

ФЕДЕРАЛЬНОЕ АГЕНСТВО ПО ОБРАЗОВАНИЮ
Государственное образовательное учреждение высшего профессионального образования
**«Национальный исследовательский
Томский политехнический университет»**

В. И. Коновалов

ИДЕНТИФИКАЦИЯ И ДИАГНОСТИКА СИСТЕМ

*Рекомендовано Сибирским региональным учебно-методическим
центром высшего профессионального образования для межву-
зовского использования в качестве учебного пособия для сту-
дентов, обучающихся по специальности 220201.65 «Управление
и информатика в технических системах»*

Издательство
Томского политехнического университета
2010

УДК 681.51.015
ББК 32.965
К64

Коновалов В. И.

К64 Идентификация и диагностика систем: учебное пособие /
В.И. Коновалов; Томский политехнический университет. –
Томск: Изд-во Томского политехнического университета, 2010. –
163 с.

ISBN 978-5-98298-631-3

В учебном пособии рассмотрены различные формы математических моделей, методы идентификации объектов управления в статике и динамике. Изложены основы регрессионного анализа и планирования эксперимента. Отдельный раздел посвящен различным аспектам диагностики технических систем. Приведены примеры решения отдельных задач моделирования, идентификации и диагностики.

Учебное пособие подготовлено на кафедре автоматики и компьютерных систем ТПУ. Предназначено для студентов направления «Управление в технических системах»

УДК 681.51.015
ББК 32.965

Рецензенты

Доктор технических наук, профессор ТУСУР
А.М. Кориков

Доктор технических наук, директор НПО «Сибгеоинформатика»
А.И. Рюмкин

ISBN 978-5-98298-631-3

© ГОУ ВПО «Томский политехнический университет», 2010
© Коновалов В.И., 2010
© Оформление. Издательство Томского политехнического университета, 2010

ПРЕДИСЛОВИЕ

Курс «Идентификация и диагностика систем» является логическим продолжением курса «Теория автоматического управления». Теория моделирования и идентификации является частью современной теории управления. Подобный курс изучается во всех высших учебных заведениях, готовящих специалистов по управлению в технических системах. Об этом говорит список авторов приведенной в конце пособия учебной литературы: Д. Гроп [8] – профессор университета штата Колорадо, США; К. Спиди [23] – профессор университета Нового Южного Уэльса, Австралия; Дж. Мелса [22] – зав. кафедрой электротехники, США; Л. Льюнг [13] – профессор Массачусетского технологического института (США) и университета г. Линчепинга (Швеция). Достаточно много литературы отечественных авторов, однако вузовского учебника нет. Возможно, это объясняется тем, что программа курса весьма обширна и включает в себя разнообразные методы и направления. Как правило, каждый автор подробно излагает те вопросы, которые ему ближе, не касаясь многих других. На кафедре автоматики и телемеханики в 1981 году было выпущено учебное пособие [11] по близкой дисциплине «Моделирование и идентификация объектов управления», но многие разделы в нем изложены очень сжато, а вопросы диагностики вообще не затрагиваются.

Интересна история первой отечественной публикации по идентификации и технической диагностике. Книга профессора В.А. Тимофеева называлась «Теория и практика анализа результатов наблюдений над техническими объектами, работающими в условиях опытной эксплуатации» и была написана в специальной тюрьме в подмосковной деревне Марфино (так называемой «шарашке») в 1945 – 1949 гг. Сейчас это заведение известно благодаря роману А.И. Солженицына «В круге первом». Много позже книга В.А. Тимофеева была издана в ЛЭТИ (1960 г.) и переведена на английский язык [33].

Настоящее пособие предназначено в первую очередь студентам направления «Управление в технических системах».

1. Терминология

Рассмотрим основные термины, используемые в дальнейшем.

Модель – условный образ объекта исследования, получаемый для того, чтобы отобразить характеристики объекта, существенные для исследователя. Модели могут быть физическими (например, модель самолета для продувки в аэродинамической трубе) и математическими. В дальнейшем речь будет идти о математических моделях. По своему виду математические модели могут быть [25]:

- символные (в виде математических формул);
- графические;
- операционно-описательные (заданные, например, в виде алгоритмов);
- топологические или иконографические (задаются в виде графа или некоторой схемы).

Моделирование – метод исследования процессов или явлений на их моделях (математических или физических). Иногда под моделированием также подразумевают процесс построения модели.

Математическое моделирование – метод исследования процессов или явлений путем построения их математических моделей и исследования этих моделей с помощью вычислительной техники.

Имитационное моделирование – метод математического моделирования, при котором используют прямую подстановку чисел, имитирующих внешние воздействия (часто случайные), параметры и переменные процессов, в математические модели объектов.

Идентификация модели – определение параметров и структуры математической модели, обеспечивающей наилучшее совпадение выходных координат объекта и модели при одинаковых входных воздействиях. Это определение соответствует ГОСТ 20913-75. Приведём более простое определение: идентификация – процедура построения модели объекта по результатам измерения и обработки входных и выходных сигналов объекта. Подход к построению модели на основе идентификации называют также экспериментальным подходом, в отличие от аналитического, когда модель выводится на основании основных законов физики, химии, электротехники, материального или энергетического баланса.

«Черный ящик» – система, у которой при неизвестной внутренней организации, структуре и поведении элементов имеется возможность наблюдать реакцию выходных величин на изменение входных воздействий. Если структура объекта известна, то используют термин «серый ящик».

Параметрическая идентификация – определение параметров модели по заданной ее структуре.

Априорная модель – модель, построенная до начала специальных экспериментальных исследований.

Апостериорная модель – модель, полученная или уточненная по результатам экспериментальных исследований.

Техническая диагностика – прикладная наука, основными задачами которой являются распознавание состояния системы, определение причин нарушения ее работоспособности, установление вида и места дефекта, а также прогнозирование его изменения.

Диагностическая модель – математическая модель, с помощью которой выявляются закономерности и характер связей между фактическим состоянием диагностируемого объекта и некоторыми измеряемыми параметрами.

2. Классификация методов идентификации

В зависимости от принятого критерия классификации можно выделить и сгруппировать подходы и методы идентификации по-разному. Рассмотрим разные виды классификации [3, 11, 32].

1. Классификация по объему исходной информации об исследуемом объекте:

- методы идентификации в широком смысле, когда неизвестна структура объекта (используется также термин «непараметрическая идентификация»);
- методы параметрической идентификации, когда стоит задача оценки параметров модели известной структуры, используется также термин «идентификация в узком смысле».

2. Классификация по виду эксперимента:

- Методы активного эксперимента. Имеется возможность целенаправленно формировать входные воздействия для исследуемого объекта. Для получения статических моделей существует целое научное направление, которое так и называется – «Планирование эксперимента».
- Методы пассивного эксперимента. При этом исследователь может наблюдать и обрабатывать входные и выходные сигналы объекта, но не может вмешаться в его функционирование. Отметим, что пассив-

ный эксперимент возможен почти всегда, а вот активный эксперимент для многих исследуемых объектов и процессов провести нельзя.

3. Классификация в зависимости от вида критерия, по которому оценивается близость модели к реальному объекту. Обычно используется среднеквадратичное отклонение между выходом модели и объекта, но могут быть и другие подходы, и, соответственно, другие методы идентификации.

4. Классификация по оперативности получения модели:

- Методы ретроспективной (послеопытной) идентификации. В этом случае вначале проводится эксперимент, собираются и затем обрабатываются статистические данные, в итоге получаем модель объекта.

- Методы адаптивной идентификации, или идентификация в темпе со временем. Алгоритмы идентификации включаются в состав системы управления или диагностики. Модель объекта пересчитывается с появлением новых данных.

5. Классификация по типу исследуемого объекта или его модели. Вообще, класс модели должен соответствовать изучаемому объекту, но часто модель получают более простую, чем реальный объект. В теории автоматического управления, например, вместо нелинейной модели используют линеаризованную модель, описывающую поведение системы не во всем диапазоне ее работы, а только в окрестности рабочей точки. В то же время могут быть объекты, где учет нелинейности очень важен, и, естественно, модель такой системы должна быть нелинейной. Итак, исследуемые объекты могут быть:

- линейные и нелинейные;
- стационарные и нестационарные;
- одномерные и многомерные;
- с сосредоточенными и с распределенными параметрами;
- непрерывные и дискретные;
- статические и динамические;
- детерминированные и стохастические.

Стохастические объекты предполагают наличие внутри себя некоторой неопределенности, так что для оценки поведения таких объектов требуется применение вероятностных методов. В то же время, при обработке результатов измерений, почти всегда приходится иметь дело со случайными ошибками и погрешностями, но сам исследуемый объект остается при этом детерминированным.

6. Классификация по виду математической модели. В теории управления используются различные виды математического описания

одного и того же объекта: дифференциальные уравнения, передаточные функции, импульсные переходные функции, переходные функции, частотные характеристики. Поэтому можно классифицировать методы идентификации по тому, нахождение какого вида модели они нацелены.

7. Классификация по используемому математическому аппарату. В предисловии уже говорилось о большом разнообразии методов идентификации. Для построения математических моделей могут использоваться корреляционные методы, регрессионный анализ, частотные методы, теория оценивания, графоаналитические методы и многие другие разделы современной теории управления.

Вопросы и задания для самопроверки к разделу 2

1) Дайте определение идентификации. 2) Как можно получить модель, не прибегая к идентификации? 3) Когда целесообразно применять оперативные методы идентификации? 4) Преимущества и недостатки активного и пассивного экспериментов. 5) Пример динамической модели линейного стационарного многомерного объекта. 6) Пример динамической нелинейной одномерной модели. 7) Пример статической нелинейной одномерной модели. 8) Перечислите известные Вам виды математических моделей линейных динамических систем.

3. Выделение объекта моделирования в пространстве параметров

Одной из особенностей технологических процессов является большое количество технологических параметров, которыми характеризуется процесс. На начальном этапе моделирования необходимо выделить объект исследования в пространстве параметров. Полученная таким образом качественная иконографическая модель носит название параметрической [25]. При анализе любого технологического процесса как объекта управления принято выделять:

- выходные переменные, определяющие состояние технологического процесса или качество получаемой продукции (y);
- управляющие воздействия на объект (u);
- возмущающие воздействия, которые могут быть контролируемые (f) или не измеряемыми (v).

Математическая модель связывает выходные переменные и входные. Причем, в зависимости от цели использования модели, входными

могут быть только управляющие переменные, или и управляющие, и измеряемые возмущения, или, наконец, все входные воздействия. Если в модель входят не измеряемые возмущения, то такая модель может быть использована для изучения поведения объекта в разных условиях, однако эту модель нельзя использовать в системе управления – у нас просто не будет информации о существующих в данный момент возмущениях v .

Рассмотрим в качестве примера некоторые технологические процессы с целью получения качественных моделей управления.

Теплообменник [1]. Существует множество конструкций таких аппаратов. Они широко применяются во многих технологических процессах. Могут быть две цели функционирования теплообменника: нагреть холодный теплоноситель; охладить горячий теплоноситель. Рассмотрим простейший теплообменник с противотоком теплоносителей и передачей тепла через стенку. В процессе теплообмена холодный теплоноситель должен нагреться до определенной температуры.

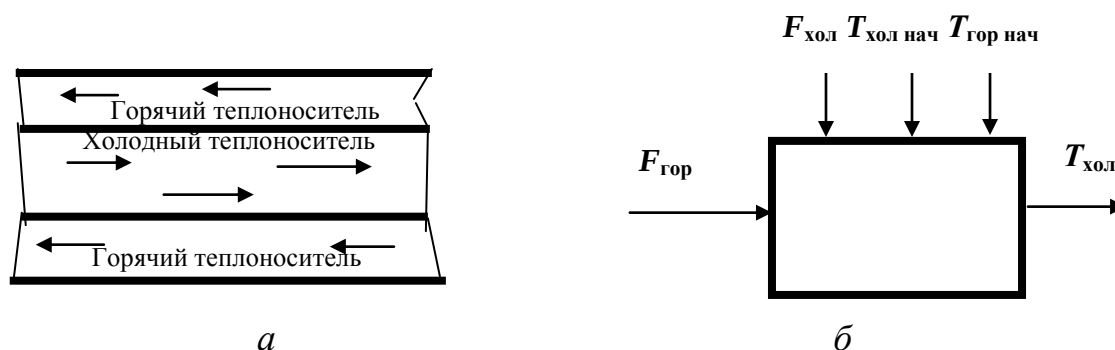


Рис. 1. Теплообменник (а) как объект управления (б)

Следовательно, выходной переменной объекта управления является температура холодного теплоносителя на выходе теплообменника $T_{\text{хол}}$, в качестве управления обычно используется расход горячего теплоносителя $F_{\text{гор}}$, а возмущающими воздействиями, которые можно контролировать, являются: $F_{\text{хол}}$ – расход холодного теплоносителя, $T_{\text{гор нач}}$ – начальная температура горячего теплоносителя на входе в теплообменник, $T_{\text{хол нач}}$ – начальная температура холодного теплоносителя на входе в теплообменник. При более детальном изучении этого процесса можно указать на дополнительные возмущающие воздействия, например на температуру окружающей среды. К числу неконтролируемых возмущений можно отнести изменение теплопроводности стенок теплообменника из-за налипания продуктов или накипи.

Химический реактор непрерывного действия. Это химико-технологический аппарат, в котором в результате реакции, например, двух различных веществ A и B , образуется вещество C , обладающее новыми свойствами. Упрощенная схема такого реактора приведена на рисунке 2, *а*. Допустим, что реакция идет с выделением тепла и для охлаждения реактор снабжен змеевиком. Компоненты A , B поступают в реактор непрерывно. Также непрерывно выводится продукт C и поступает вода для охлаждения.

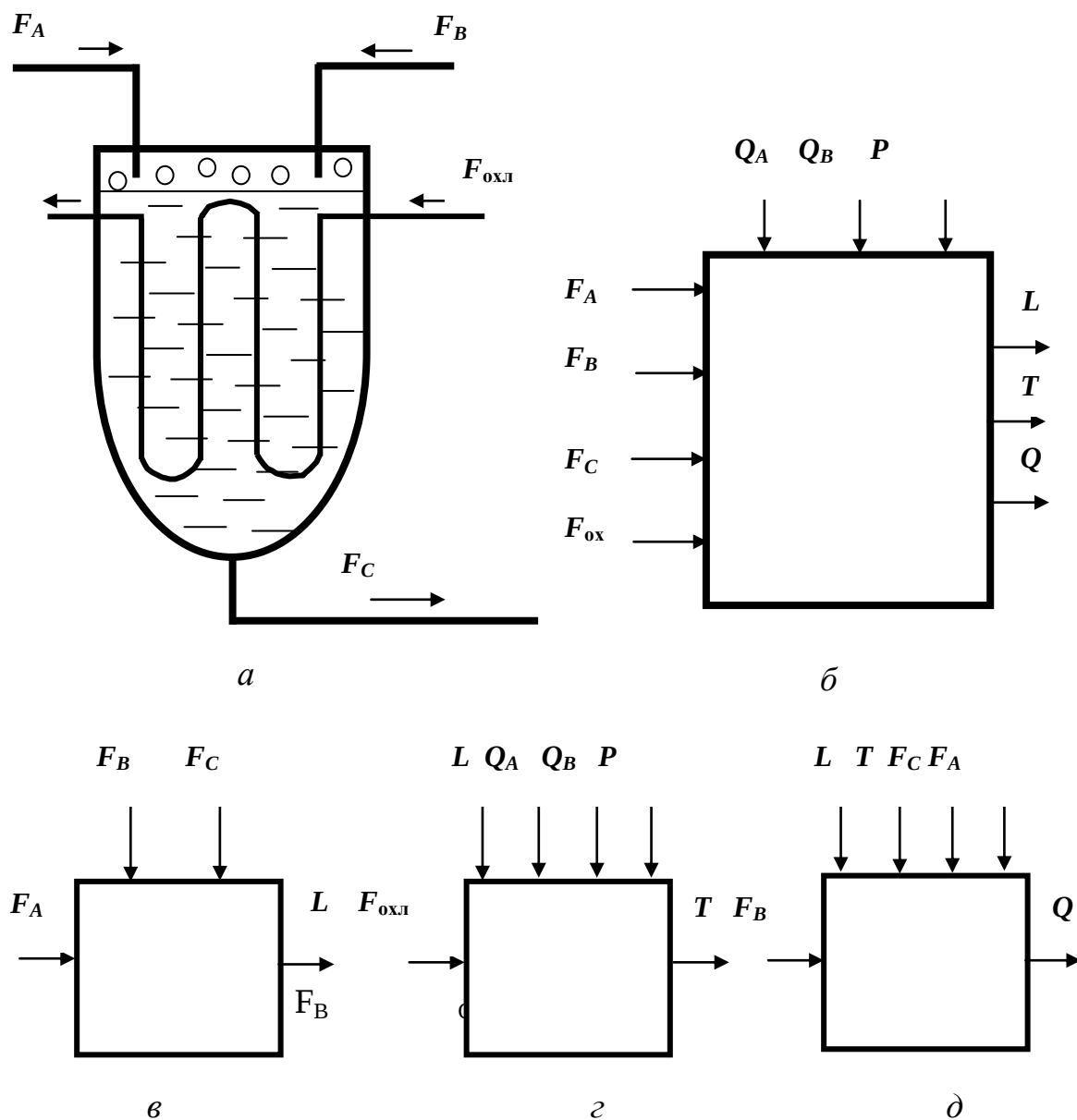


Рис. 2. Химический реактор как объект управления

Выходными параметрами рассматриваемого объекта управления будут: температура (T) в реакторе, уровень (L) в реакторе, состав выходного продукта Q , например, содержание вещества A в продукте C . Параметры T , L определяют состояние технологического процесса, а параметр Q – качество продукции. Для простоты, мы рассматриваем только один параметр Q , хотя для реальных процессов таких параметров может быть несколько. Возможными управляющими воздействиями на процесс могут быть расходы компонентов A , B в реактор и продукта C из реактора, а также расход охладителя. Не обязательно все они могут использоваться для управления. Обычно для управления одним параметром используется одно воздействие, если управляющих воздействий больше, система будет намного сложнее. В теории управления такие системы носят название переопределенных. И наоборот, может возникнуть ситуация, когда с помощью одного управляющего воздействия пытаются управлять двумя и более выходными параметрами. Такие системы неэффективны, в них возможны конфликтные ситуации и потеря управляемости.

Возможными возмущающими воздействиями может быть процентный состав потоков A , B (например, крепость кислоты или спирта). Для охлаждения существенным фактором является давление воды в сети P . Общая параметрическая модель, на которой показаны основные параметры процесса с разбивкой на управляющие, возмущающие и выходные, приведена на рис. 2, б.

Особенностью этого процесса является то, что в нем можно выделить отдельные локальные объекты управления малой размерности. Такое разбиение показано на рис. 2, в, г, д. Для задачи регулирования уровня абсолютно неважными факторами будет температура или процентный состав компонентов. Модель регулирования уровня связывает уровень с расходами компонентов A , B в реактор и расходом продукта C из реактора. Вспомните школьные задачи про бассейн, в который вливается вода и из которого выливается. Один из расходов будет управлением, больше и не требуется для регулирования одного выходного параметра, остальные являются возмущающими воздействиями. В нашем примере в качестве управляющего воздействия, от которого зависит уровень, выбран расход компонента A (F_A).

Изменение температуры в реакторе будет зависеть от того, насколько активно будет протекать реакция, т. е. от состава компонентов A , B , от массы реагирующих компонентов (а она напрямую связана с

уровнем L), от того, насколько эффективно идет процесс охлаждения (от параметров $F_{\text{охл}}$, P). Естественно, управляющим воздействием будет расход воды на охлаждение $F_{\text{охл}}$.

На рис. 2, δ представлена параметрическая модель для параметра Q , на качественном уровне выделены входные параметры – возмущающие (стрелки сверху) и управляющий – расход компонента F_B .

В результате анализа выделены на качественном уровне отдельные модели, в каждую из которых входит один выходной параметр, один управляющий, и несколько возмущающих. Видно, что один из возможных управляющих параметров – F_C в нашем случае учитывается как возмущающее воздействие.

Вопросы и задания для самопроверки к разделу 3

1) Как можно классифицировать параметры, характеризующие технологический процесс? 2) В емкость, снабженную мешалкой, непрерывно поступает холодная и горячая вода, из емкости также непрерывно ведется откачка теплой воды. Представьте качественную модель такого технологического процесса, выделив возможные локальные объекты управления.

4. Математические модели объектов идентификации

В случае если связь между входными и выходными сигналами системы чисто алгебраическая и объект не может обладать запасенной энергией, то такой объект назовем статическим. Объекты, описываемые дифференциальными, интегральными, интегродифференциальными уравнениями с запасом энергии, назовем динамическими. Примеры статических моделей (y – выходная переменная, x_j – входные переменные):

$y = 10 x$ – одномерная линейная модель;

$y = 10 x_1 + 15 x_2$ – многомерная линейная модель;

$y = 10 x_1 + 15 x_1 x_2 + 20 x_1^2$ – многомерная нелинейная модель.

Пусть задана динамическая модель в виде дифференциального уравнения, необходимо записать модель статики. Для этого приравняем все производные нулю и записываем статическое уравнение, связывающее вход – выход.

Дано: $y''' + 2 y'' + 4 y' + 10 y = 5 x + 17 x'$. Уравнение статики:
 $10 y = 5 x$, или $y = 0,5 x$.

В теории управления при решении разных задач зачастую используются различные формы динамических моделей: дифференциальные

уравнения, передаточные функции, импульсные переходные функции, переходные функции, частотные характеристики, уравнения состояния. Рассмотрим эти формы моделей и связь между ними [3, 11, 24].

1. Дифференциальное уравнение n -го порядка:

$$y^{(n)} + a_{n-1}y^{(n-1)} + \dots + a_0y = b_m x^{(m)} + \dots + b_1 x' + b_0 x.$$

В правой части уравнения стоит входная переменная и ее производные. Для физически реализуемых объектов порядок правой части не выше, чем левой. В правой части могут находиться несколько переменных, например управляющих и возмущающих. Тогда имеем многомерную модель. Если коэффициенты уравнения зависят от времени, модель – нестационарная. Для систем, обладающих транспортным запаздыванием (когда воздействие влияет на объект не сразу, а через время задержки τ), используют уравнения с запаздывающим аргументом:

$$y^{(n)} + a_{n-1}y^{(n-1)} + \dots + a_0y = b_1 x'(t - \tau) + b_0 x(t - \tau).$$

Решить дифференциальное уравнение n -го порядка – значит найти, как меняется выходной сигнал и его производные до $(n - 1)$ -го порядка включительно во времени при определенных начальных условиях под воздействием $x(t)$.

2. Система дифференциальных уравнений первого порядка. Если необходимо решить дифференциальное уравнение с помощью какого-либо программного пакета, возникает необходимость представить модель в виде системы уравнений первого порядка, т. к. все они работают именно с такой формой моделей. Переход осуществляется введением новых переменных: выходной сигнал y обозначим как переменную y_1 , его первую производную – как переменную y_2 и так далее. Число уравнений в системе равно порядку исходного дифференциального уравнения.

Пример. Задано дифференциальное уравнение 3-го порядка:

$$y''' + 7y'' + 4y = 21x.$$

Найдем модель в виде системы дифференциальных уравнений первого порядка. Обозначим $y = y_1$, и будем записывать уравнения системы, вводя новые переменные вместо производных:

$$y'_1 = y_2;$$

$$y'_2 = y_3;$$

$$y'_3 = -4y_1 - 7y_3 + 21x.$$

Обратите внимание, что в третьем уравнении системы для этого примера отсутствует переменная y_2 , т. к. первой производной в заданном дифференциальном уравнении не было.

Если бы в исходном дифференциальном уравнении были производные входного сигнала $x(t)$, все они присутствовали бы в правой части последнего уравнения системы. Рассмотрим алгоритм перехода к системе уравнений первого порядка, не содержащей производных входного сигнала. Порядок действий следующий.

- Записываем систему уравнений первого порядка, но в каждое из уравнений добавляем входную переменную $x(t)$, умноженную на неизвестный коэффициент.
- Сворачиваем полученную систему снова в одно уравнение, избавляясь от новых переменных y_2, y_3 , и т. д.
- Из условия равенства правых частей исходного и только что полученного уравнений находим неизвестные коэффициенты.

После выполнения этих пунктов будем иметь систему дифференциальных уравнений первого порядка, эквивалентную исходному дифференциальному уравнению, причем в каждом из уравнений системы будет присутствовать входная переменная без ее производных.

Рассмотрим пример. Дано уравнение: $y'' + 7y' + 4y = 2x + 5x'$. Требуется записать модель в виде системы уравнений первого порядка без x' . Записываем систему 2 уравнений первого порядка в соответствии с первым шагом алгоритма:

$$y'_1 = y_2 + q_1 x;$$

$$y'_2 = -4y_1 - 7y_2 + q_2 x.$$

Здесь через q_1 и q_2 обозначены неизвестные пока коэффициенты. Теперь выразим y_2 из первого уравнения системы и подставим во второе:

$$y_2 = y'_1 - q_1 x;$$

$$y''_1 - q_1 x' = -4y_1 - 7(y'_1 - q_1 x) + q_2 x.$$

Сгруппируем в левой части уравнения выходные переменные, в правой – входные:

$$y''_1 + 7y'_1 + 4y_1 = q_1 x' + (q_2 + 7q_1)x.$$

Из условия равенства коэффициентов при x и x' в исходном и только что полученном уравнении получаем систему двух уравнений с двумя неизвестными:

$$q_1 = 5;$$

$$q_2 + 7q_1 = 2;$$

$$\text{Отсюда } q_1 = 5, \quad q_2 = -33.$$

Окончательный ответ:

$$y'_1 = y_2 + 5x;$$

$$y'_2 = -4y_1 - 7y_2 - 33x.$$

3. Модель в пространстве состояний. Понятие состояния введено американским ученым в области теории управления Р. Калманом в 60-х годах XX века. Переменные состояния – внутренние переменные системы, которые и определяют ее состояние в текущий момент времени. Если система описывается дифференциальным уравнением n -го порядка, то такими переменными могут быть сам выходной сигнал и его производные до $(n - 1)$ -й включительно. Векторно-матричное уравнение, которое связывает вектор производных состояния, вектор переменных состояния и векторы внешних воздействий (управления, возмущения), называется уравнением состояния. Концепция этого подхода предполагает, что возможно не все переменные состояния можно измерить и поэтому уравнение состояния дополняется уравнением наблюдения, которое связывает вектор переменных наблюдения и вектор переменных состояния системы. Таким образом, модель системы в пространстве состояний имеет следующий вид:

$$\dot{X} = AX + BU;$$

$$Y = HX,$$

где X – вектор состояния; $\dot{X}$ – вектор производных переменных состояния; A – матрица состояния, квадратная, размерностью $n \times n$; B – матрица управления; U – вектор управления; Y – вектор наблюдения; H – матрица наблюдения.

Если рассматриваем одномерную систему, то вектор управления превращается в скаляр, а матрица управления становится вектором-столбцом. Если наблюдается только сам выходной сигнал y , то матрица наблюдения вырождается в вектор-строку. Если все переменные состояния доступны для измерения, то матрица наблюдения H будет единичной матрицей, т. е. уравнение наблюдения фактически не нужно.

Пример. Дано дифференциальное уравнение:

$$y''' + 7y'' + 5y' + 3y = 2u_1 + 4u_2.$$

Запишем модель в форме уравнения состояния. Предположим, что переменными состояния будут сам сигнал y и его производные и что они доступны измерению, т. е. вектор состояния и вектор наблюдения совпадают: $X = Y$. Вначале перейдем к системе дифференциальных уравнений первого порядка.

$$x'_1 = x_2;$$

$$x'_2 = x_3;$$

$$x'_3 = -3x_1 - 5x_2 - 7x_3 + 2u_1 + 4u_2.$$

Теперь эту систему необходимо записать в векторно-матричном виде. При умножении матрицы на вектор каждая строка матрицы умножается на вектор по правилу: результат равен сумме произведений каждого элемента строки на соответствующий элемент столбца.

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -3 & -5 & -7 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 2 & 4 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}.$$

Если перемножить в правой части матрицы на векторы, то получим исходную систему трех уравнений первого порядка:

$$\dot{x}_1 = 0x_1 + 1x_2 + 0x_3 + 0u_1 + 0u_2;$$

$$\dot{x}_2 = 0x_1 + 0x_2 + 1x_3 + 0u_1 + 0u_2;$$

$$\dot{x}_3 = -3x_1 - 5x_2 - 7x_3 + 2u_1 + 4u_2.$$

4. Передаточная функция. В отечественной литературе по теории управления используется очень часто. Передаточная функция $W(s)$ есть отношение изображения выходного сигнала системы $Y(s)$ к изображению входного сигнала $X(s)$ при нулевых начальных условиях:

$$W(s) = \frac{Y(s)}{X(s)}.$$

Передаточная функция легко может быть получена из дифференциального уравнения. Дано уравнение:

$$y''' + 7y'' + 5y' + 3y = 2x + 4x'.$$

Преобразуем по Лапласу правую и левую части уравнения, полагая нулевыми начальные условия:

$$s^3Y(s) + 7s^2Y(s) + 5sY(s) + 3Y(s) = 2X(s) + 4sX(s) \text{ или } (s^3 + 7s^2 + 5s + 3)Y(s) = (2 + 4s)X(s).$$

Отсюда

$$Y(s) = \frac{2 + 4s}{s^3 + 7s^2 + 5s + 3} X(s) \text{ или } W(s) = \frac{2 + 4s}{s^3 + 7s^2 + 5s + 3}.$$

Для физически реализуемых систем порядок полинома в числителе передаточной функции всегда меньше или равен порядку полинома в знаменателе.

Приведем пример обратного перехода – от передаточной функции к дифференциальному уравнению. Дано:

$$W(s) = \frac{2}{s^2 + 5s + 1}.$$

Перепишем в виде $\frac{Y(s)}{X(s)} = \frac{2}{s^2 + 5s + 1}$ или $(s^2 + 5s + 1) Y(s) = 2X(s)$.

Применим к правой и левой частям этого уравнения обратное преобразование Лапласа и перейдем от изображений к функциям времени:

$$X(s) \rightarrow x(t), \quad Y(s) \rightarrow y(t), \quad sY(s) \rightarrow \frac{dy(t)}{dt}, \quad s^2Y(s) \rightarrow \frac{d^2y(t)}{dt^2}.$$

В итоге получим дифференциальное уравнение

$$y'' + 5y' + y = 2x.$$

Найдем связь между передаточной функцией и моделью объекта в пространстве состояний. Итак, дана модель в пространстве состояний:

$$\dot{X} = AX + BU;$$

$$Y = HX.$$

Преобразуем эти уравнения по Лапласу:

$$sX(s) = AX(s) + BU(s);$$

$$Y(s) = HX(s).$$

Все переменные, которые были функциями времени, стали их изображениями по Лапласу, матрицы же коэффициентов остались неизменными. Выразим $X(s)$ из первого уравнения системы и подставим во второе:

$$sX(s) - AX(s) = BU(s);$$

$$[sE - A] X(s) = BU(s);$$

$$X(s) = [sE - A]^{-1} BU(s);$$

$$Y(s) = H[sE - A]^{-1} BU(s).$$

Здесь E – единичная матрица, эквивалент единицы в матричной алгебре.

Из последней формулы следует, что $W(s) = H[sE - A]^{-1} B$.

Пример. Дана модель в пространстве состояний. Найти модель в форме передаточной функции.

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -2 & -4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} 1 \\ -2 \end{bmatrix} u;$$

$$y = \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix};$$

$$sE - A = s \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \begin{bmatrix} 0 & 1 \\ -2 & -4 \end{bmatrix} = \begin{bmatrix} s & -1 \\ 2 & s+4 \end{bmatrix}.$$

Найдем обратную матрицу. Алгоритм нахождения обратной матрицы следующий.

1. Вместо каждого из элементов ставится его алгебраическое дополнение. Алгебраическое дополнение, в свою очередь, есть определитель матрицы, полученной при вычеркивании столбца и строки, на пересечении которых стоит этот элемент. Кроме того, определитель необходимо умножить на $(-1)^{i+j}$, i, j – номер строки и столбца, соответственно.

2. Полученную матрицу необходимо транспонировать.

3. Каждый элемент полученной таким образом матрицы поделить на определитель исходной матрицы.

После выполнения первого пункта алгоритма получим:

$$\begin{bmatrix} s+4 & -2 \\ +1 & s \end{bmatrix}.$$

После выполнения второго пункта получим:

$$\begin{bmatrix} s+4 & +1 \\ -2 & s \end{bmatrix}.$$

Определитель исходной матрицы $\Delta = s(s+4) + 2 = s^2 + 4s + 2$.

В итоге получим

$$W(s) = \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} s+4 & 1 \\ -2 & s \end{bmatrix} \begin{bmatrix} 1 \\ -2 \end{bmatrix} \frac{1}{s^2 + 4s + 4} = \frac{s+2}{s^2 + 4s + 4}.$$

5. Импульсная переходная (весовая) функция $g(t)$. Импульсная переходная функция (ИПФ) – реакция линейного стационарного динамического объекта на воздействие в виде дельта-функции. Зная ИПФ и входное воздействие $x(t)$, можно вычислять реакцию динамической системы $y(t)$ по известному интегральному уравнению свертки:

$$y(t) = \int_0^t g(t-\tau)x(\tau)d\tau.$$

Передаточная функция $W(s)$ является изображением импульсной переходной функции:

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

Зная аналитическое выражение для ИПФ, можно решить этот интеграл и найти $W(s)$. Однако проще воспользоваться правилами операционного исчисления, зная соответствие между функциями времени и их изображениями по Лапласу.

Пример. ИПФ $g(t) = e^{-2t} + e^{-t}$. Функции времени e^{-at} соответствует изображение $1/(a+s)$. Поэтому

$$W(s) = \frac{1}{2+s} + \frac{1}{1+s} = \frac{2s+3}{s^2+3s+1}.$$

Для решения обратной задачи – определения ИПФ по передаточной функции в общем случае необходимо пользоваться теорией вычетов. Для случая, когда полюса передаточной функции простые, не равны нулю и не кратные, используется формула Хэвисайда:

$$W(s) = \frac{N(s)}{M(s)} \div \sum_{i=1}^k \frac{N(s_i)}{M'(s_i)} e^{s_i t},$$

где s_i – корни полинома $M(s)$ или полюса передаточной функции; k – число полюсов.

Для нашего примера $s_1 = -2$, $s_2 = -1$. В итоге получаем:

$$g(t) = \frac{2(-2)+3}{2(-2)+3} e^{-2t} + \frac{2(-1)+3}{2(-1)+3} e^{-t} = e^{-2t} + e^{-t}.$$

6. Переходная функция $h(t)$. Она представляет собой реакцию объекта на единичное ступенчатое воздействие. Переходная функция $h(t)$ и ИПФ $g(t)$ связаны соотношениями

$$g(t) = \frac{dh(t)}{dt} \quad \text{или} \quad h(t) = \int_0^t g(v) dv.$$

Так как изображение единичной ступенчатой функции есть $1/s$, то изображение переходной функции будет

$$H(s) = W(s) \frac{1}{s} = \frac{N(s)}{M(s)} \frac{1}{s}.$$

Для этого случая – один нулевой корень в знаменателе, остальные корни простые, не кратные, ненулевые, формула Хэвисайда будет иметь вид

$$H(s) = \frac{N(s)}{M(s)} \frac{1}{s} \div \frac{N(0)}{M(0)} + \sum_{i=1}^k \frac{N(s_i)}{s_i M'(s_i)} e^{s_i t} = h(t).$$

7. Частотные характеристики. Частотные характеристики показывают, как изменится амплитуда сигнала на выходе системы и сдвиг сигнала по фазе для входного сигнала определенной частоты. Если интересно только изменение амплитуды, используют амплитудную частотную характеристику (АЧХ), при изучении фазовых сдвигов на разных частотах применяют фазовую частотную характеристику (ФЧХ). Реакцию системы в комплексе на гармонические воздействия разной частоты

ты можно определить с помощью амплитудно-фазовой частотной характеристики (АФЧХ). Совокупность всех точек, принадлежащих АФЧХ, носит название годографа частотной характеристики. Для построения АФЧХ обычно выделяют вещественную часть этой характеристики – вещественную частотную характеристику (ВЧХ) и мнимую ее часть (МЧХ). АФЧХ получают из передаточной функции заменой комплексной переменной s на $j\omega$. Далее следует обязательно избавиться от мнимых чисел в знаменателе, выделив таким образом ВЧХ и МЧХ. Формулы, связывающие частотные характеристики между собой:

$$W(j\omega) = P(\omega) + jQ(\omega);$$

$$W(j\omega) = A(\omega)e^{j\varphi(\omega)};$$

$$A^2(\omega) = P^2(\omega) + Q^2(\omega);$$

$$\varphi(\omega) = \arctg \frac{Q(\omega)}{P(\omega)}.$$

Здесь $W(j\omega)$ – АФЧХ; $A(\omega)$ – АЧХ; $P(\omega)$ – ВЧХ; $Q(\omega)$ – МЧХ; $\varphi(\omega)$ – ФЧХ.

8. Дискретные модели. При моделировании систем и объектов управления часто используют дискретные модели. На это есть, по мнению Я.З. Цыпкина, следующие причины. Во-первых, непрерывные задачи можно аппроксимировать дискретными с любой степенью точности. Во-вторых, развитие цифровой техники для управления привело к усилению внимания к дискретным моделям. И наконец, математическое исследование дискретных объектов и алгоритмов идентификации во многих случаях проще, чем непрерывных. Наиболее удобны при моделировании на компьютере модели в виде разностных уравнений. Рассмотрим описание непрерывного объекта с помощью дискретной модели.

Рассмотрим пример. Дана модель в виде дифференциального уравнения: $Ty'(t) + y(t) = kx(t)$. Найдем модель этого объекта в виде разностного уравнения. Для этого нужно заменить производную с помощью формулы конечной разности и вместо непрерывных значений переменных взять их значения в дискретные моменты времени. Формула передней конечной разности имеет вид

$$y'(t) = \frac{y[n+1] - y[n]}{\Delta T},$$

где ΔT – интервал дискретизации.

Получим разностное уравнение:

$$T \frac{y[n+1] - y[n]}{\Delta T} + y[n] = k \cdot x[n];$$

$$y[n+1] = \left(1 - \frac{\Delta T}{T}\right) y[n] + \frac{\Delta T}{T} k \cdot x[n] = \alpha \cdot y[n] + \beta \cdot k \cdot x[n].$$

Для устойчивой модели обязательно выполнение условия $\alpha + \beta = 1$, кроме того, необходимо правильно выбрать интервал дискретизации:

$$\Delta T \ll T.$$

Можно использовать для замены производной другую формулу конечной разности:

$$y'(t) = \frac{y[n] - y[n-1]}{\Delta T}.$$

В этом случае получим разностное уравнение несколько иной формы:

$$T \frac{y[n] - y[n-1]}{\Delta T} + y[n] = k \cdot x[n];$$

$$y[n] = \frac{T}{T + \Delta T} y[n-1] + \frac{\Delta T}{T + \Delta T} k \cdot x[n] = \alpha_1 \cdot y[n-1] + \beta_1 \cdot k \cdot x[n].$$

Так как «истинным» объектом моделирования был объект непрерывный, то две полученных разностных модели являются как бы приближением к истине, но с разных сторон. Вот числовой пример. Пусть в исходном дифференциальном уравнении $T = 5$, $k = 1$. При $\Delta T = 1$ в разных формулах конечной разности получим такие разностные уравнения:

$$y[n+1] = 0,8y[n] + 0,2k \cdot x[n];$$

$$y[n] = 0,833y[n-1] + 0,167k \cdot x[n].$$

Как видно, есть значимые отличия в коэффициентах уравнений. Возьмем меньшее значение интервала дискретизации $\Delta T = 0,01$. Теперь уравнения будут такими:

$$y[n+1] = 0,998y[n] + 0,002k \cdot x[n];$$

$$y[n] = 0,998004y[n-1] + 0,001996k \cdot x[n].$$

Различие между коэффициентами стало почти незначимым.

Аналогично можно получить для дифференциального уравнения второго порядка модель в виде разностного уравнения. При этом вторую производную необходимо заменить дискретной формулой

$$y'' = \frac{\frac{y[n+2] - y[n+1]}{\Delta T} - \frac{y[n+1] - y[n]}{\Delta T}}{\Delta T} = \frac{y[n+2] - 2y[n+1] + y[n]}{\Delta T^2}.$$

В итоге получим разностное уравнение второго порядка вида $y[n+2] = \alpha \cdot y[n+1] + \beta \cdot y[n] + \gamma \cdot k \cdot x[n]$.

Для устойчивого объекта необходимо равенство единице суммы коэффициентов α , β , γ . В принципе, также можно получить дискретные модели и для систем высокого порядка, но это потребует очень громоздких выкладок. Гораздо проще получить дискретную модель состояния, ведь в этом случае мы будем иметь дело только с производными первого порядка.

Рассмотрим пример. Имеем модель в виде системы дифференциальных уравнений первого порядка:

$$y_1' = y_2;$$

$$y_2' = y_3;$$

$$y_3' = -a_0 y_1 - a_1 y_2 - a_2 y_3 + b x.$$

Получим дискретную модель:

$$\frac{y_1[n+1] - y_1[n]}{\Delta T} = y_2[n];$$

$$\frac{y_2[n+1] - y_2[n]}{\Delta T} = y_3[n];$$

$$\frac{y_3[n+1] - y_3[n]}{\Delta T} = -a_0 y_1[n] - a_1 y_2[n] + (1 - a_2) y_3[n] + b x[n].$$

Окончательный вид модели в векторно-матричной форме:

$$\begin{bmatrix} y_1[n+1] \\ y_2[n+1] \\ y_3[n+1] \end{bmatrix} = \begin{bmatrix} 1 & \Delta T & 0 \\ 0 & 1 & \Delta T \\ -\Delta T a_0 - \Delta T a_1 & 1 - \Delta T a_2 & 1 \end{bmatrix} \begin{bmatrix} y_1[n] \\ y_2[n] \\ y_3[n] \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ \Delta T b \end{bmatrix} x[n].$$

Вопросы и задания для самопроверки к разделу 4

1) Как получить модель в форме дифференциального уравнения, если известна модель в форме передаточной функции? 2) Как получить передаточную функцию, если известна импульсная переходная функция системы? 3) Известна передаточная функция системы $W(s) = \frac{2 + 4s}{s^3 + 7s^2 + 5s + 3}$. Чему равно установившееся значение y переходной функции? 4) Чему равен скачок в начальный момент времени y импульсной реакции? 5) Как записать статическую модель, описывающую установившийся режим САУ, если известна динамическая модель системы в форме передаточной функции?

5. Модели динамических нелинейных систем

При идентификации нелинейных динамических объектов используется два основных подхода. Первый основан на получении математического описания линеаризованной модели, адекватной объекту в определенной области. Этот подход базируется на применении методов линеаризации в малых приращениях, гармонической и статистической линеаризации. Обычно эти методы рассматриваются при изучении теории автоматического управления.

Во втором подходе математическая модель исследуемого объекта рассматривается как существенно нелинейная и идентификация заключается в отыскании характеристик его линейных и нелинейных элементов. При этом структура объекта должна быть в какой-то мере известна [3]. Рассмотрим две возможные структуры нелинейного динамического объекта. Первая изображена на рис. 3.

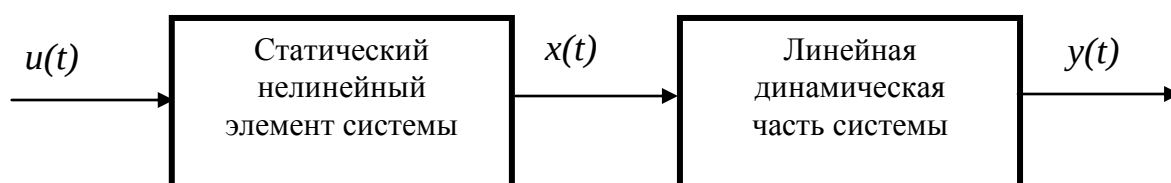


Рис. 3. Структура объекта с моделью Хаммерштейна

Если нелинейность в таком объекте можно описать степенным рядом, то модель нелинейной системы носит название модели Хаммерштейна. Итак, промежуточный сигнал $x(t)$ равен:

$$x(t) = b_0 + b_1 u(t) + b_2 u^2(t) + b_3 u^3(t) + \dots$$

Линейная часть системы описывается известным интегральным уравнением

$$y(t) = \int_0^t g(t - \tau) x(\tau) d\tau.$$

Общая модель, связывающая вход–выход динамической нелинейной системы, имеет вид

$$y(t) = \int_0^t g(t - \tau) [b_0 + b_1 u(\tau) + b_2 u^2(\tau) + \dots] d\tau.$$

При идентификации необходимо будет найти коэффициенты аппроксимации нелинейности степенным рядом и определить импульсную переходную функцию линейной части $g(t)$.

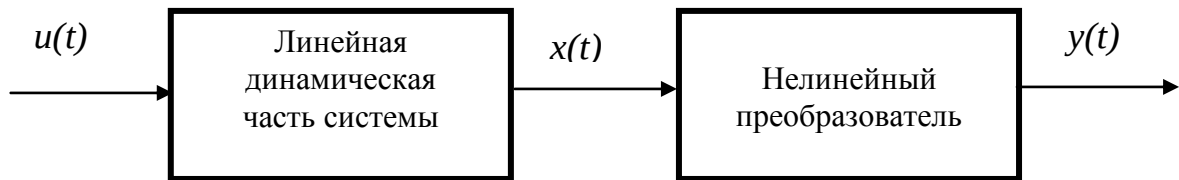


Рис. 4. Структура нелинейного объекта с моделью в форме функционального ряда Вольтерра

На рис. 4 изображен объект с другой структурой. Вначале входной сигнал проходит через динамическое линейное звено, а потом подвергается нелинейному преобразованию $F[x(t)]$:

$$y(t) = F\left\{\int_0^t g(t - \tau) u(\tau) d\tau\right\}.$$

Поведение такого объекта можно описать с помощью бесконечного функционального ряда Вольтерра (ударение на последнем слоге):

$$\begin{aligned} y(t) = & \int_0^t g(t - \tau) u(\tau) d\tau + \int_0^t \int_0^t g(t - \tau_1, t - \tau_2) u(\tau_1) u(\tau_2) d\tau_1 d\tau_2 + \\ & + \int_0^t \int_0^t \int_0^t g(t - \tau_1, t - \tau_2, t - \tau_3) u(\tau_1) u(\tau_2) u(\tau_3) d\tau_1 d\tau_2 d\tau_3 + \dots \end{aligned}$$

Функции, стоящие под интегралами, носят название ядер разложения. Определение ядер и будет в данном случае целью идентификации.

При решении конкретных задач ограничиваются двумя – тремя членами ряда Вольтерра. В зависимости от свойств нелинейного преобразования ядра могут быть сепарабельными, когда ядро можно представить в виде произведения отдельных составляющих $g(t_1, t_2) = g(t_1)g(t_2)$, или несепарабельными, когда такое разбиение сделать невозможно. Например, если у объекта, изображенного на рис. 4, нелинейным преобразователем будет квадратор, то общая динамическая модель системы может быть ограничена двумя членами разложения Вольтерра с сепарабельными ядрами.

Вопросы и задания для самопроверки к разделу 5

- 1) Чем по структуре отличается модель Хаммерштейна от модели Вольтерра?
- 2) Если известно, что после линейного динамического звена сигнал возводится в квадрат, какую модель следует применить для описания такого объекта?
- 3) Чем отличаются сепарабельные и несепарабельные ядра в разложении Вольтерра?

6. Модели динамических нестационарных систем

Существует много примеров, когда технические объекты с течением времени изменяют свои свойства. Летящий самолет или ракета, у которых выгорает топливо, меняется масса, а следовательно и динамические свойства. Химический процесс, активность которого существенно зависит от свойств катализатора, а эти свойства зависят от времени. Динамику таких объектов можно описать дифференциальными уравнениями, у которых будут переменные коэффициенты, т. е. коэффициенты, зависящие от времени. Если описывать объект моделями в частотной области, то у частотных характеристик также появится зависимость от времени. Модель нестационарной системы в форме импульсной переходной функции $g(\theta, \tau)$ будет выглядеть как функция времени подачи входного сигнала τ и времени наблюдения за выходом системы θ . Для стационарного объекта импульсная реакция зависит от их разности $(\theta - \tau)$.

При моделировании нестационарных систем с использованием традиционных форм математических моделей часто пользуются методом замороженных коэффициентов. Например, решают дифференциальное уравнение для определенного временного сечения, считая параметры уравнения постоянными. Затем переходят к другому сечению времени, соответственно изменяют коэффициенты модели и снова ре-

шают дифференциальное уравнение. Путь достаточно длительный и громоздкий.

Рассмотрим специальную форму математической модели, придуманную для анализа и синтеза нестационарных динамических систем. Она основана на спектральной теории нестационарных систем, разработанной в 70-х годах на кафедре П-1 МВТУ им. Баумана В.В. Солодовниковым и В.В. Семеновым [26].

Дана функция времени $f(t)$ на интервале $[0, T]$. Представим ее в виде разложения в обобщенный ряд Фурье:

$$f(t) = \sum_{i=0}^{\infty} F_i \Psi_i(t),$$

где F_i – коэффициенты разложения в обобщенный ряд Фурье;

$\Psi_i(t)$ – базисные функции.

В отличие от классического ряда Фурье, при разложении в обобщенный ряд в качестве базисных функций $\Psi_i(t)$ используются не только тригонометрические функции, но и полиномы Лежандра, Чебышёва. Основное требование к базисным функциям – ортогональность на интервале, где определена функция времени. То есть прежде чем пользоваться стандартными ортогональными полиномами, например полиномами Лежандра, надо преобразовать их, чтобы они стали ортогональными на интервале $[0, T]$. Таким образом, для нашего случая базисные функции должны соответствовать условию

$$\int_0^T \Psi_i(t) \Psi_j(t) dt = 0, \text{ при } i \neq j.$$

Коэффициенты разложения F_i можно определить по известной формуле для расчета коэффициентов ряда Фурье:

$$F_i = \int_0^T \Psi_i(t) f(t) dt.$$

Для этого, естественно, нужно знать аналитическое описание функции $f(t)$. Если известны значения функции в точках, то коэффициенты разложения можно вычислить методом наименьших квадратов, решая обычную задачу аппроксимации. На практике в формуле для разложения в обобщенный ряд Фурье (ОРФ) ограничиваются конечным числом членов разложения. Всегда можно задаться допустимой погрешностью аппроксимации и ограничиться, например, 7 – 10 коэффициентами разложения.

Таким образом, если четко знать временной интервал T , на котором задана функция времени, выбрать ортогональный базис и определить

число коэффициентов разложения, то вся информация о функции времени будет сосредоточена в коэффициентах разложения. Записав коэффициенты в виде вектора F , получим характеристику сигнала $f(t)$, названную в спектральной теории нестационарной спектральной характеристикой (НСХ) сигнала.

$$F(T) = \begin{bmatrix} F_0(T) \\ F_1(T) \\ \dots \\ F_n(T) \end{bmatrix}.$$

Обратим теперь внимание на импульсную переходную функцию нестационарной системы $g(\theta, \tau)$. Разложим эту функцию в ОРФ на интервале $[0, T]$ по времени τ . Получим вектор коэффициентов, которые являются функцией времени θ . Рассматривая каждый из таких коэффициентов как функцию от θ , разложим их в ОРФ на том же интервале времени $[0, T]$. В итоге получим матрицу коэффициентов, которые уже не зависят от времени θ и τ , а зависят от системы ортогональных полиномов и времени T , на котором исследуется поведение системы. Полученная таким образом модель называется двумерной нестационарной передаточной функцией (ДНПФ) $W(T, T)$. Передаточной эта функция называется потому, что она устанавливает связь между нестационарными спектральными характеристиками входного $X(T)$ и выходного сигнала системы $Y(T)$:

$$Y(T) = W(T, T)X(T).$$

В классической теории управления многие задачи анализа и синтеза линейных стационарных систем решаются с использованием моделей в форме передаточных функций. При этом, благодаря операционному исчислению, операции дифференцирования, интегрирования заменяются простой алгеброй. И для нестационарных систем использование моделей в форме ДНПФ также упрощает решение многих задач, правда здесь используется уже матричная алгебра. В книгах по спектральному методу приведены ДНПФ типовых звеньев относительно разных базисных функций – дифференцирующего, интегрирующего и усилительного с переменным коэффициентом усиления $a(\theta) = \theta$. ДНПФ обычного усилителя с постоянным коэффициентом усиления записывается в виде диагональной матрицы со значением коэффициента усиления на главной диагонали. Таким образом, если задана модель динамической нестационарной системы в форме дифференциального уравнения, то можно легко записать модель этой же системы в спектральной области.

Пример. Дана модель системы:

$$a(\theta)y'(\theta) + y(\theta) = bx(\theta).$$

Это апериодическое звено первого порядка, но с переменным коэффициентом усиления $a(\theta)$. Запишем его в спектральной области:

$$A(T,T)P(T,T)Y(T) + Y(T) = BX(T),$$

где $A(T,T)$ – ДНПФ звена с переменным коэффициентом усиления;

$P(T,T)$ – ДНПФ дифференцирующего звена;

B – ДНПФ усилительного звена с коэффициентом, равным b (диагональная матрица);

$Y(T)$, $X(T)$ – нестационарные спектральные характеристики выходного и входного сигналов системы на интервале времени $[0, T]$.

Выразив из этого уравнения $Y(T)$, можно записать его решение в спектральной области (при нулевых начальных условиях):

$$Y(T) = [A(T,T)P(T,T) + E]^{-1} BX(T).$$

В приведенной выше формуле «запрятана» ДНПФ исследуемого объекта: $W(T,T) = [A(T,T)P(T,T) + E]^{-1} B$.

Таким образом, получили ДНПФ не только из импульсной переходной функции, но и из дифференциального уравнения.

Вопросы и задания для самопроверки к разделу 6

1) Будут ли ортогональны на интервале $[-1, +1]$ функции $f_1(t) = 1$ и $f_2 = t$? 2) Будут ли ортогональны эти же функции на интервале $[0, +1]$? 3) В чем отличие обобщенного ряда Фурье от классического ряда Фурье? 4) Каким образом из импульсной переходной функции нестационарного объекта можно получить двумерную нестационарную передаточную функцию? 5) Что представляет собой нестационарная спектральная характеристика сигнала?

7. Аналитический подход к построению моделей

Если известна конструкция объекта управления и физическая сущность протекающих в нем процессов, модель можно найти аналитически – вывести уравнения, связывающие входные и выходные переменные объекта. Наиболее часто для получения моделей используются уравнения материального и энергетического баланса [12, 16, 31].

Материальный баланс в статике: сумма всех входящих и выходящих материальных потоков равна 0. Входящие потоки F_i суммируются со знаком плюс, выходящие – со знаком минус:

$$\sum_{i=1}^n F_i = 0.$$

Приведенное выше уравнение может применяться для расчета не измеряемых расходов. Например, имеется 2 входных потока – один измеряемый F_1 , другой не измеряемый F_2 и измеряемый выходной поток F_3 . Тогда для косвенного измерения потока F_2 может быть использована модель

$$F_2 = F_3 - F_1.$$

Отметим, что для расчета конкретных потоков используются различные математические выражения, определяемые методом измерения и конструкцией датчика. Широко используют измерительные устройства, в которых информацией о расходе является перепад на сужающем устройстве – диафрагме. В этом случае объемный расход F ($\text{м}^3/\text{с}$) рассчитывается по модели

$$F = \alpha \cdot \varepsilon \frac{\pi \cdot d^2}{4} \sqrt{\frac{2\Delta P}{\rho}},$$

где ΔP – перепад давления на сужающем устройстве, $\text{кгс}/\text{м}^2$;

d – диаметр отверстия сужающего устройства, мм;

ρ – плотность, $\text{кг}/\text{м}^3$;

ε – поправочный множитель на расширение сжимаемой среды, для несжимаемых жидкостей $\varepsilon = 1$;

α – коэффициент, зависящий от конструкции диафрагмы.

В системах управления часто важным является не собственно абсолютное значение расхода, а его изменение. В составе встроенного программного обеспечения многих управляющих контроллеров есть алгоритмический блок, вычисляющий значение расхода в процентах по значению перепада давления, которое также задается в процентах:

$$F = 10\sqrt{\Delta P}.$$

Уравнение материального баланса в динамике можно выразить такой словесной формулой: *мгновенное изменение количества вещества = приход – расход*.

Рассмотрим возможные модели объектов управления в системах регулирования уровня.

Пример 1. Объект управления представляет собой емкость с одним входным потоком и одним выходным потоком. При этом в качестве управляющего воздействия в системе автоматического регулирования (САР) уровня может использоваться входной поток, и тогда выходной поток будет являться возмущением, если он не зависит от уровня, а определяется, например, производительностью насоса, который выкачивает жидкость из емкости. Другой вариант построения САР: входной по-

ток является возмущением, выходной – управляющим воздействием. На рис. 5 приведены возможные функциональные схемы автоматизации.

Уравнение материального баланса:

$$T \frac{dL}{dt} = F_{BX} - F_{ВЫХ}.$$

Видно, что размерность в правой части уравнения при объемном расходе – м³/с. Чтобы размерность левой части уравнения совпадала с правой, коэффициент T должен измеряться в м². Это есть площадь сечения емкости. Для первого варианта построения системы автоматического регулирования модель в форме передаточной функции по каналу управления

$$W(s) = \frac{1}{Ts}.$$

Модель по каналу возмущения

$$W(s) = -\frac{1}{Ts}.$$

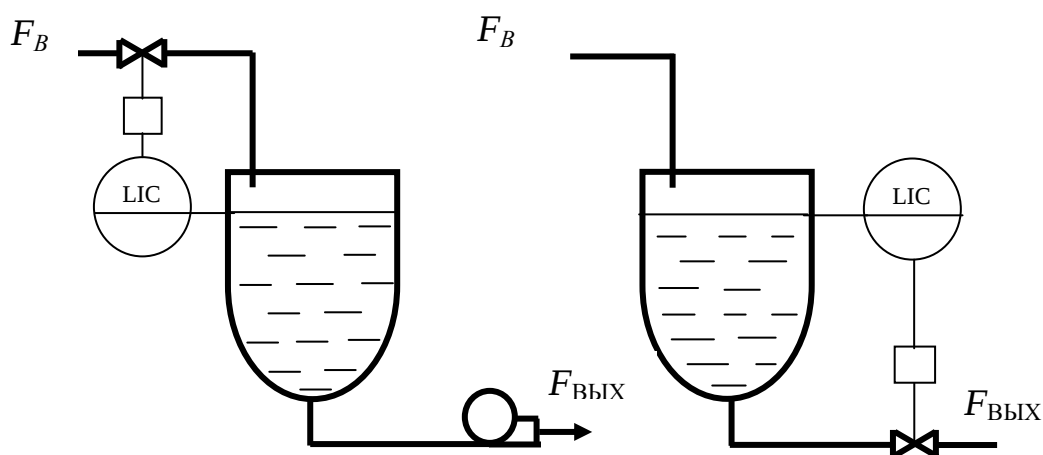


Рис. 5. Функциональные схемы систем автоматического управления уровнем

Для второго варианта системы автоматического регулирования модели по каналу управления и возмущения поменяются местами. Видно, что объект управления для систем, приведенных на рис. 5, с точки зрения теории автоматического регулирования является идеальным интегрирующим звеном с постоянной времени T .

Рассмотрим немного измененную технологическую схему. Пусть в первом варианте системы отсутствует насос, а жидкость из емкости вытекает самотеком. Теперь уравнение материального баланса будет вы-

глядеть по-другому. Выходной поток не является уже независимой переменной, он будет зависеть от того, насколько наполнена емкость:

$$T \frac{dL}{dt} = F_{\text{ВХ}} - b\sqrt{L},$$

где b – коэффициент, зависящий от свойств жидкости.

Модель объекта в данном случае описывается нелинейным дифференциальным уравнением первого порядка. Зависимость выходного потока от уровня приведена на рис. 6.

Если модель объекта управления должна быть использована для построения системы автоматического регулирования, то, скорее всего, необходимо знать поведение объекта в окрестности рабочей точки, соответствующей заданному значению уровня $L_{\text{РАБ}}$. Аппроксимируем нелинейную статическую характеристику в окрестности рабочей точки прямой линией, тангенс угла наклона которой определяется отношением приращения выходного расхода $\Delta F_{\text{ВЫХ}}$ к приращению уровня ΔL :

$$\text{tg } \alpha = \Delta F_{\text{ВЫХ}} / \Delta L.$$

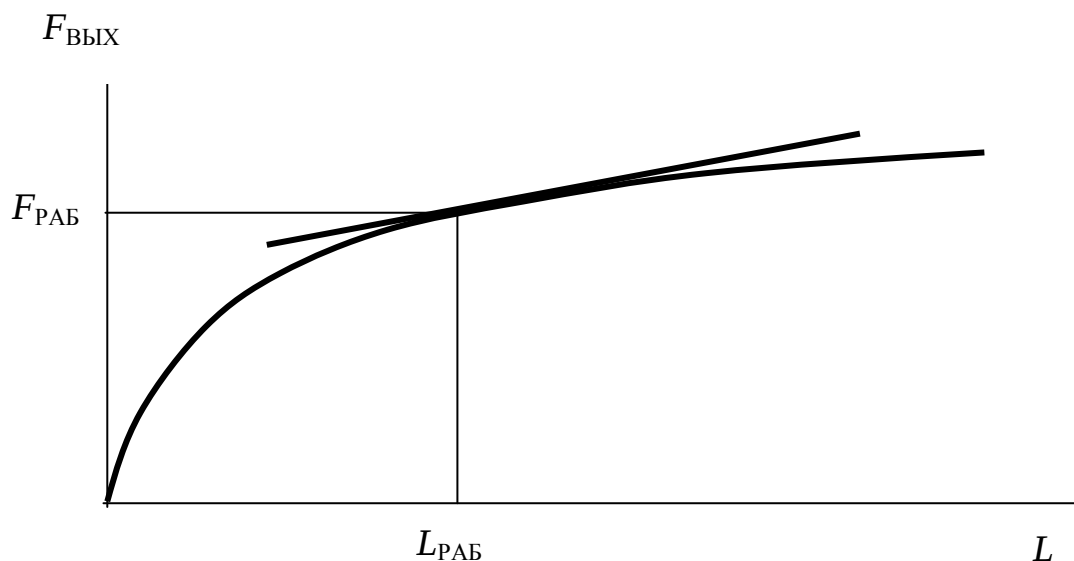


Рис. 6. Зависимость выходного потока от уровня

Запишем уравнение материального баланса в приращениях, в окрестности рабочей точки, соответствующее линейной статической характеристике:

$$T \frac{dh}{dt} = f_{\text{ВХ}} - \text{tg} \alpha \cdot h,$$

где $f_{BX} = F_{BX} - F_{BX,РАБ}$ – приращение управляющего воздействия относительно его рабочего значения $F_{BX,РАБ}$; $h = L - L_{РАБ}$ – приращение выходной переменной относительно ее рабочего значения $L_{РАБ}$.

Перепишем уравнение материального баланса в более удобном виде:

$$\frac{T}{tg\alpha} \frac{dh}{dt} + h = \frac{1}{tg\alpha} f_{BX} \quad \text{или} \quad T_1 \frac{dh}{dt} + h = k f_{BX},$$

где $T_1 = \frac{T}{tg\alpha}, \quad k = \frac{1}{tg\alpha}.$

Таким образом, линеаризованная динамическая модель для емкости со свободным стоком соответствует апериодическому звену первого порядка с постоянной времени T_1 и коэффициентом усиления k .

Пример 2. Объект управления представляет собой смеситель, в котором с помощью мешалки перемешиваются два входных потока: поток F_1 с концентрацией определенного вещества C_1 и поток F_2 с концентрацией такого же вещества C_2 . Выходной поток F_c имеет концентрацию C_c , более равномерную, конечно, чем у входных потоков. Технологический процесс непрерывный, т. е. входные и выходной потоки являются непрерывными. Запишем модель процесса смешения на основе материального баланса. В качестве мгновенно изменяющейся переменной возьмем выходную переменную объекта управления – концентрацию выходного потока C_c :

$$V \frac{dC_c}{dt} = F_1 C_1 + F_2 C_2 - F_c C_c,$$

где F_1, F_2, F_c – объемные потоки, $\text{м}^3/\text{с}$;

C_1, C_2, C_c – концентрации, $\text{кг}/\text{м}^3$;

V – рабочий объем смесителя, м^3 .

Если за переменные входные величины принять концентрации входных потоков, то модель может быть записана в таком виде:

$$T \frac{dy(t)}{dt} + y(t) = k_1 C_1(t) + k_2 C_2(t),$$

где $y(t)$ – выходная переменная C_c ;

T – постоянная времени, $T = V / F_c$;

k_1, k_2 – коэффициенты усиления по входным переменным,

$$k_1 = \frac{F_1}{F_c}, \quad k_2 = \frac{F_2}{F_c}.$$

То есть в этом случае при моделировании мы предполагаем, что входные потоки не меняются и равны определенным величинам, а их концентрации могут изменяться.

При моделировании технологических процессов часто используют типовые математические модели идеального перемешивания и идеального вытеснения. Модель идеального перемешивания описывает непрерывно действующий смеситель с одним входным потоком. Передаточная функция такого объекта имеет вид:

$$W(s) = \frac{1}{Ts + 1}, \quad T = \frac{V}{F}.$$

Здесь V – объем смесителя, а F – рабочее значение потока, который проходит через него, входной поток в данном случае равен выходному. Модель идеального вытеснения представляет собой звено идеального запаздывания с передаточной функцией

$$W(s) = e^{-T_3 s},$$

$$T_3 = L/w,$$

где T_3 – время транспортного запаздывания,

L – длина технологической линии, м,

w – линейная скорость потока, м/с.

С помощью типовых моделей возможно получать модели реальных технологических процессов. Допустим, технологический процесс состоит из двух последовательно включенных смесителей, соединенных достаточно длинной трубой. Модель в виде передаточной функции для такого объекта собирается из типовых моделей:

$$W(s) = \frac{e^{-T_3 s}}{(Ts + 1)^2}.$$

Рассмотрим теперь неидеальный смеситель (что обычно и бывает). Допустим, часть потока (половина) перемешивается по всему объему смесителя. Другая половина перемешивается где-то на дне, в относительно небольшом объеме V_1 . Тогда общая модель процесса может быть получена как параллельное соединение двух идеальных смесителей с разными постоянными времени T и T_1 :

$$W(s) = \frac{0,5}{(Ts + 1)} + \frac{0,5}{(T_1 s + 1)} = \frac{0,5(T_1 + T)s + 1}{(Ts + 1)(T_1 s + 1)}.$$

Рассмотрим применение энергетического баланса для построения моделей. Энергетический баланс в статике: сумма всех входящих и выходящих энергетических потоков равна 0. Входящие потоки Q_i суммируются со знаком плюс, выходящие – со знаком минус:

$$\sum_{i=1}^n Q_i = 0.$$

Энергетический баланс в статике используется для расчета потребления тепла. В любую систему, потребляющую тепло, оно доставляется с потоком теплоносителя. При этом система потребления тепла может быть закрытой, когда входной поток (прямой) $F_{\text{ПР}}$ равен выходному потоку (обратному) $F_{\text{ОБР}}$, или открытой, когда выходной поток меньше входного. Например, если есть возможность слива воды из батареи отопления, имеем дело с открытой системой. Модель расчета потребления тепла для открытой системы:

$$Q = [F_{\text{ПР}}(t_{\text{ПР}} - t_0) - F_{\text{ОБР}}(t_{\text{ОБР}} - t_0)]C\rho,$$

где $F_{\text{ВХ}}$, $F_{\text{ОБР}}$ – прямой и обратный потоки, м³/с;

$t_{\text{ПР}}$, $t_{\text{ОБР}}$ – температура этих потоков, °С;

t_0 – температура на входе источника тепла, обычно 5 °С или 0 °С,

ρ – плотность, кг/м³;

C – удельная теплоемкость, Дж/кг град.

Приведенная выше формула соответствует ГОСТ и используется для расчета тепла в теплосчетчиках.

Для закрытой тепловой системы модель проще:

$$Q = F C \rho (t_{\text{ПР}} - t_{\text{ОБР}}) Q.$$

Энергетический баланс в динамике можно выразить следующей словесной формулой: *мгновенное изменение энергетического состояния объекта = приход энергии – расход энергии.*

Рассмотрим пример применения энергетического баланса для построения математической модели, описывающей изменение температуры в аппарате-смесителе, в котором смешиваются два входных потока: горячий и холодный (рис. 7). В отличие от ранее рассмотренного процесса смешения, теперь нас интересует тепловой режим работы смесителя.

Будем считать, что благодаря перемешиванию содержимого аппарата температура выходного потока θ соответствует средней температуре во всем объеме аппарата. Самое общее уравнение энергетического баланса:

$$MC \frac{d\theta}{dt} = Q_{\text{ВХ}} - Q_{\text{ВЫХ}},$$

где M – масса вещества в смесителе, кг;

C – удельная теплоемкость, Дж/кг град;

θ – температура вещества в смесителе, °С;

$Q_{\text{ВХ}}, Q_{\text{ВЫХ}}$ – входной и выходной потоки тепла.

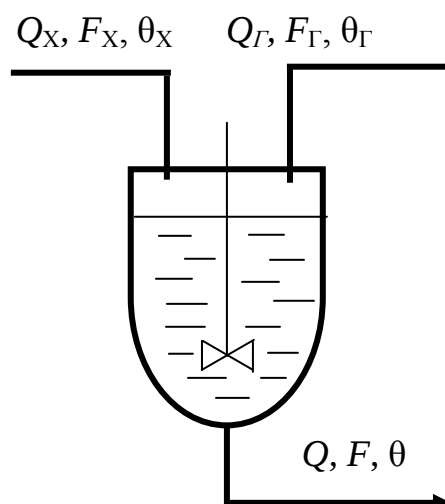


Рис. 7. Смеситель

В приведенном выше общем уравнении энергетического баланса входной тепловой поток складывается из тепла, поступающего с горячим теплоносителем, и тепла, поступающего с холодным теплоносителем. Выходной тепловой поток состоит из тепла, которое рассеивается в окружающее пространство через стенки аппарата, и тепла, которое уходит вместе с выходным потоком. Сделаем два упрощения. Во-первых, не будем учитывать рассеиваемое тепло, обычно это относительно малая величина. Во-вторых, будем рассматривать тепловой баланс относительно средней температуры выходного потока. Если температура регулируется автоматически, то это значение температуры задается в качестве задания для системы автоматического регулирования. Тогда входной холодный поток надо учитывать со знаком «минус», т. к. он понижает температуру в смесителе. Уравнение баланса примет вид:

$$MC \frac{d\theta}{dt} = Q_G - Q_X - Q,$$

где $Q_G = CF_G \rho \theta_G$; $Q_X = CF_X \rho \theta_X$; $Q = CF \rho \theta$;

F_G, F_X, F – объемные горячий, холодный и выходной потоки, м³/с;

Q_G, Q_X, Q – тепловые входные и выходные потоки, Дж/с;
 $\theta_G, \theta_X, \theta$ – температура холодного, горячего и выходного потоков.

Учитывая, что $M = \rho V$, получим следующее уравнение:

$$\frac{V}{F} \frac{d\theta}{dt} + \theta = \frac{F_G \theta_G}{F} - \frac{F_X \theta_X}{F}.$$

При практическом использовании этой модели следует учесть, что будет принято за входные переменные, а что будет задано в виде постоянных значений. Обычно переменными являются входные потоки, а их температуры задаются в виде числовых значений.

Вопросы и задания для самопроверки к разделу 7

1) Дайте формулировку материального баланса в динамике. 2) Имеется емкость, куда непрерывно поступает вода и непрерывно идет откачка воды из емкости. Какой моделью описывается такой объект управления, если управлением является расход воды в ёмкость, а выходной переменной – уровень? 3) А если управление – расход воды из ёмкости? 4) Запишите эти модели в форме дифференциального уравнения и в форме передаточной функции. 5) Как будет выглядеть переходная характеристика такого объекта управления? 6) Какой будет модель относительно уровня, если управление – расход воды в емкость, а из емкости вода бежит самотеком? 7) С какой целью проводится упрощение этой модели?

8. Идентификация динамических объектов методами активного эксперимента

Первые реализованные на практике методы идентификации были основаны на использовании ступенчатых, гармонических и импульсных воздействий. Получив реакцию на эти воздействия и проведя ее обработку, можно получить модель в виде дифференциального уравнения или передаточной функции. При этом структура объекта предполагается известной и результатом идентификации является определение коэффициентов модели.

8.1. Идентификация с помощью переходной функции

Переходная функция является реакцией линейного динамического объекта на единичное ступенчатое воздействие $1(t)$. У идеального ступенчатого сигнала время нарастания сигнала равно нулю, что вообще-то физически невозможно. Поэтому $1(t)$ является идеализированной ап-

проксимацией реального ступенчатого воздействия. Иногда идентификацию с использованием ступенчатых возмущений можно проводить в режиме пассивного эксперимента. В этом случае надо дождаться, когда на исследуемый объект будут поданы ступенчатые возмущения. Например, смена режима работы и, соответственно, изменение задания на регулятор, пуск или останов объекта.

Переходную функцию получают в виде непрерывной кривой или в виде массива значений выходного сигнала, отстоящих друг от друга на интервал дискретизации. Если аппроксимировать переходную функцию аналитическим выражением, можно затем продифференцировать его и получить импульсную переходную функцию. В инженерной практике получили распространение графоаналитические методы, позволяющие по виду переходной функции рассчитать коэффициенты модели известной структуры.

Рассмотрим некоторые из этих методов на примере конкретных моделей.

1. Апериодическое звено первого порядка.

Модель объекта в форме дифференциального уравнения:

$$T \frac{dy(t)}{dt} + y(t) = kx(t).$$

Искомые параметры: коэффициент усиления k и постоянная времени T . Известно, что реакция такого объекта на ступенчатое воздействие описывается выражением

$$y(t) = k(1 - e^{-t/T})x_0,$$

где x_0 – значение ступенчатого скачка на входе объекта.

Если $x_0 = 1$, то $y(t) = h(t) = k(1 - e^{-t/T})$.

Определим значение коэффициента усиления, разделив установившееся значение $y_{уст}$ на значение x_0 : $k = y_{уст}/x_0$. Постоянная времени может быть найдена двумя способами.

Первый способ. Запишем выражение для $y(t)$ при значении времени $t=T$. Получим

$$y(t) = k(1 - e^{-1})x_0 = k(1 - 0,37)x_0 = 0,63kx_0 = 0,63y_{уст}.$$

Методику оценки значения T иллюстрирует рис. 8. Необходимо по оси $y(t)$ отложить 63 % от установившегося значения выходного сигнала. Соответствующее значение переходной функции будет при значении времени, равном постоянной T .

Второй способ. Проведем касательную к переходной функции в точке $t=0$ (рис. 8). Получим

$$\left. \frac{dy(t)}{dt} \right|_{t=0} = \frac{k}{T} e^{-t/T} x_0 \Big|_{t=0} = \frac{kx_0}{T} = \frac{y_{уст}}{T}$$

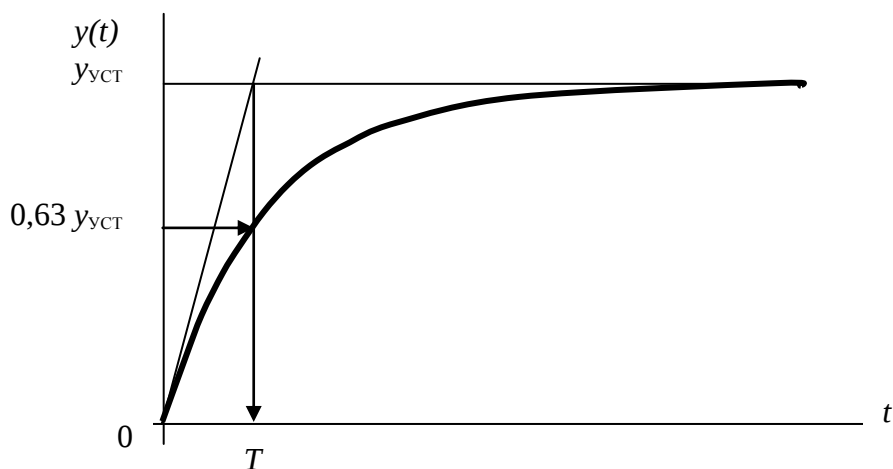


Рис. 8. Определение постоянной времени апериодического звена первого порядка

Т. е. тангенс угла наклона касательной есть отношение установившегося значения выходного сигнала к значению постоянной времени (рис. 8).

На рис. 9 показано определение параметров апериодического звена первого порядка с запаздыванием. Обратите внимание, что значение постоянной времени отсчитывается от момента появления выходного сигнала, отличного от нуля, время от нуля до этого момента – чистое запаздывание τ .

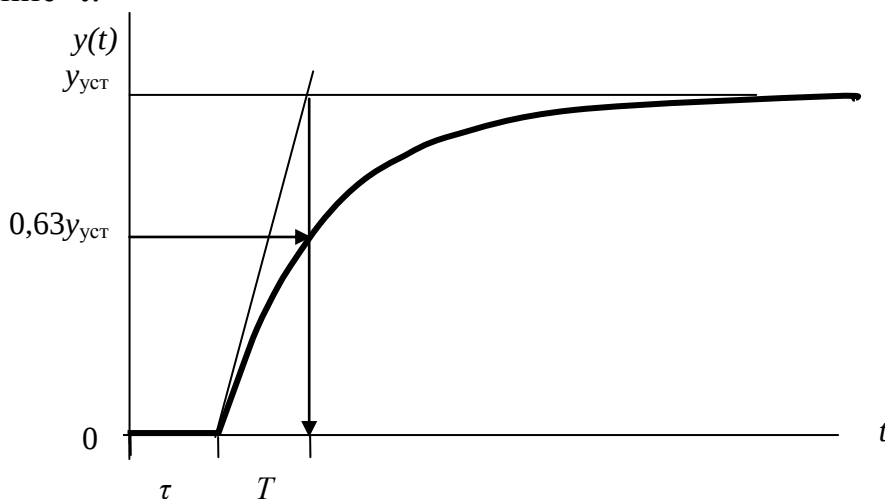


Рис. 9. Определение постоянной времени апериодического звена первого порядка с запаздыванием

Модель этого объекта в форме дифференциального уравнения:

$$T \frac{dy(t)}{dt} + y(t) = kx(t - \tau),$$

в виде передаточной функции $W(s) = \frac{ke^{-\tau s}}{Ts + 1}$.

Существует методика определения параметров апериодического звена 1-го порядка по «кусочку» переходной функции [32]. На рис. 10 показана часть реакции на ступенчатый сигнал, наблюдение велось от момента времени t_1 до времени t_2 .

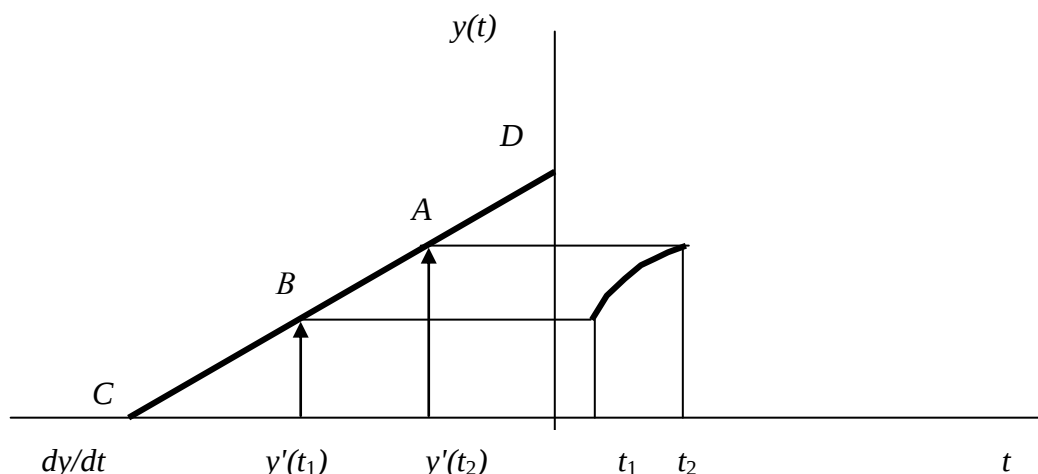


Рис. 10. Определение параметров апериодического звена первого порядка по части переходной функции

Измерим значения производных переходной функции в моменты t_1 и t_2 , отложим их по оси dy/dt и проведем прямую через точки A и B . Прямая соответствует уравнению

$$\frac{dy(t)}{dt} = \frac{kx_0 - y(t)}{T}.$$

Рассмотрим точки пересечения этой прямой с осями dy/dt и y . В точке D $dy/dt = 0$, отсюда

$$k = \frac{y_D}{x_0}.$$

В точке C $y(t)=0$, поэтому $Ty'_C = kx_0$, или $T = \frac{kx_0}{y'_C}$. Здесь y_D – значение $y(t)$ в точке D , а y'_C – значение производной в точке C .

значений корней.

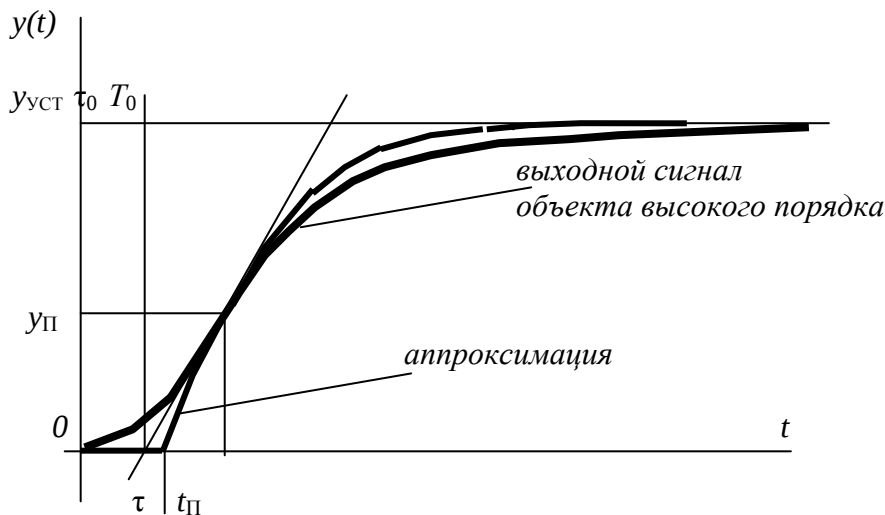


Рис. 11. Аппроксимация переходной функции апериодического звена высокого порядка

При идентификации таких объектов их аппроксимируют более простыми звеньями с передаточной функцией [21]

$$W_A(s) = \frac{ke^{-\tau s}}{(T_1 s + 1)(T_2 s + 1)^n} \text{ или } W_A(s) = \frac{ke^{-\tau s}}{(T_1 s + 1)^n}.$$

Существуют инженерные методики определения значений T_1 , T_2 , n , τ с помощью специальных таблиц и значений T_0 , τ_0 , t_{Π} , U_{Π} , взятых из графика переходной функции исследуемого объекта. Значение коэффициента усиления k определяется как $k = \frac{y_{уст}}{x_0}$, x_0 – значение ступенча-

того воздействия на входе объекта. Рассмотрим наиболее простой случай, когда апериодическое звено высокого порядка будем аппроксимировать апериодическим звеном первого порядка с запаздыванием. Необходимо определить T_1, τ . Значение $n = 1$. Критерий приближения переходной функции аппроксимирующего звена $h_A(t)$ к реальной характеристике $h(t)$ имеет вид:

$$h_A(0) = h(0);$$

$$\begin{aligned}h_{A \text{ уст}} &= h_{\text{уст}}; \\h_A(t_{\Pi}) &= h(t_{\Pi}); \\h'_A(t_{\Pi}) &= h'(t_{\Pi}).\end{aligned}$$

Рассмотрим рис. 11. Очевидно, что два первых условия выполняются: $y_A(0) = y(0)$, $y_{A \text{ уст}} = y_{\text{уст}}$. Запишем значение выходного сигнала аппроксимирующего звена в точке t_{Π} :

$$y_A(t) = 1 - e^{-t/T_1} = \frac{y_{\Pi}}{y_{\text{уст}}}. \text{ Значение производной в этой же точке:}$$

$$+ \frac{1}{T_1} e^{-t/T_1} = \frac{y_{\text{уст}}}{T_0} \quad \text{или} \quad e^{-t/T_1} = \frac{T_1 y_{\text{уст}}}{T_0}.$$

Подставим выражение для экспоненты в первое уравнение:

$$1 - \frac{T_1 y_{\text{уст}}}{T_0} = \frac{y_{\Pi}}{y_{\text{уст}}}, \text{ и выразим отсюда } T_1:$$

$$T_1 = \frac{T_0}{y_{\text{уст}}} \left(1 - \frac{y_{\Pi}}{y_{\text{уст}}}\right). \text{ Осталось найти время чистого запаздывания } \tau.$$

Для этого найдем вначале время t , прошедшее от момента появления выходного сигнала звена первого порядка, отличного от нуля, до момента времени t_{Π} . Запишем значение производной выходного сигнала звена первого порядка для времени t :

$$e^{-t/T_1} = \frac{T_1 y_{\text{уст}}}{T_0} \quad \text{или} \quad e^{+t/T_1} = \frac{T_0}{T_1 y_{\text{уст}}}.$$

$$\text{Отсюда } \frac{t}{T_1} = \ln\left(\frac{T_0}{T_1 y_{\text{уст}}}\right) \quad \text{или} \quad t = T_1 \ln\left(\frac{T_0}{T_1 y_{\text{уст}}}\right). \text{ Тогда } \tau = t_{\Pi} - t.$$

Аппроксимация промышленных объектов управления звеньями первого порядка с запаздыванием широко используется на практике. Так, многие инженерные методы расчета настроек автоматических регуляторов предполагают, что объект управления задается следующими параметрами: постоянной времени, чистым запаздыванием и коэффициентом усиления.

3. Колебательное звено второго порядка. Характеристическое уравнение такого объекта имеет комплексные корни с отрицательной действительной частью. Дифференциальное уравнение объекта в общем виде:

$$a_2 y''(t) + a_1 y'(t) + a_0 y(t) = b_0 x(t).$$

Более удобно представить коэффициенты дифференциального уравнения в виде зависимости от параметров, имеющих определенный физический смысл:

$$a_2 = 1, a_1 = 2\xi\omega_0, a_0 = \omega_0^2, b_0 = \omega_0^2 k,$$

где ξ – коэффициент демпфирования, $0 < \xi < 1$;

ω_0 – частота свободных колебаний;

k – коэффициент усиления.

При коэффициенте демпфирования $\xi = 1$ имеем аperiodический переходный процесс, при его значении $\xi = 0$ получим незатухающий процесс с постоянными колебаниями. График переходной функции колебательного звена (рис. 12) описывается следующим аналитическим выражением:

$$y(t) = k \left[1 - \frac{e^{-\xi\omega_0 t}}{\sqrt{1-\xi^2}} \sin(t\omega_0 \sqrt{1-\xi^2} + \arccos\xi) \right] x_0.$$

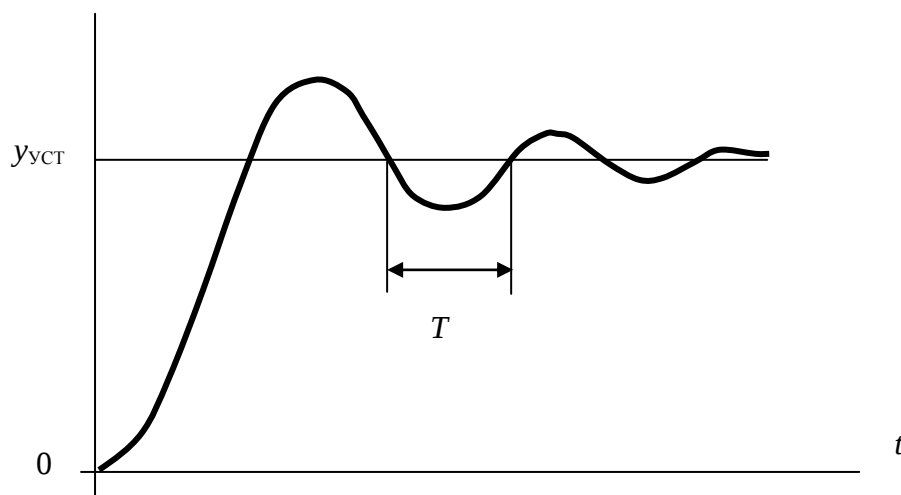


Рис. 12. Переходная функция колебательного звена

Для определения коэффициента демпфирования существуют различные методы. Приводится, например, кривая, связывающая перерегулирование с ξ [8]. Искомый коэффициент ξ удобно определить путем сравнения двух смежных площадей полувольт S_1 и S_2 колебательной составляющей переходной функции:

$$\ln \frac{S_1}{S_2} = \frac{\pi\xi}{\sqrt{1-\xi^2}}.$$

Для определения величины параметра ω_0 необходимо непосредственно по кривой переходного процесса найти собственную круговую частоту колебаний ω по формуле

$$\omega = \frac{\pi}{T}.$$

Частота ω связана с параметром ω_0 соотношением $\omega = \omega_0 \sqrt{1 - \xi^2}$, откуда при известном ξ получим $\omega_0 = \frac{\omega}{\sqrt{1 - \xi^2}}$. Коэффициент усиления

$$k = \frac{y_{уст}}{x_0}.$$

4. Интегрирующие звенья. Передаточная функция звена $W(s) = \frac{1}{Ts}$, таким образом, при идентификации необходимо определить всего один параметр – значение T . На рис. 13 приведены реакции идеального интегрирующего звена и интегрирующего звена с запаздыванием на ступенчатое воздействие.

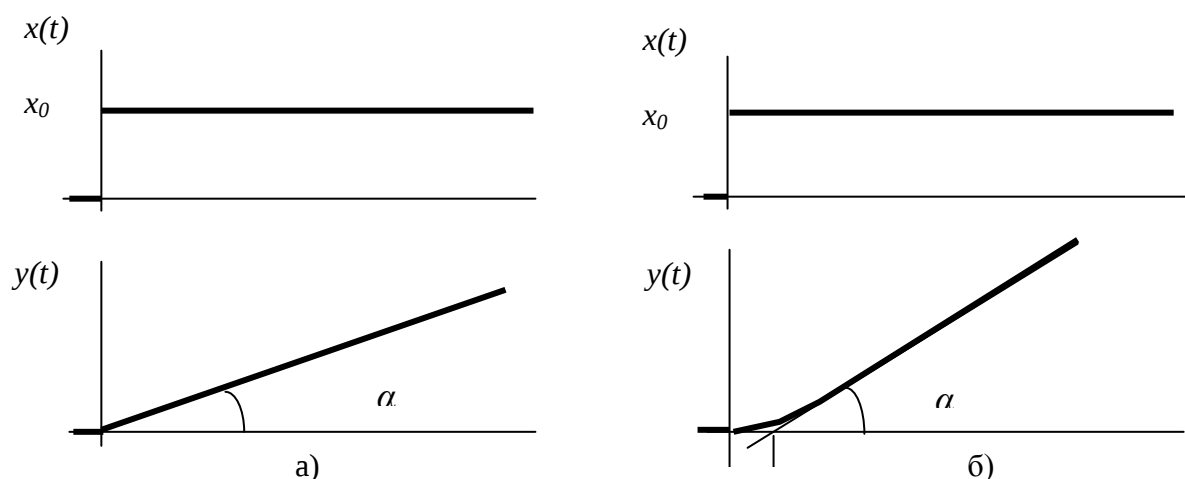


Рис. 13. Переходные функции интегрирующ^τ звеньев: *a* – идеальное; *б* – реальное (с запаздыванием)

Перейдем от $W(s)$ к модели в форме дифференциального уравнения.

$$\frac{1}{Ts} = \frac{Y(s)}{X(s)}; TsY(s) = X(s), Ty'(t) = x(t).$$

Заменим производную отношением приращений: $y'(t) = \frac{\Delta y}{\Delta t}$. Учитываем, что $\frac{\Delta y}{\Delta t} = \operatorname{tg} \alpha$, где α – угол наклона переходной функции. Отсюда $T = \frac{x_0}{\operatorname{tg} \alpha}$.

Для реальных интегрирующих звеньев переходная характеристика вначале имеет небольшой наклон (объект как бы разгоняется), а затем уже изменяется с постоянным углом α (рис. 13, б). Передаточная функция такого объекта:

$$W(s) = \frac{e^{-\tau s}}{Ts}.$$

Время запаздывания τ определяется так, как показано на рис. 13, б.

Рассмотрим пример экспериментального определения модели, описывающей поведение уровня в барабане парового котла. На первый взгляд, имеем объект, представляющий собой идеальное интегрирующее звено с двумя входными воздействиями: расходом питательной воды и расходом пара от котла. Однако если увеличить, например скачком, расход пара, то в действительности уровень сначала быстро возрастает («вскипает») и лишь спустя некоторое время начинает меняться в «правильном» направлении [21]. Объясняется это тем, что в экранных трубах и барабанах котлов высокого давления находится не вода, а пароводяная смесь, объем которой зависит от давления. Увеличение открытия парового клапана приводит к немедленному падению давления над поверхностью испарения в барабане, объем пароводяной смеси увеличивается, что проявляется во временном увеличении уровня. Такой же эффект наблюдается и при увеличении подачи питательной воды, но в другом направлении (рис. 14).

График изменения уровня с учетом эффекта «вскипания» обозначен $L(t)$. Выполним следующие вспомогательные действия:

1. проведем из начала координат линию $A(t)$ под углом α ;
2. измерим расстояние L_1 от линии $A(t)$ до оси времени в момент появления положительного значения переходной функции $L(t)$;

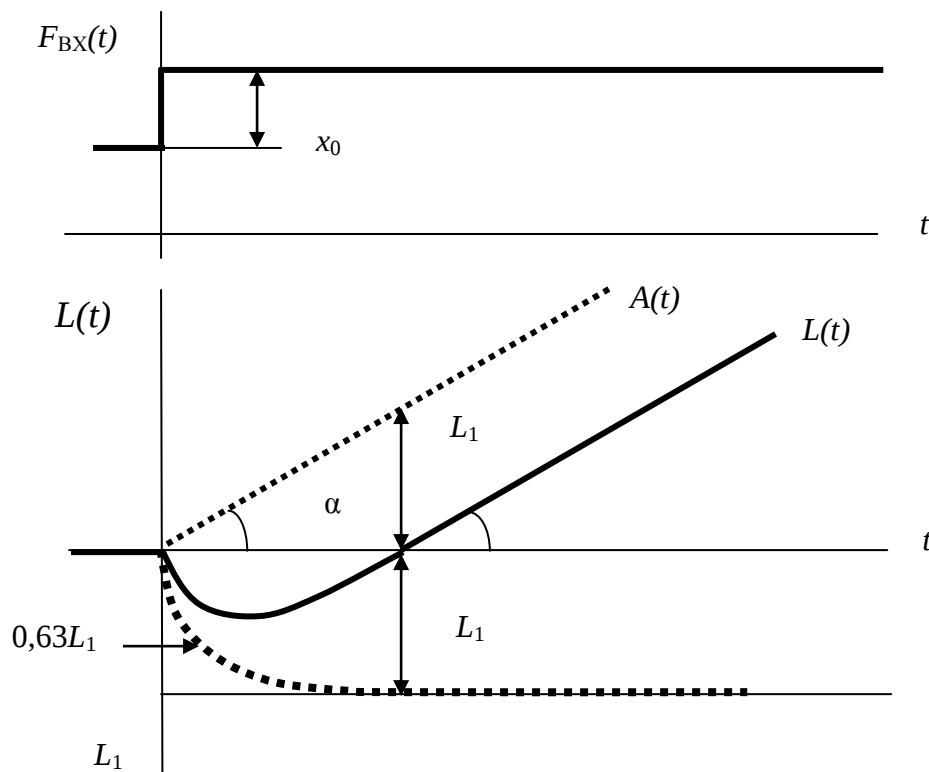


Рис. 14. Изменение уровня в барабане парового котла

3. отложим значение L_1 вниз от начала координат;

4. построим переходную функцию апериодического звена первого порядка $F(t)$, которая выходит из начала координат и достигает установившегося значения в момент перехода $L(t)$ через ноль.

Тогда $L(t) = A(t) - F(t)$. Поскольку объект линейный, такое же соотношение справедливо для передаточных функций: $W(s) = W_{\text{и}}(s) - W_{\text{а}}(s)$, где $W(s)$ – звено с переходной функцией $L(t)$;

$W_{\text{и}}(s)$ – интегрирующее звено с переходной функцией $A(t)$;

$W_{\text{а}}(s)$ – апериодическое звено с переходной функцией $F(t)$.

Таким образом,

$$W(s) = \frac{1}{T_1 s} - \frac{k}{T_2 s + 1} = \frac{1 + (T_2 - kT_1)s}{T_1 s(T_2 s + 1)},$$

где $T_1 = \frac{x_0}{tg\alpha}$, $k = \frac{L_1}{x_0}$, T_2 определяется по ранее изложенной методике оценки постоянной времени для апериодического звена первого порядка.

8.2. Частотный метод идентификации

Частотный метод – один из первых методов идентификации. Предназначен для применения в лабораторных условиях. Предполагается, что имеется возможность искусственного воздействия на объект управления синусоидальными колебаниями различной частоты. Исследуемый объект – линейный, стационарный. Должна быть известна структура объекта, т. е. общий вид дифференциального уравнения или передаточной функции. В результате идентификации должны быть определены коэффициенты модели.

Существует две разновидности метода [32]. В первом случае предполагается, что имеется возможность измерения амплитуды входного сигнала, амплитуды выходного сигнала и фазового сдвига между сигналом на входе и выходе. Известна частота генерируемого на входе сигнала. Итак, пусть на вход подается сигнал $x(t) = M\sin(\omega_1 t)$, на выходе измеряется сигнал $y(t) = N\sin(\omega_1 t + \varphi)$. Амплитудная частотная характеристика на частоте ω_1 равна: $A(\omega_1) = N/M$. Зная сдвиг по фазе φ , легко определить значения вещественной частотной характеристики $P(\omega_1)$ и мнимой частотной характеристики $Q(\omega_1)$ на частоте ω_1 :

$$Q(\omega_1) = A(\omega_1) \sin \varphi, \quad P(\omega_1) = A(\omega_1) \cos \varphi.$$

Назовем полученные значения экспериментальными – $Q_{\text{ЭКС}}(\omega_1)$ и $P_{\text{ЭКС}}(\omega_1)$. Пусть исследуемый объект описывается следующей передаточной функцией:

$$W(s) = \frac{1 + b_1 s + \dots + b_m s^m}{a_0 + a_1 s + \dots + a_n s^n}.$$

Заменим s на $j\omega$ и запишем комплексную АФЧХ для частоты ω_1 :

$$W(j\omega_1) = \frac{1 + b_1(j\omega) + \dots + b_m(j\omega)^m}{a_0 + a_1(j\omega) + \dots + a_n(j\omega)^n} = P(\omega_1) + jQ(\omega_1).$$

Теперь можем приравнять теоретические выражения ВЧХ и МЧХ, полученные из передаточной функции их экспериментальным значениям:

$$P(\omega_1) = P_{\text{ЭКС}}(\omega_1);$$

$$Q(\omega_1) = Q_{\text{ЭКС}}(\omega_1).$$

Таким образом, имеем два независимых уравнения для определения неизвестных коэффициентов модели. Всего таких коэффициентов $k = m + n + 1$. Чтобы получить систему из k независимых уравнений, необходимо провести эксперимент на $k/2$ различных частотах. В случае

если число коэффициентов модели нечетное, можно предварительно определить коэффициент усиления $1/a_0$, подав на вход исследуемого объекта постоянный сигнал. Один из возможных методов заключается также в проведении заведомо большого числа экспериментов на различных частотах, а затем решение переопределенной системы уравнений методом наименьших квадратов.

Рассмотрим пример. Исследуемый объект описывается передаточной функцией $W(s) = \frac{1}{a_0 + a_1 s}$. Необходимо определить два неизвест-

ных коэффициента a_0 и a_1 . Достаточно провести один эксперимент на частоте ω_1 , в итоге имеем экспериментальные значения ВЧХ и МЧХ, P_1 и Q_1 . Приравняем выражение для комплексной АФЧХ полученным значениям:

$$\frac{1}{a_0 + j\omega_1 a_1} = P_1 + jQ_1.$$

Выделим из этого равенства вещественную и мнимую части:

$$P_1 a_0 - Q_1 \omega_1 a_1 = 1;$$

$$Q_1 a_0 - P_1 \omega_1 a_1 = 0.$$

Решая систему уравнений, получим ответ:

$$a_0 = \frac{P_1}{P_1^2 + Q_1^2}; \quad a_1 = -\frac{Q_1}{\omega_1(P_1^2 + Q_1^2)}.$$

Вторая разновидность частотного метода идентификации. Известна частота сигнала, подаваемого на вход объекта, измеряется амплитуда входного сигнала $M(\omega)$ и выходного сигнала $N(\omega)$. Угол сдвига по фазе между входным и выходным сигналом неизвестен, не измеряется. В этом случае из передаточной функции выводим теоретическое выражение для АЧХ и приравниваем его полученному экспериментальному значению $A = N/M$ для частоты ω , на которой проводился эксперимент. Таким образом, в этом случае для определения k неизвестных коэффициентов модели потребуется столько же экспериментов на k частотах.

Рассмотрим тот же пример. Для определения неизвестных коэффициентов a_0 и a_1 подаем на вход объекта сигналы двух разных частот: $x(\omega_1) = M \sin \omega_1 t$ и $x(\omega_2) = M \sin \omega_2 t$. Замерив значения амплитуды выходного сигнала $N_1(\omega_1)$ и $N_2(\omega_2)$, рассчитаем экспериментальные значения АЧХ: $A_1 = N_1(\omega_1)/M$, $A_2 = N_2(\omega_2)/M$. Найдем теоретическое выражение для АЧХ:

$$\frac{1}{a_0 + j\omega a_1} = \frac{(a_0 - j\omega a_1)}{(a_0 - j\omega a_1)} = \frac{a_0}{a_0^2 + \omega^2 a_1^2} - j \frac{\omega a_1}{a_0^2 + \omega^2 a_1^2};$$

$$A(\omega) = \sqrt{P^2(\omega) + Q^2(\omega)} = \sqrt{\frac{1}{a_0^2 + \omega^2 a_1^2}}.$$

Запишем систему уравнений для определения коэффициентов модели:

$$\sqrt{\frac{1}{(a_0^2 + \omega_1^2 a_1^2)}} = A_1;$$

$$\sqrt{\frac{1}{(a_0^2 + \omega_2^2 a_1^2)}} = A_2,$$

или

$$A_1^2(a_0^2 + \omega_1^2 a_1^2) = 1, \quad A_2^2(a_0^2 + \omega_2^2 a_1^2) = 1.$$

Полученная система уравнений линейна относительно квадратов коэффициентов. Её решение:

$$a_0 = \frac{1}{A_1 A_2} \sqrt{\frac{A_2^2 \omega_2^2 - A_1^2 \omega_1^2}{\omega_2^2 - \omega_1^2}},$$

$$a_1 = \frac{1}{A_1 A_2} \sqrt{\frac{A_1^2 - A_2^2}{\omega_2^2 - \omega_1^2}}.$$

Отметим, что для более сложных объектов система уравнений будет обычной нелинейной системой, решение которой не так просто, как в этом примере. Это своего рода «расплата» за уменьшение информации об объекте, получаемой при идентификации, – отсутствие измерений сдвига сигнала по фазе.

Вопросы и задания для самопроверки к разделу 8

1) Запишите модель апериодического звена 1-го порядка с запаздыванием в форме дифференциального уравнения и в форме передаточной функции. 2) Как будет выглядеть для этого звена импульсная переходная функция? 3) Какие физические объекты описываются апериодическим звеном 1-го порядка? 4) Какая информация должна измеряться на входе и выходе идентифицируемого объекта при использовании частотного метода? 5) Какая информация должна быть предварительно известна об идентифицируемом объекте при использовании частотного метода? 6) Сколько экспериментов необходимо провести для идентификации объекта с передаточной функцией $W(s) = \frac{k_1 + k_2 s}{(T_1 s^2 + T_2 s + 1)}$ по одной и по второй разновидности частотного метода?

9. Характеристики случайных процессов

При решении задач идентификации и диагностики зачастую приходится иметь дело с сигналами, представляющими собой случайные процессы. Возникает задача расчета статистических характеристик, причем в качестве исходной информации используются дискретные значения, снятые через определенный интервал дискретизации. Рассмотрим основные статистические характеристики, их свойства и применение в задачах идентификации и диагностики [7, 20, 29].

Математическое ожидание. Допустим, на определенном интервале времени получена выборка N значений процесса $x_1, x_2, \dots, x_N$. Среднее значение, или математическое ожидание, находим по формуле

$$m_x = \frac{1}{N} \sum_{i=1}^N x_i.$$

Если известны частоты p_j , с которыми в выборке присутствуют значения x_j , то формула для расчета математического ожидания примет такой вид:

$$m_x = \frac{1}{N} \sum_{j=1}^K p_j x_j,$$

где K – число групп чисел в выборке, которые отличаются друг от друга.

В алгоритмах, которые работают в темпе со временем, необходимо пересчитывать математическое ожидание с появлением каждого нового измерения. Для этого удобно использовать рекуррентную формулу [5]:

$$m_x[N+1] = m_x[N] + \frac{1}{N+1} (x_{N+1} - m_x[N]).$$

Здесь $m_x[N]$ и $m_x[N+1]$ – величины математического ожидания, вычисленные по N и по $N+1$ значениям x , соответственно.

Дисперсия. С помощью дисперсии можно оценить разброс относительно среднего.

$$D_x = \frac{1}{N-1} \sum_{i=1}^N (x_i - m_x)^2 \quad \text{или} \quad D_x = \frac{1}{N} \sum_{i=1}^N (x_i - m_x)^2.$$

Первая формула не даст результата, если проведено лишь одно измерение. И действительно, как можно оценить в такой ситуации разброс значений? Вторая формула обычно используется при малых выборках [7].

Рекуррентное оценивание дисперсии при появлении $(N+1)$ -го измерения

$$D_x[N+1] = D_x[N] + \frac{1}{N+1} [(x_{N+1} - m_x[N+1])^2 - D_x[N]].$$

Среднеквадратическое отклонение σ (СКО). СКО несет в себе ту же информацию, что и дисперсия, но измеряется в тех же единицах, что и исследуемый процесс.

$$\sigma = \sqrt{D_x} \text{ или } D_x = \sigma^2.$$

Коэффициент корреляции (или ковариации) r_{xy} . Иногда используется термин «парный коэффициент корреляции». Он характеризует степень тесноты связи между переменными x и y :

$$r_{xy} = \frac{1}{N-1} \sum_{i=1}^N (x_i - m_x)(y_i - m_y).$$

Применяется также обозначение $r_{xy} = \text{cov}(x, y)$. Эту характеристику нельзя применять для сравнительной оценки тесноты связи разных параметров, т. к. полученные значения будут зависеть от единиц измерения этих параметров, а они могут быть абсолютно различными.

Нормированный коэффициент взаимной корреляции. А вот эта характеристика может применяться для сравнительной оценки тесноты связи разных параметров, т. к. ее значение всегда лежит в пределах от минус единицы до плюс единицы.

$$\rho_{xy} = \frac{\text{cov}(x, y)}{\sigma_x \sigma_y} = \frac{\frac{1}{N-1} \sum_{i=1}^N (x_i - m_x)(y_i - m_y)}{\sqrt{D_x D_y}}, \quad -1 \leq \rho_{xy} \leq +1.$$

Если исследуемые параметры x и y меняются в одном направлении (при увеличении одного – другой тоже увеличивается), значение нормированного коэффициента корреляции будет положительным. Если, например, при уменьшении x значения y увеличиваются, то такая корреляция будет отрицательной. Таким образом, сильная корреляция может быть как положительной, так и отрицательной. Если же значение ρ_{xy} будет равно плюс или минус единице, это означает, что между этими параметрами имеется не случайная, а функциональная зависимость.

Рассмотрим подробнее вопросы применения нормированного коэффициента корреляции. На начальном этапе построения модели часто стоит задача выбора входных параметров объекта, которые оказывают наибольшее влияние на выходную величину. Степень влияния можно оценить по нормированному коэффициенту корреляции, но для кор-

ректного использования этого показателя необходимо, чтобы эти входные переменные не влияли друг на друга, т. е. были не коррелированы между собой.

Первый вопрос при проведении корреляционного анализа: существует ли вообще корреляционная связь между параметрами? Итак, выдвигаем статистическую гипотезу: нормированный коэффициент корреляции $\rho \equiv 0$. Противоположная гипотеза: $\rho \neq 0$. В качестве критерия при принятии одной либо другой гипотезы используется t -критерий Стьюдента. Рассчитаем значение критерия на основе экспериментальных данных:

$$t_{\text{ЭКС}} = \frac{|\hat{\rho}| \sqrt{N-2}}{\sqrt{1-\hat{\rho}^2}}.$$

Зададим определенный уровень значимости α (это наибольшая вероятность появления события, которое мы считаем практически невозможным). Пусть $\alpha = 0,05$ (5 %). По таблицам распределения Стьюдента для числа степеней свободы $N-2$ и при $\alpha = 0,05$ получим критическое значение критерия $t_{\text{КРИТ}}$. Если $t_{\text{ЭКС}} < t_{\text{КРИТ}}$, справедлива гипотеза об отсутствии корреляционной связи.

Второй вопрос, который может возникнуть: а каков доверительный интервал коэффициента корреляции? Итак, пусть в результате обработки N измерений получена оценка $\hat{\rho}$. Насколько можно ей верить? Доверительный интервал можно также определить с использованием t -критерия Стьюдента:

$$\hat{\rho} - t(\alpha/2) \frac{1-\hat{\rho}^2}{\sqrt{N}} < \rho < \hat{\rho} + t(\alpha/2) \frac{1-\hat{\rho}^2}{\sqrt{N}},$$

где $t(\alpha/2)$ – значение t -критерия для уровня значимости $\alpha/2$.

Приведенная выше формула показывает, в каких пределах лежит значение коэффициента корреляции с доверительной вероятностью $p = 1 - \alpha$. Рассмотрим числовой пример. Пусть $N = 100$, $\rho = 0,5$, $\alpha = 0,05$, значение t -критерия при таких условиях ≈ 2 . Получаем:

$$0,5 - 2 \frac{1-0,25}{10} < \rho < 0,5 + 2 \frac{1-0,25}{10}, \text{ или } 0,35 < \rho < 0,65.$$

Таким образом, с вероятностью 95 % мы утверждаем, что значение коэффициента корреляции лежит в пределах от 0,35 до 0,65.

Автокорреляционная функция. Автокорреляционная функция (АКФ) $R_{xx}(\tau)$ показывает тесноту связи между значениями сигнала, отстоящими друг от друга на определенное время τ . Значение корреляционной функции убывает с ростом аргумента тем быстрее, чем быстрее

изменяется во времени случайный процесс $x(t)$, поскольку при этом ослабляется связь между значениями x , отстоящими на фиксированный промежуток времени τ . Так как вычисление АКФ в нашем случае ведется по дискретным значениям сигнала, отстоящим друг от друга на интервал дискретизации ΔT , то $\tau = k\Delta T$, $k = 0, 1, 2, 3, \dots$, т. е. АКФ будет функцией от дискретного аргумента k . При $k = 1$ получим величину тесноты связи между измерениями сигнала, отстоящими друг от друга на интервал ΔT , при $k = 2$ – на $2\Delta T$ и т. д. Расчетная формула:

$$R_{xx}(k) = \frac{1}{N-1-k} \sum_{i=1}^{N-k} (x_i - m_x)(x_{i+k} - m_x), \quad k \ll N.$$

При $k = 0$ получаем значение дисперсии сигнала: $R(0) = D_x$. Это наибольшее значение АКФ. При $k \rightarrow \infty$ значение АКФ стремится к нулю. АКФ является четной функцией, $R_{xx}(k) = R_{xx}(-k)$. Часто АКФ аналитически описывают в виде затухающей экспоненты.

Иногда для вычисления АКФ используют не центрированные значения сигналов. Обозначим такую АКФ $G_{xx}(k)$:

$$G_{xx}(k) = \frac{1}{N-1-k} \sum_{i=1}^{N-k} x_i x_{i+k}.$$

Найдем связь между $R_{xx}(k)$ и $G_{xx}(k)$. Запишем формулу для АКФ, перемножив скобки под знаком суммы, а также используя символ усреднения $M\{\}$:

$$\begin{aligned} R_{xx}(k) &= M\left\{\sum (x_i x_{i+k} - m_x x_{i+k} - m_x x_i + m_x^2)\right\} = \\ &= M\left\{\sum x_i x_{i+k}\right\} - m_x M\left\{\sum x_{i+k}\right\} - m_x M\left\{\sum x_i\right\} + m_x^2 = G_{xx}(k) - m_x^2 - m_x^2 + m_x^2. \end{aligned}$$

В итоге: $R_{xx}(k) = G_{xx}(k) - m_x^2$. Видно, что АКФ для нецентрированных сигналов будет иметь ту же форму, что и классическая АКФ, но при этом будет смещена на квадрат математического ожидания.

Взаимная корреляционная функция. Взаимная корреляционная функция (ВКФ) $R_{xy}(\tau)$ характеризует тесноту связи между значениями двух сигналов, отстоящими друг от друга на время τ . При работе с дискретными значениями сигналов аргумент $\tau = k\Delta T$. Значения ВКФ вычисляются по формулам:

$$R_{xy}(k) = \frac{1}{N-1-k} \sum_{i=1}^{N-k} (x_i - m_x)(y_{i+k} - m_y), \quad k \ll N,$$

или

$$R_{yx}(k) = \frac{1}{N-1-k} \sum_{i=1}^{N-k} (y_i - m_y)(x_{i+k} - m_x).$$

ВКФ не обязательно принимает максимальное значение при аргументе $k = 0$. Наоборот, для динамических систем вследствие инерции максимальная связь между сигналом на выходе и сигналом на входе наступает спустя какое-то время. ВКФ не обладает свойствами четности, т. е. $R_{xy}(\tau) \neq R_{yx}(\tau)$. С увеличением значения аргумента ВКФ стремится к нулю. ВКФ также можно рассчитывать с использованием нецентрированных сигналов:

$$G_{xy}(k) = \frac{1}{N-1-k} \sum_{i=1}^{N-k} x_i y_{i+k}.$$

По аналогии с уже рассмотренным выводом для АКФ, можно получить связь между ВКФ для центрированных и нецентрированных сигналов:

$$R_{xy}(k) = G_{xy}(k) - m_x m_y.$$

Спектральная плотность. Спектральная плотность $S_{xx}(\omega)$ получается из АКФ путем преобразования Фурье:

$$S_{xx}(\omega) = \int_{-\infty}^{+\infty} R_{xx}(t) e^{-j\omega t} dt.$$

Поскольку АКФ является четной функцией, выражение для спектральной плотности удобно представить в тригонометрической форме:

$$S_{xx}(\omega) = 2 \int_0^{\infty} R_{xx}(t) \cos(\omega t) dt.$$

Спектральную плотность называют ещё энергетическим спектром сигнала. Площадь на графике спектральной плотности, ограниченная частотами ω_1, ω_2 , показывает мощность, выделяемую сигналом в данном частотном диапазоне.

Одна из абстрактных математических моделей сигнала – «белый шум» – предполагает равномерное рассеивание мощности во всем диапазоне частот. На графике спектральная плотность «белого шума» изображается прямой линией. АКФ «белого шума» есть дельта-функция.

На рис. 15 в сравнении показаны АКФ и спектральные плотности двух различных сигналов, отличающихся по своему частотному спектру.

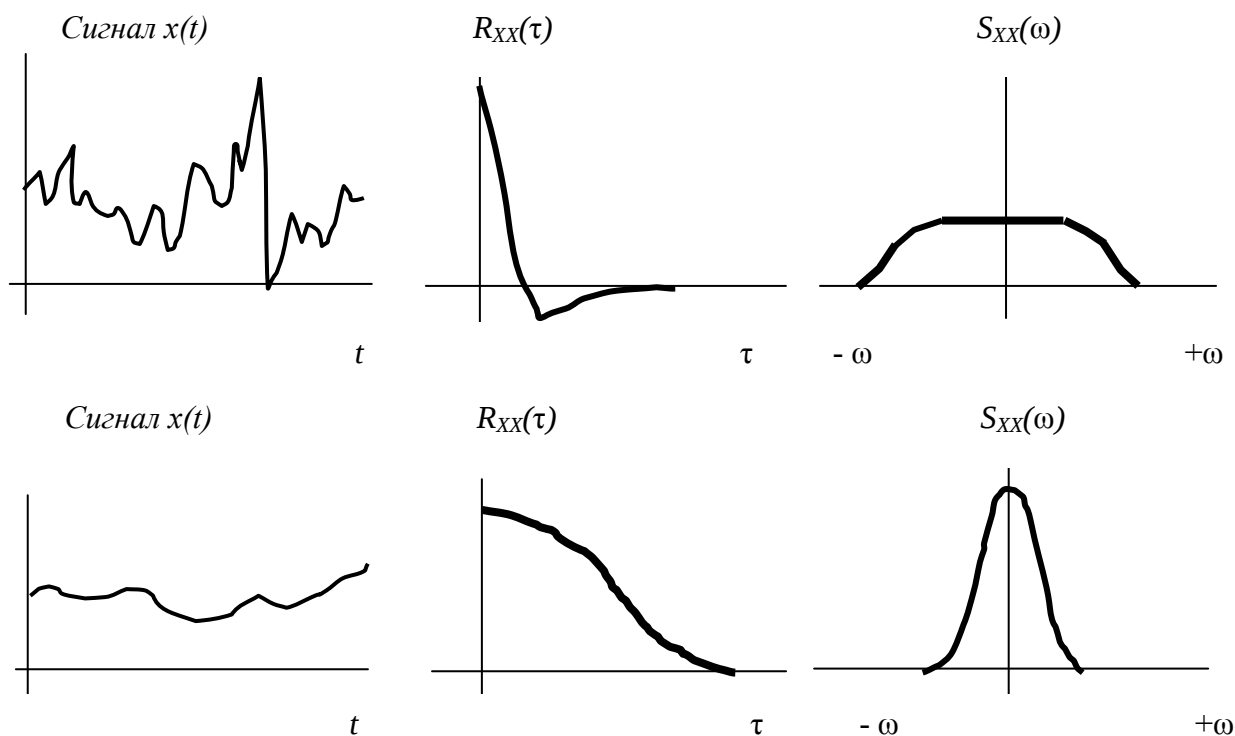


Рис. 15. Сравнительный вид сигналов, АКФ и спектральных плотностей

В качестве частотной статистической связи между двумя стационарными случайными процессами $x(t)$ и $y(t)$ применяется взаимная спектральная плотность, представляющая изображение Фурье взаимной корреляционной функции:

$$S_{xy}(\omega) = \int_{-\infty}^{+\infty} R_{xy}(t) e^{-j\omega t} dt.$$

Вопросы и задания для самопроверки к разделу 9

1) Какую информацию несет в себе АКФ? 2) Как называется значение АКФ при аргументе = 0? 3) Какую информацию несет в себе ВКФ? 4) Могут ли быть значения ВКФ отрицательными? 5) Может ли быть значение АКФ при нулевом аргументе меньше, чем при ненулевом аргументе? 6) Может ли быть значение ВКФ при нулевом аргументе меньше, чем при ненулевом аргументе? 7) Почему для сравнительной оценки тесноты связи используется нормированный коэффициент корреляции, а не обычный коэффициент корреляции? 8) Какое требование должно выполняться относительно входных параметров, если исследуется их влияние на выходной показатель с помощью нормированного коэффициента взаимной корреляции? 9) Какой статистический критерий используется для оценки значимости корреляции? 10) Что нужно задать при использовании этого критерия?

10. Предварительная обработка экспериментальных данных

Прежде чем пользоваться экспериментальными данными для идентификации, следует быть уверенным в их достоверности. Необходимо правильно выбрать интервал дискретизации, избавиться от искажений, вносимых помехами при измерении входных и выходных сигналов. Ниже рассмотрены эти вопросы более подробно.

10.1. Дискретизация сигналов

Сигналы, используемые для решения задач идентификации и диагностики, в подавляющем большинстве случаев являются непрерывными. Первая задача, которая возникает – определение интервала дискретизации ΔT . При обработке сигналов с помощью имеющихся на объекте микропроцессорных систем управления минимальное значение интервала дискретизации определяется тактом работы контроллера. Например, в системе Delta-V фирмы Emerson опрос параметров ведется с частотой 200 мс, а обновление информации на трендах и мнемосхемах – с частотой 1 с. У отечественного контроллера Ремиконт Р-130 такт опроса задается от 0,2 с до 2 с. Естественно, что такт времени при архивировании данных должен быть больше, чем такт работы контроллера.

Рассмотрим некоторые подходы для определения интервала дискретизации.

Применение теоремы Котельникова. В 1933 году В.А. Котельников сформулировал следующую теорему (в зарубежной литературе она известна как «теорема об отсчетах»): если функция $f(t)$ непрерывна и частотный спектр ее не содержит составляющих с частотой, превышающей F Гц, то она полностью определяется совокупностью ординат, отстоящих друг от друга на расстоянии в $1/2F$ с. Таким образом, необходимо предварительно провести анализ частотных свойств сигнала, определить наибольшую частоту в спектре F Гц, чему соответствует угловая частота $\omega=2\pi F$, а затем уже определять интервал дискретизации. Такой подход, скорее всего, может быть использован в задачах диагностики, когда проводится анализ спектральных свойств сигнала и при изменении этих свойств делается вывод о ненормальном функционировании объекта и о причинах этого. В задачах идентификации использование теоремы Котельникова потребует значительной предварительной рабо-

ты по определению F и даст оценку интервала дискретизации с большим запасом.

Оценка интервала дискретизации по максимальной скорости изменения сигнала. Изучая физические принципы функционирования исследуемого объекта или обращаясь к специалистам в данной предметной области, можно узнать предельную скорость изменения сигнала в самом экстремальном случае. Далее, задавшись допустимой максимальной ошибкой дискретизации за счет замены непрерывного сигнала его дискретными значениями, можно определить значение ΔT .

Пример. Сигнал, представляющий собой изменение температуры во времени, может меняться с максимальной скоростью $20\text{ }^{\circ}\text{C}/\text{мин}$. Диапазон изменения температуры – от нуля до $200\text{ }^{\circ}\text{C}$. Считаем, что приемлемой максимальной ошибкой за счет дискретизации будет ошибка в 1% , или в абсолютных единицах – $2\text{ }^{\circ}\text{C}$. Тогда, исходя из известной максимальной скорости изменения температуры, $\Delta T = 1/10$ мин.

Оценка интервала дискретизации по допустимой среднеквадратичной ошибке дискретизации. Методика предложена Э. Л. Ицковичем, при допущении, что в промежутках между замерами используется ступенчатая интерполяция сигнала. Обозначим среднеквадратичную ошибку дискретизации δ . Эта ошибка связана с корреляционной функцией сигнала R_{xx} :

$$\delta^2 = 2(R_{xx}[0] - R_{xx}[j\Delta t]),$$

где Δt – интервал дискретизации, использованный при расчете корреляционной функции.

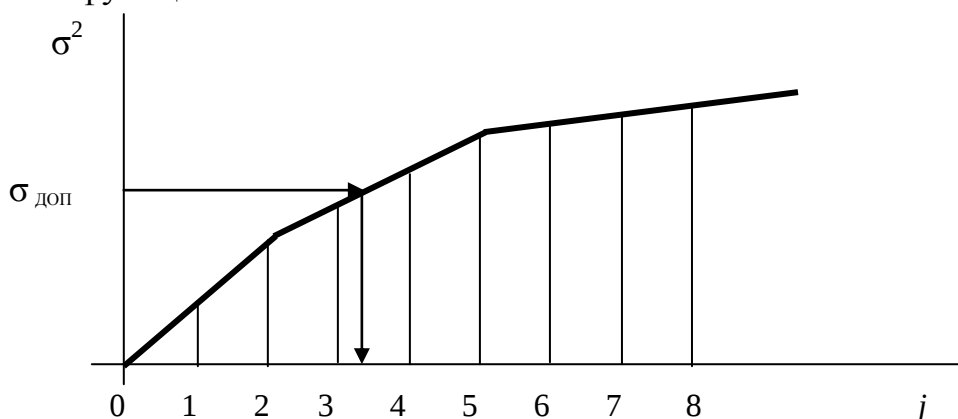


Рис. 16. Определение интервала дискретизации по допустимой СКО

Построив график, соответствующий этой формуле, и задавшись допустимой ошибкой δ , определим необходимый интервал дискрети-

зации ΔT (рис. 16). Если получилось так, что $\Delta T < \Delta t$, то необходимо пересчитать корреляционную функцию при меньшем Δt , построить график заново и найти ΔT .

10.2. Фильтрация сигналов

Когда говорят о фильтрации сигналов, используемых для идентификации, обычно имеют в виду фильтрацию от высокочастотных помех. Поэтому часто вместо термина «фильтрация» используется термин «сглаживание». Эта задача решается с помощью применения низкочастотных фильтров. Постановка задачи: на вход объекта поступает сигнал $z(t)$, состоящий из полезного сигнала $x(t)$ и аддитивной помехи $n(t)$. Сигнал на выходе фильтра $\hat{x}(t) \rightarrow x(t)$ (рис. 17).



Рис. 17. Задача фильтрации

По способу выполнения фильтры бывают аналоговые и дискретные. Дискретные фильтры могут быть реализованы схемно, на дискретных элементах и в виде алгоритма, реализованного в контроллере или компьютере. Рассмотрим некоторые алгоритмы фильтрации.

Фильтр скользящего среднего.

$$\hat{x}[j] = \frac{1}{k} \sum_{i=0}^{k-1} z[j-i].$$

Настроечным параметром фильтра является длина скользящего интервала k . Чем больше длина интервала, тем сильнее сглаживающие свойства фильтра.

Пример. Пусть имеются данные по измерениям сигнала на входе фильтра (табл. 1). Имитируем работу фильтра. Пусть длина скользящего интервала равна 4. Тогда алгоритм начнет работу, начиная с четвертой точки:

$$\hat{x}[4] = \frac{4+4+2+6}{4} = 4 ; \quad \hat{x}[5] = \frac{4+2+6+7}{4} = 4,75 ;$$

$$\hat{x}[6] = \frac{2+6+7+2}{4} = 4,25.$$

Таблица 1

Числовые данные к примеру

Номер измерения	1	2	3	4	5	6
$z[j]$	4	4	2	6	7	2

Фильтр экспоненциального сглаживания. Для непрерывных сигналов аналоговый фильтр описывается следующей передаточной функцией:

$$W(s) = \frac{\gamma}{\gamma + s},$$

где γ – параметр фильтра.

То есть фильтр представляет собой апериодическое звено первого порядка с коэффициентом усиления равным единице и постоянной времени $T=1/\gamma$. Найдем выражение для алгоритма экспоненциального сглаживания, работающего с дискретными значениями сигнала. Запишем дифференциальное уравнение, соответствующее передаточной функции фильтра:

$$\frac{d\hat{x}(t)}{dt} + \gamma\hat{x}(t) = \gamma z(t).$$

Перейдем к разностному уравнению, заменив производную формулой первой конечной разности при интервале дискретизации ΔT :

$$\frac{\hat{x}[n+1] - \hat{x}[n]}{\Delta T} + \gamma\hat{x}[n] = \gamma z[n+1]$$

или $\hat{x}[n+1] = (1 - \Delta T\gamma)\hat{x}[n] + \Delta T\gamma z[n+1]$.

При $\Delta T = 1$ получим следующее выражение:

$$\hat{x}[n+1] = (1 - \gamma)\hat{x}[n] + \gamma \cdot z[n+1].$$

Формула рекуррентная: для получения результата в текущий такт времени используется результат работы фильтра на предыдущем такте. Настраиваемый параметр γ определяет сглаживающие свойства фильтра, или полосу пропускания. Чем меньше γ , тем сильнее сглаживание и уже полоса пропускания фильтра.

Посмотрим, как работает фильтр экспоненциального сглаживания по данным, приведенным в табл. 1. Пусть $\gamma = 0,5$. Для работы алгоритма необходимо задать начальное значение $\hat{x}[0]$, которое вообще-то неиз-

вестно. Поэтому начнем работу фильтра со второй точки, приняв $\hat{x}[1] = z[1]$:

$$\begin{aligned}\hat{x}[2] &= 0,5 \cdot 4 + 0,5 \cdot 4 = 4; & \hat{x}[3] &= 0,5 \cdot 4 + 0,5 \cdot 2 = 3; \\ \hat{x}[4] &= 0,5 \cdot 3 + 0,5 \cdot 6 = 4,5; & \hat{x}[5] &= 0,5 \cdot 4,5 + 0,5 \cdot 7 = 5,75; \\ \hat{x}[6] &= 0,5 \cdot 5,75 + 0,5 \cdot 2 = 3,875.\end{aligned}$$

Робастные фильтры. Такие фильтры совмещают в себе выполнение двух задач:

- сглаживание;
- отсеивание грубых выбросов.

На рис. 18 показано, как будет сглаживаться сигнал, содержащий аномальные выбросы обычным низкочастотным фильтром и робастным фильтром. Низкочастотный фильтр представляет собой динамическое звено, и на грубый выброс он будет реагировать как на импульсный сигнал, соответственно этому выйдет выходной сигнал фильтра.

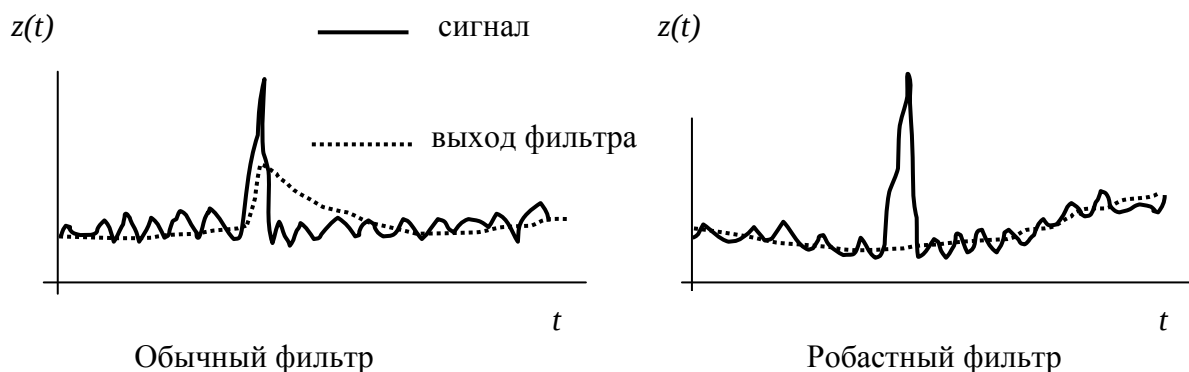


Рис. 18. Пример работы фильтров

Фильтр релейно-экспоненциального сглаживания 1-го порядка. Фильтр является робастным, рекуррентным. Построен на базе фильтра экспоненциального сглаживания с добавлением свойства робастности:

$$\hat{x}[n] = \hat{x}[n-1] + \gamma \Phi\{\Delta \hat{x}[n]\},$$

где $\Delta \hat{x}[n] = z[n] - \hat{x}[n-1]$,

$$\Phi = \Delta \hat{x}[n], \text{ если } |\Delta \hat{x}| \leq \beta,$$

$$\Phi = \beta \cdot \text{sign}(\Delta \hat{x}[n]), \text{ если } |\Delta \hat{x}| > \beta.$$

Настроенные параметры фильтра β и γ , причем γ определяет сглаживающие свойства фильтра, а β – релейные свойства фильтра. Па-

параметр β – это значение скачка сигнала за один такт дискретизации, которое считается достоверным.

Медианный одноступенчатый фильтр. Самый простой из робастных фильтров, основан на определении медианы по k последним измерениям зашумленного сигнала:

$$\hat{x}[n] = \text{med}\{z[n], z[n-1], \dots, z[n-k+1]\}.$$

Обычно значение k выбирается нечетным $k=3, 5, 7$, чтобы не проводить дополнительных операций при определении медианы. Правило вычисления медианы:

- упорядочить значения $z[i]$ по возрастанию;
- если число значений нечетно, медиана равна центральному значению этой последовательности чисел;
- при четном числе значений медиана равна полусумме двух центральных чисел.

Примеры вычисления медианы:

$$\text{med}\{8, 2, 4\}=4; \text{med}\{7, 7, 12, 2, 2\}=7; \text{med}\{5, 1, 1\}=1; \text{med}\{2, 4, 6, 2\}=3.$$

Рассмотрим, как будет работать медианный фильтр по трем точкам, если имеются измерения, приведенные в табл. 2.

Таблица 2

Числовые данные к примеру работы медианного фильтра

№ измер- ний	1	2	3	4	5	6	7	8	9	10	11	12
$z[i]$	1	0	2	100	3	4	6	7	100	100	4	3

$$\hat{x}[3] = \text{med}\{2, 0, 1\} = 1;$$

$$\hat{x}[4] = \text{med}\{100, 2, 0\} = 2;$$

$$\hat{x}[5] = \text{med}\{3, 100, 2\} = 3;$$

$$\hat{x}[6] = \text{med}\{4, 3, 100\} = 4 \text{ и т. д.}$$

Обратим внимание, что значение сигнала, равное 100 на девятом и на десятом шаге, отфильтровано не будет. Чтобы избавиться от выбросов длительностью в два такта, потребуется медианный фильтр по пяти точкам.

Двухступенчатый медианный фильтр. Алгоритм его работы следующий: скользящий интервал разбивается на N подинтервалов, тоже скользящих. На каждом из таких подинтервалов определяется отфильтрованное значение с помощью медианного фильтра первой степени. Полученные значения сглаживаются фильтром второй степени. Рассмотрим работу фильтра на примере данных из табл. 2. Фильтр двух-

ступенчатый, каждая ступень работает по трем точкам. Первое отфильтрованное значение получим только для девятого шага:

$$\hat{x}[9] = \text{med}\{\text{med}[100, 7, 6], \text{med}[4, 3, 100], \text{med}[2, 0, 1]\} = \text{med}\{7, 4, 1\} = 4;$$

$$\hat{x}[11] = \text{med}\{\text{med}[4, 100, 100], \text{med}[7, 6, 4], \text{med}[3, 100, 2]\} = \text{med}\{100, 6, 3\} = 6;$$

$$\hat{x}[12] = \text{med}\{\text{med}[3, 4, 100], \text{med}[100, 7, 6], \text{med}[4, 3, 100]\} = \text{med}\{4, 7, 4\} = 4.$$

Мажоритарный фильтр. Такой фильтр используется в некоторых контроллерах при вводе сигналов. Реализуется схемно. Алгоритм работы следующий:

- проводится троекратный опрос измеряемого сигнала;
- значение каждого разряда в итоговом коде определяется «голосованием» – по большинству нулей или единиц.

Такой фильтр обезопасит систему от единичной случайной ошибки, но при высокочастотных помехах можно получить смещенное значение сигнала. Рассмотрим такой пример. Пусть АЦП имеет восемь разрядов и результаты измерения такие:

1 опрос	1	0	1	1	0	1	1	1
2 опрос	0	1	1	1	1	0	1	0
3 опрос	1	0	1	0	1	1	0	1
Выход фильтра	1	0	1	1	1	1	1	1

Если перевести двоичный код в десятичную систему, то получим следующие значения: первый опрос равен 183, второй равен 122, третий равен 173, выход фильтра равен 191. Видно, что полученное значение существенно отклонилось от среднего.

Вопросы и задания для самопроверки к разделу 10

Какие настроечные параметры у фильтра скользящего среднего и фильтра экспоненциального? 2) Приведите принципиальный вид АЧХ этих фильтров при разных значениях настроечных параметров. 3) Чем отличаются робастные фильтры от обычных? 4) Приведите примеры робастных фильтров. 5) Какой самый простой медианный фильтр надо применить, если выброс держится в течение 2 тактов времени? 6) Одного такта? 7) Трех тактов? 8) К какому типу фильтров по своим частотным свойствам относятся медианные фильтры?

11. Оценка параметров статических моделей

Под оцениванием параметров понимается экспериментальное определение значений параметров модели исследуемого объекта. Пред-

полагается, что структура модели известна. Однако в результате оценивания параметров структура модели может измениться. Например, при получении значения оцениваемого параметра равным нулю. Основные требования к процедуре оценивания:

- оценки должны быть точными, на их качество должны оказывать малое влияние аддитивные помехи и неучтенные нелинейности;
- оценки должны вычисляться быстро, с использованием простых алгоритмов.

Рассмотрим основные свойства оценок и наиболее популярные алгоритмы оценивания.

11.1. Основные свойства оценок параметров

Принято выделять следующие свойства оценок параметров, получаемых на основе статистических данных:

- несмещенность;
- эффективность;
- состоятельность.

Оценка называется несмещенной, если ее среднее значение равно значению оцениваемого параметра. Обозначим истинное значение параметра как θ , а его выборочную оценку на основании статистических данных как $\hat{\theta}$. Тогда для несмещенной оценки

$$M\{\hat{\theta}\} = \theta.$$

Если оценка смещена, то $M\{\hat{\theta}\} \neq \theta$.

Оценка параметра называется эффективной, если обладает минимальной дисперсией $D\{\hat{\theta}\}$ из всех возможных оценок.

Состоятельной оценкой называется такая оценка, которая сходится по вероятности к истинному значению θ , т. е.

$$\text{при } n \rightarrow \infty \quad P\{|\hat{\theta} - \theta| > \varepsilon\} \rightarrow 0,$$

где ε – любое сколь угодно малое число.

11.2. Оценка параметров методом наименьших квадратов (МНК)

Постановка задачи. Задана структура модели статического объекта (см. рис. 19):

$$y_M = b_1 x_1 + b_2 x_2 + \dots + b_k x_k.$$

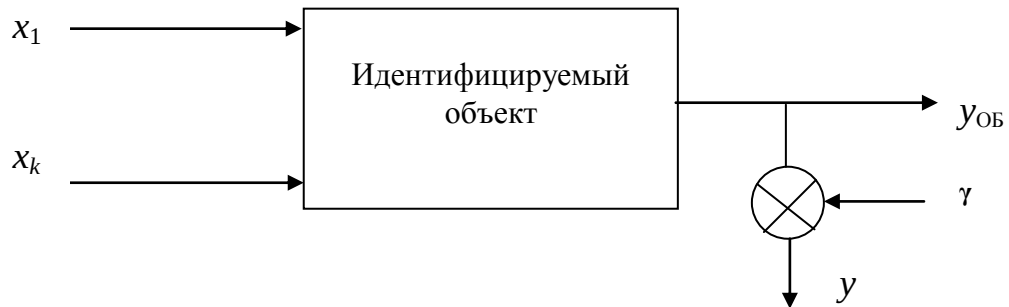


Рис. 19. Структура идентифицируемого объекта

Проведено N измерений входных переменных x_i и выходной переменной y , причем $N > k$, где k – число неизвестных коэффициентов модели. Предположим, что входные переменные измеряются без помех, а выходная переменная может измеряться с помехой. Разность $y - y_M = \varepsilon$ называется невязкой. Невязка включает в себя и ошибку измерения y , и все несоответствие между моделью и реальным объектом.

Чтобы МНК-оценки были несмещенными, эффективными и состоятельными, должны быть выполнены следующие условия:

- ошибки наблюдения за выходной переменной не имеют систематических составляющих;
- дисперсия этих ошибок постоянна и не зависит от номера наблюдения;
- ошибки наблюдения не коррелированы между собой.

МНК по сути своей является методом вычислительной математики и не предъявляет каких-либо требований к закону распределения ошибок.

Итак, необходимо определить коэффициенты модели $b_1, b_2, \dots, b_k$ из условия

$$J = \sum_{i=1}^N \varepsilon_i^2 = \sum_{i=1}^N (y_i - y_{M_i})^2 \Rightarrow \min.$$

Фактически это задача на поиск экстремума функции, зависящей в данном случае от переменных $b_1, b_2, \dots, b_k$. Решение находим из условия

$$\frac{\partial J}{\partial b_1} = 0; \quad \frac{\partial J}{\partial b_2} = 0; \quad \dots; \quad \frac{\partial J}{\partial b_k} = 0.$$

В результате имеем систему k уравнений с k неизвестными, которая решается любым из известных методов.

Пример. Модель задана в виде $y_M = b_0 + b_1 x$, необходимо найти коэффициенты b_0, b_1 по N измеренным значениям y и x . Для этого

случая $J = \sum_{i=1}^N (y_i - b_0 - b_1 x_i)^2$;

$$\frac{\partial J}{\partial b_0} = \sum_{i=1}^N (-2y_i + 2b_0 + 2b_1 x_i) = 0, \quad \frac{\partial J}{\partial b_1} = \sum_{i=1}^N (-2x_i y_i + 2b_0 x_i + 2b_1 x_i^2) = 0,$$

или

$$\begin{cases} N \cdot b_0 + b_1 \sum_{i=1}^N x_i = \sum_{i=1}^N y_i; \\ b_0 \sum_{i=1}^N x_i + b_1 \sum_{i=1}^N x_i^2 = \sum_{i=1}^N x_i y_i. \end{cases}$$

Получили систему двух уравнений с двумя неизвестными. Эту же систему можно записать в векторно-матричной форме:

$$\begin{bmatrix} N & \sum_{i=1}^N x_i \\ \sum_{i=1}^N x_i & \sum_{i=1}^N x_i^2 \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^N y_i \\ \sum_{i=1}^N x_i y_i \end{bmatrix}.$$

Решив систему, найдем неизвестные b_0, b_1 .

МНК в матричной форме. В современной литературе МНК обычно описывают в матричном виде. Дело в том, что МНК в матричной форме сразу дает готовый алгоритм, по которому можно получить значения оцениваемых параметров.

Введем обозначения:

$$Y = \begin{bmatrix} y[1] \\ y[2] \\ \dots \\ y[N] \end{bmatrix}; \quad X = \begin{bmatrix} x_1[1] & x_2[1] & \dots & x_k[1] \\ x_1[2] & x_2[2] & \dots & x_k[2] \\ \dots & \dots & \dots & \dots \\ x_1[N] & x_2[N] & \dots & x_k[N] \end{bmatrix}.$$

В матрице X сосредоточена вся информация о входных сигналах, в векторе Y – все данные об измерениях выходного сигнала. Обозначим вектор коэффициентов модели B , вектор значений выходного сигнала модели Y_M , вектор значений невязок E :

$$B = \begin{bmatrix} b_1 \\ b_2 \\ \dots \\ b_k \end{bmatrix}; \quad Y_M = \begin{bmatrix} y_M[1] \\ y_M[2] \\ \dots \\ y_M[N] \end{bmatrix}; \quad E = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \dots \\ \varepsilon_N \end{bmatrix}, \quad Y_M = X \cdot B.$$

Тогда критерий МНК в матричной форме имеет вид:

$J = E^T E \Rightarrow \min$, или $J = (Y - Y_M)^T (Y - Y_M)$. Требуется так определить вектор B , чтобы значение критерия J было минимальным. Условие, при котором критерий минимален:

$$\frac{\partial J}{\partial B} = 0 \quad \text{или} \quad \frac{\partial J}{\partial B^T} = 0.$$

Найдем выражение для оценки вектора B .

$$J = (Y - Y_M)^T (Y - Y_M) = (Y - XB)^T (Y - XB) = \\ = (Y^T - B^T X^T)(Y - XB) = Y^T Y - Y^T XB - B^T X^T Y + B^T X^T XB;$$

$$\frac{\partial J}{\partial B} = 0 - Y^T X - 0 + B^T X^T X = 0, \quad B^T X^T X = Y^T X \quad \text{или, транспонировав}$$

правую и левую части этого уравнения, получим окончательно:

$$X^T XB = X^T Y.$$

Это векторно-матричное уравнение дает сразу систему уравнений, решая которую и найдем неизвестные коэффициенты. Можно записать выражение для оценки вектора коэффициентов с помощью обратной матрицы:

$$B = (X^T X)^{-1} (X^T Y).$$

Матрица $X^T X$ является квадратной, симметричной, размерностью $k \cdot k$, $X^T Y$ – вектор размерностью k .

Рассмотрим случай, когда с помехой измеряются сигналы на входе объекта. Обозначим матрицу значений помех

$$F = \begin{bmatrix} f_1[1] & f_2[1] & \dots & f_k[1] \\ f_1[2] & f_2[2] & \dots & f_k[2] \\ \dots & \dots & \dots & \dots \\ f_1[N] & f_2[N] & \dots & f_k[N] \end{bmatrix}.$$

Эта матрица по своей структуре полностью совпадает с матрицей значений входных сигналов X . Помехи считаем аддитивными. Запишем выражение для оценки вектора коэффициентов B для этого случая:

$$B = [(X + F)^T (X + F)]^{-1} (X + F)^T Y = \\ = [X^T X + F^T X + X^T F + F^T F]^{-1} [X^T Y + F^T Y].$$

В классическую формулу матричного МНК добавились матрицы $F^T X$, $X^T F$, $F^T F$, $F^T Y$. Рассмотрим, что за информация в них скрыта и какой вклад они могут внести в искажение получаемой модели объекта.

Вектор $F^T Y = \text{cov}(f_j, y)$, $j = 1, 2, \dots, k$. Такая корреляция может возникнуть, если исследуемый объект является управляемым: оператор неверно оценил значение сигнала на входе (из-за помехи), вмешался в управление, в итоге поменялось значение выхода. Ситуация достаточно экзотическая – во-первых, при идентификации исследуют сам объект, без управляющей подсистемы, а во-вторых, управление обычно ведется по отклонению: главное – изменение выходного сигнала. Таким образом, вектор $F^T Y$ значимо не должен отличаться от нуля.

Элементы матриц $F^T X$, $X^T F$ несут в себе одну и ту же информацию – $\text{cov}(x_j, f_i)$. Недиagonalные элементы однозначно равны нулю, они показывают связь между помехами при измерениях одних сигналов и значениями совсем других сигналов. Диагональные элементы показывают корреляционную связь между величиной сигнала и помехой при его измерении. При правильно организованном эксперименте по сбору данных такой корреляции не должно быть.

Рассмотрим матрицу $F^T F$. Её недиагональные элементы показывают корреляцию между помехами на разных входах объекта, что маловероятно. А вот диагональные элементы матрицы – это значения дисперсии помех по каждому из входов и, если помехи есть, то эти значения никогда не будут равны нулю.

Таким образом, даже в условиях идеально проведенного эксперимента имеем отличие от формулы МНК для случая без помех на входе. Полученные значения коэффициентов модели окажутся смещенными. Практический вывод: при идентификации главное внимание уделять точности измерения входных сигналов. Ошибки при измерении выходного сигнала менее критичны, чем ошибки при измерении входа объекта.

Пример оценки коэффициентов модели МНК. Известна структура модели $y = b_0 + b_1 x_1 + b_2 x_2$. Необходимо найти вектор коэффициентов

$$B = \begin{bmatrix} b_0 \\ b_1 \\ b_2 \end{bmatrix}. \text{ Проведены четыре измерения входных и выходной переменных:}$$

ных:

№	x_1	x_2	y
1	1	2	-1
2	2	1	2
3	3	1	3
4	0	2	-2

Сформируем матрицу X , вектор Y . При коэффициенте b_0 поставим фиктивные значения переменной, равные единице:

$$X = \begin{bmatrix} 1 & 1 & 2 \\ 1 & 2 & 1 \\ 1 & 3 & 1 \\ 1 & 0 & 2 \end{bmatrix}, \quad X^T = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 0 \\ 2 & 1 & 1 & 2 \end{bmatrix}, \quad Y = \begin{bmatrix} -1 \\ 2 \\ 3 \\ -2 \end{bmatrix}.$$

$$X^T X = \begin{bmatrix} 4 & 6 & 6 \\ 6 & 14 & 7 \\ 6 & 7 & 10 \end{bmatrix}, \quad X^T Y = \begin{bmatrix} 2 \\ 12 \\ -1 \end{bmatrix}.$$

Система уравнений в матричной форме

$$\begin{bmatrix} 4 & 6 & 6 \\ 6 & 14 & 7 \\ 6 & 7 & 10 \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} 2 \\ 12 \\ -1 \end{bmatrix}.$$

Решим ее методом Крамера:

$$b_0 = \frac{\Delta_1}{\Delta}, \quad b_1 = \frac{\Delta_2}{\Delta}, \quad b_2 = \frac{\Delta_3}{\Delta}.$$

$$\Delta = \begin{vmatrix} 4 & 6 & 6 \\ 6 & 14 & 7 \\ 6 & 7 & 10 \end{vmatrix} = 4, \quad \Delta_1 = \begin{vmatrix} 2 & 6 & 6 \\ 12 & 14 & 7 \\ -1 & 7 & 10 \end{vmatrix} = 8, \quad \Delta_2 = \begin{vmatrix} 4 & 2 & 6 \\ 6 & 12 & 7 \\ 6 & -1 & 10 \end{vmatrix} = 4,$$

$$\Delta_3 = \begin{vmatrix} 4 & 6 & 2 \\ 6 & 14 & 12 \\ 6 & 7 & -1 \end{vmatrix} = -8.$$

$$\text{Ответ: } b_0 = 2, \quad b_1 = 1, \quad b_2 = -2.$$

11.3. Взвешенный метод наименьших квадратов

Одной из модификаций МНК является взвешенный метод наименьших квадратов. В этом случае каждый квадрат невязки, входящий в выражение для критерия МНК, берётся со своим весом:

$$J = \sum_{i=1}^N \varepsilon_i^2 q_i.$$

В матричном виде критерий записывается следующим образом:

$$J = E^T Q E = (Y - XB)^T Q (Y - XB),$$

где Q – диагональная матрица весовых коэффициентов $q_1, q_2, \dots, q_N$.

Найдем выражение для оценки вектора коэффициентов B .

$$J = (Y - XB)^T Q (Y - XB) = Y^T Q Y - Y^T Q X B - B^T X^T Q Y + B^T X^T Q X B.$$

$$\frac{\partial J}{\partial X^T} = -X^T Q Y + X^T Q X B = 0.$$

Отсюда

$$B = (X^T Q X)^{-1} (X^T Q Y).$$

Это и есть выражение для оценки коэффициентов модели взвешенным методом наименьших квадратов. Можно использовать также матричную форму для системы уравнений, решая которую, найдем B :

$$(X^T Q X) B = X^T Q Y.$$

При использовании этого метода оценивания возникает вопрос, как назначать весовые коэффициенты q_j ? Есть два подхода [32]. Первый: связать значения весовых коэффициентов с точностью проведения измерений. Допустим, есть группа измерений, для которых заведомо известно, что они проводились в худших условиях, чем остальные. Исключить их – нарушится непрерывность выборки и уменьшится её объем. Такие измерения используются с меньшим весом, менее единицы. Если известна дисперсия каждого измерения выходного сигнала δ_j^2 , то весовые коэффициенты могут быть равными $1/\delta_j^2$ или k/δ_j^2 . Другой подход – связать значения весовых коэффициентов с номером измерения, т. е. со временем. Если характеристики исследуемого объекта «плывут» во времени, то целесообразно более ранним измерениям присваивать меньший вес. Зависимость весовых коэффициентов от времени в принципе может быть различной – ступенчатой, линейной с разным наклоном, в виде слабой экспоненты.

11.4. Рекуррентный метод наименьших квадратов

Классический МНК при автоматизированной обработке информации обладает недостатком: в момент получения результатов новых измерений необходимо начинать решать всю задачу заново. Этому недостатку лишен рекуррентный МНК. Рассмотрим вначале этот метод на примере очень простой модели: $y = kx$. В результате N измерений была произведена оценка параметра k классическим МНК:

$$k_N = \frac{\sum_{i=1}^N x_i y_i}{\sum_{i=1}^N x_i^2}.$$

Допустим, мы получили результаты еще одного измерения x_{N+1}, y_{N+1} . Получим рекуррентную формулу вида $k_{N+1} = k_N + f(x_{N+1}, y_{N+1})$. Т. е. для получения уточненного значения параметра модели берем старое значение плюс некоторую «добавку», которая зависит от новых измерений. Запишем формулу МНК для $(N+)$ -го измерения:

$$k_{N+1} = \frac{\sum_{i=1}^N x_i y_i + x_{N+1} y_{N+1}}{\sum_{i=1}^N x_i^2 + x_{N+1}^2}.$$

Используем формальный прием: добавим к этой формуле одно и то же выражение со знаком плюс и со знаком минус:

$$\begin{aligned} k_{N+1} &= \frac{\sum_{i=1}^N x_i y_i}{\sum_{i=1}^N x_i^2} + \frac{\sum_{i=1}^N x_i y_i + x_{N+1} y_{N+1}}{\sum_{i=1}^N x_i^2 + x_{N+1}^2} - \frac{\sum_{i=1}^N x_i y_i}{\sum_{i=1}^N x_i^2} = \\ &= k_N + \frac{\sum_{i=1}^N x_i^2 (\sum_{i=1}^N x_i y_i + x_{N+1} y_{N+1}) - \sum_{i=1}^N x_i^2 \cdot \sum_{i=1}^N x_i y_i - \sum_{i=1}^N x_i y_i \cdot x_{N+1}^2}{\sum_{i=1}^N x_i^2 \cdot (\sum_{i=1}^N x_i^2 + x_{N+1}^2)} = \\ &= k_N + \frac{\sum_{i=1}^N x_i^2 \cdot (\sum_{i=1}^N x_i y_i + x_{N+1} y_{N+1} - \sum_{i=1}^N x_i y_i - k_N x_{N+1}^2)}{\sum_{i=1}^N x_i^2 \cdot (\sum_{i=1}^N x_i^2 + x_{N+1}^2)} = \end{aligned}$$

$$=k_N + \frac{x_{N+1}(y_{N+1} - k_N x_{N+1})}{\sum_{i=1}^N x_i^2 + x_{N+1}^2}.$$

Получили требуемую рекуррентную формулу. Обратите внимание, что в числителе стоит разность между измеренным выходным сигналом на $(N+1)$ -м шаге и его расчетным значением, вычисленным по «старой» модели. Если эта разность равна нулю, то модель идеальна и $k_{N+1} = k_N$.

Рассмотрим многомерную модель, вектор коэффициентов определен по N измерениям матричным МНК:

$$B_N = (X^T X)^{-1} (X^T Y).$$

Наступил $(N+1)$ -й такт времени и получены измерения входных сигналов $x_j[N+1]$ и выходного сигнала $y[N+1]$. Найдем рекуррентную формулу для пересчета коэффициентов модели вида $B_{N+1} = B_N$ плюс добавка за счет новой полученной информации.

Обозначим вектор новых значений входных сигналов

$$x^T = \begin{bmatrix} x_1[N+1] \\ x_2[N+1] \\ \dots \\ x_k[N+1] \end{bmatrix}.$$

Сформируем клеточные матрицы $\begin{bmatrix} X^T & x^T \end{bmatrix}$ и $\begin{bmatrix} X \\ x \end{bmatrix}$, а также вектор $\begin{bmatrix} Y \\ y_{N+1} \end{bmatrix}$. Запишем формулу МНК для $(N+1)$ измерений с применением только что введенных обозначений:

$$B_{N+1} = \left\{ \begin{bmatrix} X^T & x^T \end{bmatrix} \begin{bmatrix} X \\ x \end{bmatrix} \right\}^{-1} \begin{bmatrix} X^T & x^T \end{bmatrix} \begin{bmatrix} Y \\ y_{N+1} \end{bmatrix}.$$

По правилу умножения клеточных матриц имеем:

$$B_{N+1} = [X^T X + x^T x]^{-1} [X^T Y + x^T y_{N+1}].$$

Проблема в вычислении обратной матрицы. Однако для обращения матриц такой структуры, как в приведенном выше выражении, существует формула Вутбери [23]:

$$(A + kb^T b)^{-1} = A^{-1} - \frac{A^{-1}b^T b A^{-1}}{bA^{-1}b^T + k^{-1}}.$$

В справедливости этой формулы можно убедиться, умножив ее слева на $(A + kb^T b)$, в итоге получим тождество $E = E$. И ещё: знаменатель $bA^{-1}b^T + k^{-1}$ в этой формуле представляет собой число, т. к. b – вектор-строка, а b^T – вектор-столбец. Запишем выражение для оценки вектора коэффициентов модели, применив формулу Вутбери:

$$\begin{aligned} B_{N+1} &= [(X^T X)^{-1} - \frac{(X^T X)^{-1} x^T x (X^T X)^{-1}}{x(X^T X)^{-1} x^T + 1}] \cdot [X^T Y + x^T y_{N+1}] = \\ &= (X^T X)^{-1} X^T Y - \frac{(X^T X)^{-1} x^T x (X^T X)^{-1}}{x(X^T X)^{-1} x^T + 1} X^T Y + (X^T X)^{-1} x^T y_{N+1} - \\ &\quad - \frac{(X^T X)^{-1} x^T x (X^T X)^{-1}}{x(X^T X)^{-1} x^T + 1} \cdot x^T y_{N+1} = \\ &= B_N + (X^T X)^{-1} x^T [y_{N+1} - \frac{x B_N}{x(X^T X)^{-1} x^T + 1} - \frac{x(X^T X)^{-1} x^T}{x(X^T X)^{-1} x^T + 1} \cdot y_{N+1}] = \\ &= B_N + \frac{(X^T X)^{-1} x^T}{x(X^T X)^{-1} x^T + 1} \cdot [x(X^T X)^{-1} x^T y_{N+1} + y_{N+1} - x B_N - x(X^T X)^{-1} x^T y_{N+1}] = \\ &= B_N + \frac{(X^T X)^{-1} x^T}{x(X^T X)^{-1} x^T + 1} \cdot [y_{N+1} - x B_N]. \end{aligned}$$

Последнее выражение и есть формула рекуррентного метода наименьших квадратов в матричном виде. Если реализовывать этот алгоритм программно, то он будет состоять из следующей последовательности операций:

1. Получение новых данных: $y_{N+1}, x_1[N-1], \dots, x_k[N-1]$.
2. Расчет вектора коэффициентов B_{N+1} .
3. Запись значений B_{N+1} на место B_N .

4. Пересчет по формуле Вутбери обратной матрицы с учетом измерений на $(N+1)$ -м шаге.
5. Запись нового значения обратной матрицы на место прежнего значения.

Таким образом, при реализации алгоритма используется не только собственно формула рекуррентного МНК, но и рекуррентная формула пересчета обратной матрицы.

11.5. Текущий метод наименьших квадратов

Текущий МНК – это обычный МНК на скользящем интервале времени. Применяется для идентификации нестационарных объектов, когда для расчета коэффициентов модели используется не вся информация об объекте, а только собранная за последнее время. Критерий оценивания

$$J = \sum_{i=0}^{n-1} \varepsilon^2[j-i] \Rightarrow \min, \quad j=n, n+1, n+2, \dots$$

Здесь n – длина скользящего интервала.

Выбор длины скользящего интервала и является основной проблемой при применении этого метода оценивания. При малом интервале объем выборки будет недостаточен для получения несмещенной оценки коэффициентов модели. Большое значение n не позволит, возможно, отследить изменение свойств исследуемого объекта.

11.6. Оценка параметров методом наименьших произведений

По сути дела, это еще одна модификация МНК, хотя некоторые считают его более общим методом, чем МНК. Коэффициенты модели находятся из условия

$$J = \sum_{i=1}^{N-m} \varepsilon_i \varepsilon_{i+m} \Rightarrow \min.$$

Видно, что при $m=0$ получаем критерий метода наименьших квадратов. При использовании метода необходимо задать конкретное значение сдвига статистических данных – m . Найдем выражение для оценки коэффициентов модели. Запишем критерий оценивания в матричном виде:

$$J = (Y - XB)^T (Y_m - X_m B) \Rightarrow \min,$$

или

$$J = Y^T Y_m - B^T X^T Y_m - Y^T X_m B + B^T X^T X_m B \Rightarrow \min.$$

Обозначения X, B, Y аналогичны обозначениям, которые использовались в матричном МНК,

$$Y_m = \begin{bmatrix} y_{1+m} \\ y_{2+m} \\ \dots \\ y_N \end{bmatrix}, X_m = \begin{bmatrix} x_1[1+m] & \dots & \dots & x_k[1+m] \\ x_1[2+m] & \dots & \dots & x_k[2+m] \\ \dots & \dots & \dots & \dots \\ x_1[N] & \dots & \dots & x_k[N] \end{bmatrix}.$$

Обратите внимание, что вектора Y и Y_m имеют одинаковую размерность, так же как и матрицы X и X_m .

Найдем вектор B , обеспечивающий минимум критерия:

$$\frac{\partial J}{\partial B^T} = 0 - X^T Y_m - 0 + X^T X_m B = 0.$$

Отсюда система уравнений для оценки вектора коэффициентов модели:

$$X^T X_m B = X^T Y_m.$$

Оценка B с помощью обратной матрицы:

$$B = (X^T X_m)^{-1} X^T Y_m.$$

Рассмотрим возможный эффект от применения метода наименьших произведений [15]. Пусть с помехами измеряются входные сигналы. Ранее мы анализировали все составляющие, которые дополнительно появляются в формуле для оценивания коэффициентов модели при использовании МНК. Итак, при использовании МНК наибольший вклад в искажение результатов вносит корреляционная матрица помех $F^T F$, у которой по диагонали стоят дисперсии помех по каждому из входных каналов. При использовании метода наименьших произведений вместо этой матрицы будет матрица $F_m^T F$.

Теперь по диагонали стоит не дисперсия, а значение автокорреляционной функции помехи при аргументе, равном m . Если это значение будет намного меньше дисперсии, то возможен определенный эффект от применения метода. Но здесь надо иметь в виду, что полезная информация о входных сигналах сосредоточена теперь в корреляционной матрице $X_m^T X$, которая отличается от используемой в МНК матрицы $X^T X$. Это отличие должно быть по возможности меньше, а отличие в корреляционных матрицах помех, наоборот, возможно больше. Именно в этом случае можно получить коэффициенты модели менее смещенные, чем при использовании МНК. Рис. 20 иллюстрирует варианты выбора значения сдвига – удачный и неудачный.

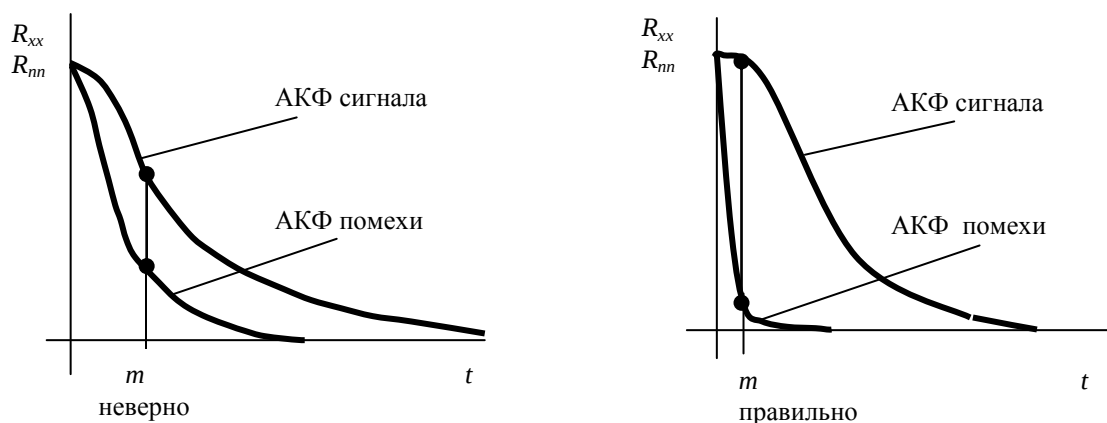


Рис. 20. Определение сдвига по АКФ сигнала и помехи

Рассмотрим пример. Зависимость $y = 2 \cdot x_1 + 3 \cdot x_2$. Зададим значения входных сигналов и вычислим значения на выходе. Имеем следующий набор данных:

№	x_1	x_2	y
1	1	1	5
2	2	1	7
3	1	3	11
4	0	1	3

Внесем искажения в значения входных сигналов и получим такие данные:

№	x_1	x_2	y
1	1,2	0,8	5
2	1,8	1,2	7
3	1,2	2,8	11
4	-0,2	1,2	3

Определим теперь по имеющимся данным коэффициенты модели с помощью метода наименьших произведений. Значение сдвига $m = 1$.

$$X_m = \begin{bmatrix} 1,8 & 1,2 \\ 1,2 & 2,8 \\ -0,2 & 1,2 \end{bmatrix}, \quad X = \begin{bmatrix} 1,2 & 0,8 \\ 1,8 & 1,2 \\ 1,2 & 2,8 \end{bmatrix}, \quad Y = \begin{bmatrix} 7 \\ 11 \\ 3 \end{bmatrix}.$$

$$X^T X_m = \begin{bmatrix} 4,08 & 7,92 \\ 2,32 & 7,68 \end{bmatrix}, \quad X^T Y_m = \begin{bmatrix} 31,8 \\ 27,2 \end{bmatrix}.$$

Имеем систему двух уравнений с двумя неизвестными коэффициентами:

$$\begin{bmatrix} 4,08 & 7,92 \\ 2,32 & 7,68 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} 31,8 \\ 27,2 \end{bmatrix}.$$

Решив эту систему, получим ответ:

$$b_1 = 2,222; \quad b_2 = 2,87.$$

Как видно, коэффициенты отклонились от «истинных», но ненамного. Если применить для определения коэффициентов МНК, взяв матрицу X такой же размерности – три первых измерения, получим

$$b_1 = 1,897; \quad b_2 = 3,11.$$

11.7. Оценка параметров модели методом максимального правдоподобия

Рассмотренные выше методы оценивания относятся к вычислительной математике. Метод максимального правдоподобия родом из несколько иной области знаний – из математической статистики.

В математической статистике при изучении какого-либо явления пользуются такими понятиями, как генеральная совокупность, выборка. При этом генеральная совокупность – это вся мыслимая совокупность объектов, удовлетворяющих определенным условиям. При статистическом исследовании из этой генеральной совокупности извлекают конечное число объектов N , которые образуют выборку. Обычно элементом выборки является результат измерения случайной величины. Если случайная величина является дискретной, т. е. может принимать ограниченное число значений R , то число всевозможных выборок размера N ограничено и равно R^N . На выборочном пространстве может быть задана вероятностная мера, которая определяет вероятность получения той или иной выборки. Такая мера называется функцией правдоподобия.

Рассмотрим случай, когда измеряемые величины являются непрерывными. Получена выборка $(x_1, x_2, \dots, x_N)$, в которой значения, входящие в выборку, представляют собой взаимно независимые случайные величины с одинаковой плотностью вероятности (законом распределения) $f(x)$. Такая выборка представляет собой N -мерную случайную величину, а плотность ее распределения будет функцией правдоподобия

$$L(x_1, x_2, \dots, x_N) = f(x_1) \cdot f(x_2) \cdot \dots \cdot f(x_N).$$

Очевидно, что функция правдоподобия будет зависеть от того, из какого распределения, $f(x)$ или $\varphi(x)$, будет извлечена выборка. В свою очередь, законы распределения задаются с помощью определенных параметров. Например, нормальный закон распределения задается двумя параметрами: математическим ожиданием и дисперсией.

Поэтому функция правдоподобия обычно представляется как условная функция распределения при каком-то значении параметра θ (или параметров) и записывается в виде:

$$L_N(\theta) = f_N(x_1, x_2, \dots, x_N / \theta) = \prod_{i=1}^N f_i(x_i / \theta).$$

Обозначение функции правдоподобия в виде $L_N(\theta)$ показывает, что функция правдоподобия есть функция параметра распределения. Аналогично можно записать функцию правдоподобия, зависящую от нескольких параметров.

А теперь рассмотрим, как использовать функцию правдоподобия для идентификации. В качестве случайных величин, входящих в выражение для функции правдоподобия, будем использовать невязки ε — разность между выходом модели и объекта. Должен быть задан или заранее известен закон распределения $f(\varepsilon)$. Невязка будет определяться тем, насколько верно подобраны коэффициенты модели $b_1, b_2, \dots, b_k$. Таким образом, функция правдоподобия применительно к задаче идентификации принимает такой вид

$$L_N(b_1, \dots, b_k) = f_N(\varepsilon_1, \varepsilon_2, \dots, \varepsilon_k / b_1, \dots, b_k) = \prod_{i=1}^N f(\varepsilon_i / b_1, \dots, b_k).$$

Постановка задачи: необходимо найти коэффициенты модели $b_1, b_2, \dots, b_k$ таким образом, чтобы функция правдоподобия $L_N(b_1, \dots, b_k)$ была максимальной. Это задача на поиск экстремума функции многих переменных. Решение находится из условия:

$$\begin{cases} \frac{\partial L_N(b_1, \dots, b_k)}{\partial b_1} = 0; \\ \dots \\ \frac{\partial L_N(b_1, \dots, b_k)}{\partial b_k} = 0. \end{cases}$$

Для удобства вычислений часто используется логарифмическая функция правдоподобия

$$l_N(b_1, \dots, b_k) = \ln(L_N(b_1, \dots, b_k)).$$

Тогда оценка коэффициентов модели будет определяться исходя из условия

$$\begin{cases} \frac{\partial l_N(b_1, \dots, b_k)}{\partial b_1} = 0; \\ \dots\dots\dots \\ \frac{\partial l_N(b_1, \dots, b_k)}{\partial b_k} = 0. \end{cases}$$

Рассмотрим применение метода максимального правдоподобия для идентификации на примере решения простейшей задачи. Имеется объект управления с одним входом x и выходом y . Задана структура модели $y_M = kx$. Проведено три измерения: x_1, x_2, x_3 , и соответственно y_1, y_2, y_3 . Требуется оценить коэффициент усиления k методом максимального правдоподобия.

Для применения метода необходимо сделать предположение о законе распределения ошибки $\varepsilon = y - y_M$. Обычно считается, что ошибка распределена по нормальному закону с математическим ожиданием m , равным нулю и постоянной для всех измерений дисперсией σ^2 .

Плотность распределения $f(\varepsilon)$ для нормального закона распределения

$$f(\varepsilon) = \frac{1}{\sqrt{2\pi}\sigma} e^{-[(\varepsilon-m)^2 / 2\sigma^2]}.$$

Учтем, что $\varepsilon_i = y_i - kx_i$, $m=0$, и составим функцию правдоподобия для нашего случая

$$\begin{aligned} L(k) &= f(\varepsilon_1) f(\varepsilon_2) f(\varepsilon_3) = \\ &= \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^3 \cdot e^{-[(y_1 - kx_1)^2 / 2\sigma^2]} \cdot e^{-[(y_2 - kx_2)^2 / 2\sigma^2]} \cdot e^{-[(y_3 - kx_3)^2 / 2\sigma^2]}. \end{aligned}$$

Перейдем к логарифмической функции правдоподобия

$$l(k) = \ln(L(k)) = 3\ln\left(\frac{1}{\sqrt{2\pi}\sigma}\right) - \frac{(y_1 - kx_1)^2}{2\sigma^2} - \frac{(y_2 - kx_2)^2}{2\sigma^2} - \frac{(y_3 - kx_3)^2}{2\sigma^2}.$$

Оценку параметра k найдем из условия $\frac{\partial l(k)}{\partial k} = 0$:

$$\frac{\partial l(k)}{\partial k} = \frac{1}{2\sigma^2} (2y_1x_1 - 2kx_1^2 + 2y_2x_2 - 2kx_2^2 + 2y_3x_3 - 2kx_3^2) = 0.$$

$$\text{Отсюда } k = \frac{\sum_{i=1}^3 y_i x_i}{\sum_{i=1}^3 x_i^2}.$$

Если бы дисперсия ошибки в каждом эксперименте была различной и принимала бы значения $\sigma_1, \sigma_2, \sigma_3$, то формула для оценивания k приняла бы такой вид:

$$k = \frac{\sum_{i=1}^3 y_i x_i \sigma_i^2}{\sum_{i=1}^3 x_i^2 \sigma_i^2}.$$

Рассмотрим многомерный случай. Необходимо оценить вектор коэффициентов модели

$$B = \begin{bmatrix} b_1 \\ \dots \\ b_k \end{bmatrix}.$$

Невязки между выходом модели и объекта ε_i распределены по нормальному закону, с нулевым математическим ожиданием и дисперсией в каждой точке эксперимента σ_i^2 . Всего проведено N экспериментов. Введем понятие корреляционной матрицы ошибок P . Эта матрица является диагональной, так как ошибки в разные моменты измерений не связаны между собой:

$$P = \begin{bmatrix} \sigma_1^2 & 0 & 0 & 0 \\ 0 & \sigma_2^2 & 0 & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \sigma_N^2 \end{bmatrix}.$$

Запишем функцию правдоподобия:

$$L(B) = \frac{1}{(2\pi)^{N/2} \sigma_1 \cdot \sigma_2 \dots \sigma_N} e^{-\frac{1}{2} \{ [Y - XB]^T P^{-1} [Y - XB] \}}.$$

Максимум функции $L(B)$ будет достигаться при минимуме:

$$J(B) = [Y - XB]^T P^{-1} [Y - XB].$$

Из условия $\frac{\partial J(B)}{\partial B} = 0$ или $\frac{\partial J(B)}{\partial B^T} = 0$ найдем искомое выражение:

$$B = (X^T P^{-1} X)^{-1} (X^T P^{-1} Y).$$

Сравним полученную оценку с оценкой по методу взвешенных наименьших квадратов. Здесь матрица весовых коэффициентов P^{-1} представляет собой обратную матрицу дисперсий ошибок наблюдения в предположении, что они распределены по нормальному закону. Во взвешенном МНК весовая матрица не связана с законом распределения и могла быть назначена исходя из разных предположений. Отметим, что если дисперсия ошибок наблюдения во всех точках эксперимента одинакова, то формула для оценки коэффициентов модели будет такой же, как для МНК:

$$B = (X^T X)^{-1} (X^T Y).$$

Итак, отличие метода максимального правдоподобия от МНК в том, что задается предпосылка о законе распределения ε_i . В МНК проводится минимизация суммы квадратов невязок независимо от статистических предположений. Одно только это обстоятельство делает метод привлекательным, поскольку, применение МНК, не требует знания статистических методов. С другой стороны, то обстоятельство, что метод всегда дает ответ, может оказаться опасным, потому как ответ может оказаться лишенным смысла.

Вопросы и задания для самопроверки к разделу 11

1) Какие условия надо выполнить, чтобы оценка МНК была несмещенной, эффективной и состоятельной? 2) Влияет ли помеха с нулевым математическим ожиданием и нормальным законом распределения во входных сигналах на смещенность оценок МНК? 3) В чем удобство матричного МНК по сравнению с традиционным? 4) Определите коэффициенты модели $y = a_1 x + a_2 x^2$ с помощью МНК по следующим данным. Измерения x : 1; 2; 0; 2; 1. Соответствующие измерения y : 3; 10; -2; 8; 5. 5) Запишите критерий ВМНК в обычной и в матричной форме. 6) Какие подходы существуют к заданию весовых коэффициентов? 7) Получите формулу оценки параметров по ВМНК в матричной форме. 8) Определите коэффициенты модели $y = a_0 + a_1 x$ с помощью ВМНК по следующим данным. Измерения x : 1; 2; 0; 1. Соответствующие измерения y : 6; 6; 2; 4. Известны дисперсии ошибок измерения y : $D_1 = 1$; $D_2 = 0,5$; $D_3 = 0,25$; $D_4 = 1$.

12. Регрессионный анализ

Среди известных методов прикладной математической статистики регрессионный анализ занимает исключительное положение. Это связано не только с тем, что регрессионный анализ является одним из самых распространенных методов обработки результатов наблюдений, но и с тем, что он, по существу, служит основой целого ряда других разделов математической статистики, и прежде всего планирования экспериментов, дисперсионного анализа, многомерного статистического анализа [9].

12.1. Метод регрессионного анализа

Регрессия – закон изменения условного математического ожидания одной случайной величины в зависимости от значений другой. На рис. 21 показано поле регрессии, т. е. результаты измерения двух переменных x и y и прямая, построенная в соответствии с уравнением регрессии. Видно, что она в среднем описывает результаты эксперимента, но только в среднем! В данном случае уравнение регрессии имеет вид $y = b_0 + b_1x$, где b_0, b_1 – параметры уравнения регрессии.

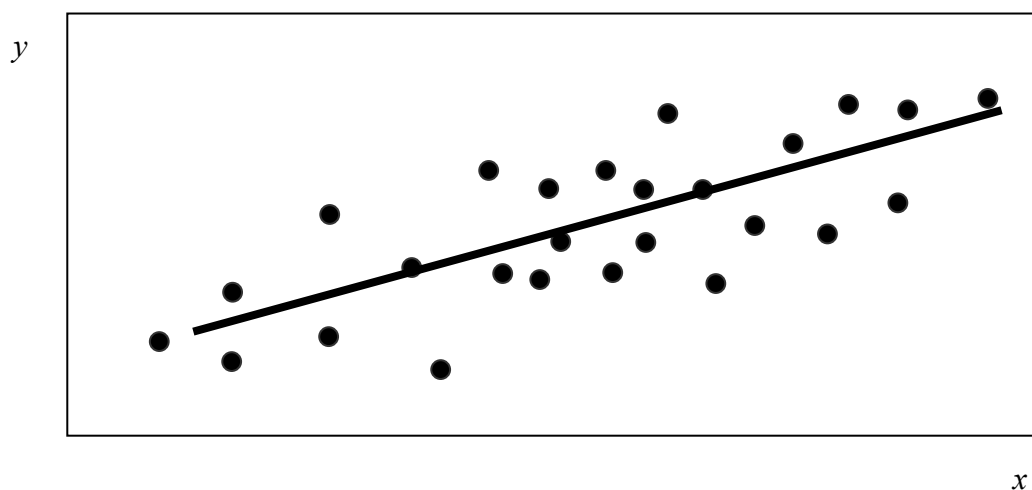


Рис. 21. Поле корреляции и линия регрессии

Математическое исследование оценок функции регрессии и изучение качества этих оценок по данным эксперимента составляет содержание регрессионного анализа.

Если изучаемая случайная величина изменяется под влиянием многих случайных величин, то в этом случае говорят о множественной регрессии. Пример уравнения множественной регрессии, когда случайная величина y зависит от трех других случайных величин:

$$y = b_1x_1 + b_2x_2 + b_3x_3.$$

На практике нередко используются регрессионные уравнения, нелинейные по факторам, т. е. в качестве переменных в уравнении могут присутствовать произведения входных факторов $x_i x_j$, их значения в виде отрезка степенного ряда. В то же время такие регрессионные уравнения являются линейными по параметрам. Путем замены входных переменных можно прийти к обычному виду уравнения множественной регрессии. Например, структура исходного уравнения имеет такой вид:

$$y = b_1x_1 + b_2x_2 + b_{12}x_1x_2 + b_{20}x_2^2.$$

Введем новые факторы: $x_3 = x_1x_2$, $x_4 = x_2^2$. Тогда уравнение множественной регрессии будет иметь обычный линеаризованный вид:

$$y = b_1x_1 + b_2x_2 + b_3x_3 + b_4x_4,$$

$$\text{где } b_3 = b_{12}, \quad b_4 = b_{20}.$$

Существует также понятие авторегрессии. В уравнении авторегрессии исследуемая величина является функцией от своих же предыдущих значений: $y[n] = b_1y[n-1] + b_2y[n-2] + b_3y[n-3]$. Такие уравнения часто используются в задачах прогнозирования.

В дальнейшем будем рассматривать задачи множественной регрессии. Основные предпосылки регрессионного анализа:

Ошибка регрессии $\varepsilon_i = y_i - b_1x_1[i] - b_2x_2[i] - \dots - b_kx_k[i]$ есть случайная величина с нулевым математическим ожиданием и дисперсией (неизвестной) σ^2 .

Ошибки регрессии в разные моменты измерения некоррелированы между собой, так что $\text{cov}(\varepsilon_i, \varepsilon_j) = 0$ при $i \neq j$.

Величина ε_j есть нормально распределенная случайная величина с нулевым математическим ожиданием и дисперсией σ^2 .

Все действия в регрессионном анализе базируются на выполнении этих предпосылок.

Первая задача, которую необходимо решить при регрессионном анализе, – оценка параметров (коэффициентов) регрессионной модели. Для этого используется метод наименьших квадратов:

$$B = (X^T X)^{-1}(X^T Y).$$

А далее разворачивается регрессионный анализ. Собственно, анализ начинается там, где закончился МНК. Для проведения анализа необходимо знать дисперсию ошибки в каждой точке эксперимента σ^2 . Однако она точно неизвестна, т. к. нельзя продублировать условия опыта в каждой точке множество раз. Поэтому довольствуются оценкой этой величины с помощью так называемой остаточной дисперсии $S_{\text{ост}}^2$:

$$\hat{\sigma}^2 = S_{\text{ост}}^2 = \frac{1}{N-k} \sum_{i=1}^N (y_i - b_1 x_1[i] - \dots - b_k x_k[i])^2.$$

То же выражение в матричной форме:

$$S_{\text{ост}}^2 = \frac{1}{N-k} (Y - XB)^T (Y - XB).$$

Здесь X – матрица значений всех входных переменных во всех точках эксперимента. Структура этой матрицы рассмотрена и раскрыта при изучении метода наименьших квадратов.

Одна из первых задач анализа – оценка дисперсии полученных коэффициентов регрессионного уравнения. Сами числовые значения, полученные МНК, ни о чём не говорят. Коэффициент, равный 0,1, со среднеквадратичным отклонением 0,01, заслуживает большего доверия, чем коэффициент, равный 10, но со среднеквадратичным отклонением, равным 5. Введем понятие матрицы дисперсий оценок параметров регрессионной модели $C\{B\}$. Элементы вне главной диагонали этой матрицы представляют собой значения коэффициентов корреляции между коэффициентами модели:

$$C\{B\} = \begin{bmatrix} D\{b_1\} & \text{cov}\{b_1, b_2\} & \dots & \text{cov}\{b_1, b_k\} \\ \text{cov}\{b_2, b_1\} & D\{b_2\} & \dots & \text{cov}\{b_2, b_k\} \\ \dots & \dots & \dots & \dots \\ \text{cov}\{b_k, b_1\} & \text{cov}\{b_k, b_2\} & \dots & D\{b_k\} \end{bmatrix}.$$

Оценка этой матрицы определяется следующим образом:

$$\hat{C}\{B\} = S_{\text{ост}}^2 (X^T X)^{-1}.$$

Можно рассчитать отдельно дисперсию каждого коэффициента регрессионной модели:

$$D\{b_j\} = S_{\text{ост}}^2 \frac{\Delta_{jj}}{\Delta},$$

где Δ – определитель матрицы $X^T X$,

Δ_{jj} – определитель минора, полученного вычеркиванием j -го столбца и j -й строки в матрице $X^T X$.

Корреляция входных переменных между собой увеличивает дисперсию коэффициентов модели. Она может быть причиной неверного знака коэффициента, что приводит к неправильной интерпретации модели.

После того как рассчитаны оценки коэффициентов уравнения регрессии и оценки их дисперсий, необходимо проверить найденные коэффициенты b_j на значимость. Используем методику проверки статистических гипотез.

Нулевая гипотеза H_0 : коэффициент $b_j = 0$.

Альтернативная гипотеза H_1 : коэффициент b_j значимо отличен от нуля.

В качестве статистики при проверке этих гипотез используется t -статистика Стьюдента. Вычисляется значение t -критерия для каждого из коэффициентов модели:

$$t_j = \frac{|b_j|}{\sqrt{D\{b_j\}}}.$$

Если рассчитанное значение критерия при выбранном уровне значимости α и известном числе степеней свободы $N - k$ больше табулированного значения $t_{\text{крит}}$, то принимается гипотеза H_1 , т. е. коэффициент модели считается значимым. В противном случае, если рассчитанное значение критерия меньше табличного, принимается гипотеза H_0 .

Если один или несколько коэффициентов оказались незначимыми, то необходимо исключить из модели факторы при этих коэффициентах и для новой структуры модели пересчитать оставшиеся коэффициенты заново.

Другой подход оценки значимости коэффициентов основан на использовании F -критерия Фишера. F -критерий представляет собой отношение дисперсий. Если последовательно включать в модель новые факторы, можно оценить, значимым ли является при этом уменьшение остаточной дисперсии за счет этого и отсюда делать выводы о включении новых факторов в регрессионную модель.

Пользуясь t -распределением Стьюдента, можно определить доверительные интервалы для каждого из коэффициентов регрессионной модели:

$$b_j = \hat{b}_j \pm t_{\text{крит}} \sqrt{D\{b_j\}}.$$

Значение $t_{\text{крит}}$ берется для доверительной вероятности $p = 1 - \alpha/2$ и для числа степеней свободы $N - m$, где m – число значимых коэффициентов модели.

Одной из ветвей регрессионного анализа является исследование остатков. Остатки ε_j – это разность между наблюдаемым значением y_j и значением, рассчитанным по регрессионному уравнению для всех точек эксперимента. После того как определена регрессионная модель, становятся известными остатки. Имея теперь числовые данные, можно ответить на вопрос: выполнены ли предпосылки регрессионного анализа, перечисленные в начале этого раздела. Если все они выполнены, то мы имеем одно из доказательств, что анализ проведен верно. Если выполнение исходных предпосылок анализа не подтвердилось, то нет доверия полученным результатам. Необходимо вернуться к началу анализа: вводить новые факторы, менять структуру модели, собирать другие статистические данные и т. п.

Процедуры проверки остатков носят графический характер, легко выполнимы и позволяют обычно установить без затруднений, что предположения нарушены. Основные виды графиков следующие:

- общий;

- в зависимости от времени;

- в зависимости от предсказываемых по модели значений $y_{\text{модели}}$;

- в зависимости от независимых переменных x_j .

Рассмотрим эти графики подробнее.

Общий график – это гистограмма, составленная из значений ошибок регрессии. Уже по внешнему виду гистограммы можно определить, соблюдается нормальный закон распределения или нет, является ли нулевым математическое ожидание значений ε_j . Если есть сомнения в нормальности распределения, можно привлечь на помощь процедуру проверки статистических гипотез. Задав определенную доверительную вероятность и применив критерий Пирсона, решить, принадлежат ли остатки нормальному закону распределения.

График временной последовательности. В этом случае остатки выстраиваются по времени, от первого момента измерения до последнего (рис. 22).

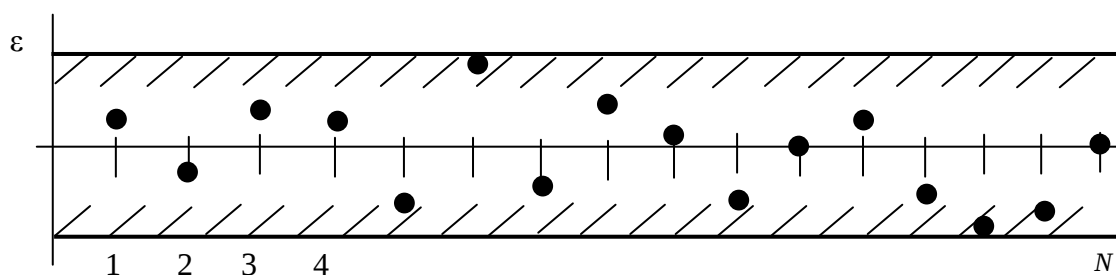


Рис. 22. График остатков в зависимости от времени

Если остатки расположены примерно так, как показано на рисунке, т. е. находятся в определенной зоне и изображения остатков как бы сливаются в горизонтальную полосу, то эффект времени не влияет на данные. Так и должно быть. Если же видно, что полоса, куда входят остатки, зависит от времени, то это говорит о том, что в модель должен войти фактор времени. Еще одна возможная ситуация: ширина полосы меняется во времени. Допустим, при первых значениях разброс был небольшой, а затем увеличивается. То есть дисперсия не постоянна, а растет со временем и потому для оценки параметров регрессионной модели следует применить взвешенный метод наименьших квадратов.

График зависимости остатков от значений $y_{\text{модели}}$. По горизонтальной оси теперь откладываем значения $y_{\text{модели}}[i] = b_1 x_1[i] + \dots + b_k x_k[i]$. Полученное поле точек не должно выглядеть как зависимость остатков от значений выхода модели. Если такая зависимость просматривается, это говорит о неадекватности полученной регрессионной модели. Если видно непостоянство дисперсии, то следует применить взвешенный МНК.

Семейство графиков зависимости остатков от независимых переменных $x_1, \dots, x_k$. Полученное поле точек также не должно напоминать некую функциональную зависимость. Если такая зависимость прослеживается, то это скорее всего говорит о том, что в модель необходимо ввести дополнительные члены от данного фактора x , например квадратичные. При непостоянстве дисперсии рекомендуется пересчитать параметры модели с использованием взвешенного МНК.

Оценка адекватности модели. Распространенным методом оценки адекватности является использование множественного коэффициента корреляции. Рассмотрим этот показатель, его связь с известными уже статистическими характеристиками и его свойства. В дисперсионном

анализе фундаментальную роль играет следующее дисперсионное уравнение:

$$\sum (y_i - m_y)^2 = \sum (y_i - y_{\text{модели}}[i])^2 + \sum (y_{\text{модели}}[i] - m_y)^2.$$

В переводе на словесный язык эта формула говорит о следующем: (сумма квадратов относительно среднего) равна (сумма квадратов относительно регрессии) плюс (сумма квадратов, обусловленная регрессией). Сумма квадратов относительно среднего известна нам по эксперименту, обозначим ее S_y^2 . Сумма квадратов относительно регрессии по-другому называется остаточной дисперсией $S_{\text{ост}}^2$. Чем меньше эта величина, тем лучше модель. Последняя составляющая S_p^2 показывает, какая часть дисперсии объясняется регрессией.

Перепишем дисперсионное уравнение:

$$S_y^2 = S_{\text{ост}}^2 + S_p^2.$$

Обозначим коэффициент множественной корреляции R . Его квадрат R^2 называется коэффициентом детерминации и показывает, насколько неслучайной является зависимость между выходной переменной и входными факторами регрессионной модели:

$$R^2 = \frac{S_p^2}{S_y^2}.$$

Выразим R^2 через остаточную дисперсию и дисперсию выходной переменной:

$$R^2 = 1 - \frac{S_{\text{ост}}^2}{S_y^2} \quad \text{или} \quad R = \sqrt{1 - \frac{S_{\text{ост}}^2}{S_y^2}},$$

где
$$S_y^2 = \frac{1}{N-1} \sum_{i=1}^N (y_i - m_y)^2.$$

Основные свойства коэффициента множественной корреляции:

1. $0 \leq R \leq 1$.
2. Минимальное значение множественного коэффициента корреляции $R=0$ соответствует случаю полного отсутствия корреляционной связи между выходной переменной y и входными факторами. При этом усредненная дисперсия «регрессионных остатков» в точности равна общей вариации выходного показателя.
3. Максимальное значение множественного коэффициента корреляции $R = 1$ соответствует полному отсутствию варьирования «регрессионных остатков», т.е. $\varepsilon_i = 0$. Это означает наличие чисто функциональной связи между выходной переменной и входными факторами.

Таким образом, значение множественного коэффициента корреляции позволяет судить, насколько полно полученная регрессионная модель описывает имеющиеся статистические данные.

Если существует возможность повторения измерений при каждом фиксированном значении аргумента, а такое возможно при использовании методов планирования эксперимента, для оценки адекватности модели применяют F -критерий.

Если заранее известна величина дисперсии σ^2 остаточной случайной компоненты ε , например исходя из характеристик точности используемых измерительных приборов, то для оценки адекватности также может быть использован критерий, равный отношению остаточной дисперсии к известной σ^2 .

Последним этапом регрессионного анализа является проверка работоспособности модели. При этом мы должны сравнить, насколько хорошо полученная модель аппроксимирует измеренные значения выходной переменной исследуемого объекта. Для этого обязательно используются другие статистические данные – не те, на которых было получено уравнение регрессии. Выборка, на которой были рассчитаны коэффициенты модели, носит название обучающей. Работоспособность модели оценивается по экзаменующей (или контрольной) выборке. Эти две выборки не должны пересекаться. Рекомендуется в ходе эксперимента сразу набрать достаточное количество статистических данных, чтобы еще до начала анализа поделить их на обучающую и контрольную выборки. В качестве критерия работоспособности используется любой известный критерий качества аппроксимации – среднеквадратичная ошибка (чаще всего), средняя ошибка, максимальная ошибка. Рекомендуется также исследовать остатки на контрольной выборке.

Может быть ситуация, что после всех действий по построению регрессионной модели, она окажется неработоспособной. Обсудим возможные причины этого.

1. Неправильно задана структура модели. Возможно, в модель должны входить такие факторы, которые даже не рассматривались при анализе. Или факторы, входящие в модель линейно, должны быть заданы в виде нелинейных функций.

2. Наблюдения имеют узкий диапазон изменений, что приводит к незначимости оценок соответствующих коэффициентов модели. Малый интервал варьирования значений параметров вообще характерен для пассивного эксперимента, и это один из его недостатков.

3. Погрешности измерений входных факторов, некачественная предварительная обработка информации.

4. При исследовании в первую очередь технических и технологических объектов не учитывается инерционность систем. Значение сигнала на выходе в таких системах определяется не только входными воздействиями, но и предыдущим состоянием объекта.

5. Входные переменные коррелированы с ошибками измерения выходной величины, что имеет место при наличии обратной связи на идентифицируемом объекте. Обратная связь может осуществляться не только техническим устройством, но и человеком. Может возникнуть такая цепочка событий: неправильная оценка состояния объекта из-за ошибки измерения y , принятие решения об управлении, изменение входной управляющей переменной.

12.2. Трансцендентная регрессия

В некоторых прикладных задачах имеет место существенная нелинейная зависимость между выходными и входными переменными. Можно аппроксимировать такую зависимость в виде отрезка степенного ряда, при этом надо будет находить значительное число коэффициентов. Иногда удобно использовать трансцендентные уравнения вида

$$y = b_0 b_1^x \quad \text{или} \quad y = b_0 x^{b_1}.$$

Постановка задачи: проведены измерения y, x . Необходимо оценить коэффициенты b_0, b_1 . Проведем замену переменных, прологарифмировав предварительно трансцендентные уравнения.

$$\ln y = \ln b_0 + x \ln b_1;$$

$$\ln y = \ln b_0 + b_1 \ln x.$$

Обозначим

$$\ln y = z, \quad \ln b_0 = a_0, \quad \ln b_1 = a_1, \quad \ln x = t.$$

Получим линейные уравнения относительно новых переменных z, t :

$$z = a_0 + a_1 x, \quad z = a_0 + b_1 t.$$

Коэффициенты a_0, a_1, b_1 находятся методом наименьших квадратов. По значениям a_0, a_1 затем восстанавливаются коэффициенты b_0, b_1 . Отметим, что полученные таким образом коэффициенты модели будут смещенными.

12.3. Получение уравнения множественной регрессии методом Брандона

Перед проведением регрессионного анализа проводится предварительное исследование факторов $x_1, x_2, \dots, x_k$ по степени влияния на выходной показатель y . Могут использоваться априорные сведения об объекте или результаты корреляционного анализа. Допустим, фактор x_1 оказывает наибольшее влияние, следующим идет фактор x_2 и т. д.

Регрессионное уравнение по методу Брандона записывается в виде произведения

$$y = af_1(x_1) \cdot f_2(x_2) \cdot \dots \cdot f_k(x_k),$$

где $f_j(x_j)$ – любая функция, тип её определяется по виду поля корреляции.

Рассмотрим пример. В регрессионное уравнение входит два фактора:

$$y = af_1(x_1) \cdot f_2(x_2).$$

Построим поле корреляции между переменной y и переменной x_1 (рис. 23). По расположению точек принимаем гипотезу о линейной связи между переменными:

$$y = a_1 + b_1x_1.$$

По статистическим данным находим коэффициенты a_1, b_1 методом наименьших квадратов.

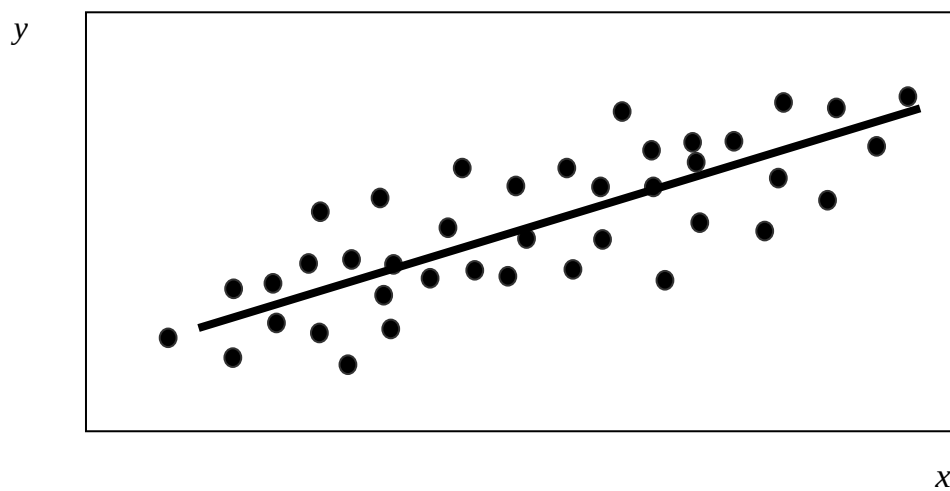


Рис. 23. Поле корреляции между переменными y и x_1

Зная эти коэффициенты, определим вспомогательную переменную y_1 :

$$y_1 = \frac{y}{a_1 + b_1 x_1} = a \cdot f_2(x_2).$$

Строим поле корреляции между переменными y_1 и x_2 . Допустим, точки расположены так, как показано на рис. 24.

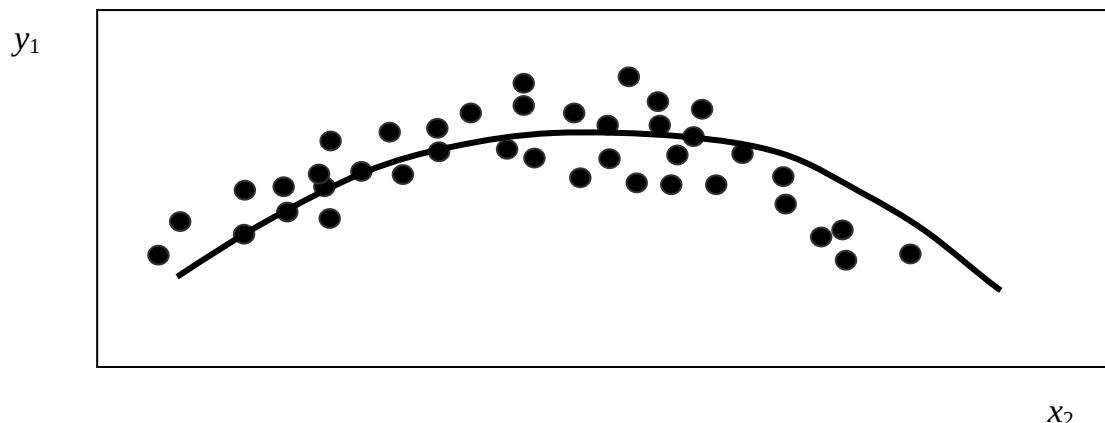


Рис. 24. Поле корреляции между переменными y_1 и x_2

Видно, что в этом случае связь между переменными скорее всего описывается отрезком параболы

$$y = a_2 + b_2 x_2 + c_2 x_2^2.$$

Определим коэффициенты a_2 , b_2 , c_2 методом наименьших квадратов. Затем находим значения вспомогательной переменной

$$y_2 = \frac{y_1}{a_2 + b_2 x_2 + c_2 x_2^2}, \text{ и по этим значениям вычисляем коэффици-}$$

ент a :

$$a = \frac{1}{N} \sum_{i=1}^N y_2[i],$$

N – объем выборки.

В итоге получаем модель вида

$$y = a(a_1 + b_1 x_1)(a_2 + b_2 x_2 + c_2 x_2^2).$$

Перемножив скобки, получим обычное регрессионное уравнение, нелинейное по параметрам, к которому можно применить изложенные выше методы регрессионного анализа.

Вопросы и задания для самопроверки к разделу 12

1) Перечислите по порядку действия при выполнении регрессионного анализа. 2) Можно ли с помощью регрессионного анализа получить модель $y = a_1x + a_2x^2$? 3) Какая статистическая гипотеза проверяется с помощью критерия Стьюдента? 4) Какая информация должна быть задана для получения критического (табличного) значения этого критерия? 5) Для чего применяется и в каких пределах может изменяться коэффициент множественной корреляции? 6) Каким требованиям должны удовлетворять остатки при регрессионном анализе? 7) По каким причинам модель может оказаться не адекватной исследуемому объекту?

13. Планирование эксперимента

Алгоритмы планирования эксперимента применяются для построения математических моделей управления статических объектов. Их использование требует проведения активного эксперимента на исследуемом объекте. Планирование эксперимента широко применяется для получения моделей в области биологии, медицины, химии, физике, металлургии [6, 5].

Теория планирования эксперимента базируется на методах математической статистики, которые используются для обработки статистических данных, полученных в результате эксперимента. При планировании эксперимента входы исследуемого объекта варьируются таким образом, чтобы извлечь наибольшую информацию из измеренных значений. Входные переменные в теории планирования эксперимента называются факторами. Если факторы подобраны верно, то модель получается адекватной объекту и, кроме того, описывает его поведение на значительном интервале изменения входных переменных. Возможность получения качественных моделей – главная причина применения планирования эксперимента при идентификации. Недостаток этого метода в том, что не во всех случаях он применим, есть сколько угодно объектов, для которых нельзя самостоятельно устанавливать значения входных переменных и повторять эксперименты при одинаковых условиях. Для таких объектов приходится довольствоваться идентификацией с применением методов пассивного эксперимента.

В отличие от традиционного подхода к эксперименту, в планировании эксперимента используются два новых положения: положение о рандомизации и положение о многофакторности.

Положение о рандомизации. Рандомизация – введение случая в эксперимент. Изучению любого явления обычно мешает целый ряд воз-

мущающих факторов. Можно пытаться исключать эти факторы, но все-таки полностью убрать их не удастся. В планировании эксперимента не требуется исключать мешающие факторы, а, наоборот, вводить их в эксперимент, сделав случайными. Таким образом, мешающие факторы создают как бы шумовой фон эксперимента. Если же мешающий фактор вносит систематическую ошибку, то следует признать его существенным для исследуемого объекта и ввести его дополнительно в модель.

Положение о многофакторности. На выходной показатель действует ряд входных факторов. Можно выбрать две стратегии при проведении опытов. Первая состоит в том, чтобы изменять один фактор, стабилизируя остальные. Затем меняем другой фактор, обеспечив неизменность всех остальных. В теории планирования эксперимента используется другая стратегия: варьируются одновременно все факторы.

Рассмотрим применение этих двух стратегий на примере взвешивания на весах трех предметов: A , B , C . Классическая стратегия: выверяем весы и по очереди взвешиваем предметы. По второй стратегии ставим на весы сразу все предметы. Ход экспериментов иллюстрирует таблица. Присутствие предмета на весах обозначено знаком «плюс», отсутствие – знаком «минус».

Таблица 3

Эксперимент со взвешиванием

№ Опы-та	Первая стратегия				Вторая стратегия			
	A	B	C	Показания весов	A	B	C	Показания весов
1	-	-	-	Y_0	+	+	+	Y_1
2	+	-	-	Y_1	-	-	+	Y_2
3	-	+	-	Y_2	-	+	-	Y_3
4	-	-	+	Y_3	+	-	-	Y_4

Определим вес предмета A и погрешность измерения по первой стратегии. $A = y_1 - y_0$. Пусть среднеквадратичная погрешность одного измерения равна σ . Тогда дисперсия ошибки при взвешивании A равна:

$$D\{A\} = D\{y_1 - y_0\} = 2\sigma^2$$

Вес тела A по второй стратегии определится в результате вычисления по формуле:

$$A = \frac{y_1 + y_4 - y_2 - y_3}{2}.$$

Оценим дисперсию ошибки:

$$D\{A\} = D\left\{\frac{y_1 + y_4 - y_2 - y_3}{2}\right\} = \frac{1}{4} D\{y_1 + y_4 - y_2 - y_3\} = \frac{4\sigma^2}{4} = \sigma^2.$$

Таким образом, во втором случае точность измерения оказалась в два раза выше.

Рассмотрим общие требования при использовании планирования эксперимента. Должна быть задана структура модели:

$$y = b_0 + b_1 x_1 + \dots + b_k x_k.$$

Выходная величина и входные переменные – факторы $x_1, \dots, x_k$ должны быть доступны измерению. Заданному набору значений входных факторов должно соответствовать определенное значение y (с точностью до ошибки измерения). При проведении эксперимента входные факторы должны быть управляемы, т. е. устанавливаться на заданный уровень и быть неизменными во время опытов.

Модель, которая должна быть найдена, описывает исследуемый объект в заданной области факторного пространства. Выбору структуры модели предшествует тщательный анализ априорной информации. По каждому фактору выбирается основной уровень. Варьирование фактора будет идти в большую и меньшую сторону от основного уровня. Основной уровень соответствует значению функционирования объекта в нормальном режиме.

Рассмотрим самый простой случай планирования эксперимента – планирование на двух уровнях. В этом случае входные переменные будут принимать значения, симметричные относительно основного уровня. Для каждого фактора должен быть определен интервал варьирования.

Пример. Пусть одним из факторов x_j является температура в технологическом аппарате. За основной уровень принято значение $x_{j0} = 50$ °С. Интервал варьирования по этому фактору $a_j = 10$ °С, тогда нижнее значение фактора будет 40 °С, верхнее значение фактора равно 60 °С.

Для упрощения обработки экспериментальных данных используют кодированные значения факторов. При этом нижнее значение фактора кодируется как «-1», а верхнее значение как «+1». Переход от натуральных значений x_{jn} к кодированным значениям x_j осуществляется по формуле

$$x_j = \frac{x_{jn} - x_{j0}}{a_j}.$$

Для рассмотренного выше примера верхний уровень равен:

$$x_j = \frac{60 - 50}{10} = +1, \text{ нижний уровень фактора } x_j = \frac{40 - 50}{10} = -1.$$

Эксперимент, в котором реализуются все возможные сочетания уровней факторов, называется полным факторным экспериментом (ПФЭ). При изменении факторов на двух уровнях план эксперимента называется планом типа 2^K , K – число факторов.

Число опытов для ПФЭ типа 2^K равно: $N = 2^K$. Матрица планирования, или план эксперимента, задает координаты точек, в которых ставятся опыты, т. е. определяет область экспериментирования. Матрица планирования записывается для кодированных значений факторов.

Рассмотрим 2 факторную задачу. $N=2^2=4$. Матрица планирования приведена в табл. 4.

Таблица 4
Матрица планирования для 2 факторов

Номер опыта	X_1	X_2
1	-1	-1
2	+1	-1
3	-1	+1
4	+1	+1

План эксперимента для двух факторов можно представить графически. Каждая комбинация факторов представляет собой вершину прямоугольника с координатами (в кодированных значениях): $(-1, -1)$, $(-1, +1)$, $(+1, -1)$, $(+1, +1)$.

Приведенная матрица планирования представляет собой матрицу исходных данных по входным переменным. При изучении МНК и регрессионного анализа она обозначалась как матрица X .

Матрица планирования ПФЭ на 2 уровнях обладает следующими свойствами:

1. Свойство симметричности плана: $\sum_{i=1}^N x_{ij} = 0$ для $j=1, \dots, k$;

k – число факторов.

2. Свойство нормировки плана: $\sum_{i=1}^N x_{ij}^2 = N$.

3. Свойство ортогональности матрицы X : $\sum_{i=1}^N x_{ij}x_{il} = 0$ при $j \neq l$.

4. Свойство ротатабельности: дисперсия выходной величины y одинакова для всех точек факторного пространства, лежащих на одинаковых расстояниях от центра.

В примере с двумя факторами план ПФЭ включает в себя 4 опыта. А неизвестных коэффициентов – два. Можно усложнить модель, дополнив матрицу планирования до 4 столбцов. Включим в модель постоянный коэффициент b_0 и нелинейность, связанную с эффектом взаимодействия двух факторов x_1 и x_2 . Уравнение модели примет вид

$$y = b_0 + b_1x_1 + b_2x_2 + b_{12}x_1x_2.$$

Матрица планирования примет такой вид:

№	X_0	X_1	X_2	X_1X_2
1	+1	−1	−1	+1
2	+1	+1	−1	−1
3	+1	−1	+1	−1
4	+1	+1	+1	+1

Появились два новых столбца. Значение переменной при коэффициенте b_0 всегда равно +1, значения произведения факторов в четвертом столбце получаются произведением кодированных значений каждого фактора для соответствующего опыта. Новая матрица также удовлетворяет перечисленным выше свойствам, кроме свойства симметричности плана для первого столбца фиктивных значений X_0 .

Вычисление коэффициентов модели проводится методом наименьших квадратов. Вектор коэффициентов модели B можно определить по известной формуле:

$$B = (X^T X)^{-1} X^T Y.$$

Обратная матрица с учетом свойств матрицы планирования:

$$X^T X = \begin{bmatrix} N & 0 & \dots & 0 \\ 0 & N & \dots & 0 \\ \dots & \dots & \dots & 0 \\ 0 & 0 & \dots & N \end{bmatrix}, \quad (X^T X)^{-1} = \begin{bmatrix} 1/N & 0 & \dots & 0 \\ 0 & 1/N & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1/N \end{bmatrix}.$$

$$\text{Вектор-столбец } X^T Y = \begin{bmatrix} \sum_{i=1}^N x_{i0} y_i \\ \dots \\ \sum_{i=1}^N x_{i1} x_{i2} y_i \end{bmatrix}.$$

Таким образом, для расчета коэффициентов модели можно применить простые формулы:

$$b_0 = \frac{1}{N} \sum_{i=1}^N y_i, \quad b_j = \frac{1}{N} \sum_{i=1}^N x_{ij} y_i, \quad b_{jl} = \frac{1}{N} \sum_{i=1}^N x_{ij} x_{il} y_i, \quad j \neq l.$$

Выясним, можно ли включить в модель квадрат переменной – значения x_j^2 . Столбец в матрице планирования будет состоять из кодированных значений, равных +1. Таким образом, он будет неотличим от фиктивных значений x_0 и значение коэффициента модели при квадрате переменной совпадет со значением коэффициента b_0 . Если попытаемся ввести в модель кубические значения переменных, то соответствующие столбцы матрицы планирования совпадут со столбцами, в которых стоят значения самих переменных. Таким образом, модель, получаемая с помощью ПФЭ, может включать в себя нулевой коэффициент, входные факторы, взаимодействия факторов (их произведения).

Полученные оценки коэффициентов модели являются оценками чувствительности выходной величины модели к вариации каждой переменной. Вычисленные оценки в силу ортогональности матрицы планирования являются не зависимыми друг от друга. Это позволяет исключить из модели незначимый коэффициент без пересчета других коэффициентов. Сравните, в регрессионном анализе было не так!

Понятие взаимодействия факторов можно интерпретировать следующим образом: если одинаковому приращению первого фактора соответствуют разные приращения y , в зависимости от того, на каком уровне находился второй фактор, то эти два фактора взаимодействуют между собой.

После того как вычислены оценки коэффициентов модели, необходимо воспользоваться методами регрессионного анализа для решения следующих задач:

- а) отобрать значимые коэффициенты модели;
- б) оценить пригодность модели для описания экспериментальных данных.

Выполнение этих этапов возможно на основе вычисленной ошибки эксперимента. Оценку дисперсии ошибки эксперимента получают дублированием опытов в точках факторного пространства. При этом порядок проведения опытов в каждой серии повторов подбирается с учетом положения о рандомизации, т. е. порядок проведения опытов при каждом повторении определяется случайным образом. Рекомендуется для этого использовать генератор случайных чисел. Допустим, каждый эксперимент из N опытов повторялся m раз. Всего, таким образом, было проведено $(N \cdot m)$ измерений. Результаты эксперимента можно представить в виде таблицы. Для примера взят ПФЭ типа 2^3 , взаимодействия факторов отсутствуют.

Таблица 5

Статистические данные по ПФЭ 2^3

№	Матрица планирования				Выход	Среднее $\bar{y}_i$	Дисперсия S^2_i
	X_0	X_1	X_2	X_3	$y_{i1} \ y_{i2} \dots \dots \dots y_{im}$		
1	+1	-1	-1	-1	$y_{11} \ y_{12} \dots \dots \dots y_{1m}$	$\bar{y}_1$	S^2_1
2	+1	-1	-1	+1	$y_{21} \ y_{22} \dots \dots \dots y_{2m}$	$\bar{y}_2$	S^2_2
...	...	...	...	...	...	...	...
8	+1	+1	+1	+1	$y_{81} \ y_{82} \dots \dots \dots y_{8m}$	$\bar{y}_3$	S^2_8

Для корректного использования методов регрессионного анализа необходимо убедиться, что каждый из восьми экспериментов проходил в одинаковых условиях. Для этого необходимо проверить предпосылку об однородности дисперсий в каждой из строк таблицы. Оценка однородности проводится с помощью критерия Кохрана:

$$G = \frac{S_{i\max}^2}{\sum_{i=1}^N S_i^2}.$$

Если рассчитанное значение G меньше табличного значения $G_{\text{табл}}$, взятого для заданного уровня значимости и числа степеней свободы $(m - 1, N - 1)$, то дисперсии опытов можно считать однородными.

Далее рассчитываются коэффициенты модели по формулам, приведенным выше. Проверка значимости коэффициентов проводится по t -критерию Стьюдента, как и в обычном регрессионном анализе, но дисперсия коэффициентов модели вычисляется по-другому, с учетом дублирования опытов в точках факторного пространства. Итак, сформулируем нулевую гипотезу $H_0 : b_j = 0$. Альтернативная гипотеза $H_1 : b_j \neq 0$.

Дисперсия каждого из коэффициентов модели в силу свойств матрицы планирования является одинаковой и определяется следующим образом:

$$D\{b_j\} = \frac{S_e^2}{N},$$

S_e^2 – дисперсия воспроизводимости, которая определяется по формуле

$$S_e^2 = \frac{1}{N(m-1)} \sum_{i=1}^N \sum_{k=1}^m (y_{ki} - \bar{y}_i)^2.$$

Если расчетное значение t -критерия $t = \frac{|b_j|}{\sqrt{D\{b_j\}}} \leq t_{\text{крит}}$, то справед-

лива нулевая гипотеза. Табличные значения критерия берутся для выбранного уровня значимости и числа степеней свободы $N(m - 1)$.

Для модели со значимыми коэффициентами должна быть проверена гипотеза о пригодности этой модели для описания экспериментальных данных. Для этого используется критерий Фишера:

$$F = \frac{S_{\text{ост}}^2}{S_e^2}.$$

Остаточная дисперсия $S_{\text{ост}}^2$ вычисляется с использованием средних значений $\bar{y}_i$, рассчитанных по итогам повторов каждого i -го опыта:

$$S_{\text{ост}}^2 = \frac{1}{N - d} \sum_{i=1}^N (\bar{y}_i - y_{\text{модели}}[i])^2,$$

где d – число значимых коэффициентов модели.

Если расчетное значение критерия Фишера меньше табличного значения, то модель считается пригодной. Табличное критическое зна-

чение F -критерия берется для выбранного уровня значимости и числа степеней свободы $N(m - 1)$.

Для того чтобы пользоваться полученной моделью, подставляя в нее значения входных факторов в натуральных единицах измерения, необходимо ее преобразовать, выразив входные кодированные переменные через натуральные переменные. При этом, естественно, изменятся коэффициенты модели.

Понятие о дробном факторном эксперименте (ДФЭ). В том случае, когда эффект взаимодействия факторов отсутствует, можно в матрице планирования вместо произведения факторов вставить новый фактор. За счет этого можно уменьшить число экспериментов.

Пример. Составлен план ПФЭ типа 2^2 . Выяснилось, что взаимодействия факторов нет, т. е. составляющая x_1x_2 в модели отсутствует. На место столбца x_1x_2 в матрицу планирования помещен новый фактор x_3 . Таким образом, по результатам 4 экспериментов получена модель

$$y = b_0 + b_1x_1 + b_2x_2 + b_3x_3.$$

Если бы сразу ставилась задача ПФЭ для трех входных факторов, число опытов было бы равно $2^3 = 8$, т.е. в два раза больше. Планы экспериментов в ДФЭ принято называть репликами от ПФЭ. В нашем примере имеем дело с полуреplikой.

Следует учесть, что ДФЭ приводит к смешанным оценкам. Оценка коэффициента b_1 смешана со взаимодействием x_2x_3 , оценка b_2 – со взаимодействием x_1x_3 , оценка b_3 – со взаимодействием x_1x_2 .

Вопросы и задания для самопроверки к разделу 13

1) На каких принципах базируется методика ПФЭ? 2) Дайте понятие основного уровня, интервала варьирования. 3) Как и зачем кодируются значения факторов? 4) Приведите пример полной матрицы планирования для модели $y = a_1x + a_2x_2 + a_{12}x_1x_2$. 5) Перечислите свойства матрицы планирования. 6) Сколько опытов должно быть проведено для ПФЭ на 2 уровнях при числе факторов, равном 4? 7) Можно ли с помощью ПФЭ идентифицировать модель $y = a_0 + a_1x + a_2x_2 + a_3x_1^2$? (Ответ должен быть доказательным). 8) Перечислите по порядку все действия при ПФЭ. 9) С помощью какого метода оценивания рассчитываются коэффициенты модели? 10) В каком порядке проводятся опыты? 11) Почему в регрессионном анализе при обнаружении незначимых коэффициентов оставшиеся коэффициенты пересчитываются, а в ПФЭ – не пересчитываются? 12) Какую самую сложную модель можно получить в результате ПФЭ типа 2^3 ? 13) Какая статистическая гипотеза проверяется с помощью критерия Фишера? 14) Какая информация должна быть задана для получения критического (табличного) значения этого критерия?

14. Уравнение Винера–Хопфа

Уравнение Винера–Хопфа [3, 8, 10, 19, 20, 32] играет большую роль в идентификации динамических объектов. В литературе его называют также «основным уравнением идентификации» или уравнением статистической идентификации. Это уравнение, полученное, кстати говоря, Винером при решении задач не идентификации, а оптимальной фильтрации, связывает модель динамического объекта и статистические характеристики входных и выходных сигналов.

Рассмотрим одномерный линейный стационарный динамический объект. Зависимость между его входом $x(t)$ и выходом $y(t)$ может быть задана в виде известного интегрального уравнения:

$$y(t) = \int_0^{\infty} g(\tau)x(t-\tau)d\tau,$$

где $x(t) = 0$ при $t < 0$,

$g(t)$ – импульсная переходная функция объекта.

Для динамических систем импульсная переходная функция с течением времени стремится к нулю и после завершения переходного процесса становится практически нулевой. Поэтому приведенное выше интегральное уравнение можно записать в виде

$$y(t) = \int_0^T g(\tau)x(t-\tau)d\tau,$$

где T – время переходного процесса.

Задачей идентификации является определение импульсной переходной функции $g(t)$ по известным входному и выходному сигналам.

Если учесть ненулевые начальные условия и рассматривать пассивный эксперимент, то функции $x(t)$ и $y(t)$ являются случайными сигналами, к тому же измеренными на относительно небольшом временном интервале. Такой информации будет явно недостаточно для определения модели объекта. Другое дело, если на интервале времени $[0, T]$ сигнал $x(t)$ будет меняться не случайно, а целенаправленно. Тогда и выходной сигнал будет изменяться существенно и интегральное уравнение свертки может быть использовано для идентификации. Но это будет уже активный эксперимент.

Преобразуем интегральное уравнение свертки, чтобы в него входили не сами сигналы, а их статистические характеристики.

Запишем выражение для взаимной корреляционной функции стационарного случайного процесса, используя обозначение $M\{\dots\}$ как усреднение по времени:

$$R_{yx}(v) = M\{[y(t) - m_y][x(t - v) - m_x]\}.$$

Подставим в эту формулу значение $y(t)$ из уравнения свертки, записанного для бесконечного интервала времени:

$$R_{yx}(v) = M\left\{\left[\int_0^{\infty} g(\tau)x(t - \tau)d\tau - m_y\right][x(t - v) - m_x]\right\}.$$

Учитывая, что $m_y = \int_0^{\infty} g(\tau)m_x d\tau$, получим

$$R_{yx}(v) = \int_0^{\infty} g(\tau)M\{[x(t - \tau) - m_x][x(t - v) - m_x]\}d\tau.$$

Выражение под знаком $M\{\dots\}$ есть не что иное, как автокорреляционная функция входного сигнала, зависящая от сдвига сигналов по времени $\tau - v$ или $v - \tau$, все равно, так как функция четная:

$$R_{xx}(v - \tau) = M\{[x(t - \tau) - m_x][x(t - v) - m_x]\}.$$

В итоге получим искомое уравнение

$$R_{yx}(v) = \int_0^{\infty} g(\tau)R_{xx}(v - \tau)d\tau.$$

Для практического использования заменим в уравнении Винера–Хопфа бесконечный верхний предел на T :

$$R_{yx}(v) = \int_0^T g(\tau)R_{xx}(v - \tau)d\tau.$$

Это уравнение связывает модель динамической системы в форме импульсной переходной функции, автокорреляционную функцию входного сигнала и взаимнокорреляционную функцию между входным и выходным сигналом. Уравнение Винера–Хопфа вполне можно использовать для идентификации в условиях пассивного эксперимента. Допустим, переходный процесс в системе затухает за время $T = 2$ часа. Но чтобы получить корреляционные функции, определенные на этом интервале времени, мы использовали данные пассивного эксперимента, например за 20 сут. Таким образом, уравнение Винера–Хопфа содержит в себе несравнимо большую информацию, чем интегральное уравнение свертки (при пассивном эксперименте!).

Однако первое применение уравнения Винера–Хопфа для идентификации связано как раз с активным экспериментом, проводимым в ла-

бораторных условиях. Допустим, имеется генератор «белого шума», который можно подключить ко входу исследуемого объекта.

Автокорреляционная функция входного сигнала в этом случае равна:

$$R_{xx}(v - \tau) = D\delta(v - \tau),$$

где D – интенсивность белого шума;

$\delta(v - \tau)$ – дельта-функция.

В соответствии со свойствами дельта-функции получим:

$$R_{yx}(v) = \int_0^{\infty} g(\tau)R_{xx}(v - \tau)d\tau = \int_0^{\infty} g(\tau)D\delta(v - \tau)d\tau = Dg(v).$$

Отсюда $g(v) = \frac{1}{D} R_{yx}(v)$. То есть взаимнокорреляционная функция между выходным сигналом и входным сигналом в виде «белого шума» будет повторять по своей форме импульсную переходную функцию исследуемого объекта.

Рассмотрим влияние помех при измерении сигналов $x(t)$ и $y(t)$ на определение импульсной переходной функции. Вначале будем считать, что с помехой измеряется только выходной сигнал: $\hat{y}(t) = y(t) + f(t)$. Тогда

$$\hat{y}(t) = f(t) + \int_0^{\infty} g(\tau)x(t - \tau)d\tau.$$

Умножив это выражение на $x(t - v)$ и усреднив результат по времени, получим:

$$R_{\hat{y}x}(v) = R_{fx}(v) + \int_0^{\infty} g(\tau)R_{xx}(v - \tau)d\tau.$$

Наличие помехи привело к тому, что в уравнении Винера–Хопфа появилось дополнительное слагаемое – взаимнокорреляционная функция входного сигнала и помехи при измерении выхода. Если исследуемый объект не охвачен обратной связью, то такой корреляции не должно существовать.

Рассмотрим случай, когда с помехой измеряется входной сигнал: $\hat{x}(t) = x(t) + n(t)$. Тогда

$$y(t) = \int_0^{\infty} g(\tau)[x(t - \tau) + n(t - \tau)]d\tau =$$

$$= \int_0^{\infty} g(\tau)x(t-\tau)d\tau + \int_0^{\infty} g(\tau)n(t-\tau)d\tau.$$

Умножив это выражение на $[x(t-v) + n(t-v)]$ и усреднив результат по времени, получим следующее уравнение:

$$R_{yx}(v) + R_{yn}(v) = \int_0^{\infty} g(\tau)[R_{xx}(v-\tau) + R_{xn}(v-\tau) + R_{nx}(v-\tau) + R_{nn}(v-\tau)]d\tau.$$

При правильно поставленном эксперименте связь между значениями сигналов и помехами при их измерениях должна отсутствовать. Так же, скорее всего, не будет взаимосвязи между значениями выходного сигнала и ошибками при измерении входа. Тогда

$R_{xn} = R_{nx} = 0$, $R_{yn} = 0$, и уравнение примет вид:

$$R_{yx}(v) = \int_0^{\infty} g(\tau)[R_{xx}(v-\tau) + R_{nn}(v-\tau)]d\tau.$$

Из этого уравнения видно, что помеха при измерении входного сигнала приводит всё-таки к искажению интегрального уравнения, и как следствие – к погрешности в определении импульсной переходной функции. Наименьшая погрешность будет при помехе типа «белого шума». А чем больше АКФ помехи будет отличаться от дельта-функции (АКФ «белого шума»), тем большее искажение вносится в интегральное уравнение.

Вопросы и задания для самопроверки к разделу 14

1) Запишите уравнение Винера–Хопфа и дайте определение всех его составляющих. 2) При решении какой задачи было получено это уравнение? 3) В чем принципиальное отличие этого уравнения от интегрального уравнения вход–выход? 4) Уравнение Винера–Хопфа применяется при активном или пассивном эксперименте? 5) Как получить модель, используя уравнение Винера–Хопфа, если входной сигнал – белый шум? 6) Раскройте понятие «белый шум».

15. Решение уравнения Винера–Хопфа

Итак, имеем интегральное уравнение Винера–Хопфа, известные АКФ входного сигнала и ВКФ между выходным и входным сигналом. Задача: как определить модель объекта, т. е. входящую в уравнение импульсную переходную функцию.

С точки зрения классической математики рассматриваемое уравнение относится к классу интегральных уравнений Фредгольма I рода [32]. В общем случае уравнение Фредгольма записывается в виде

$$\lambda \int_a^b R(t, s) G(s) ds = f(t),$$

где λ – параметр, в уравнении идентификации его значение равно 1;

$R(t, s)$ – ядро уравнения, в уравнении идентификации ядром уравнения является автокорреляционная функция, являющаяся симметричной вещественной функцией разности аргументов (времён);

$G(s)$ – искомая функция (импульсная переходная);

$f(t)$ – свободный член уравнения Фредгольма (ВКФ).

Таким образом, для решения уравнения идентификации могут быть применены методы, разработанные в теории интегральных уравнений и вычислительной математике для решения уравнений Фредгольма I рода. Различают временные и спектральные методы. Временные методы непосредственно связаны с рассмотрением интегрального уравнения статистической идентификации, спектральные методы предполагают предварительное применение преобразования Фурье к интегральному уравнению.

Одним из наиболее простых методов решения уравнения является конечно-разностный метод. Рассмотрим уравнение Винера–Хопфа с конечным верхним пределом, равным T , и заменим интеграл на сумму, используя в качестве метода численного интегрирования метод прямоугольников. Получим:

$$R_{yx}(v) = \sum_{i=0}^n R_{xx}(v - \tau_i) g(\tau_i) \Delta T,$$

где ΔT – интервал дискретизации, в дальнейшем для простоты $\Delta T = 1$.

Итак, время τ задается теперь своими дискретными значениями. Будем в тех же точках задавать время v_i , $i=0, \dots, n$. Получим следующую систему уравнений:

$$\begin{bmatrix} R_{yx}(v_0) \\ R_{yx}(v_1) \\ \dots \\ R_{yx}(v_n) \end{bmatrix} = \begin{bmatrix} R_{xx}(0) & R_{xx}(\Delta T) & \dots & R_{xx}(n\Delta T) \\ R_{xx}(\Delta T) & R_{xx}(0) & \dots & R_{xx}[(n-1)\Delta T] \\ \dots & \dots & \dots & \dots \\ R_{xx}(n\Delta T) & R_{xx}[(n-1)\Delta T] & \dots & R_{xx}(0) \end{bmatrix} \begin{bmatrix} g(\tau_0) \\ g(\tau_1) \\ \dots \\ g(\tau_n) \end{bmatrix}.$$

Обозначим в приведенном выше выражении матрицу значений АКФ как R_{xx} , вектор значений ВКФ как R_{yx} и вектор значений импульсной реакции как G . Тогда система уравнений в векторно-

матричной форме будет иметь вид $R_{yx} = R_{xx}G$, а ее решение относительно значений импульсной реакции

$$G = R_{xx}^{-1}R_{yx}.$$

Матрица R_{xx} из-за аргумента, равного разности значений времени, имеет своеобразный вид. По диагонали стоят значения дисперсии, а в разные стороны от диагонали значения АКФ в моменты времени ΔT , $2\Delta T$, $3\Delta T$ и т. д. Так как АКФ является симметричной, то $R_{xx}(\Delta T) = R_{xx}(-\Delta T)$.

Рассмотрим пример на формирование матрицы значений АКФ. Пусть известны значения АКФ: $R_{xx}(0) = 10$, $R_{xx}(1) = 8$, $R_{xx}(2) = 2$, $R_{xx}(3) = 1$. Требуется составить матрицу значений АКФ для решения интегрального уравнения размерностью 4×4 . Ответ:

$$R_{xx} = \begin{bmatrix} 10 & 8 & 2 & 1 \\ 8 & 10 & 8 & 2 \\ 2 & 8 & 10 & 8 \\ 1 & 2 & 8 & 10 \end{bmatrix}.$$

После того как получен график значений импульсной переходной функции в точках, может возникнуть задача аппроксимации этих данных с помощью какой-либо аналитической функции. Подходы к решению задач аппроксимации, методы и алгоритмы изучаются обычно в курсе вычислительной математики.

Другим подходом к решению интегрального уравнения является использование проекционных методов. Проекционные методы предполагают замену точного уравнения приближенным за счет аппроксимации решения и последующее решение системы алгебраических уравнений. К числу проекционных методов относится и метод наименьших квадратов.

Представим импульсную реакцию в виде разложения по системе линейно независимых базисных функций $F_j(t)$:

$$g(t) = \sum_{i=0}^k c_i F_i(t).$$

Если определить коэффициенты c_i , то получим аналитическое описание импульсной переходной функции. Сформируем матрицу F , состоящую из значений различных базисных функций в разные моменты времени:

$$F = \begin{bmatrix} F_0(t_0) & \dots & F_k(t_0) \\ \dots & \dots & \dots \\ F_0(t_n) & \dots & F_k(t_n) \end{bmatrix}$$

Число коэффициентов c_i меньше, чем отсчетов времени, поэтому матрица F является прямоугольной.

Обозначим вектор искомых коэффициентов $C = \begin{bmatrix} c_0 \\ \dots \\ c_k \end{bmatrix}$. Тогда вектор

значений импульсной реакции равен: $G = FC$.

Подставим это выражение в формулу $R_{yx} = R_{xx}G$ и сформируем вектор невязок ε_i как разность между левой и правой частями этого уравнения:

$$R_{xx}FC - R_{yx} = E,$$

где $E = \begin{bmatrix} \varepsilon_0 \\ \varepsilon_1 \\ \dots \\ \varepsilon_n \end{bmatrix}$.

Применим МНК. Постановка задачи: найти вектор коэффициентов C таким образом, чтобы сумма квадратов невязок была минимальна. Критерий МНК в матричной форме

$$E^T E = (R_{xx}FC - R_{yx})^T (R_{xx}FC - R_{yx}) \Rightarrow \min.$$

Результат решения задачи:

$$C = [(R_{xx}F)^T (R_{xx}F)]^{-1} (R_{xx}F)^T R_{yx}.$$

Рассмотрим еще один способ применения МНК для нахождения значений импульсной реакции. Заменим в интегральном уравнении

$y(t) = \int_0^T g(\tau)x(t-\tau)d\tau$ интеграл на сумму. Дискретные значения τ_i берем через такт времени $\Delta\tau$ на интервале протекания переходного процесса:

$$y(t) = \sum_{i=0}^n x(t-\tau_i)g(\tau_i)\Delta\tau.$$

Дискретные значения времени t будем задавать на интервале много большем, чем интервал, на котором были заданы значения импульс-

ной переходной функции. В итоге получим следующую систему уравнений

$$\begin{bmatrix} y(t_0) \\ y(t_1) \\ \dots \\ y(t_N) \end{bmatrix} = \begin{bmatrix} x(t_0) & x(t_0 - \Delta t) & \dots & x(t_0 - n\Delta t) \\ x(t_1) & x(t_1 - \Delta t) & \dots & x(t_1 - n\Delta t) \\ \dots & \dots & \dots & \dots \\ x(t_N) & x(t_N - \Delta t) & \dots & x(t_N - n\Delta t) \end{bmatrix} \begin{bmatrix} g(0) \\ g(\Delta t) \\ \dots \\ g(n\Delta t) \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \dots \\ \varepsilon_N \end{bmatrix},$$

где $N \gg n$, а вектор значений ε_i представляет собой вектор невязок. Невязка возникает за счет дискретизации интегрального уравнения и всех других несоответствий. Обозначим в этом уравнении вектор значений выходного сигнала Y , матрицу значений входного сигнала X . Минимизируем сумму квадратов невязок с помощью МНК. В итоге получим следующее уравнение для оценки вектора значений импульсной переходной функции G :

$$G = (X^T X)^{-1} (X^T Y).$$

Фактически мы получили аналог уравнения Винера–Хопфа в дискретном виде. Здесь $X^T X$ – корреляционная матрица R_{xx} , только без усреднения по N реализациям, а $X^T Y$ – вектор значений ВКФ, так же без усреднения.

Вопросы и задания для самопроверки к разделу 15

1) Каким образом в принципе можно решить уравнение статистической идентификации? 2) Пусть АКФ имеет такие значения в точках 0,1,2,3,4: $R_{xx}(0) = 10$; $R_{xx}(1) = 7$; $R_{xx}(2) = 4$; $R_{xx}(3) = 2$; $R_{xx}(4) = 1$. Запишите матрицу значений АКФ размерностью 5 x 5 при численном решении уравнения.

16. Проблема некорректности при решении уравнения статистической идентификации

При решении уравнения Винера–Хопфа численными методами вместо гладкой кривой $g(t)$ во многих случаях можно получить кривую, имеющую скачкообразный характер и даже отдаленно не напоминающую импульсную переходную функцию исследуемого объекта (рис. 25).

Если попытаться увеличить число точек, в которых рассчитываются значения $g(t)$, получим еще более неутешительную картину.

Эти факты объясняются тем, что решение интегрального уравнения относится с точки зрения математики к классу некорректно постав-

ленных задач [27]. Некорректность в данном случае является математическим термином и не означает, что исходная задача была поставлена неверно.

Классическое определение корректности задачи дал французский математик Ж. Адамар. Задача считается корректно поставленной, если выполняются три следующих условия:

1. Решение задачи $Ax = y$ существует, т. е. существует x , удовлетворяющий условию $Ax = y$ при заданных A и y .

2. Решение единственно, т. е. искомая функция единственна для конкретных A и y .

3. Решение устойчиво, т. е. при непрерывном изменении исходных данных решение меняется непрерывно, или бесконечно малым вариациям исходных данных соответствуют бесконечно малые вариации решения.

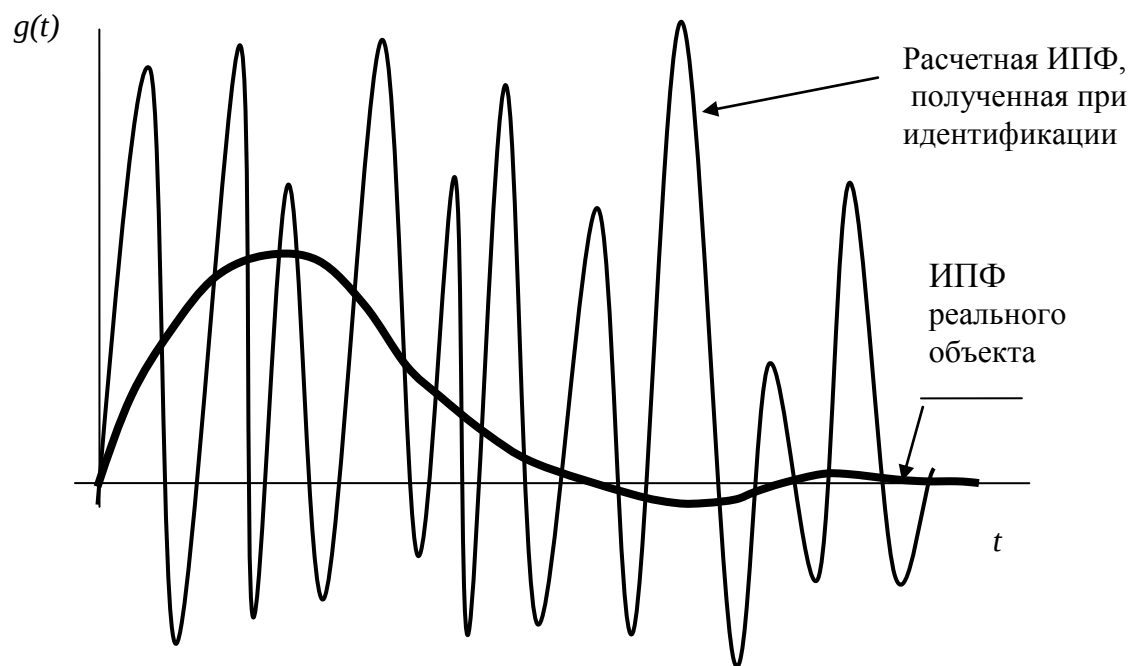


Рис. 25. Проявление некорректности при решении уравнения статистической идентификации

Задача определения импульсной переходной функции из уравнения статистической идентификации некорректна из-за невыполнения третьего условия: малые ошибки в задании исходных данных — в АКФ и ВКФ — приводят к большим ошибкам в решении. Степень неустойчи-

ности в значительной мере зависит от гладкости ядра $R_{xx}(v - \tau)$. Чем более гладкой является эта функция, тем выше получается коэффициент усиления ошибки. Гладкость ядра зависит от коррелированности сигнала. Вообще решение уравнения Фредгольма I рода в математике считается классической некорректной задачей.

Алгебраическая некорректность (неустойчивость) – именно с этим видом некорректности приходится иметь дело при численном решении уравнения идентификации – связана с тем, что определитель обращаемой матрицы равен или близок к нулю. Если увеличивать размер обращаемой матрицы, в расчете на то, что чем меньше интервал дискретизации, тем точнее будет решение, то строки и столбцы такой матрицы станут еще более близки друг другу и, соответственно, определитель ее также будет стремиться к нулю.

Долгое время считалось, что решение некорректно поставленных задач не имеет практической значимости. Однако в 60-х годах XX века необходимость их решения возникла не только в теории управления, но и в физике и в других научных приложениях.

Большой вклад в решение некорректно поставленных задач внесли отечественные ученые. Наиболее известен метод регуляризации академика А.Н. Тихонова. Если бы математики получали Нобелевские премии, то несомненно он был бы одним из лауреатов. В одной из своих работ А.Н.Тихонов [27] на очень простом примере иллюстрирует неустойчивость решения некорректной задачи. Пример такой: решение системы двух уравнений с двумя неизвестными:

$$\begin{cases} z_1 + 7z_2 = 5; \\ \sqrt{2}z_1 + \sqrt{98}z_2 = \sqrt{50}. \end{cases}$$

Система имеет точное решение: $z_1 = 0,1$, $z_2 = 0,7$, и это легко проверить.

При решении системы на ЭВМ с точностью до 100 знаков после запятой получено значение $z_1 = 0,00\dots$. При точности 300 знаков после запятой $z_1 = 1,6\dots$. При точности 500 знаков после запятой значение $z_1 = 5,\dots$.

Рассмотрим идею метода регуляризации. Допустим, при решении какой-то некорректной задачи минимизируется функционал J . Тихонов предложил искать решение в результате минимизации стабилизирующего функционала $S = J + \alpha F$, где F носит название регуляризирующего функционала, а α является численным параметром регуляризации.

Вернемся к численному решению задачи идентификации. Сформируем квадратичный критерий J и регуляризующий критерий F :

$$J = (R_{yx} - R_{xx}G)^T (R_{yx} - R_{xx}G).$$

$$F = G^T G.$$

Тогда

$$S = (R_{yx} - R_{xx}G)^T (R_{yx} - R_{xx}G) + \alpha G^T G \Rightarrow \min.$$

Минимум S достигается при условии $\frac{\partial S}{\partial G} = 0$ или $\frac{\partial S}{\partial G^T} = 0$. В результате получаем следующее выражение для определения вектора значений импульсной переходной функции:

$$G = (R_{xx}^2 + \alpha E)^{-1} R_{xx} R_{yx},$$

где E – единичная матрица;

α – параметр регуляризации.

Сравним эту формулу с той, которая была получена при численном решении интегрального уравнения статистической идентификации: $G = R_{xx}^{-1} R_{yx}$. То есть имеем другой вычислительный алгоритм. При использовании метода регуляризации обычно возникает вопрос о численном значении параметра регуляризации α . В ряде работ рекомендуется увязывать это значение с точностью измерения сигналов. Однако не всегда эта информация известна. Наиболее простым способом является экспериментальное определение параметра α . Дело в том, что зависимость критерия качества идентификации – среднеквадратичного отклонения между выходом объекта и модели от параметра регуляризации – имеет очень пологий экстремум. Достаточно просчитать задачу несколько раз при разных α и можно попасть в зону приемлемых значений параметра регуляризации (рис. 26).

Другой метод регуляризации заключается в замене исходного уравнения идентификации близким ему уравнением. Идея метода предложена М.М. Лаврентьевым. Выражение для расчета импульсной реакции при этом имеет вид

$$G = (R_{xx} + \alpha E)^{-1} R_{yx}.$$

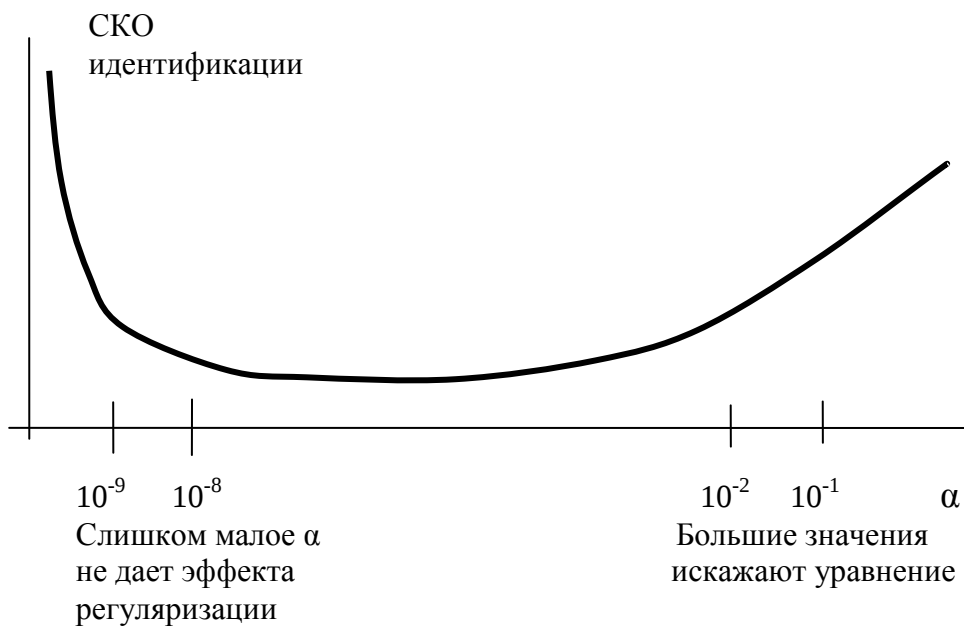


Рис. 26. Зависимость качества идентификации от параметра регуляризации

Еще одним подходом к решению некорректных задач является использование разложения сигналов по системе ортогональных полиномов. При использовании таких разложений обычно только несколько первых коэффициентов при базисных функциях несут в себе основную информацию об изменении сигналов. Вместо множества значений импульсной реакции необходимо искать небольшое количество коэффициентов разложения. При этом матрица, которую необходимо обращать при решении задачи идентификации, имеет небольшую размерность и определитель ее не равен нулю.

Вопросы и задания для самопроверки к разделу 16

1) По какой причине решение уравнения Винера–Хопфа относится к классу некорректных задач? 2) Как проявляется некорректность при численном решении уравнения на ЭВМ? 3) Как ставится задача нахождения импульсной переходной функции с применением метода регуляризации? 4) Как определить параметр регуляризации α ? Почему точность решения уравнения Винера–Хопфа не увеличивается с уменьшением интервала дискретизации?

17. Уравнение статистической идентификации в частотной области

В современной теории управления для описания динамических свойств систем и объектов управления широко используются частотные представления. Описание систем в частотной области и определение частотных характеристик во многих случаях оказывается более удобным и экономичным.

Аналогом таких статистических характеристик, как АКФ и ВКФ, в частотной области являются спектральная плотность сигнала S_{xx} и взаимная спектральная плотность S_{yx} . Часто автокорреляционная функция аналитически описывается в виде затухающей экспоненты:

$$R_{xx}(t) = De^{-\mu|t|}, \quad -\infty < t < +\infty.$$

Найдем выражение для спектральной плотности для этого случая:

$$\begin{aligned} S_{xx}(\omega) &= D \int_{-\infty}^{+\infty} e^{-\mu|t|} e^{-j\omega t} dt = D \int_{-\infty}^0 e^{+\mu t} e^{-j\omega t} dt + D \int_0^{+\infty} e^{-\mu t} e^{-j\omega t} dt = \\ &= D \int_{-\infty}^0 e^{(\mu-j\omega)t} dt + D \int_0^{\infty} e^{(-\mu-j\omega)t} dt = \\ &= D \frac{e^{(\mu-j\omega)t}}{\mu-j\omega} \Big|_{-\infty}^0 + D \frac{e^{(-\mu-j\omega)t}}{-(\mu+j\omega)} \Big|_0^{+\infty} = \frac{D}{\mu-j\omega} + \frac{D}{\mu+j\omega} = \frac{2D\mu}{\mu^2 + \omega^2}. \end{aligned}$$

Таким образом, график спектральной плотности представляет собой так называемую колоколообразную характеристику:

$$S_{xx}(\omega=0) = \frac{2D}{\mu}.$$

Если искать взаимную спектральную плотность, то в общем случае получим характеристику, содержащую вещественную и мнимую составляющие, т. к. ВКФ несимметрична относительно оси y .

Найдем уравнение статистической идентификации в частотной области. Запишем уравнение Винера–Хопфа с бесконечным верхним пределом:

$$R_{yx}(v) = \int_0^{\infty} g(\tau) R_{xx}(v-\tau) d\tau.$$

Применим к правой и левой частям уравнения преобразование Фурье:

$$\int_{-\infty}^{+\infty} e^{-j\omega v} R_{yx}(v) dv = \int_0^{\infty} g(\tau) \left[\int_{-\infty}^{+\infty} e^{-j\omega v} R_{xx}(v - \tau) dv \right] d\tau.$$

В левой части уравнения имеем выражение для взаимной спектральной плотности. Правую часть умножим на $e^{-j\omega\tau} \cdot e^{+j\omega\tau} = 1$. Получим

$$\begin{aligned} S_{yx}(j\omega) &= \int_0^{\infty} g(\tau) \left[\int_{-\infty}^{+\infty} e^{-j\omega v} \cdot e^{+j\omega\tau} \cdot e^{-j\omega\tau} R_{xx}(v - \tau) dv \right] d\tau = \\ &= \int_0^{\infty} g(\tau) e^{-j\omega\tau} \left[\int_{-\infty}^{+\infty} e^{-j\omega(v-\tau)} R_{xx}(v - \tau) dv \right] d\tau. \end{aligned}$$

Выражение в квадратных скобках в последней формуле есть спектральная плотность входного сигнала $S_{xx}(\omega)$. Учтем также, что

$$\int_0^{\infty} g(\tau) e^{-j\omega\tau} d\tau = W(j\omega) \text{ — передаточная функция объекта.}$$

В итоге получаем уравнение статистической идентификации в частотной области:

$$S_{yx}(j\omega) = W(j\omega) S_{xx}(\omega).$$

Формальное решение задачи идентификации выглядит так:

$$W(j\omega) = \frac{S_{yx}(j\omega)}{S_{xx}(\omega)}.$$

К сожалению, использование этого уравнения для идентификации может привести к получению физически нереализуемой модели. Другими словами, может оказаться, что порядок числителя передаточной функции выше, чем порядок знаменателя. Это означает, что сигнал на выходе системы появляется прежде, чем был подан сигнал на ее вход. Решением уравнения статистической идентификации с учетом физической реализуемости занимались такие известные ученые, как Винер, Боде, Шеннон. Впервые такое решение было получено при синтезе оптимального линейного фильтра. Вообще задачи идентификации и оптимальной фильтрации имеют много общего между собой. Собственно, разница в том, что при синтезе оптимального фильтра в уравнении ста-

статистической идентификации вместо ВКФ входного и выходного сигналов рассматривают ВКФ входного сигнала фильтра и наилучшего, желаемого выходного сигнала этого фильтра.

Получим решение статистического уравнения идентификации в частотной области с учетом физической реализуемости модели [14].

Запишем уравнение Винера–Хопфа для случая, когда переменная v меняется от минус бесконечности до плюс бесконечности:

$$R_{yx}(v) = A(v) + \int_0^{\infty} g(\tau) R_{xx}(v - \tau) d\tau,$$

где вспомогательная составляющая $A(v)$ соответствует отрицательным значениям $v < 0$;

$R_{yx}(v)$ определена на интервале от $-\infty$ до $+\infty$.

Взаимная корреляционная функция со временем стремится к нулю, и при положительном аргументе, и при отрицательном. Таким образом, вспомогательная функция $A(v)$ должна стремиться к нулю при $v \rightarrow -\infty$. Представим эту функцию в виде суммы затухающих экспонент:

$$A(v) = \sum A_i e^{\alpha_i v}, \quad v < 0.$$

Если найти изображение этой функции по Лапласу, получим

$$L\{A(v)\} = \sum \frac{A_i}{p - \alpha_i}.$$

Все корни (полюса) лежат в правой полуплоскости, если привести выражение к общему знаменателю, он будет состоять из произведения скобок, в каждой из которых присутствует знак «минус». Поэтому обозначим $L\{A(v)\} = A^-(p)$.

Будем искать передаточную функцию идентифицируемого объекта в классе устойчивых, физически реализуемых систем. Полюса такой передаточной функции должны лежать в левой полуплоскости, знаменатель ее будет состоять из скобок, в каждой из которых имеется знак «плюс». Обозначим искомую функцию как $W^+(p)$. Преобразуем исходное интегральное уравнение по Лапласу. С учетом введенных обозначений имеем:

$$S_{yx}(p) = A^-(p) + W^+(p) S_{xx}(p)$$

или

$$S_{yx}(p) - W^+(p) S_{xx}(p) = A^-(p).$$

Спектральная плотность $S_{xx}(p)$ является функцией четной, в ее знаменателе есть и правые, и левые корни. Проведем операцию факторизации, представив спектральную плотность в виде произведения со-

ставляющей $S_{xx}^+(p)$ с левыми корнями на составляющую $S_{xx}^-(p)$ с правыми корнями.

Факторизация:

$$S_{xx}(p) = S_{xx}^+(p) \cdot S_{xx}^-(p).$$

Получим

$$S_{yx}(p) - W^+(p) S_{xx}^+(p) S_{xx}^-(p) = A^-(p)$$

или

$$S_{xx}^-(p) \left[\frac{S_{yx}(p)}{S_{xx}^-(p)} - W^+(p) S_{xx}^+(p) \right] = A^-(p).$$

Разобьем слагаемое в квадратной скобке $\frac{S_{yx}(p)}{S_{xx}^-(p)}$ на сумму двух составляющих с правыми корнями и с левыми корнями. Такая операция называется сепаратизацией. Тогда

$$S_{xx}^-(p) \left[\left(\frac{S_{yx}(p)}{S_{xx}^-(p)} \right)^- + \left(\frac{S_{yx}(p)}{S_{xx}^-(p)} \right)^+ - W^+(p) S_{xx}^+(p) \right] = A^-(p).$$

Проанализируем полученное выражение. В левой части стоит выражение с правыми корнями. Тогда в правой части тоже должно быть выражение с правыми корнями. То есть все составляющие с левыми корнями и имеющие обозначение «+» должны уничтожиться. Это будет при условии

$$\left(\frac{S_{yx}(p)}{S_{xx}^-(p)} \right)^+ - W^+(p) S_{xx}^+(p) = 0.$$

Отсюда получим формулу для определения передаточной функции физически реализуемой системы:

$$W^+(p) = \left(\frac{S_{yx}(p)}{S_{xx}^-(p)} \right)^+ / S_{xx}^+(p).$$

Рассмотрим пример на определение передаточной функции физически реализуемого фильтра. Постановку задачи иллюстрирует рис. 27. На вход фильтра поступает сигнал $z(t)$, представляющий собой аддитивную смесь полезного сигнала $x(t)$ и помехи $n(t)$.

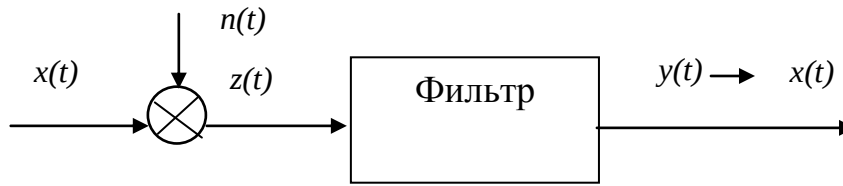


Рис. 27. Задача построения оптимального фильтра

Выходной сигнал фильтра $y(t)$ в наилучшем случае должен повторять полезный сигнал $x(t)$.

Задача. Даны автокорреляционные функции полезного сигнала $R_{xx}(t)$ и помехи $R_{nn}(t)$. Найти передаточную функцию физически реализуемого оптимального фильтра. Пусть

$$R_{xx}(t) = 0,5e^{-|t|}, \quad \text{т. е. } \mu = 1, \quad D = 0,5.$$

$$R_{nn}(t) = 2e^{-5|t|}, \quad \mu = 5, \quad D = 2.$$

Найдем спектральную характеристику полезного сигнала (см. ранее рассмотренный пример):

$$S_{xx}(\omega) = \frac{2D\mu}{\mu^2 + \omega^2}. \quad \text{Запишем спектральную плотность как функцию}$$

от переменной p , $p = j\omega$, тогда $\omega^2 = -p$.

$$S_{xx}(p) = \frac{2D\mu}{\mu^2 - p^2} = \frac{2 \cdot 0,5}{\mu^2 - p^2} = \frac{1}{\mu^2 - p^2}.$$

$$S_{nn}(p) = \frac{2 \cdot 2 \cdot 5}{25 - p^2} = \frac{20}{25 - p^2}.$$

Спектральная плотность сигнала на входе фильтра

$$S_{zz}(p) = S_{xx}(p) + S_{nn}(p) = \frac{1}{1 - p^2} + \frac{20}{25 - p^2} = \frac{45 - 21p^2}{(1 - p^2)(25 - p^2)}.$$

Рассмотрим, как изменится формула для определения передаточной функции применительно к задаче фильтрации. Взаимная корреляционная функция $R_{yx}(\tau)$, с учетом того что сигнал на выходе фильтра должен быть равен $x(t)$, а помеха некоррелирована с полезным сигналом, определится следующим образом:

$$R_{yx}(\tau) = M\{[x(t) + n(t)][x(t + \tau)]\} = R_{xx}(\tau) + R_{xn}(\tau) = R_{xx}(\tau).$$

Таким образом, в формуле для определения $W(p)$ вместо $S_{yx}(p)$ должна стоять $S_{xx}(p)$, а вместо спектральной плотности входного сигнала

ла $S_{xx}(p)$ – спектральная плотность суммы полезного сигнала и помехи $S_{zz}(p)$. Получим следующее выражение:

$$W^+(p) = \frac{\left(\frac{S_{xx}(p)}{S_{zz}^-(p)}\right)^+}{S_{zz}^+(p)}.$$

Проводим операцию факторизации:

$$S_{zz}(p) = S_{zz}^-(p) \cdot S_{zz}^+(p) = \frac{\sqrt{45} - \sqrt{21}p}{(1-p)(5-p)} \cdot \frac{\sqrt{45} + \sqrt{21}p}{(1+p)(5+p)}.$$

$$\text{Тогда } \frac{S_{xx}(p)}{S_{zz}^-(p)} = \frac{1}{1-p^2} \cdot \frac{(1-p)(5-p)}{\sqrt{45} - \sqrt{21}p} = \frac{5-p}{(1+p)(\sqrt{45} - \sqrt{21}p)}.$$

Проведем операцию сепаратизации. Представим только что найденную характеристику в виде суммы двух слагаемых. В числителе слагаемых запишем пока неизвестные коэффициенты a и b :

$$\frac{5-p}{(1+p)(\sqrt{45} - \sqrt{21}p)} = \frac{a}{1+p} + \frac{b}{\sqrt{45} - \sqrt{21}p} = \frac{a\sqrt{45} - a\sqrt{21}p + b + bp}{(1+p)(\sqrt{45} - \sqrt{21}p)}.$$

Определим неизвестные коэффициенты из условия равенства числителей в левой и правой частях этого уравнения. Приравнявая значения при одинаковых степенях p , получим следующую систему уравнений:

$$\begin{cases} 5 = a\sqrt{45} + b; \\ -1 = a\sqrt{21} + b. \end{cases}$$

Решив систему, получим $a \approx 0,53$, а значение коэффициента b в последующих расчетах не понадобится. Итак,

$$\left(\frac{S_{xx}(p)}{S_{zz}^-(p)}\right)^+ = \frac{0,53}{1+p}.$$

$$\text{Ответ: } W^+(p) = \frac{\frac{0,53}{1+p}}{\frac{\sqrt{45} - \sqrt{21}p}{(1+p)(5+p)}} = \frac{0,53(5+p)}{\sqrt{45} + \sqrt{21}p}.$$

Ещё один пример. Полезный сигнал и помеха по своим статистическим характеристикам неотличимы друг от друга:

$$S_{xx}(p) = \frac{1}{1-p^2}, \quad S_{nn}(p) = \frac{1}{1-p^2}.$$

Чему должна быть равна передаточная функция оптимального фильтра? Ответить на этот вопрос можно сразу же, на основании здравого смысла, не выполняя никаких выкладок. Но все-таки проверим, какой будет ответ.

$$S_{zz}(p) = \frac{1}{1-p^2} + \frac{1}{1-p^2} = \frac{2}{1-p^2}.$$

Операция факторизации:

$$S_{zz}(p) = S_{zz}^-(p)S_{zz}^+(p) = \frac{\sqrt{2}}{1-p} \cdot \frac{\sqrt{2}}{1+p}.$$

Найдем $\frac{S_{xx}(p)}{S_{zz}^-(p)} = \frac{1}{1-p^2} \cdot \frac{1-p}{\sqrt{2}} = \frac{1}{\sqrt{2}(1+p)}$. В этом выражении корни знаменателя – только левые, поэтому сепаратизацию делать нет необходимости, $\frac{S_{xx}(p)}{S_{zz}^-(p)} = \left(\frac{S_{xx}(p)}{S_{zz}^-(p)}\right)^+$. Найдем окончательный ответ:

$$W^+(p) = \frac{1}{\sqrt{2}(1+p)} \cdot \frac{(1+p)}{\sqrt{2}} = \frac{1}{2}.$$

Такой, собственно, результат и должен быть. На входе суммируются два неотличимых сигнала, полезный и вредный. Что должен делать фильтр? Ослабить сигнал в два раза.

Фильтр, полученный в рассмотренных выше примерах, носит название фильтра Винера. Назовем основные особенности винеровского метода фильтрации [2]:

1. Предполагается, что априорная информация о сигналах и шуме задана в терминах корреляционных функций и спектральных плотностей. При этом возникают задачи факторизации и сепаратизации спектральных плотностей.
2. Определение оптимального фильтра Винера разработано применительно к непрерывным сигналам и объектам и недостаточно хорошо подходит для цифровых вычислений.
3. Поиск оптимального фильтра заканчивается определением передаточной функции $W(p)$.
4. Метод не распространяется на нестационарные сигналы.
5. Его можно использовать для объектов со многими входами и выходами, но это усложняет вычисления.

Вопросы и задания для самопроверки к разделу 17

1) Запишите аналог уравнения Винера в частотной области и дайте определение всех его составляющих. 2) Как это уравнение можно получить? 3) Какие трудности возникнут при использовании этого уравнения для идентификации? 4) Каков физический смысл спектральной плотности? 5) Пусть АКФ описывается уравнением $R_{xx}(t) = 10e^{-2|t|}$, получите выражение для спектральной плотности. 6) Если АКФ представляет собой дельта-функцию, какова будет спектральная плотность этого сигнала? 7) Получите выражение для получения физически реализуемой передаточной функции. 8) Будет ли физически реализуемо звено $W(s) = \frac{10}{s^2 + s - 2}$? 9) На вход фильтра подается полезный сигнал, имеющий спектральную плотность $S_{xx}(s) = \frac{10}{1 - s^2}$, с полезным сигналом суммируется помеха, имеющая спектральную плотность $S_{nn}(s) = 2$. Найдите передаточную функцию физически реализуемого фильтра.

18. Фильтр Калмана–Бьюси

Редкая серьезная книга по идентификации обходится без описания этого фильтра [13, 14, 30]. При этом обычно рассматривается фильтр для многомерной системы сложной структуры, используются многочисленные матричные обозначения, и это отпугивает читателей от изучения калмановской фильтрации. Фильтр Калмана–Бьюси решает ту же задачу, что и фильтр Винера. Главное отличие – фильтр использует модель объекта в пространстве состояний. Должны быть заданы уравнение состояния и уравнение наблюдения. Все параметры этих уравнений должны быть известны. Задача состоит в том, чтобы по наблюдаемым значениям восстановить переменную или переменные состояния.

Рассмотрим одномерный, или скалярный, фильтр Калмана для дискретных сигналов и на этом простом примере попытаемся выяснить все его особенности.

Даны уравнение состояния и уравнение наблюдения:

$$\begin{aligned}x[k+1] &= ax[k], \\ y[k] &= hx[k] + v[k].\end{aligned}\tag{1}$$

Здесь $y[k]$ – наблюдаемый сигнал;

$x[k]$ – параметр состояния;

h – число, характеризующее свойства измерителя;

a – число, определяющее свойства объекта;

$v[k]$ – значение случайной помехи при измерении сигнала.

Представим реальный объект, который может соответствовать приведенной выше модели. Допустим, где-нибудь на Земле Франца-Иосифа стоит датчик, измеряющий температуру воздуха и периодически посылающий эту информацию по радиоканалу в метеоцентр. Таким образом, $x[k]$ – это температура в градусах Цельсия. Коэффициент a зависит от скорости и направления изменения температуры и, естественно, от интервала дискретизации. Этот коэффициент, по-видимому, должен перенастраиваться в зависимости от того, приближается зима или приближается лето, а также от времени суток. Коэффициент может быть как больше единицы, так и меньше. Значение $y[k]$ – это та информация, которая приходит в метеоцентр от автоматического датчика. Скорее всего, это определенный код и параметр h является коэффициентом масштабирования, связывающим получаемый код с привычными градусами Цельсия. Кроме того, этот коэффициент может учитывать возможное ослабление сигнала при передаче его от чувствительного элемента датчика до приемника.

Требуется по известным измерениям $y[k]$, $k=1, 2, \dots, k$ определять наилучшую в смысле минимума среднеквадратичной ошибки оценку параметра состояния $\hat{x}[k]$ в момент времени, соответствующий k -му измерению.

В литературе описаны различные способы вывода расчетных выражений фильтра Калмана–Бьюси: исходя из критерия максимума правдоподобия, на основе формулы Байеса для плотности вероятности параметров состояния, с использованием метода наименьших квадратов. Остановимся на последнем подходе.

С помощью метода наименьших квадратов и с использованием измерений наблюдаемой величины найдем наилучшую в среднеквадратичном смысле оценку параметра $\hat{x}[k]$. В чисто учебных целях ограничимся тремя измерениями: $y[1]$, $y[2]$, $y[3]$. Запишем систему из трех уравнений наблюдения для разных тактов времени:

$$\begin{cases} y[1] = hx[1] + v[1]; \\ y[2] = hx[2] + v[2]; \\ y[3] = hx[3] + v[3]. \end{cases} \quad (2)$$

Преобразуем эту систему таким образом, чтобы в каждом из уравнений присутствовало значение $x[3]$, используя для этого уравнение состояния:

$$\begin{cases} y[1] = ha^{-2}x[3] + v[1]; \\ y[2] = ha^{-1}x[3] + v[2]; \\ y[3] = hx[3] + v[3]. \end{cases} \quad (3)$$

Найдем выражение для расчета $x[3]$, исходя из минимума квадратичного критерия $J = \sum_{i=1}^3 v^2[i] \Rightarrow \min$. Обратите внимание, что в задаче идентификации на основе такого же критерия отыскивались коэффициенты модели, а здесь, в задаче фильтрации, должно быть найдено значение переменной состояния.

Обозначим для удобства $ha^{-2} = c_1$, $ha^{-1} = c_2$, $h = c_3$. Тогда критерий можно переписать в таком удобном виде:

$$J = \sum_{i=1}^3 (y[i] - c_i x[3])^2 \Rightarrow \min.$$

Условие минимума критерия J : $\frac{\partial J}{\partial x[3]} = 0$. Возьмем производную

$$\frac{\partial J}{\partial x[3]} = -2 \sum_{i=1}^3 c_i y[i] + 2x[3] \sum_{i=1}^3 c_i^2 = 0.$$

Отсюда получим

$$\hat{x}[3] = \frac{\sum_{i=1}^3 c_i y[i]}{\sum_{i=1}^3 c_i^2} = \frac{hy[3] + ha^{-1}y[2] + ha^{-2}y[1]}{h^2 + (ha^{-1})^2 + (ha^{-2})^2}. \quad (4)$$

Теперь усложним задачу. Считаем, что получено измерение $y[4]$ и известна только что полученная оценка $\hat{x}[3]$. Требуется с учетом этих данных получить рекуррентное выражение для оценки $\hat{x}[4]$. Исходная информация для решения этой задачи сосредоточена в системе следующих четырех уравнений:

$$\begin{aligned} y[1] &= ha^{-3}x[4] + v[1]; \\ y[2] &= ha^{-2}x[4] + v[2]; \\ y[3] &= ha^{-1}x[4] + v[3]; \\ y[4] &= hx[4] + v[4]. \end{aligned}$$

Запишем выражение для оценки $\hat{x}[4]$ по аналогии с ранее представленным выражением (4), выделив «добавку», полученную на 4-м такте времени:

$$\hat{x}[4] = \frac{ha^{-1}y[3] + ha^{-2}y[2] + ha^{-3}y[1] + hy[4]}{(ha^{-1})^2 + (ha^{-2})^2 + (ha^{-3})^2 + h^2} = \frac{\sum_{i=1}^3 c_i y[i] a^{-1} + hy[4]}{\sum_{i=1}^3 c_i^2 a^{-2} + h^2}. \quad (5)$$

Преобразуем это выражение с целью получения рекуррентной формулы для оценки $\hat{x}[4]$. Похожие выкладки делались при рассмотрении рекуррентного метода наименьших квадратов:

$$\hat{x}[4] = \frac{\sum_{i=1}^3 c_i y[i] a^{-1}}{\sum_{i=1}^3 c_i^2 a^{-2} + h^2} + \frac{hy[4]}{\sum_{i=1}^3 c_i^2 a^{-2} + h^2} = \frac{a \sum_{i=1}^3 c_i y[i]}{\sum_{i=1}^3 c_i^2 + a^2 h^2} + \frac{a^2 hy[4]}{\sum_{i=1}^3 c_i^2 + a^2 h^2}.$$

Применим формальный прием: прибавим к этой формуле одно и то

же выражение $\frac{a \sum_{i=1}^3 c_i y[i]}{\sum_{i=1}^3 c_i^2}$ со знаком «плюс» и со знаком «минус». Получим:

$$\begin{aligned} \hat{x}[4] &= \frac{a \sum_{i=1}^3 c_i y[i]}{\sum_{i=1}^3 c_i^2 + a^2 h^2} + \frac{a^2 hy[4]}{\sum_{i=1}^3 c_i^2 + a^2 h^2} + \frac{a \sum_{i=1}^3 c_i y[i]}{\sum_{i=1}^3 c_i^2} - \frac{a \sum_{i=1}^3 c_i y[i]}{\sum_{i=1}^3 c_i^2} = \\ &= \frac{a \sum_{i=1}^3 c_i y[i] \cdot \sum_{i=1}^3 c_i^2 + a^2 hy[4] \sum_{i=1}^3 c_i^2 + a \sum_{i=1}^3 c_i y[i] \cdot \sum_{i=1}^3 c_i^2 + a^3 h^2 \sum_{i=1}^3 c_i y[i] - a \sum_{i=1}^3 c_i y[i] \sum_{i=1}^3 c_i^2}{\sum_{i=1}^3 c_i^2 (\sum_{i=1}^3 c_i^2 + a^2 h^2)} + \\ &\quad - \frac{a^3 h^2 \sum_{i=1}^3 c_i y[i]}{\sum_{i=1}^3 c_i^2 (\sum_{i=1}^3 c_i^2 + a^2 h^2)}. \end{aligned}$$

Сгруппируем члены числителя следующим образом:

$$\hat{x}[4] = \frac{a \sum_{i=1}^3 c_i y[i] (\sum_{i=1}^3 c_i^2 + a^2 h^2)}{\sum_{i=1}^3 c_i^2 (\sum_{i=1}^3 c_i^2 + a^2 h^2)} + \frac{\sum_{i=1}^3 c_i^2 (a^2 h y[4] - a^3 h^2 \frac{\sum_{i=1}^3 c_i y[i]}{\sum_{i=1}^3 c_i^2})}{\sum_{i=1}^3 c_i^2 (\sum_{i=1}^3 c_i^2 + a^2 h^2)}.$$

Сократим одинаковые элементы в числителе и знаменателе, используем выражение для $\hat{x}[3]$ и получим:

$$\hat{x}[4] = a \cdot \hat{x}[3] + \frac{a^2 h}{\sum_{i=1}^3 c_i^2 + a^2 h^2} (y[4] - h a \cdot \hat{x}[3]). \quad (6)$$

Получили формулу для оценки переменной состояния в 4 такт времени. Результат складывается из двух составляющих. Первая составляющая – это расчет по модели, с использованием значения переменной состояния на предыдущем шаге. Вторая составляющая зависит от разности между измерением $y[4]$ и расчетом $h \cdot a \cdot \hat{x}[3]$ и берется с определенным весом. Если модель, связывающая соседние по времени значения переменной состояния, так хороша, что эта разность равна нулю, то оценка $\hat{x}[4]$ определяется только по модели $\hat{x}[4] = a \hat{x}[3]$. Эта формула, полученная с помощью рекуррентного МНК, своего рода «полуфабрикат» фильтра Калмана–Бьюси. Чтобы получить его в окончательном виде, нужно ввести в полученное выражение статистические характеристики помехи и ошибки оценивания.

Рассмотрим отдельно уравнение наблюдения, или уравнение измерителя:

$$y[3] = h x[3] + v[3]$$

или

$$\frac{y[3]}{h} - x[3] = \frac{v[3]}{h}. \quad (7)$$

Разность в левой части этого уравнения есть ошибка оценивания $e[3]$:

$$\frac{y[3]}{h} - x[3] = \hat{x}[3] - x[3] = e[3].$$

Дополним приведенное выше уравнение измерителя уравнениями для 1-го и 2-го тактов времени, т. е. мы как бы прогнозируем оценку $\hat{x}[3]$ по измерениям $y[1]$ и $y[2]$:

$$\begin{aligned}\frac{y[1]}{ha^{-2}} - x[3] &= \frac{1}{ha^{-2}} v[1] ; \\ \frac{y[2]}{ha^{-1}} - x[3] &= \frac{1}{ha^{-1}} v[2] ; \\ \frac{y[3]}{h} - x[3] &= \frac{1}{h} v[3] .\end{aligned}\tag{8}$$

Применив ранее введенные обозначения, каждое из этих уравнений можно записать как

$$e_i[3] = \frac{1}{c_i} v[i] \text{ или } c_i e_i[3] = v[i], \quad i = 1, 2, 3.$$

Просуммировав три уравнения, получим

$$\sum_{i=1}^3 c_i e_i[3] = \sum_{i=1}^3 v[i].\tag{9}$$

Обозначим дисперсию ошибки оценивания $e[3]$ как $P[3]$, а дисперсию помехи при измерении наблюдаемого сигнала как δ_v^2 . Учтем, что помехи $v[i]$ не зависимы друг от друга и имеют нулевое математическое ожидание. Ошибки оценивания e_i так же не зависимы друг от друга. Применяв формулу для получения дисперсии линейной функции некоррелированных случайных величин, получим выражение

$$\sum_{i=1}^3 c_i^2 \cdot P[3] = \delta_v^2 ,$$

отсюда

$$P[3] = \frac{\delta_v^2}{\sum_{i=1}^3 c_i^2} .\tag{10}$$

Итак, ранее с помощью МНК нашли оценку переменной состояния $\hat{x}[3]$. Найденное теперь значение $P[3]$ характеризует качество этой оценки. Значение $P[3]$ зависит от характеристики измерителя h , от свойств объекта a и от дисперсии помехи при измерении значений y . Затем была найдена с применением рекуррентного МНК оценка $\hat{x}[4]$.

Поставим задачу найти с использованием такого же подхода дисперсию ошибки оценивания значения $\hat{x}[4]$ – величину $P[4]$.

По итогам трех измерений было получено выражение

$$\sum_{i=1}^3 c_i^2 = \frac{\delta_v^2}{P[3]}. \quad (11)$$

По результатам 4 измерений будем иметь

$$(ha^{-3})^2 + (ha^{-2})^2 + (ha^{-1})^2 + h^2 = \sum_{i=1}^3 c_i^2 a^{-2} + h^2 = \frac{\delta_v^2}{p[4]},$$

откуда $P[4] = \frac{\delta_v^2}{\sum_{i=1}^3 c_i^2 a^{-2} + h}$. Выразив $\sum_{i=1}^3 c_i^2$ через $P[3]$, получим:

$$P[4] = \frac{\delta_v^2 a^2 P[3]}{\delta_v^2 + h^2 a^2 P[3]}. \quad (12)$$

Получили рекуррентное выражение для расчета дисперсии ошибки оценивания на четвертом шаге.

Заменяем в рекуррентной формуле для оценки $\hat{x}[4]$ значения $\sum_{i=1}^3 c_i^2$ на выражение (11) и обобщим полученные результаты для произвольного такта времени. Запишем два рекуррентных выражения, представляющих собой скалярный фильтр Калмана–Бьюси.

$$\hat{x}[k+1] = a\hat{x}[k] + \frac{a^2 h P[k]}{a^2 h^2 P[k] + \delta_v^2} (y[k+1] - ha\hat{x}[k]);$$

$$P[k+1] = \frac{\delta_v^2 a^2 P[k]}{\delta_v^2 + h^2 a^2 P[k]}.$$

Отметим, что для рекуррентного вычисления оптимальной оценки переменной состояния с помощью фильтра Калмана–Бьюси, необходимо задать начальные значения $\hat{x}[0]$ и $P[0]$. Применяется два способа задания начальных условий.

Первый способ. Значение $\hat{x}[0]$ задается произвольно, тогда $P[0]$ – достаточно большое число. Это означает, что при произвольном значении $\hat{x}[0]$ дисперсия ошибки оценивания в начальный момент времени может быть весьма велика. В дальнейшем при пересчете $P[k+1]$ по рекуррентной формуле ее значение быстро убывает. Например, пусть $a = h = 1$, $\delta_v^2 = 1$, $P[0] = 100$. Тогда $P[1] = 0,99$, $P[2] = 0,497$, $P[3] = 0,332$.

Второй способ. Перед началом использования алгоритма фильтрации проводятся предварительные расчеты по собранным статистическим данным с использованием МНК – примерно так, как выше определялись оценки $\hat{x}[3]$ и $P[3]$ по трем предварительным измерениям.

Распространим полученные выше уравнения фильтра Калмана–Бьюси на более сложные объекты.

1. Одномерная система, но модель генерации сигнала включает в себя генератор «белого шума». В этом случае имеем не только помеху при передаче сигнала по каналу измерения, но и «шум» у датчика. Должна быть задана дополнительная статистическая характеристика – дисперсия «белого шума» δ_g^2 . Модель объекта:

$$x[k+1] = ax[k] + g[k];$$

$$y[k] = hx[k] + v[k].$$

Выражение для фильтра Калмана–Бьюси:

$$\hat{x}[k+1] = a\hat{x}[k] + \frac{h(a^2P[k] + \delta_g^2)}{a^2h^2P[k] + \delta_v^2 + h^2\delta_g^2} (y[k+1] - ha\hat{x}[k]);$$

$$P[k+1] = \frac{\delta_v^2(a^2P[k] + \delta_g^2)}{\delta_v^2 + h^2a^2P[k] + h^2\delta_g^2}.$$

2. Состояние объекта характеризуется вектором переменных $X[k]$, но измерению доступна только переменная $y[k]$. Модель имеет следующий вид:

$$X[k+1] = AX[k];$$

$$y[k] = h^T X[k] + v[k],$$

где A – матрица состояния размерностью $n \times n$;

h^T – вектор-строка, характеризующая измерительное устройство;

n – размерность вектора состояния.

Уравнение фильтра имеет теперь векторно-матричную форму:

$$\hat{X}[k+1] = A\hat{X}[k] + AP[k]A^T h(h^T AP[k]A^T h + \delta_v^2)^{-1} (y[k+1] - h^T A\hat{X}[k]);$$

$$P[k+1] = AP[k]A^T - AP[k]A^T h(h^T AP[k]A^T h + \delta_v^2)^{-1} h^T AP[k]A^T.$$

Последнее выражение получено на основе рекуррентной формулы Вутбери для обращения матрицы. Здесь $P[k]$ есть корреляционная матрица вектора ошибок оценивания. Например, пусть вектор $X[k]$ состоит из двух переменных:

$$X[k] = \begin{bmatrix} x_1[k] \\ x_2[k] \end{bmatrix}, \text{ тогда } P[k] = \begin{bmatrix} D_{e1} & \text{cov}(e_1, e_2) \\ \text{cov}(e_2, e_1) & D_{e2} \end{bmatrix}.$$

По диагонали у этой матрицы стоит дисперсия ошибок оценивания параметров x_1 и x_2 , вне диагонали стоят коэффициенты корреляции между ошибками в оценивании этих параметров. При задании начальных значений матрица $P[0]$ берется диагональной.

3. Состояние объекта характеризуется вектором переменных $X[k]$, объект является управляемым, измерению доступна только переменная $y[k]$. Модель имеет следующий вид:

$$X[k+1] = AX[k] + bu[k],$$

$$y[k] = h^T X[k] + v[k],$$

где $u[k]$ – управляющее воздействие;

b – вектор управления.

Сравним этот случай с вариантом 1. Теперь вместо «белого шума» на переменные состояния будет влиять известное управляющее воздействие. Следовательно, в выражениях для фильтра вместо дисперсии δ_g^2 появится составляющая $(bu[k])^2 = b^T u^2[k] b$.

4. Классический многомерный случай, чаще всего встречающийся в учебной и научной литературе. Объект характеризуется вектором переменных состояния $X[k]$, вектором управляющих воздействий $U[k]$, вектором переменных наблюдения $Y[k]$. Модель объекта имеет вид:

$$X[k+1] = AX[k] + BU[k],$$

$$Y[k] = HX[k] + V[k].$$

Здесь B – матрица управления;

H – матрица наблюдения;

$V[k]$ – вектор помех по каждому из каналов измерения.

В формулах для фильтра в этом случае вместо дисперсии помехи δ_v^2 появится диагональная матрица дисперсии помех по каждому каналу D_v . Формулы имеют довольно громоздкий вид, например [30].

5. Объект является нестационарным. В этом случае матрица состояния A и если объект управляем, то и матрица управления B , состоят из коэффициентов, меняющихся с течением времени. Тогда к алгоритму фильтрации должны быть добавлены алгоритмы идентификации, которые пересчитывают коэффициенты матриц A, B на каждом шаге. В качестве метода идентификации могут использоваться рекуррентный или текущий МНК.

Вопросы и задания для самопроверки к разделу 18

1) В какой форме задается модель процесса при использовании фильтра Калмана? 2) Сколько выражений входят в вычислительную процедуру фильтра Калмана? 3) Для чего они используются? 4) Какая информация обязательно используется в фильтре Калмана? 5) Как могут быть заданы эти данные на первом шаге работы фильтра? 6) Основные особенности фильтра Калмана по сравнению с фильтром Винера?

19. Параметрическая идентификация динамических объектов методами пассивного эксперимента

При решении этой задачи предполагается известной структура исследуемого объекта и математическая форма модели. Такими формами могут быть дифференциальное уравнение, передаточная функция или разностное уравнение. Задача состоит в определении коэффициентов этих моделей.

19.1. Определение параметров дифференциального уравнения

Рассмотрим применение МНК для идентификации параметров дифференциального уравнения на простейшем примере. Допустим, что исследуемым объектом является апериодическое звено первого порядка:

$$Ty'(t) + y(t) = kx(t).$$

По данным пассивного эксперимента требуется определить параметры объекта – постоянную времени T и коэффициент усиления k . Попробуем применить схему, использованную при идентификации статических объектов. Схема такова: формируем разность между выходом реального объекта и выходом модели $e(t) = y(t) - y_M(t)$. Далее записываем квадратичный критерий и, минимизируя его, находим параметры модели. Но для динамической системы такой подход является неоправданно сложным. Дело в том, что $y_M(t)$ есть решение дифференциального уравнения, являющееся нелинейной функцией от искомых параметров:

$$y_M(t) = y_0 e^{-\frac{t}{T}} + \int_0^t e^{-\frac{t-\tau}{T}} kx(\tau) d\tau,$$

где y_0 – начальное условие.

Поэтому для применения МНК предпочтительнее другая схема. Берем исходное дифференциальное уравнение, но вместо переменных $y(t), y'(t), x(t)$ подставляем измеренные в дискретные моменты времени значения

$$Ty'(t_i) + y(t_i) \approx kx(t_i), \quad i = 1, 2, \dots, n.$$

Обратим внимание, что необходимо измерять не только значения выходного и входного сигналов, но и производную $y'(t)$. Разность между левой и правой частями уравнения будет невязкой:

$$e(t_i) = y(t_i) + Ty'(t_i) - kx(t_i).$$

Критерий МНК $J = \sum_{i=1}^n e(t_i)^2 \Rightarrow \min$. Условия, из которых определяются значения параметров модели:

$$\frac{\partial J}{\partial T} = 0, \quad \frac{\partial J}{\partial k} = 0.$$

$$J = \sum_{i=1}^n (y(t_i) + Ty'(t_i) - kx(t_i))^2;$$

$$\frac{\partial J}{\partial T} = 2 \sum_{i=1}^n (y(t_i) + Ty'(t_i) - kx(t_i))y'(t_i) = 0;$$

$$\frac{\partial J}{\partial k} = -2 \sum_{i=1}^n (y(t_i) + Ty'(t_i) - kx(t_i))x(t_i) = 0.$$

В итоге получаем следующую систему уравнений, линейных относительно искомых параметров:

$$\begin{cases} T \sum_{i=1}^n y'(t_i)^2 - k \sum_{i=1}^n x(t_i)y'(t_i) = - \sum_{i=1}^n y(t_i)y'(t_i); \\ T \sum_{i=1}^n y'(t_i)x(t_i) - k \sum_{i=1}^n x^2(t_i) = - \sum_{i=1}^n y(t_i)x(t_i). \end{cases}$$

Система решается любым из известных способов. Рассмотренный подход может быть в принципе применен и для оценки параметров модели, описываемой дифференциальным уравнением высокого порядка. Однако при этом возникают трудности, связанные с тем, что реально с допустимой точностью можно измерить только сами сигналы и первые две их производные в лучшем случае. Одним из способов решения задачи в такой ситуации является применение к каждому члену исходного дифференциального уравнения интегрального оператора. При интегрировании порядок исходного уравнения уменьшится на единицу, но

появятся составляющие вида $\int_0^t y(t)dt$ и $\int_0^t x(t)dt$. В литературе [32] отмечается, что такой подход пригоден для моделей до третьего порядка включительно, так как теперь возникнет новая неприятность, связанная с измерением начальных условий при пассивном эксперименте, которые могут быть ненулевыми.

Еще один подход основан на применении так называемых модулирующих функций. Исходное дифференциальное уравнение умножается на модулирующую функцию, определенную на интервале наблюдения за объектом. Из условия равенства нулю производных от модулирующей функции получается система алгебраических уравнений, в которую входят данные только о самих сигналах $y(t)$ и $x(t)$.

19.2. Определение параметров разностных уравнений

Рассмотрим оценку коэффициентов разностного уравнения методом наименьших квадратов на примере. Пусть задана модель в виде следующего разностного уравнения:

$$y[n+2] = ay[n+1] + by[n] + du[n],$$

где y – выходной сигнал объекта, u – входной сигнал, a, b, d – неизвестные коэффициенты, которые надо найти. Допустим, было проведено 10 измерений:

$$\begin{cases} y[1], y[2], \dots, y[10]; \\ u[1], u[2], \dots, u[10]. \end{cases}$$

Запишем невязки между левой и правой частями разностного уравнения для каждого такта времени:

$$y[3] - ay[2] - by[1] - du[1] = e[1];$$

$$y[4] - ay[3] - by[2] - du[2] = e[2].$$

.....

$$y[10] - ay[9] - by[8] - du[8] = e[8].$$

Обратите внимание, что получено 8 уравнений при 10 точках измерений. Введем следующие обозначения:

$$Y = \begin{bmatrix} y[3] \\ \dots \\ y[10] \end{bmatrix}, \quad B = \begin{bmatrix} a \\ b \\ d \end{bmatrix}, \quad E = \begin{bmatrix} e[1] \\ \dots \\ e[8] \end{bmatrix}, \quad X = \begin{bmatrix} y[2] & y[1] & u[1] \\ \dots & \dots & \dots \\ y[9] & y[8] & u[8] \end{bmatrix}.$$

Тогда критерий МНК – сумма квадратов невязок запишется в виде $J = E^T E = (Y - XB)^T (Y - XB)$. Для оценки вектора коэффициентов B можно использовать выражение, полученное при идентификации статических объектов МНК: $B = (X^T X)^{-1} X^T Y$, но матрица X теперь формируется по-другому (см. пример выше). Искомые коэффициенты могут быть найдены также при решении системы уравнений $(X^T X)B = X^T Y$. Для идентификации может быть использован и взвешенный МНК.

Рассмотрим еще один пример на формирование матрицы X и вектора Y при идентификации объекта с запаздыванием. Дана структура разностного уравнения:

$$y[n+1] = ay[n] + bu[n-2].$$

Сдвинем дискретные значения на 2 такта и получим следующее разностное уравнение, которое равносильно заданному:

$$y[n+3] = ay[n+2] + bu[n].$$

Проведено 10 измерений значений входного и выходного сигнала, т. е. $n=1, 2, \dots, 10$. Тогда

$$Y = \begin{bmatrix} y[4] \\ \dots \\ y[10] \end{bmatrix}, \quad X = \begin{bmatrix} y[3] & u[1] \\ \dots & \dots \\ y[9] & u[7] \end{bmatrix}, \quad B = \begin{bmatrix} a \\ b \end{bmatrix}.$$

19.3. Определение параметров передаточной функции с использованием аппроксимации сигналов ортогональными полиномами

В этом методе идентификации модель определяется сразу в форме передаточной функции $W(p)$. Структура модели задана, коэффициенты надо найти. Пусть $F_i(t)$, $i=0, 1, \dots, \infty$ есть система полиномов, ортогональных на интервале наблюдения за входным и выходным сигналами объекта $[0, T]$. Условие ортогональности:

$$\int_0^T F_i(t) F_j(t) dt = \begin{cases} 0, i \neq j; \\ p_i, i = j. \end{cases}$$

Если значения $p_i = 1$, то полиномы называются ортонормированными. Любая функция времени на заданном интервале времени может

быть представлена в виде разложения по системе ортогональных полиномов, например

$$x(t) = \sum_{i=0}^{\infty} C_i F_i(t) \approx \sum_{i=0}^n C_i F_i(t).$$

При решении практических задач в разложении берется конечное число членов, исходя из требуемой точности аппроксимации сигнала. Известны системы ортогональных полиномов Лежандра, Лягерра, Чебышёва, тригонометрические системы и т. д. Коэффициенты разложения C_i могут быть найдены как коэффициенты обобщенного ряда Фурье:

$$C_i = \int_0^T x(t) F_i(t) dt, \quad i=0, 1, \dots, n.$$

Если аналитическое выражение сигнала неизвестно, то коэффициенты разложения могут быть найдены с помощью метода наименьших квадратов. Перейдем к задаче идентификации. Входной и выходной сигналы аппроксимированы с помощью разложения по ортогональным полиномам:

$$x(t) = \sum_{i=0}^n C_i F_i(t); \quad y(t) = \sum_{i=0}^m D_i F_i(t).$$

Найдем изображения сигналов по Лапласу. Наиболее просто это получится, если используются степенные ортогональные полиномы, например полиномы Лежандра. Получим

$$X(p) = \sum_{i=0}^n C_i F_i(p); \quad Y(p) = \sum_{i=0}^m D_i F_i(p).$$

Здесь $F_i(p)$ – изображения по Лапласу ортогональных функций $F_i(t)$.

$$\text{Тогда } W(p) = \frac{Y(p)}{X(p)} = \frac{\sum_{i=0}^m D_i F_i(p)}{\sum_{i=0}^n C_i F_i(p)}.$$

Пример 1. Пусть в результате аппроксимации имеем $x(t) = C_1 t + C_0$, $y(t) = D_1 t + D_2 t^2$. Найдем изображения функций времени

$$X(p) = \frac{C_1}{p^2} + \frac{C_0}{p} = \frac{C_1 + C_0 p}{p^2};$$

$$Y(p) = \frac{D_1}{p^2} + \frac{2D_2}{p^3} = \frac{D_1 p + 2D_2}{p^3};$$

$$W(p) = \frac{Y(p)}{X(p)} = \frac{D_1 p + 2D_2}{p(C_0 p + C_1)}.$$

Пример 2. Простой, потому легко проверяемый. На вход системы подан ступенчатый сигнал $1(t)$, на выходе наблюдается постоянно возрастающий сигнал $y(t) = 0,5 \cdot t$. Сигналы простые, нет необходимости в их аппроксимации. Изображения сигналов:

$$X(p) = \frac{1}{p}; \quad Y(p) = \frac{1}{2p^2}.$$

$$W(p) = \frac{1}{2p^2} \cdot \frac{p}{1} = \frac{1}{2p}.$$

Идентифицируемый объект – интегрирующее звено с постоянной интегрирования $T = 2$.

При использовании этого метода идентификации при некачественной аппроксимации сигналов может получиться решение в виде физически нереализуемого звена, когда порядок числителя передаточной функции окажется больше порядка ее знаменателя.

Вопросы и задания для самопроверки к разделу 19

1) В чем принципиальное отличие в применении МНК при оценке параметров уравнения статики и дифференциального уравнения? 2) Пусть имеем дифференциальное уравнение: $ay'' + by' + y = kx + dx'$. Запишите критерий МНК для этого случая. 3) Запишите условия, которые обеспечат минимизацию критерия для этого случая. 4) Какие трудности возникнут при идентификации этого объекта? 5) Каков путь их преодоления? 6) Покажите, как будет решаться с помощью матричного МНК задача оценки коэффициентов такого разностного уравнения: $y[n+3] = a_2 y[n+2] + a_1 y[n+1] + a_0 y[n] + b_1 x[n+1] + b_0 x[n]$; проведены по 100 измерений входного и выходного сигналов: $x[1], \dots, x[100]$; $y[1], \dots, y[100]$. 7) Определите коэффициенты модели $y[n+1] = b_1 y[n] + b_2 x[n]$ по 4 измерениям y и x , проведенным в одни и те же моменты времени; y : 1, 3, 3, 7; x : 1, 0, 2, 0.

20. Адаптивные алгоритмы идентификации

Основная особенность таких алгоритмов заключается в том, что в них реализуется принцип подстройки модели к объекту по признакам близости поведения. В момент появления таких алгоритмов все элемен-

ты системы идентификации, включая модель, реализовывались средствами аналоговой техники. Затем, с появлением цифровых вычислительных устройств, некоторые функции, такие как вычисление критерия качества модели или автоматическая её подстройка, стали реализовываться алгоритмически. В настоящее время все методы адаптивной идентификации реализуются с помощью цифровой вычислительной техники, что приводит к некоторой размытости границ между различными способами идентификации [3].

Адаптивные алгоритмы идентификации обычно входят в состав адаптивных систем управления. Характерными особенностями таких систем являются следующие: система работает при наличии неопределенных или изменяющихся условий; получает информацию об этих условиях в процессе работы; сразу же использует эту информацию для изменения своего поведения.

Применение адаптивных методов к решению задачи идентификации позволяет определить характеристики исследуемого объекта в условиях его нормального функционирования, используя информацию об объекте и приложенных к нему воздействиях по мере её поступления, преодолеть трудности, связанные с отсутствием достаточно полной априорной информации. Полученная при этом модель объекта непосредственно используется для улучшения его работы. Обычно в адаптивных алгоритмах решается задача параметрической идентификации, т. е. структура модели предполагается известной и неизменной, а уточнению подлежат коэффициенты модели. Адаптивные алгоритмы идентификации в литературе часто называют методами идентификации с настраиваемыми моделями, так как перед включением в работу алгоритма идентификации начальная модель должна быть задана, пусть даже с весьма непохожими на реальные значения коэффициентами. На рис. 28 приведена общая схема адаптивной идентификации.

Настройка модели может рассматриваться как задача поиска экстремума функции $J(B)$ в пространстве параметров. Здесь B – вектор этих параметров, или, по-другому, коэффициентов модели. В качестве критерия обычно рассматривается функция ошибки $e(t) = y_{об}(t) - y_{мод}(t)$. Это может быть мгновенное значение квадрата ошибки $J_1 = e^2(t)$ или среднеквадратичная ошибка на скользящем интервале времени $J_2 = \int_{t-T}^t e^2(t)dt$. Может быть использована также сред-

неинтегральная абсолютная ошибка $J_3 = \int_{t-T}^t |e(t)| dt$. Критерии J_2 и J_3 можно модифицировать путем введения весовой функции, зависящей от времени.

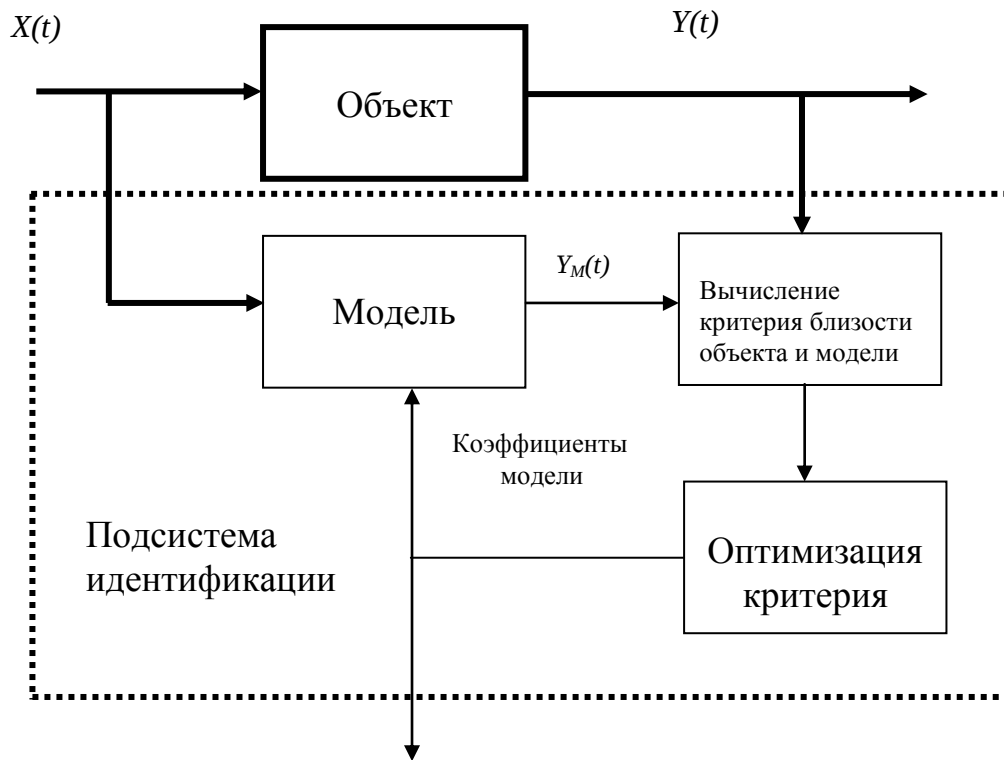


Рис. 28. Структурная схема системы адаптивной идентификации

Для достижения экстремума используются различные методы поиска экстремума функции многих переменных. Обычно эти методы изучаются в курсах «Методы оптимизации» или «Оптимальные и адаптивные системы». Наиболее распространены метод покоординатного спуска Гаусса–Зейделя, градиентные методы, методы случайного поиска. В градиентных методах для организации движения к экстремуму используется понятие градиента, т. е. векторной производной:

$$\text{grad}J(B) = \nabla J(B) = \left[\frac{\partial J}{\partial b_1}, \frac{\partial J}{\partial b_2}, \dots, \frac{\partial J}{\partial b_k} \right].$$

Градиент указывает направление наискорейшего роста функции, а обратное направление называется направлением скорейшего спуска. При использовании градиентных методов для идентификации с приме-

нением указанных выше критериев подстройка параметров идет в направлении скорейшего спуска.

Важной составной частью любого алгоритма является критерий останова работы алгоритма. Для останова поиска применяют следующие правила:

- выполняют остановку через определенное число шагов;
- выполняют остановку, когда приращение функции $J(B)$ станет меньше заданной величины;
- выполняют остановку, когда при любом из возможных шагов не происходит улучшения функции.

Рассмотрим непоисковые алгоритмы адаптивной идентификации [5,18].

Обычно такие алгоритмы идентификации представляют собой рекуррентные выражения. Новые значения коэффициентов модели пересчитываются на каждом шаге с использованием прежних значений и новой информации, полученной о сигналах на входе и выходе объекта:
$$B[n+1]=B[n]+F(\text{новая информация о поведении объекта на } (n+1) \text{ шаге}).$$

Если в алгоритме для расчета коэффициентов на $(n+1)$ шаге используются значения сигналов на этом шаге и значения коэффициентов, вычисленные на предыдущем такте, такой алгоритм называется одношаговым. Здесь вся информация, используемая для расчета, сдвинута по времени на один шаг. Если же в алгоритме используются коэффициенты, рассчитанные один, два, несколько тактов назад, или значения сигналов на входе и выходе объекта за несколько прошедших моментов времени, такие алгоритмы называются многошаговыми. Из рассмотренных ранее алгоритмов рекуррентный метод наименьших квадратов является алгоритмом одношаговым и рекуррентным, а применение текущего МНК приводит к алгоритму многошаговому и нерекуррентному. Некоторые из алгоритмов предназначены для работы со стационарными объектами. Такие алгоритмы способны уточнить, настроить значения параметров модели, но не способны отследить их при их изменении. Другие алгоритмы можно с успехом применять для идентификации нестационарных объектов.

Наиболее известным одношаговым непоисковым алгоритмом идентификации является алгоритм стохастической аппроксимации или, как его часто называют, алгоритм Роббинса–Монро. Этот алгоритм реализует, по существу, метод последовательного градиентного поиска в

условиях помех. Оценка вектора коэффициентов B на $(n+1)$ шаге получается в методе стохастической аппроксимации следующим образом:

$$B[n+1] = B[n] - \gamma_n F_n,$$

где γ_n – числовая последовательность, обеспечивающая сходимость алгоритма;

F_n – вектор-функция, представляющая собой градиент функционала $J(B)$.

Для обеспечения сходимости алгоритма числовая последовательность γ_n должна удовлетворять следующим условиям:

$$\lim_{n \rightarrow \infty} \gamma_n = 0;$$

$$\lim_{n \rightarrow \infty} \sum_{k=1}^n \gamma_k = \infty;$$

$$\lim_{n \rightarrow \infty} \sum_{k=1}^n \gamma_k^2 < \infty.$$

Таким условиям удовлетворяет последовательность $1/n$ или k/n , где k – любое положительное число. Пример последовательности, которая не удовлетворяет условиям сходимости:

$$1, \frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{16}, \dots$$

Рассмотрим пример алгоритма стохастической аппроксимации для конкретного случая. Модель статического объекта: $y = b_1 x_1 + b_2 x_2$. Необходимо оценивать на каждом шаге значения коэффициентов модели b_1, b_2 . Критерий качества идентификации примем следующий:

$$J_n(B) = \frac{1}{2} (y[n] - y_{\text{модели}}[n])^2. \text{ Запишем критерий для } (n+1) \text{ шага:}$$

$$J = \frac{1}{2} (y[n+1] - b_1[n]x_1[n+1] - b_2[n]x_2[n+1])^2.$$

$$\text{Найдем } F_n = \frac{\partial J}{\partial B[n]} = \begin{bmatrix} \frac{\partial J}{\partial b_1[n]} \\ \frac{\partial J}{\partial b_2[n]} \end{bmatrix},$$

$$\frac{\partial J}{\partial b_1[n]} = -\frac{2}{2} (y[n+1] - b_1[n]x_1[n+1] - b_2[n]x_2[n+1])x_1[n+1],$$

$$\frac{\partial J}{\partial b_2[n]} = -\frac{2}{2} (y[n+1] - b_1[n]x_1[n+1] - b_2[n]x_2[n+1])x_2[n+1].$$

В итоге получаем следующие выражения для оценки коэффициентов модели по методу стохастической аппроксимации:

$$b_1[n+1] = b_1[n] + \gamma_n (y[n+1] - b_1[n]x_1[n+1] - b_2[n]x_2[n+1])x_1[n+1];$$

$$b_2[n+1] = b_2[n] + \gamma_n (y[n+1] - b_1[n]x_1[n+1] - b_2[n]x_2[n+1])x_2[n+1].$$

На основе полученных выражений можно записать формулу в векторно-матричном виде для общего случая, когда оценивается вектор коэффициентов $B[n+1]$:

$$B[n+1] = B[n] + \gamma_n (y[n+1] - B^T[n]X[n+1])X[n+1],$$

где $X[n+1]$ – вектор значений входных сигналов в $(n+1)$ такт времени.

Алгоритм может быть применен для подстройки коэффициентов разностного уравнения. Например, для модели $y[n+1] = ay[n] + du[n]$

вектор коэффициентов $B^T[n] = [a[n] \quad d[n]]$, а $X[n+1] = \begin{bmatrix} y[n+1] \\ u[n+1] \end{bmatrix}$.

Очень важно при использовании этого алгоритма выполнять условия сходимости. Покажем это на числовом примере. Объект исследования задан уравнением $y = a + bx$. Истинные коэффициенты $a = b = 1$. Начальные значения коэффициентов, с которыми начинает работу алгоритм на первом шаге. $a = b = 0,5$. Последовательность чисел, обеспечивающая сходимость $\gamma = 1, \frac{1}{2}, \frac{1}{3}, \dots$. На пятом шаге работы алгоритма получим $a = 1,355$; $b = 0,86$. А теперь для этой же модели зададим постоянное значение $\gamma = 1$. На 4-м шаге работы алгоритма $a = 4$, $b = 9$. На 5-м шаге $a = -15$, $b = -29$, т. е. получаем явно расходящийся процесс.

Рассмотренный алгоритм применим для идентификации стационарных процессов, причем для начальной подстройки коэффициентов модели. При большом числе тактов работы алгоритма значение коэффициента γ_n станет очень малым и новая информация об объекте практически не будет влиять на подстройку коэффициентов модели. В случае если свойства объекта меняются во времени, необходимо периодически перезапускать алгоритм заново, начиная с первого шага.

В литературе [3] приводятся данные сравнительного анализа работы алгоритмов при разных вариантах задания γ_n . Если $\gamma_n = \frac{1}{n}$, то процесс подстройки медленный. Приближение к «истинному» значению

коэффициента модели идет, с одной стороны, начиная с некоторого значения n , алгоритм может практически перестать приближаться к истине. При $\gamma_n = \frac{k}{n}$ процесс подстройки имеет скачкообразный характер (рис. 29).

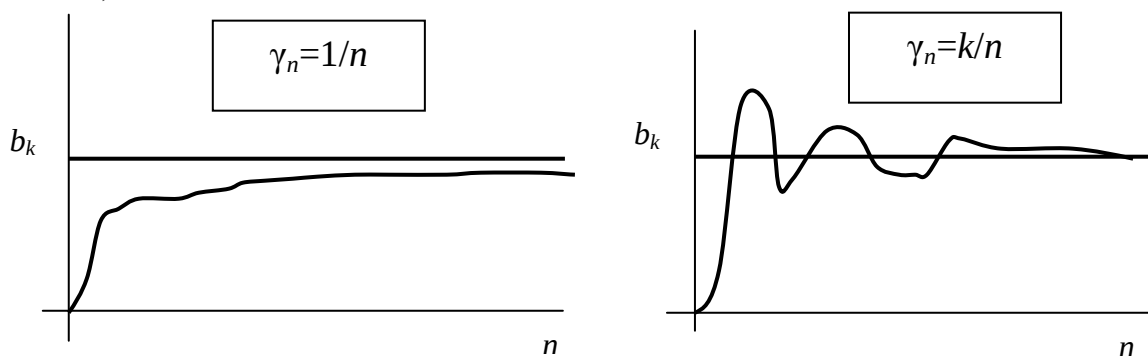


Рис. 29. Процесс подстройки коэффициентов модели алгоритмом стохастической аппроксимации

При практической реализации любых итерационных алгоритмов важнейшим показателем их работоспособности является скорость сходимости, позволяющая выяснить возможности и область применения этих алгоритмов. Повышение скорости сходимости связано в общем случае с усложнением структуры алгоритма.

На скорость сходимости существенно влияет наличие корреляционной связи между входными воздействиями. При сильной коррелированности сходимость существенно замедляется. В то же время отмечено, что при использовании стандартизованных входных и выходных переменных сходимость улучшается. Для стандартизации переменных их надо центрировать и нормировать:

$$x_{cm}[n] = \frac{x[n] - m_x[n]}{\sigma_x[n]}, \quad y_{cm}[n] = \frac{y[n] - m_y[n]}{\sigma_y[n]}.$$

При этом возникает дополнительная задача – с приходом новых измерений сигналов рекуррентно пересчитывать значение математического ожидания и среднеквадратичного отклонения. Такие рекуррентные формулы были приведены в разделе, посвященном статистическим характеристикам сигналов.

Одним из способов повышения сходимости для алгоритма Роббинса–Монро [3] является следующий. Числовая последовательность

$\gamma_n = \frac{k}{n}$ не меняется в зоне, удаленной от истинных значений коэффициентов модели. Вблизи же такой зоны коэффициенты, обеспечивающие сходимость, уменьшаются по обычному принятому правилу. Критерием приближения к зоне истинных значений является монотонное изменение коэффициентов модели. Если же они то увеличиваются, то уменьшаются, до оптимума еще далеко.

Рассмотрим некоторые другие одношаговые алгоритмы адаптивной идентификации. Достаточно известным является алгоритм Качмажа и его модификации:

$$B[n+1] = B[n] + \frac{y[n+1] - B^T[n]X[n+1]}{\|X[n+1]\|^2} X[n+1];$$

$$B[n+1] = B[n] + \frac{y[n+1] - B^T[n]X[n+1]}{\|X[n+1]\|^2 + k} X[n+1].$$

Здесь сходимость достигается путем деления ошибки идентификации на квадрат нормы значений входных сигналов. Квадрат нормы есть сумма квадратов значений входных сигналов.

Алгоритм Качмажа может применяться для идентификации нестационарных объектов, так как здесь нет множителя $\gamma_n = \frac{k}{n}$, уменьшающего приращение для рассчитываемых коэффициентов модели с течением времени.

На рис. 30 видно, как шла постройка двух коэффициентов объекта этим алгоритмом. Истинные значения коэффициентов увеличивались во времени по линейному закону.

Дальнейшим развитием этих алгоритмов являются так называемые релаксационные алгоритмы. При наличии неполной информации, т. е. неточно измеряемых входных и выходных параметрах и наличии неконтролируемых возмущений, вектор обновления коэффициентов модели $\Delta B[n+1] = B[n+1] - B[n]$ наряду с полезной информацией содержит составляющую, обусловленную указанными выше вредными факторами.

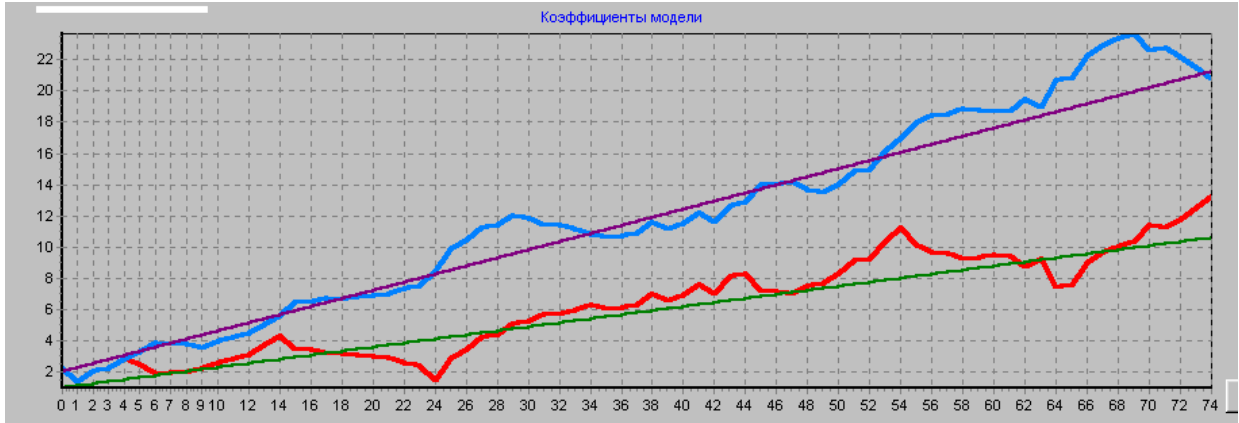


Рис. 30. Подстройка меняющихся параметров объекта алгоритмом Качмажа

С учетом этого шаг $\Delta B[n+1]$ необходимо сделать с некоторым весом q_n , зависящим от помех и возмущений. В частном случае значение q может быть постоянным. Для обеспечения сходимости величина множителя q должна быть в пределах $0 < q < 2$. Примеры релаксационных алгоритмов:

$$B[n+1] = B[n] + q_n \frac{y[n+1] - B^T[n]X[n+1]}{\|X[n+1]\|^2} X[n+1];$$

$$B[n+1] = B[n] + q_n \frac{y[n+1] - B^T[n]X[n+1]}{X^T[n+1] \text{sign}X[n+1]} \text{sign}X[n+1].$$

В последнем алгоритме сходимость обеспечивается использованием знаковых функций. На рис. 31 показана работа релаксационного алгоритма Качмажа при $q=0,2$. Если увеличивать значение этого коэффициента, процесс подстройки будет более колебательным (рис. 32).

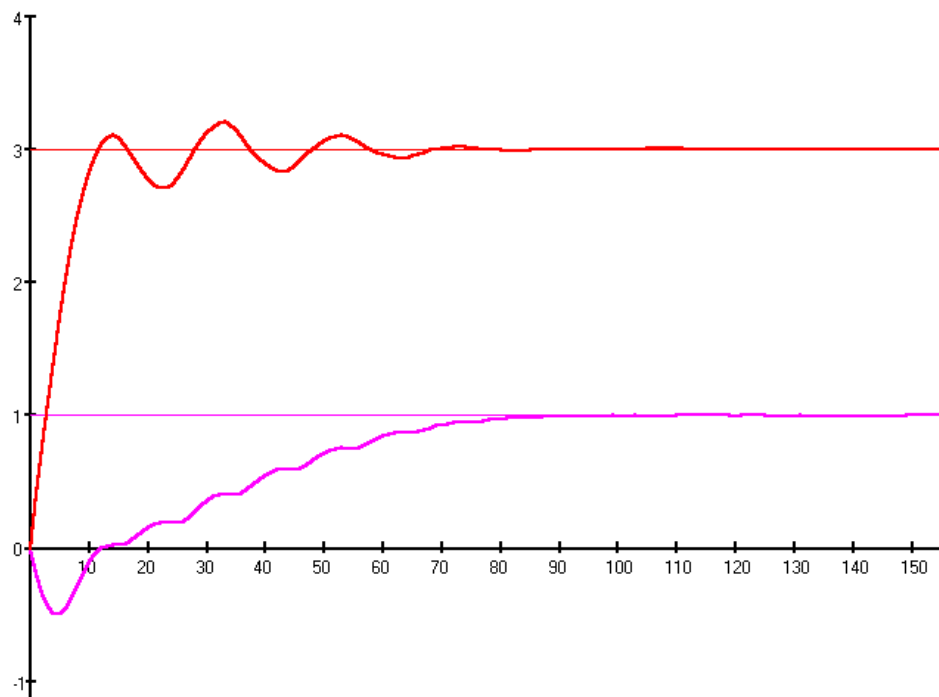


Рис. 31. Процесс подстройки коэффициентов стационарного объекта релаксационным алгоритмом при $q = 0,2$

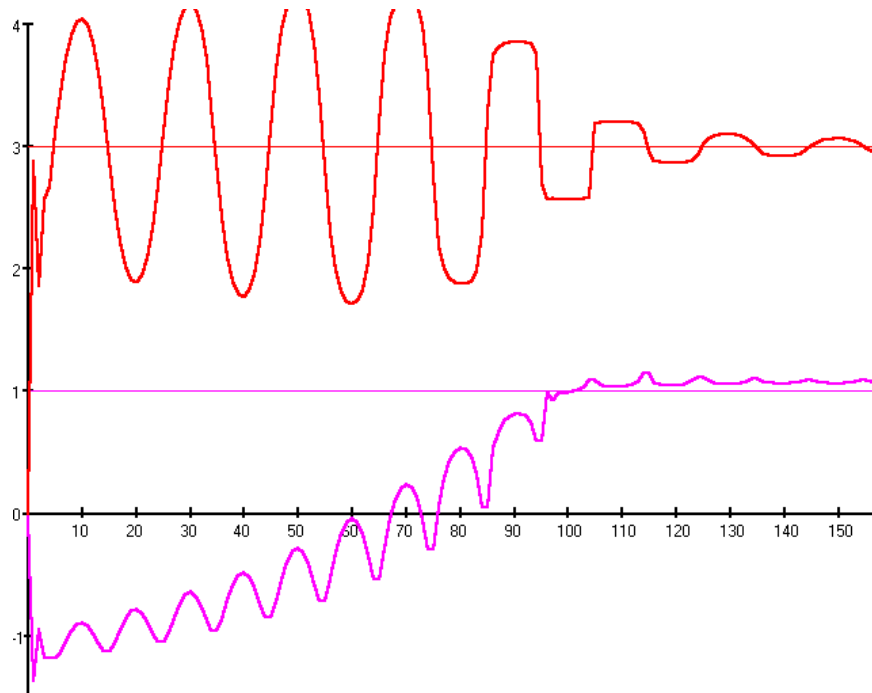


Рис. 32. Процесс подстройки коэффициентов стационарного объекта релаксационным алгоритмом при $q = 1,6$

В специальной литературе описано много различных одношаговых алгоритмов, отличающихся друг от друга методами обеспечения сходимости, скоростью сходимости, работоспособностью в условиях помех.

Быстросходящийся алгоритм Бунича характеризуется статической характеристикой типа «зона нечувствительности». При попадании параметров модели в эту зону их подстройка прекращается, но при выходе из зоны снова идет пересчет коэффициентов:

$$B[n+1] = B[n] + F\{y[n+1] - B^T[n]X[n+1]\}.$$

Обозначим рассогласование между выходом объекта и модели $y[n+1] - B^T[n]X[n+1] = e$, тогда функция $F\{\cdot\}$ в этом алгоритме равна:

$$F\{e\} = \begin{cases} e - p, & \text{если } e > p, \\ 0, & \text{если } |e| \leq p, \\ e + p, & \text{если } e < -p, \end{cases}$$

где p – зона нечувствительности, которая и является для этого алгоритма настроечным параметром.

Алгоритм применялся для оперативного прогнозирования конечной температуры стали в конвертере. Входными переменными были параметры чугуна, вес лома, расход кислорода на плавку и расход извести на плавку.

В заключение этого раздела приведем пример одного из двухшаговых алгоритмов идентификации:

$$B[n+1] = B[n] + q_n (y[n+1] - B^T[n]X[n+1]) \frac{X[n+1]\|X[n]\|^2 - X[n](X^T[n+1]X[n])}{\|X[n+1]\|^2\|X[n]\|^2 - (X^T[n+1]X[n])^2}.$$

Как видно, расчетное выражение в этом случае намного сложнее, чем для одношаговых алгоритмов.

Вопросы и задания для самопроверки к разделу 20

1) Запишите алгоритм Роббинса–Монро в скалярном виде применительно к модели $y = a_0 + a_1x$. 2) Запишите алгоритм Качмажа в скалярном виде применительно к модели $y = a_1x_1 + a_2x_2$. 3) Какая информация должна быть задана перед началом работы алгоритма Роббинса–Монро? 4) Приведите пример последовательности g_n , обеспечивающей сходимость алгоритма стохастической аппроксимации? 5) Пример последовательности, не обеспечивающей сходимость? 6) За счет чего обеспечивается сходимость в алгоритме Качмажа? 7) Какой из этих 2 алгоритмов не применим для идентификации нестационарных объектов и почему? 8) Эти алгоритмы одношаговые или многошаговые?

21. Диагностика систем

Диагноз в переводе с греческого «диагнозис» означает распознавание, определение [17]. В медицине, например, это – определение состояния человека, в технике – определение состояния технического устройства. Объект, состояние которого определяется, называется объектом диагноза. Диагноз представляет собой процесс исследования объекта диагноза. В результате завершения такого исследования формируются заключения вида: объект исправен, объект неисправен, в объекте имеется такая-то неисправность.

21.1. Общие понятия и определения

Диагностика есть отрасль знаний, включающая в себя теорию и методы организации процессов диагноза, а также принципы построения средств диагноза. Если объектами диагноза являются технические объекты, то говорят о технической диагностике.

Диагноз состояния объекта может быть непрерывным или периодическим. Во многих случаях диагностика сводится к оценке, исправен объект или нет, т. е. необходимо проводить проверку исправности объекта. Исправность предполагает полное отсутствие в объекте всякого рода неисправностей и дефектов.

Менее полной, чем проверка исправности, является проверка работоспособности. В работоспособном объекте могут оставаться необнаруженными мелкие неисправности, не препятствующие работе объекта. Например, резервированное устройство может быть работоспособным, несмотря на наличие неисправностей в резервных компонентах или связях.

Проверка правильности функционирования объекта отвечает на вопрос, правильно ли работает объект в данный момент времени, при данных внешних условиях и данном режиме работы.

Одной из важнейших задач диагностики является поиск неисправностей. При этом может быть достигнута разная степень детализации причины неисправности, или, говоря по-другому, глубина поиска. По принципу организации систем диагностики их можно разделить:

- на системы тестового диагноза;
- системы функционального диагноза.

При тестовом диагнозе имеется возможность подачи на объект специальных тестовых воздействий. При этом объект диагноза под-

ключается к системе диагностики и лишен возможности выполнять свои рабочие функции. Чем-то эти два метода диагностики напоминают методы активного и пассивного экспериментов в идентификации.

В качестве иллюстрации системы тестового диагноза рассмотрим следующий пример. Объект диагноза – электрический кабельный жгут. Определение его технического состояния осуществляется после его изготовления на этапе контроля с помощью специальной аппаратуры диагностики. Точками подключения жгута к этой аппаратуре являются все контакты штекерных разъемов и корпуса разъемов. Проверяются:

- правильность распайки;
- сопротивление изоляции между разными цепями;
- сопротивление изоляции между цепями и корпусами разъемов.

Проводятся также высоковольтные испытания изоляции. При испытании целостности цепей элементарная проверка состоит в подаче определенного напряжения на один из контактов той или иной цепи и проверке наличия этого напряжения на других контактах. Средства диагноза содержат несколько источников напряжения, устройства для измерения потенциалов и токов, устройство коммутации. Физической моделью исправного жгута является аппаратура для задания правильных (допустимых) значений ответов на элементарные проверки, а также цепи и переключатели для проведения различных проверок. Блок расшифровки результатов проверок представляет собой устройство сравнения фактических значений ответов с правильными.

В дальнейшем будем рассматривать методы и системы функционального диагноза. В этом случае необходимо определять текущие значения параметров, характеризующих объект диагноза, сравнивать их с эталонными значениями и выносить суждение о текущем состоянии объекта. Сам объект при этом выполняет свои функции.

Объектами диагноза могут быть устройства и системы разного уровня сложности. Например, в качестве объекта диагноза можно рассматривать технологический процесс со всем своим оборудованием и автоматикой: реакторами, колоннами, теплообменниками, насосами, датчиками, регуляторами, исполнительными механизмами. И, с другой стороны, объектом диагноза может быть отдельное техническое устройство: насос, электродвигатель, трансформатор, компрессор и т. д.

Задачи, решаемые при диагностике, можно разбить на две категории:

- оценивание параметров;
- распознавание образов.

Распознавание образов – процесс, при котором на основании непосредственных измерений первичных признаков решается задача о принадлежности объекта к определенному классу или определяется какая-то характеристика объекта, недоступная непосредственному измерению. Первичными признаками являются в данном случае значения измеряемых параметров, которые характеризуют состояние диагностируемого объекта: температура, давление, уровень, вибрация и т. д.

Методы распознавания образов нашли применение не только в задачах технической диагностики, но и при распознавании зрительных образов, речевых сигналов, при обработке радиолокационных и гидроакустических сигналов и т. п.

21.2. Виды неисправности технических систем

Основной целью диагностики является в конечном итоге повышение качества функционирования объекта диагноза. При наличии развитой встроенной системы диагностики возможно осуществление практически непрерывного контроля состояния отдельных узлов и системы в целом. Это во многих случаях позволяет избежать проведения планово-предупредительных ремонтов, которые обычно назначаются через определенные интервалы времени.

По мнению ряда авторов, система планово-предупредительных ремонтов, с одной стороны, уменьшает вероятность аварий, а с другой – может увеличить вероятность появления неисправностей в межремонтный период [4]. Это происходит, в частности, из-за того, что при переборке машин и конструкций могут появляться перекосы, люфты, нарушения соосности, что и приводит к сокращению срока жизни механизмов. На рис. 33 приведена типичная зависимость вероятности выхода из строя технического устройства от времени наработки без проведения профилактических ремонтов.

Использование системы диагностики позволяет выявлять дефекты на стадии их зарождения, предотвращать аварийные ситуации и сдвигать время планово-предупредительных ремонтов на время, большее чем T_p .

Неисправность, появившаяся в объекте диагностики в процессе его функционирования, приводит к отказу. Отказом называется потеря системой или её частью способности выполнять требуемые функции.

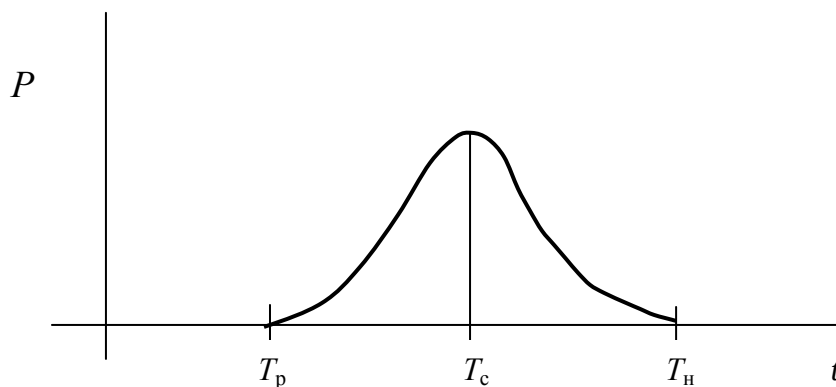


Рис 33. График вероятности выхода из строя технического устройства

Обозначения на рис. 33:

T_p – время гарантированной работы устройства;

T_n – наибольшее время наработки до выхода из строя;

T_c – среднее наиболее вероятное время поломки.

По скорости появления отказы принято делить на постепенные и внезапные. Постепенные отказы возникают вследствие медленного изменения параметров, например из-за износа, когда ухудшение технического состояния системы происходит постепенно. При внезапных отказах, по сути, происходит переход внешне не наблюдаемых количественных изменений в качественные.

По другой классификации отказы делятся на полные и частичные. При полном отказе система перестаёт выполнять свои функции. При частичном отказе система продолжает работать, но с худшим качеством. В сложных системах могут быть ситуации, когда полный отказ некоторого узла тем не менее существенно не влияет на функционирование системы в целом и наоборот, когда такой же отказ приводит к катастрофическим последствиям. Пример: отказ одного двигателя у четырехмоторного и одномоторного самолетов.

По физической природе отказы делятся на скачкообразные (перемежающиеся) и постоянные. При скачкообразных отказах происходит потеря некоторых функций устройством на очень короткий промежуток времени, причем после отказа оно вновь возвращается к полному выполнению своих функций. При полных отказах система прекращает свою работу.

Отказы возникают под влиянием таких воздействий, как:

- механические, в том числе статические, динамические, циклические;
- термические;
- химические.

Эти воздействия приводят:

- к износу элементов и узлов оборудования;
- трещинам, разрывам, деформации металла;
- коррозии и эрозии;
- нарушениям изоляции и крепления;
- снижению прочности конструкций.

Но не только эксплуатационные факторы могут быть причиной отказов оборудования. Причинами могут быть ошибки, допущенные при проектировании, нарушения технологии при изготовлении устройств, ошибки и нарушения при монтаже или сборке системы.

21.3. Диагностические модели и процедура диагностики

Диагностические модели должны отражать функциональные связи между характеристиками наблюдаемых сигналов и внутренними процессами в объектах диагноза [4].

Допустим, состояние объекта диагноза можно оценить набором измеряемых диагностических параметров r_j , представляющих собой вектор

$$R = [r_1, r_2, \dots, r_n].$$

Значения признаков $r_1, \dots, r_n$ либо непосредственно измеряются, либо вычисляются по значениям измеряемых сигналов. Кроме того, эти признаки должны быть связаны с некоторыми внутренними параметрами или характеристиками объекта диагноза a_j . Изменение значений a_j может привести к отказу объекта либо к какой-то неисправности в объекте диагноза. Каждый из диагностических признаков связан с внутренними характеристиками объекта:

$$r_j = f_j(a_1, \dots, a_m).$$

Обозначим вектор внутренних характеристик объекта $A = [a_1, \dots, a_m]$. Тогда в общем виде можно записать:

$$R = F\{A\},$$

где F – оператор, характеризующий преобразование параметров технического состояния a_j в пространство диагностических признаков.

Задача диагностики при этом формально может быть выражена обратной зависимостью:

$$A = F^{-1}\{R\}.$$

Это выражение, представляющее в самом общем виде связь между пространством параметров, характеризующим состояние объекта диагноза и пространством диагностических признаков, называется диагностической моделью. В задачах диагностики эта модель задается с помощью специальных математических методов. Обычно это не те методы, которые применяются при идентификации. Диагностическая модель может быть задана графом причинно-следственных связей, построена с использованием различного типа сетей, задана с использованием алгебры логики или на основе нечеткой логики, наконец, это может быть операционно-описательная модель.

Решение задачи диагностики, таким образом, предполагает выполнение следующих действий:

1. Измерение и предварительная обработка сигналов, поступающих от объекта диагноза.
2. Оценка первичных диагностических признаков на основе информации о сигналах от объекта диагноза.
3. Определение текущих значений параметров, характеризующих состояние объекта диагноза по диагностической модели.
4. Классификация состояния объекта диагноза на основании сравнения текущих параметров диагностической модели с известными (эталонными) значениями, характеризующими то или иное состояние объекта.

В частном случае вектор диагностических признаков R может совпадать с вектором измеряемых сигналов и тогда значения сигналов и являются диагностическими признаками. Например, для такого объекта диагноза, как химико-технологический процесс, диагностическими признаками будут измеряемые с помощью датчиков, установленных в различных точках процесса, значения температур, давлений, уровней, расходов. Именно по совокупности значений этих параметров можно вынести суждение, является ли текущее состояние процесса оптимальным, нормальным, предаварийным или аварийным.

При диагностике локальных технических устройств составляющими вектора R являются обычно различные характеристики, возможно, всего лишь одного сигнала, снимаемого с объекта диагностики. Такие информационные признаки можно разделить на временные и частотные. На рис. 34 изображены сигналы, меняющиеся со временем, у кото-

рых существенно изменились: среднее значение сигнала, дисперсия сигнала, частотные свойства при неизменном среднем значении и дисперсии.

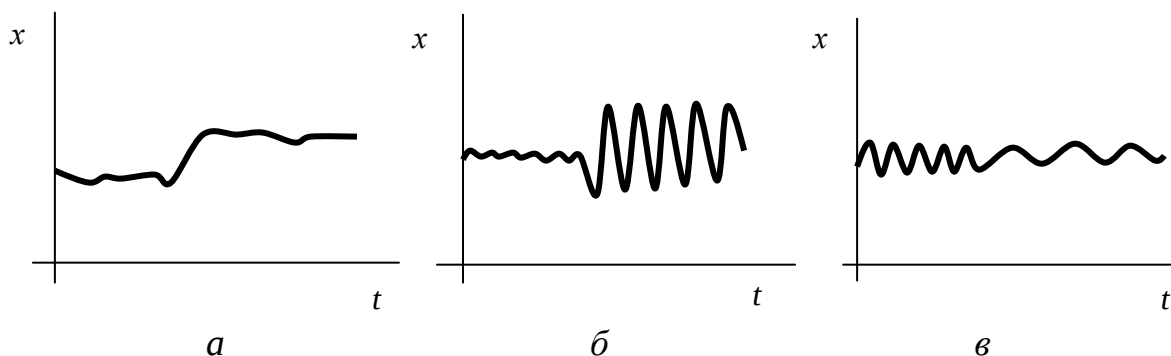


Рис. 34. Сигналы с различными диагностическими признаками

Особенно широкое применение в задачах диагностики находят частотные признаки. Например, компонентами вектора R могут быть определенные составляющие частотного спектра сигнала. Усиление либо ослабление какой-либо из этих составляющих может быть сигналом о появлении определенной неисправности у диагностируемого объекта.

21.4. Применение аппарата проверки статистических гипотез в диагностике

Так же как и в идентификации, при решении задач диагностики находят применение самые разнообразные математические методы. В монографии Д. Химмельблау (США) «Обнаружение и диагностика неполадок в химических и нефтехимических процессах» [29] большое внимание уделено аппарату проверки статистических гипотез. Обычно известны значения диагностических признаков, соответствующие нормальной работе оборудования. По экспериментальным данным можно получить значения этих же признаков на текущий момент времени. Естественно, эти данные в большей или меньшей степени будут отличаться от эталонных значений. Возникает вопрос, существенны ли эти различия? Если несущественны, то диагностируемое оборудование исправно, в противном случае можно утверждать, что обнаружена неисправность.

Итак, в терминах математической статистики при проверке гипотез подвергается испытанию некоторое утверждение – гипотеза H_0 в сравнении с другими утверждениями – гипотезами $H_1, H_2, \dots$. В самом простом случае альтернативной может быть тоже одна гипотеза. Например, гипотезой H_0 может быть утверждение, что значение некоторого признака $P=16$. Альтернативные гипотезы $H_1: P < 16$; $H_2: P > 16$. Или испытываемая гипотеза может быть выражена словесно: H_0 : «Оборудование исправно», альтернативная гипотеза H_1 : «Есть неисправность».

Последовательность действий при проверке гипотез. Основываясь на предположении, что нулевая гипотеза верна, вычисляют определенный статистический критерий по экспериментальным данным и проверяют, попадает ли вычисленное значение в область принятия гипотезы. Если оно не попадает в эту область, нулевая гипотеза отвергается, а принимается альтернативная гипотеза H_1 . В противном случае H_0 принимается, а H_1 отвергается. При проверке гипотезы задаются уровнем значимости критерия α , обычно значение α выбирается произвольным, но достаточно малым, например $\alpha = 0,05$ (5 %) или $\alpha = 0,01$ (1 %). Значение α определяет вероятность появления события, которое считается с точки зрения здравого смысла невозможным.

При проверке гипотез можно допустить ошибку. Различают ошибки двух видов (первого и второго родов).

Ошибка первого рода. Возникает, когда гипотеза верна, но отвергается (существует ложная тревога).

Ошибка второго рода. Эта ошибка возникает, когда гипотеза неверна, но принимается. То есть с точки зрения диагностики – неполадка существует, но она не обнаружена (нет тревоги).

Появление ошибки первого рода связано с тем, что значение уровня значимости α выбрано отличным от нуля, значение α и говорит о вероятности появления такой ошибки. Вероятность появления ошибки второго рода β также зависит от α , и попытка уменьшить вероятность появления ошибки первого рода увеличивает вероятность ошибки второго рода.

При классификации состояния объекта диагноза могут быть следующие варианты суждений:

1. Вероятность назвать исправный объект неисправным (α).
2. Вероятность назвать исправный объект исправным ($1 - \alpha$).
3. Вероятность назвать неисправный объект неисправным ($1 - \beta$).
4. Вероятность назвать неисправный объект исправным (β).

В литературе [29] приведен следующий пример применения процедуры проверки гипотез в целях диагностики. Технологическое оборудование: теплообменник, в котором идет нагрев (кипячение радиоактивной жидкости). Важно обнаружить, нет ли внутренней протечки из паровой трубы в межтрубное пространство. Кипятильник проверялся путем введения газообразного индикатора-изотопа в пар, подаваемый в кипятильник и взятием проб в месте выхода нагреваемой жидкости. Любая утечка будет обнаружена присутствием изотопа в пробах. Эксперимент по пяти пробам дал следующие результаты.

Среднее число единиц счета, приходящееся на пробу, составило: до введения индикатора $\bar{x}_B = 21,3 \cdot 10^4$; после введения индикатора $\bar{x}_A = 22,7 \cdot 10^4$. Известно значение среднеквадратичного отклонения $\sigma_x = 5020$.

Видно, что полученные экспериментально значения $\bar{x}_B$ и $\bar{x}_A$ имеют один и тот же порядок и отличаются незначительно. Вопрос: случайное это различие или неслучайное?

Проверяем нулевую гипотезу H_0 о равенстве двух выборочных средних значений: объемы выборок одинаковы и равны n ($n = 5$), выборочные дисперсии одинаковы. Для решения задачи используем T -критерий:

$$T_{\text{набл}} = \frac{|\bar{x}_A - \bar{x}_B|n}{\sigma_x \sqrt{2n}}.$$

Если значение $T_{\text{набл}}$ меньше критического значения критерия для выбранного уровня значимости и известного числа степеней свободы n , то принимается гипотеза H_0 (неисправности нет), иначе справедлива гипотеза $\bar{x}_A \neq \bar{x}_B$, т. е. есть протечка в теплообменнике.

$$T_{\text{набл}} = \frac{1,4 \cdot 10^4 \cdot 5}{5020 \sqrt{10}} = 4,409.$$

$$T_{\text{крит}} (\alpha/2 = 0,025) = 2,306.$$

Так как $T_{\text{набл}} > T_{\text{крит}}$, то гипотеза H_0 отвергается, а принимается альтернативная гипотеза H_1 : теплообменник неисправен, существует протечка.

21.5. Диагностика дискретных устройств

Многие объекты диагноза могут быть отнесены к классу дискретных комбинационных устройств. Для этих объектов значения выходных координат однозначно определяются только значениями входных координат. Общепринятой идеализацией при рассмотрении дискретных комбинационных устройств является предположение о том, что изменение значений выходных координат происходит мгновенно вслед за вызвавшими их изменение значениями дискретных сигналов на входе. Это условие равносильно отсутствию в устройстве контуров с обратными связями, элементов памяти и линий задержки. Наибольшее распространение в технике нашли двоичные дискретные устройства.

Для построения математических моделей комбинационных устройств используется два подхода: функциональный и структурный [17].

При функциональном подходе исследуемое устройство рассматривается как «черный ящик», т. е. как объект, внутренняя организация или структура которого неизвестна или не имеет значения. Исправное комбинационное устройство осуществляет однозначное преобразование двоичных слов на входах X_i в двоичные слова на выходах Z_j . Выходные функции $z_1, z_2, \dots, z_k$ являются булевыми функциями независимых входных переменных $x_1, x_2, \dots, x_n$. Для задания булевых функций часто используются таблицы истинности. Таблица истинности является функциональной математической моделью комбинационного устройства. От таблицы истинности можно перейти к аналитической функциональной математической модели комбинационного устройства, представив его выходные функции в виде формул булевой алгебры.

Задача диагностики будет состоять в проверке правильности функционирования комбинационного устройства в соответствии с известной моделью.

Структурный подход к построению модели предполагает учет не только выполняемых системой функций, но и её внутренней организации или структуры. В качестве структурной математической модели обычно принимается логическая сеть или схема, которая также может быть описана на языке теории графов.

Рассмотрим пример использования функциональной логической модели для диагностики конкретного технического устройства. В числе технологического оборудования, используемого на производстве метанола, имеется быстродействующая редуционно-охладительная установка (БРОУ). Она предназначена для получения пара среднего давления из

линии пара высокого давления (105 кг/см^2). Установка автоматически срабатывает при понижении давления в коллекторе пара среднего давления. Основным узлом установки является клапан с гидравлическим приводом. Для создания в гидравлической системе необходимого давления масла используется маслостанция. Стабилизация давления обеспечивается автоматически за счет периодического включения и отключения насоса. За давлением масла осуществляется контроль. Система защиты «наси́льно» включает двигатель насоса при падении давления масла ниже допустимого и выключает двигатель при повышении давления выше допустимого предела. Требуется вести контроль за системой защиты, обнаружив неполадки в ее работе в момент их возникновения. Неполадки могут быть связаны как с неисправностями в исполнительной системе (не сформировался сигнал на двигатель, не включился электродвигатель), так и с отказами дискретных датчиков давления.

Таблица истинности, описывающая логику работы системы защиты, приведена ниже.

В таблице истинности приведены 5 строк для исправного состояния системы и 2 строки для неисправного состояния (для примера), остальные – не приводятся. Первая строка соответствует ситуации, когда сформирована команда на включение из-за аварийно низкого давления, пятая строка – сформирована команда на отключение из-за аварийно-высокого давления.

Строки 2, 3 и 4 соответствуют нормальной работе маслосистемы, когда насос включается и отключается в результате работы автоматического релейного регулятора. При работе регулятора используются датчики P_H и P_L , состояние которых в данной таблице истинности не показано. В строках 6 и 7 для примера приведены варианты неисправной работы системы защиты: одновременно сработали датчики аварийно-низкого и аварийно-высокого давления; одновременно (причем без видимых причин) сформированы сигналы на отключение и включение двигателя насоса.

Составим логическую формулу, описывающую исправное состояние рассматриваемого комбинационного устройства, используем обозначения входов a, b, c, d ; обозначение выхода – z .

$$z = ab\bar{c}\bar{d} + \bar{a}b\bar{c}\bar{d} + \bar{a}\bar{b}\bar{c}\bar{d} + \bar{a}\bar{b}c\bar{d} + \bar{a}\bar{b}cd.$$

После упрощения булевой формулы получим:

$$z = b\bar{c}\bar{d} + \bar{a}\bar{b}c + \bar{a}\bar{b}\bar{c}\bar{d}.$$

Таблица 6

Таблица истинности для системы защиты

№	P_{LL}	Команда «включить»	Команда «отключить»	P_{HH}	Выход
	a	b	c	d	z
1	1	1	0	0	1
2	0	1	0	0	1
3	0	0	0	0	1
4	0	0	1	0	1
5	0	0	1	1	1
6	1	0	0	1	0
7	0	1	1	0	0
8–16	...	...	...	...	0

Алгоритм диагностики: если $z=true$, то система исправна, иначе – неисправна.

21.6. Контроль и диагностика насосных установок в нефтегазодобыче

При добыче нефти используется два типа насосных установок:

- штанговые глубинные насосные установки (ШГНУ);
- электроцентробежные насосные установки (ЭЦНУ).

ШГНУ обеспечивают добычу нефти наиболее дешевым из механизированных способов. Наземное оборудование установки включает в себя электропривод, редуктор, станок-качалку; подземное – насосно-компрессорные трубы, плунжерный насос, колонну штанг и т. д. Станок-качалку часто показывают в телерепортажах о нефтяниках.

При эксплуатации ШГНУ одной из важных задач является контроль подземного оборудования, прямой доступ к которому практически невозможен.

Одним из методов контроля за состоянием ШГНУ является использование динамограмм. Динамограмма связывает усилие P на штоке с ходом штанги (поршня насоса) h . Для получения информации о P и

h используются датчики усилия и датчики хода перемещения. По виду динамограммы можно оценить, находится ШГНУ в работоспособном состоянии или нет (допустим, произошел обрыв штанг).

Из-за больших динамических усилий датчики усилия и перемещения довольно часто выходят из строя. Замена датчиков, а также их калибровка в режиме нормального функционирования ШГНУ невозможна и требует остановки установки. Это обстоятельство является серьезным недостатком метода контроля с использованием динамограмм. Кроме того, этот метод не позволяет распознать такую неполадку в работе ШГНУ, как неуравновешенность станка-качалки.

Уравновешивание или балансировка станка-качалки выполняется с помощью специальных противовесов. От правильной балансировки зависит экономичность и надежность работы ШГНУ. В литературе [28] показано, что при неуравновешенности порядка 10 % потери электроэнергии могут достигать 15 – 20 %.

Перспективным направлением диагностики ШГНУ является контроль за электрическими параметрами сети питания электродвигателя. На рис. 35 изображена схема измерения входных электрических характеристик ШГНУ. Первичными датчиками являются датчик тока (ДТ) и датчик напряжения (ДН). Датчики расположены в шкафу управления и не испытывают динамических и атмосферных воздействий, как датчики усилия и перемещения. Устройство диагностики (УД) обрабатывает измеряемые сигналы, при необходимости дополнительно вычисляет такие электрические характеристики, как потребляемая активная, реактивная и полная мощность, передает информацию о состоянии диагностируемого объекта на центральный пульт управления (ЦПУ).

На рис. 36 приведены токограммы, показывающие зависимость действующего значения тока от хода поршня насоса, отсчитанного во времени. Здесь $T_{\text{кач}}$ – период качания станка-качалки. Из приведенных графиков видно, что частотный спектр сигнала для уравновешенного и неуравновешенного станка существенно отличается. Из приведенных графиков также видно, что другим признаком для диагноза могут быть средние значения тока на полупериоде качания. Для диагностики ШГНУ могут быть использованы ваттметрограммы или варметрограммы, показывающие зависимость потребляемой активной и, соответственно, реактивной мощности ШГНУ от времени, а также вольтамперограммы, показывающие изменение полной потребляемой мощности во времени.

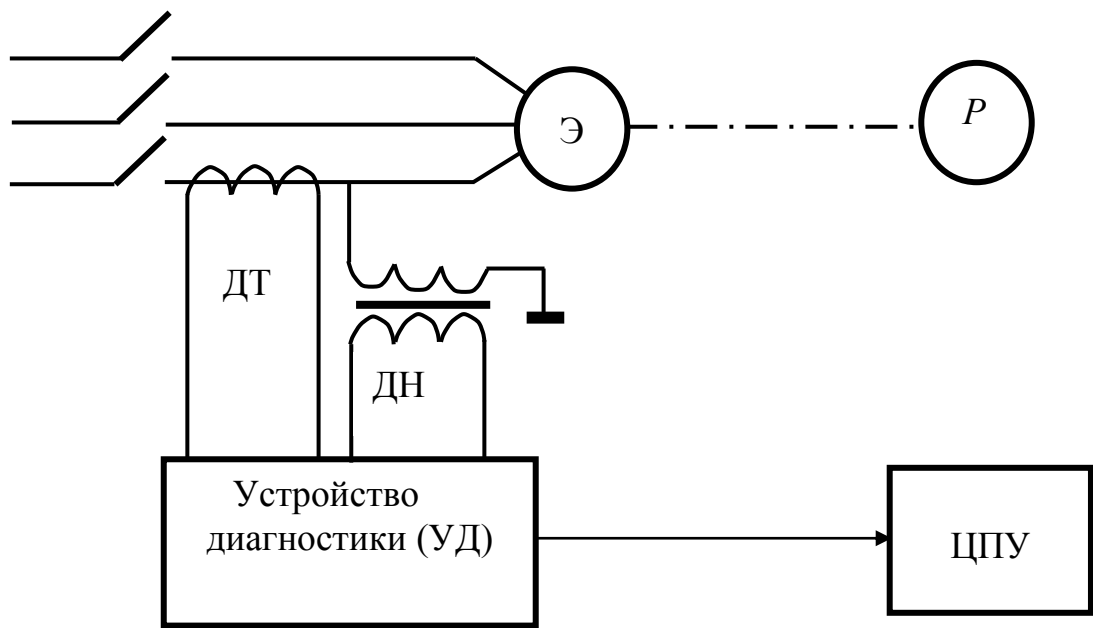


Рис. 35. Схема измерения электрических характеристик ШГНУ

Диагностика ШГНУ по электрическим параметрам является новой задачей. При ее решении могут возникнуть такие сложности, как возможная потеря информации при ее воспроизведении по цифровым отсчетам, правильность определения электрических характеристик по мгновенным значениям тока и напряжения, большие объемы обрабатываемых данных. Обширная библиография по этой проблеме приведена в монографии Е. И. Гольдштейна и его учеников [28].

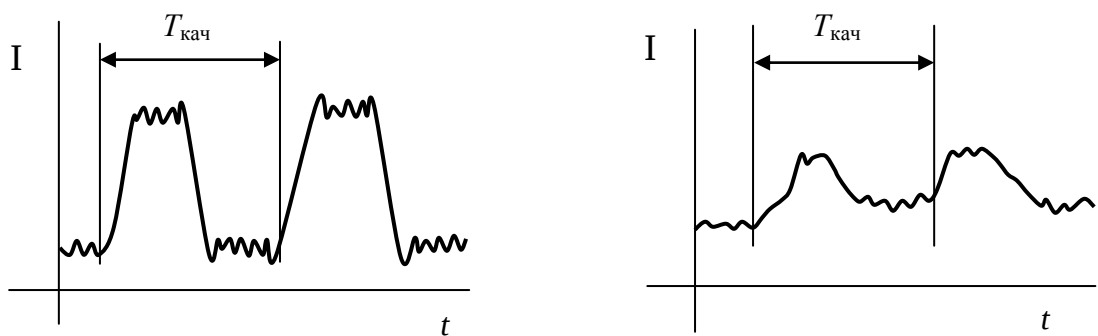


Рис. 36. Токограммы уравновешенного и неуравновешенного станка-качалки

Для диагностики ЭЦНУ информативными характеристиками являются спектральные плотности токов, напряжений, активной мощности. Исследования показали, что спектральные портреты для каждой конкретной установки существенно отличаются друг от друга.

Вопросы и задания для самопроверки к разделу 21

1) Сформулируйте задачу диагностики. 2) Перечислите по порядку этапы решения этой задачи. 3) Какие переменные связывает между собой диагностическая модель? 4) Перечислите возможные диагностические признаки для непрерывного сигнала. 5) Перечислите возможные диагностические признаки для сигнала, представляющего собой последовательность импульсов. 6) Какие сигналы могут использоваться при диагностике ШГНУ? 7) Какие диагностические признаки можно использовать у этих сигналов?

Заключение

В этом учебном пособии предпринята попытка изложить основные вопросы, связанные с построением моделей исследуемых объектов, идентификацией и диагностикой. Отметим, что некоторые разделы этой книги могут быть развернуты в отдельные самостоятельные курсы. Достаточно только упомянуть регрессионный анализ. Например, популярная монография Н. Дрейпера и Г. Смита «Прикладной регрессионный анализ» издана в 2 томах общим объемом более 600 страниц. Не охваченными остались вопросы использования теории чувствительности в целях идентификации, типовая идентификация Н.С. Райбмана.

Следует отметить, что построение модели исследуемой системы всегда является вспомогательным (инструментальным) средством при решении многих задач: предсказания, управления, имитационного моделирования, восстановления наблюдений, проектирования фильтров и др. Качество полученных моделей определяется качеством данных и правильно заданной структурой модели. Никакие методы не позволят получить хорошую модель при низком соотношении сигнал/шум или структуре модели, в которую не входят основные факторы, определяющие функционирование исследуемого объекта.

Список литературы

1. Александровский Н.М., Егоров С.В., Кузин Р.Е. Адаптивные системы автоматического управления сложными технологическими процессами. – М.: Энергия, 1973. – 272 с.
2. Адаптивные фильтры: пер. с англ. / под ред. К.Ф.Н. Коуэна и П.М. Гранта. – М.: Мир, 1988 – 392 с.
3. Алексеев А.А., Имаев Д.Х., Кузьмин Н.Н., Яковлев В.Б. Теория управления: учебник. – СПб.: Изд-во СПбГЭТУ ЛЭТИ, 1999. – 435 с.
4. Алексеев А.А., Солодовников А.А. Диагностика в технических системах управления: учебное пособие для ВТУЗов / под ред. В.Б. Яковлева. – СПб.: Изд-во СПбГЭТУ ЛЭТИ, 1997. – 188 с.
5. Автоматизированные системы управления технологическими процессами и оптимальное управление / под ред. В.И. Салыги. – Харьков: Вища школа, 1976. – 180 с.
6. Адлер Ю.П. Введение в планирование эксперимента. – М.: Металлургия, 1969. – 159 с.
7. Айвазян С.А., Енюков И.С., Мешалкин Е.Д. Прикладная статистика: Исследование зависимостей: справочное издание / под ред. С.А. Айвазяна. – М.: Финансы и статистика, 1985. – 487 с.
8. Гроп Д. Методы идентификации систем: пер. с англ. – М.: Мир, 1979. – 304 с.
9. Дрейпер Н., Смит Г. Прикладной регрессионный анализ: пер. с англ. – М.: Статистика, 1973. – 392 с.
10. Дейч А.М. Методы идентификации динамических объектов. – М.: Энергия, 1997. – 240 с.
11. Коновалов В.И. Идентификация объектов управления. – Томск: Изд-во ТПИ, 1981. – 90 с.
12. Липатов Л.Н. Типовые процессы химической технологии как объекты управления. – М.: Химия, 1973.
13. Льюнг Л. Идентификация систем: пер. с англ./ под ред. Я.З. Цыпкина. – М.: Наука, 1991. – 432 с.
14. Методы классической и современной теории автоматического управления: учебник в 5 т.: Т.2: Статистическая динамика и идентификация систем автоматического управления / под ред. К.А. Пупкова, Н.Д. Егупова. 2-е изд. – М.: Изд-во МГТУ им. Н.Э. Баумана, 2004. – 640 с.
15. Обувалин М.И., Макаревич И.И., Астахова К.В. Определение зависимостей по методу наименьшего произведения // Идентифи-

- кация и аппаратура для статистических исследований. Труды I Всесоюзного симпозиума по статистическим проблемам в технической кибернетике. – М.: Наука, 1970. – с. 39–47.
16. Ордынцев В.М. Математическое описание объектов автоматизации. – М.: Машиностроение, 1965. – 360 с.
 17. Основы технической диагностики. Кн. I. Модели объектов, методы и алгоритмы диагноза / под ред. П.П. Пархоменко. – М.: Энергия, 1976. – 464 с.
 18. Перельман И.И. Оперативная идентификация объектов управления. – М.: Энергоиздат, 1982. – 272 с.
 19. Райбман Н.С. Что такое идентификация? – М.: Наука, 1970. – 119 с.
 20. Райбман Н.С., Чадеев В.М. Построение моделей процессов производства. – М.: Энергия, 1975. – 376 с.
 21. Ротач В.Я. Теория автоматического управления теплоэнергетическими процессами: учебник для вузов. – М.: Энергоатомиздат, 1985. – 296 с.
 22. Сейдж П., Мелса Дж. Идентификация систем управления: пер. с англ. – М.: Наука, 1974. – 248 с.
 23. Спиди К., Браун Р., Гудвин Дж. Теория управления. Идентификация и оптимальное управление: пер. с англ. – М.: Мир, 1973. – 248 с.
 24. Справочник по теории автоматического управления / под ред. А.А. Красовского. – М.: Наука, 1987. – 712 с.
 25. Справочник проектировщика АСУ ТП / под ред. Г.Л. Смилянского. – М.: Машиностроение, 1983. – 527 с.
 26. Солодовников В.В., Семенов В.В. Спектральная теория нестационарных систем управления. – М.: Наука, 1974. – 335 с.
 27. Тихонов А.Н., Арсенин В.Я. Методы решения некорректных задач. – М.: Наука, 1974. – 224 с.
 28. Функциональный контроль и диагностика электротехнических и электромеханических систем и устройств по цифровым отсчетам мгновенных значений тока и напряжения / под ред. Е.И. Гольдштейна. – Томск: Печатная мануфактура, 2003. – 240 с.
 29. Химмельблау Д. Обнаружение и диагностика неполадок в химических и нефтехимических процессах: пер. с англ. – Л.: Химия, 1983. – 352 с.
 30. Эйкхофф П. Основы идентификации систем управления: пер. с англ. – М.: Мир, 1975. – 684 с.
 31. Эрриот Н. Регулирование производственных процессов: пер. с англ. – М.: Энергия, 1967. – 480 с.

32. Юсупов Р.М. Элементы теории идентификации технических объектов. – Л.: Изд-во Министерства обороны, 1974. – 202 с.
33. Яковлев В.Б. Мои воспоминания. – СПб.: Изд-во СПбГЭТУ «ЛЭТИ», 2005. – 354 с.

ОГЛАВЛЕНИЕ

ПРЕДИСЛОВИЕ.....	3
1. ТЕРМИНОЛОГИЯ.....	3
2. КЛАССИФИКАЦИЯ МЕТОДОВ ИДЕНТИФИКАЦИИ	5
3. ВЫДЕЛЕНИЕ ОБЪЕКТА МОДЕЛИРОВАНИЯ В ПРОСТРАНСТВЕ ПАРАМЕТРОВ	7
4. МАТЕМАТИЧЕСКИЕ МОДЕЛИ ОБЪЕКТОВ ИДЕНТИФИКАЦИИ	11
5. МОДЕЛИ ДИНАМИЧЕСКИХ НЕЛИНЕЙНЫХ СИСТЕМ.....	22
6. МОДЕЛИ ДИНАМИЧЕСКИХ НЕСТАЦИОНАРНЫХ СИСТЕМ.....	24
7. АНАЛИТИЧЕСКИЙ ПОДХОД К ПОСТРОЕНИЮ МОДЕЛЕЙ	27
8. ИДЕНТИФИКАЦИЯ ДИНАМИЧЕСКИХ ОБЪЕКТОВ МЕТОДАМИ АКТИВНОГО ЭКСПЕРИМЕНТА	35
8.1. Идентификация с помощью переходной функции	35
8.2. Частотный метод идентификации	45
9. ХАРАКТЕРИСТИКИ СЛУЧАЙНЫХ ПРОЦЕССОВ.....	48
10. ПРЕДВАРИТЕЛЬНАЯ ОБРАБОТКА ЭКСПЕРИМЕНТАЛЬНЫХ ДАННЫХ	54
10.1. Дискретизация сигналов.....	54
10.2. Фильтрация сигналов	56
11. ОЦЕНКА ПАРАМЕТРОВ СТАТИЧЕСКИХ МОДЕЛЕЙ.....	60
11.1. Основные свойства оценок параметров.....	61
11.2. Оценка параметров методом наименьших квадратов (МНК)	61
11.3. Взвешенный метод наименьших квадратов	67
11.4. Рекуррентный метод наименьших квадратов	68
11.5. Текущий метод наименьших квадратов	71
11.6. Оценка параметров методом наименьших произведений	71
11.7. Оценка параметров модели методом максимального правдоподобия	74
12. РЕГРЕССИОННЫЙ АНАЛИЗ.....	79
12.1. Метод регрессионного анализа	79
12.2. Трансцендентная регрессия	87
12.3. Получение уравнения множественной регрессии методом Брандона	88

13. ПЛАНИРОВАНИЕ ЭКСПЕРИМЕНТА.....	90
14. УРАВНЕНИЕ ВИНЕРА–ХОПФА	99
15. РЕШЕНИЕ УРАВНЕНИЯ ВИНЕРА–ХОПФА.....	102
16. ПРОБЛЕМА НЕКОРРЕКТНОСТИ ПРИ РЕШЕНИИ УРАВНЕНИЯ СТАТИСТИЧЕСКОЙ ИДЕНТИФИКАЦИИ	106
17. УРАВНЕНИЕ СТАТИСТИЧЕСКОЙ ИДЕНТИФИКАЦИИ В ЧАСТОТНОЙ ОБЛАСТИ.....	111
18. ФИЛЬТР КАЛМАНА–БЬЮСИ.....	118
19. ПАРАМЕТРИЧЕСКАЯ ИДЕНТИФИКАЦИЯ ДИНАМИЧЕСКИХ ОБЪЕКТОВ МЕТОДАМИ ПАССИВНОГО ЭКСПЕРИМЕНТА	127
19.1. ОПРЕДЕЛЕНИЕ ПАРАМЕТРОВ ДИФФЕРЕНЦИАЛЬНОГО УРАВНЕНИЯ	127
19.2. ОПРЕДЕЛЕНИЕ ПАРАМЕТРОВ РАЗНОСТНЫХ УРАВНЕНИЙ	129
19.3. ОПРЕДЕЛЕНИЕ ПАРАМЕТРОВ ПЕРЕДАТОЧНОЙ ФУНКЦИИ	130
С ИСПОЛЬЗОВАНИЕМ АППРОКСИМАЦИИ СИГНАЛОВ ОРТОГОНАЛЬНЫМИ ПОЛИНОМАМИ	130
20. АДАПТИВНЫЕ АЛГОРИТМЫ ИДЕНТИФИКАЦИИ.....	132
21. ДИАГНОСТИКА СИСТЕМ	143
21.1. ОБЩИЕ ПОНЯТИЯ И ОПРЕДЕЛЕНИЯ.....	143
21.2. ВИДЫ НЕИСПРАВНОСТИ ТЕХНИЧЕСКИХ СИСТЕМ	145
21.3. ДИАГНОСТИЧЕСКИЕ МОДЕЛИ И ПРОЦЕДУРА ДИАГНОСТИКИ	147
21.4. ПРИМЕНЕНИЕ АППАРАТА ПРОВЕРКИ СТАТИСТИЧЕСКИХ ГИПОТЕЗ В ДИАГНОСТИКЕ.....	149
21.5. ДИАГНОСТИКА ДИСКРЕТНЫХ УСТРОЙСТВ	152
21.6. КОНТРОЛЬ И ДИАГНОСТИКА НАСОСНЫХ УСТАНОВОК	154
В НЕФТЕГАЗОДОБЫЧЕ.....	154
ЗАКЛЮЧЕНИЕ	157
СПИСОК ЛИТЕРАТУРЫ	158

Учебное издание

КОНОВАЛОВ Виктор Иванович

ИДЕНТИФИКАЦИЯ И ДИАГНОСТИКА СИСТЕМ

Учебное пособие

Научный редактор *доктор технических наук,*
профессор Г.П. Цапко

Редактор *О.Н. Свинцова*

Компьютерная верстка *В.И. Коновалов*

Дизайн обложки *Т.А. Фатеева*

Подписано к печати 05.11.2010. Формат 60х84/16. Бумага «Снегурочка».

Печать XEROX. Усл.печ.л. 9,01. Уч.-изд.л. 8,16.

Заказ . Тираж 100 экз.



Национальный исследовательский Томский политехнический университет
Система менеджмента качества

Томского политехнического университета сертифицирована
NATIONAL QUALITY ASSURANCE по стандарту ISO 9001:2008



ИЗДАТЕЛЬСТВО



ТПУ. 634050, г. Томск, пр. Ленина, 30
Тел./факс: 8(3822)56-35-35, www.tpu.ru