной идентификации удается не всегда. Во-первых, значительная часть объектов не допускает дополнительного искусственного раскачивания, т. к. при этом может произойти потеря устойчивости, выход на недопустимые режимы и т. п. Во-вторых, многие объекты в режиме натурных испытаний и нормального функционирования оказываются постоянно под воздействием управляющих и внешних возмущений. При этом из выходного сигнала объекта трудно выделить составляющую, вызванную пробными воздействиями. В-третьих, формирования тестовых сигналов и искусственное "раскачивание" объекта требует дополнительных затрат энергии.

В связи с указанным, в реальных условиях при идентификации таких объектов с помощью автоматизированных систем обработки информации применяют методы пассивной идентификации.

Иногда прибегают к компромиссному варианту – к смешанным методам идентификации. При этом на естественные входные сигналы накладывают, если это возможно, "малый" пробный сигнал, например

сигнал типа "белого шума" небольшой интенсивности, или гармонические сигналы малой амплитуды и т.д. В дальнейшем будут излагаться в основном методы пассивной идентификации.

### 3.3.3. Адаптивные и неадаптивные методы идентификации

Запишем показатель среднего риска $J(\hat{\theta})$ в виде

$$J(\hat{\theta}) = M_x\{Q(x,\hat{\theta})\},$$

где

$$M_x\{Q(x,\theta)\} = \int_{Cx} Q(x,\hat{\theta})p(x)dx. \qquad (3.20)$$

Допустим, что плотность распределения $p(x)$ неизвестна, а априори известны лишь реализации величины $Q(x,\hat{\theta})$, определяемые по результатам измерений выходных сигналов $x$ и $\hat{x}$.

Если функционал $J(\hat{\theta})$ дифференцируем, то минимум его определяется условием

$$\nabla_\theta J(\hat{\theta}) = 0, \qquad (3.21)$$

где $\nabla_\theta$ – градиент функционала $J(\hat{\theta})$

$$\nabla_\theta J(x,\hat{\theta}) = \left\{\frac{\partial J}{\partial \hat{\theta}_1},\ldots,\frac{\partial J}{\partial \hat{\theta}_n}\right\}.$$

С учетом (3.22) необходимое условие минимума функционала $J(\hat{\theta})$ можно записать в форме

$$\nabla_{\hat{\theta}} J(\hat{\theta}) = M_x\{\nabla_{\hat{\theta}} Q\ (x,\hat{\theta})\} = 0, \qquad (3.22)$$

где

$$\nabla_{\hat{\theta}} Q(x,\hat{\theta}) = \left\{\frac{\partial Q}{\partial \hat{\theta}_1},\ldots,\frac{\partial Q}{\partial \hat{\theta}_n}\right\}.$$

В (3.22) градиент функционала $J(\hat{\theta})$ неизвестен, а известны лишь реализации $\nabla_{\hat{\theta}} Q(x,\hat{\theta})$.

Тогда для получения оценки параметра $\hat{\theta}$ на t-ом шаге можно использовать следующий алгоритм адаптации

$$\hat{\theta}(t) = \hat{\theta}(t-1) - \tilde{A}(t)\nabla_{\hat{\theta}} Q(x(t),\hat{\theta}(t-1)), \qquad (3.23)$$

или в суммарной форме

$$\hat{\theta}(t) = \hat{\theta}(0) - \sum_{i=1}^{t} \tilde{A}(i) \nabla_{\hat{\theta}} Q\left(x(i), \hat{\theta}(t-1)\right). \qquad (3.24)$$

В приведенном алгоритме *Г(t)* – матрица, определяющая величину очередного шага итеративного процесса и зависящая в общем случае от номера шага и от значений оценок вектора $\hat{\theta}$ на предыдущих шагах. Различные формы алгоритмов адаптации вида (3.23) отличаются друг от друга конкретным выбором матрицы *Г(t)*.

Характерной особенностью приведенного адаптивного алгоритма параметрической идентификации является накопление и немедленное использование текущей информации для устранения неопределенности из-за недостаточной априорной информации с целью минимизации показателя близости $J(\hat{\theta})$.

В противном случае метод идентификации будем называть неадаптивным.

### 3.3.4. Другие методы идентификации

В зависимости от темпа использования получаемой апостериорной информации различают методы идентификации в реальном масштабе времени (оперативная идентификация) и методы ретроспективной (послеопытной) идентификации. Оперативная идентификация необходима, в первую очередь, в системах управления. В современной теории идентификации наиболее изучены методы ретроспективной идентификации. Основная особенность разрабатываемых оперативных методов заключается в том, что их алгоритмы строятся на базе рекуррентных формул, позволяющих уточнять оценки структуры или параметров модели по мере получения каждой порции информации (данных) о входных и выходных сигналах объекта.

По степени привлечения к построению математической модели математической статистики различают статистические и нестатистические методы. Первая группа методов характеризуется вычислением и использованием оценок законов распределения или их числовых характеристик (например, корреляционных моментов или функций, спектральных плотностей). При использовании методов второй группы непосредственно методы математической статистики не используются. Чаще всего входные и выходные сигналы объекта подвергаются функциональным преобразованиям, сглаживающим высокочас-

тотные помехи, накладываемые на измеряемые сигналы. При этом процессы, протекающие в исследуемом объекте, могут рассматриваться как детерминированные.

Кроме указанного, методы идентификации могут делиться на замкнутые и разомкнутые. Общая схема замкнутого метода показана на рис. 3.2. Замыкание заключается в сравнении выходных сигналов объекта и модели и в перестройке структуры или параметров модели при минимизации показателя близости (адекватности) $J(x, \hat{x})$. Но иногда необходимость в сравнении выходных сигналов модели и объекта и в проверке адекватности модели отпадает. Это характерно обычно для активных методов идентификации, когда при подаче на вход объекта специальных тестовых сигналов по его реакции непосредственно определяются частотные, импульсные, переходные или другие типовые характеристики. От замкнутых и разомкнутых методов идентификации следует отличать методы построения моделей объектов, включенных в замкнутые и разомкнутые контуры (цепи) управления. Упрощенные схемы таких включений показаны на рис. 3.3 и 3.4. Идентификация в замкнутом контуре значительно сложнее, чем в разомкнутом. Размыкание же цепи обратной связи в условиях нормального функционирования обычно недопустимо из-за возможного нарушения режима работы (потеря устойчивости, недопустимые колебания и т. д.).




Рис. 3.3. Упрощенная схема разомкнутого контура

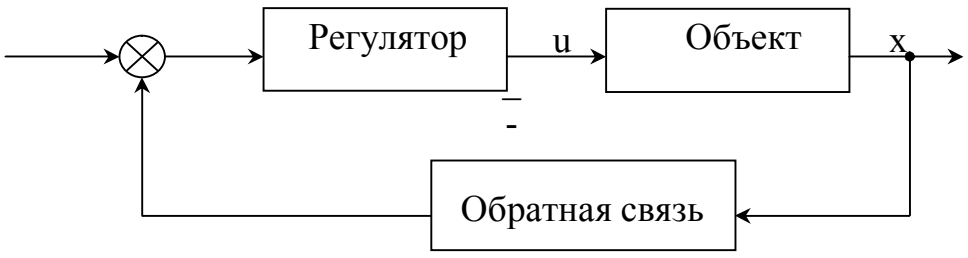



Рис. 3.4. Упрощенная схема замкнутого контура

Выше приведена классификация методов идентификации по наиболее общим признакам, характеризующим объект и условия его функционирования. На практике, как уже было указано, наличие массы более частных признаков привело к появлению огромного числа методов идентификации, отличающихся вычислительными схемами, корректностью, представлением входных и выходных сигналов объекта, способом учета начальных условий и т. д. В связи с этим пока еще не удается строго формализовать возможное множество методов и разработать научно обоснованные процедуры целенаправленного синтеза оптимальных методов. Можно лишь попытаться составить перечень основных разумных требований, которые должны учитываться при выборе того или иного метода построения математической модели объекта. Так, для реализации в системах автоматизированной обработки информации методов идентификации объектов по результатам измерения их (объектов) входных и выходных сигналов необходимо учитывать следующие рекомендации:

1. Идентификация должна осуществляться в основном в режиме нормальной эксплуатации объекта;

2. Методы идентификации должны быть быстродействующими из-за ограниченности времени наблюдения и времени на принятие решения по результатам идентификации;

3. Методы должны быть ориентированы на реализацию на ЭВМ, как правило, на универсальных;

4. Методы должны быть относительно простыми в реализации на современных ЭВМ;

5. Методы должны обладать достаточной точностью;

6. Они должны обеспечивать идентификацию объектов, находящихся как в разомкнутых, так и в замкнутых цепях;

7. Методы должны быть некритичными к помехам.

## 3.4. Об условиях идентифицируемости объектов

### 3.4.1. Проблема идентифицируемости

Теория и практика построения моделей объектов показывает, что не во всех случаях структура и параметры объекта могут быть восстановлены. Принципиальная возможность построения модели объекта на

основании имеющейся априорной информации и данных о входных и выходных переменных характеризуется понятием идентифицируемости.

Качественные и количественные требования к процессам в системе (входные и выходные сигналы, переменные состояния), а также к структурным характеристикам объекта и обратной связи, если рассматривается замкнутая система, формируют условия идентифицируемости.

К настоящему времени универсальные понятия и условия идентифицируемости применительно к широкому классу динамических объектов и их моделей не существует. Предложено и разрабатывается несколько подходов, базирующихся на идеях и методах теории управления и теории оценивания. Так как методы идентификации в широком смысле (структурная идентификация) развиты недостаточно, то в большинстве из известных подходов структура модели предполагается известной.

Группа понятий и условий идентифицируемости, развитая на методологии теории управления, не зависит от метода идентификации и оперирует результатами точных измерений входных и выходных сигналов идентифицируемого объекта. Эти условия идентифицируемости аналогичны условиям управляемости и наблюдаемости и непосредственно связаны с ними.

Понятия управляемости и наблюдаемости относят в теории управлении к фундаментальным системным понятиям. Поэтому по аналогии с этим, указанные понятия и условия идентифицируемости будем называть (возможно, несколько условно) системными.

Группа понятий и условий идентифицируемости, опирающаяся в основном на теорию оценивания, связана, как правило, с рассмотрением зашумленных входных и выходных сигналов и в значительной мире зависит от выбранных критериев и метода идентификации. При этом зачастую предполагается также разрешимость задачи идентификации по точным измерениям входных и выходных сигналов.

Эта группа понятий и условий идентифицируемости формулируется, как правило, с использованием аппарата теории вероятностей и математической статистики. В связи с этим данные понятия и условия будем называть вероятностными.

Более или менее завершенные исследования проблемы идентифицируемости получены пока только для линейных систем.

### 3.4.2. Системные условия идентифицируемости разомкнутых систем

В наиболее общем виде линейная разомкнутая система описывается уравнениями

$$\begin{aligned} x(t+1) &= Ax(t) + Bu(t), \qquad x(0) = x_0, \\ y(t) &= Cx(t), \end{aligned} \tag{3.25}$$

где $x$ – $n$-мерный вектор переменных состояния (фазовых координат); $u$-$m$-мерный вектор управления; $y$ – $s$-мерный вектор наблюдения; $A, B, C$ – постоянные матрицы соответственно размерностей $nxn, nxm, sxp$.

Рассмотрим вначале систему (объект), описываемую уравнением

$$x(t+1) = Ax(t), \qquad x(0) = x_0. \tag{3.26}$$

Предполагается, что вектор $x(t)$ доступен измерению (наблюдению). Требуется найти условия идентифицируемости матрицы $A$ на основе измерений $x(t)$.

Допустим, что произведены измерения $x(0), x(1), ..., x(n)$.

Составим соотношения

$$\begin{aligned} &x(1) = Ax(0), \\ &x(2) = Ax(1) = A^2x(0), \\ &\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ &x(n) = Ax(n-1) = A^n x(0) \end{aligned} \tag{3.27}$$

Перепишем (3.27) в виде

$$[x(1),...,x(n)] = A[x(0), x(1)..., x(n-1)] = [Ax(0), A^2x(0)..., A^{n-1}x(0)]. \tag{3.28}$$

Очевидно, что для однозначного определения из матричного уравнения (3.28) матрицы А необходимо, чтобы матрица

$$R = [x(0),..., x(1)..., x(n-1)] = [x(0), A^2x(0)..., A^{n-1}x(0)] \tag{3.29}$$

была неособенной.

Назовем матрицу (3.29) матрицей $(0, x_n)$ идентифицируемости. В данном определении $0$ означает отсутствие входного сигнала (исходная система является однородной, она возбуждается только ненулевым начальным условием), а $x_n$ означает, что измеряется вектор состояния $x$ размерности $n$.

Определение. Система (3.26) называется $(0, x_n)$ –идентифицируемой, – если матрица $R$ является неособенной.

Физически условие (*0, $x_n$*) – идентифицируемости означает, что начальное значение *x(0)* возбуждает все колебания системы.

Недостатком условия (*0, $x_n$*) – идентифицируемости является то, что оно связано с вектором *x(t)* состояния, который обычно недоступен измерениям. Поэтому большой практический интерес представляет получение условия идентифицируемости системы вида

$$\begin{aligned} x(t+1) &= Ax(t), \qquad x(0) = x_0, \\ y(t) &= c^T x(t), \end{aligned} \tag{3.30}$$

где *y(t)* – скалярная наблюдаемая величина; с – вектор наблюдения.

Пусть произведены измерения *y(1), y(2), .... y(n)* и составлены соотношения

$$\begin{aligned} y(1) &= c^T x(1) = c^T Ax(0), \\ y(2) &= c^T x(2) = c^T A^2 x(0), \\ &\cdots\cdots\cdots\cdots \\ y(n) &= c^T x(n) = c^T A^n x(0). \end{aligned} \tag{3.31}$$

Запишем (3.31) в виде

$$s_n = QAx(0), \tag{3.32}$$

где

$$s_n = \begin{bmatrix} y(1) \\ y(2) \\ . \\ y(n) \end{bmatrix}, \quad Q = \begin{bmatrix} c^T \\ c^T A \\ . \\ c^T A^{n-1} \end{bmatrix}.$$

Составим соотношения вида (3.32) для измерений *x(2), x(3), …x(n+l)*

$$\begin{aligned} y(2) &= c^T x(2) = c^T Ax(1), \\ y(3) &= c^T x(3) = c^T A^2 x(1), \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ y(n+1) &= c^T x(n+1) = c^T A^n x(1) \end{aligned}$$

или в матричной форме

$$s_{n+1} QAx(1) = QA^2 x(0),$$

где

$$s_{n+1}^T = [y(2), y(3), \ldots, y(n+1)].$$

В результате можно получить матричное соотношение вида

$$D = \left[s_n, s_{n+1}, \ldots, s_{2n-1}\right] = [\,QAx(0), QA^2x(0), \ldots, QA^nx(0)\,] = QAR, \qquad (3.33)$$

где $R$ – матрица $(0, x_n)$ – идентифицируемости.

В результате из (3.33) находим

$$A = Q^{-1}DR^{-1}. \qquad (3.34)$$

Матрица Q в теории управления называется матрицей наблюдаемости.

Определение. Система (3.30) называется наблюдаемой, если ранг матрицы $Q$ равен $n$.

Физически выполнение условия наблюдаемости означает, что все колебания системы можно наблюдать через переменную $y$.

Из выражения (3.34) следует, что условия идентифицируемости системы (3.30), которое в соответствии с принятым подходом назовем условием $(0, y)$ идентифицируемости, включает одновременно требования наблюдаемости и $(0, x_n)$ – идентифицируемости.

Определение. Система (3.30) называется $(0, y)$ – идентифицируемой, если она одновременно наблюдаема и $(0, x_n)$ – идентифицируема.

Дальнейшее усложнение систем приводит к рассмотрению систем со скалярным управлением и измерением вектора $y\ (t)$

$$x(t+1) = Ax(t) + bu(t), \qquad x(0) = 0. \qquad (3.35)$$

Ставится задача идентификации матрицы $A$ и векторного параметра $b$. Очевидно это задача $(0, x_n)$ – идентификации.

Пусть проведены измерения $x(1), \ldots, x(n+1), u(0), u(1), \ldots, u(n)$, причем $u_0 \neq 0$. Составим соотношения

$$\begin{aligned}
&x(1) = bu(0),\\
&x(2) = Ax(1) + bu(1) = Abu(0) + bu(1),\\
&\ldots\ldots\ldots\ldots\\
&x(n+1) = Ax(n) + bu(n) =\\
&= A^n bu(0) + A^{n-1}bu(1) + Abu(n-1) + bu(n).
\end{aligned} \qquad (3.36)$$

Вектор $b$ может быть определен из первого уравнения системы (3.36) при $u_0 \neq 0$. Для определения матрицы $A$ составим уравнения

$$\begin{aligned}
&Ax(1) = x(2) - bu(1),\\
&Ax(2) = x(3) - bu(2),\\
&\ldots\ldots\ldots\ldots\\
&Ax(n) = x(n+1) - bu(n)
\end{aligned}$$

или в матричной форме

$$A[x(1),..., x(n)] = [x(2) - bu(1),..., x(n+1) - bu(n)]. \qquad (3.37)$$

В результате получаем, что для идентифицируемости матрицы А необходимо, чтобы матрица

$$G_1 = [x(1),..., x(n)] \qquad (3.38)$$

была неособенной.

С учетом соотношений (3.36) матрицу $G_1$ запишем в виде

$$G_1 = [bu(0), Abu(0) + bu(1),..., A^n bu(0) + A^{n-1} bu(1) + ... + bu(n)]$$

или

$$\begin{aligned} G_1 &= u(0),[b, Ab,..., A^n b] + u(1)[0, b, Ab,...A^{n-1}] + \\ &+ u(2)[0,0, b, Ab,..., A^{n-2} b] + ... + u(n)[0,...,0b]. \end{aligned} \qquad (3.39)$$

Анализ матрицы $G_1$ показывает, что неособенность (ранг) определяется свойствами первого слагаемого, т. е. матрицы

$$G = [b, Ab, A^2 b,..., A^n b].$$

В теории управления матрица G называется матрицей управляемости.

Определение. Система (3.35) называется управляемой, если ранг матрицы *G* равен *n*.

Физически выполнение условия управляемости означает, что управление *u* возбуждает все колебания системы (объекта).

Таким образом, получаем следующее условие идентифицируемости системы (3.35).

Определение. Система (3.35) называется $(u, x_n)$ – идентифицируемой при $u(0) \neq 0$ если она является управляемой.

Рассмотрим теперь систему

$$x(t+1) = Ax(t) + Bu(t), \qquad x(0) = x_0, \qquad (3.40)$$

где В – матрица размерности nxm; *u* – вектор управления размерности m.

Пусть проведены измерения *x(0), x(1), ..., x(n+m) и u (0), u(1), ..., u(n+m-1)*. Составим соотношения

$$\begin{aligned} x(1) &= Ax(0) + Bu(0), \\ x(2) &= Ax(1) + Bu(1), \\ &..... \\ x(n+m) &= Ax(n+m-1) + Bu(n+m-1), \end{aligned} \qquad (3.41)$$

которое можно записать в форме

$$\left[x(1), x(2), \ldots, x(n+m)\right] = \left[AB\right] = \begin{bmatrix} x(0), x(1), \ldots, x(n+m-1) \\ u(0), u(1), \ldots, u(n+m-1) \end{bmatrix}. \qquad (3.42)$$

Тогда для определения матрицы $A$ и $B$ из матричного уравнения (3.42) необходимо и достаточно, чтобы матрица

$$G_2 = \begin{bmatrix} x(0), x(1), \ldots, x(n+m-1) \\ u(0), u(1), \ldots, u(n+m-1) \end{bmatrix} \qquad (3.43)$$

была неособенной.

Это условие можно рассматривать в качестве обобщающего условия $(u_m, x_n)$ – идентифицируемости системы (3.40).

С помощью соответствующих преобразований матрицы $G_2$ можно получить условие $(u_m, x_n)$ – идентифицируемости в терминах управляемости, наблюдаемости и $(0, x_n)$ – идентифицируемости. Это условия приведены в работах [10,14,16,18]. Там же выведены условия идентифицируемости, которые обобщают все приводимые выше условия идентифицируемости линейных разомкнутых систем.

### 3.4.3. Системные условия идентифицируемости линейных замкнутых систем

В наиболее общем виде уравнения линейных замкнутых систем можно записать в виде

$$\begin{aligned} x(t+1) &= Ax(t) + Bu(t), \\ u(t) &= \tilde{A}z(t) + y(t), \\ z(t) &= \hat{E}z(t) + \hat{A}y(t), \\ y(t) &= Cx(t), \quad x(0) = x_0, \end{aligned} \qquad (3.44)$$

где первое уравнение описывает объект управления, третье уравнение описывает динамику обратной связи, причем $z(t)$ – $\ell$-мерный вектор состояния звена обратной связи; $K$ – невырожденная матрица переходных состояний звена обратной связи; $B$ – постоянная матрица размерности $\ell$ xs . Постоянная матрица $\Gamma$ имеет размерность $m$x$\ell$ .

Ввиду сложности исследований системы (3.40) ограничимся рассмотрением так называемой свободной замкнутой системы вида –

$$\begin{aligned} x(t+1) &= Ax(t) + Bu(t), \\ u(t) &= \tilde{A}z(t), \\ z(t) &= \hat{E}z(t) + \hat{A}y(t), \\ y(t) &= x(t), \qquad x(0) = x_0. \end{aligned} \tag{3.45}$$

Требуется, как и ранее, получить условия идентифицируемости матриц $A$ и $B$ объекта при измерении векторов $u$ и $x$, являющихся соответственно входными и выходными сигналами объекта.

Условия $(u_m, x_n)$ – идентифицируемости замкнутой системы получаются по аналогии с получением условий $(u_m, x_n)$ – идентифицируемости разомкнутой системы. С использованием измерений *x(0), x(1),...,x(n+m), u(0), u(1),...,u(n+m-1)* составляются соотношения

$$\begin{aligned} x(1) &= Ax(0) + Bu(0), \\ x(2) &= Ax(1) + Bu(1), \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ x(n+m) &= Ax(n+m-1) + Bu(n+m-1) \end{aligned}$$

или

$$[x(1), x(2), \ldots, x(n+m)] = [AB] = \begin{bmatrix} x(0), & x(1), \ldots, x(n+m-1) \\ u(0), & u(1), \ldots, u(n+m-1) \end{bmatrix}.$$

При этом условие $(u_m, x_n)$ – идентифицируемости имеет вид

$$det \begin{bmatrix} x(0), & x(1), \ldots, x(n+m-1) \\ u(0), & u(1), \ldots, u(n+m-1) \end{bmatrix} \neq 0. \tag{3.46}$$

Детальные исследования условия (3.46) являются достаточно громоздкими, поэтому они здесь не даются. Приведем лишь без вывода некоторые следствия из общей теоремы о $(u_m, x_n)$ – идентифицируемости замкнутой системы (3.45).

Следствие 1. Необходимым условием разрешимости задачи $(u_m, x_n)$ -идентифицируемости является выполнение неравенства $\ell \geq m$, т. е. размерность звена обратной связи должна быть не меньше размерности вектора входного сигнала объекта. Отсюда, в частности, следует, что для объектов с одним входом минимально допустимая размерность цепи обратной связи равна единице.

Следствие 2. При $x(0) = 0$ для разрешимости задачи $(u_m, x_n)$ – идентификации необходимо, чтобы объект был управляемым.

Следствие 3. При $x(0)=0$ управляемость объекта является необходимым и достаточным условием разрешимости задачи $(u_m, x_n)$ – идентификации.

Для стохастического случая детерминистическая идентифицируемость является необходимой, но недостаточной для стохастической (вероятностной) идентифицируемости, когда вход искажается шумами. Поэтому в последнем случае идентифицируемость следует понимать как принципиальную возможность восстановления параметров в некотором статистическом смысле.

При рассмотрении качественной стороны идентифицируемости ставится задача принципиальной возможности получения состоятельных оценок. В этом смысле понятие идентифицируемости непосредственно можно связать с понятием взаимно исключающих классов. Количественная сторона идентифицируемости заключается в изучении различных свойств характеристик процессов оценивания с использованием тех или иных методов и процедур при решении задач идентификации. Для последующего изложения задач стохастической идентифицируемости систем, оценивания параметров представим некоторые понятия теории вероятности и математической статистики.

## 3.5. Элементы теории вероятности и математической статистики

### 3.5.1. Введение

Теория вероятностей и математическая статистика, несмотря на то, что в действительности это две самостоятельные области науки, знания часто рассматривают как одно целое, и поэтому в научных монографиях, учебниках и курсах лекции излагают совместно. Однако, хотя математическая статистика в значительной мере и опирается на понятия теории вероятностей, её определяют как науку, когда говорят о накоплении, систематизации и анализе фактов, или экспериментальных данных. С таким подходом представим некоторые положения теории вероятностей и математической статистики.

Экспериментальное исследование является одним из важнейших инструментов познания явлений природы и присущих им закономерностей.

Любое явление природы есть результат взаимодействия множества факторов. Если в ходе эксперимента зафиксировать значения всех этих факторов, то тем самым будет установлен *полный комплекс условий*, при которых проводится эксперимент (наблюдение). Фиксируя различные полные комплексы условий, в конечном итоге можно получить строго детерминированную зависимость, связывающую результаты наблюдений над изучаемым явлением со значениями рассматриваемых факторов. Однако на практике учесть все факторы, оказывающие то или иное воздействие на свойства изучаемого явления, невозможно. При этом ограничиваются заданием лишь *основного комплекса условий*, т. е. включением в рассмотрение факторы, в наибольшей степени влияющие на данное явление, относя действие остальных к случайным помехам.

Таким образом, чем сложнее объект исследования, чем больше разнородных и взаимозависимых факторов оказывают на него свое воздействие, чем менее ярко проявляют себя основные доминирующие факторы, тем сильнее сказывается влияние помех. В такой ситуации результаты наблюдений будут отличаться друг от друга (и может быть, весьма значительно) даже при неизменном основном комплексе условий. Точное предсказание этих результатов при фиксированных значениях основных факторов оказывается невозможным, поскольку результаты наблюдений будут меняться от опыта к опыту некоторым случайным образом. При этом может выясниться, что, хотя результат отдельного эксперимента нельзя предсказать с достаточной точностью, совокупность наблюдений подчиняется определенным неслучайным закономерностям. Обычно эти закономерности проявляются тем отчетливее, чем больше число наблюдений при данном основном комплексе условий реализовано. В этом случае можно найти модель, описывающую изучаемое явление и взаимозависимости между его отдельными сторонами в среднем, т. е. построить *статистическую модель* явления.

Теория вероятностей изучает объективные закономерности массовых случайных явлений, проявляющиеся при многократном воспроизведении заданного основного комплекса условий. С каждым из случайных явлений связаны определенные *события*, которые могут осуществляться или не осуществляться в результате опыта. Событие, которое при заданном основном комплексе условий:

- обязательно произойдет, называется *достоверным*;
- не может произойти, называется *невозможным*;
- может произойти, а может и не произойти, называется *случайным*.

Количественной мерой объективной возможности осуществления события при фиксированном основном комплексе условий является *вероятность* этого события.

Во многих практических ситуациях *качественное* описание случайных явлений в терминах событий, когда отмечается лишь факт их наличия или отсутствия, оказывается недостаточным. По этой причине целесообразнее представлять результаты опытов *количественно* в виде некоторой *случайной величины.* Случайная величина есть величина определенной физической размерности, принимающая в результате эксперимента то или иное числовое значение, которое в принципе нельзя предсказать, исходя из основного комплекса условий проведения эксперимента.

Способ количественного определения вероятностей попадания случайной величины в произвольные части области ее возможных значений может быть задан с помощью *закона распределения вероятностей* случайной величины. Закон распределения вероятностей каждому возможному значению дискретной случайной величины или некоторому интервалу значений непрерывной случайной величины ставит в соответствие вероятность того, что случайная величина примет это возможное значение или попадет в данный интервал [7,13, 34].

Аналитическим выражением законов распределения служат *функции распределения вероятностей*.

### 3.5.2. Функции распределения вероятностей случайной величины

Интегральной функцией распределения вероятностей случайной величины x называется функция

$$F(x) = P\{x<x\}, \qquad (3.47)$$

показывающая зависимость вероятности того, что случайная величина $x$ не превышает некоторый уровень $x$ от величины этого уровня.

Отметим основные свойства интегральной функции распределения:

1) $F(-\infty) = P\{x < -\infty\}=0$ – как вероятность невозможного события;

2) $F(+\infty) = P\{x < +\infty\}=1$– как вероятность достоверного события;

3) $0 \le F(x) \le 1$ – для всех $x$;

4) $F(x)$ неубывающая функция $x$; т. е. $F(x_2) \ge F(x_1)$, если $x_2 > x_1$.

С помощью интегральной функции распределения легко определить вероятность попадания случайной величины в любую часть области ее возможных значений. В частности, вероятность попадания в интервал с граничными значениями a, *b* равна разности значений функции распределения, вычисленных в этих двух точках, т. е.

$$P\{a<\mathrm{x}<b\} = F(b) - F(a). \qquad (3.48)$$

Аналогично

$$P\{\mathrm{x}>a\} = 1 - F(a). \qquad (3.49)$$

В большинстве практических задач предполагается, что интегральные функции распределения непрерывных случайных величин дифференцируемы во всей области их возможных значений. Тогда закон распределения вероятностей может быть описан в аналитической форме с помощью *дифференциальной функции распределения вероятностей p(x):*

$$p(x) = \frac{dF(x)}{dx} = \lim_{\Delta x \to 0} \frac{P\{x < \mathrm{X} < x + \Delta x\}}{\Delta x};\ \Delta x > 0. \qquad (3.50)$$

Если $\Delta x$ – достаточно малая величина, то

$$p(x) \cong \frac{P\{x < \mathrm{X} < x + \Delta x\}}{\Delta x}, \qquad (3.51)$$

т. е. функция *p(x)* равна отношению вероятности попадания случайной величины внутрь интервала $(x, x + \Delta x)$ к длине этого интервала $\Delta x$. Поэтому зачастую эту функцию *называют функцией плотности распределения вероятностей или функцией плотности вероятности*.

Укажем основные свойства функции *p(x):*

$$p(x) \ge 0;\ \int_{-\infty}^{\infty} p(x)dx = 1;\ \int_{-\infty}^{\infty} p(x)dx = F(x). \qquad (3.52)$$

С помощью дифференциальной функции распределения также нетрудно найти вероятность нахождения случайной величины внутри любой части из области ее возможных значений. Например,

$$P\{a_1<\mathrm{x}<a_2\} = \int_{-\infty}^{a_2} p(x)dx, \qquad (3.53)$$

$$P\{\mathrm{x}<a\} = \int_{-\infty}^{a} p(x)dx; \quad P\{\mathrm{x}>b\} = \int_{b}^{\infty} p(x)dx.$$

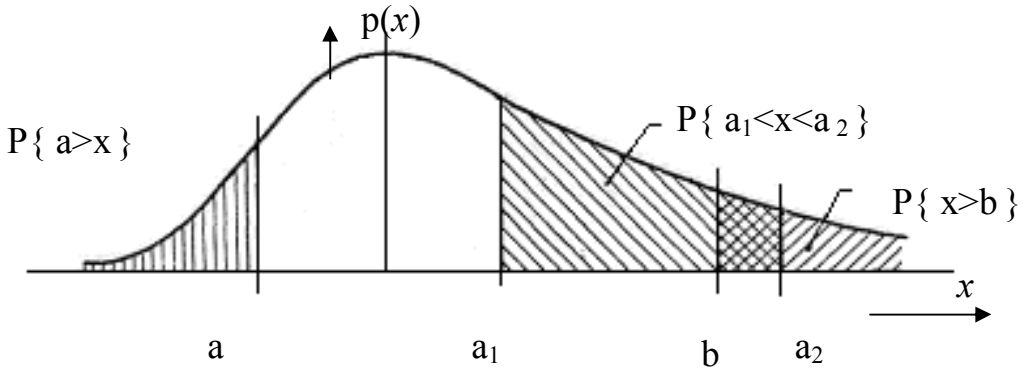



Рис. 3.5. Возможный вид дифференциальной функции распределения вероятностей

Для непрерывной случайной величины вероятность можно определить как относительную площадь под кривой плотности распределения p*(x).* На рис. 3.5 изображена некоторая функция *p(x);* штриховкой отмечены участки, соответствующие вероятностям того, что $x < a,\ x > b,\ a_1 < x < a_2$.

Отметим, что вероятность события, заключающегося в том, что непрерывная случайная величина примет какое-то конкретное заранее заданное значение равное нулю:*P{x=a}=0*, хотя это, конечно, возможное событие.

### 3.5.3. Числовые характеристики случайных величин и их функций распределения

Интегральная и дифференциальная функции распределения являются исчерпывающими вероятностными характеристиками случайной величины. Однако зачастую достаточно указать лишь некоторые *числовые параметры*, характеризующие отдельные существенные черты распределения случайной величины.

В основу получения большинства числовых характеристик случайной величины положено понятие *математического ожидания* функций случайной величины.

Если x – случайная величина, то произвольная функция *g(*x*)* случайной величины x также является случайной величиной. Тогда математическое ожидание этой функции, обозначаемое M{(g(x)} определяется соотношением:

$$M\{g(x)\} = \int_{-\infty}^{\infty} g(x)\,p(x)dx. \qquad (3.54)$$

Математическое ожидание $M\{g(x)\}$ есть результат *вероятностного осреднения* функции $g(x)$, т. е. осреднения ее с весом, равным плотности вероятности $p(x)$. Математическое ожидание $M\{g(x)\}$ для заданного вида функции $g(x)$ и плотности $p(x)$ есть некоторое неслучайное число.

Воспользовавшись понятием математического ожидания, укажем ряд числовых характеристик случайной величины:

а) начальный момент $k$-го порядка $m_k\{x\}$ есть математическое ожидание функции $g(x) = x^k$:

$$m_k\{x\} = m\{x^k\} = \int_{-\infty}^{\infty} x^k\,p(x)dx. \qquad (3.55)$$

В частности, если $k=1$, получаем *математическое ожидание* или среднее значение самой случайной величины x

$$m_1\{x\} = M\{x\} = \int_{-\infty}^{\infty} xp(x)dx = m, \qquad (3.56)$$

б) центральный момент $k$-го порядка $\mu_k\{x\}$ есть математическое ожидание функции $g(x)=(x - m_x)^k$:

$$\mu_k\{x\} = M\{(x - m_x)^k\} = \int_{-\infty}^{\infty} (x - m_x)^k\,p(x)dx. \qquad (3.57)$$

Второй центральный момент $\mu_k\{x\}$ называется дисперсией случайной величины x и обозначается $\sigma_x^2$ или $D\{x\}$:

$$D\{x\} = M\{(x - m_x)^2\} = \int_{-\infty}^{\infty} (x - m_x)^2\,p(x)dx = \sigma_x^2. \qquad (3.58)$$

Легко показать также, что

$$\sigma_x^2 = m_2\{x\} - m_x^2. \qquad (3.59)$$

Величину $\sigma_x = \sqrt{\sigma_X^2}$ называют средним квадратическим отклонением случайной величины.

Математическое ожидание $m_x$ и среднее квадратическое отклонение $\sigma_x$ (или дисперсия $\sigma_x^2$) являются наиболее употребительными числовыми характеристиками случайной величины. Эти характеристики полностью определяют одну из наиболее важных функций рас-

пределения, а именно, так называемое нормальное распределение. Однако и для других распределений указанные характеристики дают возможность в первом приближении оценить характер функций распределения, указывая центр рассеяния случайной величины ($m_x$) и степень этого рассеяния ($\sigma_x$ или $\sigma_x^2$).

Для характеристики вариабельности случайной величины иногда используется коэффициент вариации $\delta$:

$$\delta = \sigma_x / m_x. \qquad (3.60)$$

Значения $m_x$ и $\sigma_x$ используются для центрирования и нормирования случайной величины x с помощью преобразований

$$\overset{\circ}{x} = x - m_x, \quad \overset{\circ}{x}_0 = \frac{x - m_x}{\sigma_x}.$$

При этом величина $\overset{\circ}{x}$ называется центрированной, а $\overset{\circ}{x}_0$ нормированной (стандартизованной) случайной величиной.

Очевидно, что $m_{\overset{\circ}{x}} = 0;\ m_{\overset{\circ}{x}0} = 0;\ \sigma^2_{\overset{\circ}{x}0} = 1$.

Наряду с указанными моментными числовыми характеристиками для описания вида функций распределения используются и немоментные характеристики. В частности, в качестве характеристики центра рассеяния случайной величины может применяться медиана *Me*, а в качестве характеристики степени рассеяния случайных величин, ограниченных с обеих сторон, размах $\Delta$. Медианой *Me* называется такое значение случайной величины, для которой $P\{x < Me\} = P\{x > Me\} = ½$, т. е.

$$\int_{-\infty}^{Me} p(x)dx = \int_{Me}^{\infty} p(x)dx = F\{Me\} = \frac{1}{2}.$$

Для симметричного относительно математического ожидания $m_x$ распределения $Me=m_x$.

Размах $\Delta$ ограниченной с обеих сторон случайной величины есть разность между ее предельными значениями – максимальным и минимальным: $\Delta = x_{max} - x_{min}$.

### 3.5.4. Многомерные распределения

Когда исследуется совокупность n случайных величин, то их вероятностные свойства можно описать с помощью n-мерной совместной функции распределения – интегральной $F(x_1,x_2,...,x_n)$ или дифференциальной $p(x_1,x_2,...,x_n)$.

Вектор $\boldsymbol{X}$ является случайным вектором

$$\boldsymbol{X}=[x_1,x_2,...x_n]^T,$$

если его компоненты $x_i$ $(i=\overline{1,n})$ являются случайными величинами. Случайные векторы можно полностью охарактеризовать функцией распределения вероятностей, которая определяется следующим образом:

*Определение.* Совместная функция распределения вероятностей вектора $\boldsymbol{X}$ определяется выражением

$$F(x_1,x_2,...,x_n)=P\{x_1\le x_1,\ x_2\le x_2,...x_n\le x_n\}, \qquad (3.61)$$

где $P\{A\}$ – вероятность события А. Для удобства выражение (3.61) часто будет записываться в виде

$$F(X)=P\{\boldsymbol{X}\le X\}.$$

Совместная функция распределения вероятностей является монотонной, неубывающей функцией по каждому из аргументов. Кроме того, из определения следует, что

$$F(-\infty,-\infty,....,-\infty)=0,$$
$$F(+\infty,+\infty,....,+\infty)=1.$$

*Определение.* Совместная плотность вероятности вектора X определяется

$$p(X)=\lim_{\substack{\Delta x_1\to 0\\ \cdot\\ \cdot\\ \Delta x_n\to 0}}\frac{P\{x_1<X_1\le x_1+\Delta x_1,...,x_n<X_n\le x_n+\Delta x_n\}}{\Delta x_1\Delta x_2...\Delta x_n}= \qquad (3.62)$$
$$=\partial^n F(X)/\partial x_1...\partial x_n.$$

Это определение представляет собой прямое обобщение одномерного случая.

Обратно функцию распределения вероятностей можно выразить через плотность вероятности следующим образом:

$$F(X)=\int_{-\infty}^{x} p(X)dX=\int_{-\infty}^{x_1}...\int_{-\infty}^{x_n} p(X)dx_1dx_2...dx_n, \qquad (3.63)$$

где $\int_{-\infty}^{X} \{\cdot\}\, dX$ – сокращенное обозначение *n* – кратного интеграла. Для вычисления вероятности совместная плотность вероятности *f(X)* должна быть умножена на объем $\Delta x_1 \Delta x_2 ... \Delta x_n$ (или $\Delta X$ ).

Условные плотности вероятности. В задачах идентификации, распознавания образов наиболее часто встречаются два вида условных плотностей вероятности. Они различаются типами событий, от которых зависит случайный вектор.

Условная плотность вероятности при фиксированном случайном событии.

*Определение.* Условная функция распределения вероятностей вектора *X* при фиксированном событии $\omega$ определяется следующим образом:

$$F(X/\omega) = P\{\boldsymbol{X} \le X/\omega\}. \tag{3.64}$$

где $P\{A/B\}$ – условная вероятность события А определенная в предположении, что произошло событие В. Эта вероятность равна

$$P\{A/B\} = P\{A \cdot B\} / P\{B\}, \tag{3.65}$$

где $A \cdot B$ – совместное событие *A* и *B*.

*Определение.* Условная плотность вероятности вектора *X* при фиксированном событии $\omega$ определяется следующим образом:

$$p(X/\omega) = \partial^n F(X/\omega) / \partial x_1 \partial x_2 ... \partial x_n. \tag{3.66}$$

Если предположить, *что достоверное событие* $\Omega$ *состоит из m независимых событий* $\omega_1, \omega_2, ..., \omega_m$ *, то безусловная плотность вероятности вектора X* имеет вид

$$p(X) = \sum_{i=1}^{m} p(X/\omega_i) F(\omega_i). \tag{3.67}$$

Здесь $F(\omega_i)$ используется вместо $P\{\omega_i\}$.

*Условная плотность вероятности при фиксированной случайной величине.*

*Определение.* Условная плотность вероятности вектора *X* при фиксированном значении *Y* случайного вектора *Y* определяется выражением

$$p(X/Y) = \lim_{\Delta Y \to 0} p(X/Y < \boldsymbol{Y} \le Y + \Delta Y). \tag{3.68}$$

Ясно, что (3.68) является частным случаем условной плотности вероятности, определенной выше. Здесь событие $\omega$ заменено событием $(Y < \boldsymbol{Y} \le Y + \Delta Y)$, а $\Delta Y \to 0$.

Легко показать, что f(X/Y) можно выразить в более удобном виде

$$p(X/Y) = p(X,Y)/p(Y), \qquad (3.69)$$

где $p(X,Y)$ – совместная плотность вероятности всех компонент векторов $X$ и $Y$, которая, кроме того, может быть записана как

$$p(X,Y) = f(x_1, x_2, ..., x_n, y_1, y_2, ..., y_k). \qquad (3.70)$$

Выражение $p(Y)$ представляет собой *маргинальную плотность вероятности* вектора $Y$ и получается интегрированием совместной плотности вероятности

$$p(Y) = \int_{\Omega} p(X,Y)dX. \qquad (3.71)$$

Теорема. *Условные плотности вероятности p(X/Y) и p(Y/X) связаны между собой следующим образом:*

$$p(X/Y) = \frac{p(Y/X)p(X)}{p(Y)}. \qquad (3.72)$$

Это *теорема Байеса*, которая легко доказывается, если в выражении (3.69) поменять местами $X$ и $Y$. Теорема Байеса играет важную роль в теории оценивания. Замена $Y$ событием $\omega_i$ дает следующее обобщение этой теоремы:

$$p(X/\omega_1) = \frac{p(\omega_1/X)p(X)}{p(\omega_1)}. \qquad (3.73)$$

Это соотношение является основным в теории проверки гипотез.

*Плотности вероятности, зависящие от неслучайных параметров.*

Иногда можно выразить в явной форме зависимость плотности вероятности случайного вектора $X$ от неслучайного вектора параметров $\theta$. Обозначим такую зависимость через $p(X/\theta)$. Вектор $\theta$ является неслучайным вектором; однако его значение может быть неизвестно.

Пример. Пусть $x$ – нормально распределенная случайная величина с математическим ожиданием $m$. Тогда можно записать плотность вероятности x в виде

$$p(x/m) = (2\pi)^{-1/2}\ \sigma^{-1} exp\left\{-\frac{1}{2}(x-m)^2/\sigma^2\right\}.$$

Этот тип плотности вероятности играет важную роль в задаче оценивания неслучайных параметров.

Параметры распределения. Случайный вектор $\boldsymbol{X}$ полностью характеризуется его функцией распределения вероятностей или плотностью вероятности. Однако часто эти функции не могут быть легко

определены или их математические выражения оказываются слишком сложными для использования на практике. Поэтому иногда предпочтительнее воспользоваться менее полными, но проще вычисляемыми характеристиками случайного вектора.

*Вектор математического ожидания.*

Одной из наиболее важных числовых характеристик случайного вектора $\boldsymbol{X}$ является вектор математического ожидания или среднее значение случайного вектора $\boldsymbol{X}$.

*Определение.* Вектор математического ожидания случайного вектора $\boldsymbol{X}$ определяется выражением

$$q = M\{X\} = \int_{\Omega} Xp(X)dX, \tag{3.74}$$

где областью интегрирования $\Omega$ является все пространство векторов $\boldsymbol{X}$.

Значение $i$-й компоненты $q_i$, вектора $q$ можно вычислить по формуле

$$q_i = \int_{\Omega} x_i p(X)dX = \int_{-\infty}^{+\infty} x_i p(x_i)dx_i, \tag{3.75}$$

где $p(x_i)$ – *маргинальная плотность вероятности* $i$-го компонента вектора $\boldsymbol{X}$:

$$p(x_i) = \underbrace{\int_{-\infty}^{+\infty}\ldots\int_{-\infty}^{+\infty}}_{n-1} p(X)dx_1\ldots dx_{i-1}\,dx_{i+1}\ldots dx_n. \tag{3.76}$$

Таким образом, каждая компонента вектора q действительно вычисляется как математическое ожидание отдельной переменной по маргинальной одномерной плотности вероятности.

*Определение.* Условный вектор математического ожидания случайного вектора $\boldsymbol{X}$ при фиксированном значении $\boldsymbol{Y}$ равен

$$q = M\{\boldsymbol{X}/\boldsymbol{Y}\} = \int_{\Omega} Xp(X/Y)dX. \tag{3.77}$$

Это определение получается из (3.74), если $p(\mathrm{X})$ заменить на условную плотность $p(\mathbf{X}/\mathbf{Y})$.

*Дисперсия.*

Другим важным набором числовых характеристик распределения является множество элементов ковариационной матрицы, характеризующих дисперсию распределения.

*Определение*. Ковариационная матрица определяется выражением

$$\sum = M\{(X-q)(X-q)^T\} = M\left\{\begin{bmatrix} x_1 - q_1 \\ \vdots \\ x_n - q_n \end{bmatrix}[x_1 - q_1 ... x_n - q_n]\right\} =$$

$$\begin{bmatrix} M\{(x_1 - q_1)(x_1 - q_1)\} ... M\{(x_1 - q_n)(x_n - q_n)\} \\ \vdots \\ M\{(x_n - q_n)(x_1 - q_1)\} ... M\{(x_n - q_n)(x_n - q_n)\} \end{bmatrix} = \tag{3.78}$$

$$\begin{bmatrix} \sigma_{11}^2 ... \sigma_{1n}^2 \\ \vdots \\ \sigma_{n1}^2 ... \sigma_{nn}^2 \end{bmatrix}.$$

Элементами $\sigma_{ij}^2$ этой матрицы являются числа

$$\sigma_{ij}^2 = M\{(x_i - q_i)(x_j - q_j)\}, \quad ij = \overline{1,n}. \tag{3.79}$$

Диагональными элементами ковариационной матрицы являются *дисперсии* отдельных случайных величин, а недиагональными элементами – *ковариации* пары случайных величин $x_i$ и $x_j$. Следует отметить, что ковариационная матрица является симметричной. Это свойство позволяет в дальнейшем применять результаты из теории симметрических матриц.

Соотношение (3.78) часто преобразуют к следующему виду:

$$\sum M\{XX^T\} - M\{X\}q^T - qM\{X^T\} + qq^T = S - qq^T, \tag{3.80}$$

где

$$V[r(t)] = \begin{bmatrix} M\{r_1(t)\ r_1(t)\} & .... & M\{r_1(t)\ r_n(t)\} \\ ..... & ..... & .. & ..... & ..... & ..... & . \\ M\{r_n(t)\ r_1(t)\} & .... & M\{r_n(t)\ r_n(t)\} \end{bmatrix}. \tag{3.81}$$

Вычисление (3.80) производится непосредственно. Выражение S (3.81) называют *автокорреляционной матрицей* или, иногда, *матрицей разброса* вектора **X**. Уравнение (3.80) выражает связь между ковариационной матрицей и автокорреляционной матрицей и показывает, что они по существу содержат одинаковое количество информации.

Иногда удобно нормировать ковариационную матрицу, преобразуя ее элементы в *коэффициенты корреляции* по формуле

$$r_{ij} = \frac{\sigma_{ij}^2}{\sigma_{ii}\,\sigma_{jj}}. \qquad (3.82)$$

Тогда

$$\Sigma = \tilde{A} R \tilde{A}, \qquad (3.83)$$

где

$$\tilde{A} = \begin{bmatrix} \sigma_{11} & 0 & \dots & 0 \\ 0 & \sigma_{22} & & \\ \vdots & \vdots & & \\ 0 & \cdots & & \sigma_{nn} \end{bmatrix}, \qquad R = \begin{bmatrix} 1 & & r_{ij} \\ & \ddots & \\ r_{ij} & & 1 \end{bmatrix}, \qquad |\, r_{ij} \,| \le 1. \qquad (3.84)$$

Таким образом, матрицу $\Sigma$ можно представить комбинацией матриц: диагональной матрицы дисперсий и матрицы коэффициентов корреляции. Будем называть R *корреляционной матрицей*. Так как дисперсии зависят от масштаба системы координат, то корреляционная матрица содержит существенную информацию о взаимосвязях между случайными величинами.

*Некоррелированные, ортогональные, независимые случайные векторы.*

*Определение*. Два случайных вектора $X_i$ и $X_j$ называют:

*некоррелированными*, если $M\{X_i^T X_j\} = M\{X_i^T\}\, E\{X_j\}$; (3.85)

*ортогональными,* если $M\{X_i^T X_j\} = 0$; (3.86)

*независимыми*, если $p(X_i, X_j) = p(X_i)\, p(X_j)$. (3.87)

Имеется несколько существенных связей между этими тремя условиями.

1) Независимость является более общим условием, чем некоррелированность. Первое означает выполнение равенства (3.85) для каждого $X_i$ и $X_j$ в то время как второе представляет собой только интегральное свойство плотности вероятности $p(X_i, X_j)$. Если случайные векторы $X_i$ и $X_j$ – независимы, то они некоррелированы, т. к.

$$M\{X_i^T X_j\} = \int_\Omega \int_\Omega X_i^T X_j\, p(X_i, X_j)\, dX_i\, dX_j =$$

$$\left\{\int_\Omega X_i^T p(X_i)\, dX_i\right\}\left\{\int_\Omega X_j\, p(X_j)\, dX_j\right\} = M\{X_i^T\} M\{X_j\}. \qquad (3.88)$$

Обратное неверно.

2) Если математическое ожидание вектора $X_i$ или вектора $X_j$ равно нулю, то «некоррелированные» случайные векторы эквивалентны «ортогональным».

3) Если все векторы $X_i$ взаимно ортогональны, то

$$M\left\{\left\|\sum_{i=1}^{N} X_i\right\|^2\right\} = \sum_{i=1}^{N} M\left\{\|X_j\|^2\right\} + \sum_{\substack{i=1 \\ i\neq j}}^{N}\sum_{j=1}^{N} M\{X_i^T X_j\} = = \sum_{j=1}^{N} M\{\|X_i\|^2\}, \quad (3.89)$$

где $\|\cdot\|$ – *длина* или *норма* вектора. Будем использовать норму, определяемую выражением

$$\|X\|^2 = X^T X = \sum_{i=1}^{n} x_i^2 . \quad (3.90)$$

Для $i \neq j$ все математические ожидания $\mathrm{M}\{\mathrm{X_i^T X_j}\}$ равны нулю, что видно из условия ортогональности (3.86).

### 3.6. Понятие вероятностной идентифицируемости

Группа понятий вероятностной идентифицируемости достаточно многочисленна, т. к. она определяется фактически числом методов идентификации. Если оценки точности моделей рассматривать с учетом использования параметрических ошибок $r(t) = \theta - \hat{\theta}(t)$, то понятие идентифицируемости можно связать с асимптотической последовательностью оценок параметров и их состоятельностью, т. е. возможностью оценивания истинных параметров, когда, например, последовательность оценок $\hat{\theta}(t)$ сходится к $\theta$ по вероятности $\hat{\theta}(t) \xrightarrow{p} \theta$, почти наверное $\hat{\theta}(t) \xrightarrow{п.н.} \theta$ или в среднеквадратическом $\hat{\theta}(t) \xrightarrow{с.к.} \theta$.

При решении задач оценивания для практических целей в основном ограничиваются использованием таких моментных характеристик оценок $\hat{\theta}$, как

- математическое ожидание $M\{\hat{\theta}\}$;

- ковариационная матрица $V(\hat{\theta})$ или матрица ковариационных ошибок

$$V[r(t)] = M\left\{[\hat{\theta}(t) - \theta][\hat{\theta}(t) - \theta]^T\right\} = M\left\{r(t) r^T(t)\right\},$$

где $r(t)=\hat{\theta}(t)-\theta$ – вектор параметрической ошибки. В развернутой форме представима как (3.81).

Данные характеристики используются при анализе процессов построения оценок по данным доступной наблюдению информации. Отменим, что существуют свойства оценок, которые можно определять независимо от методов оценивания, из которых перечислим следующие.

Несмещенные оценки. Оценка $\hat{\theta}(t)$ вектора параметров $\theta$, полученная. По последовательности наблюдений при $t\to\infty$, называется несмещенной, если

$$M\left\{\hat{\theta}(t)\right\}=\theta.$$

В противном случае разность $M\left\{\hat{\theta}(t)\right\}-\theta$ есть вектор смещения, а $\left\| M\left\{\hat{\theta}(t)-\theta\right\}\right\|^2$ – квадрат модуля этого вектора, который также можно рассматривать как расстояние смещения.

Состоятельные оценки. Оценка $\hat{\theta}(t)$ вектора параметров $\theta$ называется состоятельной, если она сходится по вероятности с оцениваемым параметром, т. е.

$$\lim_{t\to\infty} P\left\{\left\|\hat{\theta}(t)-\theta\right\|\le\varepsilon\right\}=1,\quad \varepsilon>0.$$

Состоятельные оценки в среднеквадратическом. Оценка $\hat{\theta}(t)$ вектора параметров $\theta$ называется состоятельной в среднеквадратическом, если

$$\lim_{t\to\infty} M\left\{\left\|\hat{\theta}(t)-\theta\right\|^2\right\}=0.$$

Отметим, что из состоятельности в среднеквадратическом следует просто состоятельность. Это аналогично свойству: из сходимости в среднеквадратическом следует сходимость по вероятности. Однако из просто состоятельности не следует состоятельность в среднеквадратическом.

Эффективные оценки. Если для всех несмещенных оценок $\theta_1(t),\theta_2(t)$, которые имеют дисперсии $\sigma^2_{\theta_1},\sigma^2_{\theta_2}$ при $t\to\infty$ имеет место соотношение

$$\sigma^2_{\theta 1}=M\left\{\left\|\hat{\theta}_1(t)-\theta\right\|^2\right\}\le M\left\{\left\|\hat{\theta}_2(t)-\theta\right\|^2\right\}=\sigma^2_{\theta 2},$$

то $\hat{\theta}(t)$ называется эффективной оценкой или оценкой с минимальной дисперсией.

На основании приведенных свойств оценок $\hat{\theta}(t)$ вектора параметров $\theta$ можно определить следующие понятия идентифицируемости для стохастического случая.

*Определение.* Объект называется идентифицируемым, если последовательность оценок параметров объекта сходится к истинным значениям параметров, т. е.

$$\lim_{t\to\infty}\hat{\theta}(t)=\theta,$$

где $\hat{\theta}(t)$ – оценка, получаемая на *t*-ом шаге; $\theta$ – истинное значение параметров.

Данное понятие идентифицируемости связано с рассмотрением асимптотического поведения оценок параметров. Так как в обсуждаемой постановке задача идентификации является стохастической, то сходимость должна рассматриваться как вероятностная. При этом можно использовать сходимость по вероятности, с вероятностью, равной единице (почти наверно), сходимость в среднем квадратическом смысле или по распределению.

Так как сходимость по вероятности сводится к состоятельности, то можно ввести следующее понятие идентифицируемости.

*Определение*. Объект называется идентифицируемым, если оценки его параметров являются состоятельными.

Для векторного параметра указанная идентифицируемость означает, что ковариационная матрица ошибок оценки сходится к нулевой матрице.

Приведенной определение идентифицируемости в общем случае, имеет смысл только тогда, когда структуры исследуемого объекта и выбранной модели совпадают. Кроме того, требование состоятельности является довольно жёстким. В частности, состоятельной может быть только несмещенная оценка. В то же время практически удовлетворительными могут оказаться незначительно смещенные оценки. В связи с указанным более приемлемым может оказаться понятие идентифицируемости, связанное с требованием сходимости оценок параметров, а некоторую область, включающую в себя истинные значения параметров. При этом сходимость параметров в область может косвенно оцениваться с помощью меры p(x, $\hat{x}$) близости выходных сигналов объекта *x* и модели $\hat{x}$.

*Определение.* Объект называется идентифицируемым, если последовательность оценок $\hat{\theta}(t)$ параметров $\theta$ сходится с вероятностью, равней единице, в область

$$C_\theta \left\{\theta : \theta \in C_\theta, \lim_{i\to\infty} p(x, \hat{x}) = 0\right\}, \tag{3.91}$$

т. е.

$$P\left\{\theta(t) \in C_\theta\right\} = 1,\ t \to \infty. \tag{3.92}$$

Данное понятие идентифицируемости снимает ограничения, связанные с требованием состоятельности оценок и с требованием тождества структур объекта и его модели. Очевидно, что это понятие непосредственно не связано с истинными значениями идентифицируемых параметров. Если же область $C_\theta$ состоит только из одной точки, то рассматриваемое понятие эквивалентно понятию идентифицируемости, введенному через состоятельность оценок.

На практике начинает использоваться понятие идентифицируемости, связанное с экстремальными свойствами показателя близости объекта и модели.

*Определение.* Объект называется полностью идентифицируемым, если показатель близости имеет единственный (глобальный) экстремум (минимум или максимум). В случае нескольких экстремумов объект называется локально идентифицируемым.

Анализ используемых в теории идентификации вероятностных условий идентифицируемости показывает, что практически все они зависят от системных условий идентифицируемости. Более того, условия системной идентифицируемости составляют необходимые условия разрешимости задачи вероятностной идентифицируемости.

Более детальный анализ условий идентифицируемости (как системных, так и вероятностных) показывает их определенную зависимость от свойств входных (возбуждающих) сигналов системы и объекта. Считают, что для идентификации входной сигнал должен принадлежать к классу постоянно возбуждающих.

Для целей практической идентификации реальных технологических процессов минимально необходимым условием является то, что входной сигнал должен устойчиво возбуждать процесс в период измерений.

Если для идентификации частотной передаточной функции используется преобразование Фурье или спектральный анализ, т. е.

$$W(j\omega)=\frac{F[x(t)]}{F[u(t)]}=\frac{X(j\omega)}{U(j\omega)} \qquad (3.93)$$

или

$$W(j\omega)=\frac{F[k_{ux}(\tau)]}{F[k_{uu}(\tau)]}=\frac{S_{ux}(j\omega)}{S_{uu}(j\omega)},$$

где $k_{uu}(\tau)$, $k_{ux}(\tau)$ – авто- и взаимокорреляционные функции, то очевидно, что входной сигнал $U(j\omega)$ и его спектральная плотность $S(\omega)$ не должны обращаться в нуль на частотах $\omega_{min} \le \omega \le \omega_{max}$, для которых определяется $W(j\omega)$. Для этого метода идентификации устойчиво возбуждающими сигналами будут те, которые возбуждают процесс в необходимом диапазоне частот. Такие сигналы могут быть непериодическими, полигармоническими или стохастическими.

Для детерминированного случая дискретная импульсная переходная функция $\omega(t)$, определяющая связь между выходными $x(t)$ и входными *u(t)* сигналами записывается выражением (2.72). При определении дискретной импульсной переходной функции $\omega(t)$ с помощью корреляционного анализа для $t = 0,1,..., N$ сумма свертки может быть представлена уравнением вида

$$k_{ux}(t) = \sum_{i=0}^{N} \omega(i) k_{uu}(t-i),$$

которое используется как основное соотношение между входным и выходным сигналами и при этом так же определяется следующей системой уравнений

$$\begin{bmatrix} k_{ux}(0) \\ k_{ux}(1) \\ ..... \\ k_{ux}(N) \end{bmatrix} = \begin{bmatrix} k_{uu}(0) & 0 & ... & 0 \\ k_{uu}(1) & k_{uu}(0) & ... & 0 \\ .... & ..... & .... & ... \\ k_{uu}(N) & k_{uu}(N-1) & ... & k_{uu}(0) \end{bmatrix} \begin{bmatrix} \omega(0) \\ \omega(1) \\ ..... \\ \omega(N) \end{bmatrix}$$

или

$$k_{ux} = K_{uu}^{-1} \; k_{ux}.$$

Тогда непосредственное значение $\omega(i)$ при *i=N+1* может быть найдено непосредственным решением *N + l* уравнений

$$\omega = K_{uu}^{-1} \; k_{ux}.$$

Условием устойчивого возбуждения в этом случае является существование обратной матрицы $K_{uu}^{-1}$, и, следовательно, $K_{uu}$ не должна быть сингулярной, т. е.

$$det\, K_{uu} \neq 0.$$

### 3.7. Информация в статистике

Понятие информации может быть определено по-разному. Однако требования, которым должно удовлетворять понятие информации, введенное Фишером, таковыя:

1. Информация, содержащаяся в выборке наблюдений, должна возрастать с увеличением числа наблюдений. Удвоение числа событий должно удваивать количество доступной информации, если события независимы;

2. Информация должна быть связана с тем, что мы изучаем в эксперименте. Данные, которые не имеют отношения к проверяемой гипотезе (или измеряемым параметрам), не должны содержать какой-либо информации. Конечно, те же самые данные могут содержать информацию о других параметрах или гипотезах;

3. Информация должна быть связана с точностью: чем больше количество информации, тем выше точность эксперимента.

Очевидно, величина, обладающая такими свойствами, будет ценным инструментом при обработке данных. Сейчас трудно найти эксперимент, имеющий так мало первоначальных данных, что все они могут быть опубликованы или даже записаны. Поэтому нужен критерий для эффективного преобразования или обработки данных. При этом необходимо стремиться к максимальному сокращению первоначальных данных, совместимых с минимальной потерей информации.

Пусть $X_k$ представляет собой случайную последовательность наблюдений, а $p(X|\theta)$ является совместной плотностью вероятности случайных величин $X_k$, параметризованной неизвестным вектором параметров $\theta \in \Omega_\theta$, где $\Omega_\theta$ предполагается компактным множеством в метрическом пространстве. Полагаем, что $\theta$ означает истинное значение параметра, и $p(X|\theta)$ является истинной совместной плотностью вероятности, в соответствии с которой распределены $X$.

Функция правдоподобия. Рассмотрим выборку из независимых наблюдений *X,* скажем, $X_1,...,X_n$. Ими могут быть n событий, найденных в эксперименте, где каждому событию соответствует измерение величин. Совместная ф.п.в. всех *X* вследствие независимости равна:

$$L(X|\theta) = L(X_1,...,X_n|\theta) = \prod_{i=1}^{n} p(X_i|\theta). \qquad (3.94)$$

Обычно принято называть $L(X|\theta)$ функцией правдоподобия, если она рассматривается только как функция $\theta$ при *X* фиксированных и равных тем, которые зарегистрированы в эксперименте. Будем рассматривать её как функции и *X* и $\theta$.

Функция результатов наблюдения (статистика). Введем новую случайную переменную в виде $Y = \Phi(X_1,...,X_n)$. Любая такая функция $\Phi$ называется функцией результатов наблюдений (или статистикой). Например, среднее $\overline{X}$ есть статистика.

Определение информации Фишера. Объем информации о параметре $\theta$, содержащейся в наблюдении *X,* определяется следующим выражением (при условии, его существования):

$$I_X(\theta) = M\left[\left(\frac{\partial \ln L(X|\theta)}{\partial\theta}\right)^2\right] = \int_{\Omega_\theta}\left(\frac{\partial \ln L(X|\theta)}{\partial\theta}\right)^2 L(X|\theta)dX. \qquad (3.95)$$

Если $\theta$ – *k*-мерное, формула приобретает вид

$$[I_X(\theta)]_{ij} = M\left[\frac{\partial \ln L(X|\theta}{\partial\theta_i}\frac{\partial \ln L(X|\theta)}{\partial\theta_j}\right] =$$
$$= \int_{\Omega_\theta}\left[\frac{\partial \ln L(X|\theta}{\partial\theta_i}\frac{\partial \ln L(X|\theta)}{\partial\theta_j}\right]L(X|\theta)dX. \qquad (3.96)$$

Таким образом, в общем случае $I_x(\theta)$ есть матрица размерности $k \times k$, которая называется информационной матрицей Фишера.

Имеются два довода в пользу такого, на первый взгляд, произвольного определения: во-первых, оно работает, т. е. оно обладает требуемыми от информации свойствами; и, во-вторых, оно очень тесно связано с максимально доступной точностью оценки величины $\theta$ при заданных *X*.

Свойства информации. Для того, чтобы сделать информацию полезным понятием, определение (3.95) должно быть дополнено следующими двумя условиями:

1. $\Omega_\theta$ не зависит от $\theta$.

2. $L(X|\theta)$ обладает необходимыми свойствами, позволяющими коммутировать операторам $\partial^2 / \partial\theta_i \partial\theta_j$ и $\int dX$.

При выполнении первого условия выполняется, как правило, и второе условие.

Предполагая, что первое и второе условия выполнены, просто показать, что информация (3.96) может быть записана в виде

$$[I_X(\theta)]_{ij} = -M\left[\frac{\partial^2}{\partial\theta_i \partial\theta_j} ln\, L(X|\theta\right].$$

Свойство аддитивности немедленно следует из второго условия. Информация об одном наборе параметров, полученная в разных экспериментах, может быть сложена. В частности, информация, содержащаяся в одном событии, связана информацией, содержащейся в $N$ событиях, соотношением

$$I_N(\theta) = NI_1(\theta). \tag{3.97}$$

Это замечательное свойство может быть очень полезным вследствие соотношения между информацией и точностью. Здесь оно может быть проиллюстрировано на простом примере.

Пример. Пусть переменная $X$ распределена по нормальному закону с известной дисперсией $\sigma^2$ и неизвестным средним $\mu$. Тогда информация, содержащаяся в одном наблюдении $X$, равна

$$I_1(\mu) = -\int_{-\infty}^{\infty} \frac{\partial^2}{\partial\mu^2}\left(-\frac{(X-\mu)^2}{2\sigma^2} - ln\,\sigma\sqrt{2\pi}\right)\frac{1}{\sqrt{2}\sigma\pi} \times exp\left[-\frac{1}{2}\frac{(X-\mu)^2}{\sigma^2}\right]dX =$$

$$+\int_{-\infty}^{\infty}\frac{1}{\sigma^2}\frac{1}{\sqrt{2\pi}\sigma} \times exp\left[-\frac{1}{2}\frac{(X-\mu)^2}{\sigma^2}\right]dX = 1/\sigma^2.$$

Соответственно для $N$ наблюдений $X_i$:

$$I_1(\mu) = -\int_{-\infty}^{\infty}\ldots-\int_{-\infty}^{\infty}\frac{\partial^2}{\partial\mu^2}\left[\sum_{i=1}^{N}\left(-\frac{(X_i-\mu)^2}{\sigma^2}\right) - ln\,\sigma\sqrt{2\pi}\right] \times \prod_{i=1}^{N}\frac{1}{\sqrt{2}\sigma\pi} exp\left[-\frac{1}{2}\frac{(X_i-\mu)^2}{\sigma^2}\right]dX =$$

$$= \frac{N}{\sigma^2} = NI_1(\mu). \tag{3.98}$$

Таким образом, $I_N(\mu)$ равно обратному значению дисперсии $\overline{X}$, а само $\overline{X}$ является обычной оценкой среднего $\mu$ генеральной совокупности. Для такого случая можно показать, что соотношение

(3.98) дает обратное значение минимальной дисперсии, которая может быть получена при любом способе оценки $\mu$.

Рассмотрим другое важное свойство информации, а именно применимость этого понятия к задачам обработки данных. С этой целью введем понятие достаточных статистик.

Достаточные статистики. Говорят, что статистика $T=\Phi(X)$ достаточна для $\theta$, если условная функция плотности $X$ при данном $T$ $f(X|T)$ независима от $\theta$. $T$ и $\theta$ могут быть многомерными и иметь различные размерности.

Если $T$ является достаточной статистикой для $\theta$, то любая строго монотонная функция $T$ также будет достаточной статистикой для $\theta$. Важность понятия достаточности для обработки данных следует из определения: в $T$ содержится столько же информации о $\theta$, сколько и в первоначальных данных $X$. Иными словами, любая другая функция тех же самых данных не содержит какую-либо дополнительную информацию о $\theta$.

Очевидно, что множество измерений $T=X$ достаточно для $\theta$, поскольку оно содержит всю первоначальную информацию. Набор достаточных статистик будет полезен только в случае, когда он приводит к сокращению объема экспериментальных данных.

Из самого определения следует способ проверки достаточности, но предпочтительнее (из практических соображений) использовать следующий результат.

Необходимое и достаточное условие того, что $T(X)$ является достаточной статистикой для $\theta$, состоит в том, что функция правдоподобия может быть представлена в виде

$$L(X \mid \theta) = g(T,\theta)h(X), \qquad (3.99)$$

где 1) $h(X)$ не зависит от $\theta$ и 2) $g(T,\theta)$ пропорционально $A(T \mid \theta)$ условной плотности вероятности распределения T при заданном $\theta$.

Во многих случаях условие 2 выполняется. Например, если множество значений $X$ не зависит от $\theta$, можно записать:

$$A(T|\theta) = \int\limits_{\substack{\Omega_\theta \\ T=const}} L(X|\theta)dX = \int\limits_{\substack{\Omega_\theta \\ T=const}} g(T,\theta)h(X)dX = g(T,\theta) \int\limits_{\substack{\Omega_\theta \\ T=const}} h(X)dX,$$

т. е. в этом случае условие 2 вытекает из условия 1. В общем же случае необходимо проверять выполнимость обоих условий, прежде чем использовать $T(X)$ в качестве достаточной статистики.

Примеры. Предположим, что $X$ распределено по нормальному закону $\mathrm{N}(\mu,\sigma^2)$. Тогда функция правдоподобия равна:

$$L(X\mid\mu,\sigma^2)=(2\pi\sigma^2)^{-n/2}\,exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(X_i-\mu)^2\right]=(2\pi\sigma^2)^{-n/2}$$
$$\times exp\left[-\frac{n}{2\sigma^2}\sum_{i=1}^{n}(\overline{X}-\mu)^2\right]exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(X_i-\overline{X})^2\right], \qquad (3.100)$$

где $\overline{X}$ – среднее значение.

Если $\sigma^2$ известно, то $\overline{X}$ распределено по нормальному закону $N(\mu,\sigma^2/n)$, а функция правдоподобия (3.100) может быть представлена в виде соотношения (3.99)

$$g(\overline{X}\mid\mu)=\left(\frac{n}{2\sigma^2}\right)^{1/2}exp\left[-\frac{n}{2\sigma^2}(\overline{X}-\mu)^2\right].$$

Очевидно, что условия 1 и 2 выполняются. Поэтому $\overline{X}$ является достаточной статистикой для $\mu$.

Когда, с другой стороны, $\mu$ известно, а $\sigma^2$ неизвестно, легко показать, что статистика

$$\hat{\sigma}^2=\frac{1}{n}\sum_{i=1}^{n}\left(X_i-\mu\right)^2$$

достаточна для $\sigma^2$.

## 3.8. Структурная идентификация

### 3.8.1. Проблема структурной идентификации

Структурная идентификация является чрезвычайно сложной и до настоящего времени мало разработанной. Это объясняется большим разнообразием классов (видов) математических моделей технических систем и возможностью описать модели одного и того же класса различными операторами. Напомним, что модели объединяются в классы по многим свойствам, основными из которых являются такие, как сосредоточенность или распределенность параметров, линейность или нелинейность, статичность или динамичность, стацио-

нарность или нестационарность, непрерывность или дискретность, детерминированность или стохастичность.

Пусть $\omega_i$ означает определенное свойство системы, а $\overline{\omega_i}$ – альтернативное (противоположное) свойство. Тогда элемент $\Omega_j$ множества

$$\Omega = \{\Omega_j : (\omega_1 U \overline{\omega}_1) \bigcap (\omega_2 U \overline{\omega}_2) \bigcap ... \bigcap (\omega_s U \overline{\omega}_s)\} \qquad (3.101)$$

и определяет логически обоснованную комбинацию свойств системы, т. е. класс системы.

В формуле (3.101) через знак $U$ обозначена логическая операция "или" (дизъюнкции), через знак $\bigcap$ – операция "и" (конъюнкция). Множество классов $\Omega$ имеет $S = 2^s$ элементов, соответствующих различным классам. Примерами классов могут быть комбинации "системы с сосредоточенными параметрами-линейные-динамические-стационарные-не-прерывные-детерминированные" или "системы с сосредоточенными параметрами-нелинейные динамические-стационарные-дискретные-стохас-тические" и т. д.

Отметим, что выбор класса модели в общем случае не адекватен решению задачи структурной идентификации. Дело в том, что для описания одного и того же класса моделей могут быть использованы различные операторы, т. е. число возможных операторов превышает число классов моделей.

Так, для математического описания класса "системы с сосредоточенными параметрами-нелинейные-динамические-стационарные-непрерывные-детерминированные" могут быть использованы операторы Винера, Гаммерштейна, ряды Вольтерра и т. д. Последнее объясняется многообразием типов нелинейных систем и их математических моделей. Конечно, понятие класса систем и соответственно множества классов можно было бы расширить за счет рассмотрения операторов в качестве элементов этого класса. Но при этом для описания множества $\Omega$ вместо формулы (3.101), сконструированной с привлечением двузначной логики, пришлось бы обратиться к формуле, построенной на основе многозначной логики. Очевидно, и число элементов множества $\Omega$ сильно возрастет, что в значительной мере может усложнить задачу структурной идентификации. Поэтому более предпочтительным является пока многоэтапная процедура структурной идентификации, на каждом этапе которой по каждому признаку (свойству) $\omega_i$ решается двухальтернативная задача выбора (принятие решения).

В настоящее время универсальных (на все случаи объектов) методов и алгоритмов структурной идентификации не существует. Не последнюю роль при определении оператора модели идентифицируемого объекта играют опыт, интуиция, здравый смысл, умение анализировать, обобщать и использовать имеющуюся априорную информацию. Исходное математическое описание изучаемого объекта во многих случаях можно получить на основе законов физики, механики, химии и т. д.

Например, для составления математических моделей движущихся объектов (самолетов, ракет, искусственных спутников Земли и т. д.) широко используются такие законы механики, как первый, второй и третий законы Ньютона.

Для получения моделей электрических устройств применяются законы Ома, Кирхгофа и т. д.

Уточнение моделей, полученных аналитическим путем при использовании указанных и других законов, обычно производят на основе экспериментальных данных о входных и выходных сигналах объекта. Характер уточнений может быть различный. Возможна, например, такая ситуация. Допустим, что аналитический подход позволил нам получить модель объекта в виде обыкновенного неоднородного линейного дифференциального уравнения с постоянными коэффициентами с точностью до порядка уравнения и значений коэффициентов

$$x^{(n)} + \sum_{i=1}^{n-1} a_i x^{(i)} = \sum_{j=0}^{m} b_j u^{(j)}.$$

При этом структурная идентификация должна завершиться определением величины $n$ и $m$. Известно, что некоторые методы идентификации позволяют одновременно оценивать как структуру, так и параметры оператора модели. В основном это касается линейных объектов. В частности, структура и параметры модели определяются при применении метода, основанного на аппроксимации входных и выходных сигналов объекта ортогональными рядами, или метода типовой идентификации полученного на основе корреляционного (непараметрического) подхода.

### 3.8.2. Последовательный метод принятия решений в задаче выбора класса модели

В соответствии с выражением (3.101)3, процесс решения задачи выбора (оценивания) класса $\Omega_j$ модели можно разделить на отдельные

этапы таким образом, чтобы на каждом очередном *i*-ом этапе устанавливалась истинность свойства $\omega_i$ или $\overline{\omega}_i$. Допустим, что в соответствии с принятой классификацией $\omega_1$ – сосредоточенность параметров ($\overline{\omega}_1$ – распределенность); $\omega_2$ – линейность оператора модели ($\overline{\omega}_2$ – нелинейность); $\omega_3$ – статичность ($\overline{\omega}_3$ – динамичность); $\omega_4$ – стационарность ($\overline{\omega}_4$ – нестационарность); $\omega_5$ – непрерывность ($\overline{\omega}_5$ – дискретность); $\omega_6$ – детерминированность ($\overline{\omega}_6$ – стохастичность). Тогда структурную граф – схему последовательного оценивания класса модели можно представить в виде, изображенном на рис. 3.6. Очевидно, что число возможных классов моделей при рассматриваемом подходе определяется формулой $S = 2^s$, где $S$ – число признаков классификации.

В нашем случае *s=6*, поэтому *S=64*.

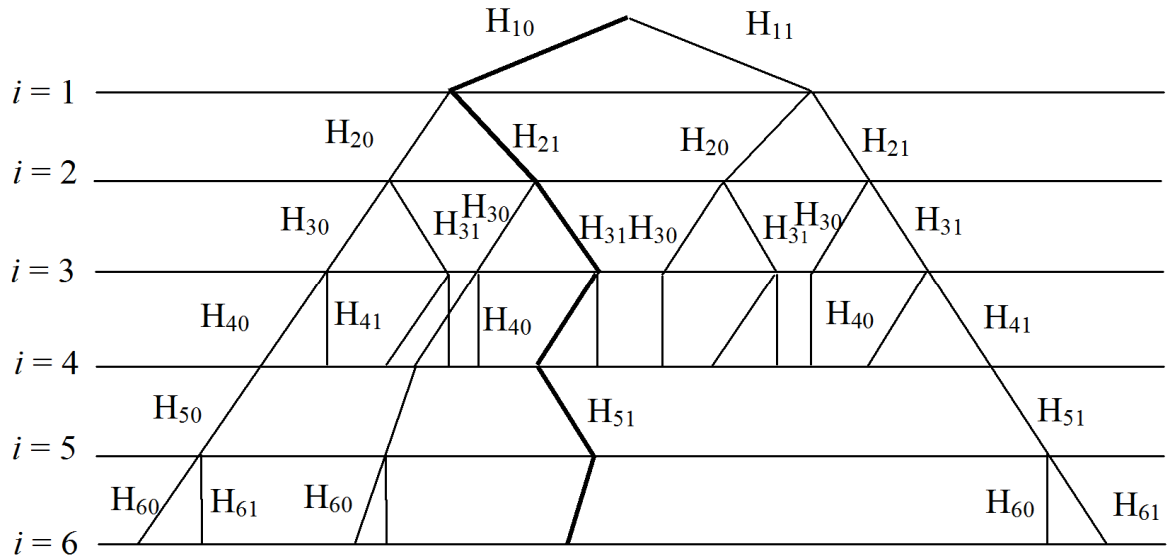



Рис. 3.6. Структура граф-схемы последовательного оценивания класса модели

На рис. 3.6 жирными линиями выделен класс модели, описывающей объекты с сосредоточенными параметрами, нелинейные, динамические, стационарные, дискретные, детерминированные.

Заметим, что наличие соответствующей априорной информации заметно упрощает решение задачи выбора классов модели последовательным методом. Существование только двух альтернативных решений на каждом *i*-ом этапе также облегчает выбор класса модели. Как уже было отмечено, задача структурной идентификации при ис-

пользовании данных о входных и выходных сигналах объекта носит статистический характер. В связи с этим вполне естественным является стремление использовать для решения двух альтернативных задач методы статистической проверки гипотез.

При таком подходе на каждом *i*-м этапе идентификации последовательным методом выдвигается нулевая или основная гипотеза $H_{i0}$ и противоречащая нулевой конкурирующая или альтернативная гипотеза $H_{i0}$. Например, если нулевая гипотеза состоит в предположении, что исследуемый объект является линейным, то конкурирующая гипотеза состоит в том, что объект является нелинейным.

В математической статистике и теории идентификации универсальных методов статистической проверки гипотез для решения задач структурной идентификации в целом не существует. Обычно применяют известные в математической статистике параметрические и непараметрические методы проверки гипотез для установления истинности свойств $\omega_i$ или $\overline{\omega_i}$ на отдельных этапах последовательного метода идентификации. Кроме того, в последнее время начали разрабатываться специальные методы проверки гипотез решения задач структурной идентификации. Возможности методов статистической проверки гипотез рассмотрим на примере одной типовой задачи.

### 3.8.3. Проверка линейности статического объекта

Допустим, что мы установили статичность объекта и хотим на очередном этапе идентификации определить его линейность (или нелинейность). Для этого воспользуемся тем фактом, что в некоторых случаях выходной сигнал линейного объекта сохраняет вид закона распределения входного сигнала. Так, выходной сигнал линейного объекта

$$x = au + b,$$

где $u$ – случайный нормально распределенный входной сигнал с плотностью

$$p(u) = \frac{1}{\sqrt{2}\pi\sigma}\exp\left(\frac{(u-\overline{u})^2}{2\sigma_u^2}\right)$$

имеет также нормальное распределение

$$p(x) = \frac{1}{\sqrt{2\pi}|a|\sigma_u}\exp\left(\frac{(x-b-a\overline{u})^2}{2a^2\sigma_u^2}\right).$$

Таким образом, производя проверки нормальности выходного сигнала *x*, можно сделать вывод о линейности объекта. Проверка нормальности сводится к проверке принадлежности статистической функции распределения сигнала *x* к семейству нормальных распределений. Эту проверку можно проводить несколькими методами статистической проверки гипотез о законах распределения, отличающихся друг от друга видом меры рассогласования между статистическими и гипотетическими законами распределения (в литературе эти меры не совсем корректно называются критериями). Наиболее часто применяются методы Колмогорова, Пирсона, Смирнова и др.

Эти методы для решения задач согласия (проверки гипотез о законах распределения) являются универсальными. В статистике, кроме того, разработаны специальные методы, учитывающие особенности нормального распределения. В частности, весьма простыми и удобными являются методы, основанные на оценивании коэффициентов асимметрии и эксцесса. Известно, что у нормально распределенных случайных величин коэффициент асимметрии $a_u = \mu_3 / \sigma_u^3$, где $\mu_3$ – центральный момент третьего порядка, равный нулю, а коэффициент эксцесса $e_u = \mu_4 / \sigma_u^4$, где $\mu_4$ – центральный момент четвертого порядка, равный трем. Для подтверждения нормальности распределения выходного сигнала объекта проверяется гипотеза о равенстве коэффициентов симметрии и эксцесса выходного сигнала соответственно нулю и трем.

Указанный подход к проверке линейности применим и к динамическим объектам, если измерения входного *u* и выходного *x* сигналов проводить в установившемся режиме.

# ГЛАВА 4. НЕПАРАМЕТРИЧЕСКАЯ ИДЕНТИФИКАЦИЯ

Существует много способов определения динамических характеристик неизвестного линейного объекта или системы. Ниже в целях упрощения предполагается, что:

1) система имеет единственный вход и единственный выход;

2) система является линейной и стационарной;

3) аддитивная помеха, действующая на выходной сигнал системы, не зависит от входного сигнала и представляет собой стационарный случайный процесс.

При этих предположениях структура модели может быть представлена так, как и изображено на рис. 4.1. Тогда задача сводится к определению одной их следующих характеристик:

1) импульсного перехода или весовой функции *g(t);*

2) частотной характеристики *G(jω);*

3) передаточной функции *G(s).*

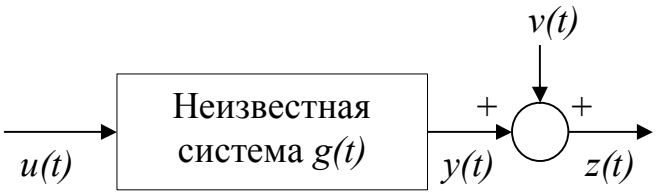



Рис. 4.1. Модель для идентификации линейной системы с одним входом и одним выходом

Следуя работам [20, 21, 23, 25, 26, 28], приведем основные соображения и теоретические положения, лежащие в основе наиболее представляемых методов.

## 4.1. Частотный анализ

А. Частотная характеристика. При отсутствии помех выходной сигнал системы *y(t),* показанной на рис 4.1, выражается через входной сигнал *u(t)* и импульсную переходную функцию *g(t)* с помощью интеграла свертки

$$y(t) = \int_0^\infty g(\tau)u(t-\tau)d\tau \ . \qquad (4.1)$$

Теперь предположим, что входной сигнал синусоидален; тогда *u(t)* можно представить в виде

$$u(t) = A\exp[(\sigma + j\omega)t] = A\exp(st), \tag{4.2}$$

где А – амплитуда, а $\omega$ – угловая частота, измеренная в рад/с. Подстановка выражения (4.2) в интеграл свертки (4.1) дает

$$y(t) = \int_0^\infty g(\tau)\,exp(-st) =$$

$$= A\,exp(st)\int_0^\infty g(\tau)\,exp(-st)\,d\tau = \tag{4.3}$$

$$= u(t)\int_0^\infty g(\tau)\,exp(-st)\,d\tau. \tag{4.4}$$

Теперь отношение преобразований Лапласа выходного и входного сигналов, которое называется передаточной функцией системы, может быть представлено как

$$G(s) = \frac{Y(s)}{U(s)} = \int_0^\infty g(\tau)\exp(st)\exp(-s\tau)\,d\tau. \tag{4.5}$$

Таким образом, этот интеграл не зависит от $t$ и определяет передаточную функцию системы $G(s)$ через ее импульсную переходную функцию $g(t)$. Если в формулу (4.5) в качестве аргумента преобразований Лапласа подставить комплексную переменную $j\omega$, то получится установившееся значение частотной характеристики системы

$$G(j\omega) = \frac{Y(j\omega)}{U(j\omega)} = \int_0^\infty g(\tau)\exp(-j\omega)\,d\tau\ . \tag{4.6}$$

Поскольку импульсная переходная функция системы $g(t)$ должна быть равна нулю при $t < 0$, то нижний предел интегрирования можно продлить до - $\infty$, так что

$$G(j\omega) = \int_0^\infty g(\tau)\,exp(-j\omega)\,dt. \tag{4.7}$$

Из этой формулы видно, что установившееся значение частотной характеристики системы тождественно преобразованию Фурье от импульсной переходной функции системы.

Рассматривая комплексную величину $G(j\omega)$ как вектор, получим представление $G(j\omega)$ в полярных координатах отношение

амплитуд (амплитудная характеристика) и фазовый сдвиг (фазовая характеристика).

$$\begin{aligned} |G(j\omega)| &= \frac{|Y(j\omega)|}{|U(j\omega)|} = S(\omega), \\ \angle G(j\omega) &= \angle \frac{|Y(j\omega)|}{|U(j\omega)|} = \varphi(\omega). \end{aligned} \tag{4.8}$$

Здесь передаточная функция системы определяется на основе частотного анализа.

*Б. Частотный анализ методом диаграмм Боде.* Передаточная функция стационарной линейной системы обычно записывается в виде отношения двух многочленов от *s*:

$$G(s) = \frac{bm^{sm} + \ldots + b_{1s} + b_0}{an^{sn} + \ldots + a_{1s} + a_0}. \tag{4.9}$$

Хорошо известно, что при отсутствии комплексных полюсов и нулей выражение (4.9) можно разложить на элементарные дроби вида

$$G(s) = \frac{k_1}{s + s_2} + \frac{k_2}{s + s_2} + \ldots + \frac{k_n}{s + s_n}, \tag{4.10}$$

где $s_1, s_2, \ldots s_n$ – полюсы, а $k_1, k_2 \ldots k_n$ – вычеты *G(s).* Такое представление иногда называют разложением Хевисайда, которое является чрезвычайно полезным при вычислении обратного преобразования Лапласа.

Вторую форму записи можно получить посредством факторизации числителя и знаменателя передаточной функции:

$$G(s) = \frac{K(s + c_1) + (s + c_1)\ldots(s + c_m)}{(s + s_1)(s + s_1)\ldots(s + s_n)}. \tag{4.11}$$

Здесь $c_1, c_2, \ldots c_m$ – нули *G(s),* а *K* – коэффициент усиления. Такое представление называют разложением Боде или полюсно-нулевым представлением.

Отметим, что если полюсы и нули функции *G(s),* определяемой формулой (4.9), не являются чисто вещественными, то приведенное разложение должно включать члены второго порядка. Так, например, при наличии в формуле (4.9) комплексно-сопряженных полюсов $s_j = \alpha + j\beta$ и $s_{j+1} = \alpha - j\beta$ числитель *G(s*), можно представить в виде

$$(s+s_j)(s+s_{j+1})\prod_{\substack{i=1\\ i\neq j,j+1}}^{n}(s+s_i)=$$

$$\left[(s-\alpha)^2+\beta^2\prod_{\substack{i=1\\ i\neq j,j+1}}^{n}(s+s_{j)}\right]= \tag{4.12}$$

$$\frac{1}{T^2}\left[(Ts)^2+2\xi Ts+1\right]\prod_{\substack{i=1\\ i\neq j,j+1}}^{n}(s+s_i),$$

где $T=1/\sqrt{\alpha^2+\beta^2}$ , $\xi=-a/\sqrt{2\alpha^2+\beta^2}$ – коэффициент затухания.

Если значения $c_i$ и $s_i$ не равны нулю, то можно преобразовать формулу (4.11) к виду

$$G(s)=\frac{K'(1+Q_1s)(1+Q_2s)...(1+Q_ms)}{(1+T_1s)(1+T_2s)...(1+T_ns)}, \tag{4.13}$$

где $$K'=K\frac{c_1c_2...c_m}{s_1s_2...s_n},\quad T_i=\frac{1}{s_i},\quad Q_i=\frac{1}{c_i}\ . \tag{4.14}$$

Тогда частотная характеристика получается путем подстановки в формуле (4.13) $j\omega$ вместо $s$

$$G(j\omega)=\frac{K'(1+jQ_2\omega)(1+jQ_2\omega)...(1+jQ_m\omega)}{(1+jT_1\omega)(1+jT_2\omega)...(1+jT_n\omega)}. \tag{4.15}$$

Частотную характеристику удобно определять с помощью диаграммы Боде, потому что:

1) амплитудную характеристику для $G(j\omega)$можно аппроксимировать кусочно-линейно ломаной;

2) поскольку масштаб измерения частоты $\omega$ логарифмический, то в диаграмму укладывается широкий диапазон значений $\omega$ ;

3) умножение амплитуды частотной характеристики $G(j\omega)$ на диаграмме Боде сводится к простому сложению. Преобразуя формулу (4.15), получим

$$\lg\left|G(j\omega)\right|=\lg\left|K'\right|+\sum_{i=1}^{m}\lg\left|1+jQ_i\omega\right|-\sum_{i=1}^{n}\lg\left|1+jT_i\omega\right|, \tag{4.16}$$

$$\varphi(\omega)=\sum_{i=1}^{m}arctg(Q_i\omega)-\sum_{i=1}^{n}arctg(T_i\omega)\,. \tag{4.17}$$

Эти соотношения называют уравнениями частотной характеристики. Уравнение (4.16) описывает зависимость амплитуды от частоты, а (4.17) – зависимость фазы от частоты. Используя формулу (4.15), частотную характеристику $G(j\omega)$ можно определить на основе частотных характеристик передаточных функций простейших звеньев $1/(1+T_i s)$ и $(1+Q_i s)$. Таким образом, важно знать, как ведут себя частотные характеристики звеньев, $1/(1+Ts)$ и $(1+Ts)$.

В соответствии с формулами (4.16) и (4.17) для $K/(1+Ts)$ находим

$$lg|G(j\omega)| = lg|K| - 0{,}5lg\left[1+(T\omega)^2\right], \quad \varphi(\omega) = -arctg(T\omega). \qquad (4.18)$$

Хорошо известны следующие асимптотические свойства диаграммы Боде для такого звена в левой (область низких частот) и правой (область высоких частот) частях диапазона:

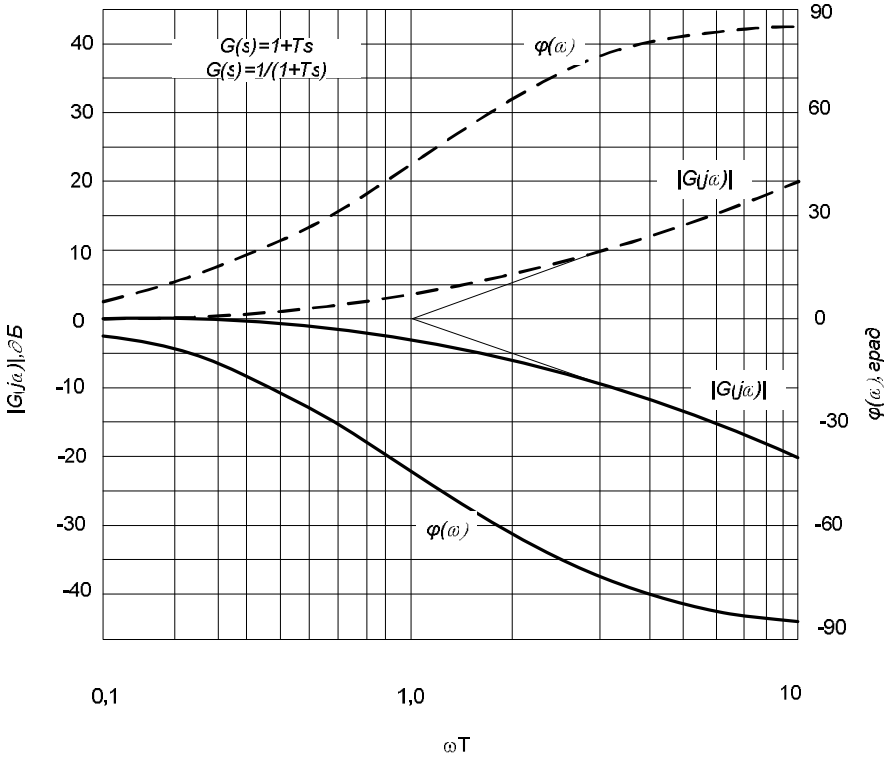



Рис. 4.2. Диаграммы Боде для звеньев $G(s)=1+Ts$ и $G(s)=1/(1+Ts)$

$$\begin{gathered} \omega T << \text{(левая часть диапазона)} \\ \lg|G(j\omega)| = \lg|K|,\ \varphi(\omega) = 0^0, \\ \omega T >> 1 \ \text{(правая часть диапазона)} \\ \lg|G(j\omega)| = \lg|K|/T\omega,\ \varphi(\omega) = 90^0 \\ \omega T = 1 : \lg|G(j\omega)| = \frac{\lg|K|}{\sqrt{2}},\ \varphi(\omega) = -45^0. \end{gathered} \qquad (4.19)$$

На рис. 4.2 приведена диаграмма Боде для звена *1/(1+Ts)*. Частота $\omega = 1/T$ называется частотой среза. Примерно так же выглядит график частотной характеристики звена *(1+Ts*), показанный на рис. 4.2 штриховой линией. В случае, когда полюсы или нули *G(s)* являются комплексно-значными величинами, рассматриваемое разложение *G(s)* включает сомножители второго порядка типа (4.12). Поэтому нужно знать поведение частотной характеристики и для передаточной функции звена второго порядка.

Рассмотрим частотную характеристику звена

$$x(t) = \sum_{i=0}^{\infty} \omega(i) u(t-i) \,. \tag{4.20}$$

Здесь предполагается, что $T > 0$ и *0<ξ<1*. Используя это соотношение в случаях, аналогичных рассмотренным в (4.19), получаем следующие асимптотические выражения:

$$\begin{gathered} \omega T << 1 : \lg\left|G(j\omega)\right| = 0, \;\; \varphi(\omega) = 0^0 \,, \\ \omega T >> 1 : \lg\left|G(j\omega)\right| = \lg[1/(\omega T)^2], \;\; \varphi(\omega) = -180^0, \\ \omega T = 1 : \lg\left|G(j\omega)\right| = \lg[1/2\xi\omega T], \;\; \varphi(\omega) = -90^0 . \end{gathered} \tag{4.21}$$

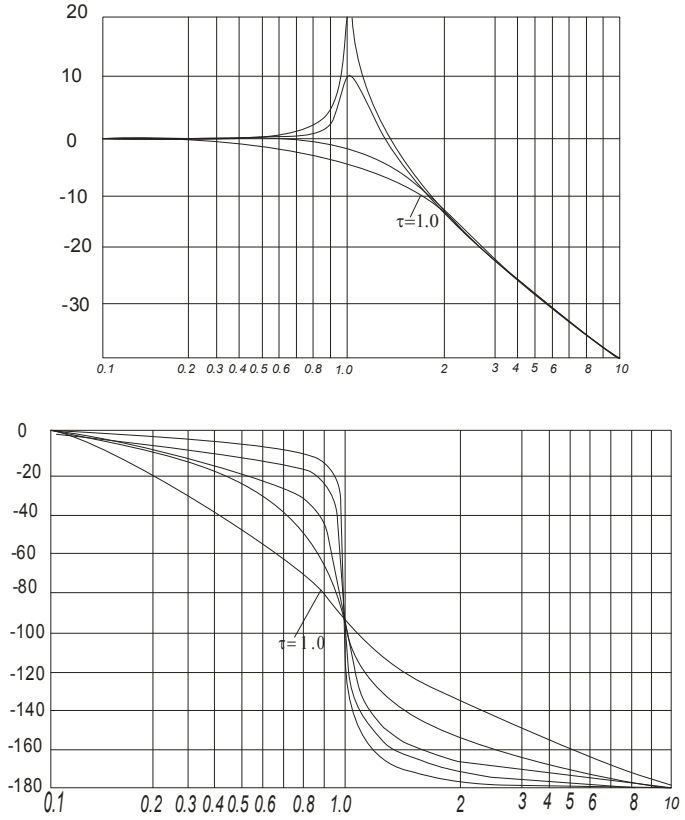



Рис. 4.3. Диаграмма Боде для звена *G(s)=1/[(Ts)²+2ξTs+1]*

На рис. 4.3 показана диаграмма Боде для различных значений ξ из формулы (4.20). Асимптотические характеристики в области низких частот совпадают с характеристиками звена первого порядка, однако в области высоких частот сдвиг по фазе вдвое больше. Кроме того, при малом коэффициенте затухания максимум резонанса достигается в окрестности частоты среза.

## 4.2. Анализ переходных процессов

Как было отмечено выше, с помощью интеграла свертки выходной сигнал системы выражается через ее импульсную переходную функцию. Таким образом, если импульсная переходная функция задана, то частотная характеристика может быть определена по формуле (4.6). Ниже описывается процедура построения частотной характеристики по заданной импульсной переходной функции.

*А. Аппроксимация частотной характеристики с помощью конечной суммы.* Если известна импульсная переходная функция системы, заданная, например, дискретным набором ординат, представляющих значения $g(t)$ в точках $t=i\theta,\ i=0,1,\ldots,\infty$, где $\theta$ – период выборки, то выражение (4.6) может быть приближенно представлено суммой вида

$$G(j\omega) \approx \theta \sum_{i=0}^{\infty} g(i\theta) \exp[-j\omega(i\theta)]. \qquad (4.22)$$

Используя формулу Эйлера, можно представить экспоненту в виде

$$\exp[-j\omega(i\theta)] = \cos(i\omega\theta) - j\sin(i\omega\theta). \qquad (4.23)$$

Подстановка выражения (4.22) в (4.23) дает

$$G(j\omega) \approx \theta \sum_{i=0}^{\infty} g(i\theta)\cos(i\omega\theta) - j\theta \sum_{i=0}^{\infty} g(i\theta)\sin(i\omega\theta). \qquad (4.24)$$

Если система асимптотически устойчива, так что $g(t) \to 0$ при $t \to 0$, то импульсная переходная функция может быть усечена при некоторым t=nθ. Тогда выражения (4.24) преобразуется к виду

$$G(j\omega) \approx \theta \sum_{i=0}^{\infty} g(i\theta)\cos(i\omega\theta) - j\theta \sum_{i=0}^{\infty} g(i\theta)\sin(i\omega\theta). \qquad (4.25)$$

Это соотношение определяет частотную характеристику.

При проведении расчетов по формуле (4.25) с использованием вычислительных машин необходимо принимать во внимание эффект наложения (aliasing). Период выборки $\theta$ нужно определить так, чтобы не пропустить резких изменений $g(t)$. Чтобы получилось хорошее

согласие с истинной частотной характеристикой, период выборки по теореме выборки должен удовлетворять следующему неравенству:

Период выборки ≤ Половина рассматриваемой частотной области. Отсюда следует, что если рассматриваемая область частот задается в виде максимальной частоты $\overline{\omega}$ (рад/с), то наибольший период выборки, при котором имеет место хорошее согласие с истиной характеристикой, равен [23, 28]

$$\theta = 2\pi/\overline{\omega}. \qquad (4.26)$$

Так как время установлено Т импульсной переходной функции известно, т. е. известна точка, в которой $g(i\theta)=0=g(T)$, то минимальное возможное значение $n$ равно

$$n=T/\theta. \qquad (4.27)$$

*Б. Определение частотной характеристики.* Передаточная функция линейной динамической системы может быть представлена в виде отношения двух многочленов от частоты. Допустим сначала, что форма идентифицируется передаточной функцией определяется выражением

$$\hat{G}(j\omega)=\frac{b_0+b_1(j\omega)+\ldots+b_m(j\omega)^m}{1+a_1(j\omega)+\ldots+a_n(j\omega)^n}=\frac{\hat{B}(j\omega)}{\hat{A}(j\omega)}. \qquad (4.28)$$

Для интерполяции экспериментальных данных функциями такого вида предложено несколько методов. Наиболее известный подход состоит в определении таких оценок неизвестных коэффициентов $a_i$ и $b_i$ передаточной функции, чтобы сумма квадратов отклонений оценок $G(j\omega)$ значений частотной характеристики от экспериментальных данных была минимальной.

Непараметрической моделью линейной стационарной динамической системы являются ее амплитудная и фазовая характеристики:

$$S=S(\omega),\ \psi=\psi(\omega),\ 0<\omega<\infty, \qquad (4.29)$$

т. е. зависимость амплитуды $S$ и фазы $\psi$ колебаний выхода объекта от частоты $\omega$ единичного гармоничного возмущения.

Связь между импульсной переходной функцией $\omega(t)$ и амплитудно-фазовыми характеристиками реализуется соотношениями:

$$S(\omega)\cos\psi(\omega)=\int_0^\infty \omega(t)\cos\omega t dt,$$

$$S(\omega)\sin\psi(\omega)=-\int_0^\infty \omega(t)\sin\omega t dt,$$

откуда получаем

$$S(\omega)\cos\psi(\omega)=\left[\left(\int_0^\infty \omega(t)\cos\omega t dt\right)^2+\left(\int_0^\infty \omega(t)\sin\omega t dt\right)^2\right]^{\frac{1}{2}},$$

$$\psi(\omega)=arctg\left[-\frac{\int_0^\infty \omega(t)\sin\omega t dt}{\int_0^\infty \omega(t)\cos\omega t dt}\right].$$

Импульсная переходная функция выражается через амплитудно-фазовые характеристики (4.29) следующей зависимостью:

$$\omega(t)=-\frac{2}{\pi}\int_0^\infty S(\omega)\cos\psi(\omega)\cos\omega t dt. \qquad (4.30)$$

Таким образом, определение амплитудно-фазовых свойств объекта дает возможность вычислить его импульсную переходную функцию. И наоборот, значение импульсной переходной функции дает возможность определить (4.29).

Связь моделей описание выше модели естественно связаны между собой. Так непрерывная параметрическая модель (4.28) связана с непараметрической (4.29) следующим образом [20, 23, 28]

$$\begin{aligned} S(\omega)&=\sqrt{\frac{a^2+b^2}{c^2+d^2}},\\ \psi(\omega)&=arctg\frac{bc-ad}{ac+bd}, \end{aligned} \qquad (4.31)$$

где

$$\begin{aligned} a&=b_0-b_2\omega^2+b_4\omega^4-...,\\ b&=b_1\omega-b_3\omega^3+b_s\omega^s-...,\\ c&=a_0-a_2\omega^2+a_4\omega^4-...,\\ d&=a_1\omega-a_3\omega^3+a_s\omega^s-..., \end{aligned} \qquad (4.32)$$

а коэффициенты $a_i$ *(i=0,...,p-1)* и $b_j$*(j=0,...,1)* соответствуют модели (4.28).

Таким образом, зная параметры непрерывной параметрической модели (4.28), легко определить характеристики непараметрической модели (4.29) и с помощью (4.30) импульсную переходную функцию *g(t)*.

Если же известна одна из амплитудно-фазовых характеристик непараметрической модели (4.29), то параметры непрерывной модели

$a_i$ ,$b_j$ можно оценить в результате решения одной из следующих задач минимализации:

$$I_S(C) = \int_0^\infty \left[ S(\omega) - \sqrt{\frac{a^2 + b^2}{c^2 + d^2}} \right]^2 d\omega \to \min_C, \tag{4.33}$$

$$I_\psi(C) = \int_0^\infty \left[ \psi(\omega) - arctg \frac{bc - ad}{ac + bd} \right]^2 d\omega \to \min_C, \tag{4.34}$$

где функции *a, b, c* и *d* определены в (4.35), а вектор искомых параметров с параметрической модели имеет вид (4.2).

Пример 4.1. Рассмотрим задачу восстановления модели (4.1) по наблюдениям амплитудно-частотной характеристики объекта *S(ω)* в случае, когда его структура представляет собой апериодическое звено. Модель, таким образом, имеет вид: $\dot{y} + a_1 y = b_0 x$.

Для идентификации параметров $a_1$ и $b_0$ модели воспользуемся первым выражениям из (4.29). Для апериодической структуры с учетом (4.31), т.е. при $a=b_0$, $b=0$, $c=a$, $d=w$, получаем следующее выражение для минимизируемой функции:

$$I_S(a_0, b_0) = \int_0^\infty \left[ S(\omega) - \frac{b_0}{\sqrt{a_0^2 + \omega^2}} \right]^2 d\omega. \tag{4.35}$$

Так как наблюдения дискретны, то естественно записать этот интеграл в виде суммы

$$I_S(a_1, b_0) = \sum_{i=0}^{10} \left[ S(\omega_i) - \frac{b_0}{\sqrt{a_1^2 + \omega_i^2}} \right]^2. \tag{4.36}$$

## 4.3. Определение передаточной функции по временным характеристикам объекта

Уравнения связи между входными и выходными переменными можно записать в различной форме. Наиболее универсальными из них являются дифференциальные уравнения объекта (2.21), (2.31).

Широко используется также уравнение связи между входом и выходом типа интеграла свертки (интеграл Дюамеля):

$$y(t) = \int_0^t x(\tau) w(t - \tau) d\tau = \int_0^t w(\tau) x(t - \tau) d\tau \ , \tag{4.37}$$

где $w(\tau)$ – функция веса объекта, т. е. реакция объекта на входной сигнал в виде дельта функции

$$\delta(t)=\begin{cases}0 & \text{ïðè } t\neq 0\\ \infty & \text{ïðè } t=0\end{cases};\qquad \int_{-\infty}^{\infty}\delta(t)dt=1. \tag{4. 38}$$

Дифференциальные уравнения (ДУ) и передаточная функция являются наиболее общими формами связи между переменными состояния на входе и выходе линейной системы. Но в реальных условиях часто наблюдаются только сигналы управления и реакции систем, по которым необходимо получить уравнение связи.

Таким образом, с помощью эксперимента можно получить график, определяющий частное решение при определенном входном сигнале. Затем, аппроксимировав аналитическим выражением полученные реализации, можно построить ДУ заданной структуры и записать его в одной из форм, приведенных выше (т. е. получить модель объекта).

Большое распространение получили методы идентификации детерминированных объектов путем определения переходной характеристики $h(t)$ по кривой разгона при ступенчатом изменении управления на входе

$$h(t)\approx\frac{y(t)-y_0}{u-u_0}, \tag{4.39}$$

где $y(t)$ – изменение выходной величины объекта при подаче на его вход ступенчатого управления $u$ (кривая разгона), $y_0$ и $u_0$ установившиеся значения выхода и входа объекта до начала проведения эксперимента.

Если объект управления не допускает изменения выходной координаты, то на его вход вместо ступенчатого воздействия подают единичный импульс или серию импульсов.

При проведении эксперимента по снятию кривой разгона необходимо тщательно изолировать объект от случайных возмущений, стремиться, как можно точнее воспроизвести заданную форму возмущения на входе, дублировать проведение экспериментов по снятию кривой разгона для различных начальных значений $y_0$ и $u_0$.

Рассмотрим определение передаточной функции объекта по кривой разгона логарифмическим методом. Преимущество этого метода состоит в том, что результаты идентификации получаются в виде аналитического выражения, хорошо поддающегося дальнейшей машинной обработке. Методика идентификации в этом случае [26]

заключается в аппроксимации переходной характеристики аналитическим выражением типа

$$h(t) = k_0 + \sum_{k=1}^{n} C_k e^{p_k t}, \qquad (4.40)$$

где $k_0 = \dfrac{y(\infty) - y_0}{u - u_0}$ – коэффициент передачи объекта $y(\infty)$ – установившееся значение выходной величины объекта, соответствующее частному решению его ДУ и определяемое вынужденным движением под действием входного сигнала, остальные слагаемые определяют свободные движения и представляют общее решение однородного ДУ объекта, $C_к$ – постоянные интегрирования, $p_к$ – корни характеристического уравнения.

Положим, для определенности, что в (4.39) имеется один действительный корень, два комплексно сопряженных и два кратных корня. Логарифмируя (4.39), получим

$$z = ln\left[h(t) - k_0\right] = ln\left[\sum_{k=1}^{5} C_k e^{p_k t}\right]. \qquad (4.41)$$

Для устойчивого объекта свободные движения с течением времени стремятся к нулю, причем время переходного процесса будет определяться корнем, имеющим минимальную действительную часть (например, действительным корнем). Тогда начиная с некоторого момента времени слагаемыми в (4.41), имеющими большие действительные части можно пренебречь и приближенно записать

$$z = \ln\left[h(t) - k_0\right] \approx \ln C_1 + p_1 t\ . \qquad (4.42)$$

Уравнение (4.42) является асимптотой (4.41) при $t \to \infty$.

Если теперь в полулогарифмическом масштабе построить график уравнения (4.41), откладывая на оси абсцисс время на оси ординат $z=ln[h(t)-k_o]$, то неизвестные коэффициенты $C_1$, $p_1$ легко определить графически, как показано на рис 4.4.

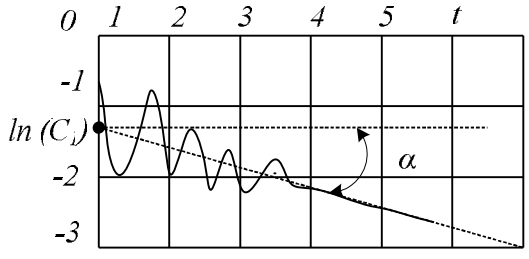



Рис. 4.4. График определения неизвестных параметров $C_1$,$p_1$

Ордината асимптоты при $t=0$ равна $ln((C_1)$, а $p_1= tan(\alpha)$. После чего из (4.4) можно исключить одно слагаемое соответствующее действительному корню, тогда

$$z_1 = \ln\left[h(t) - k_0 - C_1 e^{p_1 t}\right] = \ln\left[\sum_{k=1}^{4} C_k e^{p_k t}\right]. \qquad (4.43)$$

Аналогично, начиная с некоторого момента времени слагаемыми в (4.43), имеющими большие действительные части можно пренебречь и приближенно записать

$$h(t) - k_0 - C_1 e^{p_1 t} \approx C_2 e^{-\lambda t} \sin(\omega t + \varphi). \qquad (4.44)$$

Логарифмируя, получим

$$z_1 = \ln\left[h(t) - k_0 - C_1 e^{p_1 t}\right] \approx \ln(C_2) - \lambda t + \ln\left[\left|\sin(\omega t + \varphi)\right|\right]. \qquad (4.45)$$

Уравнение прямой на графике (рис. 4.5), проходящей через точки в которых $sin\ (\omega t + \varphi) = 1$ имеет вид:

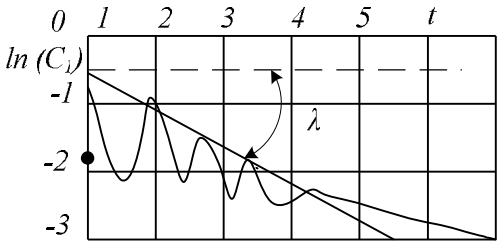



Рис. 4.5. График определения неизвестных параметров $C_2$, λ

Неизвестные параметры $C_2$ и λ находятся так же, как и в предыдущем случае.

Круговую частоту ω и начальную фазу φ можно определить из очевидного условия $sin(\omega t + \varphi) = \pm 1$

$$\frac{\pi}{2} + (i-1)\pi = \omega t_i + \varphi, \qquad (4.46)$$

где $t_i$ – значения времени при которых ординаты переходной характеристики принимают экстремальные значения.

Минимизируя квадрат невязки (4.46) по $\omega$ и $\varphi$, получим систему алгебраических уравнений для их определения

$$\sum_{i=1}^{m} t_i^2 f + \sum_{i=1}^{m} t_i b = \frac{1}{4}\sum_{i=1}^{m}(2i-1)t,$$

$$\sum_{i=1}^{m} t_i f + mb = \frac{1}{4}\sum_{i=1}^{m}(2i-1), \qquad (4.47)$$

где $f = \frac{\omega}{2\pi}$ – неизвестная частота, $b = \frac{\varphi}{2\pi}$ – относительная фаза, $m$ – число экстремальных значений переходной характеристики.

Коэффициенты последних слагаемых от корней, имеющих кратность – два, в уравнении переходной характеристики (4.40) получают из выражения аналогичного (4.43)

$$z_2 = \ln\left[h(t) - k_0 - C_1 e^{p_1 t} - C_2 e^{-\lambda t}\sin(\omega t + \varphi)\right] = \ln\left[\sum_{k=1}^{2} C_k e^{p_k t}\right]. \qquad (4.48)$$

Приближенно при больших $t$ можно записать

$$z_2 \approx \ln(C_3) + \ln(t) + pt\,, \qquad (4.49)$$

где $p=p_4=p_5$ – корень двойной кратности.

Асимптотой этого уравнения будет прямая, тангенс угла наклона которой будет равен $p$.

Если теперь простроить в логарифмических координатах функцию $z_2 - pt$, то ее асимптотой будет $\ln(t) + \ln(C_3)$. Тангенс угла наклона асимптоты дает кратность корня. Ордината асимптоты при $t=0$ равна $\ln(C_3)$.

Зная корни выражение для переходной характеристики нетрудно определить и передаточную функцию через преобразование Лапласа

$$W(p) = pL\left[h(t)\right] = p\int_0^\infty h(t)e^{-pt}dt\,. \qquad (4.50)$$

Удовлетворительные результаты при использовании этого метода получаются в том случае, если корни характеристического уравнения далеко отстоят друг от друга. Желательно, чтобы каждый следующий корень был в 2–3 раза меньше предыдущего. Следует отметить, что точность определения характеристик объекта с помощью временных характеристик невысока. Качество идентификации снижают случайные помехи, искажающие реакцию объекта и неточность аппроксимации объекта линейной моделью. Все это не позволяет считать эти методы перспективными.

Пример 4.2. Проведем идентификацию объекта с передаточной функцией

$$W(p) = \frac{2{,}5}{(p + 0{,}1)(p^2 + 6p + 25)}\,.$$

Корни характеристического уравнения объекта будут равны $p_1$ = - 0,1; $p_2$ = -3 + $j$ 4; $p_3$ = -3- $j$ 4, а его коэффициент передачи $\kappa_o$=1.

Зная передаточную функцию объекта нетрудно построить его переходную характеристику $h(t)$ (рис. 4.6), используя преобразование Лапласа

$$h(t) = L\left[\frac{W(p)}{p}\right].$$

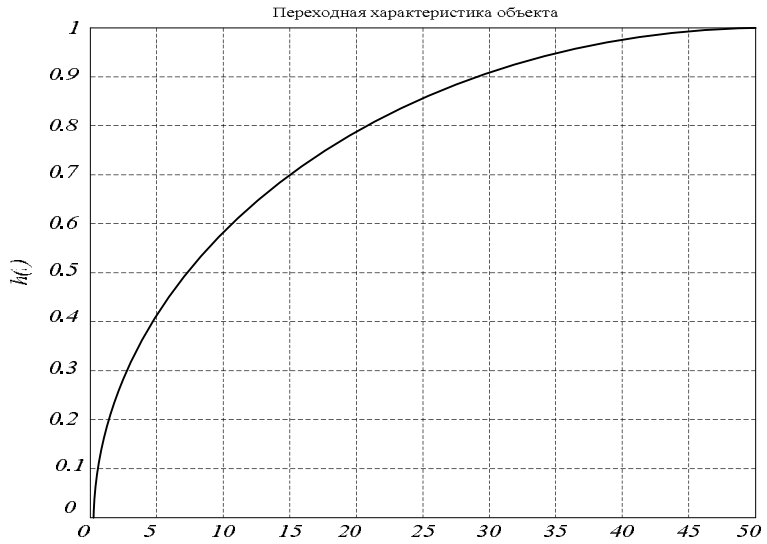



Рис. 4.6. Переходная характеристика $h(t)$

По переходной характеристике определяем коэффициент передачи $k_o = h(\infty) = 1$, и строим по формуле (4.41) график относительной переходной характеристики в полулогарифмическом масштабе (рис. 4.7).




Рис. 4.7. График относительной переходной характеристики

По графику находим значение коэффициента $C_1$= 1,0242 и оценку первого корня характеристического уравнения $p_1$= - 0,1, которая совпадает с его истинным значением.

Исключаем из переходной характеристики слагаемое, соответствующее действительному корню и строим по формуле (4.45), новую переходную характеристику (рис. 4.8).

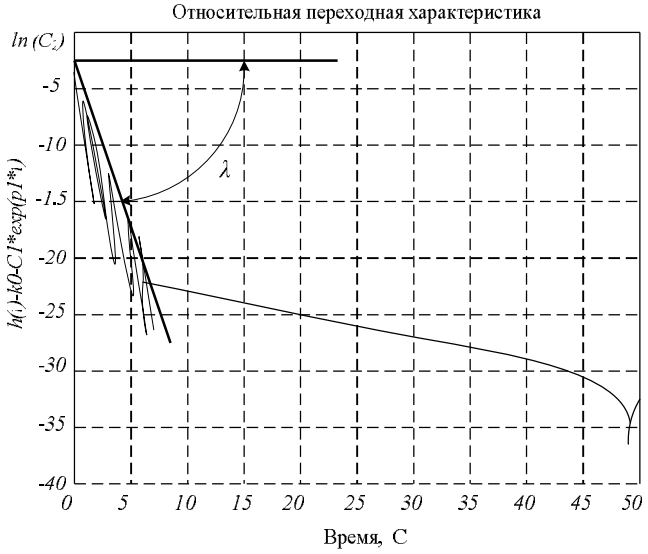



Рис. 4.8. Переходная характеристика по формуле (4.45)

Из графических построений и по формулам (4.48) находим оценки параметров объекта $\hat{p}_2 = -2{,}9959 + j3{,}0084$; $\hat{p}_3 = -2{,}9959 - j3{,}0084$, хорошо совпадающие с истинными значениями.

## 4.4. Определение передаточной функции объекта по частотным характеристикам

Этот метод идентификации относится к непараметрическим методам, т. к. вначале экспериментально снимают частотные характеристики объекта, а затем уже по полученным экспериментальным характеристикам вычисляют передаточную функцию. Основные трудности при проведении экспериментов заключаются в определении рабочего диапазона частот и дрейфе оси колебаний на выходе объекта. Чаще всего область рабочих частот задается ориентировочно, и наибольшее внимание уделяется диапазону, в котором сдвиг по фазе между входным и выходным гармоническими сигналами составляет $180^0$.

При снятии частотных характеристик используют различные методы воздействия на объект.

Метод синусоидальной волны предполагает подачу на вход объекта гармонических колебаний. На каждой из выбранных в пределах рабочего диапазона частот проводится отдельный опыт. На входе исследуемого объекта возбуждаются колебания выбранной частоты. Процессу колебаний дают установиться и, когда ось колебаний, их форма и амплитуда станут неизменными, измеряют амплитуды входных и выходных колебаний и фазовый сдвиг между ними. Частное от деления амплитуды выходных колебаний на амплитуду входных колебаний дает амплитуду частотной характеристики на взятой частоте, сдвиг по фазе – ординату фазовой частотной характеристики.

Основным затруднением при использовании этого метода является необходимость возбуждения колебаний большой мощности, имеющих правильную синусоидальную форму.

Поэтому чаще применяют метод «прямоугольной» волны. При этом по прибору, измеряющему входную величину объекта, градуируются три положения регулирующего органа, изменяющего эту величину. Среднее положение соответствует значению входной величины, при котором стабилизируется режим работы объекта перед началом испытаний (ось колебаний); два других равноотстоят от среднего положения. Перед началом опыта регулирующий орган устанавливается в среднее положение и удерживается там до установления в объекте стабильного режима. Затем регулирующее устройство через равные промежутки времени, соответствующие полупериоду избранной частоты, переводится из одного крайнего положения в другое и обратно. Эти переключения продолжаются до тех пор, пока на выходе объекта колебания *y(t)* не примут установившиеся формы. После чего производят регистрацию этих колебаний.

На основе полученных осциллограмм колебаний на выходе объекта проводят их гармонический анализ, ограничиваясь вычислением амплитуд и фаз первой и третьей гармоник

$$b_k = \frac{1}{\pi}\int_{-\pi}^{\pi} y(t)\cos\left(k\frac{2\pi}{T}t\right)dt,$$

$$c_k = \frac{1}{\pi}\int_{-\pi}^{\pi} y(t)\sin\left(k\frac{2\pi}{T}t\right)dt, \qquad (4.51)$$

$$a_k = \sqrt{b_k^2 + c_k^2},$$

$$\varphi_k = arctg\left(\frac{b_k}{c_k}\right),$$

где $T$ – период колебаний; $k$ – номер гармоники.

В том случае если значения выходной величины известны только в дискретные, равноотстоящие моменты времени интегралы в (4.51) заменяются суммами

$$b_k = \frac{1}{N}\sum_{I=0}^{N-1} y_i \cos\left(2\pi\frac{ik}{N}\right),$$

$$c_k = \frac{1}{N}\sum_{I=0}^{N-1} y_i \sin\left(2\pi\frac{ik}{N}\right), \qquad (4.52)$$

где $N$ – число дискрет выходного сигнала.

Гармонический анализ входного прямоугольного сигнала приводит к такому выражению

$$u(t) = \frac{4A}{\pi}\left[\sin\left(\frac{2\pi}{T}t\right) + \frac{1}{3}\sin\left(3\frac{2\pi}{T}t\right) + \frac{1}{5}\sin\left(5\frac{2\pi}{t}t\right) + ...\right], \qquad (4.53)$$

где $A$ – амплитуда прямоугольной волны.

Вычисляют амплитуды и фазы входных и выходных гармонических характеристик на выбранной частоте, как отношение амплитуд гармонических составляющих на выходе и входе объекта и значения фазочастотной характеристики, как соответствующий фазовый сдвиг $\varphi_k$.

Для повышения точности определения частотных характеристик рекомендуется использовать при гармоническом анализе только первую гармонику.

Для того чтобы не допускать больших отклонений регулируемых составляющих, можно вычислить значения амплитудно-частотной величины в процессе проведения опытов переключения регулирующего органа, который осуществляют в моменты времени, когда регулируемая (выходная) величина достигает заранее выбранные граничные значения. Для реализации такого режима используют двухпозиционный регулятор с регулируемой зоной нечувствительности. Изменяя величину зоны нечувствительности и знак обратной связи можно изменять частоту возникающих в такой системе автоколебаний.

Эксперимент обычно начинают на частоте, при которой фазовый сдвиг между входным и выходным сигналами равен $\pi$. Это дос-

тигается при нулевой зоне нечувствительности позиционного регулятора и отрицательной обратной связи. Увеличивая зону нечувствительности при отрицательной обратной связи, уменьшают частоту автоколебаний. Увеличивая зону нечувствительности при положительной обратной связи, увеличивают частоту автоколебаний.

Результаты экспериментов обрабатываются так же, как и в методе прямоугольной волны. Основным недостатком рассмотренных методов является длительное время эксперимента, затрачиваемое в основном на ожидание установившегося режима колебаний и получения достаточного для аппроксимации частотных характеристик значений их ординат. Для ускорения экспериментов иногда на вход объекта подается сумма гармонических составляющих разных частот.

Установившиеся колебания выходной величины также подвергают гармоническому анализу и сразу находят несколько ординат частотной характеристики. Однако в этом случае требуется специальный источник полигармонического воздействия и линейность объекта.

Определение аналитического выражения передаточной функции по частотным характеристикам может быть осуществлено несколькими способами. Рассмотрим один из них [23, 26] позволяющий аналитически вычислить выражение для передаточной функции по дискретным ординатам вещественной частотной характеристики.

Будем искать выражение для передаточной функции в виде ряда

$$W(p) = \sum_{k=0}^{\infty} A_k \left( \frac{1-p}{1+p} \right)^k . \qquad (4.54)$$

Тогда делая подстановку

$$tg\left( \frac{\varphi}{2} \right) = \omega \qquad (4.55)$$

выражение для вещественной частотной характеристики можно записать в виде

$$U(\varphi) = \sum_{k=0}^{\infty} A_k \cos(k\varphi) . \qquad (4.56)$$

Если теперь провести гармонический анализ экспериментально полученной вещественной частотной характеристики, то в результате будут получены неизвестные коэффициенты передаточной функции *Ак.*

Ограничиваясь значащими членами ряда (4.56) можно записать выражение передаточной функции объекта в виде (4.54).

Ограничение на использование данного метода вычисления передаточной функции относятся к самовыравнивающимся, устойчивым и неминимально-фазовым объектам. В том случае если объект не отвечает этим требованиям, то при вычислении вещественной частотной характеристики необходимо учесть влияние интегрирующих звеньев и звеньев запаздывания на вид этой характеристики, и проводить ее расчет с учетом этого влияния.

Компенсация запаздывания осуществляется выражением

$$U_ç(\omega) = U(\omega)\cos\omega\tau - V(\omega)\sin\omega\tau, \qquad (4.57)$$

а компенсация интегрирующих звеньев –

$$U_u(\omega) = \frac{V(\omega)}{\omega}, \qquad (4.58)$$

где $U(\omega)$ и $V(\omega)$ – экспериментальные вещественная и мнимая частотные характеристики.

## 4.5. Статистическая идентификация линейных устойчивых стационарных объектов (метод корреляционных функций) [20, 38]

### 4.5.1. Интегральное уравнение Фредгольма 1-го рода, определяющее импульсную переходную функцию идентифицируемого объекта.

Рассмотренная в предыдущем параграфе детерминированная идентификация линейных стационарных устойчивых объектов имеет недостатки, два из которых оказывают существенное влияние на точность решения задачи:

- наличие источника ошибок идентификации, порожденных неконтролируемыми шумами *n(t),* которые приводят к тому, что выходной процесс определяется формулой

$$x(t) = x_y(t) + x_n(t);$$

- наличие источника ошибок за счет $x^0 = (x(0), x'(0), \ldots, x^{(n-1)}(0)) \neq 0$.

Статистическая идентификация позволяет устранить влияние указанных факторов. Положим, что на вход стационарного линейного

объекта, имеющего установившийся режим работы, поступает случайный стационарный процесс *u(t),* тогда

$$x(t) = \int k(\lambda)u(t-\lambda)d\lambda\,, \tag{4.59}$$

$$M\left[u(t-\tau)x(t)\right] = R_{xu}(\lambda) = M\left[u(t-\tau)\int_0^\infty k(\lambda)u(t-\lambda)d\lambda\right] = M\left[u(t-\lambda)\int_0^\infty k(\lambda)u(t-\lambda)d\lambda\right] =$$

$$= \int_0^\infty k(\lambda)M\left[x(t-\tau)x(t-\lambda)\right]d\lambda = \int_0^\infty k(\lambda)M\left[R_{uu}(t-\lambda-t+\tau)\right]d\lambda = \int_0^\infty R_{uu}(\tau-\lambda)k(\lambda)d\lambda = R_{xu}(\lambda).$$

Имеем, в основе статистических методов определения импульсных переходных функций лежит интегральное уравнение

$$R_{xu}(\lambda) = \int_0^\infty R_{uu}(\tau-\lambda)k(\lambda)d\lambda, \quad -\infty < \lambda > \infty, \tag{4.60}$$

позволяющее по корреляционной функции сигнала *u(t)* на входе $R_{uu}(\tau)$ исследуемого объекта и взаимной корреляционной функции $R_{xu}(\tau)$ между выходом *x(t)* и входам *u(t)* найти импульсную переходную функцию $k(\lambda)$.

Предположим, что к объекту приложено не только воздействие *u(t),* но и помеха *n(t).* Тогда справедлива зависимость

$$x(t) = \int_0^\infty u(t-\tau)k(\lambda)d\tau + \int_0^\infty n(t-\tau)k_n(\tau)d\tau\,. \tag{4.61}$$

Определить $k(\tau)$ из последней формулы методами, описанными в предыдущем параграфе, достаточно трудно, т. к. второй член в правой части является источником погрешности.

Умножим обе части формулы (4.61) на $u(t-\lambda)$ и произведем усреднение по времени

$$M\left[x(t)u(t+\lambda)\right] = \int_0^\infty M\left[u(t-\tau)u(t-\lambda)\right]k(\tau)d\tau + \int_0^\infty M\left[n(t-\tau)u(t+\lambda)\right]k_n(\tau)dt_\theta\,.$$

В результате получаем уравнение (4.60), которое остается справедливым при наличии помехи *n(t),* воздействующей на объекту при условии, что она не коррелирована с сигналом *u(t).*

Статистические методы определения импульсной переходной функции подразделяются на требующие и не требующие искусственных воздействий. Первая группа методов основана на том, что к реальному входному сигналу добавляется искусственный вспомогательный сигнал *m(t)* с известной корреляционной функцией. Обычно

сигналом *m(t)* служит белый шум. При отсутствии корреляции между реальным входным сигналом и *m(t)* имеем

$$R_{xm}(\tau) = \int_0^\infty R_{mm}(\tau - \lambda)k(\lambda)d\lambda . \qquad (4.62)$$

Если *m(t)* – белый шум, то $R_{mm}(\tau)$ является дельта-функцией, т. е.

$$R_{mm}(\tau) = 2\pi S_0 \delta(\tau)$$

и выражение (4.62) сводится к виду

$$R_{xm}(\tau) = 2\pi S_0 k(\tau) .$$

Из последней зависимости следует, что определение импульсной переходной функции не требует решения интегрального уравнения. Однако в ряде случаев введение искусственного шума в систему нежелательно и тогда необходимо пользоваться уравнением (4.62).

Отметим, что нахождение импульсной переходной функции из интегрального уравнения (4.62) эквивалентно ее оценке по критерию минимума среднеквадратической ошибки [15, 20, 38].

Действительно, пусть

$$\varepsilon(t) = x(t) - \int_0^\infty k(\tau)u(t-\tau)d\tau , \qquad (4.63)$$

$$J = \frac{1}{T}\int_0^T \varepsilon^2(t)dt . \qquad (4.64)$$

Если функция $k(\tau)$ минимизирует критерий (4.64), то справедливо следующее представление: $k(\tau) = \tilde{k}(\tau) + \alpha k_\alpha(\tau)$, где $k_\alpha(\tau)$ – произвольная функция; $k_\alpha(\tau) = 0$ при $\tau < 0$.

В соответствии с формулой (4.64) можно записать известное выражение $\frac{dJ}{d\alpha}|_{\alpha=0} = 0$.

Поскольку

$$J = \frac{1}{T}\int_0^T \left\{ x(t) - \int_0^\infty \left[\tilde{k}(\tau) + \alpha k_\alpha(\tau)\right]u(t-\tau)d\tau \right\}^2 dt,$$

то

$$\frac{dJ}{d\alpha}|_{\alpha=0} = \frac{2}{T}\int_0^T \left\{ x(t) - \int_0^\infty \tilde{k}(\tau)u(t-\tau)d\tau \right\} \int_0^T \int_0^\infty k_\alpha(\tau)u(t-\tau)d\tau dt = 0 .$$

Отсюда следует

$$\frac{1}{T}\int_0^T\left\{\left[x(t)-\int_0^\infty \widetilde{k}(\tau)u(t-\tau)d\tau\right]\int_0^\infty k_\alpha(\tau)u(t-\tau)d\tau\right\}dt=0$$

или, что то же самое

$$\frac{1}{T}\int_0^T x(t)\int_0^\infty k_\alpha(\tau)u(t-\tau)d\tau dt-\frac{1}{T}\int_0^T\int_0^\infty \widetilde{k}(\tau_1)u(t-\tau_1)d\tau_1\int_0^\infty k_\alpha(\tau_2)u(t-\tau_2)d\tau_2 dt=0$$

получаем

$$\int_0^\infty k_\alpha(\tau_2)\left\{\frac{1}{T}\int_0^T x(t)u(t-\tau)dt-\int_0^\infty \widetilde{k}(\tau_1)d\tau_1\frac{1}{T}\int_0^T u(t-\tau_1)u(t-\tau_2)dt\right\}d\tau_2=0\,.$$

Далее находим

$$\frac{1}{T}\int_0^T x(t)u(t-\tau)dt=\int_0^\infty \widetilde{k}(\lambda)d\lambda\frac{1}{T}\int_0^T u(t-\tau)u(t-\tau)dt$$

или

$$\widetilde{R}_{xu}(\tau,T)=\int_0^\infty R_{uu}(\tau-\lambda,T)\widetilde{k}(\lambda)d\lambda\,.$$

Условие эргодичности процессов при $T\to\infty$ позволяет записать интегральное уравнение (4.62), называемое уравнением Винера-Хопфа**.** Оно принадлежит к классу уравнений Фредгольма 1-го рода.

Существуют формулы, позволяющие выбрать интервал наблюдений *T,* Например, если сигнал *x(t)* имеет автокорреляционную функцию вида $R_{xx}(\tau)=D_{xx}e^{\alpha|\tau|}\cos\omega_0\tau$, то $T\geq 20(\frac{\alpha}{\alpha^2+\omega_0^2}+\frac{1}{\alpha})$, если же $R_{xx}(\tau)=e^{\alpha|\tau|}$, то $T\geq 40/\alpha$.

Получены формулы, при помощи которых можно, задаваясь корреляционной функцией, полученной на основе опыта статистической обработки аналогичных процессов, и, исходя из заданной среднеквадратичной ошибки, выбрать значение *T*.

В соответствии с уравнением (4.62) для вычисления *k(τ)* необходимо выполнить два основных этапа (рис. 4.9):

1) определение корреляционных функций $R_{uu}(\tau)$ и $R_{xu}(\tau)$*;*

2) решение интегрального уравнения Винера-Хопфа.

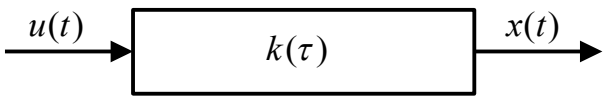



Рис. 4.9. К выводу интегрального уравнения Винера-Хопфа

Справедливы следующие положения [15, 22, 23].

Пусть $k(\tau)$ – импульсная переходная функция линейной стационарной системы, а $R_{хи}(\tau)$ и $R_{ии}(\tau)$ – взаимные корреляционные функции входного и выходного сигналов и автокорреляционная функция входного сигнала соответственно. Тогда $k(\tau)$ удовлетворяет следующему уравнению свертки:

$$R_{хи}(\tau)=\int_0^{\infty} k(\tau)R_{ии}(\tau-\lambda)d\lambda,\ \tau\in(-\infty,+\infty).$$

Тогда $k(\tau)$ удовлетворяет уравнению Винера-Хопфа

$$R_{хи}(\tau)=\int_0^{\infty} k(\tau)R_{ии}(\tau-\lambda)d\lambda,\ \tau\in(0,+\infty).$$

Пусть

$$R_{ии}(\tau)=\begin{cases}R_{ии}^{+}(\tau),\tau\ge 0;\\ R_{ии}^{-}(\tau),\tau<0;\end{cases}\qquad R_{хи}(\tau)=\begin{cases}R_{хи}(\tau),\tau\ge 0;\\ R_{хи}(\tau),\tau<0;\end{cases}$$

тогда ИПФ удовлетворяет уравнениям

$$R_{хи}^{+}(\tau)-R_{хи}^{-}(\tau)=\int_0^{\infty} k(\lambda)\left[R_{ии}^{+}(\tau-\lambda)-R_{ии}^{-}(\tau-\lambda)\right]d\lambda,\tau\ge 0,$$

$$R_{хи}^{-}(\tau)=\int_0^{\infty} k(\lambda)R_{ии}(\tau-\lambda)d\lambda,\tau<0.$$

Пусть $R_{ии}^{-}(\tau)$ и $R_{хи}^{-}(\tau)$ – аналитические функции, допускающие аналитическое продолжение на дополнительную полуось, и выполнено соотношение $R_{хи}^{+}(\tau)\ne R_{хи}^{-}(\tau)$ при $\tau\ge 0$. Тогда справедлива следующая формула для передаточной функции *W(s)* линейной системы:

$$W(s)=\frac{R_{хи}^{+}(s)-R_{хи}^{-}(s)}{R_{ии}^{+}(s)-R_{ии}^{-}(s)},$$

где через $R_{хи}^{-}(s),\ R_{хи}^{-}(s),\ R_{ии}^{+}(s)$ и $R_{ии}^{-}(s)$ обозначены преобразования Лапласа от функций $R_{хи}^{-}(\tau),\ R_{хи}^{-}(\tau),\ R_{ии}^{+}(\tau)$ и $R_{ии}^{-}(\tau)$ соответственно.

Пусть функция *F(t)* определяется формулой

$$F(\tau)=\int_0^{\infty} k(\lambda)R_{ии}^{-}(\tau-\lambda)d\lambda,$$

тогда уравнение Винера-Хопфа может быть представлено в виде уравнения, в котором передаточная функция определяется по формуле

$$W(s)=\frac{R_{xu}^{+}(s)-F(s)}{R_{uu}^{+}(s)-R_{uu}^{-}(s)},$$

где *F(s)* – преобразование Лапласа от функции F(t).

Пусть функции $R_{uu}^{+}(\tau-\lambda)$ и $R_{uu}^{-}(\tau-\lambda)$ допускают представление

$$R_{uu}^{+}(\tau-\lambda)=\sum_{i=1}^{l}d_i(\tau)v_i(\lambda) \text{ при } \tau\geq\lambda,$$

$$R_{uu}^{-}(\tau-\lambda)=\sum_{i=1}^{l}v_i(\tau)d_i(\lambda) \text{ при } \tau<\lambda,$$

а постоянные $c_i$ вычисляются по формуле $c_i=\int_0^{\infty}k(\lambda)d_i(\lambda)d\lambda$.

Тогда передаточная функция системы определяется формулой

$$W(s)=\frac{R_{xu}^{+}(s)-\sum_{i=1}^{l}c_iV_i(s)}{R_{uu}^{+}(s)-R_{uu}^{-}(s)}.$$

Метод идентификации объектов управления, рассмотренный выше, справедлив для следующих двух случаев:

1) объект изолирован, т. е. не замкнут обратной связью; внутренние помехи присутствуют или отсутствуют;

2) объект замкнут обратной связью; внутренние помехи отсутствуют. Для первого случая изложенный метод справедлив в силу принятой гипотезы о независимости внутренних шумов от внешних воздействий.

Перейдем теперь к рассмотрению общего случая: объект, замкнутый обратной связью, работает в условиях внутренних помех (рис. 4.10) [20].

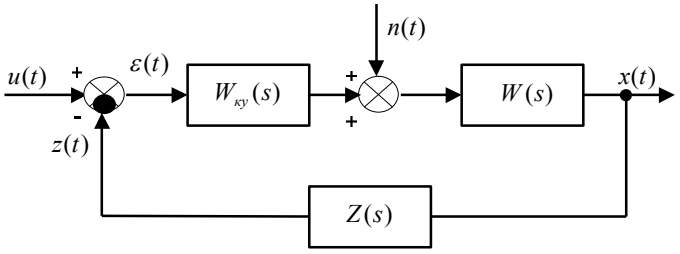



Рис. 4.10. Структурная схема автоматического управления с внутренней помехой

Уравнения этой системы могут быть представлены в следующем виде:

$$x(s) = W_{ky}(s)W(s)E(s) + W(s)N(s), \tag{4.65}$$

$$E(s) = u(s) - Z(s)x(s), \tag{4.66}$$

где $W_p(s) = W_{ky}(s)W(s), k_p(\tau)$ – соответственно передаточная и импульсная переходная функции прямой цепи; $W(s)$, $k(\tau)$ – передаточная и импульсная переходная функции объекта по отношению к внутреннему шуму; $Z(s), l(\tau)$ – передаточная и импульсная переходная функции обратной связи.

Из уравнения (4.29) получаем

$$R_{x\varepsilon}(\tau) = \int_0^\infty R_{\varepsilon\varepsilon}(\tau - \vartheta)k_p(\vartheta)d\vartheta + \int_0^\infty R_{n\varepsilon}(\tau - \vartheta)k(\vartheta)d\vartheta. \tag{4.67}$$

Как видно из уравнений (4.65), (4.67), введение обратной связи приводит к тому, что ошибка $\varepsilon(t)$ коррелированна с внутренними помехами $R_{n\varepsilon}(\tau) \neq 0$ и соотношение (4.62) не выполняется.

Ниже приведены возможные методы определения динамических характеристик линейных объектов при наличии обратных связей и внутренних помех [20].

Косвенный метод определения динамических характеристик объекта, замкнутого обратной связью. Найдем импульсные переходные функции замкнутой системы, а затем по ним получим динамические характеристики самого объекта.

Учитывая выражения (4.65), (4.67), имеем

$$x(s) = \hat{O}(s)E(s) + W_n(s)N(s), \tag{4.68}$$

где $\hat{O}(s), k_3(\tau)$ – соответственно передаточная и импульсная переходная функции замкнутой системы; $W_n(s), q(\tau)$ – соответственно передаточная и импульсная переходная функции системы по отношению к шуму;

$$\hat{O}(s) = \frac{W_p(s)}{1 + W_p(s)Z(s)}, \qquad W_n(s) = \frac{W(s)}{1 + W_p(s)Z(s)}.$$

Через импульсные переходные функции соотношение (4.68) запишется в виде

$$x(t) = \int_0^\infty u(t - \vartheta)k_3(\vartheta)d\vartheta + \int_0^\infty n(t - \vartheta)q(\vartheta)d\vartheta. \tag{4.69}$$

Используя последнее выражение и условие

$$R_{nu}(\tau) = 0, \tag{4.70}$$

получим

$$R_{xu}(\tau) = \int_0^\infty R_{uu}(\tau - \vartheta) k_3(\vartheta) d\vartheta. \tag{4.71}$$

Аналогичное соотношение

$$R_{Zx}(\tau) = \int_0^\infty R_{xx}(\tau - \vartheta) l(\vartheta) d\vartheta \tag{4.72}$$

можно записать для обратной связи.

Из уравнений (4.71), (4.72) находим $k_3(\tau)$ и $l(\tau)$, а по ним *Ф(s)* и *Z(s)* передаточную функцию прямой цепи определим по формуле

$$W_p(s) = \frac{\hat{O}(s)}{1 + \hat{O}(s) Z(s)}. \tag{4.73}$$

Остановимся теперь на приближенном способе.

Приближенный способ получения импульсных переходных функций объекта, замкнутого обратной связью при наличии внутренних помех.

Не нарушая общности отнесем помехи, возникающие внутри объекта, к его выходу, заменив $n(t)$ на $f(t)$, тогда

$$F(s) = W(s) N(s), \tag{4.74}$$

как показано на рис. 4.11.

Тогда

$$E(s) = \frac{1}{1 + W_{ky}(s) W(s) Z(s)} u(s) - \frac{Z(s)}{1 + W_{ky}(s) W(s) Z(s)} F(s), \tag{4.75}$$

или

$$\varepsilon(t) = \int_0^\infty u(t - \vartheta) \lambda(\vartheta) d\vartheta - \int_0^\infty f(t - \vartheta) \mu(\vartheta) d\vartheta, \tag{4.76}$$

где $\lambda(t), \mu(t)$ – импульсные переходные функции, соответствующие передаточным функциям

$$\frac{1}{1 + W_{ky}(s) W(s) Z(s)}, \quad \frac{Z(s)}{1 + W_{ky}(s) W(s) Z(s)}.$$

Используя уравнение (4.76) и условие $R_{uf}(\tau) = 0$, получим

$$R_{f\varepsilon}(\tau) = -\int_0^\infty R_{ff}(\tau - \vartheta)\mu(\vartheta)d\vartheta \,. \qquad (4.77)$$

Если помеха *f(t)* является белым шумом, то $R_{ff}(\tau) = \delta(\tau)$, где $\delta(\tau)$ – дельта-функция.

В этом случае

$$R_{f\varepsilon}(\tau) = -\int_0^\infty \mu(\vartheta)\delta(\tau + \vartheta)d\vartheta = -\mu(-\tau)\,, \qquad (4.78)$$

т. е. $R_{f\varepsilon}(\tau)$ является «перевернутой» импульсной переходной функцией замкнутой системы, как показано на рис. 4.12(*а*),

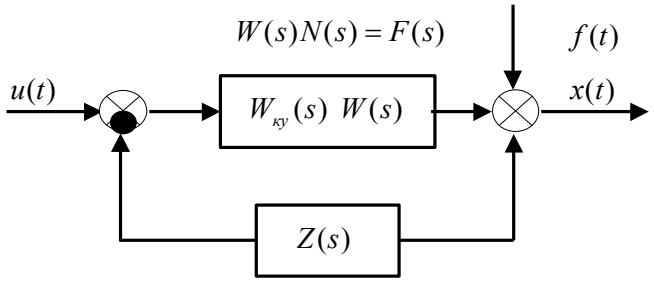



Рис. 4.11. Преобразованная структурная схема системы
с внутренней помехой

В большинстве случаев, когда помеха не является белым шумом $R_{ff}(\tau)$, полученная из эксперимента, имеет большой центральный пик и стремится, к нулю по обе стороны от $\tau = 0$. Для этого случая уравнение (4.78) интерпретировано на рис. 4.12 (*б*)**.** Вследствие некоторой протяженности $R_{ff}(\tau)$ взаимная корреляционная функция $R_{fe}(\tau)$ начинается на расстоянии *a* справа от $\tau = 0$. Чем более острый пик имеет $R_{ff}(\tau)$, тем ближе $R_{fe}(\tau)$ приближается к $-\mu(-\tau)$.

Учитывая уравнение (4.74), имеем

$$x(t) = \int_0^\infty \varepsilon(\tau - \vartheta)k_p(\vartheta)d\vartheta + f(t)\,. \qquad (4.79)$$

Используя уравнение (4.79), получим

$$R_{x\varepsilon}(\tau) = \int_0^\infty R_{\varepsilon\varepsilon}(\tau - \vartheta)k_p(\vartheta)d\vartheta + R_{f\varepsilon}(\tau)\,. \qquad (4.80)$$

Как видно из рис. 4.13, значениями $R_{fe}(\tau)$ при $\tau > a$ можно пренебречь и свести задачу определения $k_p(\tau)$ к решению уравнения

$$R_{x\varepsilon}(\tau) = \int_0^\infty R_{\varepsilon\varepsilon}(\tau - \vartheta) k_p(\vartheta) d\vartheta \qquad (4.81)$$

при

$$\tau > a\,. \qquad (4.82)$$

Условие (4.82) накладывает определенные ограничения на метод решения (4.81) [20].

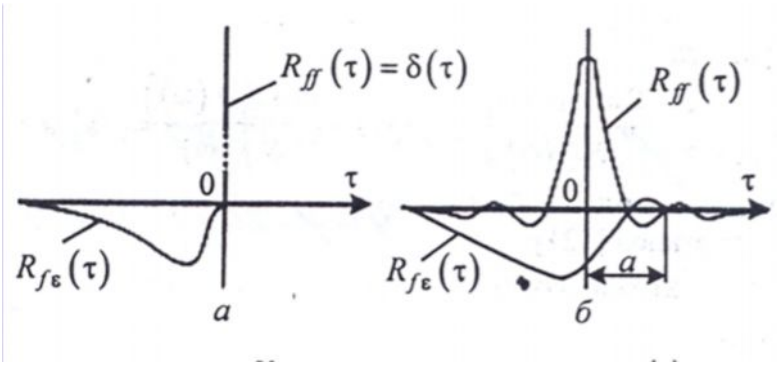



Рис. 4.12. Корреляционная функция $R_{ff}(\tau)$ и «перевернутая» импульсная переходная функция

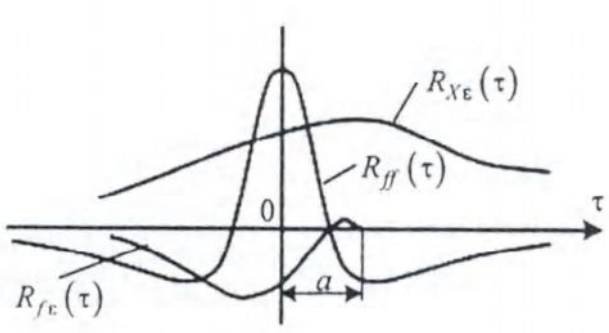



Рис. 4.13. К определению импульсной переходной функции

### 4.3.2. Методы решения уравнения Фредгольма 1-го рода

Решение интегрального уравнения (4.62) в частотной области. Умножим обе части уравнений (4.62) на $e^{jw\tau}$ и проинтегрируем по $\tau$ от $-\infty$ до $+\infty$.

В результате получим

$$\int_{-\infty}^{\infty} R_{xu}(\tau)e^{-jw\tau}d\tau = \int_{0}^{\infty} k(\theta)d\theta \int_{-\infty}^{\infty} R_{uu}(\tau-\theta)e^{-jw\tau}d\tau .$$

Произведя в правой части этого равенства замену переменной по формуле $\lambda = \tau - \theta$, будем иметь

$$\int_{-\infty}^{\infty} R_{xu}(\tau)e^{-jw\tau}d\tau = \int_{0}^{\infty} k(\theta)e^{-jw\tau}d\theta \int_{-\infty}^{\infty} R_{uu}(\lambda)e^{-jw\tau}d\lambda .$$

Вводя спектральные плотности

$$S_{xu} = \int_{-\infty}^{\infty} R_{xu}(\tau)e^{-jw\tau}d\tau ; S_{uu}(\omega) = \int_{-\infty}^{\infty} R_{uu}(\tau)e^{-jw\tau}d\tau$$

и учитывая, что

$$W(j\omega) = \int_{0}^{\infty} k(t)e^{-jwt}dt ,$$

получим

$$W(j\omega) = \frac{S_{xu}(\omega)}{S_{uu}(\omega)} . \qquad (4.83)$$

Из выражения (4.83) следует, что вещественная $P(\omega)$, мнимая $Q(\omega)$, частотные характеристики, соответствующие передаточной функции (4.83), могут быть найдены при помощи формул

$$P(\omega) = \frac{Re[S_{xu}(\omega)]}{S_{uu}(\omega)}, \quad Q(\omega) = \frac{Im[S_{xu}(\omega)]}{S_{uu}(\omega)}. \qquad (4.84)$$

Таким образом, статистический метод определения передаточной функции состоит из двух основных этапов [15, 20]:

1) определения спектральных плотностей по заданным реализациям входа *u(t)* и выхода *x(t)*;

2) определения частотных характеристик или передаточной функции при помощи формул (4.85).

О точности определения динамических характеристик. Воспользовавшись формулой (4.83) запишем импульсную переходную функцию *k(t)* следующим образом:

$$k(\tau) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{S_{xu}(\omega)}{S_{uu}(\omega)} e^{jw\tau}d\omega . \qquad (4.85)$$

Очевидно, что формула (4.85) справедлива лишь при условии, что $S_{uu}(\omega) \neq 0$ при всех значениях $\omega$. Если же на некотором интерва-

ле $(\omega_1, \omega_2)$ $S_{uu}(\omega) = 0$, а $S_{xu}(\omega) \neq 0$, то подынтегральное выражение в формуле (4.85) обращается в бесконечность в указанном интервале. При этом импульсная переходная функция *k(t)* может отличаться от истинной сколь угодно сильно. Но передаточная функция *W(jw)* будет отличаться от истинной только в интервале частот $(\omega_1, \omega_2)$, которые и не содержались в воздействии *u(t)*, т. к. в этом интервале $S_{uu}(\omega) = 0$.

Рассмотрим теперь случай, когда исходные корреляционные функции известны с некоторой ошибкой. Тогда и соответствующие спектральные плотности также будут содержать ошибку. В этом случае выражение (4.85) может быть записано в виде

$$k(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{S_{xu}(\omega) + \delta_{xu}(\omega)}{S_{uu}(\omega) + \delta_{uu}(\omega)} e^{jw\tau} d\omega, \qquad (4.86)$$

где $\delta_{xu}(\omega)$ и $\delta_{uu}(\omega)$ – спектры ошибок.

Непосредственно из рассмотрения формулы видно, что сколь угодно малая ошибка $\delta$ может вызывать сколь угодно большое искажение решения уравнения (4.85), если спектр ее распределен на тех участках оси $\omega$, где величины $S_{xu}(\omega)$ и $\delta_{xu}(\omega)$ становятся сравнимыми с $S_{uu}(\omega)$ и $\delta_{uu}(\omega)$. Из сказанного следует, что необходимое условие для точного определения функции *k(t) по* формуле (4.85) состоит в том, чтобы спектр частот входного сигнала был значительно шире предполагаемой полосы пропускания исследуемого объекта.

От любого метода определения динамической характеристики разумно требовать такого приближения $k(\tau)$, которое сводило бы разность выходных процессов реального объекта и его модели к некоторой малой величине при одном и том же процессе на входе. Действительно, только по отклонению выходных величин при одинаковом процессе на входе мы можем судить о неидентичности объектов с точки зрения их динамических характеристик.

Естественно считать критерием или мерой точности определения динамической характеристики дисперсию этой разности, т. е.

$$M\left\{\left[x_m(t) - x(t)\right]^2\right\} = M\left\{\varepsilon^2\right\}. \qquad (4.87)$$

Очевидно, что

$$M\left\{\varepsilon^2\right\} = \frac{1}{2\pi} \int_{-\infty}^{\infty} \left|W_M(jw) - W(jw)\right|^2 S_{uu}(\omega) d\omega. \qquad (4.88)$$

Введем в рассмотрение относительное среднеквадратичное отклонение

$$\delta_\varepsilon = \sqrt{\frac{\int\limits_{-\infty}^{\infty} \left|W_M(jw) - W(jw)\right|^2 S_{uu}(\omega)d\omega}{\int\limits_{-\infty}^{\infty} \left|W(jw)\right|^2 S_{uu}(\omega)d\omega}} . \qquad (4.89)$$

Физический смысл этого критерия заключается в том, что при оценке погрешности определения динамической характеристики принимается во внимание реакция объекта лишь на те частоты, которые присутствуют во входном процессе, причем с весом, пропорциональным спектральной плотности входного процесса $\mu(t)$.

Критерий (4.88) дает ответ и на вопрос, какую информацию содержит решение (4.86) уравнения (4.87) о динамике исследуемого объекта. Действительно, в случае абсолютно точного определения $k(\tau)$ имеем $\delta_\varepsilon = 0$.

Если $k(\tau)$ определена не точно, но спектр ошибки сосредоточен в той области, где $S_{uu}(\omega) = 0$, то $\delta_\varepsilon = 0$. Это означает, что по выходу мы не сможем отличить модель от реального объекта.

Если $k(\tau)$ определена не точно и спектр ошибки распределен произвольным образом, то $\delta_\varepsilon \neq 0$ и критерий (4.89) характеризует погрешность определения динамической характеристики на тех частотах, присутствие которых во входном процессе существенно. Поэтому, несмотря на возможное большое отклонение $k_M(\tau)$ от $k(\tau)$, ошибка может быть малой, и выходные процессы реального объекта и его модели будут отличаться лишь незначительно.

2. Метод квадратур. Рассмотрим решение уравнения (4.62), пользуясь методом квадратур. Представим уравнение (4.62) в виде

$$R_{xu}(\tau) = \sum_{N=0}^{N} K(n\Delta) R_{uu}(\tau - n\Delta)\Delta .$$

Обозначим

$$A = \begin{pmatrix} R_{uu}(0) & R_{uu}(\Delta) & \cdots & R_{uu}((N-1)\Delta) \\ R_{uu}(\Delta) & R_{uu}(0) & \cdots & R_{uu}((N-2)\Delta) \\ \vdots & \vdots & \ddots & \vdots \\ R_{uu}((N-1)\Delta) & R_{uu}((N-2)\Delta) & \cdots & R_{uu}(0) \end{pmatrix}, \qquad (4.90)$$

$$K=\begin{pmatrix} k(\Delta) \\ k(2\Delta) \\ k(3\Delta) \\ \vdots \\ k(N\Delta) \end{pmatrix}, \qquad Q=\begin{pmatrix} q_1 \\ q_2 \\ q_3 \\ \vdots \\ q_N \end{pmatrix}, \tag{4.91}$$

где

$$q_1=\frac{R_{xu}(i\Delta)}{\Delta}. \tag{4.92}$$

Тогда (4.89) перепишем в виде

$$AK=Q. \tag{4.93}$$

В (4.93) $A=(a_{ij})_{i,j=0}^{N-1}$ – квадратная симметричная матрица размерности $N\times N$.

Как и в случае решения уравнения свертки, задача расчета импульсной переходной функции свелась к решению системы линейных алгебраических уравнений. Последняя является симметричной с доминирующей главной диагональю. Это свойство порождено свойствами корреляционных функций, входящих в исходное уравнение Фредгольма 1-го рода (4.62).

Статистический метод идентификации устраняет недостатки, присущие детерминированному подходу; вместе с тем этот метод также имеет недостатки, определяемые статистическим содержанием эксперимента:

• время, необходимое для решения задачи идентификации, значительно увеличивается;

• имеют место ошибки построения $R_{uu}(\tau)$ и $R_{xu}(\tau)$ данным нормальной эксплуатации, определяемые величиной промежутка усреднения (известно, что ошибка определения корреляционной функции путем усреднения по времени убывает только как $\tau^{-0.5}$ и для достижения точности в несколько процентов требуется время накопления Т в 100 и более раз превышающее $\tau$ [15, 20]), а также наличием фактора не стационарности, низкой точностью измерений и др.

Пример 4.3 [15, 20]. Проиллюстрируем изложенные выше соображения на примере. Предположим, что

$$R_{uu}(\tau)=\begin{cases} e^{-\alpha\tau}-0{,}5 & \text{ïðè}\quad 0\le\tau\le T;\\ 0 & \text{ïðè}\quad \tau>T; \end{cases}$$

$$R_{xu}(\tau)=\begin{cases} 1{,}2102\sin 0{,}2618\tau+3{,}1213e^{-\alpha\tau}+2{,}3228e^{\alpha\tau}-3{,}8197 & \text{ïðè}\quad 0\le\tau\le T\\ 6{,}2022e^{-\alpha\tau}-0{,}36306\sin\dfrac{\pi}{12}(\tau-12)+0{,}3229\cos\dfrac{\pi}{12}(\tau-12)-1{,}9099 & \text{ïðè}\quad 0\le\tau\le 2T\\ 0 & \text{ïðè}\quad \tau>2T\\ \alpha=0{,}004257;\quad T=12. \end{cases}$$

Точное решение уравнения (4.62), полученное аналитически, в рассматриваемом случае имеет вид

$$k(t)=\begin{cases} sin(t\pi/2) & \text{ïðè}\quad 0\le\tau\le T;\\ 0 & \text{ïðè}\quad \tau>T. \end{cases}$$

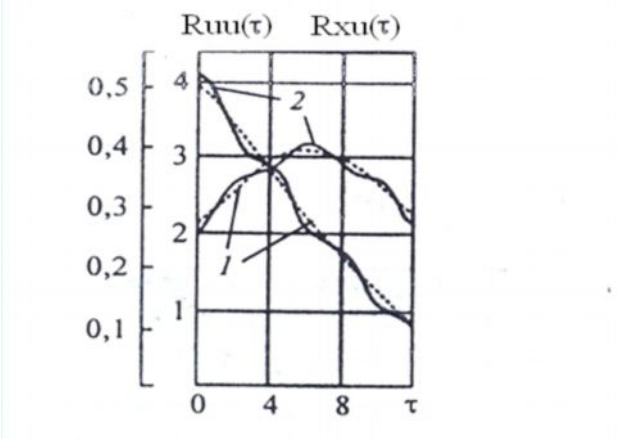



Рис. 4.14. Корреляционные и взаимокорреляционные функции

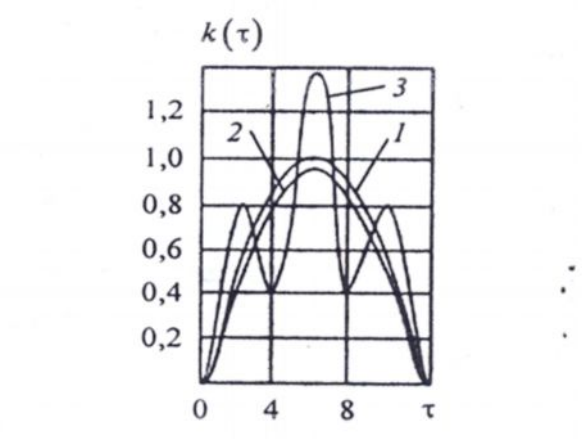



Рис. 4.15. Импульсные переходные функции при различных точностях выполнений

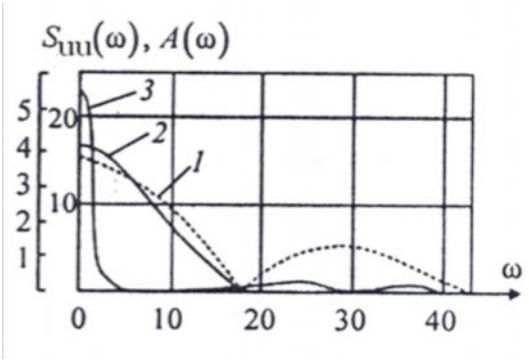



Рис. 4.16. Точные и приближенные значения амплитудной характеристики

На рис. 4.14 и 4.15 представлены графики функций $R_{uu}(\tau), R_{xu}(\tau), k(\tau)$ (кривые 1). Решая то же уравнение методом квадратур при $\Delta = 2, N = 7$, получим приближенное решение, довольно близкое к точному (рис. 4.16.). График этого решения изображен на рис. 4.15 (кривая 2). При незначительном изменении исходных данных, не превышающем 3 % от максимального значения, графики измененных функций $R_{uu}(\tau)$ и $R_{xu}(\tau)$ изображены на рис. 4.14 (кривые 2), решение же $k(\tau)$ (кривая 3 на рис. 4.15) значительно отличается от точного.

Пример 4.4 [28]. Рассмотрим задачу определения ИПФ системы, определяющей преобразование вибраций от двигателя легкового автомобиля к кузову при прямолинейном движении с постоянной скоростью (рис. 4.17).

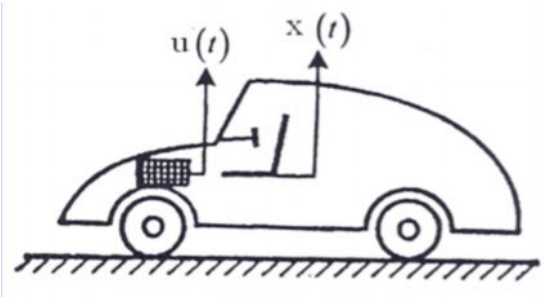



Рис. 4.17. Эксперимент по изучению нерегулярных вибраций легкового автомобиля (постоянная скорость 20 км/ч)

Считая, что входной сигнал $u$(t) и выходной процесс $x$(t) в установившемся режиме связаны соотношением $x(t) = \int_0^\infty k(\tau)u(t-\tau)d\tau$

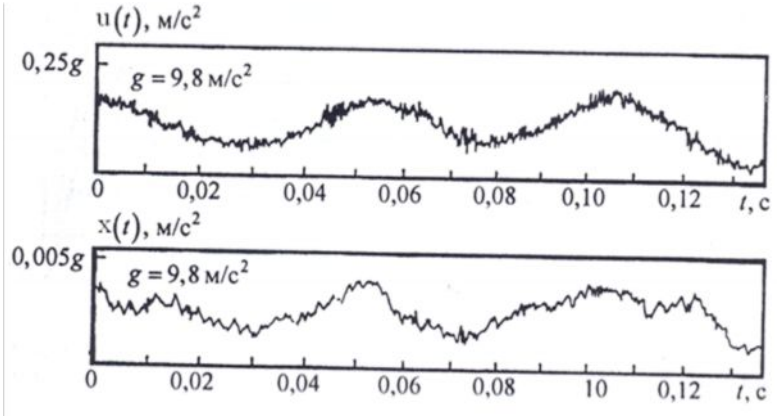



Рис. 4.18. Графики изменения ускорений вертикальной вибрации двигателя и кузова в процессе движения легкового автомобиля

определим импульсную переходную функцию $k(\tau)$ по реализациям $\{u(t)\}$ и $\{x(t)\}$, представленным на рис. 4.18. Поскольку, как видно из рис. 4.18, $\{u(t)\}$ не является белым шумом, то импульсная переходная функция определяется решением уравнения Винера-Хопфа.

Процедура получения импульсной переходной функции включает следующие этапы [28]. Оценивание $R_{uu}(i\Delta)$ и $R_{xu}(i\Delta)$ в этом эксперименте реализации $\{u(t)\}$ и $\{x(t)\}$ были дискретизованы по точкам выборки с периодом $\Delta = 0{,}002$ с; тогда $R_{uu}(i\Delta)$ и $R_{xu}(i\Delta)$ можно вычислить по формулам

$$R_{xu}(i\Delta) = \frac{1}{N}\sum_{i=0}^{N-1} u(l\Delta)u((l+i)\Delta), \qquad (4.94)$$

$$R_{xu}(i\Delta) = \frac{1}{N}\sum_{i=0}^{N-1} u(l\Delta)x((l+i)\Delta),$$

где N – число интервалов дискретизации, которое было принято равным 200. Как видно из формул (4.94), использовалась прямоугольная аппроксимация. Отметим, что хотя на практике такая аппроксимация, как правило, достаточно точна, но для систем высокого порядка она могла бы привести к ошибкам. На рис. 4.19 представлены графики функций $R_{uu}(i\Delta)$ и $R_{xu}(i\Delta)$.

2. После того как найдены $R_{uu}(i\Delta)$ и $R_{xu}(i\Delta)$ можно определить импульсную переходную функцию из уравнения (4.62). Полученная в этом эксперименте импульсная переходная функция показана на рис. 4.20.

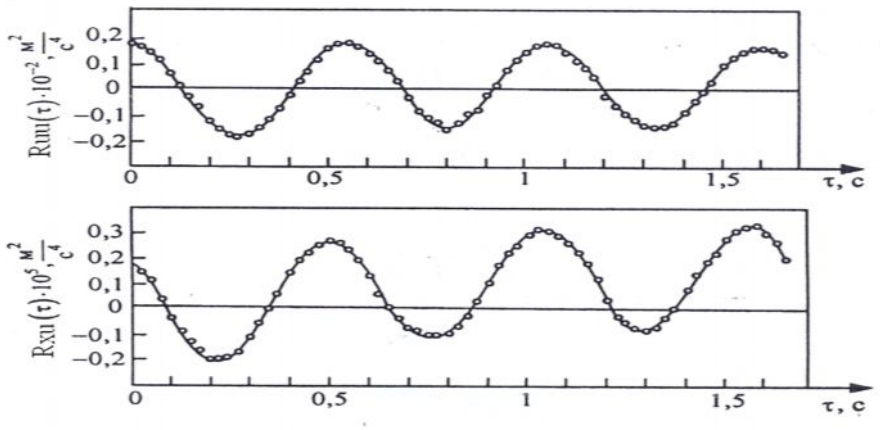



Рис. 4.19. Авто- и взаимная корреляционная функция сигналов $u(t)$ и $x(t)$

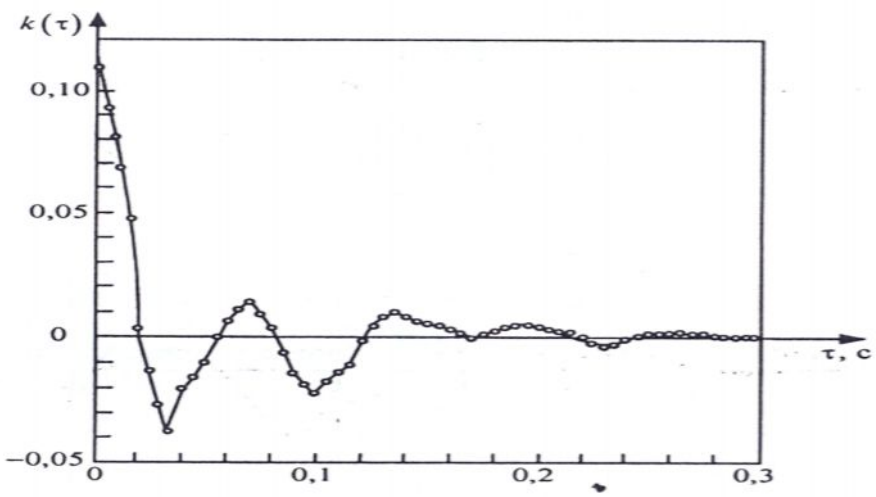



Рис. 4.20. График импульсной переходной функции системы «двигатель легкового автомобиля – кузов автомобиля» при прямолинейном движении с постоянной скоростью

Пример 4.5. Рассмотрим несколько другой подход к решению задачи идентификации, использующий аппроксимацию авто и взаимной корреляционных функций и позволяющий найти ИПФ в аналитическом виде.

Положим, что приближенное аналитическое уравнение для нормированной корреляционной функции входа *u(t)* имеет вид

$$r_{uu}(\tau) = e^{-0,008\tau} \cos\frac{\pi\tau}{200},$$

а средние квадратические отклонения входа $u(t)$ и выхода $x(t)$ равны $\sigma_x = 0.005$ и $\sigma_u = 0,33$.

На рис. 4.19 приведен график нормированной корреляционной функции входа $R_{uu}(\tau)$.

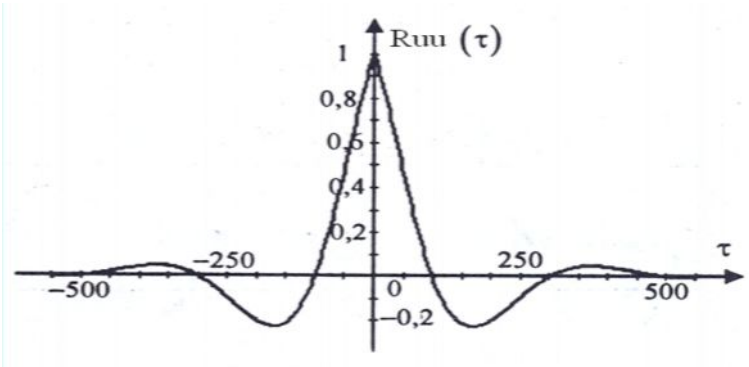



Рис. 4.21. График $Ruu(\tau)$

Взаимная корреляционная функция определяется зависимостью

$$r_{xu}(\tau)=\begin{cases}0{,}45e^{-0{,}0056\tau}\left(\cos\dfrac{\pi\tau}{200}+1{,}376\sin\dfrac{\pi\tau}{200}\right), & \tau\geq 0;\\ 0{,}45e^{-0{,}008\tau}\left(\cos\dfrac{\pi\tau}{200}+1{,}376\sin\dfrac{\pi\tau}{200}\right), & \tau\leq 0.\end{cases}$$

График взаимной корреляционной функции показан на рис. 4.22.

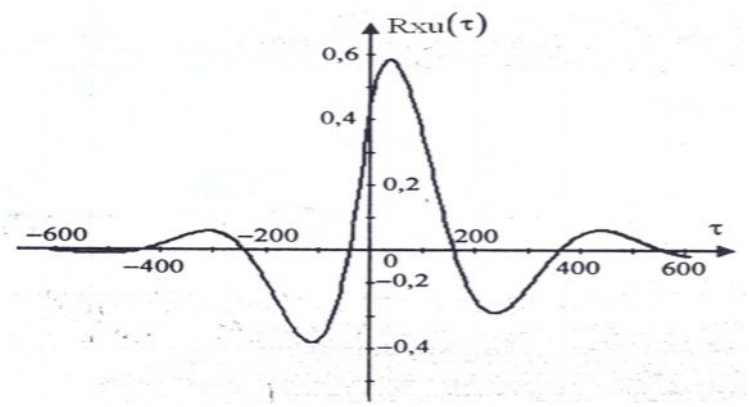



Рис. 4.22. График $Ruu(\tau)$

Из приведенных соотношений для нормированной корреляционной и взаимной корреляционной функции легко получить соответствующие корреляционные входа и взаимные корреляционные входа и выхода функции. Для рассматриваемого случая получим

$$R_{xu}(\tau)=\begin{cases}0{,}74\cdot 10^{-3}e^{-0{,}0056\tau}\left(\cos\dfrac{\pi\tau}{200}+1{,}376\sin\dfrac{\pi\tau}{200}\right), & \tau\geq 0;\\ 0{,}74\cdot 10^{-3}e^{-0{,}008\tau}\left(\cos\dfrac{\pi\tau}{200}+1{,}376\sin\dfrac{\pi\tau}{200}\right), & \tau\leq 0.\end{cases}$$

Теперь можно записать интегральное уравнение для определения ИПФ объекта:

$$0{,}74\cdot 10^{-3}e^{-0.0056\tau}\left(\cos\frac{\pi\tau}{200}+1{,}376\sin\frac{\pi\tau}{200}\right)=\int_0^{\infty}k(\lambda)e^{-0{,}008(\tau-\lambda)}\cos\frac{\pi(\tau-\lambda)}{200}d\lambda\,.$$

Выражение для ИПФ имеет вид

$$k(\tau)=8{,}5\cdot 10^{-3}\left[0{,}74\delta(\tau)+0{,}7e^{-0{,}0176\tau}-2\cdot 10^{-6}\cos 0{,}5\pi e^{-0.0056\tau}-1{,}3\sin 0{,}05\pi e^{-0{,}0056\tau}\right].$$

где $\delta$ – дельта-функция.

Применяя преобразование Лапласа, получим следующую зависимость для передаточной функции:

# ГЛАВА 4. НЕПАРАМЕТРИЧЕСКАЯ ИДЕНТИФИКАЦИЯ

Существует много способов определения динамических характеристик неизвестного линейного объекта или системы. Ниже в целях упрощения предполагается, что:

1) система имеет единственный вход и единственный выход;

2) система является линейной и стационарной;

3) аддитивная помеха, действующая на выходной сигнал системы, не зависит от входного сигнала и представляет собой стационарный случайный процесс.

При этих предположениях структура модели может быть представлена так, как и изображено на рис. 4.1. Тогда задача сводится к определению одной их следующих характеристик:

1) импульсного перехода или весовой функции *g(t);*

2) частотной характеристики *G(jω);*

3) передаточной функции *G(s).*

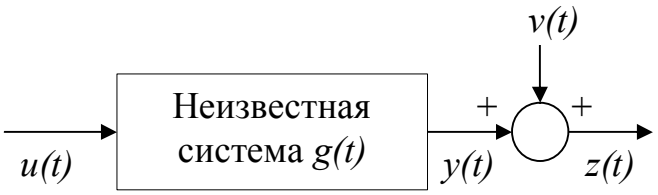



Рис. 4.1. Модель для идентификации линейной системы с одним входом и одним выходом

Следуя работам [20, 21, 23, 25, 26, 28], приведем основные соображения и теоретические положения, лежащие в основе наиболее представляемых методов.

## 4.1. Частотный анализ

А. Частотная характеристика. При отсутствии помех выходной сигнал системы *y(t),* показанной на рис 4.1, выражается через входной сигнал *u(t)* и импульсную переходную функцию *g(t)* с помощью интеграла свертки

$$y(t) = \int_0^\infty g(\tau)u(t-\tau)d\tau \ . \qquad (4.1)$$

Теперь предположим, что входной сигнал синусоидален; тогда *u(t)* можно представить в виде

$$u(t) = A\exp[(\sigma + j\omega)t] = A\exp(st), \tag{4.2}$$

где А – амплитуда, а $\omega$ – угловая частота, измеренная в рад/с. Подстановка выражения (4.2) в интеграл свертки (4.1) дает

$$y(t) = \int_0^\infty g(\tau)\,exp(-st) =$$

$$= A\,exp(st)\int_0^\infty g(\tau)\,exp(-st)\,d\tau = \tag{4.3}$$

$$= u(t)\int_0^\infty g(\tau)\,exp(-st)\,d\tau. \tag{4.4}$$

Теперь отношение преобразований Лапласа выходного и входного сигналов, которое называется передаточной функцией системы, может быть представлено как

$$G(s) = \frac{Y(s)}{U(s)} = \int_0^\infty g(\tau)\exp(st)\exp(-s\tau)\,d\tau. \tag{4.5}$$

Таким образом, этот интеграл не зависит от $t$ и определяет передаточную функцию системы $G(s)$ через ее импульсную переходную функцию $g(t)$. Если в формулу (4.5) в качестве аргумента преобразований Лапласа подставить комплексную переменную $j\omega$, то получится установившееся значение частотной характеристики системы

$$G(j\omega) = \frac{Y(j\omega)}{U(j\omega)} = \int_0^\infty g(\tau)\exp(-j\omega)\,d\tau\ . \tag{4.6}$$

Поскольку импульсная переходная функция системы $g(t)$ должна быть равна нулю при $t < 0$, то нижний предел интегрирования можно продлить до - $\infty$, так что

$$G(j\omega) = \int_0^\infty g(\tau)\,exp(-j\omega)\,dt. \tag{4.7}$$

Из этой формулы видно, что установившееся значение частотной характеристики системы тождественно преобразованию Фурье от импульсной переходной функции системы.

Рассматривая комплексную величину $G(j\omega)$ как вектор, получим представление $G(j\omega)$ в полярных координатах отношение

амплитуд (амплитудная характеристика) и фазовый сдвиг (фазовая характеристика).

$$\begin{aligned} |G(j\omega)| &= \frac{|Y(j\omega)|}{|U(j\omega)|} = S(\omega), \\ \angle G(j\omega) &= \angle \frac{|Y(j\omega)|}{|U(j\omega)|} = \varphi(\omega). \end{aligned} \tag{4.8}$$

Здесь передаточная функция системы определяется на основе частотного анализа.

*Б. Частотный анализ методом диаграмм Боде.* Передаточная функция стационарной линейной системы обычно записывается в виде отношения двух многочленов от *s*:

$$G(s) = \frac{bm^{sm} + \ldots + b_{1s} + b_0}{an^{sn} + \ldots + a_{1s} + a_0}. \tag{4.9}$$

Хорошо известно, что при отсутствии комплексных полюсов и нулей выражение (4.9) можно разложить на элементарные дроби вида

$$G(s) = \frac{k_1}{s + s_2} + \frac{k_2}{s + s_2} + \ldots + \frac{k_n}{s + s_n}, \tag{4.10}$$

где $s_1, s_2, \ldots s_n$ – полюсы, а $k_1, k_2 \ldots k_n$ – вычеты *G(s).* Такое представление иногда называют разложением Хевисайда, которое является чрезвычайно полезным при вычислении обратного преобразования Лапласа.

Вторую форму записи можно получить посредством факторизации числителя и знаменателя передаточной функции:

$$G(s) = \frac{K(s + c_1) + (s + c_1)\ldots(s + c_m)}{(s + s_1)(s + s_1)\ldots(s + s_n)}. \tag{4.11}$$

Здесь $c_1, c_2, \ldots c_m$ – нули *G(s),* а *K* – коэффициент усиления. Такое представление называют разложением Боде или полюсно-нулевым представлением.

Отметим, что если полюсы и нули функции *G(s),* определяемой формулой (4.9), не являются чисто вещественными, то приведенное разложение должно включать члены второго порядка. Так, например, при наличии в формуле (4.9) комплексно-сопряженных полюсов $s_j = \alpha + j\beta$ и $s_{j+1} = \alpha - j\beta$ числитель *G(s*), можно представить в виде

$$(s+s_j)(s+s_{j+1})\prod_{\substack{i=1\\ i\neq j,j+1}}^{n}(s+s_i)=$$

$$\left[(s-\alpha)^2+\beta^2\prod_{\substack{i=1\\ i\neq j,j+1}}^{n}(s+s_{j)}\right]= \tag{4.12}$$

$$\frac{1}{T^2}\left[(Ts)^2+2\xi Ts+1\right]\prod_{\substack{i=1\\ i\neq j,j+1}}^{n}(s+s_i),$$

где $T=1/\sqrt{\alpha^2+\beta^2}$ , $\xi=-a/\sqrt{2\alpha^2+\beta^2}$ – коэффициент затухания.

Если значения $c_i$ и $s_i$ не равны нулю, то можно преобразовать формулу (4.11) к виду

$$G(s)=\frac{K'(1+Q_1s)(1+Q_2s)...(1+Q_ms)}{(1+T_1s)(1+T_2s)...(1+T_ns)}, \tag{4.13}$$

где $$K'=K\frac{c_1c_2...c_m}{s_1s_2...s_n},\quad T_i=\frac{1}{s_i},\quad Q_i=\frac{1}{c_i}\ . \tag{4.14}$$

Тогда частотная характеристика получается путем подстановки в формуле (4.13) $j\omega$ вместо $s$

$$G(j\omega)=\frac{K'(1+jQ_2\omega)(1+jQ_2\omega)...(1+jQ_m\omega)}{(1+jT_1\omega)(1+jT_2\omega)...(1+jT_n\omega)}. \tag{4.15}$$

Частотную характеристику удобно определять с помощью диаграммы Боде, потому что:

1) амплитудную характеристику для $G(j\omega)$можно аппроксимировать кусочно-линейно ломаной;

2) поскольку масштаб измерения частоты $\omega$ логарифмический, то в диаграмму укладывается широкий диапазон значений $\omega$ ;

3) умножение амплитуды частотной характеристики $G(j\omega)$ на диаграмме Боде сводится к простому сложению. Преобразуя формулу (4.15), получим

$$\lg\left|G(j\omega)\right|=\lg\left|K'\right|+\sum_{i=1}^{m}\lg\left|1+jQ_i\omega\right|-\sum_{i=1}^{n}\lg\left|1+jT_i\omega\right|, \tag{4.16}$$

$$\varphi(\omega)=\sum_{i=1}^{m}arctg(Q_i\omega)-\sum_{i=1}^{n}arctg(T_i\omega)\,. \tag{4.17}$$

Эти соотношения называют уравнениями частотной характеристики. Уравнение (4.16) описывает зависимость амплитуды от частоты, а (4.17) – зависимость фазы от частоты. Используя формулу (4.15), частотную характеристику $G(j\omega)$ можно определить на основе частотных характеристик передаточных функций простейших звеньев $1/(1+T_i s)$ и $(1+Q_i s)$. Таким образом, важно знать, как ведут себя частотные характеристики звеньев, $1/(1+Ts)$ и $(1+Ts)$.

В соответствии с формулами (4.16) и (4.17) для $K/(1+Ts)$ находим

$$lg|G(j\omega)| = lg|K| - 0{,}5lg\left[1+(T\omega)^2\right], \quad \varphi(\omega) = -arctg(T\omega). \qquad (4.18)$$

Хорошо известны следующие асимптотические свойства диаграммы Боде для такого звена в левой (область низких частот) и правой (область высоких частот) частях диапазона:

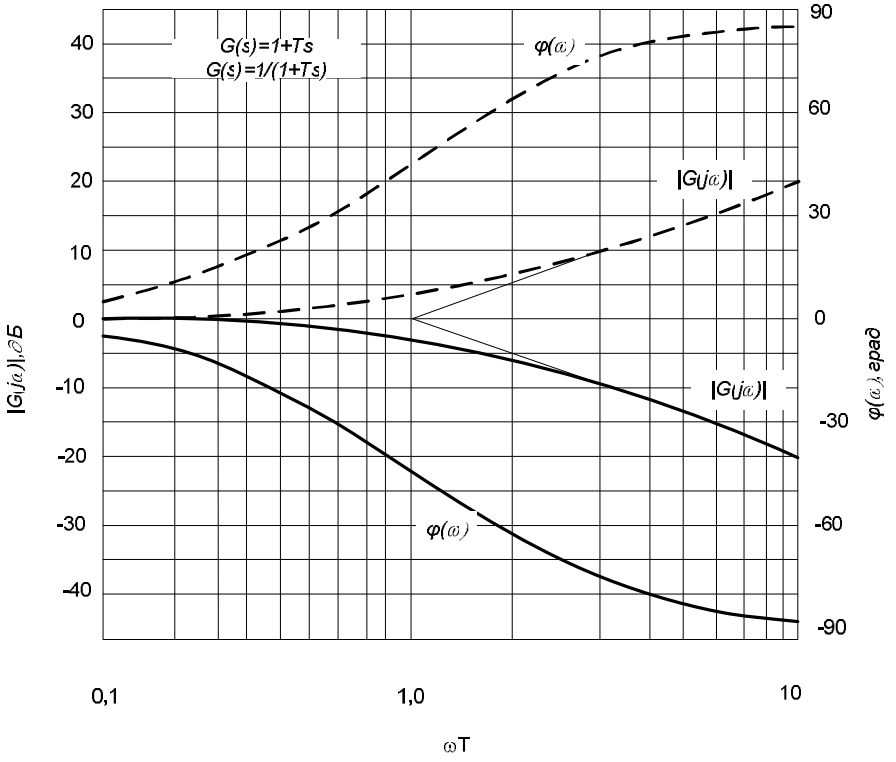



Рис. 4.2. Диаграммы Боде для звеньев $G(s)=1+Ts$ и $G(s)=1/(1+Ts)$

$$\begin{gathered} \omega T << \text{(левая часть диапазона)} \\ \lg|G(j\omega)| = \lg|K|,\ \varphi(\omega) = 0^0, \\ \omega T >> 1 \ \text{(правая часть диапазона)} \\ \lg|G(j\omega)| = \lg|K|/T\omega,\ \varphi(\omega) = 90^0 \\ \omega T = 1 : \lg|G(j\omega)| = \frac{\lg|K|}{\sqrt{2}},\ \varphi(\omega) = -45^0. \end{gathered} \qquad (4.19)$$

На рис. 4.2 приведена диаграмма Боде для звена *1/(1+Ts)*. Частота $\omega = 1/T$ называется частотой среза. Примерно так же выглядит график частотной характеристики звена *(1+Ts*), показанный на рис. 4.2 штриховой линией. В случае, когда полюсы или нули *G(s)* являются комплексно-значными величинами, рассматриваемое разложение *G(s)* включает сомножители второго порядка типа (4.12). Поэтому нужно знать поведение частотной характеристики и для передаточной функции звена второго порядка.

Рассмотрим частотную характеристику звена

$$x(t) = \sum_{i=0}^{\infty} \omega(i) u(t-i) \,. \tag{4.20}$$

Здесь предполагается, что $T > 0$ и *0<ξ<1*. Используя это соотношение в случаях, аналогичных рассмотренным в (4.19), получаем следующие асимптотические выражения:

$$\begin{gathered} \omega T << 1 : \lg\left|G(j\omega)\right| = 0, \;\; \varphi(\omega) = 0^0 \,, \\ \omega T >> 1 : \lg\left|G(j\omega)\right| = \lg[1/(\omega T)^2], \;\; \varphi(\omega) = -180^0, \\ \omega T = 1 : \lg\left|G(j\omega)\right| = \lg[1/2\xi\omega T], \;\; \varphi(\omega) = -90^0 . \end{gathered} \tag{4.21}$$

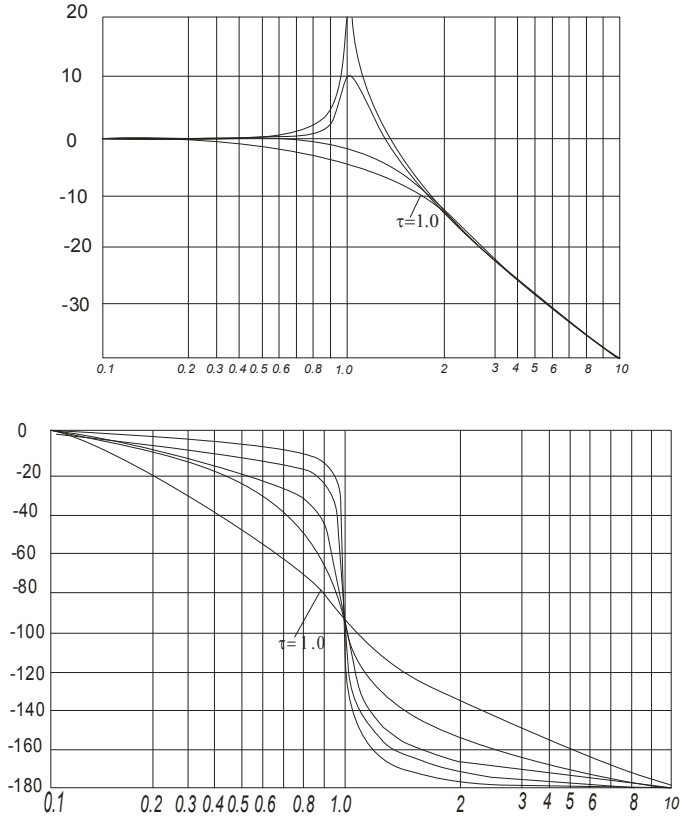



Рис. 4.3. Диаграмма Боде для звена *G(s)=1/[(Ts)²+2ξTs+1]*

На рис. 4.3 показана диаграмма Боде для различных значений ξ из формулы (4.20). Асимптотические характеристики в области низких частот совпадают с характеристиками звена первого порядка, однако в области высоких частот сдвиг по фазе вдвое больше. Кроме того, при малом коэффициенте затухания максимум резонанса достигается в окрестности частоты среза.

## 4.2. Анализ переходных процессов

Как было отмечено выше, с помощью интеграла свертки выходной сигнал системы выражается через ее импульсную переходную функцию. Таким образом, если импульсная переходная функция задана, то частотная характеристика может быть определена по формуле (4.6). Ниже описывается процедура построения частотной характеристики по заданной импульсной переходной функции.

*А. Аппроксимация частотной характеристики с помощью конечной суммы.* Если известна импульсная переходная функция системы, заданная, например, дискретным набором ординат, представляющих значения $g(t)$ в точках $t=i\theta,\ i=0,1,\ldots,\infty$, где $\theta$ – период выборки, то выражение (4.6) может быть приближенно представлено суммой вида

$$G(j\omega) \approx \theta \sum_{i=0}^{\infty} g(i\theta) \exp[-j\omega(i\theta)]. \qquad (4.22)$$

Используя формулу Эйлера, можно представить экспоненту в виде

$$\exp[-j\omega(i\theta)] = \cos(i\omega\theta) - j\sin(i\omega\theta). \qquad (4.23)$$

Подстановка выражения (4.22) в (4.23) дает

$$G(j\omega) \approx \theta \sum_{i=0}^{\infty} g(i\theta)\cos(i\omega\theta) - j\theta \sum_{i=0}^{\infty} g(i\theta)\sin(i\omega\theta). \qquad (4.24)$$

Если система асимптотически устойчива, так что $g(t) \to 0$ при $t \to 0$, то импульсная переходная функция может быть усечена при некоторым t=nθ. Тогда выражения (4.24) преобразуется к виду

$$G(j\omega) \approx \theta \sum_{i=0}^{\infty} g(i\theta)\cos(i\omega\theta) - j\theta \sum_{i=0}^{\infty} g(i\theta)\sin(i\omega\theta). \qquad (4.25)$$

Это соотношение определяет частотную характеристику.

При проведении расчетов по формуле (4.25) с использованием вычислительных машин необходимо принимать во внимание эффект наложения (aliasing). Период выборки $\theta$ нужно определить так, чтобы не пропустить резких изменений $g(t)$. Чтобы получилось хорошее

согласие с истинной частотной характеристикой, период выборки по теореме выборки должен удовлетворять следующему неравенству:

Период выборки ≤ Половина рассматриваемой частотной области. Отсюда следует, что если рассматриваемая область частот задается в виде максимальной частоты $\overline{\omega}$ (рад/с), то наибольший период выборки, при котором имеет место хорошее согласие с истиной характеристикой, равен [23, 28]

$$\theta = 2\pi/\overline{\omega}. \qquad (4.26)$$

Так как время установлено Т импульсной переходной функции известно, т. е. известна точка, в которой $g(i\theta)=0=g(T)$, то минимальное возможное значение $n$ равно

$$n=T/\theta. \qquad (4.27)$$

*Б. Определение частотной характеристики.* Передаточная функция линейной динамической системы может быть представлена в виде отношения двух многочленов от частоты. Допустим сначала, что форма идентифицируется передаточной функцией определяется выражением

$$\hat{G}(j\omega)=\frac{b_0+b_1(j\omega)+\ldots+b_m(j\omega)^m}{1+a_1(j\omega)+\ldots+a_n(j\omega)^n}=\frac{\hat{B}(j\omega)}{\hat{A}(j\omega)}. \qquad (4.28)$$

Для интерполяции экспериментальных данных функциями такого вида предложено несколько методов. Наиболее известный подход состоит в определении таких оценок неизвестных коэффициентов $a_i$ и $b_i$ передаточной функции, чтобы сумма квадратов отклонений оценок $G(j\omega)$ значений частотной характеристики от экспериментальных данных была минимальной.

Непараметрической моделью линейной стационарной динамической системы являются ее амплитудная и фазовая характеристики:

$$S=S(\omega),\ \psi=\psi(\omega),\ 0<\omega<\infty, \qquad (4.29)$$

т. е. зависимость амплитуды $S$ и фазы $\psi$ колебаний выхода объекта от частоты $\omega$ единичного гармоничного возмущения.

Связь между импульсной переходной функцией $\omega(t)$ и амплитудно-фазовыми характеристиками реализуется соотношениями:

$$S(\omega)\cos\psi(\omega)=\int_0^\infty \omega(t)\cos\omega t dt,$$

$$S(\omega)\sin\psi(\omega)=-\int_0^\infty \omega(t)\sin\omega t dt,$$

откуда получаем

$$S(\omega)\cos\psi(\omega)=\left[\left(\int_0^\infty \omega(t)\cos\omega t dt\right)^2+\left(\int_0^\infty \omega(t)\sin\omega t dt\right)^2\right]^{\frac{1}{2}},$$

$$\psi(\omega)=arctg\left[-\frac{\int_0^\infty \omega(t)\sin\omega t dt}{\int_0^\infty \omega(t)\cos\omega t dt}\right].$$

Импульсная переходная функция выражается через амплитудно-фазовые характеристики (4.29) следующей зависимостью:

$$\omega(t)=-\frac{2}{\pi}\int_0^\infty S(\omega)\cos\psi(\omega)\cos\omega t dt. \qquad (4.30)$$

Таким образом, определение амплитудно-фазовых свойств объекта дает возможность вычислить его импульсную переходную функцию. И наоборот, значение импульсной переходной функции дает возможность определить (4.29).

Связь моделей описание выше модели естественно связаны между собой. Так непрерывная параметрическая модель (4.28) связана с непараметрической (4.29) следующим образом [20, 23, 28]

$$\begin{aligned} S(\omega)&=\sqrt{\frac{a^2+b^2}{c^2+d^2}},\\ \psi(\omega)&=arctg\frac{bc-ad}{ac+bd}, \end{aligned} \qquad (4.31)$$

где

$$\begin{aligned} a&=b_0-b_2\omega^2+b_4\omega^4-...,\\ b&=b_1\omega-b_3\omega^3+b_s\omega^s-...,\\ c&=a_0-a_2\omega^2+a_4\omega^4-...,\\ d&=a_1\omega-a_3\omega^3+a_s\omega^s-..., \end{aligned} \qquad (4.32)$$

а коэффициенты $a_i$ *(i=0,...,p-1)* и $b_j$*(j=0,...,1)* соответствуют модели (4.28).

Таким образом, зная параметры непрерывной параметрической модели (4.28), легко определить характеристики непараметрической модели (4.29) и с помощью (4.30) импульсную переходную функцию *g(t)*.

Если же известна одна из амплитудно-фазовых характеристик непараметрической модели (4.29), то параметры непрерывной модели

$a_i$ ,$b_j$ можно оценить в результате решения одной из следующих задач минимализации:

$$I_S(C) = \int_0^\infty \left[ S(\omega) - \sqrt{\frac{a^2 + b^2}{c^2 + d^2}} \right]^2 d\omega \to \min_C, \tag{4.33}$$

$$I_\psi(C) = \int_0^\infty \left[ \psi(\omega) - arctg \frac{bc - ad}{ac + bd} \right]^2 d\omega \to \min_C, \tag{4.34}$$

где функции *a, b, c* и *d* определены в (4.35), а вектор искомых параметров с параметрической модели имеет вид (4.2).

Пример 4.1. Рассмотрим задачу восстановления модели (4.1) по наблюдениям амплитудно-частотной характеристики объекта *S(ω)* в случае, когда его структура представляет собой апериодическое звено. Модель, таким образом, имеет вид: $\dot{y} + a_1 y = b_0 x$.

Для идентификации параметров $a_1$ и $b_0$ модели воспользуемся первым выражениям из (4.29). Для апериодической структуры с учетом (4.31), т.е. при $a=b_0$, $b=0$, $c=a$, $d=w$, получаем следующее выражение для минимизируемой функции:

$$I_S(a_0, b_0) = \int_0^\infty \left[ S(\omega) - \frac{b_0}{\sqrt{a_0^2 + \omega^2}} \right]^2 d\omega. \tag{4.35}$$

Так как наблюдения дискретны, то естественно записать этот интеграл в виде суммы

$$I_S(a_1, b_0) = \sum_{i=0}^{10} \left[ S(\omega_i) - \frac{b_0}{\sqrt{a_1^2 + \omega_i^2}} \right]^2. \tag{4.36}$$

## 4.3. Определение передаточной функции по временным характеристикам объекта

Уравнения связи между входными и выходными переменными можно записать в различной форме. Наиболее универсальными из них являются дифференциальные уравнения объекта (2.21), (2.31).

Широко используется также уравнение связи между входом и выходом типа интеграла свертки (интеграл Дюамеля):

$$y(t) = \int_0^t x(\tau) w(t - \tau) d\tau = \int_0^t w(\tau) x(t - \tau) d\tau \ , \tag{4.37}$$

где $w(\tau)$ – функция веса объекта, т. е. реакция объекта на входной сигнал в виде дельта функции

$$\delta(t)=\begin{cases}0 & \text{ïðè } t\neq 0\\ \infty & \text{ïðè } t=0\end{cases};\qquad \int_{-\infty}^{\infty}\delta(t)dt=1. \tag{4. 38}$$

Дифференциальные уравнения (ДУ) и передаточная функция являются наиболее общими формами связи между переменными состояния на входе и выходе линейной системы. Но в реальных условиях часто наблюдаются только сигналы управления и реакции систем, по которым необходимо получить уравнение связи.

Таким образом, с помощью эксперимента можно получить график, определяющий частное решение при определенном входном сигнале. Затем, аппроксимировав аналитическим выражением полученные реализации, можно построить ДУ заданной структуры и записать его в одной из форм, приведенных выше (т. е. получить модель объекта).

Большое распространение получили методы идентификации детерминированных объектов путем определения переходной характеристики $h(t)$ по кривой разгона при ступенчатом изменении управления на входе

$$h(t)\approx\frac{y(t)-y_0}{u-u_0}, \tag{4.39}$$

где $y(t)$ – изменение выходной величины объекта при подаче на его вход ступенчатого управления $u$ (кривая разгона), $y_0$ и $u_0$ установившиеся значения выхода и входа объекта до начала проведения эксперимента.

Если объект управления не допускает изменения выходной координаты, то на его вход вместо ступенчатого воздействия подают единичный импульс или серию импульсов.

При проведении эксперимента по снятию кривой разгона необходимо тщательно изолировать объект от случайных возмущений, стремиться, как можно точнее воспроизвести заданную форму возмущения на входе, дублировать проведение экспериментов по снятию кривой разгона для различных начальных значений $y_0$ и $u_0$.

Рассмотрим определение передаточной функции объекта по кривой разгона логарифмическим методом. Преимущество этого метода состоит в том, что результаты идентификации получаются в виде аналитического выражения, хорошо поддающегося дальнейшей машинной обработке. Методика идентификации в этом случае [26]

заключается в аппроксимации переходной характеристики аналитическим выражением типа

$$h(t) = k_0 + \sum_{k=1}^{n} C_k e^{p_k t}, \qquad (4.40)$$

где $k_0 = \dfrac{y(\infty) - y_0}{u - u_0}$ – коэффициент передачи объекта $y(\infty)$ – установившееся значение выходной величины объекта, соответствующее частному решению его ДУ и определяемое вынужденным движением под действием входного сигнала, остальные слагаемые определяют свободные движения и представляют общее решение однородного ДУ объекта, $C_к$ – постоянные интегрирования, $p_к$ – корни характеристического уравнения.

Положим, для определенности, что в (4.39) имеется один действительный корень, два комплексно сопряженных и два кратных корня. Логарифмируя (4.39), получим

$$z = ln\left[h(t) - k_0\right] = ln\left[\sum_{k=1}^{5} C_k e^{p_k t}\right]. \qquad (4.41)$$

Для устойчивого объекта свободные движения с течением времени стремятся к нулю, причем время переходного процесса будет определяться корнем, имеющим минимальную действительную часть (например, действительным корнем). Тогда начиная с некоторого момента времени слагаемыми в (4.41), имеющими большие действительные части можно пренебречь и приближенно записать

$$z = \ln\left[h(t) - k_0\right] \approx \ln C_1 + p_1 t\ . \qquad (4.42)$$

Уравнение (4.42) является асимптотой (4.41) при $t \to \infty$.

Если теперь в полулогарифмическом масштабе построить график уравнения (4.41), откладывая на оси абсцисс время на оси ординат $z=ln[h(t)-k_o]$, то неизвестные коэффициенты $C_1$, $p_1$ легко определить графически, как показано на рис 4.4.

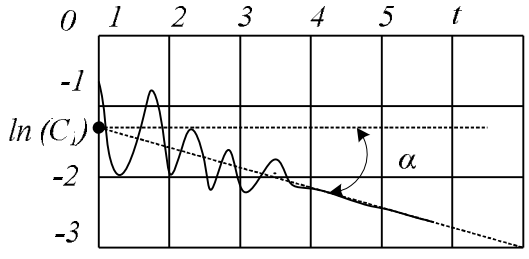



Рис. 4.4. График определения неизвестных параметров $C_1$,$p_1$

Ордината асимптоты при $t=0$ равна $ln((C_1)$, а $p_1= tan(\alpha)$. После чего из (4.4) можно исключить одно слагаемое соответствующее действительному корню, тогда

$$z_1 = \ln\left[h(t) - k_0 - C_1 e^{p_1 t}\right] = \ln\left[\sum_{k=1}^{4} C_k e^{p_k t}\right]. \qquad (4.43)$$

Аналогично, начиная с некоторого момента времени слагаемыми в (4.43), имеющими большие действительные части можно пренебречь и приближенно записать

$$h(t) - k_0 - C_1 e^{p_1 t} \approx C_2 e^{-\lambda t} \sin(\omega t + \varphi). \qquad (4.44)$$

Логарифмируя, получим

$$z_1 = \ln\left[h(t) - k_0 - C_1 e^{p_1 t}\right] \approx \ln(C_2) - \lambda t + \ln\left[\left|\sin(\omega t + \varphi)\right|\right]. \qquad (4.45)$$

Уравнение прямой на графике (рис. 4.5), проходящей через точки в которых $sin\ (\omega t + \varphi) = 1$ имеет вид:

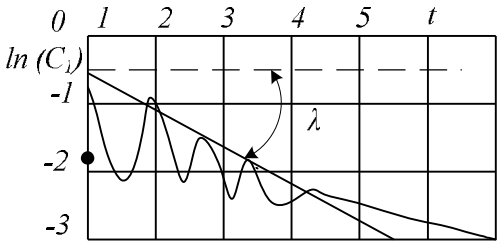



Рис. 4.5. График определения неизвестных параметров $C_2$, λ

Неизвестные параметры $C_2$ и λ находятся так же, как и в предыдущем случае.

Круговую частоту ω и начальную фазу φ можно определить из очевидного условия $sin(\omega t + \varphi) = \pm 1$

$$\frac{\pi}{2} + (i-1)\pi = \omega t_i + \varphi, \qquad (4.46)$$

где $t_i$ – значения времени при которых ординаты переходной характеристики принимают экстремальные значения.

Минимизируя квадрат невязки (4.46) по $\omega$ и $\varphi$, получим систему алгебраических уравнений для их определения

$$\sum_{i=1}^{m} t_i^2 f + \sum_{i=1}^{m} t_i b = \frac{1}{4}\sum_{i=1}^{m}(2i-1)t,$$

$$\sum_{i=1}^{m} t_i f + mb = \frac{1}{4}\sum_{i=1}^{m}(2i-1), \qquad (4.47)$$

где $f = \frac{\omega}{2\pi}$ – неизвестная частота, $b = \frac{\varphi}{2\pi}$ – относительная фаза, $m$ – число экстремальных значений переходной характеристики.

Коэффициенты последних слагаемых от корней, имеющих кратность – два, в уравнении переходной характеристики (4.40) получают из выражения аналогичного (4.43)

$$z_2 = \ln\left[h(t) - k_0 - C_1 e^{p_1 t} - C_2 e^{-\lambda t}\sin(\omega t + \varphi)\right] = \ln\left[\sum_{k=1}^{2} C_k e^{p_k t}\right]. \qquad (4.48)$$

Приближенно при больших $t$ можно записать

$$z_2 \approx \ln(C_3) + \ln(t) + pt\,, \qquad (4.49)$$

где $p=p_4=p_5$ – корень двойной кратности.

Асимптотой этого уравнения будет прямая, тангенс угла наклона которой будет равен $p$.

Если теперь простроить в логарифмических координатах функцию $z_2 - pt$, то ее асимптотой будет $\ln(t) + \ln(C_3)$. Тангенс угла наклона асимптоты дает кратность корня. Ордината асимптоты при $t=0$ равна $\ln(C_3)$.

Зная корни выражение для переходной характеристики нетрудно определить и передаточную функцию через преобразование Лапласа

$$W(p) = pL\left[h(t)\right] = p\int_0^\infty h(t)e^{-pt}dt\,. \qquad (4.50)$$

Удовлетворительные результаты при использовании этого метода получаются в том случае, если корни характеристического уравнения далеко отстоят друг от друга. Желательно, чтобы каждый следующий корень был в 2–3 раза меньше предыдущего. Следует отметить, что точность определения характеристик объекта с помощью временных характеристик невысока. Качество идентификации снижают случайные помехи, искажающие реакцию объекта и неточность аппроксимации объекта линейной моделью. Все это не позволяет считать эти методы перспективными.

Пример 4.2. Проведем идентификацию объекта с передаточной функцией

$$W(p) = \frac{2{,}5}{(p + 0{,}1)(p^2 + 6p + 25)}\,.$$

Корни характеристического уравнения объекта будут равны $p_1$ = - 0,1; $p_2$ = -3 + $j$ 4; $p_3$ = -3- $j$ 4, а его коэффициент передачи $\kappa_o$=1.

Зная передаточную функцию объекта нетрудно построить его переходную характеристику $h(t)$ (рис. 4.6), используя преобразование Лапласа

$$h(t) = L\left[\frac{W(p)}{p}\right].$$

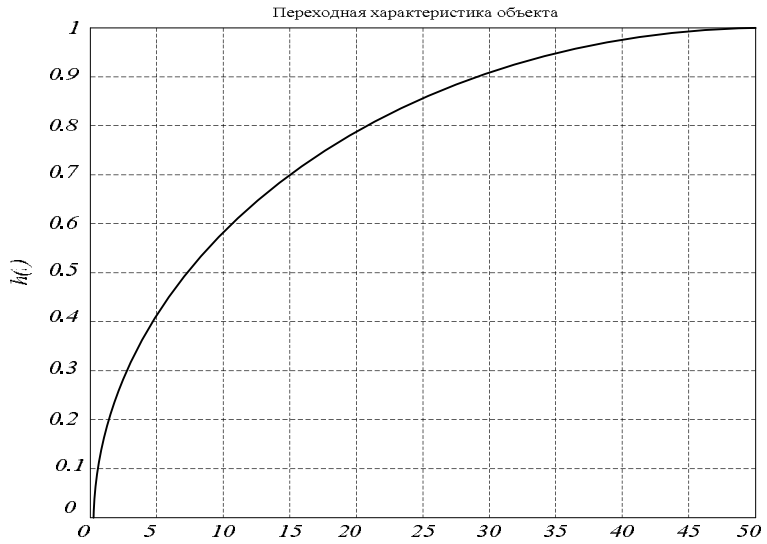



Рис. 4.6. Переходная характеристика $h(t)$

По переходной характеристике определяем коэффициент передачи $k_o = h(\infty) = 1$, и строим по формуле (4.41) график относительной переходной характеристики в полулогарифмическом масштабе (рис. 4.7).




Рис. 4.7. График относительной переходной характеристики

По графику находим значение коэффициента $C_1$= 1,0242 и оценку первого корня характеристического уравнения $p_1$= - 0,1, которая совпадает с его истинным значением.

Исключаем из переходной характеристики слагаемое, соответствующее действительному корню и строим по формуле (4.45), новую переходную характеристику (рис. 4.8).

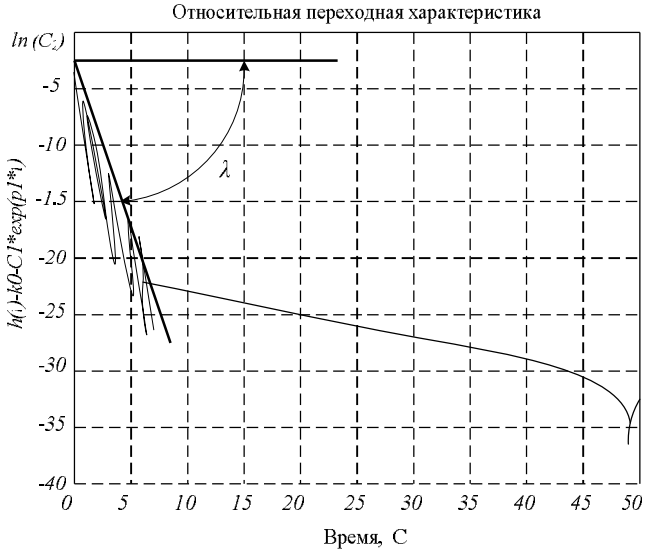



Рис. 4.8. Переходная характеристика по формуле (4.45)

Из графических построений и по формулам (4.48) находим оценки параметров объекта $\hat{p}_2 = -2{,}9959 + j3{,}0084$; $\hat{p}_3 = -2{,}9959 - j3{,}0084$, хорошо совпадающие с истинными значениями.

## 4.4. Определение передаточной функции объекта по частотным характеристикам

Этот метод идентификации относится к непараметрическим методам, т. к. вначале экспериментально снимают частотные характеристики объекта, а затем уже по полученным экспериментальным характеристикам вычисляют передаточную функцию. Основные трудности при проведении экспериментов заключаются в определении рабочего диапазона частот и дрейфе оси колебаний на выходе объекта. Чаще всего область рабочих частот задается ориентировочно, и наибольшее внимание уделяется диапазону, в котором сдвиг по фазе между входным и выходным гармоническими сигналами составляет $180^0$.

При снятии частотных характеристик используют различные методы воздействия на объект.

Метод синусоидальной волны предполагает подачу на вход объекта гармонических колебаний. На каждой из выбранных в пределах рабочего диапазона частот проводится отдельный опыт. На входе исследуемого объекта возбуждаются колебания выбранной частоты. Процессу колебаний дают установиться и, когда ось колебаний, их форма и амплитуда станут неизменными, измеряют амплитуды входных и выходных колебаний и фазовый сдвиг между ними. Частное от деления амплитуды выходных колебаний на амплитуду входных колебаний дает амплитуду частотной характеристики на взятой частоте, сдвиг по фазе – ординату фазовой частотной характеристики.

Основным затруднением при использовании этого метода является необходимость возбуждения колебаний большой мощности, имеющих правильную синусоидальную форму.

Поэтому чаще применяют метод «прямоугольной» волны. При этом по прибору, измеряющему входную величину объекта, градуируются три положения регулирующего органа, изменяющего эту величину. Среднее положение соответствует значению входной величины, при котором стабилизируется режим работы объекта перед началом испытаний (ось колебаний); два других равноотстоят от среднего положения. Перед началом опыта регулирующий орган устанавливается в среднее положение и удерживается там до установления в объекте стабильного режима. Затем регулирующее устройство через равные промежутки времени, соответствующие полупериоду избранной частоты, переводится из одного крайнего положения в другое и обратно. Эти переключения продолжаются до тех пор, пока на выходе объекта колебания *y(t)* не примут установившиеся формы. После чего производят регистрацию этих колебаний.

На основе полученных осциллограмм колебаний на выходе объекта проводят их гармонический анализ, ограничиваясь вычислением амплитуд и фаз первой и третьей гармоник

$$b_k = \frac{1}{\pi}\int_{-\pi}^{\pi} y(t)\cos\left(k\frac{2\pi}{T}t\right)dt,$$

$$c_k = \frac{1}{\pi}\int_{-\pi}^{\pi} y(t)\sin\left(k\frac{2\pi}{T}t\right)dt, \qquad (4.51)$$

$$a_k = \sqrt{b_k^2 + c_k^2},$$

$$\varphi_k = arctg\left(\frac{b_k}{c_k}\right),$$

где $T$ – период колебаний; $k$ – номер гармоники.

В том случае если значения выходной величины известны только в дискретные, равноотстоящие моменты времени интегралы в (4.51) заменяются суммами

$$b_k = \frac{1}{N}\sum_{I=0}^{N-1} y_i \cos\left(2\pi\frac{ik}{N}\right),$$

$$c_k = \frac{1}{N}\sum_{I=0}^{N-1} y_i \sin\left(2\pi\frac{ik}{N}\right), \qquad (4.52)$$

где $N$ – число дискрет выходного сигнала.

Гармонический анализ входного прямоугольного сигнала приводит к такому выражению

$$u(t) = \frac{4A}{\pi}\left[\sin\left(\frac{2\pi}{T}t\right) + \frac{1}{3}\sin\left(3\frac{2\pi}{T}t\right) + \frac{1}{5}\sin\left(5\frac{2\pi}{t}t\right) + ...\right], \qquad (4.53)$$

где $A$ – амплитуда прямоугольной волны.

Вычисляют амплитуды и фазы входных и выходных гармонических характеристик на выбранной частоте, как отношение амплитуд гармонических составляющих на выходе и входе объекта и значения фазочастотной характеристики, как соответствующий фазовый сдвиг $\varphi_k$.

Для повышения точности определения частотных характеристик рекомендуется использовать при гармоническом анализе только первую гармонику.

Для того чтобы не допускать больших отклонений регулируемых составляющих, можно вычислить значения амплитудно-частотной величины в процессе проведения опытов переключения регулирующего органа, который осуществляют в моменты времени, когда регулируемая (выходная) величина достигает заранее выбранные граничные значения. Для реализации такого режима используют двухпозиционный регулятор с регулируемой зоной нечувствительности. Изменяя величину зоны нечувствительности и знак обратной связи можно изменять частоту возникающих в такой системе автоколебаний.

Эксперимент обычно начинают на частоте, при которой фазовый сдвиг между входным и выходным сигналами равен $\pi$. Это дос-

тигается при нулевой зоне нечувствительности позиционного регулятора и отрицательной обратной связи. Увеличивая зону нечувствительности при отрицательной обратной связи, уменьшают частоту автоколебаний. Увеличивая зону нечувствительности при положительной обратной связи, увеличивают частоту автоколебаний.

Результаты экспериментов обрабатываются так же, как и в методе прямоугольной волны. Основным недостатком рассмотренных методов является длительное время эксперимента, затрачиваемое в основном на ожидание установившегося режима колебаний и получения достаточного для аппроксимации частотных характеристик значений их ординат. Для ускорения экспериментов иногда на вход объекта подается сумма гармонических составляющих разных частот.

Установившиеся колебания выходной величины также подвергают гармоническому анализу и сразу находят несколько ординат частотной характеристики. Однако в этом случае требуется специальный источник полигармонического воздействия и линейность объекта.

Определение аналитического выражения передаточной функции по частотным характеристикам может быть осуществлено несколькими способами. Рассмотрим один из них [23, 26] позволяющий аналитически вычислить выражение для передаточной функции по дискретным ординатам вещественной частотной характеристики.

Будем искать выражение для передаточной функции в виде ряда

$$W(p) = \sum_{k=0}^{\infty} A_k \left( \frac{1-p}{1+p} \right)^k . \qquad (4.54)$$

Тогда делая подстановку

$$tg\left( \frac{\varphi}{2} \right) = \omega \qquad (4.55)$$

выражение для вещественной частотной характеристики можно записать в виде

$$U(\varphi) = \sum_{k=0}^{\infty} A_k \cos(k\varphi) . \qquad (4.56)$$

Если теперь провести гармонический анализ экспериментально полученной вещественной частотной характеристики, то в результате будут получены неизвестные коэффициенты передаточной функции *Ак.*

Ограничиваясь значащими членами ряда (4.56) можно записать выражение передаточной функции объекта в виде (4.54).

Ограничение на использование данного метода вычисления передаточной функции относятся к самовыравнивающимся, устойчивым и неминимально-фазовым объектам. В том случае если объект не отвечает этим требованиям, то при вычислении вещественной частотной характеристики необходимо учесть влияние интегрирующих звеньев и звеньев запаздывания на вид этой характеристики, и проводить ее расчет с учетом этого влияния.

Компенсация запаздывания осуществляется выражением

$$U_ç(\omega) = U(\omega)\cos\omega\tau - V(\omega)\sin\omega\tau, \qquad (4.57)$$

а компенсация интегрирующих звеньев –

$$U_u(\omega) = \frac{V(\omega)}{\omega}, \qquad (4.58)$$

где $U(\omega)$ и $V(\omega)$ – экспериментальные вещественная и мнимая частотные характеристики.

## 4.5. Статистическая идентификация линейных устойчивых стационарных объектов (метод корреляционных функций) [20, 38]

### 4.5.1. Интегральное уравнение Фредгольма 1-го рода, определяющее импульсную переходную функцию идентифицируемого объекта.

Рассмотренная в предыдущем параграфе детерминированная идентификация линейных стационарных устойчивых объектов имеет недостатки, два из которых оказывают существенное влияние на точность решения задачи:

- наличие источника ошибок идентификации, порожденных неконтролируемыми шумами *n(t),* которые приводят к тому, что выходной процесс определяется формулой

$$x(t) = x_y(t) + x_n(t);$$

- наличие источника ошибок за счет $x^0 = (x(0), x'(0), \ldots, x^{(n-1)}(0)) \neq 0$.

Статистическая идентификация позволяет устранить влияние указанных факторов. Положим, что на вход стационарного линейного

объекта, имеющего установившийся режим работы, поступает случайный стационарный процесс *u(t),* тогда

$$x(t) = \int k(\lambda)u(t-\lambda)d\lambda\,, \tag{4.59}$$

$$M\left[u(t-\tau)x(t)\right] = R_{xu}(\lambda) = M\left[u(t-\tau)\int_0^\infty k(\lambda)u(t-\lambda)d\lambda\right] = M\left[u(t-\lambda)\int_0^\infty k(\lambda)u(t-\lambda)d\lambda\right] =$$

$$= \int_0^\infty k(\lambda)M\left[x(t-\tau)x(t-\lambda)\right]d\lambda = \int_0^\infty k(\lambda)M\left[R_{uu}(t-\lambda-t+\tau)\right]d\lambda = \int_0^\infty R_{uu}(\tau-\lambda)k(\lambda)d\lambda = R_{xu}(\lambda).$$

Имеем, в основе статистических методов определения импульсных переходных функций лежит интегральное уравнение

$$R_{xu}(\lambda) = \int_0^\infty R_{uu}(\tau-\lambda)k(\lambda)d\lambda, \quad -\infty < \lambda > \infty, \tag{4.60}$$

позволяющее по корреляционной функции сигнала *u(t)* на входе $R_{uu}(\tau)$ исследуемого объекта и взаимной корреляционной функции $R_{xu}(\tau)$ между выходом *x(t)* и входам *u(t)* найти импульсную переходную функцию $k(\lambda)$.

Предположим, что к объекту приложено не только воздействие *u(t),* но и помеха *n(t).* Тогда справедлива зависимость

$$x(t) = \int_0^\infty u(t-\tau)k(\lambda)d\tau + \int_0^\infty n(t-\tau)k_n(\tau)d\tau\,. \tag{4.61}$$

Определить $k(\tau)$ из последней формулы методами, описанными в предыдущем параграфе, достаточно трудно, т. к. второй член в правой части является источником погрешности.

Умножим обе части формулы (4.61) на $u(t-\lambda)$ и произведем усреднение по времени

$$M\left[x(t)u(t+\lambda)\right] = \int_0^\infty M\left[u(t-\tau)u(t-\lambda)\right]k(\tau)d\tau + \int_0^\infty M\left[n(t-\tau)u(t+\lambda)\right]k_n(\tau)dt_\theta\,.$$

В результате получаем уравнение (4.60), которое остается справедливым при наличии помехи *n(t),* воздействующей на объекту при условии, что она не коррелирована с сигналом *u(t).*

Статистические методы определения импульсной переходной функции подразделяются на требующие и не требующие искусственных воздействий. Первая группа методов основана на том, что к реальному входному сигналу добавляется искусственный вспомогательный сигнал *m(t)* с известной корреляционной функцией. Обычно

сигналом *m(t)* служит белый шум. При отсутствии корреляции между реальным входным сигналом и *m(t)* имеем

$$R_{xm}(\tau) = \int_0^\infty R_{mm}(\tau - \lambda)k(\lambda)d\lambda . \qquad (4.62)$$

Если *m(t)* – белый шум, то $R_{mm}(\tau)$ является дельта-функцией, т. е.

$$R_{mm}(\tau) = 2\pi S_0 \delta(\tau)$$

и выражение (4.62) сводится к виду

$$R_{xm}(\tau) = 2\pi S_0 k(\tau) .$$

Из последней зависимости следует, что определение импульсной переходной функции не требует решения интегрального уравнения. Однако в ряде случаев введение искусственного шума в систему нежелательно и тогда необходимо пользоваться уравнением (4.62).

Отметим, что нахождение импульсной переходной функции из интегрального уравнения (4.62) эквивалентно ее оценке по критерию минимума среднеквадратической ошибки [15, 20, 38].

Действительно, пусть

$$\varepsilon(t) = x(t) - \int_0^\infty k(\tau)u(t-\tau)d\tau , \qquad (4.63)$$

$$J = \frac{1}{T}\int_0^T \varepsilon^2(t)dt . \qquad (4.64)$$

Если функция $k(\tau)$ минимизирует критерий (4.64), то справедливо следующее представление: $k(\tau) = \tilde{k}(\tau) + \alpha k_\alpha(\tau)$, где $k_\alpha(\tau)$ – произвольная функция; $k_\alpha(\tau) = 0$ при $\tau < 0$.

В соответствии с формулой (4.64) можно записать известное выражение $\frac{dJ}{d\alpha}|_{\alpha=0} = 0$.

Поскольку

$$J = \frac{1}{T}\int_0^T \left\{ x(t) - \int_0^\infty \left[\tilde{k}(\tau) + \alpha k_\alpha(\tau)\right]u(t-\tau)d\tau \right\}^2 dt,$$

то

$$\frac{dJ}{d\alpha}|_{\alpha=0} = \frac{2}{T}\int_0^T \left\{ x(t) - \int_0^\infty \tilde{k}(\tau)u(t-\tau)d\tau \right\} \int_0^T \int_0^\infty k_\alpha(\tau)u(t-\tau)d\tau dt = 0 .$$

Отсюда следует

$$\frac{1}{T}\int_0^T\left\{\left[x(t)-\int_0^\infty \widetilde{k}(\tau)u(t-\tau)d\tau\right]\int_0^\infty k_\alpha(\tau)u(t-\tau)d\tau\right\}dt=0$$

или, что то же самое

$$\frac{1}{T}\int_0^T x(t)\int_0^\infty k_\alpha(\tau)u(t-\tau)d\tau dt-\frac{1}{T}\int_0^T\int_0^\infty \widetilde{k}(\tau_1)u(t-\tau_1)d\tau_1\int_0^\infty k_\alpha(\tau_2)u(t-\tau_2)d\tau_2 dt=0$$

получаем

$$\int_0^\infty k_\alpha(\tau_2)\left\{\frac{1}{T}\int_0^T x(t)u(t-\tau)dt-\int_0^\infty \widetilde{k}(\tau_1)d\tau_1\frac{1}{T}\int_0^T u(t-\tau_1)u(t-\tau_2)dt\right\}d\tau_2=0\,.$$

Далее находим

$$\frac{1}{T}\int_0^T x(t)u(t-\tau)dt=\int_0^\infty \widetilde{k}(\lambda)d\lambda\frac{1}{T}\int_0^T u(t-\tau)u(t-\tau)dt$$

или

$$\widetilde{R}_{xu}(\tau,T)=\int_0^\infty R_{uu}(\tau-\lambda,T)\widetilde{k}(\lambda)d\lambda\,.$$

Условие эргодичности процессов при $T\to\infty$ позволяет записать интегральное уравнение (4.62), называемое уравнением Винера-Хопфа**.** Оно принадлежит к классу уравнений Фредгольма 1-го рода.

Существуют формулы, позволяющие выбрать интервал наблюдений *T,* Например, если сигнал *x(t)* имеет автокорреляционную функцию вида $R_{xx}(\tau)=D_{xx}e^{\alpha|\tau|}\cos\omega_0\tau$, то $T\geq 20(\frac{\alpha}{\alpha^2+\omega_0^2}+\frac{1}{\alpha})$, если же $R_{xx}(\tau)=e^{\alpha|\tau|}$, то $T\geq 40/\alpha$.

Получены формулы, при помощи которых можно, задаваясь корреляционной функцией, полученной на основе опыта статистической обработки аналогичных процессов, и, исходя из заданной среднеквадратичной ошибки, выбрать значение *T*.

В соответствии с уравнением (4.62) для вычисления *k(τ)* необходимо выполнить два основных этапа (рис. 4.9):

1) определение корреляционных функций $R_{uu}(\tau)$ и $R_{xu}(\tau)$*;*

2) решение интегрального уравнения Винера-Хопфа.

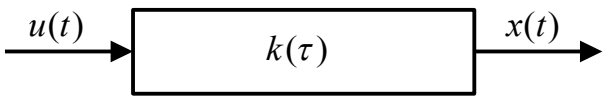



Рис. 4.9. К выводу интегрального уравнения Винера-Хопфа

Справедливы следующие положения [15, 22, 23].

Пусть $k(\tau)$ – импульсная переходная функция линейной стационарной системы, а $R_{хи}(\tau)$ и $R_{ии}(\tau)$ – взаимные корреляционные функции входного и выходного сигналов и автокорреляционная функция входного сигнала соответственно. Тогда $k(\tau)$ удовлетворяет следующему уравнению свертки:

$$R_{хи}(\tau)=\int_0^{\infty} k(\tau)R_{ии}(\tau-\lambda)d\lambda,\ \tau\in(-\infty,+\infty).$$

Тогда $k(\tau)$ удовлетворяет уравнению Винера-Хопфа

$$R_{хи}(\tau)=\int_0^{\infty} k(\tau)R_{ии}(\tau-\lambda)d\lambda,\ \tau\in(0,+\infty).$$

Пусть

$$R_{ии}(\tau)=\begin{cases}R_{ии}^{+}(\tau),\tau\ge 0;\\ R_{ии}^{-}(\tau),\tau<0;\end{cases}\qquad R_{хи}(\tau)=\begin{cases}R_{хи}(\tau),\tau\ge 0;\\ R_{хи}(\tau),\tau<0;\end{cases}$$

тогда ИПФ удовлетворяет уравнениям

$$R_{хи}^{+}(\tau)-R_{хи}^{-}(\tau)=\int_0^{\infty} k(\lambda)\left[R_{ии}^{+}(\tau-\lambda)-R_{ии}^{-}(\tau-\lambda)\right]d\lambda,\tau\ge 0,$$

$$R_{хи}^{-}(\tau)=\int_0^{\infty} k(\lambda)R_{ии}(\tau-\lambda)d\lambda,\tau<0.$$

Пусть $R_{ии}^{-}(\tau)$ и $R_{хи}^{-}(\tau)$ – аналитические функции, допускающие аналитическое продолжение на дополнительную полуось, и выполнено соотношение $R_{хи}^{+}(\tau)\ne R_{хи}^{-}(\tau)$ при $\tau\ge 0$. Тогда справедлива следующая формула для передаточной функции *W(s)* линейной системы:

$$W(s)=\frac{R_{хи}^{+}(s)-R_{хи}^{-}(s)}{R_{ии}^{+}(s)-R_{ии}^{-}(s)},$$

где через $R_{хи}^{-}(s),\ R_{хи}^{-}(s),\ R_{ии}^{+}(s)$ и $R_{ии}^{-}(s)$ обозначены преобразования Лапласа от функций $R_{хи}^{-}(\tau),\ R_{хи}^{-}(\tau),\ R_{ии}^{+}(\tau)$ и $R_{ии}^{-}(\tau)$ соответственно.

Пусть функция *F(t)* определяется формулой

$$F(\tau)=\int_0^{\infty} k(\lambda)R_{ии}^{-}(\tau-\lambda)d\lambda,$$

тогда уравнение Винера-Хопфа может быть представлено в виде уравнения, в котором передаточная функция определяется по формуле

$$W(s)=\frac{R_{xu}^{+}(s)-F(s)}{R_{uu}^{+}(s)-R_{uu}^{-}(s)},$$

где *F(s)* – преобразование Лапласа от функции F(t).

Пусть функции $R_{uu}^{+}(\tau-\lambda)$ и $R_{uu}^{-}(\tau-\lambda)$ допускают представление

$$R_{uu}^{+}(\tau-\lambda)=\sum_{i=1}^{l}d_i(\tau)v_i(\lambda) \text{ при } \tau\geq\lambda,$$

$$R_{uu}^{-}(\tau-\lambda)=\sum_{i=1}^{l}v_i(\tau)d_i(\lambda) \text{ при } \tau<\lambda,$$

а постоянные $c_i$ вычисляются по формуле $c_i=\int_0^{\infty}k(\lambda)d_i(\lambda)d\lambda$.

Тогда передаточная функция системы определяется формулой

$$W(s)=\frac{R_{xu}^{+}(s)-\sum_{i=1}^{l}c_iV_i(s)}{R_{uu}^{+}(s)-R_{uu}^{-}(s)}.$$

Метод идентификации объектов управления, рассмотренный выше, справедлив для следующих двух случаев:

1) объект изолирован, т. е. не замкнут обратной связью; внутренние помехи присутствуют или отсутствуют;

2) объект замкнут обратной связью; внутренние помехи отсутствуют. Для первого случая изложенный метод справедлив в силу принятой гипотезы о независимости внутренних шумов от внешних воздействий.

Перейдем теперь к рассмотрению общего случая: объект, замкнутый обратной связью, работает в условиях внутренних помех (рис. 4.10) [20].

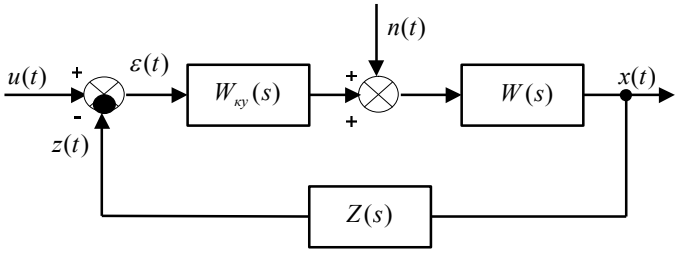



Рис. 4.10. Структурная схема автоматического управления с внутренней помехой

Уравнения этой системы могут быть представлены в следующем виде:

$$x(s) = W_{ky}(s)W(s)E(s) + W(s)N(s), \tag{4.65}$$

$$E(s) = u(s) - Z(s)x(s), \tag{4.66}$$

где $W_p(s) = W_{ky}(s)W(s), k_p(\tau)$ – соответственно передаточная и импульсная переходная функции прямой цепи; $W(s)$, $k(\tau)$ – передаточная и импульсная переходная функции объекта по отношению к внутреннему шуму; $Z(s), l(\tau)$ – передаточная и импульсная переходная функции обратной связи.

Из уравнения (4.29) получаем

$$R_{x\varepsilon}(\tau) = \int_0^\infty R_{\varepsilon\varepsilon}(\tau - \vartheta)k_p(\vartheta)d\vartheta + \int_0^\infty R_{n\varepsilon}(\tau - \vartheta)k(\vartheta)d\vartheta. \tag{4.67}$$

Как видно из уравнений (4.65), (4.67), введение обратной связи приводит к тому, что ошибка $\varepsilon(t)$ коррелированна с внутренними помехами $R_{n\varepsilon}(\tau) \neq 0$ и соотношение (4.62) не выполняется.

Ниже приведены возможные методы определения динамических характеристик линейных объектов при наличии обратных связей и внутренних помех [20].

Косвенный метод определения динамических характеристик объекта, замкнутого обратной связью. Найдем импульсные переходные функции замкнутой системы, а затем по ним получим динамические характеристики самого объекта.

Учитывая выражения (4.65), (4.67), имеем

$$x(s) = \hat{O}(s)E(s) + W_n(s)N(s), \tag{4.68}$$

где $\hat{O}(s), k_3(\tau)$ – соответственно передаточная и импульсная переходная функции замкнутой системы; $W_n(s), q(\tau)$ – соответственно передаточная и импульсная переходная функции системы по отношению к шуму;

$$\hat{O}(s) = \frac{W_p(s)}{1 + W_p(s)Z(s)}, \qquad W_n(s) = \frac{W(s)}{1 + W_p(s)Z(s)}.$$

Через импульсные переходные функции соотношение (4.68) запишется в виде

$$x(t) = \int_0^\infty u(t - \vartheta)k_3(\vartheta)d\vartheta + \int_0^\infty n(t - \vartheta)q(\vartheta)d\vartheta. \tag{4.69}$$

Используя последнее выражение и условие

$$R_{nu}(\tau) = 0, \tag{4.70}$$

получим

$$R_{xu}(\tau) = \int_0^\infty R_{uu}(\tau - \vartheta) k_3(\vartheta) d\vartheta. \tag{4.71}$$

Аналогичное соотношение

$$R_{Zx}(\tau) = \int_0^\infty R_{xx}(\tau - \vartheta) l(\vartheta) d\vartheta \tag{4.72}$$

можно записать для обратной связи.

Из уравнений (4.71), (4.72) находим $k_3(\tau)$ и $l(\tau)$, а по ним *Ф(s)* и *Z(s)* передаточную функцию прямой цепи определим по формуле

$$W_p(s) = \frac{\hat{O}(s)}{1 + \hat{O}(s) Z(s)}. \tag{4.73}$$

Остановимся теперь на приближенном способе.

Приближенный способ получения импульсных переходных функций объекта, замкнутого обратной связью при наличии внутренних помех.

Не нарушая общности отнесем помехи, возникающие внутри объекта, к его выходу, заменив $n(t)$ на $f(t)$, тогда

$$F(s) = W(s) N(s), \tag{4.74}$$

как показано на рис. 4.11.

Тогда

$$E(s) = \frac{1}{1 + W_{ky}(s) W(s) Z(s)} u(s) - \frac{Z(s)}{1 + W_{ky}(s) W(s) Z(s)} F(s), \tag{4.75}$$

или

$$\varepsilon(t) = \int_0^\infty u(t - \vartheta) \lambda(\vartheta) d\vartheta - \int_0^\infty f(t - \vartheta) \mu(\vartheta) d\vartheta, \tag{4.76}$$

где $\lambda(t), \mu(t)$ – импульсные переходные функции, соответствующие передаточным функциям

$$\frac{1}{1 + W_{ky}(s) W(s) Z(s)}, \quad \frac{Z(s)}{1 + W_{ky}(s) W(s) Z(s)}.$$

Используя уравнение (4.76) и условие $R_{uf}(\tau) = 0$, получим

$$R_{f\varepsilon}(\tau) = -\int_0^\infty R_{ff}(\tau - \vartheta)\mu(\vartheta)d\vartheta \,. \qquad (4.77)$$

Если помеха *f(t)* является белым шумом, то $R_{ff}(\tau) = \delta(\tau)$, где $\delta(\tau)$ – дельта-функция.

В этом случае

$$R_{f\varepsilon}(\tau) = -\int_0^\infty \mu(\vartheta)\delta(\tau + \vartheta)d\vartheta = -\mu(-\tau)\,, \qquad (4.78)$$

т. е. $R_{f\varepsilon}(\tau)$ является «перевернутой» импульсной переходной функцией замкнутой системы, как показано на рис. 4.12(*а*),

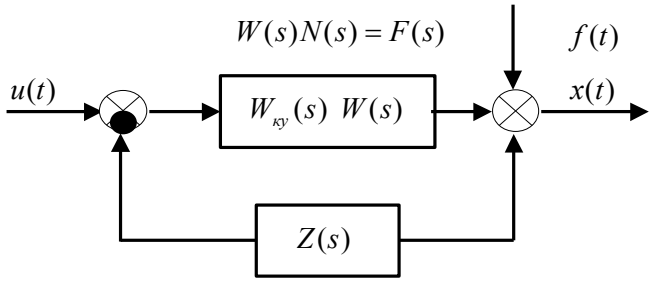



Рис. 4.11. Преобразованная структурная схема системы
с внутренней помехой

В большинстве случаев, когда помеха не является белым шумом $R_{ff}(\tau)$, полученная из эксперимента, имеет большой центральный пик и стремится, к нулю по обе стороны от $\tau = 0$. Для этого случая уравнение (4.78) интерпретировано на рис. 4.12 (*б*)**.** Вследствие некоторой протяженности $R_{ff}(\tau)$ взаимная корреляционная функция $R_{fe}(\tau)$ начинается на расстоянии *a* справа от $\tau = 0$. Чем более острый пик имеет $R_{ff}(\tau)$, тем ближе $R_{fe}(\tau)$ приближается к $-\mu(-\tau)$.

Учитывая уравнение (4.74), имеем

$$x(t) = \int_0^\infty \varepsilon(\tau - \vartheta)k_p(\vartheta)d\vartheta + f(t)\,. \qquad (4.79)$$

Используя уравнение (4.79), получим

$$R_{x\varepsilon}(\tau) = \int_0^\infty R_{\varepsilon\varepsilon}(\tau - \vartheta)k_p(\vartheta)d\vartheta + R_{f\varepsilon}(\tau)\,. \qquad (4.80)$$

Как видно из рис. 4.13, значениями $R_{fe}(\tau)$ при $\tau > a$ можно пренебречь и свести задачу определения $k_p(\tau)$ к решению уравнения

$$R_{x\varepsilon}(\tau) = \int_0^\infty R_{\varepsilon\varepsilon}(\tau - \vartheta) k_p(\vartheta) d\vartheta \qquad (4.81)$$

при

$$\tau > a\,. \qquad (4.82)$$

Условие (4.82) накладывает определенные ограничения на метод решения (4.81) [20].

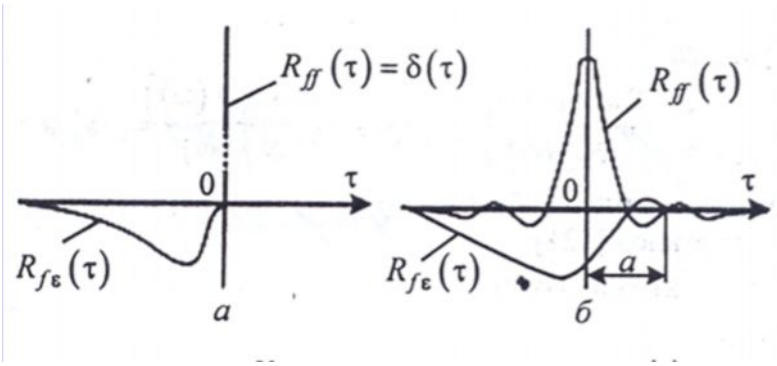



Рис. 4.12. Корреляционная функция $R_{ff}(\tau)$ и «перевернутая» импульсная переходная функция

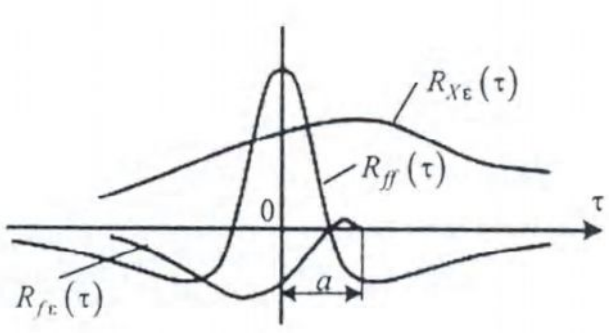



Рис. 4.13. К определению импульсной переходной функции

### 4.3.2. Методы решения уравнения Фредгольма 1-го рода

Решение интегрального уравнения (4.62) в частотной области. Умножим обе части уравнений (4.62) на $e^{jw\tau}$ и проинтегрируем по $\tau$ от $-\infty$ до $+\infty$.

В результате получим

$$\int_{-\infty}^{\infty} R_{xu}(\tau)e^{-jw\tau}d\tau = \int_{0}^{\infty} k(\theta)d\theta \int_{-\infty}^{\infty} R_{uu}(\tau-\theta)e^{-jw\tau}d\tau .$$

Произведя в правой части этого равенства замену переменной по формуле $\lambda = \tau - \theta$, будем иметь

$$\int_{-\infty}^{\infty} R_{xu}(\tau)e^{-jw\tau}d\tau = \int_{0}^{\infty} k(\theta)e^{-jw\tau}d\theta \int_{-\infty}^{\infty} R_{uu}(\lambda)e^{-jw\tau}d\lambda .$$

Вводя спектральные плотности

$$S_{xu} = \int_{-\infty}^{\infty} R_{xu}(\tau)e^{-jw\tau}d\tau ; S_{uu}(\omega) = \int_{-\infty}^{\infty} R_{uu}(\tau)e^{-jw\tau}d\tau$$

и учитывая, что

$$W(j\omega) = \int_{0}^{\infty} k(t)e^{-jwt}dt ,$$

получим

$$W(j\omega) = \frac{S_{xu}(\omega)}{S_{uu}(\omega)} . \qquad (4.83)$$

Из выражения (4.83) следует, что вещественная $P(\omega)$, мнимая $Q(\omega)$, частотные характеристики, соответствующие передаточной функции (4.83), могут быть найдены при помощи формул

$$P(\omega) = \frac{Re[S_{xu}(\omega)]}{S_{uu}(\omega)}, \quad Q(\omega) = \frac{Im[S_{xu}(\omega)]}{S_{uu}(\omega)}. \qquad (4.84)$$

Таким образом, статистический метод определения передаточной функции состоит из двух основных этапов [15, 20]:

1) определения спектральных плотностей по заданным реализациям входа *u(t)* и выхода *x(t)*;

2) определения частотных характеристик или передаточной функции при помощи формул (4.85).

О точности определения динамических характеристик. Воспользовавшись формулой (4.83) запишем импульсную переходную функцию *k(t)* следующим образом:

$$k(\tau) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{S_{xu}(\omega)}{S_{uu}(\omega)} e^{jw\tau}d\omega . \qquad (4.85)$$

Очевидно, что формула (4.85) справедлива лишь при условии, что $S_{uu}(\omega) \neq 0$ при всех значениях $\omega$. Если же на некотором интерва-

ле $(\omega_1, \omega_2)$ $S_{uu}(\omega) = 0$, а $S_{xu}(\omega) \neq 0$, то подынтегральное выражение в формуле (4.85) обращается в бесконечность в указанном интервале. При этом импульсная переходная функция *k(t)* может отличаться от истинной сколь угодно сильно. Но передаточная функция *W(jw)* будет отличаться от истинной только в интервале частот $(\omega_1, \omega_2)$, которые и не содержались в воздействии *u(t)*, т. к. в этом интервале $S_{uu}(\omega) = 0$.

Рассмотрим теперь случай, когда исходные корреляционные функции известны с некоторой ошибкой. Тогда и соответствующие спектральные плотности также будут содержать ошибку. В этом случае выражение (4.85) может быть записано в виде

$$k(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{S_{xu}(\omega) + \delta_{xu}(\omega)}{S_{uu}(\omega) + \delta_{uu}(\omega)} e^{jw\tau} d\omega, \qquad (4.86)$$

где $\delta_{xu}(\omega)$ и $\delta_{uu}(\omega)$ – спектры ошибок.

Непосредственно из рассмотрения формулы видно, что сколь угодно малая ошибка $\delta$ может вызывать сколь угодно большое искажение решения уравнения (4.85), если спектр ее распределен на тех участках оси $\omega$, где величины $S_{xu}(\omega)$ и $\delta_{xu}(\omega)$ становятся сравнимыми с $S_{uu}(\omega)$ и $\delta_{uu}(\omega)$. Из сказанного следует, что необходимое условие для точного определения функции *k(t) по* формуле (4.85) состоит в том, чтобы спектр частот входного сигнала был значительно шире предполагаемой полосы пропускания исследуемого объекта.

От любого метода определения динамической характеристики разумно требовать такого приближения $k(\tau)$, которое сводило бы разность выходных процессов реального объекта и его модели к некоторой малой величине при одном и том же процессе на входе. Действительно, только по отклонению выходных величин при одинаковом процессе на входе мы можем судить о неидентичности объектов с точки зрения их динамических характеристик.

Естественно считать критерием или мерой точности определения динамической характеристики дисперсию этой разности, т. е.

$$M\left\{\left[x_m(t) - x(t)\right]^2\right\} = M\left\{\varepsilon^2\right\}. \qquad (4.87)$$

Очевидно, что

$$M\left\{\varepsilon^2\right\} = \frac{1}{2\pi} \int_{-\infty}^{\infty} \left|W_M(jw) - W(jw)\right|^2 S_{uu}(\omega) d\omega. \qquad (4.88)$$

Введем в рассмотрение относительное среднеквадратичное отклонение

$$\delta_\varepsilon = \sqrt{\frac{\int\limits_{-\infty}^{\infty} \left|W_M(jw) - W(jw)\right|^2 S_{uu}(\omega)d\omega}{\int\limits_{-\infty}^{\infty} \left|W(jw)\right|^2 S_{uu}(\omega)d\omega}} . \qquad (4.89)$$

Физический смысл этого критерия заключается в том, что при оценке погрешности определения динамической характеристики принимается во внимание реакция объекта лишь на те частоты, которые присутствуют во входном процессе, причем с весом, пропорциональным спектральной плотности входного процесса $\mu(t)$.

Критерий (4.88) дает ответ и на вопрос, какую информацию содержит решение (4.86) уравнения (4.87) о динамике исследуемого объекта. Действительно, в случае абсолютно точного определения $k(\tau)$ имеем $\delta_\varepsilon = 0$.

Если $k(\tau)$ определена не точно, но спектр ошибки сосредоточен в той области, где $S_{uu}(\omega) = 0$, то $\delta_\varepsilon = 0$. Это означает, что по выходу мы не сможем отличить модель от реального объекта.

Если $k(\tau)$ определена не точно и спектр ошибки распределен произвольным образом, то $\delta_\varepsilon \neq 0$ и критерий (4.89) характеризует погрешность определения динамической характеристики на тех частотах, присутствие которых во входном процессе существенно. Поэтому, несмотря на возможное большое отклонение $k_M(\tau)$ от $k(\tau)$, ошибка может быть малой, и выходные процессы реального объекта и его модели будут отличаться лишь незначительно.

2. Метод квадратур. Рассмотрим решение уравнения (4.62), пользуясь методом квадратур. Представим уравнение (4.62) в виде

$$R_{xu}(\tau) = \sum_{N=0}^{N} K(n\Delta) R_{uu}(\tau - n\Delta)\Delta .$$

Обозначим

$$A = \begin{pmatrix} R_{uu}(0) & R_{uu}(\Delta) & \cdots & R_{uu}((N-1)\Delta) \\ R_{uu}(\Delta) & R_{uu}(0) & \cdots & R_{uu}((N-2)\Delta) \\ \vdots & \vdots & \ddots & \vdots \\ R_{uu}((N-1)\Delta) & R_{uu}((N-2)\Delta) & \cdots & R_{uu}(0) \end{pmatrix}, \qquad (4.90)$$

$$K=\begin{pmatrix} k(\Delta) \\ k(2\Delta) \\ k(3\Delta) \\ \vdots \\ k(N\Delta) \end{pmatrix}, \qquad Q=\begin{pmatrix} q_1 \\ q_2 \\ q_3 \\ \vdots \\ q_N \end{pmatrix}, \tag{4.91}$$

где

$$q_1=\frac{R_{xu}(i\Delta)}{\Delta}. \tag{4.92}$$

Тогда (4.89) перепишем в виде

$$AK=Q. \tag{4.93}$$

В (4.93) $A=(a_{ij})_{i,j=0}^{N-1}$ – квадратная симметричная матрица размерности $N\times N$.

Как и в случае решения уравнения свертки, задача расчета импульсной переходной функции свелась к решению системы линейных алгебраических уравнений. Последняя является симметричной с доминирующей главной диагональю. Это свойство порождено свойствами корреляционных функций, входящих в исходное уравнение Фредгольма 1-го рода (4.62).

Статистический метод идентификации устраняет недостатки, присущие детерминированному подходу; вместе с тем этот метод также имеет недостатки, определяемые статистическим содержанием эксперимента:

• время, необходимое для решения задачи идентификации, значительно увеличивается;

• имеют место ошибки построения $R_{uu}(\tau)$ и $R_{xu}(\tau)$ данным нормальной эксплуатации, определяемые величиной промежутка усреднения (известно, что ошибка определения корреляционной функции путем усреднения по времени убывает только как $\tau^{-0.5}$ и для достижения точности в несколько процентов требуется время накопления Т в 100 и более раз превышающее $\tau$ [15, 20]), а также наличием фактора не стационарности, низкой точностью измерений и др.

Пример 4.3 [15, 20]. Проиллюстрируем изложенные выше соображения на примере. Предположим, что

$$R_{uu}(\tau)=\begin{cases} e^{-\alpha\tau}-0{,}5 & \text{ïðè}\quad 0\le\tau\le T;\\ 0 & \text{ïðè}\quad \tau>T; \end{cases}$$

$$R_{xu}(\tau)=\begin{cases} 1{,}2102\sin 0{,}2618\tau+3{,}1213e^{-\alpha\tau}+2{,}3228e^{\alpha\tau}-3{,}8197 & \text{ïðè}\quad 0\le\tau\le T\\ 6{,}2022e^{-\alpha\tau}-0{,}36306\sin\dfrac{\pi}{12}(\tau-12)+0{,}3229\cos\dfrac{\pi}{12}(\tau-12)-1{,}9099 & \text{ïðè}\quad 0\le\tau\le 2T\\ 0 & \text{ïðè}\quad \tau>2T\\ \alpha=0{,}004257;\quad T=12. \end{cases}$$

Точное решение уравнения (4.62), полученное аналитически, в рассматриваемом случае имеет вид

$$k(t)=\begin{cases} sin(t\pi/2) & \text{ïðè}\quad 0\le\tau\le T;\\ 0 & \text{ïðè}\quad \tau>T. \end{cases}$$

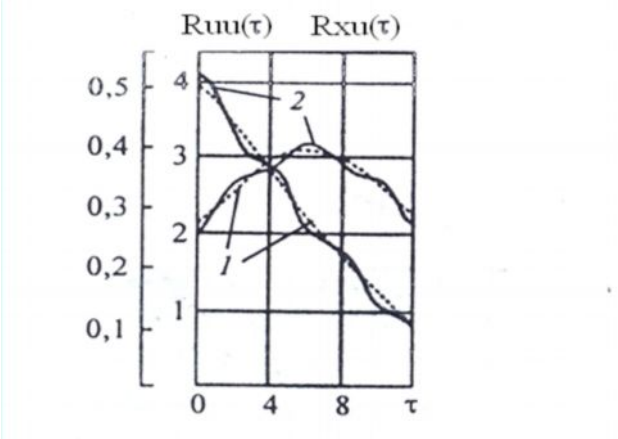



Рис. 4.14. Корреляционные и взаимокорреляционные функции

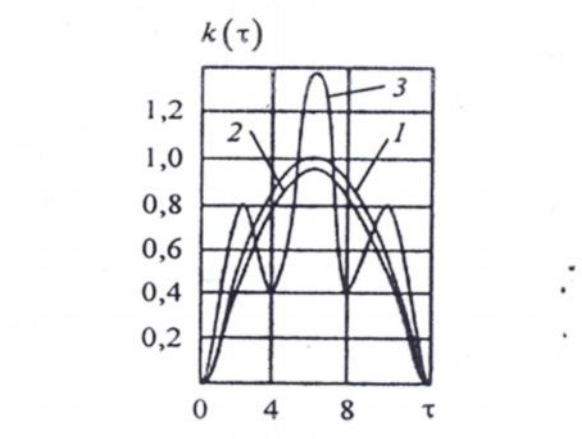



Рис. 4.15. Импульсные переходные функции при различных точностях выполнений

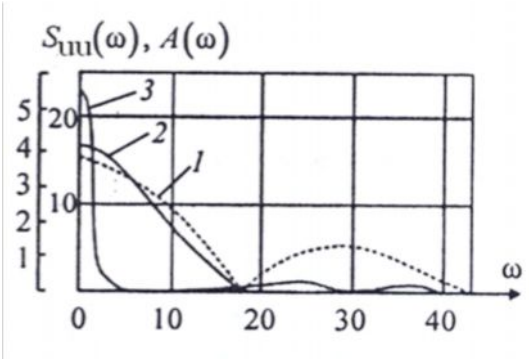



Рис. 4.16. Точные и приближенные значения амплитудной характеристики

На рис. 4.14 и 4.15 представлены графики функций $R_{uu}(\tau), R_{xu}(\tau), k(\tau)$ (кривые 1). Решая то же уравнение методом квадратур при $\Delta = 2, N = 7$, получим приближенное решение, довольно близкое к точному (рис. 4.16.). График этого решения изображен на рис. 4.15 (кривая 2). При незначительном изменении исходных данных, не превышающем 3 % от максимального значения, графики измененных функций $R_{uu}(\tau)$ и $R_{xu}(\tau)$ изображены на рис. 4.14 (кривые 2), решение же $k(\tau)$ (кривая 3 на рис. 4.15) значительно отличается от точного.

Пример 4.4 [28]. Рассмотрим задачу определения ИПФ системы, определяющей преобразование вибраций от двигателя легкового автомобиля к кузову при прямолинейном движении с постоянной скоростью (рис. 4.17).

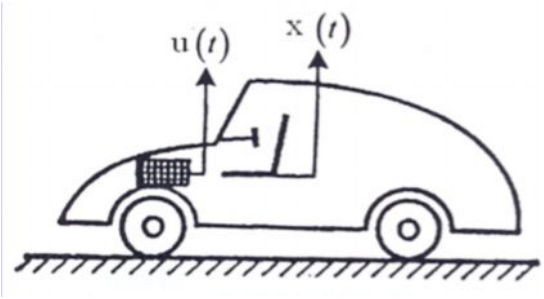



Рис. 4.17. Эксперимент по изучению нерегулярных вибраций легкового автомобиля (постоянная скорость 20 км/ч)

Считая, что входной сигнал $u$(t) и выходной процесс $x$(t) в установившемся режиме связаны соотношением $x(t) = \int_0^\infty k(\tau)u(t-\tau)d\tau$

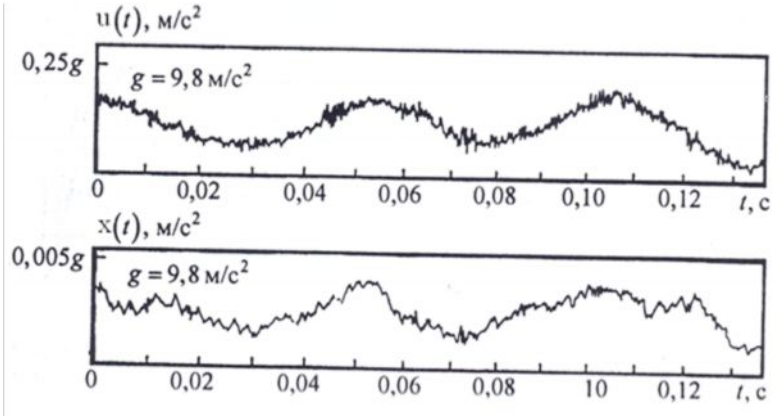



Рис. 4.18. Графики изменения ускорений вертикальной вибрации двигателя и кузова в процессе движения легкового автомобиля

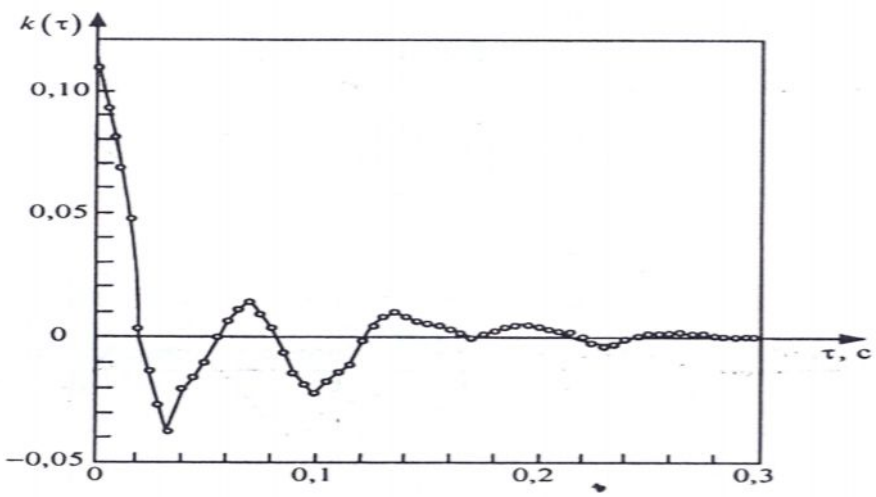



Рис. 4.20. График импульсной переходной функции системы «двигатель легкового автомобиля – кузов автомобиля» при прямолинейном движении с постоянной скоростью

Пример 4.5. Рассмотрим несколько другой подход к решению задачи идентификации, использующий аппроксимацию авто и взаимной корреляционных функций и позволяющий найти ИПФ в аналитическом виде.

Положим, что приближенное аналитическое уравнение для нормированной корреляционной функции входа *u(t)* имеет вид

$$r_{uu}(\tau) = e^{-0,008\tau} \cos\frac{\pi\tau}{200},$$

а средние квадратические отклонения входа $u(t)$ и выхода $x(t)$ равны $\sigma_x = 0.005$ и $\sigma_u = 0,33$.

На рис. 4.19 приведен график нормированной корреляционной функции входа $R_{uu}(\tau)$.

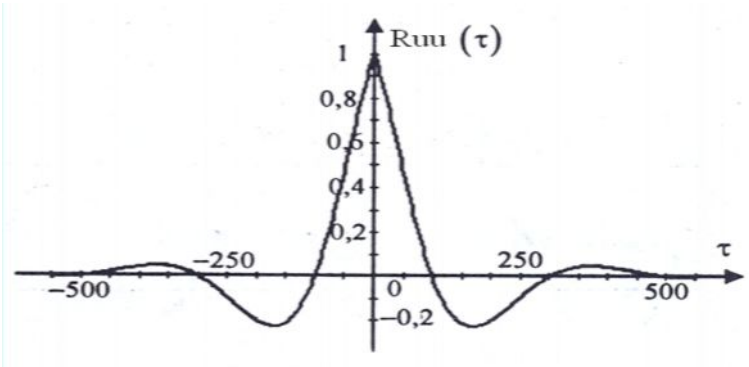



Рис. 4.21. График $Ruu(\tau)$

Взаимная корреляционная функция определяется зависимостью

$$r_{xu}(\tau)=\begin{cases}0{,}45e^{-0{,}0056\tau}\left(\cos\dfrac{\pi\tau}{200}+1{,}376\sin\dfrac{\pi\tau}{200}\right), & \tau\geq 0;\\ 0{,}45e^{-0{,}008\tau}\left(\cos\dfrac{\pi\tau}{200}+1{,}376\sin\dfrac{\pi\tau}{200}\right), & \tau\leq 0.\end{cases}$$

График взаимной корреляционной функции показан на рис. 4.22.

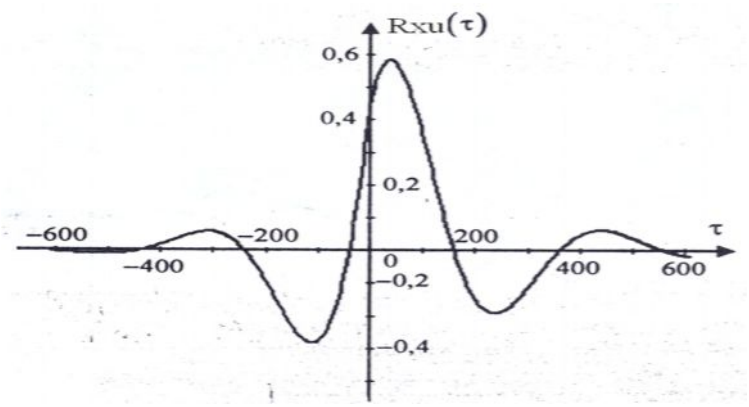



Рис. 4.22. График $Ruu(\tau)$

Из приведенных соотношений для нормированной корреляционной и взаимной корреляционной функции легко получить соответствующие корреляционные входа и взаимные корреляционные входа и выхода функции. Для рассматриваемого случая получим

$$R_{xu}(\tau)=\begin{cases}0{,}74\cdot 10^{-3}e^{-0{,}0056\tau}\left(\cos\dfrac{\pi\tau}{200}+1{,}376\sin\dfrac{\pi\tau}{200}\right), & \tau\geq 0;\\ 0{,}74\cdot 10^{-3}e^{-0{,}008\tau}\left(\cos\dfrac{\pi\tau}{200}+1{,}376\sin\dfrac{\pi\tau}{200}\right), & \tau\leq 0.\end{cases}$$

Теперь можно записать интегральное уравнение для определения ИПФ объекта:

$$0{,}74\cdot 10^{-3}e^{-0.0056\tau}\left(\cos\frac{\pi\tau}{200}+1{,}376\sin\frac{\pi\tau}{200}\right)=\int_0^{\infty}k(\lambda)e^{-0{,}008(\tau-\lambda)}\cos\frac{\pi(\tau-\lambda)}{200}d\lambda\,.$$

Выражение для ИПФ имеет вид

$$k(\tau)=8{,}5\cdot 10^{-3}\left[0{,}74\delta(\tau)+0{,}7e^{-0{,}0176\tau}-2\cdot 10^{-6}\cos 0{,}5\pi e^{-0.0056\tau}-1{,}3\sin 0{,}05\pi e^{-0{,}0056\tau}\right].$$

где $\delta$ – дельта-функция.

Применяя преобразование Лапласа, получим следующую зависимость для передаточной функции:

$$W(s) = 8{,}5 \cdot 10^{-3} \frac{0{,}74s + 0{,}04}{s + 0{,}0176} \cdot \frac{(s + 0{,}008)^2 + 0{,}0026}{(s + 0{,}0056)^2 + 0{,}0026}.$$

Дифференциальное уравнение рассматриваемого объекта будет иметь вид

$$\frac{d^3x}{dt^3} + 0{,}0288\frac{d^2x}{dt^2} + 0{,}0028\frac{dx}{dt} + 4 \cdot 10^{-5}x =$$

$$= 8{,}5 \cdot 10^{-3}\left(0{,}74\frac{d^3y}{dt^3} + 0{,}0518\frac{d^2y}{dt^2} + 0.0026\frac{dy}{dt} + 0{,}0001y\right).$$

# ГЛАВА 5. МЕТОДЫ ОЦЕНИВАНИЯ ПАРАМЕТРОВ

## 5.1. Общая схема оценивания параметров

Рассмотрим стандартную схему на рис. 5.1, являющуюся частью общей схемы метода идентификации.

В соответствии с данной схемой задача идентификации формируется следующим образом. По известным в интервале идентификации значениям последовательностей входа $\{u(t)\}$ и выхода $\{x(t)\}\,(t=1,...,N)$ определить при наличии возмущений $\nu(t)$ вектор неизвестных параметров $\theta$ объекта (процесса).

Решение задачи идентификации осуществляется с использованием невозмущенной модели объекта, на вход которой поступают заданные значения $\{u(t)\}$. Поскольку вектор параметров $\theta$ объекта неизвестен, для построения модели используется оценка параметров $\hat{\theta}$, рассчитываемая так, чтобы получить определенное соответствие между выходными сигналами модели и объекта $e(t)=x(t)-\hat{x}(t)$. Очевидно, что при действии на объект возмущений $\nu$ (t) даже при полном совпадении параметров модели и процесса $\hat{\theta}=\theta$ будет иметь место ошибка $e(t)=x(t)-\hat{x}(t)$ Из рис. 5.1 следует соотношение $e(t)=\nu(t)$ при $\hat{\theta}=\theta$.

Следовательно, если оценка $\hat{\theta}$ совпадает с действительным значением параметров, ошибка *e(t)* является реконструкцией возмущения, действующего на процесс. В этом случае для получения лучшей оценки шума необходимо иметь больший объем данных входных и выходных сигналов.

Положим, что динамическая система с дискретным временем с одним входом и одним выходом описывается линейным обобщенным регрессионным уравнением вида

$$x(t)=\sum_{i=1}^{n_a} a_i x(t-i)+\sum_{j=1}^{n_b} b_j u(t-j)+\nu(t), \qquad (5.1)$$

где *u(t)* – входная переменная; *x(t)* – выходная переменная; $\nu$ (t) – помехи измерения, которые присутствуют в каждый наблюдаемый момент времени $t$ $(t=1,...,N)$.

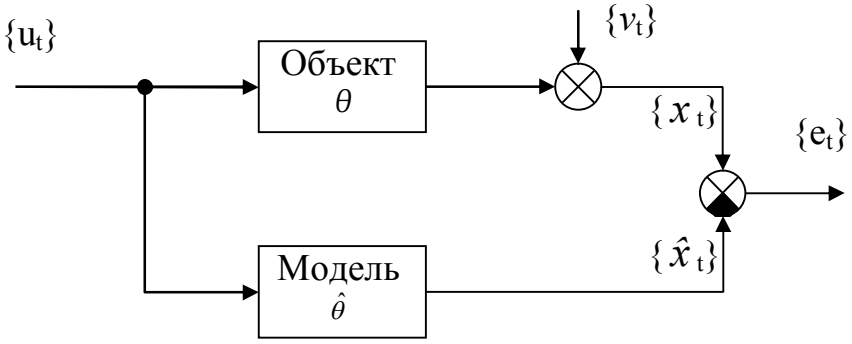



Рис. 5.1. Общая часть стандартной схемы метода идентификации

Линейный процесс (5.1) может быть также представлен в форме

$$A(z^{-1})x(t) = B(z^{-1})u(t) + v(t), \qquad (5.2)$$

где $A(z^{-1})$, $B(z^{-1})$ – полиномы в операторе обратного сдвига $z^{-1}$

$$A(z^{-1}) = 1 + a_1 z^{-1} + \ldots + a_{n_a} z^{-n_a},$$

$$B(z^{-1}) = 1 + b_1 z^{-1} + \ldots + b_{n_b} z^{-n_b}.$$

Обобщенное регрессионное уравнение (5.1) является удобной формой записи для линейной динамической системы с измеряемыми входными и выходными сигналами. Для этого случая целесообразно рассмотреть более простые регрессионные уравнения, которые могут быть получены из обобщенной формы (5.1), т. е. обычная регрессионная модель,

$$x(t) = \sum_{j=1}^{n_b} b_j u(t-j) + v(t) \qquad (5.3)$$

и линейная авторегрессионная модель

$$x(t) = \sum_{i=1}^{n_a} a_i x(t-i) + v(t) \qquad (5.4)$$

с запаздывающими выходными величинами $x(t-i)$.

Неизвестные параметры $a_i$, $b_j$ $\left(i = \overline{1, n_a},\ j = \overline{1, n_b}\right)$ могут рассматриваться как весовые коэффициенты регрессионной модели (5.1). Они могут быть оценены с помощью измеряемых величин *x(t)* и *u(t)* в моменты времени *t*.

Будем полагать, что верхние границы $n_a$, $n_b$ в (5.1) известны. Вместе с тем, без потери общности можно положить $n_a = n_b = n$.

Если проделана серия из N экспериментов, общее количество уравнений, полученных из (5.1) может быть представлено как

$$x = X\,a + X\,b + \nu = \left[X : U\right]\begin{bmatrix} a \\ b \end{bmatrix} + \nu = z^T\,\theta + \nu, \qquad (5.5)$$

где

$$x^T = \left[x(1),...,x(N)\right],\quad \theta^T = \left[a_1,...,a_n,b_1,...,b_n\right],$$

$$X = \begin{bmatrix} x(1)...x(1-n) \\ x(2)...x(2-n) \\ .\qquad .... \\ x(N)...x(N-n) \end{bmatrix}_{Nxn}, \quad U = \begin{bmatrix} u(1)..u(1-n) \\ u(2)..u(2-n) \\ ..\qquad ... \\ u(N)..u(N-n) \end{bmatrix}_{Nxn}. \qquad (5.6)$$

$$Z = \left[X : U\right], \qquad \nu^T = \left[\nu(1),...,\nu(N)\right]$$

При этом $N >> 2\,n$.

Аналогичные представления уравнений справедливы и для моделей (5.3), (5.4), которые можно записать как

$$x = Ub + \nu, \qquad x = Xa + \nu \qquad (5.7)$$

или в виде

$$x = U\theta + \nu, \qquad (5.8)$$

$$x = X\theta + \nu. \qquad (5.9)$$

Для разомкнутых систем входные величины *u(t-j)* могут рассматриваться как константы. С другой стороны, для замкнутых систем величины *u(t)* вычисляются с помощью измеряемых предыдущих значений выходных переменных. Следовательно, величины *u(t)* зависят от предыдущих величин *x(t)*, которые обусловлены шумом $\nu(t)$ и являются наблюдаемыми величинами случайной переменной *x(t)*.

Отсюда следует, что все переменные замкнутой системы, описываемой обобщенной регрессионной моделью (5.1) или ее более простыми версиями, есть случайные переменные. При условии, что шум является гауссовским процессом, его можно представить с помощью первого и второго статистических моментов, т. е. с помощью среднего и его ковариационной матрицы

$$M\left\{\nu^T\right\} = M\left\{\left[\nu(1),...,\nu(N)\right]\right\}, \qquad (5.10)$$

$$K_\nu = M\left\{\nu\nu^T\right\} = \begin{bmatrix} M\{\nu(1)\nu(1)\}...M\{\nu(N)\nu(1)\} \\ ...\qquad\qquad ...\qquad\qquad ... \\ M\{\nu(1)\nu(N)\}...M\{\nu(N)\nu(N)\} \end{bmatrix}.$$

В зависимости от природы случайных воздействий $\nu(t)$ и объема априорной информации об их вероятностных характеристиках ковариационная матрица $K_\nu$ шума в теории оценивания может быть представлена несколькими вариантами при $M\{\nu(j)\}=0$.

Первый случай – наиболее простой случай, когда априори известно, что наблюдения некоррелированы и равноточны. При этом ковариационная матрица шума определяется выражением

$$K_\nu = \sigma_\nu^2 E, \tag{5.11}$$

где E – единичная матрица; $\sigma_\nu^2$ – дисперсия случайных величин.

Второй случай – наблюдения неравноточные, но независимые. При этом

$$K_\nu = \begin{bmatrix} \sigma_{1\nu}^2 & . & 0 \\ . & . & . \\ 0 & . & \sigma_{N\nu}^2 \end{bmatrix}. \tag{5.12}$$

Третий случай – наблюдения равноточные, но коррелированные, т. е.

$$K_\nu = \sigma_\nu^2 R, \tag{5.13}$$

где R – нормированная ковариационная матрица [7, 10, 19].

Наконец, четвертый, самый общий случай – наблюдения коррелированные и неравноточные.

Для идентификации параметров системы по результатам последовательности $z^T(N)=[z(1),...,z(N)]$ входных и выходных переменных из N наблюдений над системой с параметрами $\theta$ большое значение имеет объем априорной информации. При этом предполагают, что истинное значение параметров системы $\theta$ принадлежит компактному подмножеству $\Omega_\theta$ в $H$, где $H$ - разделимое метрическое пространство с метрикой $d$.

До сих пор предполагалось, что оцениваемый параметр является неизвестной, но не случайной величиной, т. е. во всех наблюдениях случайных величин, описываемых данным распределением вероятностей, этот параметр принимает одинаковые значения. Теперь будем считать, что оцениваемый параметр является случайной величиной, вообще говоря, векторной со случайными компонентами $\theta_1,...,\theta_r$.

Случайный параметр $\theta$ возможно также оценивать по наблюдениям некоторой другой случайной последовательности $z(N)$ статистически

связанной со случайным вектором $\theta$. Предполагается, что при каждом наблюдении случайного вектора (или, другими словами, случайной последовательности) *z(N)* реализуется некоторое значение случайного вектора $\theta$. Таким образом, всевозможные наблюдения распадаются на некоторое (в общем случае бесконечное) число классов, для каждого из которых случайный вектор $\theta$ принимает свое определенное значение $\theta \in \Omega_\theta$. Каждый такой класс наблюдений случайной величины X характеризуется условной плотностью вероятности

$$p(z \mid \theta) \tag{5.14}$$

случайного вектора *z* при условии, что вектор $\theta$ принял значение, определенное для этого класса.

Будем полагать, что условная плотность вероятности $p(z \mid \theta)$ известна для всех возможных значений $\theta$. Так как параметр $\theta$ является случайной величиной, он имеет свое распределение вероятностей или плотностью вероятности $p(\theta)$. Плотность вероятности $p(\theta)$, не зависящая от наблюдений *z*, называется априорной плотностью вероятности случайного параметра $\theta$.

Всякая случайная величина $\theta$ полностью описывается своим вероятностным распределением (например, плотностью вероятности $p(\theta)$). Если эта величина статистически связана с какой-либо другой случайной величиной *X*, а случайная величина *z* в результате испытания приняла какое-либо определенной значение, то случайная величина $\theta$ будет полностью описана условной плотностью вероятности $p(\theta \mid z)$ при условии, что *z=c*. Таким образом, вся информация о случайной величине $\theta$, получаемая в результате измерений с, заключена в плотности вероятности $p(\theta \mid z)$, которая как функция от $\theta$ называется апостериорной плотностью вероятности случайного параметра $\theta$.

В этом случае для процедуры оценивания значения $\theta$, основанной на последовательности наблюдений $\{z(N)\}$, необходимо знать априорную плотность распределения вероятности наблюдений *p(z),* а также совместную плотность вероятности $p(\theta, z)$. Поскольку

$$p(\theta, z) = p(\theta \mid z)p(z) = p(\theta \mid z)p(\theta), \tag{5.15}$$

то в ряде случаев более естественно предполагать, что задана функция априорной плотности распределения $\theta$, т. е. $p(\theta)$. В задачах параметрической идентификации, в дополнение к отмеченному, методы оценивания базируются на минимизации некоторой скалярной функ-

ции потерь $Q(\theta,\hat{\theta})$, которая наилучшим в некотором смысле соответствует имеющейся информации.

Таким образом, при рассмотрении методов идентификации, возможно использовать следующий объем априорной информации такой как [1, 30]:

а) плотность вероятности шума $p(\nu)$, по которой можно рассчитывать функцию плотности распределения $p(\theta|z)$;

б) плотность распределения вероятностей идентифицируемых параметров $p(\theta)$;

в) числовая функция потерь $Q(\theta,\hat{\theta})$.

## 5.2. Байесово оценивание параметров

В тех случаях, когда статистическое описание измеряемых данных $z$ и параметров $\theta$ является полным, т. е. заданы функции плотностей распределения вероятностей $p(z|\theta)$ и $p(\theta)$, оптимальная оценка, доставляющая максимум апостериорной функции плотности распределения вероятности $p(\theta|z)$, называется байесовской.

Пусть априорная информация включает значение функций плотностей распределения вероятностей $p(\theta|z)$ и $p(\theta)$, а также $Q(\theta,\hat{\theta})$.

Рассмотрим формулу Байеса для определения функции $p(\theta|z)$

$$p(\theta|z)=\frac{p(z|\theta)p(\theta)}{p(z)}, \tag{5.16}$$

где

$$p(z)=\int_{\Omega_\theta} p(z|\theta)p(\theta)d\theta, \tag{5.17}$$

а $\Omega_\theta$ – область определения вектора $\theta$. Следует отметить, что интеграл (5.17) многомерный по $d\theta_1,...,d\theta_r$, поскольку $\theta$ – вектор и берется по всей области, где р($\theta$) отлична от нуля. Таким образом, апостериорная плотность вероятности можно определить, если известна априорная плотность $p(\theta)$. Плотность вероятности $p(z|\theta)$ предполагается всегда известной.

Тогда в соответствии с исходным уравнением априорной определенности все члены в правой части соотношения (5.16) известны. При

известной апостериорной плотности распределения $p(\theta|z)$ оценка вектора параметров $\hat{\theta}$ определяется в процессе минимизации величины

$$J(\hat{\theta}) = \int_{\Omega_\theta} Q(\theta,\hat{\theta})\, p(\theta \mid z)\, d\theta. \qquad (5.18)$$

Вывод формулы (5.18) можно получить следующим образом. Пусть оценивается вектор параметров $\theta$. Рассмотрим среднее значение потерь по различным выборкам, учитывая частоту их появления

$$r(\hat{\theta}) = \int_{\Omega_\theta} Q(\theta,\hat{\theta})\, p(\theta \mid z) dz. \qquad (5.19)$$

Данное среднее значение в математической статистике называется условной функцией риска.

Усредняя для случайного вектора $\theta$ с плотностью распределения $p(\theta)$ условную функцию риска по возможным значениям $\hat{\theta}$, получаем средний риск

$$R(\hat{\theta}) = \int_{\Omega_\theta}\int_{\Omega_z} p(\theta)\, p(z \mid \theta)\, Q(\theta,\hat{\theta})\, d\theta\, dz. \qquad (5.20)$$

Запишем последнее соотношение с учетом формулы Байеса (5.16)

$$R(\hat{\theta}) = \int_{\Omega_\theta}\int_{\Omega_z} p(z)\, p(\theta \mid z)\, Q(\theta,\hat{\theta})\, d\theta\, dz = \int_{\Omega_z} p(z) dz \int_{\Omega_\theta} Q(\theta,\hat{\theta})\, p(\theta \mid z)\, d\theta. \qquad (5.21)$$

Так как $p(z) \geq 0$ и $\int_{\Omega_z} p(z) dz \leq 1$, то для оценивания вектора $\hat{\theta}$ вместо минимизации $R(\hat{\theta})$ для (5.21) можно минимизировать внутренний интеграл (5.21)

$$J(\hat{\theta}) = \int_{\Omega_\theta} Q(\theta,\hat{\theta})\, p(\theta \mid z)\, d\theta, \qquad (5.22)$$

т. е. минимизация апостериорного риска при достижении максимума апостериорной функции плотности распределения вероятности $p(\theta|z)$ при известной выборке z.

Необходимое условие минимума функции $J(\hat{\theta})$ имеет вид

$$\frac{d}{d\hat{\theta}} \int_{\Omega_\theta} Q(\theta,\hat{\theta})\, p(\theta \mid z)\, d\theta = 0. \qquad (5.23)$$

Полученная оценка называется байесовской, поскольку для вывода требуемых соотношений использовалась формула Байеса.

Из структуры выражения (5.23) следует, что функция $J(\hat{\theta})$ является результатом усреднения функции потерь $Q(\theta,\hat{\theta})$ по апостериорному распределению $p(\theta|z)$

$$J(\hat{\theta}) = M_{\theta|z}\ \left\{ Q(\theta,\hat{\theta}) \right\}.$$

Байесовское оценивание, как это следует из его сущности, дает общую структуру решения задачи оценивания параметров и инвариантно к классам функции потерь, объему выборки Z и к виду функции плотности распределения вероятности $p(z|\theta)$, $p(\theta)$. Байесово оценивание дает наилучшие (абсолютно оптимальные) оценки для данного вектора наблюдений z, но для получения конкретных значений оценок $\hat{\theta}$ необходимо задание $p(z|\theta)$ и $p(\theta)$. Адаптивное байесово оценивание при параметрической и непараметрической неопределенности задания $p(z|\theta)$ и $p(\theta)$, связанное с заменой точной количественной меры ожидаемых потерь – апостериорного поиска – его состоятельной оценкой по имеющимся данным измерений z.

Одной из возможных оценок параметра $\theta$, получаемых на основе его апостериорной плотности вероятности, может быть то значение $\hat{\theta}$, при котором $p(\theta|z)$, достигает максимального значения. Условия, которым удовлетворяет точка максимума p $(\theta|z)$ есть оценка решения уравнения

$$\frac{\partial\, p(\theta|z)}{\partial\theta}\Big|_{\theta=\hat{\theta}(Z)} = 0, \qquad \text{или} \qquad \frac{\partial\, ln\ p(\theta|z)}{\partial\theta}\Big|_{\theta=\hat{\theta}(Z)} = 0. \quad (5.24)$$

Такую оценку называют оценкой, максимизирующей апостериорную вероятность, которую с использованием формулы Байеса (5.16) можно представить в виде

$$ln\ p(\theta|z) = ln\ p(z|\theta) + ln\ p(\theta) - ln\ p(z), \quad (5.25)$$

$$\frac{\partial\, ln\ p(\theta|z)}{\partial\theta}\Big|_{\theta=\hat{\theta}(Z)} = 0, \frac{\partial\, ln\ p(z|\theta)}{\partial\theta}\Big|_{\theta=\hat{\theta}(Z)} + \frac{\partial\, ln\ p(\theta)}{\partial\theta}\Big|_{\theta=\hat{\theta}(Z)} = 0. \quad (5.26)$$

Если

$$\frac{\partial\, ln\ p(z|\theta)}{\partial\theta} >> \frac{\partial\, ln\ p(\theta)}{\partial\theta}. \quad (5.27)$$

или, иными словами, плотность вероятности $p(\theta)$ является более «плоской» функцией, чем $p(z|\theta)$, то можно пренебречь вторым членом в выражении (5.26), и оценка, максимизирующая апостериорную вероятность, становится оценкой максимального правдоподобия.

Конкретные оценки $\hat{\theta}$ свойства определяются выбором вида функции потерь $Q(\theta,\hat{\theta})$. В теории оценивания и теории идентификации используется ряд функций потерь. Рассмотрим следующие наиболее употребительные функции потерь для скалярного случая:

1. Равномерная функция потерь

$$Q_1(\theta,\hat{\theta}) = c - \delta(\theta - \hat{\theta}), \quad c > 0, \tag{5.28}$$

где $\delta(x)$ – дельта-функция, обладающая следующими свойствами

$$\delta(x) = \begin{cases} 0, & x \neq 0, \\ \infty, & x = 0; \end{cases}$$

2. Функция потерь, пропорциональная модулю ошибки идентификации (абсолютной ошибке)

$$Q_2(\theta,\hat{\theta}) = \left|\theta - \hat{\theta}\right|; \tag{5.29}$$

3. Квадратичная функция потерь

$$Q_3(\theta,\hat{\theta}) = (\theta - \hat{\theta})^2. \tag{5.30}$$

Рассмотрим оценки скалярного параметра $\theta$, получаемые путем минимизации функции $J(\theta)$ при трех приведенных выше функциях потерь.

1. Для функции $Q_1(\theta,\hat{\theta})$ вида (5.28) имеем

$$J_1(\hat{\theta}) = \int_{\Omega_\theta} [c - \delta(\theta - \hat{\theta})]\, p(\theta \mid z)\, d\theta = c - p(\hat{\theta} \mid z).$$

В результате получаем

$$\arg \min_{\theta \in \Omega_\theta} J_1(\hat{\theta}) = \arg \max_{\theta \in \Omega_\theta} p(\theta \mid z),$$

т. е. требование минимума функции $J_1(\hat{\theta})$ сводится к максимизации функции $p(\theta|z)$. Таким образом, в случае использования функции потерь $Q_1(\theta,\hat{\theta})$ из критерия минимума среднего риска следует так называемый критерий максимума апостериорной вероятности.

2. Перейдем к рассмотрению функции потерь $Q_2(\theta,\hat{\theta}) = \left|\theta - \hat{\theta}\right|$. Пусть $\theta \in (-\infty,\infty)$, тогда

$$J_2(\hat{\theta}) = \int_{-\infty}^{\infty} \left|\theta - \hat{\theta}\right| p(\theta|z) d = \theta \int_{-\infty}^{\theta} (\theta - \hat{\theta}) p(\theta|z) d\theta - \int_{\theta}^{\infty} (\theta - \hat{\theta}) p(\theta|z) d\theta. \tag{5.31}$$

Так как необходимое условие минимума функции $J_2(\hat{\theta})$ имеет вид

$$\frac{dJ_2(\hat{\theta})}{d(\hat{\theta})} = 0,$$

то из (5.31) получаем

$$\int_{-\infty}^{\hat{\theta}} p(\theta \mid z)\, d\theta = \int_{\hat{\theta}}^{\infty} p(\theta \mid z)\, d\theta.$$

Таким образом, оптимальной оценкой параметра $\theta$ является медиана, т. е. $\hat{\theta} = Me(\hat{\theta})$.

Напомним, что медианой распределения случайной величины называется такое число Me $(\theta)$, для которого справедливы равенства

$$P(\hat{\theta} \le Me(\theta)) = P(\hat{\theta} \ge Me(\theta)) = 0.5.$$

3. Для квадратичной функции потерь выражение среднего риска принимает вид

$$J_3(\hat{\theta}) = \int_{\Omega_\theta} (\theta - \hat{\theta})^2 \; p(\theta \mid z)\, d\theta.$$

Из условия

$$\frac{dJ_3(\hat{\theta})}{d(\hat{\theta})} = 0$$

получаем

$$\int_{\Omega_\theta} (\theta - \hat{\theta}) \; p(\theta \mid z)\, d\,\theta = 0.$$

С учетом того, что $\int_{\Omega_\theta} p(\theta \mid z)\, d\theta = 1$

имеем

$$\hat{\theta} = \int_{\Omega_\theta} \theta \; p(\theta \mid z) d\theta. \qquad (5.32)$$

Из последнего соотношения следует, что баейсовская оценка параметра $\theta$ при квадратичной функции потерь совпадает с условным средним значением (математическим ожиданием) оцениваемого параметра при заданных выборочных значениях z, т. е. $\hat{\theta} = M_{\theta|z}\{\theta\}$.

Такая оценка является равномерно наилучшим решением (абсолютно оптимальным) для классов функции условной плотности распределения вероятности $p(\theta \mid z)$, для которых выполняются условия существования и непрерывности почти всюду.

Заметим, что если функция $p(\theta \mid z)$ унимодальна, т. е. имеет один максимум и симметрична относительно моды, то медиана и среднее значение случайной величины $\theta$ совпадают и равны ее моде. При этом оценки при функциях потерь $Q_2(\theta, \hat{\theta})$ и $Q_3(\theta, \hat{\theta})$ одинаковы и совпадают с наиболее вероятной оценкой.

Пример 5.1. Рассмотрим процедуру оценивания параметра $\theta$ линейного статического объекта

$$z = \theta u + \nu ,$$

где $u$ – детерминированный входной сигнал; $\theta$ и $\nu$ – независимые нормально распределенные случайные величины, причем

$$p(\theta) = \frac{1}{\sqrt{2\pi\sigma_\theta}} exp\left[ -\frac{(\theta - \hat{\theta})^2}{2\sigma_\theta^2} \right]; \qquad p(\nu) = \frac{1}{\sqrt{2\pi\sigma_\nu}} exp\left[ -\frac{\nu^2}{2\sigma_\nu^2} \right].$$

В процессе независимых измерений получены значения N-мерных векторов $z^T = [z_1,...,z_N]$ и $u^T = [u_1,...,u_N]$ при этом для условной плотности распределения $p(z|\theta)$ имеем

$$p(\theta) = \frac{1}{(\sqrt{2\pi\sigma_\nu})^N} exp\left[ -\frac{1}{2\sigma_\nu^2}(z - u\theta)^2 (z - u\theta) \right],$$

так как $\nu = z - u\,\theta, \quad \sigma_{z|\theta} = \sigma_\nu, \quad M_{z|\theta}\{z\} = u\,\theta.$

Получим оценку параметра $\theta$ с помощью формулы (5.32). Для этого надо определить функцию $p(\theta \mid z)$. Последнюю можно найти, если воспользоваться формулой Байеса (5.1) с предварительным нахождением выражения для плотности распределения *p(z)*. Этот путь является достаточно громоздким. Поэтому поступим следующим образом. Примем во внимание, что функция $p(\theta|z)$ есть плотность распределения параметра для любого *z*. Тогда *p(z)* будем учитывать постоянной величиной, необходимой для нормировки

$$\int_{-\infty}^{\infty} p(\theta \mid z)\, d\theta = 1. \tag{5.33}$$

При этом

$$p(\theta|z) = p(\theta)\, p(z - u\theta)\, p^{-1}(z) =$$

$$c\, exp\left\{ -\frac{1}{2}\left[ \frac{(\theta - \hat{\theta})^2}{\sigma_\theta^2} + \frac{(z - u\theta^T)(z - u\theta)}{\sigma_\nu^2} \right] \right\},$$

где *c* – нормировочная постоянная.

Произведем перегруппировку членов в показателе экспоненты так, чтобы получить выражение вида

$$p(\theta|z) = c\, exp\left\{ -\frac{1}{2\sigma^2}(\theta - d)^2 \right\}, \tag{5.34}$$

где

$$\sigma^2 = \frac{\sigma_\theta^2\,\sigma_\nu^2}{\sigma_\nu^2 + \sigma_\theta^2\,u^T u}; \qquad d = \frac{\hat{\theta}\,\sigma_\theta^2 + \sigma_\theta^2\,u^T z}{\sigma_\nu^2 + \sigma_\theta^2\,u^T u}.$$

Из (5.34) следует с учетом (5.33), что функция $p(\theta \mid z)$ есть плотность нормального распределения, а величина $d$ – условное среднее, являющееся искомой оценкой параметра $\theta$

$$\hat{\theta} = d = \frac{\theta\,\sigma_\nu^2 + \sigma_\theta^2 u^T z}{\sigma_\nu^2 + \sigma_\theta^2 u^T u}. \tag{5.35}$$

Для получения оценки $\hat{\theta}$ использованы как априорные данные $\hat{\theta}$, $\sigma_\theta$, $\sigma_\nu$ так и результаты измерений $z_i$, $u_i$, $i = \overline{1,N}$.

Так как для получения оценки (5.35) использовалась функция потерь $Q_3 = \left(\theta - \hat{\theta}\right)^2$, то обозначим эту оценку через $\hat{\theta}_3$. Несложно показать, что в рассматриваемой задаче $\hat{\theta}_1 = \hat{\theta}_2 = \hat{\theta}_3$, т. е. все эти оценки совпадают. Это объясняется тем, что нормальное распределение является унимодальным и симметричным.

Отметим два предельных свойства оценки (5.35). С одной стороны, если $\sigma_\theta^2 << \sigma_\nu^2 / \sum_{i=1}^{N} u_i^2$, то объем априорных сведений гораздо меньше объема данных, полученных в результате измерений, и оценка весьма близка к априорному среднему $\hat{\theta}$.

С другой стороны, если $\sigma_\theta^2 << \sigma_\nu^2 / \sum_{i=1}^{N} u_i^2$, то априорная информация незначительна и при идентификации используются в основном результаты измерений. В пределе оценки $\hat{\theta}$ определяется выражением

$$\hat{\theta} = \sum_{i=1}^{N} u_i z_i \,/ \sum_{i=1}^{N} u_i^2 \,.$$

## 5.3. Оценивание по методу максимального правдоподобия

Положим, что исходная (априорная) информация не содержит данных о плотности распределения $p(\theta)$ вектора параметров и функции потерь $Q\left(\theta, \hat{\theta}, z\right)$, т. е. предположения б) и в) отсутствуют. При этом можно положить $p(z, \theta) = const$ в области возможных значений параметра $\theta$. В этом случае для произвольного $z$ имеем

$$\max_\theta p(\theta \mid z) = \frac{const}{p(z)} \max_\theta p(z \mid \theta). \tag{5.36}$$

Оценка значения $\theta$, основанная на последовательности наблюдений z(N), является измеряемой функцией от *z(N)* и обозначается через $\hat{\theta}(N) = \hat{\theta}[N, z(N)]$. Требуется получить оценку истинного значения вектора параметров $\theta$ на основе последовательности наблюдений *z(N),* когда априори (до измерений) задана условная плотность распределения вероятности измеряемых входных и выходных переменных *z*, которую обозначим через

$$p(z \mid \theta) = p(z;\theta) = p[z(1),...,z(N);\theta].$$

Здесь $\theta$ уже не случайная величина, а неизвестный постоянный параметр; вместо вертикальной черты представляется точка с запятой. Функция $p(z;\theta)$, определяемая как функция параметров $\theta$ при фиксированном z, называется функцией правдоподобия. При этом наиболее правдоподобным считаются такие значения $\hat{\theta}$ оценок параметров $\theta$, которые доставляют максимум функций правдоподобия.




Рис. 5.3. Функции правдоподобия

На рис. 5.3 приведены несколько возможных функций правдоподобия, зависящих от одного параметра. Часто встречается наиболее желательный случай симметричной колоколообразной функции правдоподобия. В этом случае значение $\hat{\theta}$ параметра $\theta$, при котором функция правдоподобия $p(z;\theta)$ максимальна, несомненно, можно принять в качестве лучшей оценки $\theta$. Квадратный корень из дисперсии, задаваемого функцией правдоподобия распределения относительно $\hat{\theta}$, можно принять в качестве ошибки оценки. В случае асимметричного распределения $\hat{\theta}$ также является наилучшей оценкой, но дисперсия, хотя и не теряет своего математического смысла, служит неудовлетворительным параметром для описания поведения функции

правдоподобия в окрестности своего максимума. В этом случае желательно вместе с $\hat{\theta}$ приводить и само распределение $p(z;\theta)$.

Эти оценки максимального правдоподобия (ОМП) можно найти путем решения следующей системы алгебраических уравнений

$$\frac{\partial p(z;\theta)}{\partial \theta_j}\Big|_{\theta_j=\hat{\theta}_j} = 0, \quad j = \overline{1,N}. \tag{5.37}$$

В практических приложениях в силу свойства монотонности логарифмической функции максимизацию функции правдоподобия можно заменить максимизацией логарифмической функции правдоподобия:

$$L(z;\theta) = ln\, p(z;\theta).$$

Поскольку точка максимума функций $p(z;\theta)$ и $ln\, p(z;\theta)$ совпадают, то вместо системы (5.37) часто решается система уравнений

$$\frac{\partial L(z,\theta)}{\partial \theta_j}\Big|_{\theta_j=\hat{\theta}_j} = 0, \quad j = \overline{1,N}, \tag{5.38}$$

называемая уравнениями правдоподобия.

Пример 5.2. Найдем ОМП для среднего значения наблюдаемой дискретной последовательности $z = m_z + \nu$, основанную на N измерениях *z(l), .., z(N).* Здесь $m_z = M\{z(i)\}$ – среднее значение наблюдаемой и измеряемой дискретной стохастической последовательности. Ошибка измерений $\nu$ представляет собой последовательность $\nu(1),..., \nu(N)$ независимых случайных величин с гауссовским распределением с нулевым средним $M\{\nu(i)\} = 0$, дисперсией $M\{\nu^2(i)\} = \sigma_\nu^2$, когда некоррелированы и равноточны.

При выборе конечного объема $(\overline{1,N})$, уравнение относительно вектора ошибок запишем как

$$\nu = z - Em_z,$$

где $\nu^T = [\nu(1),...,\nu(N)]$, $z^T = [z(1),...,z(N)]$, $m_z = [m_z,...,m_z]$, $E$ – единичная матрица.

Плотность распределения вероятности вектора $\nu$ согласно уравнениям для случая независимых испытаний запишем в виде

$$p(\nu) = \prod_{i=1}^{N} p[\nu(i)] = \frac{1}{(2\pi)^{N/2}\sigma_\nu^N} exp\left[-\frac{\nu^T\nu}{2\sigma_\nu^2}\right] =$$

$$= \frac{1}{(2\pi)^{N/2}\sigma_\nu^2} exp\left[-\frac{1}{2\sigma_\nu^2}\sum_{i=1}^{N}(z(i)-m_z)^2\right] = p(z;m)$$

Логарифмическую функцию правдоподобия можно описать в виде

$$L\left(z;m_z\right)=ln\,p\left(z;m_z\right)=-N\,ln\sqrt{2\pi}-N\,ln\,\sigma_v-\frac{1}{2\sigma_v^2}\sum_{i=1}^{N}\left(z(i)-m_z\right)^2.$$

В этом случае уравнение правдоподобия имеет вид

$$\frac{dL\left(z;m_z\right)}{dm_z}\Big|_{m_z=\hat{m}_z}=-\frac{1}{\sigma_v^2}\sum_{i=1}^{N}\left(z(i)-\hat{m}_z\right)=0.$$

Из этого следует ОМП, определяемая как

$$\hat{m}_z=\frac{1}{N}\sum_{i=1}^{N}z(i), \tag{5.39}$$

которое соответствует эмпирическому среднему.

Аналогично можно получить оценку дисперсии $\hat{\sigma}_v^2$. Так частная производная $L(z;m_z;\sigma_v)$ по стандартному отклонению $\sigma_v$ в точке экстремума равна следующему выражению:

$$\frac{\partial L(z;m_z,\sigma_v)}{\partial\sigma_v}\Big|_{\sigma=\hat{\sigma}_v}=-\frac{N}{\hat{\sigma}_v}+\frac{1}{\hat{\sigma}_v^3}\sum_{i=1}^{N}(z(i)-\hat{m}_z)^2=0,$$

из которого следует, что

$$\hat{\sigma}^2_v=\frac{1}{N}\sum_{i=1}^{N}\left(z(i)-\hat{m}_z\right)^2.$$

Пример 5.3. Рассмотрим случай оценивания среднего значения при повторных измерениях. Измерения $x(j)$ в общем случае имеют различную точность: например, они могут быть получены с помощью разной аппаратуры или различных методов. Предположим, что ошибки измерений распределены нормально. Следовательно, отдельное измерение соответствует выборке единичного объема из нормального распределения со средним $m_x$ и дисперсией $\sigma_j^2$. Тогда плотность распределения вероятности значения $x(j)$ равна

$$p(x(j);m_x)=\frac{1}{\sqrt{2\pi\sigma_j}}exp\left(-\frac{(x(j)-m_x)}{2\sigma_j^2}\right).$$

Если произведены N измерений, то соответствующая функция правдоподобия равна

$$p(x(j);m_x)=\prod_{j=1}^{N}\frac{1}{\sqrt{2\pi\sigma_j}}exp\left(-\frac{(x(j)-m_x)}{2\sigma_j^2}\right)$$

и ее логарифм

$$L(x; m_x) = -\frac{1}{2}\sum_{j=1}^{N}\frac{(x(j) - m_x)}{2\sigma_j^2} + const.$$

Уравнение правдоподобия принимает вид

$$\frac{dL}{dm_x}|_{m_x=\hat{m}_x} = \sum_{j=1}^{N}\frac{x(j) - m_x}{\sigma_j^2} = 0.$$

Его решение:

$$\hat{m}_x = \sum_{j=1}^{N}\frac{x(j)}{\sigma_j^2} / \sum_{j=1}^{N}\frac{1}{\sigma_j^2}.$$

Результат, соответствующий максимальному правдоподобию, есть среднее от отдельных измерений, взятых с весами, обратными дисперсиями.

Пример 5.4. Найдем ОМП для вектора неизвестных параметров регрессионного процесса, основанную на серии N экспериментов, которое можно записать как (5.8)

$$x = U\theta + \nu,$$

где x – N-мерный вектор измерений выходных значений; $\theta$ –неизвестный n-мерный вектор; $\nu$ – случайный N-мерный вектор имеет нормальное распределение; U – матрица размера (N x n), элементами которой являются измеряемые входные переменные. Уравнение относительно ошибок запишем в виде

$$\nu = x - U\theta.$$

Для случая, когда шум $\nu$ имеет нормальное распределение с нулевым средним и матрицей ковариации $K_\nu$, функцию распределения запишем в следующем виде:

$$p(\nu) = (x - U\theta) = c\,exp\left[-\frac{1}{2}(x - U\theta)^T K_\nu^{-1}(x - U\theta)\right], \qquad (5.40)$$

где $K_\nu = M\{\nu^T\nu\}$ – ковариационная матрица; $c = \left[(2\pi)^N |K_\nu|\right]^{-\frac{1}{2}}$.

Логарифмическая функция правдоподобия имеет вид

$$L(x;\theta) = \ln c - \frac{1}{2}(x - U\theta)^T K_\nu^{-1}(x - U\theta). \qquad (5.41)$$

Максимальное значение логарифмической функции правдоподобия определяется из соотношения

$$\frac{\partial L(z;,\theta)}{\partial\theta}\Big|_{\sigma=\hat{\sigma}} = -U^T K_\nu^{-1} x + U^T K_\nu^{-1} U\hat{\theta} = 0. \qquad (5.42)$$

Из (5.42) следует ОМП, определяемая как

$$\hat{\theta} = (U^T K_v^{-1} U)^{-1} U^T K_v^{-1} x. \qquad (5.43)$$

Для случая, когда $K_v^{-1} = \sigma_v^2 E$ (5.43), то оценки параметров определяются согласно следующему уравнению:

$$\hat{\theta} = (U^T U)^{-1} U^T x. \qquad (5.44)$$

Как следует из (5.44) оценки максимального правдоподобия в случае нормального распределения возмущений обладают очевидными свойствами линейности относительно измерений и несмещенности.

## 5.4. Оценивание по методу наименьших квадратов

Метод наименьших квадратов (МНК) является одним из универсальных методов в математике и прикладном анализе. Широкое применение процедуры оценивания МНК получили и в задачах идентификации для целей аппроксимации различных функциональных зависимостей и сглаживания влияния помех и случайных ошибок измерения в задачах обработки результатов наблюдений.

Построение математической модели по результатам измерений входных и выходных сигналов можно рассматривать как аппроксимацию этих результатов определенной функциональной зависимостью (моделью). Так как результаты измерений содержат случайные ошибки, то имеет место проблема сглаживания.

Схему оценивания по методу наименьших квадратов можно проиллюстрировать в соответствии с рис. 5.1. Положим, что модель объекта является линейной по параметрам

$$\hat{x}(t) = \sum_{i=1}^{n} \hat{\theta}_i \, u_i(t) = u^T(t)\hat{\theta}, \qquad (5.45)$$

где $u_i(t)$ – измеряемые без ошибок входные последовательности объекта и модели; $\hat{\theta}_i$ – оценки неизвестных параметров. Для серии из N экспериментов, где $N > n$, выход модели (5.45) представим в виде

$$\hat{x} = \begin{bmatrix} \hat{x}(1) \\ \hat{x}(2) \\ \dots \\ \hat{x}(N) \end{bmatrix} = U\hat{\theta} = \begin{bmatrix} u_1(1) \dots u_n(1) \\ u_1(2) \dots u_n(2) \\ \dots \quad \dots \quad \dots \\ u_1(N) \dots u_n(N) \end{bmatrix} \begin{bmatrix} \hat{\theta}_1 \\ \hat{\theta}_2 \\ \dots \\ \hat{\theta}_n \end{bmatrix} \qquad (5.46)$$

Тогда вектор ошибки $e^T = [e(1),...,e(N)]$ между выходными сигналами объекта и модели согласно выражениям (5.8), (5.2) можно определить как

$$e = x - \hat{x} = U\theta + \nu - U\hat{\theta}.$$

Согласно принципу МНК искомые оценки $\hat{\theta}$ минимизируют сумму квадратов отклонений $\hat{x}(t) - x(t)$, т. е.

$$Q(\hat{\theta}) = e^T e = (x - U\hat{\theta})^T (x - U\hat{\theta}) \qquad (5.47)$$

Для получения оценок $\hat{\theta}$ по МНК нужно продифференцировать величину $Q(\hat{\theta})$ по $\hat{\theta}$ и из условия равенства производной нулю можно найти значения $\hat{\theta}$.

В соответствии с МНК необходимое условие минимума функции (5.47) определяется решением векторного уравнения

$$\frac{\partial Q(\hat{\theta})}{\partial \hat{\theta}} = \frac{\partial}{\partial \hat{\theta}}\left[\left(x - U\hat{\theta}\right)^T \left(x - U\hat{\theta}\right)\right] = \frac{\partial}{\partial \hat{\theta}}\left(x^T x - 2\hat{\theta}^T x + \theta^T U\hat{\theta}\right) = 2\left(U^T x - U^T U\hat{\theta}\right) = 0,$$

из которого следует справедливость следующего равенства $U^T U\hat{\theta} = U^T x.$

Если матрица $U^T U$ невырожденная, то оценка $\hat{\theta}$ вычисляется по формуле

$$\hat{\theta} = \left(U^T U\right)^{-1} U^T x. \qquad (5.48)$$

Оценка МНК (5.48) является линейной относительно вектора результатов измерений. Качество этой оценки можно характеризовать ее вероятностными, в частности математическим ожиданием, дисперсией и корреляцией между компонентами вектора $\hat{\theta}$. Представим оценку МНК (5.48) в виде

$$\hat{\theta} = \left(U^T U\right)^{-1} U^T \left(U\theta + \nu\right) = \theta + \left(U^T U\right)^{-1} U^T \nu. \qquad (5.49)$$

Определим математическое ожидание $M\{\hat{\theta}\}$. Если входной и выходной сигналы статистически независимы с помехами измерений $\nu$, то

$$M\{\hat{\theta}\} = \theta + M\{\left(U^T U\right)^{-1} U^T\} M\{\nu\} = \theta, \qquad (5.50)$$

т. к. предполагалось, что $M\{\nu\} = 0$.

Тогда условие (5.50) означает, что оценка $\hat{\theta}$ является несмещенной. В противном случае оценка $\hat{\theta}$ будет смещенной.

Значения дисперсий и корреляций между компонентами вектора оценки $\hat{\theta}$ можно определить с помощью матрицы ковариационных ошибок

$$K_{\hat{\theta}} = M\left\{(\theta - \hat{\theta})(\theta - \hat{\theta})^T\right\} = M\left\{rr^T\right\}, \qquad (5.51)$$

где $r = \theta - \hat{\theta}$ – вектор ошибки. В развернутой форме

$$K_{\hat{\theta}} = \begin{bmatrix} M\{r_1 r_2\}...M\{r_1 r_n\} \\ M\{r_2 r_1\}...M\{r_2 r_n\} \\ .....\\ M\{r_n r_1\}...M\{r_n r_n\} \end{bmatrix},$$

где $r_j = \theta_j - \hat{\theta}_j,\ \ j = \overline{1,n}.$

С учетом выражения (5.50) ковариационную матрицу (5.51) несмещенной оценки МНК можно записать в виде

$$\begin{aligned} K_{\hat{\theta}} &= M\{(QU\theta + Q\nu - \theta)(QU\theta + Q\nu - \theta)\} = \\ &= M\{(Q\nu)(Q\nu)^T\} = (U^T U)^{-1} U^T K_\nu U (U^T U)^{-1}, \end{aligned} \tag{5.52}$$

где было введено обозначение $Q = \left(U^T U\right)^{-1} U^T$.

Ковариационная матрица вектора оценок, вычисляемая для случая (5.11) относительно свойств помехи, с учетом (5.8) определяется выражением

$$K_{\hat{\theta}} = \sigma_\nu^2 (U^T U)^{-1}. \tag{5.53}$$

Если дисперсия $\sigma_\nu^2$ априори неизвестна, то она может быть оценена по формуле

$$\hat{\sigma}_\nu^2 = \frac{1}{n-1}\left(x - U\hat{\theta}\right)^t \left(x - U\hat{\theta}\right), \tag{5.54}$$

являющейся остаточной дисперсией эмпирического уравнения регрессии.

Покажем теперь, что процедуры оценивания по МНК для второго и третьего случаев моделей наблюдения (эти случаи определяются ковариационными матрицами (5.12) и (5.13)) путем специального линейного преобразования сводятся к первому случаю – модели (5.11).

Действительно, поскольку ковариационная матрица $K_\nu$ является симметричной положительно - определенной, то в соответствии с [8, 10] имеет место представление $K_\nu = \left(K_\nu^{\frac{1}{2}}\right)^T K_\nu^{\frac{1}{2}}$.

Произведем замену переменных по формуле

$$z = K_\nu^{\frac{1}{2}} x. \tag{5.55}$$

Тогда отношение (5.8) с учетом (5.11) можно представить в виде

$$z = H\theta + \varepsilon, \tag{5.56}$$

где
$$H = K_v^{\frac{1}{2}} U, \qquad \varepsilon = K_v^{\frac{1}{2}} v \tag{5.57}$$

Применим к уравнению (5.56) МНК. Для этого составим минимизируемую функцию

$$Q(\hat{\theta}) = \varepsilon^T \varepsilon. \tag{5.58}$$

В этом случае МНК оценки смогут быть представлены в виде

$$\hat{\theta} = (H^T H)^{-1} H^T z$$

или с учетом (5.57) и (5.55)

$$\hat{\theta} = (U^T K_v^{-1} U)^{-1} U K_v^{-1} x. \tag{5.59}$$

Очевидно, что для ковариационной матрицы (5.12)

$$\hat{\theta} = (U^T W U)^{-1} U W x, \tag{5.60}$$

где $W = diag\left(\frac{1}{\sigma_v^2},...,\frac{1}{\sigma_{vN}^2}\right)$, а для ковариационной матрицы (5.13)

$$\hat{\theta} = \left(U^T R^{-1} U\right) U^T R^{-1} x \tag{5.61}$$

Заметим, что минимизируемая функция (5.58) с учетом замены (5.56), (5.57) может быть представлена в виде (5.47). При этом ее использование, как и указанное преобразование (5.55) возможно только тогда, когда известна или может быть оценена ковариационная матрица $K_v$.

# ГЛАВА 6. РЕКУРЕНТНЫЕ МЕТОДЫ ИДЕНТИФИКАЦИИ

## 6.1. Основные понятия. Алгоритмы градиентного метода

Этап структурной идентификации позволяет сформировать векторы координат состояния, координат входных воздействий и выходных реакций, а также раскрыть технологию внутренних связей и выбрать структуру операторов связей. На следующем этапе стоит задача оценивания параметров этих операторов, т. е. задача параметрической идентификации.

При этом предполагают, что последовательность наблюдений $\{z(t)\} = \{u(t), x(t)\}$ входных $u(t)$ и выходных $x(t)$ сигналов порождается системой, параметризованной конечномерным вектором параметров $\theta$, который принадлежит компактному подмножеству $C_\theta$ из $R^n$.

В общем случае процесс параметрической идентификации можно представить в виде оптимизационной процедуры приближения модели известной топологии и структуры к объекту идентификации (технологическому процессу) в смысле выбранного критерия $J(\theta)$ по вектору параметров $\theta$.

Оптимизация модели по вектору параметров $\hat{\theta}$ может осуществляться как в ретроспективе, т. е. по результатам наблюдений временных процессов в координатах технологического процесса, так и в реальном масштабе времени, т. е. в ходе развития этих процессов и по мере поступления текущей информации о них. В последнем случае используются так называемые итеративные методы параметрической идентификации.

Итеративные методы идентификации, получившие название методов идентификации по настраиваемым моделям, могут обеспечить точное нахождение оценок параметров после бесконечного числа шагов или циклов оценивания. При этом приближенное решение может быть получено как результат промежуточных вычислений. Доступность результатов промежуточных вычислений обеспечивает в случае нормального функционирования объекта отслеживание его изменяющихся параметров.

Таким образом, многие задачи идентификации с настраиваемыми моделями могут быть сведены к оптимизационной задаче

$$J(\hat{\theta}) = M\{Q[e(t),\hat{\theta}]\} \to \min_{\hat{\theta} \in C_\theta}, \qquad (6.1)$$

где Q (·) – заданная функция потерь; $\hat{\theta}$ – n-мерный вектор оценок параметров; $C_\theta$ –- область допустимых значений параметров $C_\theta \in R^n$.

Минимизация критерия качества может осуществляться соответствующим выбором структуры настраиваемой модели и целенаправленным изменением ее параметров алгоритмами идентификации через блок параметрической настройки по наблюдениям входного воздействия, выходных величин объекта и модели.

При таком подходе близость структур (операторов) объекта и настраиваемой модели оценивается по степени близости их реакций на одно и то же воздействие *u(t)*. Рассматриваемый метод определения близости структур объекта и настраиваемой модели приводит к получению формального математического описания объекта, устанавливающего связь между входными и выходными переменными. В общем случае структуры объекта и настраиваемой модели могут быть различны.

Если структура объекта задана с точностью до вектора неизвестных параметров, то ставится задача оценивания параметров и близость структур объекта и модели понимается, в смысле близости истинных значений параметров $\theta$ к их оценкам $\hat{\theta}$.

Модель может быть представлена уравнением

$$\hat{x}(t) = f(t,u,x,\hat{\theta}) = f(t,z,\hat{\theta}), \qquad (6.2)$$

которое по наблюдениям *z* позволяет вычислить значение $\hat{x}(t)$, определяющее прогноз *x(t)* в момент времени *t*. Тогда задача идентификации заключается в оценке вектора $\theta$ таким образом, чтобы по наблюдениям за входными и выходными величинами *z(t)* достичь определенного соответствия модели $M(\hat{\theta})$ системе S, которое оценивается критерием качества $J(\hat{\theta})$.

Решение данной задачи можно получить с использованием различных известных методов идентификации. При этом предполагаются такие устойчивые итеративные схемы методов идентификации, чтобы существовала сходимость оценки $\hat{\theta}(t)$ к θ при $t \to \infty$. Общий вид итеративных алгоритмов для разных задач можно представить как

$$\hat{\theta}(t) = \hat{\theta}(t-1) - y(t)p(t-1). \qquad (6.3)$$

Согласно схеме данных алгоритмов, новый вектор параметров оценивается с учетом вектора параметров $\hat{\theta}$(t-1) на предыдущем шаге

и дополнительного члена $\gamma p(t\text{-}1)$, где $\gamma(t)$ скалярный коэффициент усиления, определяющий величину шага в направлении вектора $p(t\text{-}1)$, который можно определить.

В зависимости от априорной информации об исследуемом объекте вектор $p(t)$ может быть представлен как

$$p(t) = \frac{\partial J[\hat{\theta}(t)]}{\partial \hat{\theta}(t)} = \nabla J[\hat{\theta}(t)] \tag{6.4}$$

или

$$p(t) = \frac{\partial J[\varepsilon(t), \hat{\theta}(t)]}{\partial \hat{\theta}(t)} = \nabla Q[\varepsilon(t), \hat{\theta}(t)]. \tag{6.5}$$

Тогда итеративная схема алгоритма (6.3) обобщает широкий класс алгоритмов для регулярных итеративных методов (6.4) и вероятностных итеративных (рекуррентных) методов (6.5) оценки параметров.

К вероятностным рекуррентным алгоритмам относятся и алгоритмы метода стохастической аппроксимации, которые получили широкое распространение для идентификации объектов при наличии помех измерения входных и выходных переменных. Первые исследования в области стохастической аппроксимации были проведены для целей отыскания корня уравнения регрессии и получили дальнейшее развитие для решения широкого класса задач.

В общем случае вместо скаляра $\gamma(t)$ в (6.3) вводят матрицу коэффициентов усиления $\Gamma(t)$ для того, чтобы сделать различными величины шагов по различным составляющим вектора параметров $\hat{\theta}$. Соответствующим выбором коэффициента $\gamma(t)$ и матрицы $\Gamma(t)$ можно из алгоритма (6.3) получить известные алгоритмы Ньютона, наискорейшего спуска и т. д. [16, 28, 29].

При этом, например [28], для широкого класса непрерывных и дискретных алгоритмов адаптации возможны следующие представления матрицы Г(t):

- скалярная матрица усиления

$$\tilde{A}(t) = \begin{bmatrix} \gamma(t) & 0 & \dots & 0 \\ 0 & \gamma(t) & \dots & 0 \\ . & . & \dots & . \\ 0 & 0 & \dots & \gamma(t) \end{bmatrix} = \gamma(t)E, \qquad \gamma(t) > 0,$$

- диагональная матрица

$$\tilde{A}(t)=\begin{bmatrix} \gamma_1(t) & 0 & \dots & 0 \\ 0 & \gamma_2(t) & \dots & 0 \\ \cdot & \cdot & \dots & \cdot \\ 0 & 0 & \dots & \gamma_n(t) \end{bmatrix}, \qquad \gamma_i(t)>0 \quad i=\overline{1,n},$$

- матрица вида $\tilde{A}(t)=\gamma(t)\hat{A},$

где B > 0 – положительно определенная матрица, а $\gamma(t)=1/t.$

Тогда наиболее простой алгоритм адаптации – в данном примере это вычисление текущего значения вектора $\hat{\theta}$ – основан на методе градиентного спуска. Непрерывный алгоритм градиентного спуска записывается в виде

$$\hat{\theta}(t)=-\tilde{A}(t)\nabla_{\hat{\theta}}Q_t(z(t),\hat{\theta}(t)); \qquad (6.6)$$

здесь $\hat{\theta}(0)=\hat{\theta}_0$ задан; $\Gamma=\Gamma^T>0$ – положительно определенная числовая матрица, от которой зависят «коэффициенты усиления» вычислительной процедуры (6.1),

$$\nabla_{\hat{\theta}}Q(\cdot)=col\left(\frac{\partial Q}{\partial\hat{\theta}_1},.....,\frac{\partial Q}{\partial\hat{\theta}_n}\right)$$

- градиент Q по аргументу $\hat{\theta}$.

Для дискретного времени $t=t_k=k\cdot\Delta t,\Delta t=const,$ градиентный алгоритм минимизации функции $Q_1$ по $\hat{\theta}$ записывается так:

$$\hat{\theta}_{k+1}=\hat{\theta}_k-\tilde{A}_{k+1}\nabla_{\hat{b}}Q(z_{k+1,}\hat{\theta}_{t+1}). \qquad (6.7)$$

На рис. 6.1 приведены структурные схемы – эквиваленты градиентных алгоритмов (6.1) и (6.2). Эти схемы соответствуют в общем виде непрерывным (рис. 6.1, *а*) или дискретным (рис. 6.1, *б*) динамическим системам с обратной связью. Схемы включают нелинейный преобразователь $\nabla_{\hat{\theta}}Q(\cdot),$ усилитель с переменным, вообще говоря, коэффициентом усиления Г(t) или $\Gamma_{k+1}$ («матричный» усилитель) либо $\gamma(t)$ или $\gamma_k^1$ («скалярный» усилитель) и интегратор (сумматор). Широкими стрелками на рис. 6.1 обозначены векторные связи. Особенностью схем является их неавтономность.

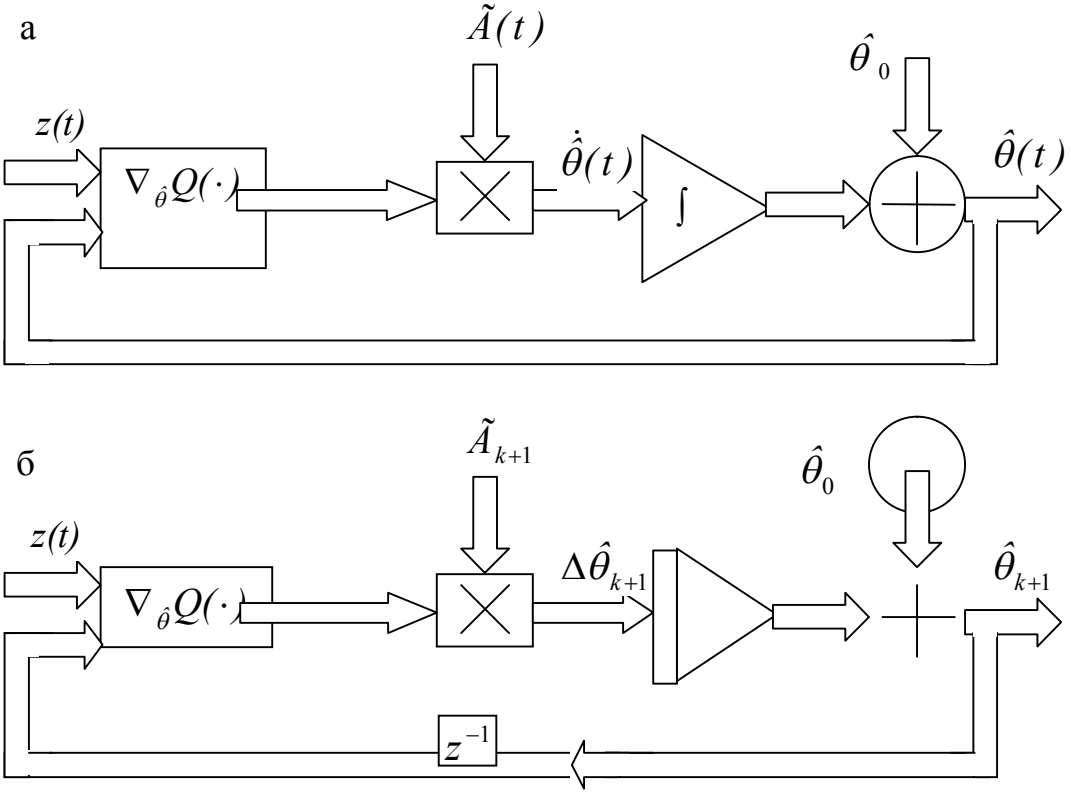



Рис. 6.1. Структурные схемы непрерывного (а) и дискретного (б) градиентных алгоритмов адаптации

Для алгоритмов адаптации необходима обновляемая текущая информация в виде измерений z(t) или $z_{k+1}$, которая и поступает в «блок» нелинейных (в общем случае) преобразований $grad_{\hat{\theta}} Q(\cdot)$. Эта информация предназначена для компенсации недостаточной текущей информации.

Матрица коэффициентов $\Gamma>0$ выбирается в зависимости от условий функционирования основного контура адаптивной системы, в т. ч. в зависимости от следующих компонентов:

а) детерминированная или стохастическая природа измеряемых переменных в составе вектора измерений z(t) для вычислений, поэтому алгоритмы градиентного спуска могут быть детерминированными или стохастическими;

б) учитывается или не учитывается динамика обобщенного настраиваемого объекта в алгоритме по схеме градиентного метода.

## 6.2. Детерминированные алгоритмы

Алгоритм наискорейшего градиентного спуска. В алгоритмах градиентного метода коррекция вычисляемых параметров осуществляется в направлении, противоположном изменению оценочной функции (в задачах оценивания) или целевой функции (в задачах настройки параметров регулятора в основном контуре). Критерий оценки принимается в виде локального квадратичного или интегрального функционала, аргумент которого в содержательном смысле адекватен решаемой задаче. Пусть, например, в задаче оценивания $\hat{\theta}_к$ – значение оценки истинного и неизвестного вектора $\theta^*$ на *к*–м шаге вычислений. Тогда дискретный алгоритм градиентного метода имеет вид

$$\hat{\theta}_{k+1} = \hat{\theta}_k - \tilde{A}_{k+1} \nabla_\theta Q(\hat{\theta}_{k+1}, z_{k+1}), \qquad (6.8)$$

здесь – оценка $\hat{\theta}_0$ задана, $\Gamma_{к+1}=\Gamma^T_{к+1}>0$ – положительно определенная матрица; $Q(.)$ – оценочная функция.

Суть алгоритма, напомним, сводится к следующей последовательности вычислений: задается начальное значение $\hat{\theta}_0$; в этой «точке» вычисляется $\nabla_\theta Q(\hat{\theta}_0)$ и при выбранной матрице $\Gamma_1$, например $\Gamma_1=\gamma_1 \cdot I$ вычисляется (корректируется) новое значение оценки $\hat{\theta}_1$. Далее все повторяется до тех пор, пока на некотором шаге $к=к^*$ при поиске $min_{\hat{\theta}} Q$ не получим, что $Q(\hat{\theta}_{\hat{e}+1}) > Q(\hat{\theta}_{\hat{e}})$. Известные особенности дискретного градиентного алгоритма – чувствительность к величине шага $\gamma$ и малая скорость сходимости.

Работоспособность алгоритмов, использующих базовый градиентный метод вычислений, зависит как от выбранной целевой функции (степени ее адекватности задаче вычислений, информативности ее относительно вычисляемых параметров), так и от выбора матрицы $\Gamma$. В некоторых случаях $\Gamma = \gamma_1 \cdot I$, где $\gamma > 0$ – постоянное число, $I$ – единичная матрица. В других случаях $\Gamma_к$ – матрица переменных во времени коэффициентов, например $\Gamma_к=\gamma_{\hat{e}} \cdot I$, $\gamma_{\hat{e}} > 0$ для всех $к=0,1,...$ В любом случае $\Gamma_к$ или $\gamma_{\hat{e}}$ нужно выбрать так, чтобы вычислительный алгоритм сходился и тем самым достигалась цель адаптивного управления.

Наиболее популярной модификацией градиентного алгоритма является алгоритм, основанный на методе наискорейшего градиентного спуска. В методе градиентного спуска выбор матрицы $\Gamma$ или скаляра $\gamma$ остается открытым. Матрица Γ или скаляр $\gamma$ существенно

влияют на динамику вычисления. Если их выбор осуществлять на каждом *k-м* шаге из условия

$$\min_{\tilde{A},\gamma} Q(\hat{\theta}_{\hat{e}+1}, z_{x-1}) = \min_{\tilde{A},\gamma} Q(\hat{\theta}_{\hat{e}-1} - \tilde{A}_{\hat{e}} \nabla_{\theta} Q(\hat{\theta}_{\hat{e}-1}, z_{\hat{e}}), z_{\hat{e}+1},),$$

то это и будет тем дополнением к алгоритму градиентного спуска, которое делает его алгоритмом наискорейшего спуска. Таким образом, этот алгоритм опять состоит в последовательности минимизации функций одной переменной: $\Gamma$ или $\gamma$.

С целью упрощения необходимых математических преобразований примем матрицу усиления $\Gamma_{к+1}$ диагональной: $\Gamma_{к+1}=diag(\gamma_{\hat{e}+1}^{i,i})$, $i=\overline{1,N}$; $k=0,1,2,....;$ начальные значения элементов $\gamma_0^{i,i}$ задаются. Выбор $\gamma_{\hat{e}+1}^{i,i}$ на каждый *k-й* итерации производим из условия

$$Q'(\hat{\theta}_{\hat{e}}, z_{\hat{e}+1}) = (\partial\hat{\theta}_{\hat{e}} / \partial\gamma_{\hat{e}}^{i,i}) \nabla_{\theta} Q(\hat{\theta}_{\hat{e}-1}, z_{\hat{e}+1}) = 0, i = \overline{1,N}. \qquad (6.9)$$

Здесь $\partial\hat{\theta}_{\hat{e}} / \partial\gamma_{\hat{e}}^{i,i}$ – диагональная матрица производных размерностью *NxN*; вектор производных $Q_{\gamma}'$ размерностью *Nxl* с элементами $\partial Q'(\hat{\theta}_{\hat{e}}, z_{\hat{e}+1}) / \partial\gamma_{\hat{e}}^{i,i})$.

Так как на промежуточных итерациях в силу алгоритма (6.8) производные оценки $\hat{\theta}_{\hat{e}}$ по элементам матрицы $\Gamma_{к-}$ ненулевые, то условие (6.9) может быть выполнено, если только

$$\nabla_{\theta} Q(\hat{\theta}_{\hat{e}}, z_{\hat{e}+1}) = \nabla_{\theta} Q(\hat{\theta}_{\hat{e}-1} - \tilde{A}_{\hat{e}}^{îïò} \nabla_{\theta} Q(\hat{\theta}_{\hat{e}-1}, z_{\hat{e}}), z_{\hat{e}+1}) = 0.$$

При числе итераций к>>1 для сходящегося процесса вычислений $\|\delta_{\hat{e}+1}\| = \|\theta_{\hat{e}+1} - \theta_{\hat{e}}\| << \|\hat{\theta}_{\hat{e}}\|$. Тогда разложение функции $\nabla_{\theta}$Q(.) в ряд Тейлора в окрестности вектора $\hat{\theta}_{\hat{e}}$ с радиусом разложения $\|\delta_{\hat{e}+1}\|$ и с удержанием линейной по $\|\delta_{\hat{e}+1}\|$ составляющей ряда приводит к приближенному равенству

$$\nabla_{\hat{\theta}} Q(\hat{\theta}_{\hat{e}}, z_{\hat{e}+1}) + \nabla_{\hat{\theta}}^2 Q(\hat{\theta}_{\hat{e}-1}, z_{\hat{e}+1}) \delta_{\hat{e}+1} \approx 0, \qquad (6.10)$$

откуда с учетом того, $\delta = -\tilde{A}\nabla_{\theta} Q$ следует квазиоптимальное (в силу приближения (6.10)) значение матрицы $\tilde{A}_{\hat{e}}^{îïò} = [\nabla_{\hat{\theta}}^2 Q(\cdot)]^{-1}$.

Положительно определенная матрица вторых производных $\nabla_{\hat{\theta}}^2 Q(\cdot)$ в (6.10) называется *матрицей Гессе* и вычисляется по формуле

$$\nabla^2_{\hat\theta} Q(\cdot) = \nabla_\theta \nabla_{\theta\grave{o}} Q(\cdot) = \frac{\partial}{\partial\hat\theta}\left(\frac{\partial Q}{\partial\hat\theta}\right)^T = \begin{bmatrix} \dfrac{\partial^2 Q}{\partial\hat\theta_1\partial\hat\theta_1},\ldots\dfrac{\partial^2 Q}{\partial\hat\theta_1\partial\hat\theta_n} \\ \ldots\ldots\ldots\ldots\ldots\ldots \\ \dfrac{\partial^2 Q}{\partial\hat\theta_n\partial\hat\theta_1},\ldots,\dfrac{\partial^2 Q}{\partial\hat\theta_n\partial\hat\theta_n} \end{bmatrix}. \qquad (6.11)$$

Из (6.10) при сопоставлении с (6.8) следует, что диагональная матрица $\Gamma_{K+1}$ в алгоритме градиентного спуска должна быть замена на обратную матрицу вторых производных оценочной функции по параметрам

$$\tilde A^{\hat e\hat a.\hat i\hat i\grave o}_{\hat e+1}\ (\nabla^2_{\hat\theta} Q(z_{\hat e+1},\hat\theta_{\hat e}))^{-1}. \qquad (6.12)$$

Анализ вычислительных особенностей алгоритма наискорейшего спуска содержится, например, в [5].

Пример. Пусть оценочная функция выбрана квадратичной:

$$Q(z_{\hat e+1},\hat\theta_{\hat e}) = 0{,}5\sigma^2_{\hat e+1} = 0{,}5(\tilde o_{\hat e+1} - \acute o_{\hat e+1})^2 = 0{,}5[(\hat\theta_{\hat e} - \theta^*)^{\grave O} z_{\hat e+1}]^2,$$

здесь $\hat\theta_{\hat e}$ – оцениваемой вектор параметров; $\theta^*$ – истинный вектор оцениваемых параметров; $z_{\kappa+1}$ – используемый вектор измерений.

Градиентный алгоритм (6.8) при заданном начальном значении $\hat\theta_0 = \hat\theta(0)$ оценки $\hat\theta$ приобретает следующий вид:

$$\hat\theta_{\hat e+1} = \hat\theta_{\hat e} - \tilde A_{\hat e+1}\sigma_{\hat e+1} z_{\hat e+1}.$$

Матрица вторых производных $\nabla^2_{\hat\theta} Q(\cdot)$ согласно (6.11) вычисляется как

$$\nabla^2_{\hat\theta} Q(\cdot) = \frac{\partial}{\partial\hat\theta}\left(\frac{\partial Q}{\partial\hat\theta}\right)^T = \frac{\partial(\hat\theta - \theta^*)^T z_{\hat e+1} z^{\grave O}_{\hat e+1}}{\partial\hat\theta} = z_{\hat e+1} z^{O}_{\hat e+1},$$

где векторы $\hat\theta$ и Z имеют nxl. В случаях квадратичных оценочных функций матрица $\nabla^2_{\hat\theta} Q(\cdot)$ не зависит от оцениваемых параметров.

Алгоритм наискорейшего спуска в рассматриваемом примере приобретает следующий вид:

$$\hat\theta_{\hat e+1} = \hat\theta_{\hat e} - (z_{\hat e+1} z^{\grave O}_{\hat e+1})^{-1} z_{\hat e+1}\sigma_{\hat e+1}).$$

Математические исследования метода указывают на его сходимость при произвольных начальных условиях и уменьшение $Q(\hat\theta, z)$ на каждой итерации. Происходит заведомо меньше, чем у градиентного процесса с постоянным шагом $\gamma$.

**Вычислительные алгоритмы, использующие метод Ньютона.** Подобно методу итераций метод Ньютона и его модификации явля-

ются общими и применимыми к решению очень широкого класса нелинейных операторных уравнений *F(x)=0*, к которым приводятся оптимизационные задачи при адаптивном управлении. Значение метода Ньютона состоит в том, что процесс решения нелинейных уравнений сводится к решению последовательности линейных задач. Это достигается выделением из исходного нелинейного уравнения его главной линейной части, если это удается сделать.

Идея метода Ньютона проста. Пусть *F(x)* – векторная функция векторной переменной *x* и *F(x)* имеет конечные производные не менее второго порядка. Пусть x* – точное решение уравнения *F(x)=0,* к таким уравнениям приводятся, например, задачи отыскания $min_{\theta}Q(\hat{\theta})$. Предложим, что $x^{*}=x_0+\varepsilon$, где $\varepsilon$ – малая векторная величина. Тогда, раскладывая F(x) по степеням $\varepsilon$ в ряд Тейлора, получаем, что $F(x^*) = F(x_0) + \nabla_x F(x_0)\varepsilon + p(\varepsilon) = 0$, где отбрасываем далее остаток $p(\varepsilon)$. Получаем линейное уравнение для $\varepsilon$: $F(x_0) + \nabla_x F(x_0)\varepsilon \approx 0$. Решая его относительно $\varepsilon$, мы найдем лишь приближенное значение $\varepsilon = \varepsilon_0$ для исходного нелинейного уравнения $F(x_0 + \varepsilon) = 0$. Добавим его к $x_0$, получим приближенное число или улучшенное по сравнению с начальным приближением $x_0$ решение

$$x_1 = x_0 + \varepsilon_0 = x_0 - [\nabla_x F(x_0)]^{-1} F(x_0). \qquad (6.13)$$

Пользуясь теми же соображениями, можно получить $x_1$ и т. д. В итоге следует правило, принадлежащее скалярной форме Ньютона:

$$x_{k+1} = x_k - [\nabla_x F(x_k)]^{-1} F(x_k),\ \text{k=0,1,2,}\ldots\ldots\ldots \qquad (6.14)$$

Из (6.14) следуют необходимые условия применения метода Ньютона: удачный выбор начального приложения $x_0$ к истинному, но неизвестному решению *x*;* обращение матрицы $\nabla_x F(x)$ на каждой итерации.

В адаптивном управлении векторным функциям F(x) соответствуют градиенты оценочных или целевых функционалов $\nabla_{\hat{\theta}} Q(\cdot)$ (или $\nabla^2_{\theta} J$) по векторному аргументу $\hat{\theta}$, так что градиенту функционалов $\nabla_{\hat{\theta}} Q(\cdot)$ (или $\nabla^2_{\hat{\theta}} J$) будут соответствовать матрицы вторых производных $\nabla^2_{\hat{\theta}} Q$ или $\nabla^2_{\theta} J$ - матрицы Гессе. Тогда алгоритмы адаптации на основе метода Ньютона приобретают вид

$$\hat{\theta}_{k+1} = \hat{\theta}_k - [\nabla^2_{\hat{\theta}} Q(\hat{\theta}_k)]^{-1} \nabla_{\hat{\theta}} Q(\hat{\theta}_k), \qquad (6.15)$$

где матрица Гессе определяется как (6.11).

Из сопоставления алгоритмов (6.8) и (6.15) следует, что алгоритм Ньютона – это алгоритм градиентного наискорейшего спуска с матрицей весовых коэффициентов, равный обратной матрице Гессе:

$$\tilde{A}_{k+1} = [\nabla^2_{\hat{\theta}} Q(\hat{\theta}_k)]^{-1}. \qquad (6.16)$$

Если в матрице (6.16) вместо текущих оценок $\hat{\theta}_k$ используют их начальное значение $\hat{\theta}_0$, то говорят о модифицированном методе Ньютона.

Если оценочная (целевая) функция $Q(\cdot)$ – квадратичная по аргументу $\hat{\theta}$, то ее градиент $\nabla_{\hat{\theta}} Q(\hat{\theta})$ будет линейной по $\hat{\theta}$ функцией, а матрица $\Gamma_k$ определяемая (6.18), не будет зависеть от вычисляемого вектора параметров $\hat{\theta}$. В этом случае алгоритм типа (6.15) называется алгоритмом Ньютона-Рафсона. Его замечательной особенностью следует считать способность приводить к истинному (оптимальному) значению параметра $\hat{\theta}$ (или $b^*$) при любом начальном значении $\hat{\theta}_0$ или $\hat{b}_0$ за одну итерацию, т. е. $\hat{\theta}_1 = \theta^*$.

Алгоритмы на основе метода Ньютона более трудоемки в вычислительном отношении и рекомендуются в сочетании с обычным градиентным алгоритмом.

Практический интерес представляют случаи детерминированных градиентных алгоритмов с квадратичными оценочными функциями, усредненными на интервале времени [0,k-1] или на скользящем интервале $[k-N_0,k]$, $k \geq N_0 >> 1$, такие эмпирические средние оценочные функции «дают жизнь» множеству практически ценных вычислительных алгоритмов в задачах адаптивного управления и идентификации.

Пусть, например, целевая функция имеет вид среднего арифметического функции $Q(\hat{\theta}) = \sigma^2(\hat{\theta})$:

$$J_{\text{эти}}(\hat{\theta}_k) = \frac{1}{k+1}\sum_{m=1}^{k+1} \sigma^2(\hat{\theta}_k, m), \qquad (6.17)$$

Где обобщенная ошибка $\sigma(\hat{\theta}_k, z_{k+1}) = y_{k+1} - \tilde{y}_{k+1} = (\hat{\theta}_k - \theta^*)^T z_{k+1}$ определена в задаче синтеза контура адаптации дискретной адаптивной системы с неявной настраиваемой моделью с точностью до множителя из алгоритма Ньютона следует алгоритм оценивания вектора параметров

$$\hat{\theta}_{k+1} = \hat{\theta}_k + (k+1)^{-1} \tilde{A}_{k+1}(y_{k+1} - \tilde{y}_{k+1}) z_{k+1}, \qquad (6.18)$$

где $\tilde{A}_{k+1} = \left( \sum_{j=1}^{k+1} z_j z_j^T \right)^{-1}$ – *матрица Калмана.* Особенностью алгоритма (6.18) является его «строгая» оптимальность, в отличие от алгоритма Ньютона при неквадратичной оценочной функции.

**Алгоритм Маркуардта.** Идея алгоритма [2,3] базируется на представлении оценок $\hat{\theta}_k$ как центров гиперсферы в пространстве параметров $\Omega_{\hat{\theta}}$ с радиусами $\Delta\hat{\theta}_k$, ограниченными величиной *r*. Отыскивается экстремум $J_t$ по $\theta$ при ограничении ($g_i(\hat{\theta}),\ i = \overline{1,m}$). В соответствии с методом множителей Лагранжа составляется функция n переменных, где *n* – размерность вектора оценок $\hat{\theta}$:

$$L(\theta, \mu) = J_t(\hat{\theta}) + \sum_{i=1}^{m} \mu_i g_i(\hat{\theta}). \qquad (6.19)$$

Здесь $\mu_i \in \overline{0,\infty}$ – числовые параметры: множители Лагранжа; $g_i(\hat{\theta}) = 0$ – функции ограничения типа равенств. В нашем случае $g_i(\hat{\theta}) = \Delta\hat{\theta}_k^T \Delta\hat{\theta}_k - r^2$. В зависимости от «произвольных» значений $r_1,....,r_m$ имеем *m* уравнений связи, как еще называют ограничения типа равенств. Далее вычисляются и приравниваются нулю частные производные $L(\hat{\theta}, \mu)$ по $\hat{\theta}$ и $\mu$:

$$\frac{\partial L}{\partial \hat{\theta}_j} = \frac{\partial J_t}{\partial \hat{\theta} j} + \sum_{i=1}^{m} \mu_i \frac{\partial g_i(\hat{\theta})}{\partial \hat{\theta}_j} = 0, \quad j \in \overline{1,n;}$$

$$\frac{\partial L}{\partial \mu_i} = g_i(\hat{\theta}) = 0, i \in \overline{1,m.} \qquad (6.20)$$

Система уравнений (6.20) разрешается относительно n+m переменных $\hat{\theta}_1,....,\hat{\theta}_n,\ \mu_1,.....,\mu_m$. В качестве аргумента функционала $J_t$ прием $\Delta\hat{\theta} = \hat{\theta}_{\ddot{e}+1} - \hat{\theta}_{\ddot{e}}$, а в силу произвола выбора $\mu_i$ на интервале от 0 до ∞ полагаем $\mu_i = \mu = const$ для всех $i \in \overline{1,m.}$ Тогда

$$\sum_{i=1}^{m=n} \mu_i \frac{\partial g_i}{\partial \Delta\hat{\theta}_j} = \mu \frac{\partial g_1}{\partial \Delta\hat{\theta}_1} + ..... + \mu \frac{\partial g_{m=n}}{\partial \Delta\hat{\theta}_n} = \mu \cdot I \frac{\partial g(\hat{\theta})}{\partial \Delta\hat{\theta}} = \mu\Delta\hat{\theta},$$

т. к. $g(\hat{\theta}) = \Delta\hat{\theta}^T \Delta\hat{\theta} + r^2$. Отсюда следует правая часть первого уравнения (6.20) в виде

$$\frac{\partial J_t}{\partial \Delta\hat{\theta}} + \mu\Delta\hat{\theta} = 0. \qquad (6.21)$$

Решение векторного уравнения (2.21) относительно $\Delta\hat{\theta}$ осуществляется рекуррентно. Для этого прежде вычислим частную производную интегральной оценочной функции

$$J_t = \int_0^t (y-\bar{y})^T H (y-\bar{y}) d\tau,$$

где $H=H^T>0$ положительно определенная матрица весов, а векторы выходных параметров $y$, $\tilde{y} = \tilde{y}(\theta)$, объекта и настраиваемой модели определены как в алгоритме Гаусса-Ньютона. Итак, имеем:

$$\begin{aligned}
&\frac{\partial J_t(\hat{\theta})}{\partial \Delta\hat{\theta}}\bigg|_{\hat{\theta}=\hat{\theta}_{k+1}} = \frac{\partial J_t(\hat{\theta}_k+\Delta\hat{\theta})}{\partial\Delta\hat{\theta}} = \frac{\partial J_t}{\partial\hat{\theta}}\frac{\partial\hat{\theta}}{\partial\Delta\hat{\theta}}\cdot 1 = \frac{\partial J_t}{\partial\hat{\theta}}\bigg|_{\hat{\theta}=\hat{\theta}_k}\cdot 1 = \\
&\frac{\partial}{\partial\hat{\theta}}\left\{+0{,}5\int_0^t (y-\tilde{y})\partial\tau\right\}_{\tilde{y}=\tilde{y}_{k+1}} = \\
&\frac{\partial}{\partial\hat{\theta}}\left\{-0{,}5\int_0^t (y-\tilde{y})H(y-\tilde{y})\partial\tau\right\}_{\tilde{y}=\tilde{y}_{k+1}} = \\
&-\int_0^t\left(\frac{\partial\tilde{y}^T}{\partial\hat{\theta}}\right)H(y-\hat{y}_k) - \frac{\partial\tilde{y}}{\partial\hat{\theta}^T}\Delta\theta\bigg|_{\hat{\theta}=\hat{\theta}_k}\partial\tau + \mu\Delta\theta = 0.
\end{aligned} \tag{6.22}$$

Здесь использовано представление $\tilde{y} = \tilde{y}(\theta)$ в момент времени $k$+1 первыми двумя членами ряда Тейлора с радиусом разложения $\Delta\hat{\theta}$. Перепишем (6.22) в равносильной форме:

$$\begin{aligned}
0 = \frac{\partial J_t}{\partial\hat{\theta}}\bigg|_{\hat{\theta}=\hat{\theta}_{k+1}} &= \mu I(\hat{\theta}_{k+1}-\hat{\theta}_k) + \int_0^t \frac{\partial\tilde{y}^T}{\partial\hat{\theta}}\bigg|_{\hat{\theta}=\hat{\theta}_k} \\
&- H\frac{\partial\tilde{y}}{\partial\hat{\theta}}\bigg|_{\hat{\theta}=\hat{\theta}_k}(\hat{\theta}_{k+1}-\hat{\theta}_k)d\tau - \int_0^t \frac{\partial\tilde{y}^T}{\partial\hat{\theta}}\bigg|_{\hat{\theta}=\hat{\theta}_k} H(y-\tilde{y}_k)d\tau.
\end{aligned}$$

Отсюда следует

$$(\hat{\theta}_{k+1}-\hat{\theta}_k)\left(\int_0^t \frac{\partial\tilde{y}^T}{\partial\hat{\theta}}\bigg|_{\hat{\theta}=\hat{\theta}_k} H\frac{\partial\tilde{y}}{\partial\hat{\theta}}\bigg|_{\hat{\theta}=\hat{\theta}_k} d\tau + \mu I\right) = \int_0^t \frac{\partial\tilde{y}^T}{\partial\hat{\theta}}\bigg|_{\hat{\theta}=\hat{\theta}_k} H(y-\tilde{y}_k)d\tau$$

В этих преобразованиях вектор y измеряется в момент *k + 1*. Окончательно имеем:

$$\hat{\theta}_{k+1} = \hat{\theta}_k + \left[\left(\int_0^t \frac{\partial\tilde{y}_T}{\partial\hat{\theta}} H\frac{\partial\tilde{y}}{\partial\hat{\theta}^T}\partial\tau\right)_{\hat{\theta}=\hat{\theta}_k} + \mu I\right]^{-1}\left(\int_0^t \frac{\partial\tilde{y}^T}{\partial\hat{\theta}} H(y_{k+1}-\tilde{y}_k)\partial\tau\right)_{\hat{\theta}=\hat{\theta}_k}. \tag{6.23}$$

Алгоритм (6.23) и есть *алгоритм Маркуардта.* При $\mu = 0$ алгоритм совпадает с алгоритмом Гаусса-Ньютона. При $\mu = \infty$ – это алгоритм наискорейшего спуска для интегрального функционала $J_t$ с нулевым шагом Г=0. Параметр $\mu$ формально определяется как решение уравнений (6.20). Практически же эти уравнения не решаются и являются лишь основанием для приведенной рекуррентной схемы (6.23).

**Рекуррентная форма алгоритма максимального правдоподобия**

Рекуррентный алгоритм вычисления оценки максимального правдоподобия можно получить, если в алгоритме Ньютона положить оценочную функцию $Q(\theta) = p(\tilde{z};\hat{\theta})$, т. е. использовать функцию правдоподобия. Таким образом, рекуррентный алгоритм максимального правдоподобия предполагает заданным вблизи $\hat{\theta}^*$ начальное значение $\hat{\theta}_0$, и эта оценка также является оценкой максимального правдоподобия в последовательности $\hat{\theta}_0,.....,\hat{\theta}_{\acute{e}},.....,\hat{\theta}^*_N$.

Далее будем использовать логарифмическую функцию правдоподобия $L(\hat{\theta})$. При статистически независимых измерениях $z_m$ с объемом выборки N логарифмическая функция $L(\hat{\theta})$ принимает вид

$$L(\hat{\theta}) = ln\, p(\tilde{z}_1,.....,\tilde{z}_N;\hat{\theta}) = \sum_{m=1}^{N} lnp(\tilde{z}_m,\hat{\theta}). \qquad (6.24)$$

Оценка максимального правдоподобия состоятельна по Фишеру, если существует единственное решение уравнения правдоподобия:

$$\frac{\partial}{\partial \hat{\theta}} \sum_{m=1}^{N} lnp(\tilde{z}_m;\hat{\theta}) = \sum_{m=1}^{N} \nabla_{\hat{\theta}}\, ln\, p(\tilde{z}_m;\tilde{\theta})\Big|_{\hat{\theta}=\hat{\theta}^*_N} = 0. \qquad (6.25)$$

Пусть $N = k \to \infty$. если k достаточно велико, то естественно считать, что $\left\|\hat{\theta}_{k+1} - \hat{\theta}_k\right\| << 1$ или $\left\|\Delta\hat{\theta}_{k+1}\right\| << 1$ и тогда существуют предпосылки для использования алгоритма Ньютона (6.15), где вместо обратной матрицы Гессе $\left[\nabla^2_{\hat{\theta}} Q(\hat{\theta}_k)\right]^{-1}$ должна быть обратная матрица Гессе $\left[\nabla^2_{\hat{\theta}}\, ln\, L(\hat{\theta}_k)\right]^{-1}$. Теперь рекуррентный алгоритм решения уравнения правдоподобия (6.25) можно записать аналогично алгоритму Ньютона (6.15):

$$\hat{\theta}_{k+1} = \hat{\theta}_k - \left( \sum_{m=1}^{N} \nabla^2_{\hat{\theta}} \ln p(\tilde{z}_m; \hat{\theta}_k) \right)^{-1} \nabla_{\hat{\theta}} \ln p(\tilde{z}_{k+1}; \hat{\theta}_k). \qquad (6.26)$$

Начальное значение $\hat{\theta}_0$ должно быть также оценкой максимального правдоподобия, близкой к $\hat{\theta}_k$, а $\hat{\theta}_k \to \hat{\theta}^* \to \theta$. Тогда функция $L(\hat{\theta})$ по ее определению есть сумма независимых и одинаково распределенных случайных величин:

$$\lim|_{k\to\infty} M\left\{ \sum_{m=1}^{k} \frac{\partial^2 \ln p(\tilde{z}_m; \hat{\theta})}{\partial \hat{\theta} \partial \hat{\theta}^T} \right\} = -J_1(\hat{\theta}),$$

где $J_1(\hat{\theta})$ – информационная матрица Фишера. В силу закона больших чисел следует

$$\sum_{m=1}^{k} \nabla^2_{\hat{\theta}} \ln p(\tilde{z}_m; \hat{\theta}) = -kJ_1(\theta) + \rho(\Delta\theta, k).$$

При $k \to \infty$ $\rho(\Delta\theta, k) \to 0$ и эмпирическое среднее $\nabla^2_{\hat{\theta}} L(\hat{\theta})$ стремится к математическому ожиданию, в данном случае к информационной матрице Фишера $J_1(\theta)$. Учитывая это, приходим к так называемому алгоритму Сакрисона или к рекуррентному алгоритму максимального правдоподобия

$$\hat{\theta}_{k+1} = \hat{\theta}_k + \frac{1}{k} J_1^{-1}(\theta) \nabla_{\hat{\theta}} \ln p(\tilde{z}_{k+1}; \hat{\theta}_k). \qquad (6.27)$$

Матрица Фишера $J_1(\theta)$ может быть заменена ее оценкой $\hat{J}_1(\hat{\theta}_k)$. В частном рассматриваемом случае, когда $\tilde{z}_1, ....., \tilde{z}_m, ....$ статистически независимы и $\tilde{z}_m \in N(0, \sigma_z^2)$, имеем

$$\nabla_{\hat{\theta}} L(\hat{\theta}_{k+1}) = \frac{\partial}{\partial \hat{\theta}} \ln p(\tilde{z}_1, ....., \tilde{z}_k; \hat{\theta}_{k-1}) = \frac{\partial}{\partial \hat{\theta}} \sum_{m=1}^{k} \tilde{z}_m^2(\hat{\theta}).$$

Тогда оценкой информационной матрицы Фишера может служить матрица, вычисляемая во времени параллельно с алгоритмом МП:

$$\hat{J}_1(\hat{\theta}) = \sum_{m=1}^{k} \left( \frac{\partial \tilde{z}_m}{\partial \hat{\theta}} \right) \left( \frac{\partial \tilde{z}_m}{\partial \hat{\theta}} \right)^T$$

или на конечном «скользящем» интервале $N_0$:

$$\hat{J}_1(\hat{\theta}_k) = \sum_{m=k-N_0}^{k} \left( \frac{\partial \tilde{z}_m}{\partial \hat{\theta}} \right) \left( \frac{\partial \tilde{z}_m}{\partial \hat{\theta}} \right)^T. \qquad (6.28)$$

Окончательно рекуррентный алгоритм вычисления оценки максимального правдоподобия принимает с учетом (6.28) следующий вид:

$$\hat{\theta}_{k+1} = \hat{\theta}_k + \frac{1}{k}\delta_{k+1},$$

$$\delta_{k+1} = \hat{J}_1^{-1}(\hat{\theta}_k)\nabla_{\hat{\theta}} \ln p(\tilde{z}_{k+1}; \hat{\theta}_k), \qquad (6.29)$$

$$\hat{J}_1(\hat{\theta}_k) = \sum_{m=k-N_0}^{k} \left(\frac{\partial \tilde{z}_m}{\partial \hat{\theta}}\right)\left(\frac{\partial \tilde{z}_m}{\partial \hat{\theta}}\right)^T.$$

Установлено, что алгоритм (6.29) сходится почти наверное [16] и оценки максимального правдоподобия $\hat{\theta}_k$ состоятельны в сильном смысле по Фишеру и эффективны. Однако нужно помнить об условиях, при которых построен алгоритм (6.29): он получен при гипотезе о статистической независимости измерений $z_1,.....,z_k$ и удачном выборе начального приближения $\hat{\theta}_0$ в малой окрестности максимума $p(z|\theta)$. Это весьма ограничивающее условие для динамических адаптивных систем. Кроме того, последующие начальные оценки $\hat{\theta}_1, \hat{\theta}_2,.....$ также могут не быть оценками максимального правдоподобия, и лишь при *k>>1* алгоритм (6.29) обладает теми свойствами, которые установлены в [16]. Поэтому возникает проблема ускорения или акселерации алгоритма (6.29) при небольших *k*.

### 6.3. Рекуррентный алгоритм МНК

Метод наименьших квадратов в рассмотренной выше форме (глава 5) обладает существенным недостаткам: при получении результатов новых измерений на t-ом шаге необходимо решение задачи производить заново, с учетом данных, полученных на всех t шагах измерения. Такой алгоритм заметно снижает оперативность (быстродействие) процесса идентификации, особенно при использовании автоматизированной системы сбора и обработки информации. Этого недостатка лишен рекуррентный МНК, идея которого сводится к следующему.

Обозначим оценку вектора $\hat{\theta}$, полученную при обработке результатов t измерений, через $\hat{\theta}$ (t). Тогда выражение (6.4.4) можем записать

$$\hat{\theta} = [U^T(t)U(t)]^{-1}U(t)x(t) =$$
$$= [U^T(t-1)U(t-1) + u(t)u^T(t)]^{-1} \cdot \qquad (6.30)$$
$$\cdot [U^T(t-1)x(t-1) + u(t)x(t)].$$

Применяя формулу обращения матриц [8, 9] к (6.30) получим

$$\hat{\theta} = \left[ \mathrm{P(t-1)} - \frac{\mathrm{P(t-1)u(t)u^T(t)P(t-1)}}{\mathrm{u^T(t)P(t-1)u(t)+1}} \right] \left[ \mathrm{U^T(t-1)x(t-1) + u(t)}x(t) \right] =$$
$$= \mathrm{P(t-1)U^T(t-1)x(t-1)} + \qquad (6.31)$$
$$+ \mathrm{P(t-1)x(t)} \frac{x(t) - \hat{\theta}^{\mathrm{T}}(t-1)\mathrm{u(t)}}{\mathrm{u^T(t)P(t-1)u(t)+1}},$$

где $P(t-1) = [U^T(t-1)U(t-1)]^{-1}$.

Таким образом, из (6.31) получаем рекуррентный алгоритм МНК

$$\hat{\theta}(t) = \hat{\theta}(t-1) - \tilde{A}(t-1)\left[x(t) - \hat{\theta}^T(t-1)u(t)\right]u(t), \qquad (6.32)$$

где $\tilde{A}(t-1) = \dfrac{[U^T(t-1)U(t-1)]^{-1}}{u^T(t)[U^T(t-1)U(t-1)]^{-1}u(t)+1}$.

В формуле (6.32) для вычисления $\theta$ по $t$-му измерению используется оценка $\hat{\theta}(t-1)$, полученная по предыдущим измерениям, а также результаты новых измерений u(t) и x(t). Оценка, получаемая по рекуррентной формуле (6.31), после любого числа N шагов (измерений) совпадает с оценкой, получаемой при обработке полной выборки из N измерений по формуле (6.4.4).

$$\tilde{A}_{k+1} = \tilde{A}_k - \frac{\tilde{A}_k z_{k+1} z_{k+1}^T \tilde{A}_x}{1 + z_{k+1}^T \tilde{A}_k z_{k+1}}. \qquad (6.33)$$

Рекуррентное вычисление матрицы $\tilde{A}_k$ начинается с момента $k_0 >> 1$, когда $\tilde{A}_{k_0}$ – невырожденная матрица. В практических приложениях принимают начальное значение $\tilde{A}_0 = \gamma_0 \cdot I$, где число $\gamma_0 >> 1$.

Алгоритм (6.32) построен на основе квадратичной целевой функции $J_{ýìï}(\hat{\theta}_k)$, и при использовании матрицы усиления $\tilde{A}_{k+1}$ в виде матрицы Калмана при $k >> 1$ называется рекуррентном алгоритмом метода наименьших квадратов (РМНК).

Обычно для промышленных объектов характерна коррелированность во времени шумов, действующих на объект. Использование обычного МНК при таком шуме, т. е. при минимизации выражения (6.6), вызывает смещение оценок параметров, увеличение дисперсии этих оценок. Ухудшение этих оценок, в свою очередь, приводит к

ухудшению свойств оценок переменных состояния *x(k)* и в итоге к снижению качества управления.

Для получения несмещенных оценок используется обобщенный МНК (ОМНК). При использовании ОМНК оцениваются параметры моделей объекта и шума на его выходе. Идентификации подвергается модель максимального правдоподобия (МП-модель) для которой связь между переменными задается уравнением

$$A(z)x(z) - B(z)z^{-d}u(z) = D(z)e(z). \tag{6.34}$$

Вводя расширенные векторы данных

$$\begin{aligned}\hat{\psi}^T(k) = [\, & y(k-1),...y(k-n),u(k-d-1),... \\ & u(k-1-d-n),\hat{v}(k-1),...\hat{v}(k-n)]\end{aligned} \tag{6.35}$$

и параметров

$$\hat{\theta}^T = [\,\hat{a}_1...\hat{a}_n,\hat{b}_1...\hat{b}_n,\hat{d}_1...\hat{d}_n\,], \tag{6.36}$$

выходной сигнал объекта можно записать через (6.34) и (6.35)

$$x(k) = \hat{\psi}^T(k)\hat{\theta}(k-1) + e(k). \tag{6.37}$$

Так как сигнал помехи *e(k)* неизвестен, то используется его оценка, $\hat{v}(k)$ определяемая из уравнения

$$\hat{v}(k) = y(k) - \hat{\psi}^T(k)\hat{\theta}(k-1). \tag{6.38}$$

Оценки параметров МП-модели вычисляются аналогично как в МНК по формулам (6.30) – (6.32).

### 6.3.1. Стохастическая аппроксимация

Методы стохастической аппроксимации являются по существу стохастическими аналогами градиентных методов. В частности, широко, известный алгоритм Кифера-Вольфовица представляет собой стохастический аналог обычного градиентного метода для отыскания экстремума (минимума) унимодальной функции $J(\hat{\theta})$. Ранее для этой задачи был получен регулярный итеративный алгоритм

$$\hat{\theta}(t) = \hat{\theta}(t-1) - \gamma(t)\nabla J\left[\hat{\theta}(t-1)\right].$$

В данном пункте рассматривается случай, когда измерения искажаются помехой, т. е.

$$Q\left[\hat{\theta}(t)\right] = J\left[\hat{\theta}(t)\right] + \mu, \tag{6.39}$$

где μ – помеха с нулевым математическим ожиданием и в общем случае неизвестным распределением, причем, μ и $\hat{\theta}$ независимы.

При этом стохастический алгоритм для отыскания корней уравнения регрессии (6.39) имеет вид [28]

$$\hat{\theta}(t) = \hat{\theta}(t-1) - \gamma(t)\nabla Q\left[\hat{\theta}(t-1)\right]. \qquad (6.40)$$

Сходимость этого алгоритма имеет место при выполнении условий

$$\lim_{t\to\infty}\gamma(t) = 0; \quad \sum_{t=1}^{\infty}\gamma(t) = \infty; \quad \sum_{t=1}^{\infty}\gamma^2(t) < \infty. \qquad (6.41)$$

Условиям (6.43) удовлетворяют, например, последовательности вида

$$\gamma(t) = \frac{\beta}{\beta + t}; \gamma(t) = \frac{c}{t^{\beta}}; c = const; \frac{1}{2} < \beta \le 1. \qquad (6.42)$$

В выражении (6.42) первое условие связано с требованием подавления помех в установившемся режиме (сходимость регулярной составляющей процесса), второе – обеспечения асимптотической несмещенности оценки параметров $\hat{\theta}$, третье – подавления суммарного влиянии помехи за весь период работы алгоритма.

Если для отыскания производной $\frac{dQ}{d\hat{\theta}}$ используется формула численного дифференцирования, например,

$$\frac{dQ}{d\hat{\theta}_j(t)} = \frac{Q[\hat{\theta}_j(t) + \Delta\hat{\theta}_j(t-1)] - Q[\hat{\theta}_j(t-1) + \Delta\hat{\theta}_j(t-1)]}{2\Delta\hat{\theta}_j(t-1)} = d_j,$$

то алгоритм стохастической аппроксимации записывается в виде

$$\hat{\theta}_j(t) = \hat{\theta}_j(t-1) - \gamma(t-1)d_j, \; j = \overline{1,n},$$

причем условия сходимости имеют форму $\lim_{t\to\infty}\gamma(t) = 0$; $\lim_{t\to\infty}\Delta\hat{\theta}_j(t) = 0$;

$$\sum_{t=1}^{\infty}\gamma(t) = \infty; \quad \sum_{t=1}^{\infty}\gamma^2(t) < \infty; \quad \sum_{t=1}^{\infty}\frac{k^t}{\Delta\hat{\theta}_j(t)} < \infty. \qquad (6.43)$$

Для алгоритмов стохастической аппроксимации также существуют различные способы увеличения скорости их сходимости.

Отметим, что алгоритмы стохастической аппроксимации позволяют также находить такие значения параметров, при которых неизвестная функция регрессии принимает заданное значение. На этом строятся, в частности, алгоритмы идентификации линейных статистических объектов. Рассмотрим, например, объект, описываемый линейной дискретной моделью

$$x(t) = \hat{\theta}^T u(t) + \nu(t), \qquad (6.44)$$

где *x(t)* – скалярный выходной сигнал; $u(t)$ – вектор измеряемого входного сигнала; $\nu(t)$ – скалярная помеха с нулевым математическим ожиданием; $\theta$ – вектор неизвестных параметров. Алгоритм стохастической аппроксимации для оценивания вектора в простейшем случае имеет вид

$$\hat{\theta}(t) = \hat{\theta}(t-1) - \gamma(t-1)\, x(t-1). \qquad (6.45)$$

Но т. к. входные сигналы измеряются, то для оценивания параметров θ можно использовать модифицированный алгоритм стохастической аппроксимации

$$\hat{\theta}(t) = \hat{\theta}(t-1) - \gamma(t-1)[\hat{\theta}^T(t-1)u(t-1)x(t)]u(t). \qquad (6.46)$$

Алгоритм (6.46) получается при минимизации функционала

$$J(\hat{\theta}) = M\{\nu(t)\} = M\{[\hat{\theta}^T u(t) - x(t)]^2\}.$$

### 6.3.2. Акселерация алгоритмов стохастической аппроксимации

Основные идеи акселерации, т. е. повышение скорости сходимости алгоритмов стохастической аппроксимации связаны с выбором начальных значений оценок $\hat{\theta}_0$ и матриц усиления Г или коэффициентов γ. Эти значения не влияют на асимптотические свойства алгоритмов, но они существенны для реальных случаев конечных к≤N с учетом достигаемой точности приближения $\hat{\theta}_N$ к истинному значению $\theta$. Увеличение скорости сходимости $\hat{\theta}_K$ к $\theta$, т. е. необходимость уменьшения *N*, обеспечивающая нужную точность приближения $\hat{\theta}_N$ к $\theta$, будем называть акселерацией алгоритмов оценивания или оптимизации, если речь идет о параметрах *b* регулятора основного контура. Акселерация может осуществляться также путем выбора такого алгоритма управления и, следовательно, входных воздействий на объект управления *u(t)*, при которых оценивание параметров модели управления в принципе протекает быстрее на принятом классе алгоритмов. Наконец, ускорение сходимости зависит от текущих значений матриц коэффициентов усиления $\tilde{A}_{\hat{e}}$ или и значений градиентов оценочных функций $\nabla_\theta Q(z_k;\theta)$.

Приведем способы акселерации алгоритмов, следующие этой идее.

Алгоритм Фабиана. Алгоритм стохастической аппроксимации запишем в виде

$$\hat{\theta}_{x+1}= \hat{\theta}_k - \frac{1}{k+1}\nabla_\theta Q(\tilde{z}_{k+1};\hat{\theta}_k) = \theta_k - \gamma_{k+1}\left\|\nabla_{\hat{\theta}} Q(\tilde{z}_{k+1};\hat{\theta}_k)\right\| sgn\,\nabla_{\hat{\theta}} Q(\tilde{z}_{k+1};\hat{\theta}_k).$$

Если $\left\|\nabla_{\hat{\theta}} Q(\cdot)\right\| > 1,$ а оценка $\hat{\theta}_K$ находится в окрестности $\theta$, то вместо умножения $\gamma_{k+1}$ на $\nabla_{\hat{\theta}} Q(\cdot)$ предполагается умножать $\gamma_{k+1}$ на знаковую функцию $sgn\nabla_{\hat{\theta}} Q(\cdot)$. Алгоритм стохастической аппроксимации в виде

$$\hat{\theta}_{k+1}= \hat{\theta}_k - \gamma_{k+1}\, sgn\nabla_{\hat{\theta}} Q(\tilde{z}_{k+1};\hat{\theta}_k) \qquad (6.47)$$

называется алгоритмом Фабиана. В работе [3] показано, что сходимость $\hat{\theta}_K$ к $\theta$ в соответствии с алгоритмом (6.47) правомерен, если только помеха имеет распределение, медиана которого совпадает с математическим ожиданием. Алгоритм Фабиана является «релейным» алгоритмом, где коэффициент усиления $\gamma = (k+1)^{-1}\, sgn\nabla_{\hat{\theta}} Q(\tilde{z}_{k+1};\hat{\theta}_k)$ меняет свой знак.

Алгоритм Кестена. Если оценка параметра $\theta$ находится в окрестности $\hat{\theta}$, то по методу Роббинса - Монро она будет флуктуировать близ $\theta$. Если же $\hat{\theta}$ не находится в окрестности $\theta$, то нет оснований ожидать, что это будет так. Это соображение использовано в алгоритме Кестена для ускорения алгоритма стохастический аппроксимации. Шаг $\gamma_k$ выбирается из последовательности значений $\{a_k\}$, удовлетворяющей теореме Роббинса - Монро :

$$\gamma_1 = a_1,\ \gamma_2 = a_2, \gamma_k = a_{s(k)}$$

$$s(k) = 2 + \sum_{i=3}^{k} l[(\theta_i - \theta_{i-1})(\theta_{i-1} - \theta_{i-2})] \qquad (6.48)$$

$$l(x) = \begin{cases} 1, \text{åñëè } x \le 0; \\ 0, \text{åñëè } x > 0. \end{cases}$$

Таким образом, всякий раз, когда $sgn(\hat{\theta}_i - \hat{\theta}_{i-1}) \neq sgn(\hat{\theta}_{i-1} - \hat{\theta}_{i-2})$, выбирается новое значение $a_{s(k)}$ из ряда допускаемых теоремой. Робинса - Монро значений, например из гармонического ряда $\gamma_k = k^{-1}$, $k$ =1, 2,…. Показано, что при выборе $\gamma_k$ по Кестену алгоритм стохастической аппроксимации сходится с вероятностью 1. В алгоритме метода максимального правдоподобия (6.48) индикаторная функция *l(x)*

удовлетворяет условиям до тех пор, пока оценка $\hat{\theta}_k$ не попадет в окрестность $\hat{\theta}^*$. При ее «проскакивании» изменяется знак шага приращения $\Delta\theta_k$, что и является сигналом для уменьшения $\gamma_k$. Таким образом, вдали от искомого решения $\hat{\theta}^*$ ускоренный алгоритм стохастической аппроксимации имеет большой коэффициент «усиления», чем обычный алгоритм с равномерным уменьшением $\gamma_k$ от шага к шагу.

Систематическое изложение способов акселерации алгоритмов, учитывающих априорную информацию об оптимальных оценках и использующих обобщенные эмпирические средние потери, приводится в монографии Я.З. Цыпкина [39]. Для формирования акселерантных алгоритмов им использована одна из разновидностей алгоритма Ньютона - Рафсона для решения систем нелинейных уравнений и их частный случай – рекуррентная форма алгоритма вычислений по методу наименьших квадратов.

### 6.4. Об адаптивных алгоритмах идентификации

В адаптивных алгоритмах идентификации получаемая (накапливаемая) при построении модели информация оперативно используется для уточнения этой модели. В свою очередь, уточняемые таким образом (в процессе работы объекта) модели получили название адаптивных моделей. Применение адаптивных методов идентификации и соответственно адаптивных моделей принципиально необходимо при управлении нестационарными объектами [10, 14, 20, 30].

Для реализации адаптивных моделей используется простые алгоритмы объектов такой большой размерности, где применить для построения оптимальные методы невозможно из-за вычислительных возможностей представлено много примеров [20], что, чем выше размерность объекта, тем более простые алгоритмы следует применять для построения его модели. Более сложные алгоритмы применяются только для того, чтобы быстрее строить модель объекта. Это значить – использовать меньшее число опытов.

В пользу простых алгоритмов адаптации говорит и следующий факт. Разница в количестве экспериментов, необходимых для получения оценок параметров объекта с заданной надежностью, для разных алгоритмов уменьшается с увеличением размерности объекта. Иначе, скорости сходимости собственности алгоритмов выравниваются.

При необходимости вариацию входных переменных всегда можно построить достаточно точную линейную модель произвольного нелинейного объекта. Это естественное предположение позволяет для некоторых объектов ограничиться линейной адаптивной моделью там, где необходимо было бы построение существенной нелинейной стационарной модели. Такими объектами являются, например, нелинейные объекты, у которых средние значения входных переменных изменяются медленно по сравнению с небольшими колебаниями относительно этих средних.

Несколько слов следует сказать относительно свойств алгоритмов адаптации. Будем различать стационарные и нестационарные алгоритмы. Будем называть алгоритм стационарным, если оценки параметров объекта, получаемые по этому алгоритму, являются функциями входных и выходных переменных объекта, но явно не зависят от времени или от известных функций времени, не являются входными переменными объекта. И, наоборот, алгоритм, в который явно входит время или известная функция времени, будем называть нестационарным. Пример нестационарного алгоритма может служить, например, такой алгоритм:

$$k_{(\tilde{\theta}+1)} = k_\theta + \frac{1}{N+1}\left[y(N+1) - k^T(N)x(N+1)\right]\theta(N+1).$$

Здесь оценки параметров объекта $k$ уточняются с течением времени, т. е. с увеличением $N$, идет уменьшение процесса оценивания параметров. Такие нестационарные алгоритмы оказываются оптимальными среди одношаговых алгоритмов при определении параметров стационарных объектов и хорошо работают даже в условиях помех. Но такие алгоритмы принципиально непригодны для слежения за периметрами нестационарного объекта. Действительно, при достаточно большом $N$ оценки коэффициентов практически перестают уточняться, в то время как параметры объекта могут изменяться при работе все время.

Существует достаточно большое число алгоритмов адаптации, отличающиеся глубиной памяти. Алгоритмы, использующие для уточнения оценок параметров результаты только одного эксперимента, будем называть одношаговыми. Алгоритмы, использующие для уточнения параметров результаты нескольких экспериментов, назо-

вем многошаговыми. Примером многошагового алгоритма адаптации может служить алгоритм текущих наименьших квадратов.

В основу адаптивных методов идентификации положены регулярные (детерминированные) или вероятностные итеративные алгоритмы градиентного типа (6.1.3).

В детерминированных алгоритмах сходимость итерационного процесса определяется свойствами целевой функции в близи оптимальной точки θ, выбором начального значения $\hat{\theta}_0$ и значениями элементов матрицы Г.

Большинство вероятностных алгоритмов адаптации базируется на методе стохастической аппроксимации. Одним из наиболее распространенных адаптивных алгоритмов идентификации параметров линейных объектов являются алгоритм Качмажа и его различные модификации.

Рассмотрим линейный объект, описываемый уравнением

$$x(t) = u^T(t)\theta + v(t), \qquad (6.49)$$

где $x(t)$ – наблюдаемый скалярный сигнал; $u(t)$ – наблюдаемый $n$-мерный входной сигнал; $\theta$ – $n$-мерный вектор идентифицируемых параметров; $v(t)$ – ненаблюдаемая помеха с нулевым средним.

Сформулируем задачу идентификации следующим образом. Допустим, что в пространстве параметров $\theta$ в $t$-$1$ момент времени получена оценка $\hat{\theta}(t-1)$,. Потребуем, чтобы следующая оценка $\hat{\theta}(t)$, минимизировала определенную меру, характеризующую "близость" оценок $\hat{\theta}(t-1)$ и $\hat{\theta}(t)$ при условии, что удовлетворяет уравнению

$$x(t) = u^T(t)\hat{\theta}(t). \qquad (6.50)$$

В качестве указанной меры "близости" используем евклидово расстояние $\rho[\hat{\theta}(t), \hat{\theta}(t-1)]$ между соответствующими оценками. Тогда задача идентификации математически формулируется как следующая задача на условный экстремум

$$\min \hat{\theta}(t) = \rho[\hat{\theta}(t), \hat{\theta}(t-1)] = \{[\hat{\theta}(t) - \hat{\theta}(t-1)]^T [\hat{\theta}(t) - \hat{\theta}(t-1)]\}^{-\frac{1}{2}} \qquad (6.51)$$

при условии $x(t) - u^T(t)\hat{\theta}(t) = 0$.

Введем функцию Лагранжа

$$L[\hat{\theta}(t), \lambda(t)] = \rho[\hat{\theta}(t), \hat{\theta}(t-1)] + \lambda(t)[x(t) - u^T(t)\hat{\theta}(t)]. \qquad (6.52)$$

Тогда решение задачи (6.51), (6.52) удовлетворяет системе уравнений

$$\frac{\partial L}{\partial \hat{\theta}(t)} = 0, \qquad \frac{\partial L}{\partial \lambda(t)} = 0$$

или после дифференцирования (6.52)

$$[\hat{\theta}(t) - \hat{\theta}(t-1)]\rho^{-1}[\hat{\theta}(t), \hat{\theta}(t-1)] - \lambda(t)u(t) = 0, \tag{6.53}$$

$$x(t) - u^T(t)\hat{\theta}(t) = 0. \tag{6.54}$$

Умножая слева обе части уравнения (6.53) на $u^T(t)$, находим

$\lambda(t) = \dfrac{u^T(t)[\hat{\theta}(t) - \hat{\theta}(t-1)]}{\rho[\hat{\theta}(t), \hat{\theta}(t-1)]u^T(t)u(t)}$ или с учетом (6.50)

$$\lambda(t) = \frac{x(t)u^T(t) - \hat{\theta}(t-1)}{\rho[\hat{\theta}(t), \hat{\theta}(t-1)]u^T(t)u(t)}. \tag{6.55}$$

Подставим выражение (6.55) для $\lambda(t)$ в уравнение (6.53) и найдем

$$\hat{\theta}(t) = \hat{\theta}(t-1) + \frac{x(t)u^T(t) - \hat{\theta}(t-1)}{u^T(t)u(t)}u(t). \tag{6.56}$$

Полученное выражение и есть алгоритм Качмажа, широко используемый для решения систем линейных алгебраических уравнений и для оценивания параметров линейных статических объектов.

Сравнение рекуррентных алгоритмов МНК, приведенных в п. 6.3 с алгоритмом Качмажа показывает их некоторую идентичность. Действительно, запишем рекуррентный алгоритм МНК (6.32) и алгоритм Качмажа соответственно в виде

$$\hat{\theta}(t) = \hat{\theta}(t-1)] + \gamma_1[x(t) - u^T(t)\hat{\theta}(t-1)]u(t),$$

$$\hat{\theta}(t) = \hat{\theta}(t-1)] + \gamma_2[x(t) - u^T(t)\hat{\theta}(t-1)]u(t),$$

где $\gamma_1 = \dfrac{[U^T(t-1) - U(t-1)]^{-1}}{u^T(t)[U^T(t-1), U(t-1)]^{-1}u(t) + 1}$, $\gamma_2 = \dfrac{1}{u^T(t)u(t)}$.

Различие сравниваемых алгоритмов заключается в структуре коэффициентов $\gamma_1$ и $\gamma_2$. В рекуррентном алгоритме МНК для вычисления коэффициента $\gamma_1$ необходимо знать матрицу *U(t-1)*, т. е. значения вектора входных сигналов в моменты *k=1,...,t-1*. В алгоритме Качмажа коэффициент $\gamma_2$ определяется значениями вектора только в данный момент времени. Последнее обстоятельство приводит к существенному уменьшению объема памяти вычислительного устройства, реализующего алгоритм Качмажа. В то же время показано, что точность этого алгоритма не намного хуже, чем алгоритм МНК.

В настоящее время известны различные модификации алгоритма Качмажа. Алгоритм (6.56) чаще применяется при отсутствии помех или малой их интенсивности. В общем случае наиболее употребительной является форма алгоритма в виде

$$\hat{\theta}(t) = \hat{\theta}(t-1) + \frac{x(t) - u^T(t)\hat{\theta}(t-1)}{\alpha + u^T(t)u(t)}u(t), \qquad (6.57)$$

где $\alpha$ – скалярный неотрицательный параметр, некоторые соображения по выбору которого можно найти в работах [9, 12, 20, 22, 23, 28].

## 6.5. Сходимость итерационного процесса идентификации

### 6.5.1. Случай детерминированной системы

Рассмотрим условия сходимости итерационного процесса идентификации по результатам измерений входного и выходного сигнала объекта с использованием моделей параллельного типа. Пусть объект идентификации описывается уравнением вида (6.49) при $\nu(t) = 0$. Предположим, что порядок системы определен, тогда параметры процесса можно оценивать с применением модели аналогичной структуры и размерности вида

$$\hat{x}(t) = \hat{\theta}^T(t-1)u(t), \qquad (6.58)$$

где $\hat{\theta}(t)$ – оценка вектора $\theta$; $\hat{x}(t)$ – оценка значения выхода $x(t)$ на $t$-ом шаге.

В основу представления рекуррентных схем оценки параметров положим метод градиента, алгоритм которого в соответствии с выражениями (6.3) и (6.4) можно записать в общем виде как

$$\hat{\theta}(t) = \hat{\theta}(t-1) - \tilde{A}(t)\nabla J[\hat{\theta}(t-1)], \qquad (6.59)$$

где $\Gamma(t)$ – матрица размера $n \times n$, определяющая величину шага.

Для критерия качества вида

$$J(\hat{\theta}) = \frac{1}{2}\varphi(t)[x(t) - \hat{x}(t)]^2, \qquad (6.60)$$

где $\varphi(t) > 0$ – мгновенный весовой коэффициент при стационарном векторе $\theta$ градиент определяется как $\nabla J(\hat{\theta}) = \varphi(t)[x(t) - \hat{x}(t)]u(t)$. Тогда общая процедура оценки вектора параметров согласно алгоритма (6.59) определяется следующим законом настройки:

$$\hat{\theta}(t) = \hat{\theta}(t-1) - \varphi(t)\tilde{A}(t)[x(t) - \hat{x}(t)]u(t). \qquad (6.61)$$

Сходимость процесса идентификации с применением алгоритма (6.61) можно показать, если ввести функцию Ляпунова вида $V(t) = r^T(t)\tilde{A}r(t)$, где $r(t) = \theta - \hat{\theta}(t)$.

Утверждение 1. Если $\varphi(t) > 0$ определить как

$$\varphi(t) = \eta(t)[\alpha(t) + u_T(t)\tilde{A}u(t)]^{-1}, \qquad (6.62)$$

где $\eta(t)$ и $\alpha(t)$ удовлетворяют условиям

$$0 < \eta(t) < 2, \qquad \alpha(t) \geq 0, \qquad (6.63)$$

то имеет место сходимость процесса идентификации.

Поскольку $u^T(t)\tilde{A}\,u(t) \leq \lambda_m u^T(t)u(t)$, где $\lambda_m$ – наибольшее собственное значение матрицы Г, то условия (6.62), (6.63) представимы в виде

$$\varphi(t) = \eta(t)[\alpha(t) + u^T(t)u(t)]^{-1},$$

$$0 < \eta(t) < \frac{2}{\lambda_m}, \quad \alpha(t) \geq 0.$$

Таким образом, алгоритм (6.61) может быть представлен в виде

$$\hat{\theta}(t) = \hat{\theta}(t-1) - \eta(t)[\alpha(t) + u^T(t)u(t)]^{-1}\tilde{A}[x(t) - \hat{x}(t)], \qquad (6.64)$$

$$0 < \eta(t) < \frac{2}{\lambda_m}, \qquad \alpha(t) \geq 0. \qquad (6.65)$$

В практических случаях алгоритм (6.64) при определенном выборе диагональных элементов $\gamma_j$; матрицы $\Gamma$ и $\eta(t)$, $\alpha(t)$, удовлетворяющих условиям (6.65), может быть использован, например, в виде следующих рекуррентных алгоритмов:

$$\hat{\theta}(t) = \hat{\theta}(t-1) - \frac{x(t) - \hat{x}(t)}{\alpha(t) + u^T(t)u(t)}u(t), \qquad (6.66)$$

$$\hat{\theta}(t) = \hat{\theta}(t-1) - \tilde{A}(t)\left[x(t) - \hat{x}(t)\right]u(t), \qquad (6.67)$$

$$\hat{\theta}(t) = \hat{\theta}(t-1) - \gamma(t)\left[x(t) - \hat{x}(t)\right]u(t), \qquad (6.68)$$

где $\Gamma$ – есть матрица размера nxn.

### 6.5.2. Случай стохастической системы

Рассмотрим оценку параметров объекта (6.49) с использованием модели (6.58). Положим, что алгоритм настройки параметров модели определен как

$$\hat{\theta}(t) = \hat{\theta}(t-1) - \gamma(t)\left[x(t) - \hat{x}(t)\right]u(t), \qquad (6.69)$$

где $\gamma(t) \geq 0$ – скалярная величина.

Определим такие условия идентификации, чтобы $\| r(t) \| \to 0$ при $t \to \infty$.

Для этой цели представим ошибку в виде

$$x(t) - \hat{x}(t) = r^T(t-1)\, u(t) + \nu(t) \tag{6.70}$$

и выберем функцию Ляпунова согласно следующему выражению

$$V(t) = \frac{1}{2}\|r(t)\|^2. \tag{6.71}$$

Положим, что $\nu(t)$ представляет собой некоррелированную последовательность случайных величин с нулевым средним, дисперсией $\sigma_\nu^2 < \infty$ и независимую от входных величин. Тогда с учетом уравнений (6.69) – (6.71) получим условие сходимости алгоритма (6.69)

$$\begin{aligned} M\{\Delta V(t)\} = M\{ & -2\gamma(t) r^T(t-1) u(t) u^T(t-1) r(t-1) + \\ & + \gamma^2(t) u^T(t) u(t) r^T(t-1) u(t) u^T(t) r(t-1) + \\ & + \gamma^2(t)\, u^T(t) u(t)\, \nu^2(t)\} < 0, \end{aligned} \tag{6.72}$$

где $M\{\Delta V(t)\} = M\{V(t) - V(t-1)\}$.

Полагая, что $\lambda_1(t)$, $\lambda_n(t)$ – соответственно наименьшее и наибольшее собственное значение матрицы $u(t)u^T(t)$, можно получить следующую оценку:

$$\begin{aligned} M\{\Delta V(t)\} < M\{ & -2\gamma(t)\lambda_1(t)\| r(t-1)\|^2 + \\ & + \gamma^2(t) u^T(t) u(t) \lambda_n(t)\| r(t-1)\|^2 + \gamma^2(t) u^T(t) u(t)\, \sigma_\nu^2\} < 0. \end{aligned} \tag{6.73}$$

Тогда будет справедливо следующее утверждение.

Утверждение 2. Если $\gamma(t) \geq 0$ определить как

$$\gamma(t) = \frac{\eta_1 \lambda(t) \| r(t) \|^2}{[\, \lambda_n(t) \| r(t) \|^2] + \sigma_\nu^2 u^T(t) u(t)}, \tag{6.74}$$

где $0 < \eta < 2$, то имеет место выполнение условия (6.73).

Таким образом, $\hat{\theta}(t)$ сходится к $\theta$ по вероятности и является состоятельной оценкой.

Поскольку величина $\| r(t) \|^2$ неизвестна, то практическая реализация алгоритма (6.69) при $\gamma(t)$ вида (6.74) невозможна. Однако в реальных условиях практической идентификации неэффективно проводить рекуррентную процедуру поиска с большим числом итераций ("до конца") с целью получения состоятельных оценок. Чаще требуется обеспечить заданную точность оценок. Данное положение выте-

кает из практических требований удовлетворения заданным значениям нормы параметрических рассогласований.

При проведении (6.74) замены $\| r(t) \|^2$ произвольной положительной величиной δ такой, что $\| r(t_0) \|^2 > \delta$ норма параметрических рассогласований будет уменьшаться до определенного конечного момента времени $t_1 > 0$ при выполнении условия (6.73). Для случая $\| r(t_0) \|^2 \le \delta$ существует вероятность нарушения условия сходимости (6.73). Данные условия показывают, что существует такой процесс сходимости в определенную область и выхода из нее, когда точность оценок ограничена и определяется для нормы параметрических рассогласований некоторым заданным значениям $\delta$.

Очевидно, что для произвольно малой величины $\delta$ процесс достижения заданной области определенной точности будет монотонным для достаточно большого периода времени. Это обусловлено тем, что для малых значений $\delta$ величина $\gamma(t)$ также является малой. В связи с этим для повышения скорости сходимости и выполнения условия (6.73) выражения для $\gamma(t)$ алгоритма (6.69) можно представить как

$$\gamma(t) = \frac{\eta_1 \lambda R(t,\delta)}{\left[\lambda_n R(t,\delta) + \sigma_v^2\right] \| u_m \|^2}, \qquad (6.75)$$

где $R(t,\delta) = k_1 e^{-k_2 t} + \delta$ ; $k_1$, $k_2$ – некоторые положительные постоянные; $\delta$ - положительная малая величина; $\| u_m \|^2$ – наибольшее значение следа матрицы $u(t)u^T(t)$ для любого t. Тогда очевидно, что при $\| r(t) \|^2 > R(t, \delta)$, имеет место условие (6.73).

Таким образом, при выборе определенного значения $\delta$ и значения $\gamma(t)$ вида (6.75) можно получать оценки параметров заданной точности, удовлетворяющие практическим требованиям.

Для детерминированного случая, когда $v(t) = 0$, как следует из (6.72), условие сходимости алгоритма (6.69) будет иметь место, если значение $\gamma(t)$ находится в интервале

$$0 < \gamma(t) < 2\left[u(t)u^T(t)\right]^{-1}. \qquad (6.76)$$

### 6.5.3. Скорость сходимости центрированного процесса

Без потери общности определения скорости сходимости процедуры оценивания параметров объекта (6.49) с использованием алгоритма (6.68) положим, что $\nu(t)=0$, а входные сигналы являются центрированными случайными величинами и удовлетворяют условиям

$$M\left\{u_i(t)u_j(t-m)\right\}=\begin{cases}\sigma^2\delta_{ij}, & m=0,\\ 0, & m\neq 0,\end{cases} \tag{6.77}$$

где $\delta_{ij}$ – символ Кронекера.

Основные результаты оценивания вектора неизвестных параметров объекта (6.49) с применением алгоритма (6.68) при выполнении условий (6.77) приведем в виде утверждений. Более глубокий анализ итеративных процессов для данного случая представлен в [28].

Утверждение 3. Если входные сигналы удовлетворяют условиям (6.77), вектор параметров модели (6.58) настраивается согласно алгоритма (6.68), то $M\left\{\| r(t)\|^2\right\}\to 0$, когда $\gamma$ удовлетворяет условию

$$0<\gamma(t)<2\left[(n+2)\sigma^2\right]^{-1}. \tag{6.78}$$

Действительно, с учетом (6.73) при $\nu(t)=0$, взяв математическое ожидание от (6.72), получим

$$M\{\Delta V(t)\}=-\left[2-\gamma(n+2)\sigma^2\right]\gamma\sigma^2 M\left\{\| r(t-1)\|^2\right\}<0. \tag{6.79}$$

Из (6.79) следует, что для $\gamma$, удовлетворяющего условию (6.77) $M\{\Delta V(t)\}\to 0$.

Утверждение 4. При выборе

$$\gamma=\left[(n+2)\sigma^2\right]^{-1} \tag{6.80}$$

значение $M\{\Delta V(t)\}$ оптимально и скорость сходимости алгоритма (5.2.11) определяется выражением

$$M\left\{\| r(t)\|^2\right\}=\left(1-\frac{1}{(n+2)}\right)^t\| r(0)\|^2. \tag{6.81}$$

Оптимальное значение $\gamma_0$ для алгоритма (6.68) определяется из уравнения (6.79) как $\frac{d}{d\gamma}\left[2\gamma\sigma^2-(n+2)\gamma^2\sigma^4\right]=0,$ из которого следует (6.80).

Тогда с учетом выражения $M\{\Delta V(t)\}=M\left\{\| r(t)\|^2-\| r(t-1)\|^2\right\}$ уравнение (6.79) при $\gamma$ равном (6.80) запишем в следующем виде:

$$M\left\{ \| r(t) \|^2 \right\} = \left( 1 - \frac{1}{(n+2)} \right) M\left\{ \| r(t-1) \|^2 \right\}. \qquad (6.82)$$

Исключая промежуточные значения $M\{ \| r(k) \|^2 \}$, $k = \overline{1,\ t-1}$, получим зависимость среднего значения квадрату нормы параметрических рассогласований после *t* шагов от начальной ошибки в виде (6.81).

Аналогичные условия можно представить и для алгоритмов (6.66), (6.10). Так например, для алгоритма (6.66) при выполнении условий (6.77), когда $\alpha = 0,\ \nu(t) = 0,$, скорость сходимости процесса оценивания максимальна и средняя ошибка после t экспериментов представлена как

$$M\left\{ \| r(t) \|^2 \right\} = \left( 1 - \frac{1}{n} \right)^t \| r(0) \|^2. \qquad (6.83)$$

Из (6.81), (6.83) следует, что с увеличением числа оцениваемых параметров и повышение точности оценок значительно возрастает время наблюдения за объектом идентификации с целью получения оценок заданной точности. При этом эффективность применения отмеченных алгоритмов по скорости сходимости при оценивании вектора параметров высокой размерности становится одинаковой. Это подчеркивает известное положение [20, 21] о том, что с увеличением сложности объекта необходимо применять более простые алгоритмы для решения задач идентификации, поскольку скорость сходимости известных алгоритмов идентификации выравнивается и определяющими становятся вопросы практической реализуемости.

## 6.6. Оценивание параметров и состояний модели объекта

Приведем начальные сведения о решении задач оценивания параметров и состояния модели объекта в контексте адаптивного управления. Особенностью таких задач является их решение в реальном времени: в темпе формирования управляющих воздействий и с использованием текущей измерительной информации без предварительного накопления. Эти два требования ограничивают применение ряда известных в теории идентификации методов, в частности методов, которые используют явные математические выражения, приводят к необходимому накоплению информации и тем самым позволяют решать системы уравнений в результате выполнения конечного числа элементарных вычислительных операций.

Для вычисления оценок параметров и состояния модели объекта в замкнутом контуре управления предпочтительнее использовать методы идентификации на основе настраиваемой модели. Эти методы основаны на использовании априорной информации о структуре объекта, начальных значениях его параметров, а уточнение этой информации в виде адекватных оценок осуществляется итеративными методами с применением текущих измерений переменных состояния или входа-выхода объекта управления.

Рассматриваемые далее настраиваемые модели предназначены для вычисления оценок параметров $\hat{\theta}(t), \theta(t)$ объекта, заданного дифференциальным или разностным уравнениями вход-выход.

Оценивание состояния (векторы состояния объекта управления) представляет собой самостоятельную проблему. Для оценивания состояния детерминированного объекта используются асимпточеские идентификаторы Калмана и Люенбегера. Вычисление вектора состояния объекта в случайной среде основано на теории фильтрации Калмана. В целом проблеме оценивания параметров и (или) состояния посвящена обширная литература, в частности публикации [4, 39, 41].

Адаптивное оценивание параметров уравнений вход-выход.

Систему идентификации будем называть адаптивной, если она обеспечивает за конечное время достижение цели идентификации для любых объектов из заданного класса и при произвольных начальных условиях. Схема адаптивной идентификации иллюстрируется рис. 6.11. где НМ – настраиваемая модель, формирующая сигнал обобщенной ошибки *e(t)* на основе входного *u (t)* и выходного *y (t)* сигналов объекта управления: БНМ – блок настройки модели, входом которой служит сигнал рассогласования $e(t)=y(t)-\hat{y}(t)$. Блок настройки модели решает задачу оптимизации неизвестных параметров $\theta$ в соответствии с выбранной подходящей оценочной функцией *Q (e(t), t)* на основе оценки $\hat{\theta}(t)$. Предполагается, что *Q(.)* выбрано так, что ее минимум достигается при условии параметрического рассогласования $\left\|\hat{\theta}(t)-\theta\right\| \to 0$ при $t>t^*$, $t \geq 0$ и интервал времени (интервал оценивания) *[0. $t^*$]* конечен. При неудачном выборе оценочной функции ее минимум достигается при $\left\|\hat{\theta}(t)-\theta\right\| \neq 0$; более того, истинное значение оцениваемых параметров $\theta$ может быть далеким от оценки $\hat{\theta}^*$, доставляющей *min Q*.

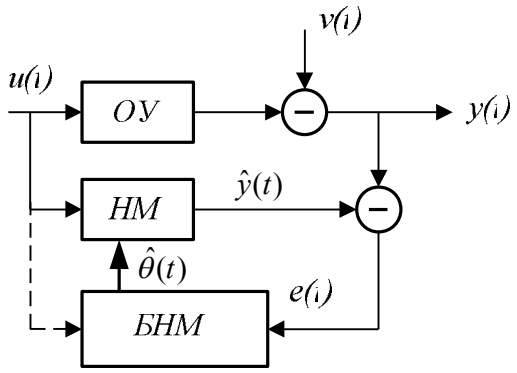



Рис. 6.11. Схема адаптивного оценивания параметров объекта управления

В этом случае оценка $\hat{\theta}^*$ – не единственная, а многообразная. Чем большее число оцениваемых параметров вычисляется, тем больше вероятность, что результаты оценивания как раз и будут многообразием. Это не имеет существенного значения, если для оптимального решения задач идентификации используются не оценки параметров $\hat{\theta}(t)$, а оценки переходных процессов настраиваемой модели $\hat{y}(t)$, для которых $|e(t)| = |y(t) - \hat{y}(t)| \xrightarrow{t\to\infty} 0$ и которые доставляют *min Q*. Но если реализации оптимального оценивания адаптивная система нуждается в оценках неизвестных текущих параметров модели объекта, $\hat{\theta}(t)$, то необходимо принять меры, исключающие не единственность решения задачи оценивания в зависимости от начальных условий процесса оценивания.

В качестве настраиваемой модели часто используется обобщенная модель объекта управления. Ее достоинство в том, что обобщенная ошибка $e(t)$ вычисления оценок параметров модели $\hat{\theta}$ является линейной функцией относительной этих оценок. При этом требование линейности связи между входами и выходами модели может не выдвигаться и не соблюдаться.

В структуре, приведенной на рис.6.12, выражение для обобщенной ошибки имеет вид

$$e(t) = \sum_{t=1}^{n} \alpha_i y_i + \sum_{j=1}^{m} \beta_i u_i ,$$

здесь $e(t)$ линейна по отношению к $\{\alpha_i\},\{\beta_i\}$, но операторы $F_j$, $G_i$ могут быть нелинейными. В случае использования обобщенной модели

выполняются равенства $\partial e(t)/\partial\alpha_i = y_i$, $\partial e(t)/\partial\beta_j = u_j$, т. е. коэффициенты чувствительности входят в обобщенную модель в явном виде.

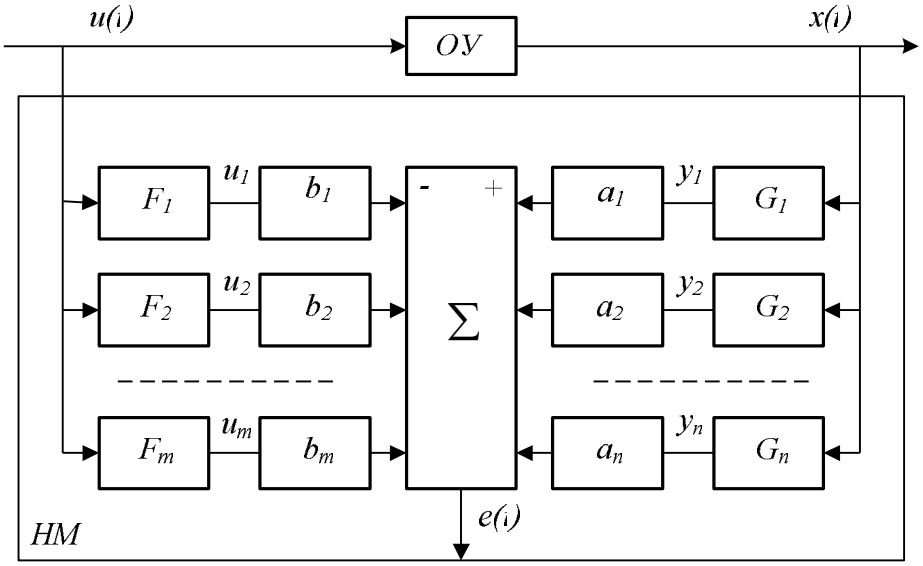



Рис. 6.12. Схема модели объекта управления

Примем $\alpha_1 = 1_3 . G_i$=1. $y_1$=y, тогда обобщенная ошибка, вычисленная в момент $t = k\cdot\Delta t$ равна

$$e(t) = e(k\cdot\Delta t) = e[k] = y[k] = \sum_{i=1}^{n}\alpha_i[k] + \sum_{j=1}^{m}\beta_i u[k]$$

или в векторной форме

$$e(k) = y(k) = y_k^T\alpha - y_k^T\beta = y[k] - [u_k^T : y_k][\frac{\beta}{-\alpha}] = y[k] - z_k^T\theta,$$

здесь $[\beta : -\alpha]^T = [\{\beta\},\{-\alpha_i\}]$ – вектор оценок параметров обобщенной настраиваемой модели; $u_k = [u_1(k),\ldots\ldots\ldots,u_m(k)]$ – вектор, составленный из m значений входного сигнала объекта u(t), $t=k\cdot\Delta t$, преобразованного операторами $F_1,F_2.....,F_m$; $y_к=[y_1(k),y_2(k),...,y_n(k)]^T$ – вектор, составленный из n значений выходного сигнала объекта $y(t)$, $t=k\Delta t$, преобразованного операторами $G_1,G_2,.....G_n$.

В случае описания модели объекта разностным уравнением

$$a_n x_{(k-n)}+a_{n-1}x_{(k-n-1)}+\ldots+x_k=b_m u_{k-m}+b_{m-1}u_{k-m-1}+\ldots+b_1 u_{k-1}$$

с применением оператора сдвига $z^{-1}\{y_к\}=y_{к-1}$ это уравнение преобразуется к виду

$$(a_n z^{-n}+a_{n-1}z^{-n+1}+\ldots+a_1 z^{-1}+1)y_k = (b_m z^{-m}+b_{m-1}z^{-m+1}+\ldots.+b_1 z^{-1})u_k.$$

В этом случае функции G и F есть операторы сдвига $z^{-1}$. Вектор оценок параметров $\theta = (\beta_1, \beta_2, ..., \beta_m; \alpha_1, \alpha_2, ..., \alpha_n)$ состоит из оценок коэффициентов разностного уравнения $b_j, a_j$, а вектор измерения $z_k$ включает переменные на входе и выходе идентифицируемого объекта управления:

$$z^T{}_k = (u_{k-1}, u_{k-2}, \ldots u_{k-m}; -y_{k-1}, -y_{k-2}, \ldots, -y_{k-n})^T.$$

Несомненными достоинствами обобщенной настраиваемой модели и ее модификации с практической точки зрения является линейность по настраиваемым параметрам и возможность использования априорной информации весьма общего вида (например, «объект приближенно линеен» или известен интервал времени затухания импульсной переходной функции). Однако обобщенная настраиваемая модель может приводить к оценкам в виде многообразий, а не к единственному решению в процессе настройки при различных начальных значениях оценок $\alpha_j^0, \beta_j^0$.

Это свойство усиливается с ростом числа оцениваемых параметров и в случаях, когда динамические операторы *G, F* есть операторы сдвига $z^{-1}$ или линейные фильтры в виде типовых динамических звеньев *G (s) F(s)*. Этого можно избежать, если операторы $F_j$ – ортогональные фильтры Лагерра, Мейкснера, а операторы $G_i=0, i>0$, $G_0=1, \alpha_0 = 1$. Использование взамен элементов задержки или типовых линейных звеньев упомянутых ортогональных фильтров расширяет возможности обобщенной модели, но одновременно порождает другую проблему: возникает необходимость решения системы нелинейных уравнений в блоке настройки модели при пересчете оценок настраиваемой модели в оценки параметров принятой математической модели объекта управления.

Настраиваемые модели в адаптивных системах рассмотренных ранее типов представляют собой частный случай обобщенной настраиваемой модели, где $G_1=0$, $\alpha_0 = 1$, а операторы $G_i=0$, $i>0$. Обобщенная схема такой модификации настраиваемой модели была дана на рис. 6.11. Реальный объект и его модель возбуждаются одним и тем же управляющим сигналом в замкнутом контуре, а ошибка настройки определяется как разности сигналов на выходах объекта и модели.

Структура обобщенной настраиваемой модели может быть распространена на случаи, когда объект функционирует при случайных воздействиях, влияние которых может быть промоделировано экви-

валентным гауссовым процессом (последовательностью). При этом моделью гауссова эквивалентного возмущения служит формирующий устойчивый динамический фильтр, возбуждаемый гауссовым процессом типа белого шума. Тогда можно составить обобщенную настраиваемую модель, где оцениваются параметры как собственно объекта, так и устойчивого формирующего фильтра. Ниже приводится теорема об оптимальной настраиваемой модели [38] по невязке $e_k = y_k - \hat{y}_k$ с использованием оценочной функции $J = M\{e_k^2\}$.

Теорема: Для минимально-фазового объекта с уравнением

$$A(z^{-1})y_k = B(z^{-1})u_k + C(z^{-1})\xi_k \qquad (6.84)$$

оптимальная обобщенная настраиваемая модель определяется в неявной форме уравнением

$$C(z^{-1})\hat{y}_k = B(z^{-1})u_{k-1} + [C(z^{-1}) - A(z^{-1})]y_k; \qquad (6.85)$$

в явной форме – рекуррентным уравнением

$$\hat{y}_k = B(z^{-1})u_k - [A(z^{-1}) - 1]y_k + [C(z^{-1}) - 1]e_к, \qquad (6.86)$$

где полиномы A ($z^{-1}$), B($z^{-1}$) и C ($z^{-1}$) заданы в виде

$$A(z^{-1}) = 1 + \sum_{j=1}^{m}\alpha_j z^{-j}; \quad B(z^{-1}) = \sum_{j=0}^{m\le n} b_j z^{-j}; \quad C(z^{-1}) = 1 + \sum_{j=1}^{m} c_j z^{-j}. \qquad (6.87)$$

Для оптимальной настраиваемой модели (6.85) или (6.86) ошибка настройки $e_к = \xi_k$ и $J^* = min_u M\{e_k^2\} = \sigma\frac{2}{\xi}$, где $\xi_k$ чисто случайная последовательность (гауссов белый шум).

Доказательство теоремы следует из условий «выбеливания» ряда $C(z^{-1})e_к = e_к + \sum_{j=1}^{l} c_j e_{k-j}$, т. е. обращение его в ряд $C(z^{-1})\xi_k$

Для этого проведем ряд преобразований:

$$C(z^{-1})e_к = [C(z^{-1}) + A(z^{-1}) - A(z^{-1})](y_k - \hat{y}_k) =$$
$$= A(z^{-1})y_k + [C(z^{-1}) - A(z^{-1})]y_k - [C(z^{-1}) + A(z^{-1}) - A(z^{-1})]\hat{y}_k =$$
$$B(z^{-1})u_k + C(z^{-1})\xi_k + [C(z^{-1}) - A(z^{-1})]y_k - C(z^{-1})\hat{y}_k.$$

Очевидно, что для обращения ряда $C(z^{-1})e_к$ в ряд $C(z^{-1})\xi_k$ необходимо и достаточно, чтобы $C(z^{-1})\hat{y}_k = B(z^{-1})u_k + [C(z^{-1}) - A(z^{-1})]y_k$, что соответствует утверждению теоремы (6.85) о неявной форме оптимальной настраиваемой модели.

Модель в явной форме, т. е. в виде рекуррентного уравнения, получается если к обеим частям (6.86) прибавить - $e_к = y_к - y_k$:

$$\hat{y}_{k.} - y_k + C(z^{-1})\,\hat{y}_{k.} = \hat{y}_{k.} - y_k + B(z^{-1})u_k + [C(z^{-1}) - A(z^{-1})]\,y_k,$$

откуда следует

$$\hat{y}_k = B(z^{-1})u_k + C(z^{-1})(\hat{y}_{k.} - y_k) - (\hat{y}_{k.} - y_k)[A(z^{-1}) - 1]\,y_k =$$
$$= B(z^{-1})u_k - [A(z^{-1}) - 1]\,y_k + [C(z^{-1}) - 1]e_к,$$

что соответствует соотношению (6.86) теоремы.

Уравнения (6.85), (6.86) задают обобщенную структуру оптимальной настраиваемой модели, если оценки параметров полиномов есть $\hat{A}(z^{-1}) \equiv A(z^{-1}),\ \hat{B}(z^{-1}) \equiv B(z^{-1}), \hat{C}(z^{-1}) \equiv C(z^{-1})$. Линейная относительно настраиваемых параметров $\hat{\theta}$ модель в векторной форме имеет вид

$$\hat{y}_k = \hat{\theta}^T(k^*\Delta t)z_k; \qquad (6.88)$$
$$\hat{\theta}_0 = \theta(0) \text{ задан.}$$

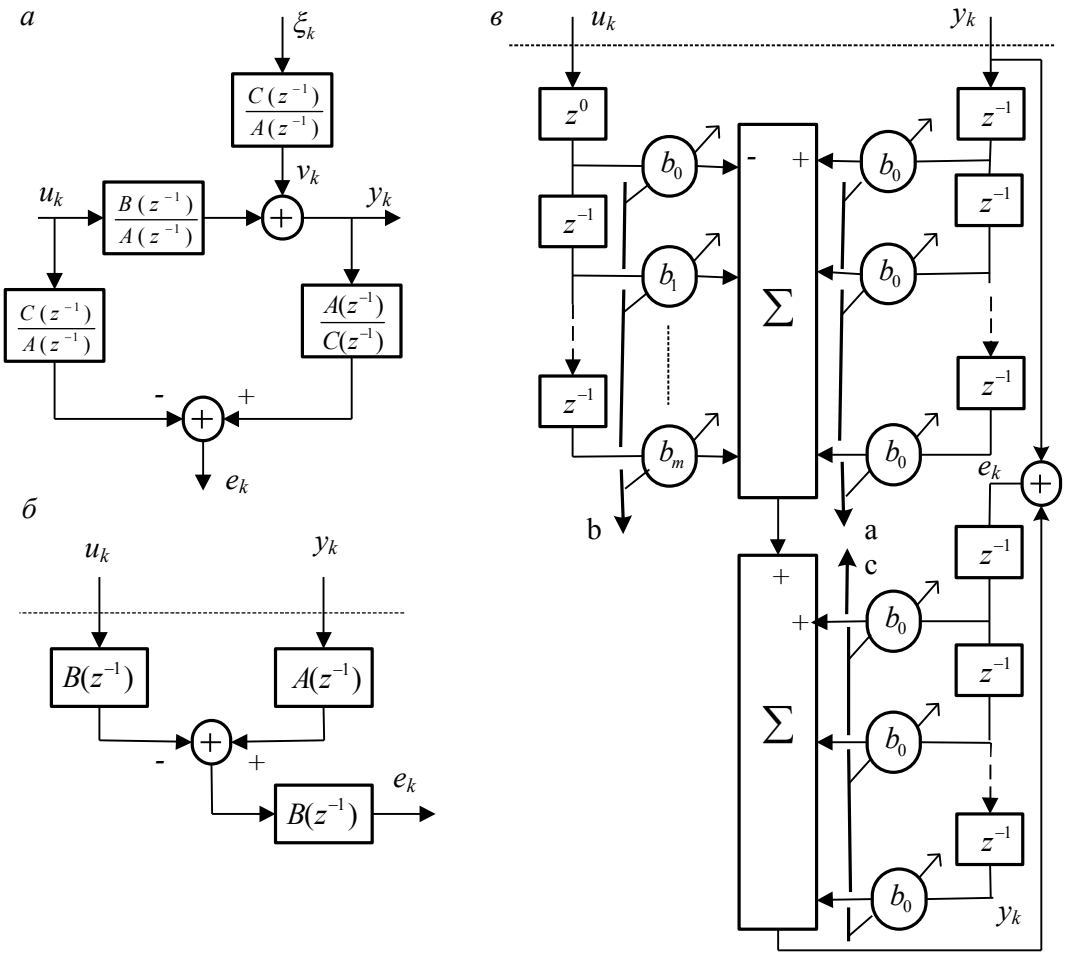



Рис. 6.13. Схема оптимальной обобщенной линейной настраиваемой модели объекта управления

*Схемы построения в соответствии с уравнениями (6.85) (а, б) и (6.86) (в)*

На рис. 6.13 изображены структурные схемы обобщенной оптимальной настраиваемой модели объекта, составленной в соответствии с уравнениями (6.85) (*а, б*) и (6.86)(*в*).

Идентификаторы состояния [4]. Модели на основе идентификаторов состояния («наблюдателей») необходимы для получения оценок $\hat{x}(t)$ вектора состояния линейного объекта $x(t)$ в основном контуре адаптивной системы. Для реализации настраиваемых моделей в этих случаях используются измерения входных и выходных переменных объекта управления и оценки матриц уравнения состояния объекта $\hat{A},\hat{B}$. Предполагается, что объект наблюдаемый и идентифицируемый. Физический смысл термина наблюдаемый состоит в возможности наблюдать (измерять) все собственные движения уравнения состояния объекта при нулевых начальных условиях. В этом смысле термин идентифицируемый обозначает возможность возбуждать собственные движения систем.

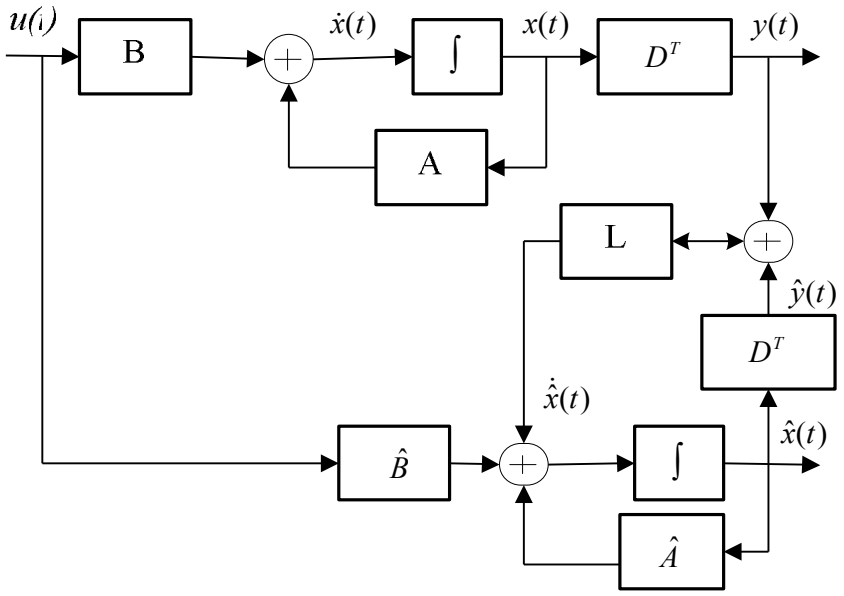



Рис. 6.14. Структурная схема асимптотического идентификатора состояния для системы с одним выходом

Структурная схема асимптотического идентификатора (без блока настройки) одномерного объекта изображена на рис 6.14 В этой схеме выходной параметр объекта $y(t)=D^Tx(t)$ сравнивается с параметром на выходе настраиваемой модели $\hat{y}(t)=D^T\hat{x}(t)$. Разность

$y(t)-\hat{y}(t)=y(t)-D^T\hat{x}(t)$. Умножается на вектор $L$ и подаются модели состояния $(\hat{A},\hat{B},D^T)$ в качестве корректирующего воздействия.

Вектор $L$ влияет на динамику процесса $[x(t)-\hat{x}(t)]\to 0$ при $t\to\infty$. Говорят, что порядок идентификатора равен $n$ ($n$ –размерность вектора состояния $x(t)$). Динамическое уравнение такого идентификатора имеет вид

$$\dot{\hat{x}}(t)=A\hat{x}(t)+L\,[y(t)-D^T\hat{x}\,(t)]+Bu(t) \qquad (6.89)$$

или

$$\dot{\hat{x}}(t)=[A-L\,D^T]\hat{x}(t)+L\,y(t)+Bu(t). \qquad (6.90)$$

Если ввести новую переменную $e(t)=\hat{x}(t)-x(t)$, то уравнение (6.90) примет следующий вид:

$$\dot{e}(t)=[A-L\,D^T]\,e(t)=\tilde{A}\,e(t). \qquad (6.91)$$

Структура уравнения (6.91) подсказывает направление синтеза необходимого корректирующего вектора $L$: матрица $\tilde{A}=A-LD^T$ должна иметь отрицательные вещественные части собственных чисел.

Асимптотический идентификатор (6.89) имеет ту же равномерность, что и оцениваемый вектор состояния. Такой идентификатор называется наблюдателем полного порядка или наблюдателем Калмана – по имени американского математика Р. Калмана, первым предложившего такой способ оценки вектора состояния линейных систем.

Особенностью наблюдателя полного порядка является ограничение свободы выбора корректирующего вектора $L$ только условиями устойчивости. Этим обстоятельством можно воспользоваться для понижения равномерности асимптотического идентификатора по крайней мере на единицу. Параметр на выходе объекта $y(t)$ совпадает с компонентой $x_n(t)$ оцениваемого вектора состояния $x(t$. Примем $\hat{x}_n(t)=y(t)=x_n(t)$, и подставим его в уравнение идентификатора (6.89). Тогда система (6.89) вырождается в систему линейных дифференциальных уравнений $(n-1)$-го порядка, т. к. последнее уравнение в системе (6.89) заменяется на тождество $\hat{x}_n(t)=y(t)$ и $\dot{\hat{x}}(t)=\overline{A}\hat{x}(t)+\overline{a}_n\,y(t)+\overline{B}\,u(t)$.

Здесь $\hat{x}=col\,(\hat{x}_1,\ldots\ldots\,\hat{x}_{n-1})$; $\overline{A}$ – подматрица равномерностью $(n-1)$x$(n-1)$ в составе матрицы A; $\overline{a}$ – вектор-столбец – последний столбец в матрице A; вектор $\overline{B}$ получается из прежнего B удалением из него последней компоненты $b_n$.

Введем вектор-ошибку $\tilde{e}(t) = \bar{x}(t) - \hat{x}(t)$ равномерностью $(n-1)$x$1$. В новых переменных векторное уравнение идентификатора примет следующий вид:

$$\dot{\bar{x}}(t) - \dot{\tilde{e}}(t) = \bar{A}[\bar{x}(t) - \tilde{e}(t)] + \bar{a}_n y(t) + \bar{B} u(t). \qquad (6.92)$$

Вектор состояния $\bar{x}(t)$ идентифицируемого объекта без компоненты $x_n(t)$ удовлетворяет системе из (n-1) уравнений:

$$\bar{x}(t) = \bar{A}\bar{x}(t) + \bar{a}_n x_n(t) + \bar{B} u(t). \qquad (6.93)$$

Уравнение (6.99) соответствует первым $(n-1)$ уравнениям и исходной системы уравнений объекта

$$\dot{\bar{x}}(t) = A\, x(t) + B\, u(t).$$

Вычтем из уравнения (6.93) уравнение (6.92). Тогда уравнение относительно $(n-1)$-мерного вектора-ошибки $\tilde{e}$(t) может быть представлено в виде

$$\dot{\tilde{e}}(t) = \bar{A}\tilde{e}(t). \qquad (6.94)$$

В этом случае проблема построения асимптотически устойчивого идентификатора состояния порядка $(n-1)$ сводится к построению неособого преобразования P, которое обеспечивает желаемый набор характеристических чисел у подматрицы $\bar{A}$ размерностью $(n-1)$x$(n-1)$ в составе матрицы $A$. В результате такого преобразовании характеристические числа матрицы $A$ не изменяются. Доказательство о возможности построения идентификатора порядка $(n-1)$ для идентифицируемой линейной системы размерностью n и примеры неособого преобразования $P$, которое обеспечивает желаемый набор характеристических чисел у подматрицы $\bar{A}$, приводятся в публикации [4].

Идентификаторы равномерностью $(n-1)$ и (в более общем случае) $(n-p)$-мерные ($p$ – число линейно независимых выходов объектов управления) получили название идентификаторов (наблюдателей) Люенбергера – по имени американского ученого, изучавшего свойства таких устройств.

Алгоритм построения $(n-1)$-мерного идентификатора Люенбергера для вычисления оценок вектора состояния $\hat{x}(t)$ линейной системы с одним и более выходами приведен в публикации [4].

Наряду с традиционными идентификаторами состояния укажем на обучаемые динамические многослойные нейронные сети как на весьма эффективный инструмент для вычисления вектора состояния при минимальной априорной информации и в случаях исследования нелинейных объектов.

# ГЛАВА 7. МЕТОДЫ ИДЕНТИФИКАЦИИ НЕЛИНЕЙНЫХ СИСТЕМ УПРАВЛЕНИЯ

В изложении методов идентификации неоднократно отмечается, что *широкий класс нелинейных систем описывается рядами Вольтерра.* Однородный оператор, действующий в С [О, *Т*], может быть записан в виде [12]

$$x(t) = \hat{O}_N(u(t)) = \int_0^T \ldots \int_0^T K_N(t,\tau_1,\ldots,\tau_N) \prod_{i=1}^{N} u(\tau_i) d\tau_i, \quad t \in [o,T]. \qquad (7.1)$$

Подобно тому, как обычные полиномы используются для аппроксимации непрерывных функций, функциональные полиномы используются для аппроксимации операторов, описывающих нелинейные системы. *Нелинейную систему, оператор которой может быть представлен функциональным полиномом, называют полиномиальной.*

Оператор системы Ф*(u(t))*, ряд которого, образованный полиномами Вольтерра, равномерно сходится к оператору

$$\hat{O}(u(t)) = \sum_{n=1}^{\infty} \hat{O}_n(u(t)),$$

называется аналитическим, а САУ – аналитической.

Важным является следующее обстоятельство: *если объект представляет собой «черный ящик», структура которого неизвестна, то при определенных допущениях для него может быть построена процедура идентификации, позволяющая приближенно описать его с помощью ряда Вольтерра* [12].

Функциональные ряды нашли применение при решении задач детерминированного и статистического анализа, частотного анализа, анализа устойчивости, синтеза нелинейных систем (при синтезе нелинейных систем может быть применен тот же подход, что и при синтезе линейных систем: требования, предъявляемые к системе, формализуются с помощью эталонного нелинейного оператора системы, после чего решается задача приближения фактического оператора к эталонному путем изменения параметров и структуры регулятора), статистической оптимизации.

Для решения задач анализа, синтеза и оптимизации систем, описываемых функциональными рядами, ключевой является проблема идентификации динамических характеристик объекта. Некоторые методы решения этой проблемы обсуждаются в настоящем параграфе.

## 7.1. Идентификация нелинейных объектов управления корреляционным методом

Интенсивное развитие средств вычислительной техники создало предпосылки для разработки новых методов идентификации, ориентированных на использование ЭВМ. Метод Винера, как указывалось в предыдущем параграфе, имеет в основном теоретическое значение и встречает большие трудности при решении инженерных задач.

Вместе с тем нашли широкое практическое применение статистические методы идентификации линейных объектов.

*Основой при решении задач статистической идентификации линейных стационарных объектов является уравнение Винера-Хопфа*

$$R_{xu}(\sigma) = \int_0^\infty k(\tau) R_{uu}(\sigma - \tau) d\tau. \qquad (7.2)$$

В частном случае, когда входной сигнал $u(t)$ представляет белый шум, определение ИПФ $k(\tau)$ исследуемого объекта упрощается.

Действительно, в этом случае соотношение (7.2), с учетом, что для белого шума корреляционная функция имеет вид $R_{uu}(\sigma) = 2\pi S_0 \delta(\sigma)$, можно записать так:

$$R_{uu}(\sigma) = \int_0^\infty k(\tau) 2\pi S_0 \delta(\sigma - \tau) dt = 2\pi S_0 k(\sigma), \qquad (7.3)$$

где $S_0$ – интенсивность белого шума.

Тогда ИПФ идентифицируемого объекта определяется через взаимную корреляционную функцию по формуле

$$k(\sigma) = \frac{1}{2\pi S_0} R_{xu}(\sigma). \qquad (7.4)$$

Структурная схема экспериментального определения функции $k(\sigma)$ представлена на рис. 7.1.

В соответствии с формулой (7.4) *ИПФ идентифицируемого линейного объекта с точностью до постоянного множителя равна взаимной корреляционной функции входного и выходного случайных сигналов объекта.*

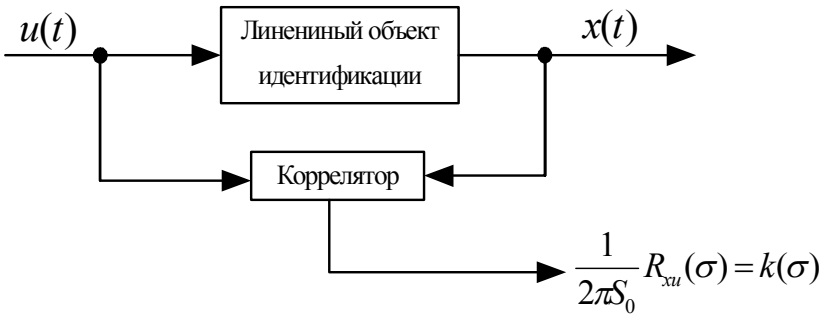



Рис. 7.1. Структурная схема идентификатора

Рассмотрим вопрос технической реализации системы идентификации. Соответствующая структурная схема представлена на рис. 7.2.

При построении идентификатора, представленного на рис. 7.2, использовалась формула, определяющая оценку взаимной корреляционной функции центрированных стационарных эргодических случайных процессов $u(t)$ и $x(t)$,

$$R_{xu}(-\sigma_i) = \frac{1}{T-\sigma_i}\int_{\sigma_i}^{T} u(t)x(t-\sigma_i)dt, \qquad (7.5)$$

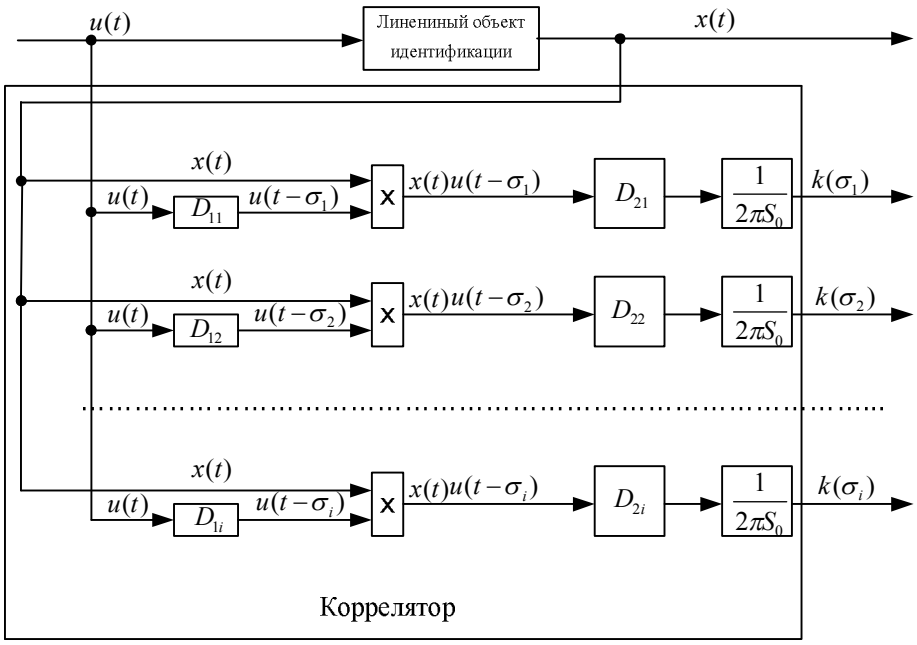



Рис. 7.2. Структурная схема идентификатора:

$D_{1i}$ – *звено с запаздыванием* $\sigma_i$; $D_{2i} = \dfrac{1}{T-\sigma_i}\displaystyle\int_{\sigma_i}^{T} x(t)u(t-\sigma_i)dt$

а также зависимость

$$R_{xu}(-\sigma_i) = R_{ux}(+\sigma_i). \qquad (7.6)$$

Этот подход, приводящий к простому решению задачи идентификации линейных объектов, Ю.В. Ли и М. Шетценом распространен на нелинейные объекты и более приемлем при решении инженерных задач, чем метод Винера [20].

С целью изложения основного содержания метода корреляции введем в рассмотрение *понятие G-функционалов.*

Выход нелинейного объекта можно представить в виде

$$x(t) = \sum_{n=0}^{\infty} G_n[k_n, u(t)], \tag{7.7}$$

где $\{\kappa_n\}$ – последовательность ядер Вольтерра; $\{G_n\}$ – полный ряд ортогональных функционалов.

Для $G$-функционалов справедливо свойство ортогональности

$$\overline{G_n[k_n, u(t)]G_m[k_m, u(t)]} = 0 \quad \text{при} \quad m \neq n, \tag{7.8}$$

где черта сверху означает осреднение по времени.

Для первых четырех $G$-функционалов справедливы зависимости

$$G_0[k_0, u(t)] = k_0;\ G_1[k_1, u(t)] = \int_0^{\infty} k_1(\tau_1)u(t-\tau_1)d\tau_1;$$

$$G_2[k_2, u(t)] = \int_0^{\infty}\int_0^{\infty} k_2(\tau_1, \tau_2)u(t-\tau_1)u(t-\tau_2)d\tau_1 d\tau_2 - 2\pi S_0 \int_0^{\infty} k_2(\tau_1, \tau_2)d\tau_2; \tag{7.9}$$

$$G_3[k_3, u(t)] = \int_0^{\infty}\int_0^{\infty}\int_0^{\infty} k_3(\tau_1, \tau_2, \tau_3)u(t-\tau_1)u(t-\tau_2)u(t-\tau_3)d\tau_1 d\tau_2 d\tau_3 -$$

$$- 6\pi S_0 \int_0^{\infty}\int_0^{\infty} k_3(\tau_1, \tau_2, \tau_3)u(t-\tau_1)d\tau_1 d\tau_2.$$

Рассмотрим задачу определения ядер произвольного порядка, для чего *введем в рассмотрение понятие белого гауссова шума с запаздыванием* (одномерным, двухмерным, $n$-мерным) [20].

Если на вход цепи с запаздыванием $\sigma$ подать белый шум, то выходной сигнал будет иметь вид

$$x_1(t) = u(t-\sigma) = \int_0^{\infty} \delta(\tau - \sigma)u(t-\tau)d\tau. \tag{7.10}$$

Аналогичную зависимость можно записать для двухмерного белого гауссова процесса с запаздыванием

$$x_1(t) = u(t-\sigma_1)u(t-\sigma_2) = \int_0^{\infty}\int_0^{\infty} \delta(\tau_1 - \sigma_1)\delta(\tau_2 - \sigma_2)u(t-\tau_1)u(t-\tau_2)d\tau_1 d\tau_2. \tag{7.11}$$

Структурные схемы соответствующих цепочек представлены на рис. 7.3, *а* и *б*.

Для $x_3(t)$ имеет место зависимость

$$x_3(t) = u(t-\sigma_1)u(t-\sigma_2)u(t-\sigma_3) =$$

$$= \int_0^\infty \int_0^\infty \int_0^\infty \delta(\tau_1-\sigma_1)\delta(\tau_2-\sigma_2)\delta(\tau_3-\sigma_3)u(t-\tau_1)u(t-\tau_2)u(t-\tau_3)d\tau_1 d\tau_2 d\tau_3. \quad (7.12)$$

Для общего случая можно записать выражение

$$x_n(t) = u(t-\sigma_1)u(t-\sigma_2)\ldots u(t-\sigma_n) =$$

$$= \int_0^\infty \int_0^\infty \ldots \int_0^\infty \delta(\tau_1-\sigma_1)\delta(\tau_2-\sigma_2)\ldots\delta(\tau_n-\sigma_n)\times \quad (7.13)$$

$$\times u(t-\tau_1)u(t-\tau_2)\ldots u(t-\tau_n)d\tau_1 d\tau_2 \ldots d\tau_n.$$

Очевидно, ядро $k_0$ определяется формулой $k_0 = \overline{x(t)}$, т. е. равно среднему значению выходного сигнала объекта.

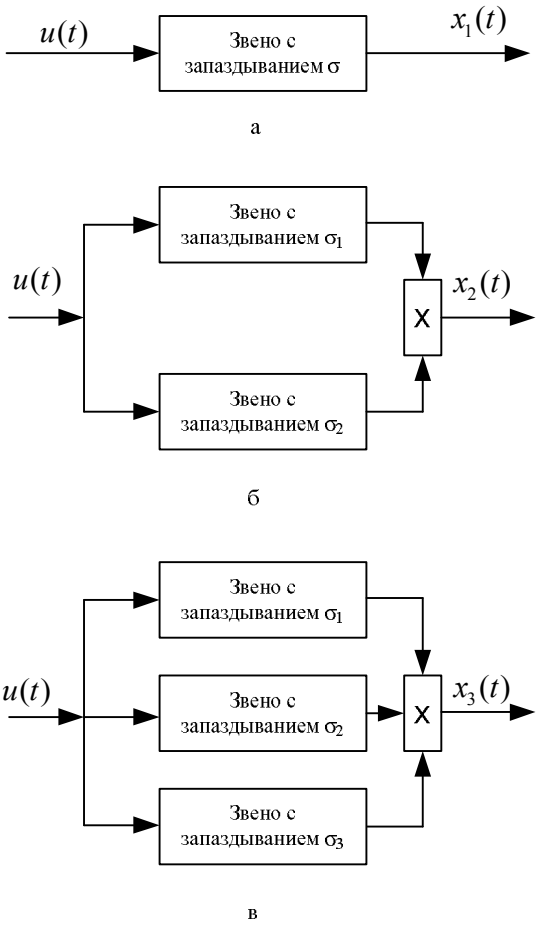



Рис. 7.3. Структурные схемы цепей с запаздыванием:
*а – одномерное запаздывание; б – двухмерное запаздывание;*
*в – трехмерное запаздывание*

Для определения $k_1(\sigma_1)$ воспользуемся устройством, структурная схема которого представлена на рис. 7.4.

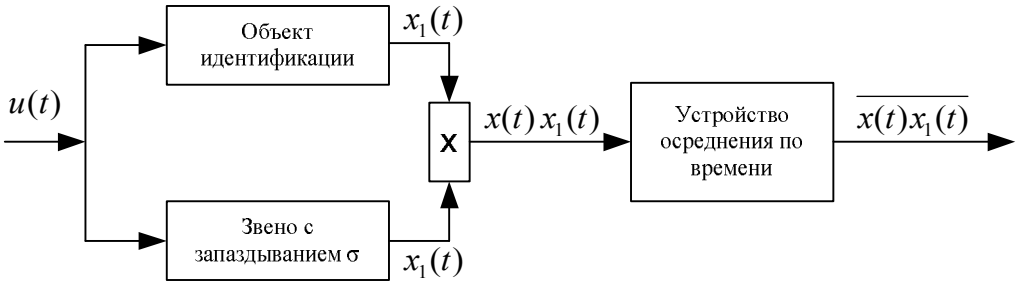



Рис. 7.4. Структурная схема идентификатора

Для системы, представленной на рис. 7.4, справедлива зависимость

$$\overline{x(t)x_1(t)} = \overline{\left\{\sum_{n=0}^{\infty} G_n[k_n, u(t)]\right\} u(t-\sigma)}. \qquad (7.14)$$

Поскольку $x_1(t) = x(t\text{-}\sigma)$ – функционал первого порядка, то, учитывая условия ортогональности (7.8), получим

$$\overline{G_n[k_n, u(t)]u(t-\sigma)} = 0 \quad \text{для всех} \quad n > 1. \qquad (7.15)$$

Для случая, когда $n = 0$, справедливо соотношение

$$\overline{G_0[k_0, u(t)]u(t-\sigma)} = \overline{k_0 u(t-\sigma)} = K\overline{u(t-\sigma)} = 0. \qquad (7.16)$$

Запишем формулу для случая, когда $n = 1$, получим

$$\overline{x(t)x_1(t)} = \overline{G_1[k_1, u(t)]x_1(t)} = \overline{\left(\int_0^{\infty} k_1(\tau_1)u(t-\tau_1)d\tau_1\right)u(t-\sigma)} = \qquad (7.17)$$

$$= \int_0^{\infty} k_1(\tau_1)\overline{u(t-\tau_1)u(t-\sigma)}d\tau_1 = \int_0^{\infty} k_1(\tau_1)2\pi S_0\delta(\sigma-\tau_1)d\tau_1 = 2\pi S_0 k_1(\sigma).$$

Отсюда следует *зависимость, позволяющая найти искомое ядро первого порядка:*

$$k_1(\sigma) = \frac{1}{2\pi S_0}\overline{x(t)u(t-\sigma)}. \qquad (7.18)$$

Если имеются цепочки с запаздываниями $\sigma_1, \sigma_2, ..., \sigma_i$, то по формуле (7.18) можно построить дискретные значения ядра первого порядка

$$k_1(\sigma) = \frac{1}{2\pi S_0}\overline{x(t)u(t-\sigma_i)}. \qquad (7.19)$$

Далее рассмотрим задачу определения ядра второго порядка $k_2(\sigma_1,\sigma_2)$, для чего воспользуемся устройством, структурная схема которого представлена на рис. 7.5.

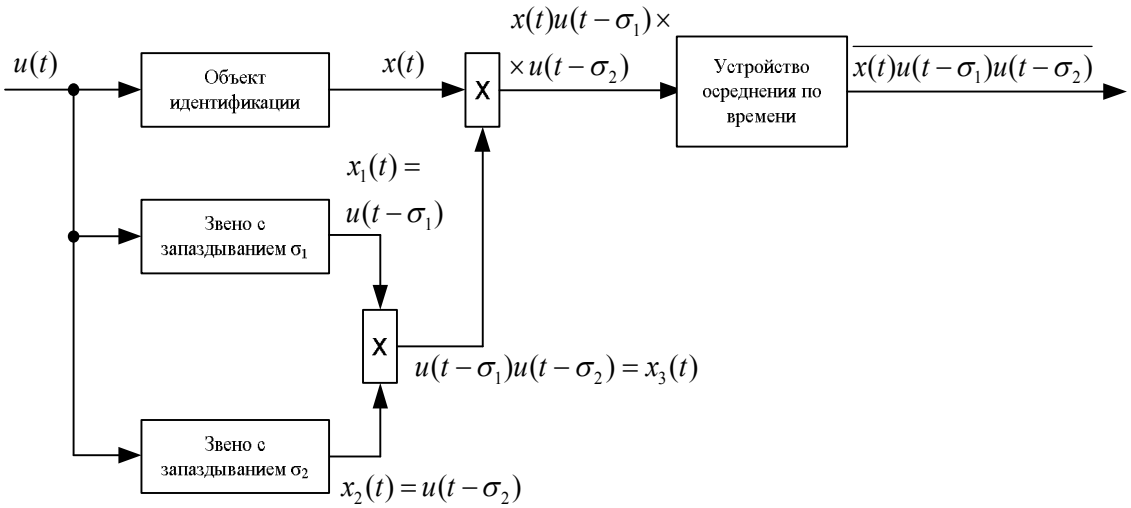



Рис. 7.5. Структурная схема идентификатора

Как и в предыдущем случае, имеем

$$\overline{x(t)x_3(t)} = \overline{x(t)u(t-\sigma_1)u(t-\sigma_2)} = \left\{\sum_{n=0}^{\infty} G_n[k_n,u(t)]\right\}u(t-\sigma_1)u(t-\sigma_2). \tag{7.20}$$

Причем

$$\overline{\left\{\sum_{n=0}^{\infty} G_n[k_n,u(t)]\right\}u(t-\sigma_1)u(t-\sigma_2)} = 0 \quad \text{для} \quad n>2. \tag{7.21}$$

Найдем выражение вида (7.21) для $n = 0$; получим

$$\overline{\left\{\sum_{n=0}^{\infty} G_n[k_n,u(t)]\right\}u(t-\sigma_1)u(t-\sigma_2)} = k_0\overline{u(t-\sigma_1)u(t-\sigma_2)} = k_0\pi S_0\delta(\sigma_1-\sigma_2). \tag{7.22}$$

Если же $n = 1$, то

$$\overline{\left\{\sum_{n=0}^{\infty} G_n[k_n,u(t)]\right\}u(t-\sigma_1)u(t-\sigma_2)} = \int_0^{\infty} k_1(\tau_1)\overline{u(t-\tau_1)u(t-\sigma_1)u(t-\sigma_2)}d\tau_1 = 0, \tag{7.23}$$

т. к. центральные моменты нечетного порядка гауссова белого шума равны нулю.

Далее запишем

$$\overline{G_2[k_2,u(t)]u(t-\sigma_1)u(t-\sigma_2)} =$$

$$= \int_0^\infty\int_0^\infty k_2(\tau_1,\tau_2)\overline{u(t-\tau_1)u(t-\tau_2)u(t-\sigma_1)u(t-\sigma_2)}d\tau_1 d\tau_2$$

$$- 2\pi S_0\int_0^\infty\int_0^\infty k_2(\tau_1,\tau_2)\overline{u(t-\sigma_1)u(t-\sigma_2)}d\tau_1 d\tau_2 =$$

$$= \int_0^\infty\int_0^\infty k_2(\tau_1,\tau_2)\Big[(2\pi S_0)^2\delta(\tau_1-\tau_2)\delta(\sigma_1-\sigma_2)+(2\pi S_0)^2\delta(\tau_1-\sigma_1)\delta(\tau_2-\sigma_2)+$$

$$+(2\pi S_0)^2\delta(\tau_1-\sigma_2)\delta(\tau_2-\sigma_1)\Big]d\tau_1 d\tau_2 - 2\pi S_0\int_0^\infty\int_0^\infty k_2(\tau_1,\tau_2)2\pi S_0\delta(\sigma_1-\sigma_2)d\tau_1 d\tau_2 = \quad (7.24)$$

$$= (2\pi S_0)^2\Bigg[\delta(\sigma_1-\sigma_2)\int_0^\infty k_2(\tau_1,\tau_2)d\tau_1 + k_2(\sigma_1,\sigma_2)+k_2(\sigma_2,\sigma_1)-$$

$$-\delta(\sigma_1-\sigma_2)\int_0^\infty k_2(\tau_2,\tau_2)d\tau_2\Bigg] = 2\cdot(2\pi S_0)^2 k_2(\sigma_1,\sigma_2) = 8\pi^2 S_0^2 k_2(\sigma_1,\sigma_2).$$

При выводе последнего соотношения использовалось свойство симметричности ядра $k_2(\tau_1,\tau_2)$.

Таким образом, находим

$$\overline{x(t)u(t-\sigma_1)u(t-\sigma_2)} = 8\pi^2 S_0^2 k_0\delta(\sigma_1-\sigma_2). \quad (7.25)$$

Из (7.25) следует

$$k_2(\sigma_1,\sigma_2) = \frac{1}{8\pi^2 S_0^2}\overline{x(t)u(t-\sigma_1)u(t-\sigma_2)} \quad \text{при} \quad \sigma_1 \neq \sigma_2. \quad (7.26)$$

*Полученная формула определяет ядро второго порядка* $k_2(\sigma_1,\sigma_2)$ *при конкретных значениях* $\sigma_1$ *и* $\sigma_2$. *Очевидно, для построения дискретных значений поверхности* $k_2(\sigma_1,\sigma_2)$ *необходимо взять соответствующее число звеньев задержки* (рис. 7.6).

При $\sigma_1 = \sigma_2$ имеет место дельта-функция. Пути исключения влияния дельта-функций на точность решения задачи идентификации указаны в [20].

Например, на практике всегда возможно к критической точке, имеющей место при $\sigma_1 = \sigma_2$, подойти на такое расстояние, которое исключило бы импульсное возмущение (рис. 7.6).

Рассуждая аналогичным образом, *для ядра третьего порядка можно получить зависимость* [20]

$$\overline{x(t)u(t-\sigma_1)u(t-\sigma_2)u(t-\sigma_3)} = 6(2\pi S_0)^3 k_3(\sigma_1,\ \sigma_2,\sigma_3) + \\ + (2\pi S_0)^2\left[\delta(\sigma_2-\sigma_3)k_1(\sigma_1)+\delta(\sigma_1-\sigma_3)k_1(\sigma_2)+\delta(\sigma_1-\sigma_2)k_1(\sigma_3)\right]. \quad (7.27)$$

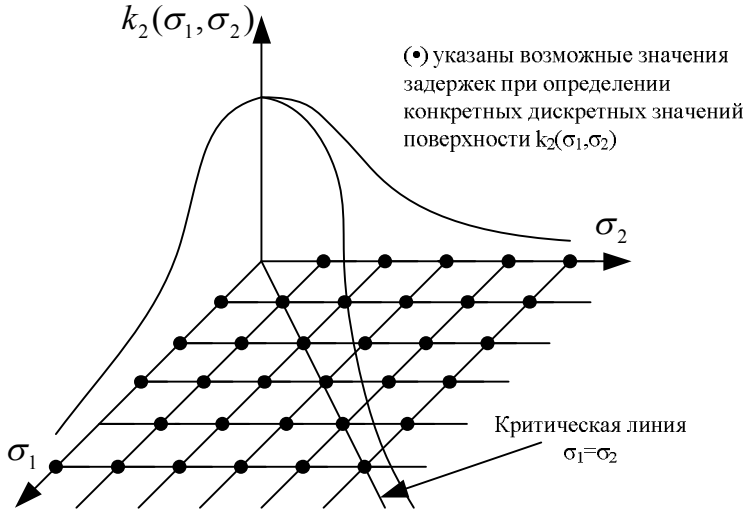



Рис. 7.6. ИПФ второго порядка

Отсюда находим $k_3(\sigma_1,\sigma_2,\sigma_3) = \dfrac{1}{48\pi^3 S_0^3}\overline{x(t)u(t-\sigma_1)u(t-\sigma_2)u(t-\sigma_3)}$ (7.28)

при $\sigma_1 \neq \sigma_2,\ \ \sigma_2 \neq \sigma_3,\ \ \sigma_3 \neq \sigma_1.$

Таким образом, *если на нелинейный объект и трехмерную цепь с запаздыванием подать белый гауссов шум, а затем произвести осреднение по времени произведения* $x(t)u(t-\sigma_1)u(t-\sigma_2)u(t-\sigma_3)$, *то ядро третьего порядка может быть выражено через центральный момент четвертого порядка*

$$\beta_{xu}^{(3)}(\sigma_1,\sigma_2,\sigma_3) = \overline{x(t)u(t-\sigma_1)u(t-\sigma_2)u(t-\sigma_3)}.$$

В [20] подробно *рассмотрены вопросы определения ядер n-го порядка, преодоления ограничений, а также расчета ядер для случая, когда на вход подается небелый гауссов сигнал.*

Рассмотренный метод корреляции более прост при технической реализации, чем метод Винера.

В отличие от метода Винера, позволяющего характеризовать ядра нелинейной системы коэффициентами Фурье разложения по функциям Лягерра (спектральной характеристикой в базисе функций

Лягерра), *метод взаимной корреляции* позволяет определить лишь дискретные значения ядер идентифицируемого объекта, по которым можно судить о свойствах объекта (например, о его инерционности).

Весьма важным преимуществом при решении инженерных задач методом корреляции является факт исключения ошибки аппроксимации и вопросов, связанных со сходимостью разложений и оценкой погрешности. В методе Винера центральным является вопрос: *сколько членов разложения по функциям Лягерра и полиномам Эрмита необходимо брать, чтобы обеспечить заданную точность при приемлемой степени сложности системы идентификации.*

Метод корреляции приспособлен для реализации на ЭВМ и использует дискретную модель G-функционалов.

*При реализации рассматриваемого метода имеется возможность использования априорной информации об объекте с целью выбора значений запаздываний* (степень инерционности объекта, его колебательность и т. д.). Наличие такой возможности *позволяет сократить объем вычислений при решении задачи идентификации.*

Как указано в ряде работ, несмотря на отмеченные достоинства метода корреляции, проблема идентификации нелинейных объектов далека от завершения. *Если иметь в виду работы, которые отражают результаты применения метода Ли и Шепцена при решении инженерных задач, то они в основном ограничены оценкой ядер нелинейной системы не выше второго порядка.* Построенные модели учитывали только нелинейности, носящие квадратичный характер.

Необходимо обратить внимание и на тот факт, что в некоторых случаях могут нарушаться условия ортогональности (7.8), что приводит к накоплению ошибок идентификации. Речь идет о невозможности идеальной реализации белого гауссова процесса. Реальный белый шум, поступающий на вход нелинейного объекта, имеет ограниченную полосу частот, т. е. его спектр ограничен. Естественно, необходимым условием проведения эксперимента является условие: *спектральная функция реального белого шума должна быть шире полосы пропускания объекта.* Вместе с тем, в этом случае имеет место различие между статистическими моментами $\alpha_{ш}(\tau_1,\tau_2,...,\tau_i)$ реального белого шума и соответствующими моментами идеального процесса при увеличении «*i*». Степень отличия указанных моментов является тем обстоятельством, которое определяет ошибки алгоритма и, в

конечном счете, размерность «*i*» ядра $k(\tau_1,\tau_2,...,\tau_i)$ идентифицируемого объекта, при которой ошибки являются допустимыми. Математически отличие реального белого шума от идеального отражается тем фактом, что, например, в (7.24) будут иметь место «искаженные» дельта-функции. Еще раз укажем на трудность вычисления ядер вдоль диагоналей и, следовательно, невозможность вычисления интегралов при $\tau_1 = \tau_2 = \tau_3 = ....$

Подробный анализ метода Ли и Шетцена проведен в [20].

## 7.2. Уравнения статистической идентификации нелинейных объектов управления

При рассмотрении метода Винера и метода корреляции, *важным является то обстоятельство, что на вход объекта идентификации подается специальный пробный сигнал — нормальный белый шум.* Во многих случаях такая возможность отсутствует. Например, *важной является задача идентификации для случая, когда объект находится в режиме нормальной эксплуатации и входной и выходной сигналы объекта отличны от белого шума.*

В некоторых случаях необходимо построить модель объекта по имеющимся записям входного и выходного случайных сигналов, отличных от белого шума, причем записи сделаны в условиях помех.

*Цель следующих параграфов – рассмотрение возможных подходов к решению задачи идентификации для случая, когда входной сигнал u(t) отличен от белого гауссова шума, например, когда для решения задачи идентификации используется информация, полученная в режиме нормальной эксплуатации объекта.*

Положим, что нелинейный объект описывается рядом Вольтерра вида

$$x(t) = \int_0^\infty k_0(\tau)d\tau + \sum_{i=1}^{N}\int_0^\infty ...\int_0^\infty k_i(\tau_1,\tau_2,...,\tau_i)\times \\ \times u(t-\tau_1)u(t-\tau_2)...u(t-\tau_i)d\tau_1...d\tau_i; \qquad (7.29)$$

ИПФ нелинейной системы удовлетворяют условиям

$$\begin{cases} \lim\limits_{\tau\to\infty} k_0(\tau) = 0; \\ \lim\limits_{\tau_1,...,\tau_i\to\infty} k_i(\tau_1,\tau_2,...,\tau_i) = 0; \end{cases} \qquad (7.30)$$

при этом выполнены условия физической реализуемости

$$k_0(\tau) = 0 \quad \text{при} \quad \tau < 0; \tag{7.31}$$

$$k_i(\tau_1, \tau_2, ..., \tau_i) = 0 \quad \text{при} \quad \tau_j < 0; \quad j = \overline{1,i}.$$

Положим, что на выходной сигнал накладывается помеха ν(*t*) (рис. 7.7).

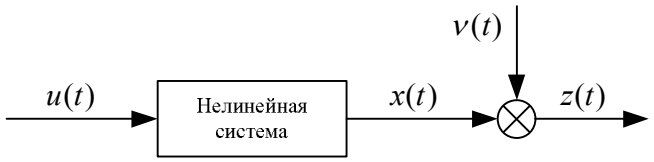



Рис. 7.7. К постановке задачи идентификации нелинейных объектов

Задача идентификации формулируется так: *построить математическую модель нелинейной системы, поведение которой описывается функциональным рядом* (7.29), *при наблюдении за случайными сигналами u*(*t*) *и z*(*t*).

*Такой подход называется статистической идентификацией.*

Практически задача идентификации сводится к определению ядер

$$k_0(\tau), k_1(\tau), k_2(\tau_1, \tau_2), ..., k_i(\tau_1, \tau_2, ..., \tau_i)$$

по статистическим характеристикам сигналов *u*(*t*) и z(*t*).

Статистические методы позволяют оценить динамические характеристики нелинейных систем по критерию минимума СКО.

В теории линейных систем широко используется корреляционный метод (рис. 7.8).

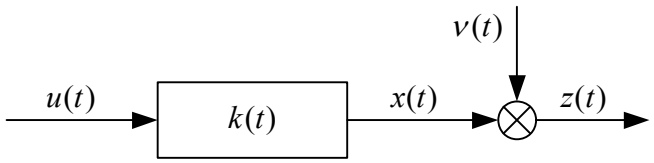



Рис. 7.8. К постановке задачи идентификации линейных объектов

Стационарность процессов ν(t) и *u*(*t*) и работа объекта в установившемся режиме влечет стационарность процесса *z*(*t*).

Необходимо определить оценку ИПФ *k*(τ) из условия минимума СКО

$$\bar{\varepsilon}^2 = J = M\left[(u(t) - z(t))^2\right] \to min. \tag{7.32}$$

Решение поставленной задачи для класса линейных объектов дается интегральным уравнением Винера-Хопфа

$$R_{uz}(\tau) = \int_0^\infty k(\theta) R_{uu}(\tau - \theta) d\theta. \qquad (7.33)$$

Подход, применяемый при решении задачи статистической идентификации линейных объектов, обобщается на решение той же задачи в классе объектов, описываемых функционалами Вольтерра [12, 20].

Учитывая описание нелинейной системы соотношением (7.29), запишем формулу, определяющую критерий (7.32):

$$M\left[\varepsilon^2(t)\right] = \bar{\varepsilon}^2 = J =$$

$$= M\left[\left( z(t) - \int_0^\infty k_0(\tau) d\tau - \sum_{i=1}^N \int_0^\infty \ldots \int_0^\infty k_i(\tau_1, \ldots, \tau_i) u(t-\tau_1) \ldots u(t-\tau_1) d\tau_1 \ldots d\tau_i \right)^2 \right] =$$

$$= M\left[\left( z^2(t) - 2z(t)\left( \int_0^\infty k_0(\tau) d\tau - \sum_{i=1}^N \int_0^\infty \ldots \int_0^\infty k_1(\tau_1, \ldots, \tau_i) u(t-\tau_1) \ldots u(t-\tau_i) d\tau_1 \ldots d\tau_i \right) + \right.$$

$$+ \left( \int_0^\infty k_0(\tau) d\tau \right)^2 + 2 \int_0^\infty k_0 d\tau \cdot \left( \sum_{i=0}^N \int_0^\infty \ldots \int_0^\infty k_i(\tau_1, \ldots, \tau_i) u(t-\tau_1) \ldots u(t-\tau_i) d\tau_1 \ldots d\tau_i \right) +$$

$$+ \sum_{i=1}^N \sum_{j=1}^N \int_0^\infty \ldots \int_0^\infty k_i(\tau_1, \ldots, \tau_i) k_j(\lambda_1, \ldots, \lambda_i) u(t-\tau_1) \ldots u(t-\tau_i) \times$$

$$\left. \times u(t-\lambda_1) \ldots u(t-\lambda_j) d\tau_1 \ldots d\tau_i d\lambda_1 \ldots d\lambda_j \right)\Big] \qquad (7.34)$$

Поскольку операции суммирования, интегрирования и математического ожидания перестановочны, то из последней зависимости получим формулу

$$J = \bar{\varepsilon}^2 = D_{zz} - 2 \sum_{i=0}^N \int_0^\infty \ldots \int_0^\infty k_i(\tau_1, \ldots, \tau_i) \alpha_{zu}^{(i)}(\tau_1, \tau_2, \ldots, \tau_i) d\tau_1 \ldots d\tau_i +$$

$$+ \sum_{i=0}^N \sum_{j=0}^N \int_0^\infty \ldots \int_0^\infty k_i(\tau_1, \ldots, \tau_i) k_j(\lambda_1, \ldots, \lambda_j) \alpha_{uu}^{(i+j)}(\tau_1, \ldots, \tau_i, \lambda_1, \ldots, \lambda_j) d\tau_1 \ldots d\lambda_j. \qquad (7.35)$$

При выводе последней зависимости с целью сокращения записи принято, что член при $i = 0$ имеет вид

$$\int_0^\infty k_0(\tau) d\tau M(z(t)); \qquad (7.36)$$

член при $i = 0, j = 0$ запишется так:

$$\int_0^\infty k_0(\tau)d\tau\int_0^\infty k_0(\lambda)d\lambda; \qquad (7.37)$$

члены при $i = 0$ и $j=\overline{1,N}$ или $i=\overline{1,N}$ и $j = 0$ однотипны и имеют вид

$$\int_0^\infty k_0(\tau)d\tau\int_0^\infty\ldots\int_0^\infty k_j(\lambda_1,\ldots,\lambda_j)\alpha_{uu}^{(j)}(\lambda_1,\lambda_2,\ldots,\lambda_j)d\lambda_1\ldots d\lambda_j; \qquad (7.38)$$

$$\alpha_{zu}^{(i)}(\tau_1,\ldots,\tau_i) = M\left[z(t)u(t-\tau_1)u(t-\tau_2)\ldots u(t-\tau_i)\right]; \qquad (7.39)$$

$$\alpha_{uu}^{(i+j)}(\tau_1,\ldots,\tau_i,\lambda_1,\ldots,\lambda_j) = M\left[u(t-\tau_1)\ldots u(t-\tau_i)u(t-\lambda_1)\ldots u(t-\lambda_j)\right], \qquad (7.40)$$

$$i,j=\overline{1,N};$$

$$\alpha_{uu}^{j}(\lambda_1,\ldots,\lambda_i) = M\left[u(t-\lambda_1)\ldots u(t-\lambda_j)\right], \quad j=\overline{1,N}. \qquad (7.41)$$

Найдем необходимые условия, которым должны удовлетворять ядра $k_i(\tau_1,\tau_2,\ldots,\tau_i)$, для того чтобы они доставляли минимум функционалу

$$J = M\left[\varepsilon^2(t)\right]. \qquad (7.42)$$

При решении задачи оптимизации применим следующие рассуждения.

Пусть $k_0^*(\tau)$, $k_1^*(\tau_1)$, $k_2^*(\tau_1,\tau_2),\ldots,k_i^*(\tau_1,\tau_2,\ldots,\tau_i),\ldots$ – минимизируем $\overline{\varepsilon}^2$ в (7.35); тогда произвольные ядра $k_0(\tau)$, $k_1(\tau_1)$, $k_2(\tau_1,\tau_2),\ldots$, $k_i(\tau_1,\tau_2,\ldots,\tau_i),\ldots$ можно представить в виде

$$k_i(\tau_1,\tau_2,\ldots,\tau_i) = k_i^*(\tau_1,\tau_2,\ldots,\tau_i) + \gamma_i\Delta k_i(\tau_1,\tau_2,\ldots,\tau_i), \qquad (7.43)$$

где $\gamma_i$ – некоторый произвольный параметр, не зависящий от $\tau_1,\tau_2,\ldots,\tau_i$; $\Delta k_i(\tau_1,\tau_2,\ldots,\tau_i)$ – произвольная функция, которая аналогично $k_i^*(\tau_1,\tau_2,\ldots,\tau_i)$ удовлетворяет условиям (7.30) и (7.31). Подставим (7.43) в (7.35); результат имеет вид

$$M\left[\varepsilon^2(t)\right] = J + \Delta J =$$

$$= M\left[z^2(t)\right] - 2\sum_{i=0}^{N}\int_0^\infty\ldots\int_0^\infty\left[k_i^*(\tau_1,\tau_2,\ldots,\tau_i) + \gamma_i\Delta k_i(\tau_1,\tau_2,\ldots,\tau_i)\right]\times$$

$$\times\,\alpha_{zu}^{(i)}(\tau_1,\tau_2,\ldots,\tau_i)d\tau_1\ldots d\tau_i + \sum_{i=0}^{N}\sum_{j=0}^{N}\int_0^\infty\ldots\int_0^\infty\left[k_i^*(\tau_1,\tau_2,\ldots,\tau_i) + \right.$$

$$\left. + \gamma_i\Delta k_i(\tau_1,\tau_2,\ldots,\tau_i)\right]\left[k_j^*(\lambda_1,\lambda_2,\ldots,\lambda_j) + \gamma_j\Delta k_j(\lambda_1,\lambda_2,\ldots,\lambda_i)\right]\times$$

$$\times\,\alpha_{uu}^{(i+j)}(\tau_1,\tau_2,\ldots,\tau_i,\lambda_1,\lambda_2,\ldots,\lambda_i)d\tau_1 d\tau_2\ldots d\tau_1 d\lambda_1 d\lambda_2\ldots d\lambda_j. \qquad (7.44)$$

Последнее выражение можно привести к форме

$$J + \Delta J = J - 2\sum_{i=0}^{N} \gamma_i J_{1(i)} + \sum_{i=0}^{N}\sum_{j=0}^{N} \gamma_i \gamma_j J_{2(ij)}, \qquad (7.45)$$

где

$$J_{1(i)} = \int_0^\infty \ldots \int_0^\infty \Delta k_i(\tau_1, \tau_2, \ldots, \tau_i)\alpha_{zu}^{(i)}(\tau_1, \tau_2, \ldots, \tau_i)d\tau_1 d\tau_2 \ldots d\tau_i; \qquad (7.46)$$

$$J_{2(ij)} = \int_0^\infty \ldots \int_0^\infty \Delta k_i(\tau_1, \tau_2, \ldots, \tau_i)\Delta k_j(\lambda_1, \lambda_2, \ldots, \lambda_j)\alpha_{uu}^{(i+j)} \times$$

$$\times(\tau_1, \ldots, \tau_i, \lambda_1, \ldots, \lambda_j)d\tau_1 \ldots d\lambda_j; \qquad (7.47)$$

$J + \Delta J$ в (7.45) – значение, принимаемое функционалом $J$ при замене $k_i(\tau_1, \tau_2, \ldots, \tau_i)$ на

$$k_i^*(\tau_1, \tau_2, \ldots, \tau_i) + \gamma_i \Delta k_i(\tau_1, \tau_2, \ldots, \tau_i). \qquad (7.48)$$

Необходимое условие минимума функционала (7.45) имеет вид

$$\left[\frac{\partial}{\partial \gamma_m}(J + \Delta J)\right]_{\gamma_0=\gamma_1=\ldots=\gamma_N=0} = 0, \quad m = \overline{0, N}. \qquad (7.49)$$

Можно показать, что условие (7.49) является не только необходимым, но и достаточным условием минимума функционала $J$. Для этого необходимо показать, что при любом выборе функций $\Delta k_i(\tau_1, \tau_2, \ldots, \tau_i)$ будет иметь место неравенство

$$(J + \Delta J)_{J_{1(i)}} \geq J. \qquad (7.50)$$

Доказательство можно найти в [12, 20].

Таким образом, *условие* (7.49) *является необходимым и достаточным условием минимума дисперсии* $M\left[\varepsilon^2(t)\right]$.

Из (7.45) легко получить систему уравнений

$$-2\int_0^\infty \ldots \int_0^\infty \Delta k_m(\tau_1, \tau_2, \ldots, \tau_m)\left[\alpha_{zu}^{(m)}(\tau_1, \tau_2, \ldots, \tau_m) - \sum_{j=0}^{N}\int_0^\infty \ldots \int_0^\infty k_j^*(\lambda_1, \lambda_2, \ldots, \lambda_j) \times\right.$$

$$\left.\times \alpha_{uu}^{(m+j)}(\tau_1, \tau_2, \ldots, \tau_m, \lambda_1, \lambda_2, \ldots, \lambda_j)d\lambda_1 \ldots d\lambda_j\right]d\tau_1, d\tau_2 \ldots d\tau_m = 0, \; m = \overline{0, N}. \qquad (7.51)$$

Равенство нулю последнего соотношения при любых функциях $\Delta k_m(\tau_1, \tau_2, \ldots, \tau_m)$ будет иметь место, если

$$\alpha_{zu}^{(m)}(\tau_1, \tau_2, \ldots, \tau_m) - \sum_{j=0}^{N}\int_0^\infty \ldots \int_0^\infty k_j^*(\lambda_1, \lambda_2, \ldots, \lambda_j) \times$$

$$\times \alpha_{uu}^{(m+j)}(\tau_1, \tau_2, \ldots, \tau_m, \lambda_1, \lambda_2, \ldots, \lambda_j)d\lambda_1 d\lambda_2 \ldots d\lambda_j = 0, \; m = \overline{0, N}. \qquad (7.52)$$

*Отсюда получаем систему линейных интегральных уравнений относительно ядер Вольтерра* $k_j^*(\lambda_1,\lambda_2,...,\lambda_j)$

$$\sum_{j=0}^{N}\int_0^\infty ...\int_0^\infty k_j^*(\lambda_1,\lambda_2,...,\lambda_j)\alpha_{uu}^{(m+j)}(\tau_1,\tau_2,...,\tau_m,\lambda_1,\lambda_2,...,\lambda_j)d\lambda_1...d\lambda_j =$$

$$= \alpha_{zu}^{(m)}(\tau_1,\tau_2,...,\tau_m),\tau_1,\tau_2,...,\tau_m \geq 0,\ \ m=\overline{0,N}. \qquad (7.53)$$

*Полученная система интегральных уравнений представляет собой обобщение интегрального уравнения Винера-Хопфа* (7.33) *для нелинейных динамических объектов. Так как уравнения* (7.53) *линейны, то системы, описываемые рядом Вольтерра, называют системами, сводимыми в статистическом смысле к линейным.*

Решение системы (7.53) определяет минимум функционала *I*.

Система (7.53) в общем случае аналитически неразрешима. *Существенные упрощения уравнений* (7.53) *имеют место в случае, если сигнал u*(*t*) *имеет нормальный закон распределения. Упрощения достигаются и при законе распределения Пуассона. Основой такого упрощения является положение о том, что моменты старших порядков случайных процессов выражаются через моменты второго порядка.*

Если *u*(*t*) – центрированный стационарный нормальный случайный процесс, то

$$\alpha_{uu}^{(2\nu+1)}(\tau_1,\tau_2,...,\tau_{2\nu+1}) = 0,\ \ \nu = 0,1,..., \qquad (7.54)$$

т. е. все моментные функции нечетного порядка равны нулю; моментные функции четного порядка представляются в виде сумм произведений моментных функций второго порядка, т. е.

$$\alpha_{uu}^{(2\nu)}(\tau_1,\tau_2,...,\tau_{2\nu}) = \sum_{n_1,...,n_{2\nu}}\alpha_{uu}^{(2)}(\tau_{n_1}-\tau_{n_2})...\alpha_{uu}^{(2)}(\tau_{n_{2\nu-1}}-\tau_{n_{2\nu}}), \qquad (7.55)$$

где суммирование производится по всем сочетаниям пар различных чисел $(n_1,n_2),(n_1,n_3),...,(n_{2\nu-1},n_{2\nu})$; $n_1,n_2,...,n_{2\nu}$ – целые положительные числа от 1 до 2ν.

*В силу условий* (7.54) *и* (7.55) *для случая, когда u*(*t*) *— стационарный центрированный нормальный случайный процесс, система линейных интегральных уравнений* (7.53) *распадается на две подсистемы, в одну из которых входят импульсные переходные функции четного порядка, в другую – нечетного.*

Если $m$ = 0, 2, 4,..., $2p$, то в левой части уравнений останутся лишь члены с четными номерами, если же $m$ = 1, 3, 5,...,$2p$+1 – то в левой части останутся члены с нечетными номерами.

Если $N = 2N_1 + 1$, то указанные две подсистемы имеют вид

$$\sum_{q=0}^{N_1} \int_0^\infty \dots \int_0^\infty k_{2q}^*(\lambda_1, \lambda_2, \dots, \lambda_{2q}) \alpha_{uu}^{(2p+2q)} \times$$

$$\times (\tau_1, \tau_2, \dots, \tau_{2q}, \lambda_1, \lambda_2, \dots, \lambda_{2q}) d\lambda_1 d\lambda_2 \dots d\lambda_{2q} = \tag{7.56}$$

$$= \alpha_{zu}^{(2p)}(\tau_1, \tau_2, \dots, \tau_{2p});$$

$$\sum_{q=0}^{N_1} \int_0^\infty \dots \int_0^\infty k_{2q+1}^*(\lambda_1, \lambda_2, \dots, \lambda_{2q+1}) \alpha_{uu}^{(2p+2q+2)}(\tau_1, \tau_2, \dots \tau_{2p+1}, \lambda_1, \lambda_2, \dots, \lambda_{2q+1}) \times \tag{7.57}$$

$$\times d\lambda_1 d\lambda_2 \dots d\lambda_{2q+1} = \alpha_{zu}^{(2p+1)}(\tau_1, \tau_2, \dots, \tau_{2p+1}),$$

где $\tau_i \geq 0, \quad i = 1, 2, \dots, 2p, (2p+1), \quad p = \overline{0, N_1}$.

Две исходные системы интегральных уравнений (7.56) и (7.57) приводятся к треугольной системе линейных интегральных уравнений вида [12,20]

$$\sum_{q=p}^{N_1} \gamma_{2p,2q} \int_0^\infty \dots \int_0^\infty k_{2q}^*(\lambda_1, \lambda_2, \dots, \lambda_{2q}) \alpha_{uu}^{(2)}(\tau_1 - \lambda_1) \dots \alpha_{uu}^{(2)}(\tau_{2p} - \lambda_{2q}) \times$$

$$\times \alpha_{uu}^{(2)}(\lambda_{2p+1} - \lambda_{2p+2}) \dots \alpha_{uu}^{(2)}(\lambda_{2q-1} - \lambda_{2q}) d\lambda_1 \dots d\lambda_{2q} = E_{2p}(\tau_1, \dots, \tau_{2p}); \tag{7.58}$$

$$\sum_{q=p}^{N_1} \gamma_{2p+1,2q+1} \int_0^\infty \dots \int_0^\infty k_{2q+1}^*(\lambda_1, \dots, \lambda_{2q+1}) \alpha_{uu}^{(2)}(\tau_1 - \lambda_1) \dots \alpha_{uu}^{(2)}(\tau_{2p+1} - \lambda_{2p+1}) \times$$

$$\times \alpha_{uu}^{(2)}(\lambda_{2p+2} - \lambda_{2p+3}) \dots \alpha_{uu}^{(2)}(\lambda_{2q} - \lambda_{2q+1}) d\lambda_1 \dots d\lambda_{2q+1} = E_{2p+1}, \tag{7.59}$$

где $\tau_1, \tau_2, \dots, \tau_{2N_1}, \tau_{2N_1+1} \geq 0; \quad \gamma_{p,q} = \dfrac{q!}{(q-p)!!};$

$$E_p(\tau_1, \tau_2, \dots, \tau_p) = \alpha_{zu}^{(p)}(\tau_1, \tau_2, \dots, \tau_p) +$$

$$+ \sum_{0 \leq 2j \leq p-1} \sum_{\alpha_1, \dots, \alpha_p} (-1)^{j+1} \alpha_{zu}^{(p-2(j+1))}(\tau_{\alpha_1}, \dots, \tau_{\alpha_{p-2(j+1)}}) \times$$

$$\times \alpha_{uu}^{(2(j+1))}(\tau_{\alpha_{p-2(j+1)+1}}, \dots, \tau_{\alpha_p}). \tag{7.60}$$

В формуле, определяющей $E_p(\tau_1, \tau_2, \dots, \tau_p)$, первое суммирование производится по всем $j$ таким, что $2j$ меньше $p$ -1, а второе – по всем перестановкам индексов $\alpha_1, \alpha_2, \dots, \alpha_p$, принимающим разные значения от 1 до $p$.

*Уравнение* (7.33) *называется уравнением статистической идентификации линейного объекта, а система* (7.58), (7.59) *носит*

*название линейной системы интегральных уравнений статистической идентификации нелинейных объектов, поведение которых описывается функционалами Вольтерра.*

### 7.3. Спектральный (параметрический) подход к решению задачи идентификации

Положим, что объект идентификации описывается функционалами Вольтерра и ядра интегрируемых с квадратом.

*При решении задачи идентификации можно использовать спектральное представление процессов и динамических характеристик объектов.* Такой подход, с одной стороны, во многих случаях *упрощает решение задачи* (например, в тех случаях, когда для аппроксимации используется небольшое число членов разложения), с другой – *позволяет получить математическую модель в удобном для практического использования виде* (например, такая форма описания позволяет достаточно просто реализовать физическую модель объекта в аналоговой или цифровой формах).

Введем обобщенное обозначение ортонормированного базиса

$$\hat{O}(t) = (\varphi_1(t), \varphi_2(t), ..., \varphi_l(t))^T, \qquad (7.61)$$

где $l$ – число удерживаемых членов разложения.

Каждое из ядер идентифицируемого объекта можно аппроксимировать многомерным ортонормированным рядом

$$k_i(\tau_1, \tau_2, ..., \tau_i) = \sum_{\nu_1=1}^{l} \sum_{\nu_2=1}^{l} ... \sum_{\nu_i=1}^{l} c^{k_i}_{\nu_1 \nu_2 ... \nu_i} \varphi_{\nu_1}(\tau_1) \varphi_{\nu_2}(\tau_2) ... \varphi_{\nu_i}(\tau_i). \qquad (7.62)$$

Постановка задачи состоит в следующем: на вход идентифицируемого объекта поступает случайный стационарный сигнал $\nu(t)$; $x(t)$ – выходной процесс, $\nu(t)$ – помеха, $z(t)$ – наблюдаемый сигнал (рис. 7.9).




Рис. 7.9. К постановке задачи идентификации

Требуется построить математическую модель объекта, т. е. рассчитать коэффициенты $c^{k_i}_{\nu_1\nu_2\ldots\nu_i}$, $i=\overline{1,N}$. при наблюдении за случайными сигналами $u(t)$ и $z(t)$.

Представление (7.62) заменим эквивалентной формой

$$k_i(\tau_1,\tau_2,\ldots,\tau_i)=\sum_{\nu=1}^{l_1}\theta_\nu^i\varphi_\nu(\tau_1,\tau_2,\ldots,\tau_i), \tag{7.63}$$

где

$$\theta=(\theta_1^1,\theta_2^1,\ldots,\theta_{l_1}^1,\theta_1^2,\theta_2^2,\ldots,\theta_{l_2}^2,\ldots,\theta_1^N,\theta_2^N,\ldots,\theta_{l_N}^N) \tag{7.64}$$

– вектор коэффициентов, подлежащий определению при решении задачи идентификации.

С учетом (7.63) ряд Вольтера

$$x(t)=\sum_{i=1}^{N}\int_0^\infty\ldots\int_0^\infty k_i(\tau_1,\tau_2,\ldots,\tau_i)u(t-\tau_1)u(t-\tau_2)\ldots u(t-\tau_i)d\tau_1d\tau_2\ldots d\tau_i, \tag{7.65}$$

характеризующий поведение объекта, запишется так:

$$x_M(t)=\sum_{i=1}^{N}\sum_{\nu=1}^{l_i}\theta_\nu^i\int_0^\infty\ldots\int_0^\infty\varphi_\nu(\tau_1,\tau_2,\ldots,\tau_i)\times$$
$$\times u(t-\tau_1)u(t-\tau_2)\ldots u(t-\tau_i)d\tau_1d\tau_2\ldots d\tau_i. \tag{7.66}$$

Если ввести обозначение

$$x_{\nu i}(t)=\int_0^\infty\ldots\int_0^\infty\varphi_\nu(\tau_1,\tau_2,\ldots,\tau_i)u(t-\tau_1)\times$$
$$\times u(t-\tau_2)\ldots u(t-\tau_i)d\tau_1d\tau_2\ldots d\tau_i,\quad i=\overline{1,N}, \tag{7.67}$$

то интегральное соотношение (7.66) принимает вид [12, 20]

$$x_M(t)=\sum_{i=1}^{N}\sum_{\nu=1}^{l_i}\theta_\nu^i x_{\nu i}(t). \tag{7.68}$$

В некоторых случаях для решения задачи идентификации (если, например, задача идентификации решается в режиме нормальной эксплуатации объекта управления) сигналы $x_{\nu i}(t)$ необходимо получать в реальном масштабе времени. Решение этой задачи практически осуществимо, если можно реализовать фильтры с ИПФ $\varphi_1(\tau),\varphi_2(\tau),\ldots,\varphi_l(\tau)$.

Запишем функционал, определяющий степень адекватности модели и объекта в виде

$$J(\theta_1^1,\theta_2^1,...,\theta_{l_N}^N)=M\left[\left(z(t)-x_M(t,\theta_1^1,\theta_2^1,...,\theta_{l_N}^N)\right)^2\right]=$$

$$=M\left[\left(z(t)-\sum_{i=1}^{N}\sum_{\nu=1}^{l_i}\theta_\nu^i x_{\nu i}(t)\right)^2\right]=$$

$$=M\left[z^2(t)-2z\left(\sum_{i=1}^{N}\sum_{\nu=1}^{l_i}\theta_\nu^i x_{\nu i}(t)\right)+\sum_{i_1=1}^{N}\sum_{i_2=1}^{N}\sum_{\nu_1=1}^{l_{i_1}}\sum_{\nu_2=1}^{l_{i_2}}\theta_{\nu_1}^{i_1}\theta_{\nu_2}^{i_2}x_{\nu_1 i_1}(t)x_{\nu_2 i_2}(t)\right]. \quad (7.69)$$

При решении задачи идентификации параметры $\theta_1^1,\theta_2^1,...,\theta_{l_N}^N$ находятся, исходя из следующего условия:

$$J\left(\theta_1^1,\theta_2^1,...,\theta_{l_N}^N\right)\to\min_{\theta}. \quad (7.70)$$

Перепишем (7.69) следующим образом:

$$J\left(\theta_1^1,\theta_2^1,...,\theta_{l_N}^N\right)=\overline{\sigma_z}^2-2\sum_{i=1}^{N}\sum_{\nu=1}^{l_i}M\left[z(t)x_{\nu i}(t)\right]+$$

$$+\sum_{i_1=1}^{N}\sum_{i_2=1}^{N}\sum_{\nu_1=1}^{l_{i_1}}\sum_{\nu_2=1}^{l_{i_2}}\theta_{\nu_1}^{i_1}\theta_{\nu_2}^{i_2}M\left[x_{\nu_1 i_1}(t)x_{\nu_2 i_2}(t)\right]=$$

$$=\overline{\sigma_z}^2-2\sum_{i=1}^{N}\sum_{\nu=1}^{l_i}\alpha_{zx_{\nu i}}+\sum_{i_1=1}^{N}\sum_{i_2=1}^{N}\sum_{\nu_1=1}^{l_{i_1}}\sum_{\nu_2=1}^{l_{i_2}}\alpha_{x_{\nu_1 i_1}x_{\nu_2 i_2}}, \quad (7.71)$$

где

$$\begin{aligned}\alpha_{zx_{\nu i}}&=M\left[z(t)x_{\nu i}(t)\right];\\ \alpha_{x_{\nu_1 i_1}x_{\nu_2 i_2}}&=M\left[x_{\nu_1 i_1}(t)x_{\nu_2 i_2}(t)\right].\end{aligned} \quad (7.72)$$

Учитывая (7.70), определяющее меру близости процессов на выходе объекта и его математической модели, можно записать

$$\frac{\partial J(\theta_1^1,\theta_2^1,...,\theta_{l_N}^N)}{\partial\theta_k^j}=0,\quad j=\overline{1,N};\quad k=\overline{1,(l_1+l_2+...+l_N)}. \quad (7.73)$$

Отсюда

$$\sum_{i=1}^{N}\sum_{\nu=1}^{l_i}\theta_\nu^i\alpha_{x_{\nu k}x_{ij}}=\alpha_{zx_{kj}}. \quad (7.74)$$

*Последняя система определяет необходимые условия экстремума*. В [12, 20] рассмотрены вопросы, связанные с достаточностью указанных условий.

Если воспользоваться соответствующими обозначениями, то систему алгебраических уравнений (7.74) можно записать в форме

$$\alpha_{xx}\theta=\alpha_{zx}. \quad (7.75)$$

Формула, определяющая вектор идентифицируемых параметров, имеет вид

$$\theta^* = \alpha_{xx}^{-1}\alpha_{zx}. \qquad (7.76)$$

Можно показать, что если система $\varphi_1(\tau),\varphi_2(\tau),...,\varphi_l(\tau).$ – линейно независима, то (7.75) имеет единственное решение.

Для расчета матриц $\alpha_{xx}$ и $\alpha_{zx}$ необходимо найти значения величин $\alpha_{zx_{vi}}$ и $\alpha_{x_{v_1 i_1} x_{v_2 i_2}}$.

Моменты $\alpha_{zx_{vi}}$ рассчитываются по формуле

$$\alpha_{zx_{vi}} = M\left[z(t)x_{vi}(t)\right] =$$

$$= M\int_0^\infty ...\int_0^\infty \varphi_v(\tau_1,\tau_2,...,\tau_i)z(t)u(t-\tau_1)u(t-\tau_2)\times...\times$$

$$\times u(t-\tau_i)d\tau_1 d\tau_2 ...d\tau_i =$$

$$= \int_0^\infty ...\int_0^\infty \varphi_v(\tau_1,\tau_2,...,\tau_i)\alpha_{zu}^{(i+1)}(\tau_1,\tau_2,...,\tau_i)d\tau_1 d\tau_2 ...d\tau_i, \qquad (7.77)$$

а для нахождения $\alpha_{x_{v_1 i_1} x_{v_2 i_2}}$ можно воспользоваться зависимостью

$$\alpha_{X_{v1'1}X_{v2'2}} = \int_0^\infty ...\int_0^\infty \varphi_{v_1}(\tau_1,\tau_2,...,\tau_{i_1})\varphi_{v_2}(\tau_1',\tau_2'...,\tau_{i_2}')\times$$

$$\times a_{uu}^{(i_1+i_2)}(\tau_1,\tau_2...,\tau_{i_1};\tau_1',\tau_2',...,\tau_{i_2}')d\tau_1 d\tau_2 ...d\tau_{i_2}.$$

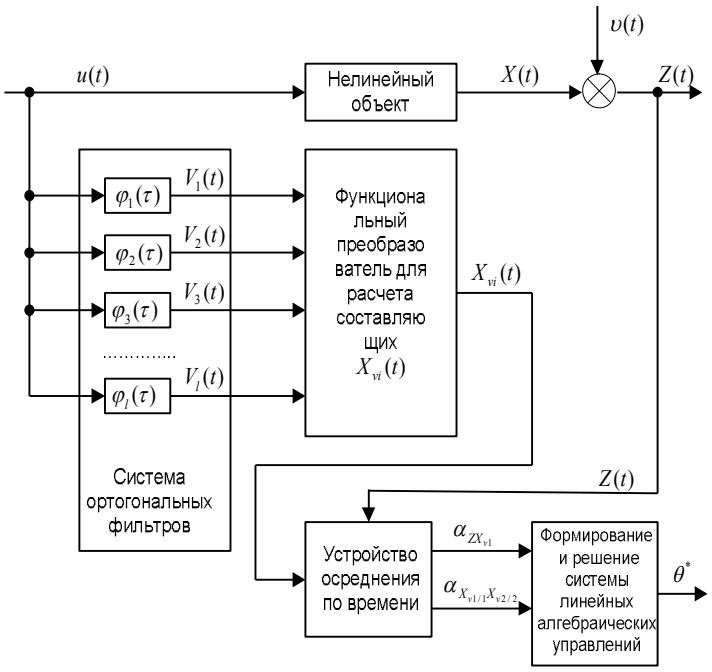



Рис. 7.10. Структурная схема системы идентификации

## 7.4. Статистические методы оценки параметров нелинейных объектов

Изложим теоретические положения, которые являются основой методов, содержание которых изложено ниже.

Положим, что объект идентификации описывается системой нелинейных дифференциальных уравнений вида

$$\begin{aligned}
\dot{x}_1 &= f(x_1, x_2, \ldots x_n; u_1, u_2, \ldots, u_m; \theta_1, \theta_2 \ldots, \theta_e; t), \\
\dot{x}_2 &= f(x_1, x_2, \ldots x_n; u_1, u_2, \ldots, u_m; \theta_1, \theta_2 \ldots, \theta_e; t), \\
&\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\
\dot{x}_n &= f(x_1, x_2, \ldots x_n; u_1, u_2, \ldots, u_m; \theta_1, \theta_2 \ldots, \theta_e; t),
\end{aligned} \qquad (7.78)$$

где $u = (u_1, u_2, \ldots, u_m)^T$ – $m$-мерный векторный случайный входной сигнал;

$x = (x_1, x_2, \ldots x_n)^T$ – $n$-мерный случайный вектор состояния;

$\theta = (\theta_1, \theta_2 \ldots, \theta_e)^T$ – вектор идентифицируемых параметров;

$\theta_i = (\theta_{1i}, \theta_{2i} \ldots, \theta_{li})^T$ – вектор номинальных значений параметров;

$\Delta\theta = (\Delta\theta_1, \Delta\theta_2 \ldots, \Delta\theta_n)^T$ – вектор отклонений фактических значений идентифицируемых параметров от номинальных, т. е. $\Delta\theta_i = \theta_i - \theta_{ii}$, $i = \overline{1,l}$.

Задача идентификации состоит в определении $\Delta\theta_i$, $i = \overline{1,l}$.

Перепишем (7.78) в векторной форме

$$\dot{x} = f(x, u, \theta, t). \qquad (7.79)$$

Очевидно, режим функционирования объекта при номинальных знамениях параметров можно определить как номинальный; он определяется векторным уравнением

$$\dot{x}_i = f(x_i, u, \theta_i, t), \quad x_i(0) = 0 \qquad (7.80)$$

или в скалярной форме

$$\begin{aligned}
\dot{x}_{1i} &= f(x_{1i}, x_{2i}, \ldots x_{ni}; u_1, u_2, \ldots, u_m; \theta_{1i}, \theta_{2i} \ldots, \theta_{ln}; t), \\
&\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\
\dot{x}_{ni} &= f(x_{1i}, x_{2i}, \ldots x_{ni}; u_1, u_2, \ldots, u_m; \theta_{1i}, \theta_{2i} \ldots, \theta_{ln}; t).
\end{aligned} \qquad (7.81)$$

Вектор переменных состояния $x = (x_1, x_2, \ldots x_n)^T$ объекта (7.78) при реальных значениях параметров $\theta_1, \theta_2 \ldots, \theta_l$ и при воздействии на

входе *m*-мерного вектора $u(t)$ может быть измерен и его можно считать известным на промежутке [0,T].

Вектор же состояния $x_i(t) = (x_{1i}(t), x_{2i}(t), \ldots x_{ni}(t))^T$ объекта (7.78) при номинальных значениях параметров $\theta_{1i}, \theta_{2i} \ldots, \theta_{li}$ и при входном сигнале $u(t)$ может быть получен в результате численного решения системы нелинейных дифференциальных уравнений (7.80) и, таким образом, его также можно считать известным на $[0,T]$.

Теперь легко найти процесс

$$\Delta x(t, \Delta\theta) = x(t, \theta) - x(t, \theta_i), \qquad (7.82)$$

который порожден вектором отклонения фактических значений идентифицируемых параметров от номинальных, т. е.

$$\Delta\theta_i = \theta_i - \theta_{ii}, \quad i = \overline{1,l}.$$

Вектор-функция $\Delta x(t, \Delta\theta)$ содержит информацию о $\Delta\theta_i, i = \overline{1,l}$ и поэтому является основой для построения метода и соответствующего алгоритма идентификации.

Известно, что наиболее совершенен аппарат решения линейных задач, поэтому при разработке метода идентификации, базирующегося на использовании $\Delta x(t, \Delta\theta)$, тем или иным путем приходят к линейным задачам. Далее рассмотрим реализации изложенного подхода, в основе которых лежат методы теории чувствительности и методы фильтрации [15, 20].

### 7.4.1. Оценка параметров нелинейных объектов методом наименьших квадратов с использованием теории чувствительности [15]

Положим, что объект идентификации описывается системой ДУ (7.78), а номинальный режим определяется ДУ (7.80)

Значения сигнала $x_i$ могут быть измерены в узлах сетки $t_k$ с шагом $\Delta$ при подаче на вход идентифицируемого объекта *m*-мерного вектора $u(t)$. Если же провести N экспериментов, то можно зафиксировать значения $x_{i(v)}(t)$ при $t = t_k$, $k = \overline{1,q}$, $v = \overline{1,N}$ (рис. 7.11).

С другой стороны, путем интегрирования уравнений (7.80) при входном сигнале $u(t)$ и номинальных значениях параметров

$\theta_{ii}, i = \overline{1,l}$ можно получить сеточную функцию $x_{ii}(t_k), k = \overline{1,q}$ рис. 7.12. За счет наличия $\Delta\theta_i = \theta_i - \theta_{ii}, i = \overline{1,l}$ будет иметь место сеточная функция ошибки $\Delta x_i(t_k, \theta) = x_1(t_k) - x_{ii}(t_k), \quad k = \overline{1.q}$ .

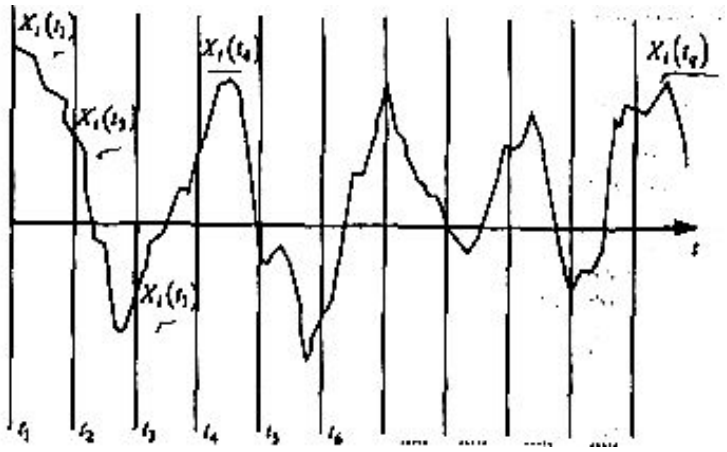


Рис. 7.11. Запись фазовой координаты $x_i(t)$,

Если провести N интегрирований системы ДУ (7.83), соответствующих N экспериментам с реальным объектом (7.78), то можно получить таблицу значений

$$\Delta x_i(t_k, \theta) = x_{1(v)}(t_k) - x_{i(v)i}(t_k), \quad k = \overline{1,q}, \quad v = \overline{1,N}. \qquad (7.83)$$

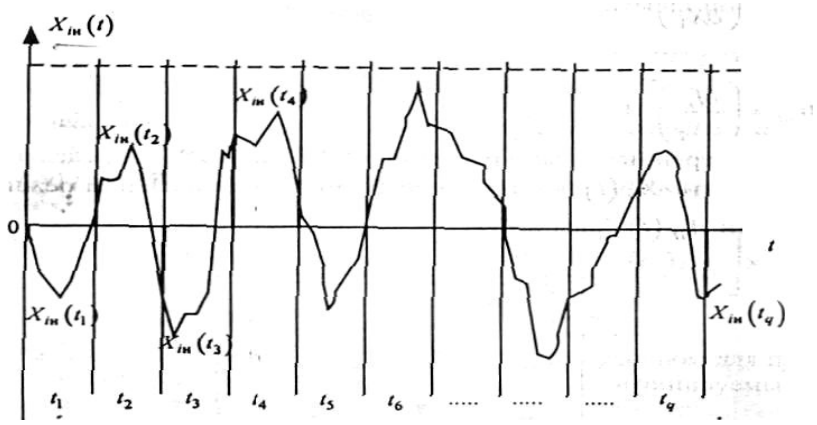


Рис. 7.12. Реализация фазовой координаты $\Delta x_{ii}(t)$,

полученная путем решения системы ДУ (7.80) при $p_i = p_{ii}, i = \overline{1,l}$

Поскольку измеренную в результате проведения эксперимента фазовую координату $x_i(t)$ можно представить в виде двух составляющих, одна из которых фазовая координата $x_{ii}(t)$, имеющая место при номинальных значениях параметров $\theta_{ii}, i=\overline{1,l}$, другая порождена отклонениями фактических значений $\Delta\theta_i=\theta_i-\theta_{ii}), i=\overline{1,l}$ от их номинальных значений при наблюдении $x_i(t)$, то справедлива зависимость

$$x_i(t)=x_{ii}(t)+\Delta x_i(t,\Delta\theta), \qquad (7.84)$$

где $\Delta x_i(t,\Delta\theta)$ – известная функция.

Важным этапом построения метода является нахождение формулы, определяющей зависимость $\Delta x_i(t,\Delta\theta)$ от величин $\Delta\theta$ в явной форме. Такую формулу можно получить при предположении, что отклонения параметров $\Delta\theta$ от номинальных значений являются малыми, и в этом случае можно применить теорию чувствительности, использование которой позволяет провести теоретическое обоснование метода, а для расчета неизвестных $\Delta\theta_i, i=\overline{1,l}$ получить линейную систему алгебраических уравнений.

Введем в рассмотрение функцию чувствительности $W_{ir}=(\partial x_i/\partial\theta_r)$; она определяется уравнениями [15]

$$\dot{W}_{ir}=\sum_{\theta=1}^{n}\left(\frac{\partial f_i}{\partial x_\theta}\right)^* W_{\theta r}+\left(\frac{\partial f_1}{\partial\theta_r}\right)^*,\quad W_{ir}(0)=0, i=\overline{1,n},\quad r=\overline{1,l}, \qquad (7.85)$$

где звездочка (*) означает, что производные и переменные вычисляются при номинальных значениях параметров.

В развернутом виде (7.85) можно записать так

$$\begin{aligned}
\dot{W}_{11}&=\left(\frac{\partial f_1}{\partial x_1}\right)^* W_{11}+\left(\frac{\partial f_1}{\partial x_2}\right)^* W_{21}+\ldots\ldots+\left(\frac{\partial f_1}{\partial x_n}\right)^* W_{n1}+\left(\frac{\partial f_1}{\partial\theta_1}\right)^*;\\
\dot{W}_{21}&=\left(\frac{\partial f_2}{\partial x_1}\right)^* W_{12}+\left(\frac{\partial f_1}{\partial x_2}\right)^* W_{21}+\ldots\ldots+\left(\frac{\partial f_1}{\partial x_n}\right)^* W_{n2}+\left(\frac{\partial f_2}{\partial\theta_2}\right)^*; \qquad (7.86)\\
&\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\\
\dot{W}_{n1}&=\left(\frac{\partial f_n}{\partial x_1}\right)^* W_{1l}+\left(\frac{\partial f_n}{\partial x_2}\right)^* W_{2l}+\ldots\ldots+\left(\frac{\partial f_1}{\partial x_n}\right)^* W_{nl}+\left(\frac{\partial f_n}{\partial\theta_1}\right)^*.
\end{aligned}$$

Системы дифференциальных уравнений (7.78) и (7.85) определяют модели, позволяющие получить $x_i(t)$ – вектор состояния в номинальном режиме и функции чувствительности $W_{ir}(t)$ (рис. 7.13).

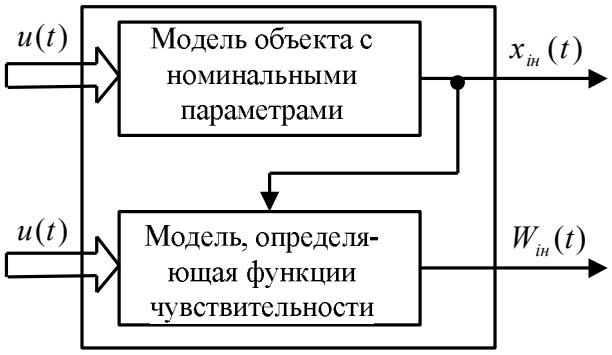



Рис. 7.13. Структурная схема вспомогательной системы

Взаимосвязанную систему двух моделей можно рассматривать как некоторую вспомогательную систему.

Учитывая следующее:

• процессы $x_{ii}(t)$ и $W_{ir}(t)$ могут быть определены путем интегрирования дифференциальных уравнений вспомогательной системы;

• отклонения параметров $\theta_i$ от их номинальных значений полагаются малыми, можно записать соотношение, определяющее $x_i(t)$ в виде

$$x_i(t) = x_{ii}(t) = x_{ii}(t) + \sum_{r=1}^{l} W_{ir}(t)\Delta\theta_r, \quad i = \overline{1,n}. \tag{7.87}$$

Положим, что проведено необходимое число экспериментов и измерено N значений одной фазовой координаты $x_i(t)$, а также процессов $x_{ii}(t)$, $W_{ir}(t)$ в точках $t_1, t_2 \ldots\ldots t_q$, т. е. известны величины

$$x_{i(v)}(t_k), \quad x_{ii(v)}(t_k), \quad W_{ir(v)}(t_k), \quad k = \overline{1,q}, \quad v = \overline{1,N}. \tag{7.88}$$

Считаем, что ошибки измерения указанных выше величин независимы и имеют нормальный закон распределения. В этом случае целесообразно для определения оценок $\Delta\theta_i$, $i = \overline{1,l}$ применить метод наименьших квадратов.

Составим квадратичную форму вида

$$I(\Delta\theta_1, \ldots\ldots, \Delta\theta_1) =$$

$$\frac{1}{N}\sum_{v=1}^{N}\left\{\sum_{k=1}^{q}\rho_k\left[x_{i(v)}(t_k) - x_{ii(v)}(t_k) - \sum_{r=1}^{l}W_{ir(v)}(t_k)\Delta\theta_r\right]^2\right\}, \tag{7.89}$$

где $\rho_k$ – весовые коэффициенты; их можно выбрать, исходя из условия $\theta_k = [\sigma_i^2(t_k)]^{-1}$, причем $\sigma_i^2(t_k)$ – дисперсии ошибок измерений.

С целью упрощения расчетов коэффициентов веса их можно положить равными между собой.

Оценки параметров идентифицируемого объекта находятся из условия

$$I(\Delta\theta_1, \Delta\theta_2, ....., \Delta\theta_l) \to \min_{\Delta\theta}. \tag{7.90}$$

В соответствии с условием (7.90), можно получить систему алгебраических уравнений относительно вектора $\Delta\theta$, для чего, исходя из необходимого условия оптимальности, надо от (7.89) взять частные производные по каждому из параметров и полученные соотношения приравнять к нулю

$$\frac{\partial I(\Delta\theta)}{\partial \Delta\theta_S} =$$

$$= \frac{2}{N}\sum_{k=1}^{N}\left\{\sum_{k=1}^{q}\rho_k\left[X_{i(v)}(t) - X_{ii(v)}(t_k) - \sum_{r=1}^{l}W_{ir(v)}(t_k)\Delta p_r\right]\times W_{is(v)}(k)\right\} = 0, s = \overline{1,l}. \tag{7.91}$$

Последнюю зависимость целесообразно записать в более удобном для практического использования виде, сгруппировав члены относительно идентифицируемых величин:

$$\sum_{r=1}^{e}\left(\sum_{k=1}^{e}\rho\frac{1}{2}\sum_{v=1}^{N}W_{ir(v)}(t_k)W_{is(v)}(t_k)\right)\Delta\theta_r =$$

$$= \sum_{k=1}^{q}\rho_k\frac{1}{N}\sum_{v=1}^{N}\left[x_{i(v)}(t_k) - x_{ii(v)}(t_k)\right]W_{is(v)}(t_k),$$

$$s = \overline{1,l}. \tag{7.92}$$

Воспользовавшись обозначениями:

$$\sum_{k=1}^{q}\rho_k\frac{1}{N}\sum_{v=1}^{N}W_{ir(v)}(t_k)W_{is(v)}(t_k) = a_{sr}; \tag{7.93}$$

$$\sum_{k=1}^{q}\rho_k\frac{1}{N}\sum_{v=1}^{N}\left[x_{i(v)}(t_k) - x_{ii(v)}(t_k)\right]W_{iS(v)}(t_k) = b_S, \tag{7.94}$$

из (7.92) получим систему алгебраических уравнений

$$\begin{aligned}
&a_{11}\Delta\theta_1 + a_{12}\Delta\theta_2 + ..... + a_{1l}\Delta\theta_3 = b_1;\\
&a_{21}\Delta\theta_1 + a_{22}\Delta\theta_2 + ..... + a_{2l}\Delta\theta_3 = b_2;\\
&\ldots\ldots\ldots

$$A \cdot \Delta\theta = B\,. \qquad (7.96)$$

Отсюда для вектора идентифицируемых параметров можно записать соотношение

$$\Delta\theta = A^{-1} \cdot B\,. \qquad (7.97)$$

Приведенные рассуждения показывают, что оценки являются линейными функциями результатов наблюдений; последние являются несмещенными и имеют наименьшую дисперсию среди всех линейных несмещенных оценок.

Число измерений определяется заданной точностью статистической обработки. Метод обобщается на нестационарные нелинейные объекты. В этом случае объект с переменными параметрами следует заменить на интервале времени $t_k, t_{k+1}$ системой уравнений с кусочно-постоянными коэффициентами и на этом интервале применить изложенный алгоритм идентификаций. Интервал $t_k, t_{k+1}$ является шагом идентификации [15].

Структурная схема системы идентификации представлена на рис. 7.14.

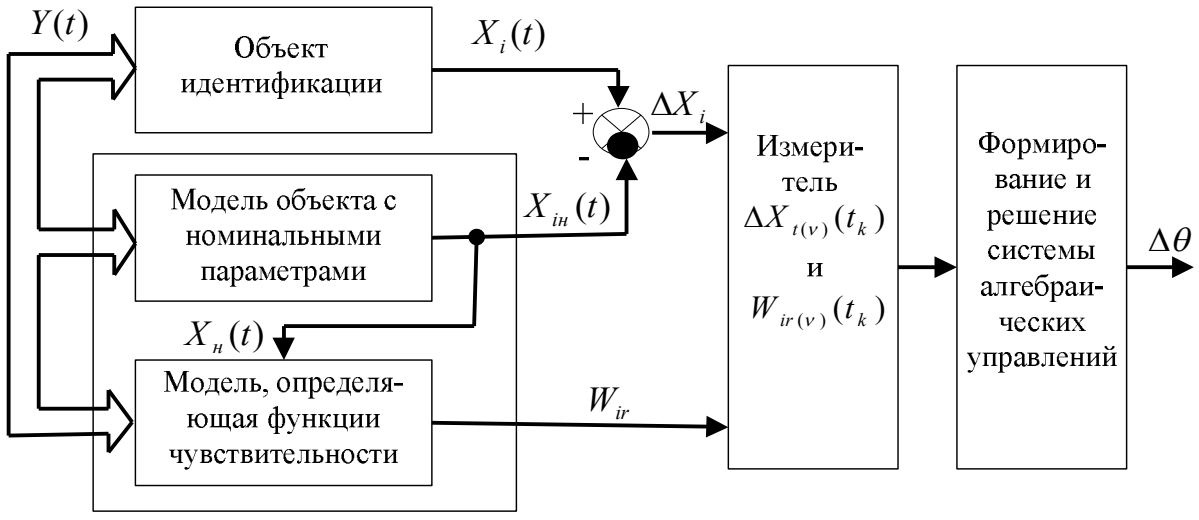



Рис. 7.14. Структурная схема системы идентификации

### 7.4.2. Оценка параметров нелинейных систем методами фильтрации

Положим [15, 20], что задано дифференциальное уравнение объекта идентификации

$$\dot{x} = f(x,\theta,t) + B(\theta,t)u(t), \qquad (7.98)$$

где $x(t)$ – $n$-мерная вектор-функция фазовых координат, $f(x,\theta,t)$ – нелинейная вектор-функция размерности $n$; $\theta$ – вектор случайных параметров, причем $\theta = \theta_i + \Delta\theta$, где $\theta_i = m_\theta$ – вектор номинальных значений параметров (эти значения заданы) и $\Delta\theta$ – вектор случайных малых отклонений; $B(\theta,t)$ – матрица коэффициентов; $u(t)$ – вектор-функция входного сигнала.

Известны необходимые статистические характеристики случайных параметров. Требуется определить оценку параметров $\theta$ по наблюдению на интервале $[0,T]$ сигнала

$$z(t) = H\Delta x(t) + v(t) \qquad (7.99)$$

по критерию минимума среднеквадратической ошибки; в формуле (7.99) $v(t)$ – помеха, $\Delta x(t) = x(t) - x_i(t)$, где $x_i(t)$ – сигнал на выходе модели объекта, имеющей номинальные значения параметров; $H$ – известная матрица.

Структурная схема объекта может быть представлена в виде, показанном на рис. 7.15.

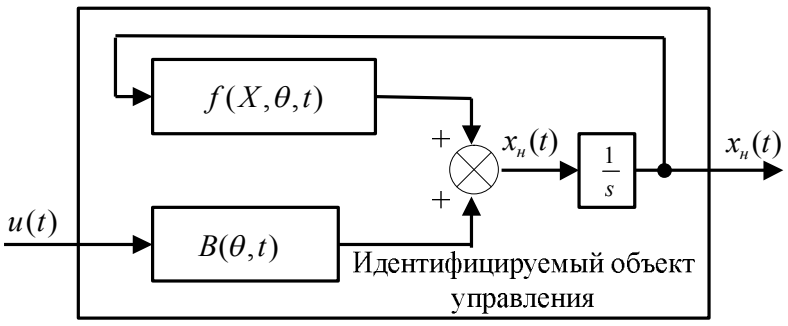



Рис. 7.15. Структурная схема объекта управления

Для решения задачи идентификации проведем некоторые преобразования. Введем в рассмотрение нелинейную модель объекта, имеющую номинальные значения параметров (полагаем, что номи-

нальные параметры системы равны своим математическим ожиданиям); уравнение такой модели имеет вид

$$\dot{x}_i = f(x_i, \theta_i t) + B(\theta_i, t)u(t)). \qquad (7.101)$$

В последнем уравнении $u(t)$ – входной сигнал, $x(t)$ – вектор фазовых координат модели.

Структурная схема для формирования сигнала $x_i(t)$ представлена на рис. 7.16.

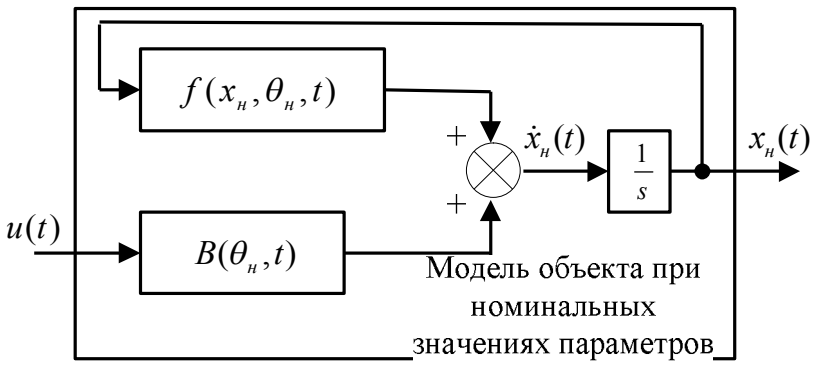



Рис. 7.16. Структурная схема модели объекта с номинальными значениями параметров

Линеаризуем нелинейную вектор-функцию $f(x,\theta,t)$ (воспользуемся рядом Тейлора), полагая при этом:

1) $f(x,\theta,t)$ и $B(\theta,t)$ – дифференцируемы по фазовым координатам и параметрам;

2) линеаризация проводится относительно опорного движения, определяемого уравнением (7.101);

3) отклонения параметров $\theta$ от номинальных значений параметров $\theta_i$ незначительны;

4) отклонения фазовых координат системы (7.98) от фазовых координат системы (7.101) малы.

Результат имеет вид [15,20]

$$\begin{aligned} f(x,\theta,t) &= f(x_i + \Delta x, \theta_i + \Delta\theta, t) = \\ &= f(x,\theta,t) + D_1(x_i, \theta_i, t)\Delta x + D_2(x_i, \theta_i, t)\Delta\theta, \end{aligned} \qquad (7.102)$$

где $D_1(x_i, \theta_i, t)$ – матрица производных векторной функции $f$ по компонентам вектора $x(t)$ при $x = x_i$, $\theta = \theta_i$; $D_2(x_i, \theta_i, t)$ – матрица производных векторной функции $f$ по компонентам вектора $\theta$ при $x = x_i$, $\theta = \theta_i$:

$$D_1(x_i,\theta_i,t)=\left\|\frac{\partial f_i(x_i,\theta_i,t)}{\partial x_j}\right\|;\ D_2(x_i,\theta_i,t)=\left\|\frac{\partial f_i(x_i,\theta_i,t)}{\partial \theta_k}\right\|;$$

$$\Delta x(t)=x(t)-x_n(t);\ \Delta\theta=\theta-\theta_i;\ i,\ j=\overline{1,n},\ k=\overline{1,l}.$$

Аналогично получим

$$B(\theta,t)=B(\theta_i,t)+D_3(\theta_i,t)\Delta\theta, \qquad (7.103)$$

где $D_3(\theta_i,t)$ – матрица производных от матрицы коэффициентов $B(\theta,t)$ по компонентам вектора $\theta$:

$$D_3(\theta_i,t)=\left\|\frac{\partial B_{ij}(\theta,t)}{\partial \theta_k}\right\| \text{ при } \theta=\theta_i, i,j\ \ j=\overline{1,n},\ k=\overline{1,l}.$$

Найдем линейное дифференциальное уравнение относительно $\Delta x(t)$

$$\Delta\dot{x}=f(x_i,\theta_i,t)+D_1(x_i,\theta_i,t)\Delta x+D_2(x_i,\theta_i,t)\Delta\theta+$$

$$+B(\theta_i,t)u(t)+D_3(\theta_i,t)\Delta\theta u(t)-f(x_i,\theta_i,t)-B(\theta_i,t)u(t)=$$

$$=D_1(x_i,\theta_i,t)\Delta x+D_2(x_i,\theta_i,t)\Delta\theta+D_3(\theta_i,t)\Delta\theta u(t). \qquad (7.104)$$

Перепишем последнее уравнение в виде

$$\Delta\dot{x}-D_1(x_i,\theta_i,t)\Delta x=(D_2(x_i,\theta_i,t)+D_3(\theta_i,t)u(t))\Delta\theta. \qquad (7.105)$$

Введем в рассмотрение матричную импульсную переходную функцию, определяемую уравнением

$$\frac{dK(t,\tau,x_i,\theta_i)}{dt}=D_1(x_i,\theta_i,t)K(t,\tau,x_i,\theta_i) \qquad (7.106)$$

при $K(t,\tau,x_i,\theta_i)=I$, где $I$ – единичная матрица.

Решение векторно-матричного дифференциального уравнения (7.105), выраженное в форме интеграла Коши, запишется в виде

$$\Delta x(t)=\int_0^t K(t,\tau,x_i,\theta_i)(D_2(x_i,\theta_i,\tau)+D_3(\theta_i,\tau)u(\tau))d\tau\cdot\Delta\theta. \qquad (7.107)$$

Воспользуемся обозначением

$$\hat{O}(t)=\int_0^t K(t,\tau,x_i,\theta_i)(D_2(x_i,\theta_i,\tau)+D_3(\theta_i,\tau)u(\tau))d\tau. \qquad (7.108)$$

Структурная схема вспомогательной системы, выходом которой является процесс $\hat{O}(t)$, представлена на рис. 7.17.

Формулу, определяющую $\Delta x(t)$ можно переписать так:

$$\Delta x(t) = \hat{O}(t)\Delta\theta . \qquad (7.109)$$

Наблюдению подвергается вектор-функция $Z(t)$, определяемый соотношением

$$Z(t) = H\Delta x(t) + v(t) = H\hat{O}(t)\Delta\theta + v(t), \qquad (7.110)$$

где $H$ – матрица, характеризующая измеритель; $v(t)$ – вектор-функция помех измерителя.

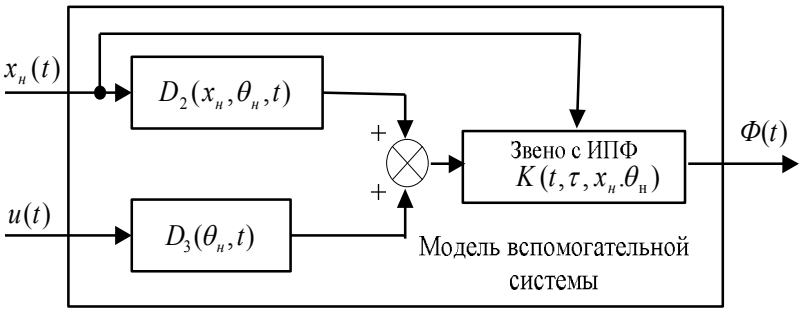



Рис. 7.17. Структурная схема вспомогательной системы

Воспользовавшись обозначением

$$S(t) = H\hat{O}(t), \qquad (7.111)$$

соотношение (7.110) перепишем в виде

$$Z(t) = S(t)\Delta\theta + v(t). \qquad (7.112)$$

Таким образом, получена формула (7.112), в которой $Z(t)$ линейно зависит от вектора случайных параметров системы $\Delta\theta$.

Поскольку имеет место случай наблюдения аддитивной помехи и полезного сигнала, линейно зависящего от вектора параметров, то оптимальная оценка параметров может быть получена в соответствии с формулой

$$\theta = (R_{\theta\theta}C + I)^{-1} R_{\theta\theta}Q, \qquad (7.113)$$

где $R_{\theta\theta}$– априорная корреляционная матрица вектора параметров; $I$ – единичная матрица;

$$C = \int_0^T K_0^T(\tau)S(\tau)d\tau; \quad Q = \int_0^T K_0^T(\tau)Z(\tau)d\tau, \qquad (7.114)$$

причем

$$\int_0^T K_0^T(\tau)Z(\tau)d\tau = S^T(t). \qquad (7.115)$$

Из (7.115) следует, что для нахождения матрицы $K_0(t)$ входящей в зависимости, определяющие матрицы C и Q, необходимо решить интегральное уравнение.

Алгоритм значительно упрощается, если $\nu(t)$ – белый шум. Тогда

$$R_{NN}(t_1,t_2) = S_0\delta(t_2 - t_1) \qquad (7.116)$$

и

$$K_0^T(t) = S^T(t)S_0^{-1};\ \ C = \int_0^T S^T(\tau)S_0^{-1}S(\tau)d\tau;\ \ Q = \int_0^T S^T(\tau)S_0^{-1}Z(\tau)d\tau. \qquad (7.117)$$

Апостериорная корреляционная матрица, характеризующая точность получения компонент вектора оценки $\Delta\theta$, имеет вид

$$R^*_{\theta\theta} = M(R_{\theta\theta}C + 1)^{-1}R_{\theta\theta}, \qquad (7.118)$$

где $M$ – оператор математического ожидания.

Структурная схема системы, позволяющая получить оценку вектора параметров: $\Delta\theta^*$ нелинейной системы, представлена на рис. 7.18.

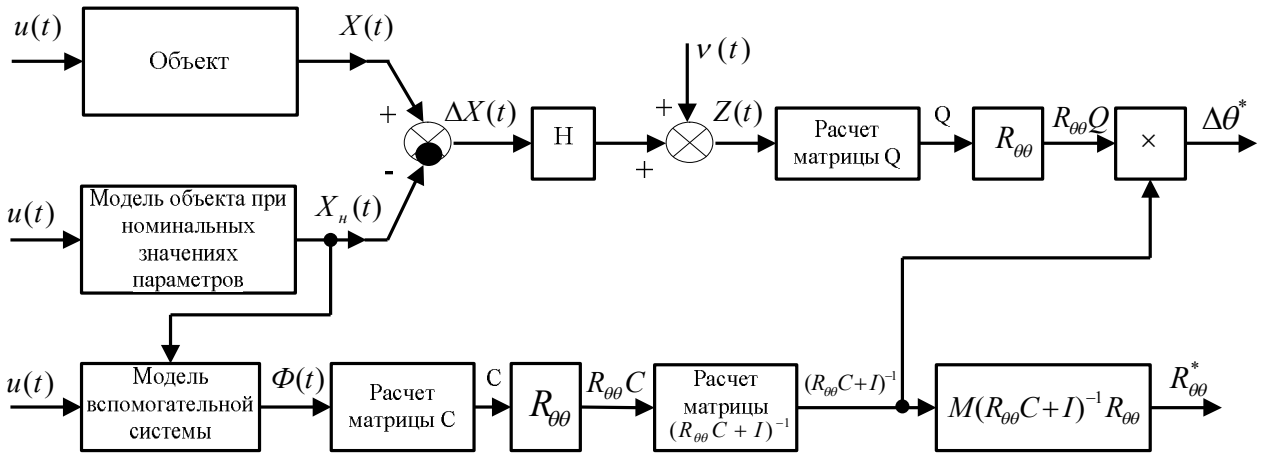



Рис. 7.18. Структурная схема системы для оценки вектора параметров $\Delta\theta$ нелинейной системы

Для получения алгоритма оценки вектора случайных параметров нелинейной системы можно использовать подход, содержание которого состоит в том, что параметры системы принимаются за дополнительные фазовые координаты расширенной нелинейной системы [15, 20].

# ГЛАВА 8. МЕТОДОЛОГИЧЕСКИЕ ПРИНЦИПЫ КОНСТРУИРОВАНИЯ НОВЫХ МЕТОДОВ ИДЕНТИФИКАЦИИ НА БАЗЕ ОБЩЕГО ПАРАМЕТРА

Как следует из выше представленных разделов в настоящее время предложено много разнообразных концепций и методов идентификации. При этом сформировался комплекс требований к предлагаемым концепциям и методам идентификации в виде следующих условий: простота реализации невысокая вычислительная эффективность; возможность использования в режиме нормальной работы; гарантированная точность решения задачи идентификации для конечных интервалов времени наблюдения; достаточно высокая скорость сходимости.

Известные методы идентификации позволяют удовлетворительно находить модели достаточно широкого класса объектов при выполнении в целом противоречивых вышеуказанных требований в том случае, когда число оцениваемых параметров не слишком велико. Однако обеспечение данных требований при оценивании вектора неизвестных параметров высокой размерности становится практически невозможным.

Более того, в общем случае неизвестны как порядок, так и запаздывание системы. Все это при решении практических задач оценивания параметров вектора большой размерности приводит к малой скорости сходимости известных методов идентификации. В тех случаях, когда время вычисления оценок параметров становится близким ко времени протекания технологического процесса, использование полученных оценок для целей адаптивного управления оказывается практически невозможным.

Трудности решения ряда важных практических задач, например, задач большой размерности, приводит к пересмотру определенных положений существующей теории и практики идентификации. В этих условиях асимптотическая точность оценок не является определяющей по сравнению с требованием быстро и возможно с меньшими затратами получить приближенное решение задачи идентификации. Решение этой задачи может быть достигнуто путем разработки таких методов и, алгоритмов, которые дают возможность получать оценки параметров моделей высокой размерности более

высокой точности при конечном числе наблюдений по сравнению с известными методами и алгоритмами.

## 8.1. Постановка задачи

Рассматривается общая задача идентификации сложной системы (объекта) *S* с некоторыми структурами, характеризуемые конечномерным вектором $\theta$, в режиме нормальной работы.

Предположение 1. Система *S* является полностью управляемой и наблюдаемой системой и представляется эквивалентной моделью $\boldsymbol{m}(\hat{\theta})$, допускающей параметризацию вектором параметров $\hat{\theta} \in R^n$, где $R^n$ – параметрическое пространство с метрикой $\rho(\theta, \hat{\theta})$.

Наблюдаются случайные временные последовательности входных воздействий $\{u(t)\}$ и выходных величин $\{x(t)\}$, порождаемых системой *S*, которые удовлетворяют определенным экспериментальным условиям *e*.

Предположение 2. Случайные процессы *(u,x)* имеют конечные дисперсии, причем $u(t) \in R^n, x(t) \in R^r, \forall t \in T$, где $T$ – множество целых чисел. Дастся совокупность $F_s$ функций $f(t, x, u, \hat{\theta}) \in R^r$.

Функция $f(t, x, u, \hat{\theta})$ по наблюдениям системы позволяет вычислять значения $\hat{x}(t / \hat{\theta})$, определяющие прогноз вектора x в момент времени *t*, т. е.

$$\hat{x}(t / \hat{\theta}) = f(t, x, u, \theta).$$

Предположение 3. Функция $f(t, x, u, \hat{\theta})$ является линейной по $\hat{\theta}$, но не обязательно линейной по *t, x, u* и представляет собой измеримую функцию, для которой имеет место соотношение

$\hat{x}(t) = f(t - \tau, z^{-\tau} x,\ z^{-\tau} u,\ \hat{\theta}) = f(t, x, u, \hat{\theta}), \forall \tau \in T$, где $z^{-\tau}\ \hat{x}(t) = x(t - \tau)$.

Так для системы с одним входом *u(t)* и одним выходом *x(t)* модель может быть представлена уравнением вида

$$\hat{x}_i(t) = \hat{\theta}^T f_i[z(t-1)], \qquad (8.1)$$

где $f_i[z(t-1)] - n_i$ – мерная вектор-функция; $\hat{\theta} - n_i$ – мерный вектор параметров модели. Элементами вектора *z(t-1)* могут быть прошлые значения входов $u(t - \hat{d}_i - 1), \ldots, u(t - \hat{d}_i - \hat{n}_i)$ и (или) выходов $x(t-1), \ldots, u(t - n_i)$ и

их произведения. Здесь $\hat{n}_i$, и $\hat{d}_i$ – порядок модели и запаздывание соответственно.

Множество всех возможных моделей, имеющих одну и ту же функцию $f(\bullet)$, но различающихся по $\hat{\theta}$, принадлежит определенному классу

$$C = \left\{ (f_i, \theta) : \hat{\theta} \in \Omega_i, i = \overline{1,s} \right\} = \left( f_i, \hat{\theta}, \hat{n}_i, \hat{d}_i \right).$$

Так модель (8.1), принадлежащая классу $C_i$ предъявляется совокупностью возможных значений параметров $\left( f_i, \hat{\theta}, \hat{n}_i, \hat{d}_i \right)$.

Предположение 4. Если последовательность образованная соотношением $v(t) = x(t) - \hat{x}(t)$ последовательности $\{ x(t) \}$ и модели (8.1) с параметрами $\left( f_i, \hat{\theta}, \hat{n}_i, \hat{d}_i \right)$ имеет нормальное распределение, то уравнение

$$\hat{x}_i(t) = \hat{\theta}^T f_i[z(t-1)] + v_i(t). \qquad (8.2)$$

описывает действительные условия функционирования системы, а параметры модели являются истинными.

Тогда задача идентификации заключается в оценке выбора $\hat{\theta}$ по экспериментальным данным ***e*** таким образом, чтобы по наблюдениям за входными и выходными величинами $\{z(t)\}$ достичь определенного соответствия модели $m\left(\hat{\theta}\right)$ системе ***s***, которое оценивается критерием качества $J\left(\hat{\theta}\right)$. Решение данной задачи возможно получить с использованием различных известных методов идентификации ***i***, которые обеспечивают сходимость оценки $\hat{\theta}(t)$ к θ в определенном смысле при $t \to \infty$.

В этом случае систему идентификации *I* можно описать четверкой {***s, m, e, i***} и оценку параметров в момент времени *t* обозначить как $\hat{\theta}(t) = \hat{\theta}(t; I)$. Отметим, что $\hat{\theta}(t; I)$ может рассматриваться как детерминированная функция или как случайная последовательность. Введем область определения значений параметров модели вида

$$P = \left\{ \hat{\theta} \in R^n; \quad \lim_{t \to \infty} \rho [\hat{\theta}(t; I), \theta] = \delta < \infty \right\}.$$

Предположение 5. Область $R^n$ является непустой. При $\delta = 0$ область *P* состоит из единственного элемента, в противном случае $\delta > 0$.

Из предположения 5 следует следующее определение.

Определение 1. Система ***s*** идентифицируема для *I*, если $\delta > 0$ и параметрический идентифицируема, если $\delta = 0$.

Данное определение приведено независимо от методов оценивания ***i*** и предполагает такие структуры рекуррентных алгоритмов, которые обеспечивают выполнение условия сходимости

$$\rho\,[\,\hat{\theta}\,(t;\,I\,),\,\theta\,] \;<\; \rho\,[\,\hat{\theta}\,(t\,-\,1;\;I\,),\theta\;]\;.$$

В практических задачах обычно имеет место малая точность исходных данных, а также несоблюдение предположений об условиях сходимости последовательности оценок $\hat{\theta}(t;I)$ к их истинным значениям $\theta$. Это вызвано выбором конструкций самих моделей для описания системы, реализацией схем алгоритмов различных методов, а также свойствами исходных данных и наблюдений, при которых сходимость $\hat{\theta}(t;I)$ к θ не гарантируется для ряда промежутков или всего интервала времени наблюдения. Особо характерно это для задачи идентификации высокой размерности, для которых чрезмерно возрастает вычислительная сложность оценивания параметров и вероятность нарушения условий точного решения задачи.

Все это приводит к необходимости введения условного понятия «объекта высокой размерности» не только от числа оцениваемых параметров, но и от свойств и условий функционирования идентифицируемого объекта и системы идентификации, т. к. качество идентификации тесно связано с количеством оцениваемых параметров, временем функционирования объекта, ограниченным числом измерений, свойствами входных сигналов и т. п.

Определение 2. Если для элементов {***s, m, e, i***} системы идентификации I параметрические оценки заданной точности не могут быть получены известными методами идентификации, то классифицируемая система *s* условно определяется как "система или объект высокой размерности".

Введенное определение не связано с известным понятием идентифицируемости системы, поскольку понятие идентифицируемости определены, исходя из принципиальной возможности оценивания параметров модели известной структуры. Так, например, идентифицируемость параметров может иметь место при выбранной структуре модели, однако, выбор метода и алгоритмов идентификации, технических средств измерения, условия функционирования объекта во многих задачах не обеспечивают заданной точности оценок параметров при значениях интервалов наблюдений. Отмеченные обстоятельства обуславливают необходимость разработки упрощенных концеп-

ций конструирования рекуррентных алгоритмов оценки параметров, характеризуемых как значительным быстродействием, так и простотой их реализации. В этом случае отмеченная выше задача идентификации может быть решена с учетом применения новых конструкций моделей $\boldsymbol{m_k}$ и схем оценивания $\boldsymbol{i_k}$ вектора параметров $\theta$. Тогда система идентификации $I_k$ определяется четверкой $\{\boldsymbol{s, m_k, e, i_k}\}$. Оценку параметров в момент времени $t$ можно обозначить как $\hat{\theta}(\mathrm{t}) = \hat{\theta}\ (\mathrm{t};\ \mathrm{I_k})$.

Требования по быстродействию конструируемых алгоритмов должны обеспечивать в виде соблюдения условия

$$\rho\,[\ \hat{\theta}\,(t;\ I\,),\ \theta] \ >\ \rho\,[\ \hat{\theta}\,(t;\ I_{\,k}\,),\theta\ ]\ , \qquad (8.3)$$

где $I, I_k$ – соответственно обозначают известные и конструируемые системы идентификации, используемые для нахождения параметра $\theta$. При этом задача оценивания с учетом требований по быстродействию тесно связана с акселеризацией оценок. Под акселеризацией оценок (по Я.З. Цыпкину) [29] было принято называть "любые способы ускорить получения оценок заданной точности и улучшать асимптотические свойства". Естественно, что решение данной задачи идентификации возможно проводить за счет усложнения реализации алгоритмов, т. е. всей структуры системы идентификации, что обычно и предлагают известные подходы.

Простоту реализуемости конструируемых алгоритмов и систем идентификации "объектов высокой размерности" необходимо обеспечивать на основе таких методологических подходов, которые позволяют создавать системы идентификации более простой структурной реализации и более эффективные по объему вычислений по сравнению с известными системами идентификации. В этом случае, если можно получить приближенные решения задачи при выполнении условий (8.3), исходные структуры, реализующие алгоритмы известных систем идентификации $I$ оценивания параметров, удовлетворяющие определенным экспериментальным условиям $\boldsymbol{e}$, можно считать условно избыточными или просто избыточными структурами относительно структур конструируемых систем идентификаций $\mathrm{I_k}$ при тех же самых условиях $\boldsymbol{e}$.

При разработке новых подходов идентификации является очевидным, чтобы они обеспечивали более высокую эффективность их использования относительно известных методов оценивания. Будем теперь полагать, что система идентификации $I_k$ с элементами $\{\boldsymbol{s, m_k, e, i_k}\}$

более проста как в реализации, так и качеству ее функционирования относительно системы идентификации $I$ с элементами $\{\boldsymbol{s, m, e, i}\}$. Тогда система идентификации $I$ является структурно-избыточной или условно-избыточной относительно системы $I_k$ при конечных интервалах времени наблюдения.

Физическая интерпретация данного положения состоит в следующем. Известные подходы идентификации предопределяют такие схемы рекуррентных алгоритмов обновления оценок, которые базируются на одновременном изменении всех параметров модели по результатам эксперимента. Практическая реализация такой процедуры представляет большие трудности, поскольку чрезмерно возрастает вычислительная сложность оценивания параметров, особенно для моделей высокой размерности.

Однако рассматриваемые системы с точки зрения решения задачи идентификации не является структурно-избыточными, несмотря на их сложность. Структурная избыточность системы идентификации $I$ имеет место только относительно такой системы идентификации $I_k$, которая адекватно решает задачу идентификации на базе новых подходов. При этом структурная реализация, вычислительная сложность для системы $I_k$ значительно проще, чем для системы $I$.

Когда имеет место параметрическая идентифицируемость, то для обеих систем идентификации $I$ и $I_k$ при $\mathrm{t} \to \infty$ оценки $\hat{\theta}(t)$, сходятся к одному и тому же параметру $\theta$ или $\rho[\hat{\theta}(\mathrm{t; I}), \theta] \to \rho[\hat{\theta}(\mathrm{t; I_k}), \theta] \to 0$.

В практических условиях оценивания параметров имеют большое значение свойства системы идентификации при конечных интервалах времени наблюдения. Если $\rho[\hat{\theta}(\mathrm{t; I}), \theta] > \rho[\hat{\theta}(\mathrm{t; I_k}), \theta]$ для всех $t \in T$, то система $I$ является условно избыточной относительно системы $I_k$ при конечных интервалах времени наблюдения. В этих случаях система $I_k$ имеет более высокую скорость сходимости и точность оценивания относительно системы $I$. При этом начальные значения не оказывают существенного влияния на асимптотические свойства системы $I_\mathrm{k}$.

Таким образом, развитие теории и практики идентификации в направлении разработки и внедрения новых методов и алгоритмов идентификации "объектов высокой размерности" (в соответствии с введенным определением) должно базироваться на таких методологических подходах, которые позволяют создавать системы идентификации более

простые в структурной реализации и эффективные в смысле свойств асимптотического поведения, чем известные системы идентификации.

В аспекте сформулированной проблемы методология, теория, математические модели и алгоритмы методов общего параметра [4, 5, 26, 27] могут служить основой для построения систем идентификации объектов и процессов широкого класса с обеспечением требований по быстродействию и простоте реализации, а также улучшения существующих методов идентификации.

## 8.2. Теоретическое обоснование условий избыточности

Предположим теперь, что система описывается уравнением вида

$$x(t) = f[\theta, u(t)] + \nu(t), \qquad (8.4)$$

где $x(t)$ – выходная переменная; $u(t)$ – вектор, элементами которого могут быть прошлые значения входных воздействий и (или) выходных величин; $\theta$ вектор неизвестных параметров; $\nu(t)$ – помеха: $f(\bullet)$ – заданная линейная или нелинейная функция.

Если модель представлена аналогичной структурой с учетом общего параметра в виде

$$\hat{x}(t) = f[\hat{\theta}, u(t), \beta(t)], \qquad (8.5)$$

то задачу оценки параметров возможно свести к определению такого вектора параметров $\hat{\theta}$ модели путем изменения общего параметра $\beta$, чтобы обеспечить в некотором смысле приближение последовательности модели $\hat{x}(t)$ к выходу системы $x(t)$. Случай, когда задача оценивания θ решается путем изменения вектора параметров $\hat{\theta}$ при заданном $\beta$, представляет собой широко известные процедуры идентификации. Возможны и комбинированные методы, сочетающие оба подхода в различных схемах идентификации метода общего параметра.

Выбор того или иного метода идентификации должен определяться с учетом вычислительных проблем, априорной информации качества метода (точность, сходимость). Наиболее важным является качество метода. Однако сравнительная оценка различных методов идентификации представляет собой большие трудности.

Сравнительный анализ качества оценивания методом общего параметра относительно известных методов проведен, когда задача идентификации, решается с использованием модели параллельного типа (рис. 8.1).

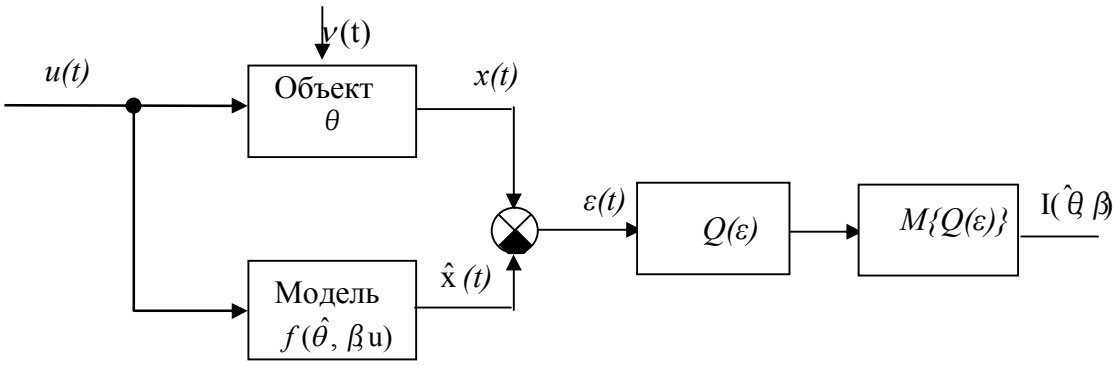



Рис. 8.1. Модель параллельного типа для идентификации объекта

Здесь процесс идентификации осуществляется на определенном интервале времени таким образом, чтобы по известным значениям входного воздействия u(t) и выходной величины объекта *x(t)* при наличии случайной ненаблюдаемой помехи $\nu(t)$ *t=1,...,N* воздействующей на объект, определить вектор неизвестных параметров θ. В реальных условиях сигналы *u(t), x(t)* искажены помехами измерения. Для решения задачи идентификации заданные значения *u(t)* подаются на вход модели объекта. Выход настраиваемой модели $\hat{x}(t)$ рассчитывается с учетом значений оценок $\hat{\theta}$ при традиционном подходе индивидуального изменения всех параметров модели по наблюдениям входного воздействия, выходных величин объекта и модели. Для предлагаемого метода осуществляется изменение параметра $\beta$. Возможны и различные комбинированные реализации данных подходов. Разность выходных сигналов объектов и настраиваемой модели $\varepsilon(t) = x(t) - \hat{x}(t)$ используется для формирования критерия качества идентификации

$$J(\hat{\theta}) = M\left\{ Q\left[ \varepsilon(t), \hat{\theta}, \beta \right] \right\},$$

где $Q(\cdot)$ – заданная функция потерь; $\hat{\theta}$ – $n$-мерный вектор оценок параметров; $\beta$ – общий параметр. Рассматриваемая схема идентификации также справедлива, когда сигналы *u(t), x(t),* $\hat{x}(t)$, $\nu(t)$ являются векторами.

Минимизация критерии качества может осуществляться соответствующим выбором структуры настраиваемой модели и целенаправленным изменением ее параметров алгоритмами идентификации. При таком подходе рассматриваемый метод приводит к получению формального математического описания объекта, устанавливающего

связь между входными и выходными переменными. В общем случае структуры объекта и настраиваемой модели могут быть различны.

Если структура объекта задана с точностью до вектора неизвестных параметров, то ставится задача оценивания параметров и близость структур объема и модели понимается смысле близости истинных значений параметров θ и их оценок $\hat{\theta}$.

Пусть модель (8.5) представлена без учета общего параметра как

$$\hat{x}(t) = \hat{\theta}^T(t-1)u(t).$$

Тогда функция потерь имеет вид

$$Q\left[\varepsilon(t),\hat{\theta}\right] = \frac{1}{2}\left[\hat{x}(t) - \hat{\theta}^T(t-1)u(t)\right]^2. \qquad (8.6)$$

Для оценивания параметров воспользуемся рекуррентной процедурой вида

$$\hat{\theta}(t) = \hat{\theta}(t-1) + \tilde{A}(t)\left[x(t) - \theta^T(t-1)u(t)\right]u(t), \qquad (8.7)$$

где $\tilde{A}(t)$ – скалярная матрица усиления.

Рассмотрим задачу оценивания параметров, когда дискретная система идентификации представлена уравнениями (8.4), (8.6), (8.7). Положим, что имеет место коррелированная связь между входными переменными. Данное условие при оценивании вектора параметров высокой размерности в условиях нормальной работы фактически всегда имеет место.

Как может быть показано, что наличие корреляции между входными переменными приводит к избыточности системы идентификации при ее функционировании, по крайней мере, на начальных этапах оценивания параметров. Этот вывод имеет большое практическое значение, особенно когда процедура идентификации ограничена конечным моментом времени наблюдений.

С этой целью рассмотрим алгоритм (8.7) и определим для него матрицу усиления в виде $\tilde{A}(t) = [\gamma E]$, где $E$ – единичная матрица, $\gamma > 0$. Если представить параметрический вектор ошибки как $r(t) = \theta - \hat{\theta}(t)$, то алгоритм (8.7) можно записать в виде

$$r(t) = \left[E - \gamma u(t)u^T(t)\right]r(t-1). \qquad (8.8)$$

Пусть входная последовательность u(t) удовлетворяет условиям

$$M\left\{u(t)u^T(t)\right\}=\sigma^2 K=\sigma^2\begin{bmatrix}1....k_{ij}\\ .....\\ .....\\ k_{ij}...1\end{bmatrix},\ M\left\{u(t)\right\}=0, \qquad (8.9)$$

где все компоненты случайного вектора u(t) имеют нулевые средние значения, равные дисперсии, и ковариации $\sigma^2 k_{ij}$. Примем $k_{ij}=k<1(i\neq j)$. Хотя такое распределение входных сигналов представляет собой частный случай, оно позволяет вскрыть ряд характерных свойств алгоритмов. При этом более существенным оказывается не вид распределения, а равенство всех недиагональных элементов матрицы *K*. Однако при неравенстве коэффициентов корреляции можно взять вместо $k_{ij}$ максимальный коэффициент корреляции *k*, что приводит к ограничениям γ сверху и для $M\{\ \|\ r(t)\ \|^2\ \}$ снизу.

Используя выражение (8.8), запишем значение квадрата нормы параметрических рассогласований в виде

$$\|r(t)\|^2=r^T(t-1)\left(E-2\gamma u(t)u^T(t)+\gamma^2\left(u(t)u^T(t)\right)^2\right)r(t-1). \qquad (8.10)$$

Тогда среднее значение для (8.10) с учетом предположений о входных воздействиях (8.9) можно записать в виде

$$M\left\{\|r(t)\|^2\right\}= \qquad (8.11)$$

$$=\left[1-2\gamma\sigma^2\left(1-k+kn\lambda^2(t-1)\right)+\gamma^2\sigma^4\left(c(1-k)+(c+2d)kn\lambda^2(t-1)\right)\right]M\left\{\|r(t-1)\|^2\right\},$$

где $c=n+2(1-k),\ d=1+(n-1)k$,

$\lambda^2(t)=\dfrac{\left(r^T(t)I\right)^2}{nr^T(t)r(t)}$, I – единичный вектор.

С учетом оптимального значения γ (t) вида

$$\gamma(t)=\frac{(1-k)+kn\lambda^2(t-1)}{\left[c(1-k)+(c+2d)kn\lambda^2(t-1)\right]\sigma^2}$$

выражение (8.11) примет вид

$$M\left\{\|r(t)\|^2\right\}=f(n,k,\lambda,t)M\|r(t-1)\|^2, \qquad (8.12)$$

где

$$f(n,k,\lambda,t)=1-\frac{\left(1-k+n\lambda^2(t-1)\right)^2}{\left[c(1-k)+(c+2d)kn\lambda^2(t-1)\right]}.$$

Как следует из (8.12) при независимых входных переменных *k=0* средняя ошибка после *t* шагов оценивания имеет значение

$$M\left\{\|r(t)\|^2\right\}=\left(1-\frac{1}{n+2}\right)^t\|r(0)\|^2. \qquad (8.13)$$

Из (8.13) видно, что ошибка идентификации уменьшается по экспоненциальному закону и скорость сходимости зависит от числа оцениваемых параметров *n*. При этом выражение (8.13) характеризует максимально возможное уменьшение ошибки на каждом шаге оценивания, когда $\gamma=\left[(n+2)\sigma^2\right]^{-1}$ выполняются предположения для свойств входных и выходных сигналов. Поэтому причины или подходы, при которых на определенных этапах оценивания n параметров имеет место выполнение условия $\left(1-\frac{1}{n+2}\right)>f(n,k,\lambda,t)$ обладает акселерантными свойствами, т. е. обуславливают возможность оценивания параметров более высокой точности, чем для (8.13).

Рассмотрим теперь выражение (8.12), (8.13). Очевидно, что для выбора оптимального значения γ необходимо знать неизвестную величину $\lambda^2(t)$ на каждом шаге. На рис. 8.2 представлена зависимость $M\{\|r(t)\|^2\}/\|r(0)\|^2$ времени для различных значений *k* при $\lambda=const$. Рассмотрим характерные случаи задачи идентификации, которые позволяют интерпретировать основные результаты оценивания параметр при наличии корреляции между входными переменными.

1. Пусть $\lambda^2(t)=0$. Тогда выражение (8.11) при оптимальном $\gamma=1/c\,\sigma^2$ от начальной ошибки после *t* шагов оценивания можно записать в виде

$$M\left\{\|r(t)\|^2\right\}=\left(1-\frac{1-k}{n+2(1-k)}\right)^t\|r(0)\|^2. \qquad (8.14)$$

Отсюда следует, что при k=l уточнение оценок не представляется возможным. При *k=0* справедливо выражение (8.13). На рис. 8.2 данные условия для (8.14) приведены соответственно линией *1* и кривой 0 при $n=10$. Все промежуточные кривые процесса точности оценивания для $0\le k\le 1$ находятся в области *I* над кривой *0*, т. е. скорость сходимости оценивания параметров с увеличением корреляции между входными переменными уменьшается. В общем случае данное положение хорошо известно в теории идентификации.

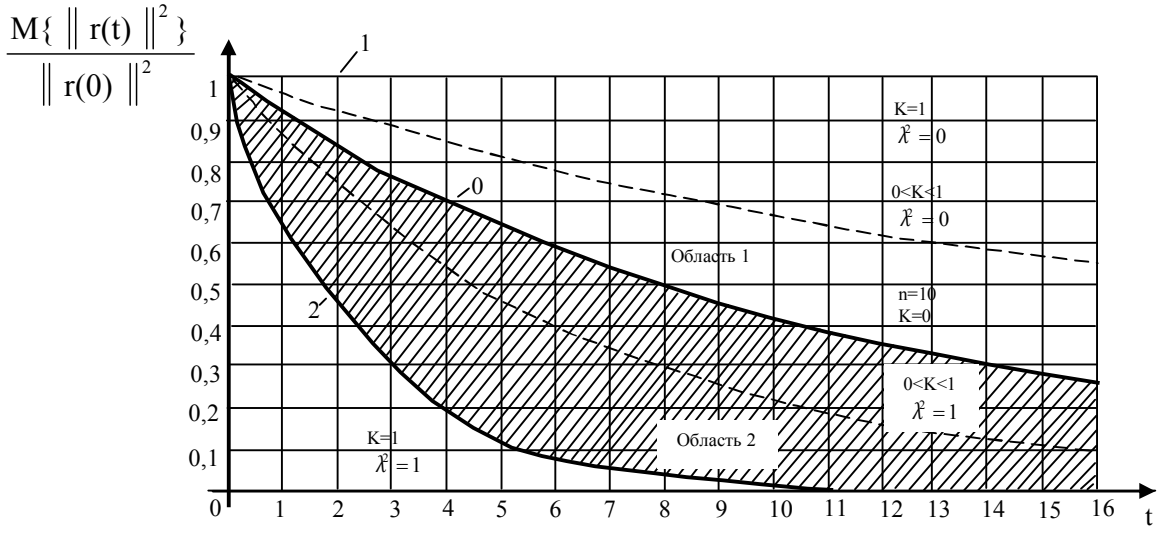



Рис. 8.2. Зависимости выражения $M\{\|r(t)\|^2\}/\|r(0)\|^2$ от времени для разных значений $K$ при $\lambda = const$

2. Пусть $\chi^2(t) = 1$. Тогда выражение (8.11) от начальной ошибки после t шагов при оптимальном $\gamma$ имеет вид

$$M\left\{\|r(t)\|^2\right\} = \left(1 - \frac{1 + (n-1)k)^2}{c(1-k) + (c+2d)kn}\right)^t \|r(0)\|^2. \qquad (8.15)$$

Из (8.15) следует, что при $k=1$ справедливо выражение

$$M\left\{\|r(t)\|^2\right\} = \left(1 - \frac{1}{3}\right)^t \|r(0)\|^2. \qquad (8.16)$$

а при $k=0$ выражение (8.13). На рис. 8.2 данные условия для (8.13) и (8.16) приведены соответственно кривыми 0 и 2. Все кривые, характеризующие точность оценивания для промежуточных значений $0 \le k \le 1$ находится в области *2* под кривой *0*. В этом случае скорость сходимости оценивания параметров с увеличением корреляции между входными переменными увеличивается. Действительно, из (8.13), (8.15) следует, что

$$\left(1 - \frac{1}{n+2}\right) > \left(1 - \frac{(1-k+kn)^2}{c(1-k)+(c+2d)kn}\right) = \left(1 - \frac{1}{m+2}\right). \qquad (8.17)$$

Таким образом, хотя для получения оценок используется алгоритм (8.7) настройки всех n параметров модели, однако, при наличии корреляции между входными переменными система идентификации фактически эквивалентна системе идентификации по скорости схо-

димости оценивания m параметров ($m<n$). Так, например, для случая, когда $k=0.5$, $k=1$ , то значения $m$, исходя из выражения (8.17), соответственно равны $m=4$, $m=1$.

При этом корреляция между входными переменными влияет на скорость сходимости оценивание параметров и определяет наличие в системе идентификации избыточность каналов настройки параметров модели ($n<m$). При этом, чем больше корреляционная зависимость, тем выше избыточность в системе идентификации и её влияние на оценивание параметров.

В этом случае векторное параметрическое управление эквивалентно скалярному управлению. Оценка параметров может быть представлена в виде $\hat{\theta}(t)=\hat{\theta}^*(t)+\beta(t)I$, где I – единичный вектор; $\beta(t)$ – настраиваемый общий параметр. При этом теперь определены два канала настройки параметров: основной векторный канал и дополнительный скалярный канал. Дополнительный канал влияет на основной канал так, что при слабой корреляции между входными переменными осуществляется настройка параметров основным векторным каналом, а при сильной корреляции – дополнительным скалярным каналом путем настройки общего параметра $\beta(t)$.

Таким образом, при увеличении корреляции между входными переменными увеличивается влияние дополнительного канала, и схема оценивания n параметров становится избыточной, т. е. может быть реализована с использованием каналов настройки параметров пониженного порядка, что фактически определено выражением (8.17).

Выше рассмотренные два случая оценивания вектора параметров высокой размерности позволили провести анализ характерных свойств процессов идентификации при наличии корреляции между входными переменными. Однако следует отметить, что значение $\lambda^2(t)$ в отличие от значения $k$ в процессе оценивания не является постоянным. Хотя, в общем случае, статистические свойства входных сигналов также изменяются в процессе идентификации. В связи с этим, основываясь на выводах случаев 1 и 2, рассмотрим оценивания параметров, когда значения $\lambda^2(t)$ являются переменными.

3. Пусть $\lambda^2(t)\neq\text{const}$. Положим значение $k=1$. Тогда при оптимальном $\gamma=\left(3n\sigma^2\right)^{-1}$ выражение (8.11) представимо в виде

$$M\left\{\| r(t)\|^2\right\} = M\left\{\left(1-\frac{\lambda^2(t-1)}{3}\right)\| r(t-1)\|^2\right\}. \qquad (8.18)$$

После t шагов оценивания выражение (8.18) от начальной ошибки можно записать в виде

$$M\left\{\| r(t)\|^2\right\} = M\left\{\prod_{j=1}^{t}\left(1-\frac{\lambda^2(j-1)}{3}\right)\right\}\| r(0)\|^2.$$

Поскольку в процессе настройки параметров $\lim\limits_{t\to\infty}\lambda^2(t)\to 0$ то среднее значение квадрата нормы параметрических рассогласований для установившегося состояния можно записать в виде

$$M\left\{\| r(t)\|^2\right\} = \left[1-\lambda^2(0)\right]\| r(0)\|^2. \qquad (8.19)$$

На рис. 8.3 представлена зависимость $M\left\{\| r(t)\|^2\right\}/\| r(0)\|^2$ от времени для различных значений k при $\lambda^2(t)\neq const$. Из (8.19) ясно, что точное оценивание возможно, когда $\lambda^2(0)=1$. Тогда кривая 2 на рис. 8.3 показывает изменение ошибки параметров для данного случая. При $\lambda^2(0)\neq 1$ ошибка оценивания в установившемся состоянии будет принимать различные значения в зависимости от начальных значений $\lambda^2(0)$. Характерные кривые, отражающие изменение ошибки оценивания при $\lambda^2(0)\neq 1$, приведены в виде кривых 3 и $3^*$. Если теперь и значение $k=l$, то согласно выражения (8.12) ошибка оценивания с увеличением числа наблюдений стремится к нулю (кривые 4 и $4^*$).

Кривые 3 и 4 на рис. 8.3 показывают, что на начальном этапе до моментов времени $t_3$, $t_4$ точность оценивания $n$ параметров при наличии корреляции между входными переменными выше, чем при отсутствии корреляции и кривые изменения ошибок расположены в области 2. В этом случае, как было отмечено выше, имеет место избыточность каналов настройки и фактически схема оценивания эквивалентна схеме пониженного порядка $m$ $(m<n)$. Уточнение ошибок оценивания после моментов времени $t_3$, $t_4$ когда кривые 3 и 4 расположены в области $I$ свойства системы идентификации меняются на противоположные, т. е. точность оценивания при наличии корреляции между входными переменными становится ниже, чем при отсутствии корреляции.

Кривые 3* и 4* находятся в области 1 и также показывают изменение ошибок при аналогичных ситуациях, что и для кривых 3 и 4

при малых значениях $\lambda(0)$. Тогда условие избыточности каналов настройки (8.17) не имеют место и точность оценивания ухудшается.

Однако в теории и практике идентификации решение задачи оценивания вектора параметров высокой размерности при корреляции между входными переменными осуществляются именно для данного случая. Это связано как с произвольным выбором начальных условий, которые определяют поведение системы в области 1 (рис. 8.3), так и стремлением получать состоятельные при большом интервале "наблюдений до конца", когда уточнение ошибок при увеличении числа наблюдений характеризуется изменением условий оценивания при переходе системы из области 2 в область 1.

Для этих случаев наличие корреляции вызывает нежелательные эффекты, которые отражаются на точность и скорость оценивания. Здесь поведение истинной системы идентификации фактически эквивалентно поведению системы более высокого порядка при отсутствии корреляции между входными сигналами.

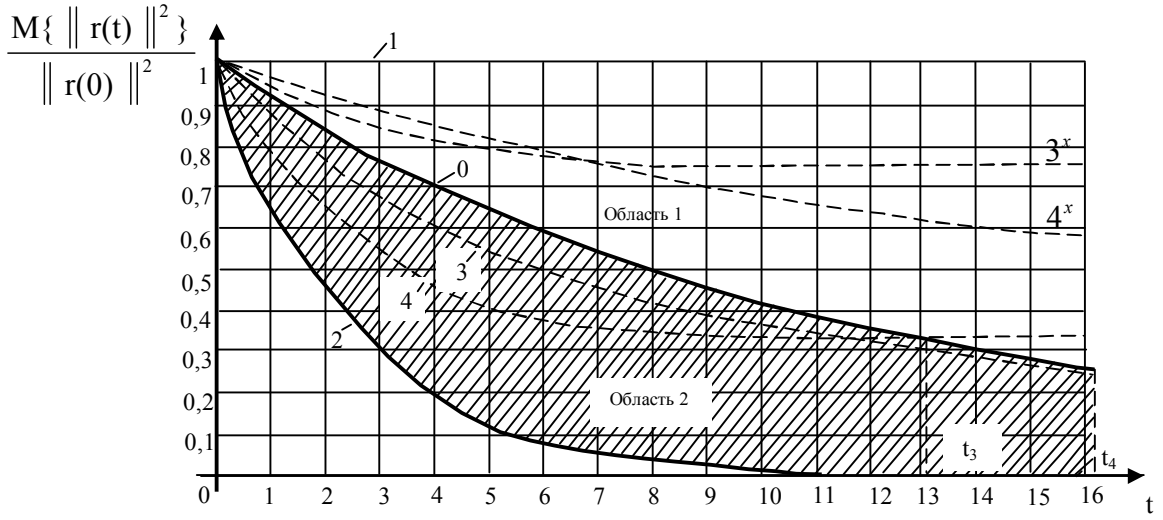



Рис. 8.3. Зависимости выражения $M\{\|r(t)\|^2\}/\|r(0)\|^2$ от времени для различных значений К при $\lambda(t) \neq const$

Таким образом, все рассмотренные выше различные причины приводят к избыточности системы идентификации объектов высокой размерности. Очевидно, что сложная природа поведения системы идентификации не исключает и другие причины, которые приводят к ее избыточности. Поэтому во многих практических ситуациях тради-

ционный подход к оцениванию вектора параметров высокой размерности одновременной настройки всех параметров модели ("канал – параметр настройки") оказывается малоэффективным. Это связано как с вычислительными проблемами, реализацией алгоритмов идентификации, так и малой скоростью сходимости.

Чтобы полностью устранить данные проблемы необходимо применять такие методы оценивания, которые основаны на стратегии понижения порядка настраиваемых параметров. Это возможно осуществить, если принять во внимание и использовать естественное явление избыточности, вызванное различными причинами, связанными как с самой системой идентификации, так и с внешней средой.

## БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. *Александров А.Г.* Оптимальные и адаптивные системы. –М.: Высш. шк. 2003. – 278 с.

2. *Андриевский Б.Р., Фрадков А.Л.* Избранные главы теории автоматического управления с примерами на языке MATLAB. –СПб.: Наука, 2000. – 475 с.

3. *Антонов В.Н., Терехов В.А., Тюкин И.В.* Адаптивное управление в технических системах. – СПб.; СПбГУ, -2001.

4. *Akhmetov D., Dote Y., Syzdykov D.* Fuzzy System Identification with General Parameter Radical Basis Function Neural Network / Smart Engineering Systems. New York; Asme Press, 1998, v.8.

5. *Ашимов А.А., Сыздыков Д.Ж.* Идентификация методом общего параметра// Справочник по теории автоматического управления / Под ред. *А.А. Красовского*. –М.: Наука, 1987, С. 263–271.

6. *Бокс Дж., Дженкис Г.* Анализ временных рядов. Прогноз и управление. –М.: Мир, 1974. – 408 с.

7. *Бородюк В.П.* Статистические методы математического описания сложных объектов. Учеб. пособие. –М.: МЭИ, 1981. – 92 с.

8. *Вольтера В.* Теория функционалов, интегральных и интегродифференциальных уравнений. –М.: Наука, 1982. – 304 с.

9. *Гроп Д.* Методы идентификации систем. –М.: Мир, 1979. – 302 с.

10. *Дмитриев А.К., Юсупов Р.М.* Идентификация и техническая диагностика. –МО СССР, 1987. – 580 с.

11. *Каминскас В.* Идентификация динамических систем по дискретным наблюдениям. ч.2. – Вильнюс: Мокслас, 1985. – 247 с.

12. Идентификация и оптимизация нелинейных стохастических систем / *Попков Ю.С., Кисчелев О.Н., Петров Н.П. и др.* – М.: Энергия, 1976. – 440 с.

13. *Кашьяп Р.Л., Рао А.Р.* Построение динамических стохастических моделей по экспериментальным данным. –М.: Мир, 1983. – 384 с.

14. *Кузовков Н.Т.* Модальное управление и наблюдающие устройства. –М.: Наука. 1976. – 184 с.

15. *Казаков И.Е.* Статистическая теория систем управления в пространстве состояний. –М.: Наука, 1975. – 270 с.

16. *Ли Р.* Оптимальные оценки, определение характеристик и управление. –М.: Наука, 1966. – 176 с.

17. *Льюнг Л.* Идентификация систем. Теория для пользователя. –М.: Наука, 1991. – 432 с.

18. *Лурье Б.Я., Энрайт П.Д.* Классические методы автоматического управления. –СПб.: БХВ – Петербург, 2004, – 520 с.

19. *Мирошник И.В., Никифоров В.О., Фрадков А.Л.* Нелинейное и адаптивное управление сложными динамическими системами. – СПб.: Наука, 2000. – 549 с.

20. Методы классической и современной теории автоматического управления: Учебник в 5-и тт.; Т2: Статистическая динамика и идентификация систем автоматического управления /Под. ред. *К.А. Пупкова, Н.Д. Егумова*. –М.: МГТУ им. Баумана, 2004. – 640 с.

21. *Рао С.Р.* Линейные статистические методы и их применение. –М.: Наука, 1968. – 548 с.

22. *Райбман Н.С., Чадеев В.М.* Построение моделей процессов производства. –М.: Энергия, 1975. – 375 с.

23. *Растригин Л.А., Моджаров Н.Е.* Введение в идентификацию объектов управления. –М.: Энергия, 1977. – 216 с.

24. *Саридис Дж.* Самоорганизующиеся стохастические системы управления. –М.: Наука, 1980. – 400с.

25. *Сейдж Э.П., Мэлс Дж.Л.* Теория оценивания и её применение в связи и управлении. –М.: Статистика, 1976. – 495 с.

26. *Семенов А.Д., Артамонов Д.В., Брюхачев А.В.* Идентификация объектов управления: Учеб. пособие. – Пенза: Пенз. ГУ, 2003. – 211 с.

27. *Сизиков В.С.* Математические методы обработки результатов измерений. –СПб.: Политехника, 2001. –346 с.

28. Современные методы идентификации систем /Под ред. *П. Эйкхоффа*. –М.: Мир, 1983. –400 с.

29. *Сыздыков Д.Ж., Юсупов Р.М.* Идентификация технических объектов. –Алматы, 1994. – 228 с.

30. *Сыздыков Д.Ж.* Методологические принципы построения систем идентификации применительно к сложным технологическим прцессам. –Алма-Ата: КазНИИНТИ, 1991. – 66 с.

31. *Сыздыков Д.Ж., Ширяева О.И.* Управление в условиях неполной определенности. – Шымкент, 2008. – 344 с.

32. *Сыздыков Д.Ж.* Идентификация в системах управления. – Алматы: Эверо, 2007. – 216 с.

33. *Солодовников В.В.* Статистическая динамика линейных систем автоматического управления. –М.: Физматгиз, 1960. – 656 с.

34. Статистические методы в экспериментальной физике /Пер. с англ. *А.И. Тяпкина*. –М.: Атомиздат, 1976. - 335 с. (Авт.: *Идье В., Драйард Д. и др.*).

35. *Хубер Дж.* П. Робастность в статистике. –М.: Мир, 1984. – 304 с.

36. *Цыпкин Я.З.* Адаптация и обучение в автоматических системах. –М.: Наука, 1968. – 400 с.

37. *Цыпкин Я.З.* Информационная теория идентификации. –М.: Наука, 1995.– 336 с.

38. *Эйкхофф П.* Основы идентификации систем управления. – М.: Мир, 1975. – 683 с.

39. *Юсупов Р.М.* Элементы теории идентификации технических объектов. –МО СССР, 1974. – 220 с.

# СОДЕРЖАНИЕ

ВВЕДЕНИЕ……...

ГЛАВА 1. ПРОБЛЕМА ИДЕНТИФИКАЦИИ……..…….
1.1. Операторные возмущающие воздействия……...…
1.2. Особенности решения задач идентификации……..……

ГЛАВА 2. МАТЕМАТИЧЕСКИЕ МОДЕЛИ ОБЪЕКТОВ ИДЕНТИФИКАЦИИ……..……
2.1. Общие сведения о математических моделях и их классификация…....
2.2. Формы записи дифференциальных уравнений ……..…
2.3. Передаточные функции и передаточные матрицы ……
2.4. Временные характеристики ……..……
2.5. Дискретные модели……
2.6 Нелинейные модели……

ГЛАВА 3. ОБЩИЕ СВЕДЕНИЯ И ПОНЯТИЯ О ЗАДАЧАХ ИДЕНТИФИКАЦИ……
3.1. Постановка задачи и особенности процесса идентификации……..……
3.2. Критерии и показатели качества идентификации …..……
3.3. Классификация методов идентификации……
3.4. Об условиях идентифицируемости объектов ……
3.5. Элементы теории вероятности и математической статистики……..…
3.6. Понятие вероятностной идентифицируемости……..…
3.7. Информация в статистике……
3.8. О структурной идентификации……..

ГЛАВА 4. НЕПАРАМЕТРИЧЕСКАЯ ИДЕНТИФИКАЦИЯ. …
4.1. Частотный анализ……
4.2. Анализ переходных процессов……
4.3. Определение передаточной функции по временным характеристикам объект……

4.4. Определение передаточной функции объекта по частотным характеристикам……
4.5. Статистическая идентификация линейных устойчивых стационарных объектов (метод корреляционных функций)……...

ГЛАВА 5. МЕТОДЫ ОЦЕНИВАНИЯ ПАРАМЕТРОВ……..…..
5.1. Общая схема оценивания параметров……..……
5.2. Байесово оценивание параметров ……
5.3. Оценивание по методу максимального правдоподобия……
5.4. Оценивание по методу наименьших квадратов ……

ГЛАВА 6. РЕКУРЕНТНЫЕ МЕТОДЫ ИДЕНТИФИКАИИ…....
6.1. Основные понятия. Алгоритмы градиентного метода…..…
6.2 Детерминированные алгоритмы……
6.3. Метод вспомогательных переменных……..…
6.4. Акселерация алгоритмов стохастической аппроксимации……
6.5. Сходимость итерационного процесса идентификации…....
6.6. Оценивание параметров и состоянии модели объекта…..…

ГЛАВА 7. МЕТОДЫ ИДЕНТИФИКАЦИИ НЕЛИНЕЙНЫХ СИСТЕМ УПРАВЛЕНИЯ……
7.1. Идентификация нелинейных объектов управления корреляционным методом……..…
7.2. Уравнения статической идентифической нелинейных систем управления……
7.3. Спектральный (параметрический) подход к решению задачи идентификации ……. ..……

Учебное издание

**Сыздыков Дастан Жаканович**

# ИДЕНТИФИКАЦИЯ СИСТЕМ АВТОМАТИЧЕСКОГО УПРАВЛЕНИЯ

Учебное пособие

Нач. РО *З.А. Губайдулина*
Редактор *Г.М. Дюсенбаева*
Компьютерная верстка *Л.П. Умирбекова*

Подписано в печать ……..2012 г.
Тираж 300 экз. Формат 60х84 1/16. Бумага типогр. №1.
Уч. изд. л. 17,4. Усл. п. л. 16,2. Заказ № … Цена договорная

Издание Казахского национального технического университета им. К.И. Сатпаева
Информационно-издательский центр КазНТУ
г. Алматы, Сатпаева, 22.

ISBN 978-601-228-378-5

