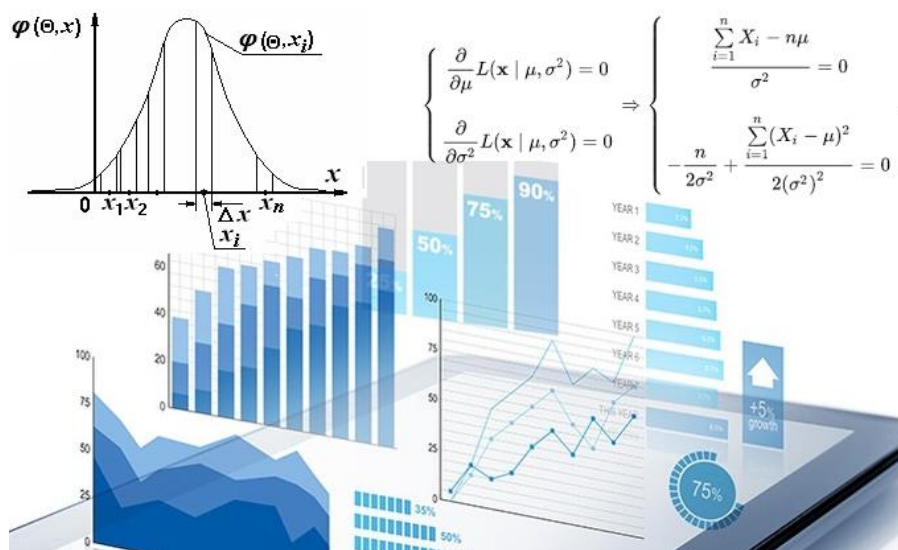


ИТМО

А.Ю. Краснов

СТАТИСТИЧЕСКИЕ МЕТОДЫ В ИНЖЕНЕРНЫХ ИССЛЕДОВАНИЯХ



Санкт-Петербург
2022

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ
ФЕДЕРАЦИИ

УНИВЕРСИТЕТ ИТМО

А.Ю. Краснов
СТАТИСТИЧЕСКИЕ МЕТОДЫ В ИНЖЕНЕРНЫХ
ИССЛЕДОВАНИЯХ

УЧЕБНО-МЕТОДИЧЕСКОЕ ПОСОБИЕ

РЕКОМЕНДОВАНО К ИСПОЛЬЗОВАНИЮ В УНИВЕРСИТЕТЕ ИТМО
по направлению подготовки 15.03.06, 27.03.04
в качестве Учебно-методического пособия для реализации основных
профессиональных образовательных программ высшего образования
бакалавриата

ИТМО

Санкт-Петербург
2022

Краснов А.Ю., Статистические методы в инженерных исследованиях – СПб: Университет ИТМО, 2022. – 119 с.

Рецензент(ы):

Николаев Николай Анатольевич, кандидат технических наук, доцент ф-та СУиР, Университета ИТМО.

Учебно-методическое пособие по дисциплине «Статистические методы в инженерных исследованиях» содержит изложение основ теории случайных процессов и статистической обработки данных применительно к системам управления, элементы теории марковских процессов, математических методов планирования эксперимента, элементы теории прогнозирования с подробным рассмотрением процедуры синтеза фильтра Калмана.

Учебное пособие рассчитано на студентов направлений Робототехника, Интеллектуальная робототехника и Управление в технических системах, однако оно может быть рекомендовано также аспирантам и молодым специалистам, которым по роду своей образовательной и практической деятельности приходится иметь дело с проблемами планирования и обработки результатов экспериментов, задачами прогнозирования управления стохастическими процессами или системами.



Университет ИТМО – национальный исследовательский университет, ведущий вуз России в области информационных, фотонных и биохимических технологий. Альма-матер победителей международных соревнований по программированию – ICPC (единственный в мире семикратный чемпион), Google Code Jam, Facebook Hacker Cup, Яндекс.Алгоритм, Russian Code Cup, Topcoder Open и др. Приоритетные направления: IT, фотоника, робототехника, квантовые коммуникации, трансляционная медицина, Life Sciences, Art&Science, Science Communication. Входит в ТОП-100 по направлению «Автоматизация и управление» Шанхайского предметного рейтинга (ARWU) и занимает 74 место в мире в британском предметном рейтинге QS по компьютерным наукам (Computer Science and Information Systems). С 2013 по 2020 гг. – лидер Проекта 5–100.

© Университет ИТМО, 2022

© Краснов А.Ю., 2022

Содержание

1	Введение	7
2	Случайные процессы	10
2.1	Математическое описание случайных процессов	11
2.2	Одномерные функции распределения и плотности вероятности случайных процессов	12
2.3	Двумерные функции распределения и плотности вероятности случайных процессов	13
2.4	Моментные характеристики случайных процессов	14
2.4.1	Начальные моментные функции	14
2.4.2	Центральные моментные функции	15
2.4.3	Математическое ожидание СП	15
2.4.4	Ковариационная и корреляционная функции СП	16
2.4.5	Взаимные функции распределения и плотности вероятностей СП	17
2.5	Некоторые классы СП	17
2.5.1	СП с независимыми значениями	18
2.5.2	СП с независимыми приращениями	18
2.5.3	Марковские СП	18
2.5.4	Стационарные СП	19
2.5.5	СП с некоррелированными приращениями	19
2.5.6	Гауссовские (нормальные) СП	20
2.5.7	Эргодические СП	20
3	Стохастические системы	22
3.1	Функциональные схемы стохастических систем управления	22
4	Математические модели стохастических динамических систем	27
4.1	Модели непрерывных стохастических динамических систем	27
4.2	Линейные стохастические системы	28
4.3	Линейные непрерывные стохастические системы со случайной структурой	30
4.4	Дискретные модели стохастических систем	31
4.5	Марковская модель чередования структурных состояний системы	32

4.6	Полумарковская модель чередования режимов в системах со случайной структурой	33
4.7	Модель динамики технических состояний системы с возможными нарушениями	34
5	Марковские процессы	37
5.1	Цепи Маркова	38
5.2	Полумарковские процессы	41
6	Датчики физических величин в системах управления	49
6.1	Методы статистического оценивания в исследовании датчиков систем управления	50
7	Введение в теорию эксперимента	59
7.1	Полный факторный эксперимент	63
7.2	Дробный факторный эксперимент	68
7.3	Обработка результатов многофакторного эксперимента	72
8	Методы прогнозирования. Фильтр Калмана	82
8.1	Методы и системы прогнозирования	82
8.2	Классификация методов прогнозирования	83
8.3	Прогнозирование как оценка состояния динамической системы. Теория фильтрации	103
8.4	Качественные показатели прогноза	105
8.5	Расчет фильтра Калмана	106

Предисловие

В учебном пособии представлены некоторые аспекты применения методов теории вероятностей математической статистики и теории случайных процессов в инженерных задачах. В первом разделе приведены базовые сведения о случайных процессах, их типах и статистических характеристиках. Во втором и третьем разделах пособия приведены примеры различных технических систем, поведение которых может характеризоваться как стохастическое, а также рассмотрены основные классы моделей таких систем (с использованием понятийного аппарата теории автоматического управления). В четвертом разделе пособия приведены теоретические сведения о марковских и полумарковских процессах и выведены их основные статистические характеристики. В пятом разделе пособия разбирается задача построения математических моделей выходных характеристик датчиков с применением аппарата регрессионного анализа и концепции правдоподобия. В шестом разделе пособия рассмотрены основы теории планирования эксперимента и последующей обработки результатов эксперимента с применением методов математической статистики. В седьмом разделе пособия приведены теоретические сведения о методах прогнозирования и процедура синтеза фильтра Калмана.

Учебное пособие рассчитано на студентов направлений Робототехника, Интеллектуальная робототехника и Управление в технических системах, однако оно может быть рекомендовано также аспирантам и молодым специалистам, которым по роду своей образовательной и практической деятельности приходится иметь дело с проблемами статистического анализа экспериментальных данных, расчета параметров марковских процессов и управления стохастическими системами и процессами.

На основе настоящего пособия может быть построен курс лекций по дисциплинам, связанным с применением статистических методов в инженерных исследованиях, а также более узкоспециальных дисциплин, таких как планирование и обработка результатов экспериментов. При построении курса лекций могут быть использованы лишь некоторые разделы настоящего пособия в зависимости от количества лекционных часов, при этом разделы, не включенные в лекционный курс, рекомендуется включить в самостоятельную работу студента для сохранения целостности восприятия материала. В качестве заданий для практических занятий рекомендуется использовать расчетные работы, охватывающие материал одного конкретного раздела дисциплины, например, расчет параметров преобразовательной характеристики датчика (см. раздел 6.1) или расчет параметров фильтра Калмана (см. раздел 8.5). В качестве самостоятельной работы студента рекомендуется изучение литературы с более подробным изложением положений теории случайных процессов [1, 2]. Для направления

подготовки "Управление в технических системах" также рекомендуется в качестве самостоятельной работы студента более подробное изучение пособия по стохастическим системам управления [5].

1 Введение

Важнейшим понятием кибернетики и теории систем является понятие системы. Система представляет собой совокупность взаимосвязанных элементов, выделенных из внешней среды, взаимодействующих между собой и со средой, имеющих общую цель функционирования и заданную либо изменяющуюся во времени структуру.

Элемент – составная часть. Разделение системы возможно по различным признакам и до различных уровней. Следует сразу различать функциональные либо конструктивные реальные элементы системы и элементы математической модели системы.

В теории управления и автоматике в качестве элементов моделей систем выступают типовые динамические звенья, а также звенья, представленные безинерционными нелинейными зависимостями.

При рассмотрении надежности систем в качестве элемента системы выбирают такую часть системы, которую далее не имеет смысла разделять при локализации отказов и их устранении.

Подсистема – часть системы, меньшая, чем сама система, но большая, чем элемент. Подсистемы выделяются по функциональному, конструктивному либо иным признакам.

Структура системы определяется составом элементов и связями между ними. Связи характеризуются направлением, интенсивностью и природой.

Цель системы – наиболее качественное функционирование системы (когда система работает наилучшим образом).

В любой системе управления (СУ) прежде всего выделяют объект управления (ОУ) и устройство управления (УУ).

В технике ОУ обычно является некоторое техническое устройство либо физическая система.

Управление – это целенаправленное воздействие на ОУ для достижения требуемого качества его функционирования.

Любой процесс управления содержит в себе следующие основные компоненты:

- формирование цели управления;
- получение информации о текущем состоянии ОУ;
- сравнение текущего состояния ОУ с требуемым и выработка управляющих воздействий;
- непосредственное приложение сформированных управляющих воздействий ко входам ОУ.

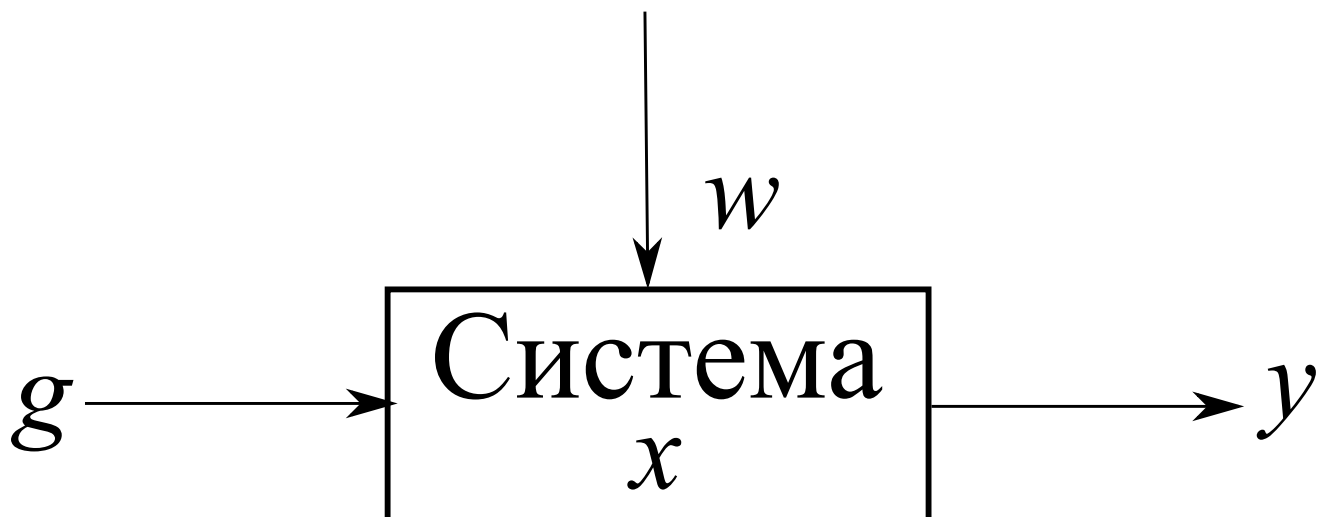


Рис. 1: Схематическое изображение модели явления (системы).

Научные модели и методы исследований, используемые в инженерной практике, разделяют на детерминированные и вероятностные, стохастические, статистические.

В природе и в функционировании технических систем есть явления, протекающие в устойчивых формах: законы движения планет, основные законы физики и т.п.

Другая группа явлений, отличная от детерминированных, характеризуется существенным влиянием множества случайных факторов. Эти явления должны изучаться статистическими методами на основе вероятностных моделей.

Модель явления (системы) есть упрощенное представление процессов и связей в данном явлении (системе), представленное в удобном для исследования и анализа форме. Модель, описанная средствами математического аппарата, называется математической моделью, при этом вектор состояния системы, как правило, обозначают x , выходной сигнал – y , задающее воздействие – g , а возмущающее воздействие – w (см. рисунок 1).

В детерминированных системах g и w представляются детерминированными (однозначно определяющимися значением аргумента) функциями времени. Для исследования детерминированных систем используют детерминированные методы и подходы. Описание детерминированных систем представляется детерминированными операторами, а начальные условия системы задаются численными значениями. Реакция детерминированных систем определяется однозначно.

В стохастических системах в операторах описания системы содержатся элементы с вероятностными свойствами. Задающее и возмущающее воздействия представляются в виде случайных процессов. Для начальных условий задают-

ся их статистические распределения. В описанных условиях реакция системы определяется с точностью до ее вероятностных характеристик. Статистические методы требуют при их применении большего объема описания системы и применения более изощренного математического аппарата.

При использовании вероятностных методов исследования систем достигается более естественное описание системы, полнее отражающее свойства системы и условия ее функционирования, удастся установить особенности системы, не учитываемые в детерминированной модели, на основе вероятностных методов удастся синтезировать УУ, обеспечивающие системе лучшие динамические характеристики.

В регуляторах для стохастических систем при их синтезе появляются дополнительные функциональные блоки.

Использование вероятностных моделей и методов позволяет учесть возможное случайное изменение структуры, режимов работы и целей функционирования системы.

Области статистических методов в инженерных исследованиях:

- статистическая теория разработки, синтеза и оценки качества систем;
- статистическая теория измерений и обработки измерительной информации;
- статистическая теория оценивания;
- статистические методы в теории надежности;
- статистические методы систем радиоавтоматики.

2 Случайные процессы

Процесс – изменение во времени некоторой (например, физической) величины. Процесс – последовательность смены событий (эволюция состояний некоторой системы).

Говоря о случайных процессах, выделяют множество задания процесса и множество его значений. $t \in T$ – некоторый скалярный либо векторный параметр, T – множество заданий случайного процесса, ξ – множество значений случайного процесса. Если t – векторный параметр, то случайная функция называется случайным полем.

Случайная функция есть измеримое отображение Ω в R^n , зависящее от параметра t : $\Omega \xrightarrow{x(t,\omega)} R^n$, где $\omega \in \Omega$ – элементарный исход.

Если параметр $t \in R^1$ и по своему физическому смыслу является временем, то такие случайные функции называют случайными процессами.

Случайные процессы в теории управления и автоматике используют при моделировании действия внешних случайных возмущений на объект управления и отражения непредсказуемых и хаотичных изменений в устройстве самого объекта управления.

К случайным процессам относят такие процессы, поведение которых до начала функционирования системы предсказать невозможно.

При выполнении многих наблюдений за функционированием системы накапливается информация о свойствах системы. Появляется возможность вычислять и предсказывать вероятности некоторых событий, но поведение системы может быть охарактеризовано только в вероятностном смысле (можно предсказать лишь степени возможности наступления событий).

Не всякий процесс с неопределенным поведением может быть отнесен к случайным процессам. Для того, чтобы процесс обоснованно мог быть отнесен к классу случайных, должны выполняться условия

- о характере реализации процесса в предстоящем наблюдении можно лишь строить предположения (однозначного суждения до начала эксперимента не может быть сформировано принципиально);
- предполагается, что количество наблюдений за системой сколь угодно большое;
- случайные величины (СВ), получаемые из случайного процесса при фиксации моментов времени, должны обладать свойством стохастической устойчивости.

Случайный процесс (СП) может быть рассмотрен с различных точек зрения:

- с точки зрения математики случайный процесс является математической моделью системы, развивающейся во времени;
- с физической точки зрения важны физические свойства процесса (броуновское движение, винеровский процесс);
- с точки зрения теории систем случайный процесс может быть использован как модель влияния случайных возмущений на систему;
- с точки зрения теории надежности случайный процесс может служить аналитическим описанием некоторых свойств технических систем.

Проводя наблюдение за случайным процессом, фиксируют реализации процесса, которые являются детерминированными функциями времени. Реализации называют также траекториями или выборочными функциями случайного процесса. Множество выборочных функций случайного процесса с заданной вероятностной мерой их появления называется ансамблем и обозначается

$$x_t = \{x(t, \omega), t \in T, \omega \in \Omega\}.$$

Будем связывать элементарное событие ω_i с появлением i -ой реализации. Выборочные функции принадлежат множеству выборочных функций $\mathfrak{L}$.

Случайный процесс есть функция двух аргументов: времени t и элементарного события ω . При фиксации момента времени t получают случайную величину, которую называют сечением случайного процесса. При фиксации элементарного события ω получают выборочную функцию (детерминированную функцию времени).

2.1 Математическое описание случайных процессов

Считается, что случайный процесс описан, если установлены статистические характеристики, достаточно полно отражающие свойства процесса. В силу того, что случайный процесс может быть представлен множеством его выборочных функций, его полное описание предполагает перечисление всех выборочных функций с указанием вероятностей появления каждой из них. Такое полное описание крайне редко достижимо, кроме того, в нем нет практической необходимости.

С другой стороны, случайный процесс при фиксации момента времени можно представить как случайную величину, а при фиксации нескольких моментов времени – систему случайных величин. Тогда для описания случайного процесса можно использовать те же аналитические средства, что и для описания случайных величин и случайных векторов [1, 2]:

- функции распределения;
- плотности вероятности;
- моментные характеристики;
- характеристические функции;
- другие средства аналитического описания случайных величин.

Случайный процесс считают описанным, если известен закон распределения системы случайных величин $x(t_1), x(t_2), \dots, x(t_n)$ для любых n произвольных моментов времени. При $n \rightarrow \infty$ достигается описание случайного процесса, эквивалентное полному описанию.

2.2 Одномерные функции распределения и плотности вероятности случайных процессов

Выделим некоторый момент времени t_1 , зафиксируем значение уровня ξ_1 . Выберем достаточно большое число реализаций N , из них выберем $n_1(t_1, \xi_1)$ реализаций, которые в момент времени t_1 прошли ниже уровня ξ_1 :

$$\left. \frac{n_1(t_1, \xi_1)}{N} \right|_{N \rightarrow \infty} \approx F_1(t_1, \xi_1) = P\{x(t_1) < \xi_1\}, \quad (1)$$

где $F_1(t_1, \xi_1)$ – одномерная функция распределения вероятностей случайного процесса.

$$f_1(t_1, \xi_1) = \frac{\partial F_1(t_1, \xi_1)}{\partial \xi_1}.$$

$$F_1(t_1, \xi_1) = \int_{-\infty}^{\xi_1} f_1(u_1, t_1) du_1.$$

$$f_1(t_1, \xi_1) d\xi_1 = P\{\xi_1 < x_1(t) \leq \xi_1 + d\xi_1\} - \text{элементарная вероятность.}$$

Функции $F_1(t_1, \xi_1)$ и $f_1(t_1, \xi_1)$ описывают поведение случайного процесса лишь в фиксированный момент времени, но этот момент фиксируется произвольным образом.

Описание случайного процесса плотностями вероятностей физически более наглядно, чем функциями распределения. Однако с помощью функций распределения можно описывать более широкий класс случайных процессов (дискретные, непрерывные, смешанные).

2.3 Двумерные функции распределения и плотности вероятности случайных процессов

Функции $F_1(t_1, \xi_1)$ и $f_1(t_1, \xi_1)$ описывают статистические свойства процесса только в один фиксированный момент времени. Для получения более полного описания процесса зафиксируем два момента времени t_1 и t_2 , выберем N реализаций случайного процесса и из их числа выберем n_2 реализаций, которые в момент времени t_1 прошли ниже уровня ξ_1 , а в момент времени t_2 прошли ниже уровня ξ_2 :

$$\left. \frac{n_2(t_1, \xi_1; t_2, \xi_2)}{N} \right|_{N \rightarrow \infty} \approx F_2(t_1, \xi_1; t_2, \xi_2) = P\{x(t_1) < \xi_1, x(t_2) < \xi_2\}; \quad (2)$$

$$f_2(t_1, \xi_1; t_2, \xi_2) = \frac{\partial^2}{\partial \xi_1 \partial \xi_2} F_2(t_1, \xi_1; t_2, \xi_2).$$

$f_2(t_1, \xi_1; t_2, \xi_2) d\xi_1 d\xi_2$ – элементарная вероятность.

Для процессов с независимыми значениями (например, гауссовских или марковских) через одно- и двумерные плотности вероятности можно выразить все остальные. Для m -мерного случая

$$\left. \frac{n_m(t_1, \xi_1; t_2, \xi_2; \dots; t_m, \xi_m)}{N} \right|_{N \rightarrow \infty} \approx F_m(t_1, \xi_1; t_2, \xi_2; \dots; t_m, \xi_m) = P\{x(t_1) < \xi_1, x(t_2) < \xi_2, \dots, x(t_m) < \xi_m\}; \quad (3)$$

$$f_m(t_1, \xi_1; t_2, \xi_2; \dots; t_m, \xi_m) = \frac{\partial^m}{\partial \xi_1 \partial \xi_2 \dots \partial \xi_m} F_m(t_1, \xi_1; t_2, \xi_2; \dots; t_m, \xi_m);$$

$f_m(t_1, \xi_1; t_2, \xi_2; \dots; t_m, \xi_m) d\xi_1 d\xi_2 \dots d\xi_m$ – элементарная вероятность.

Функции распределения вероятностей случайных процессов неотрицательны, не изменяются при любых перестановках их аргументов и обладают следующими свойствами:

- $F_1(-\infty, t_1) = P\{x(t_1) = -\infty\} = 0;$
- $F_1(\infty, t_1) = P\{x(t_1) = \infty\} = 1;$
- $F_1(\xi_1'', t_1) \geq F_1(\xi_1', t_1)$ при $\xi_1'' > \xi_1';$
- $P\{\xi_1' < x(t_1) \leq \xi_1''\} = F_1(\xi_1'') - F_1(\xi_1') = \int_{\xi_1'}^{\xi_1''} f_1(\xi_1, t_1) d\xi_1;$
- $P\{x(t_1) < \xi_1; x(t_1) < \infty\} = F_2(\xi_1, t_1; \infty, t_2) = F_1(\xi_1, t_1) = \int_{-\infty}^{\xi_1} \int_{-\infty}^{\infty} f_2(u_1, t_1; \xi_2, t_2) du_1 d\xi_2.$

Плотность вероятности случайного процесса обладает следующими свойствами:

- $f_n(\xi_1, t_1; \dots; \xi_n, t_n) \geq 0$;
- $\int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f_n(\xi_1, t_1; \dots; \xi_n, t_n) d\xi_1 \dots d\xi_n = 1$;
- $f_n(\xi_1, t_1; \dots; \xi_n, t_n) = f_n(\xi_n, t_n; \dots; \xi_1, t_1)$;
- $f_m(\xi_1, t_1; \dots; \xi_m, t_m) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f_n(\xi_1, t_1; \dots; \xi_m, t_m; \xi_{m+1}, t_{m+1}; \dots; \xi_n, t_n) d\xi_{m+1} \dots d\xi_n$

Из плотности распределения вероятностей n -го порядка можно получить любую плотность распределения вероятностей меньшей мерности интегрированием n -мерной плотности распределения вероятностей по "лишним" аргументам.

2.4 Моментные характеристики случайных процессов

- математическое ожидание;
- ковариационная и корреляционная функции;
- дисперсия;
- характеристические функции.

Математическое ожидание нелинейного преобразования случайного процесса для случайной величины x с $f(\xi)$

$$M[\varphi(x)] = \int_{-\infty}^{\infty} \varphi(\xi) f(\xi) d\xi;$$

$$M[\varphi(x_t, t)] = \int_{-\infty}^{\infty} \varphi(\xi, t) f(\xi, t) dt.$$

2.4.1 Начальные моментные функции

Начальные моментные функции определяются как функции, симметричные относительно всех своих аргументов и равные математическим ожиданиям соответствующих произведений. Одномерный начальный момент порядка ν_1 рассчитывается как

$$\alpha_{\nu_1}(t_1) = M[x^{\nu_1}(t_1)] = \int_{-\infty}^{\infty} \xi_1^{\nu_1} f_1(\xi_1, t_1) d\xi_1.$$

Двухмерный начальный момент порядка $(\nu_1 + \nu_2)$ рассчитывается как

$$\alpha_{\nu_1 \nu_2}(t_1, t_2) = M[x^{\nu_1}(t_1) \cdot x^{\nu_2}(t_2)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \xi_1^{\nu_1} \xi_2^{\nu_2} f_2(\xi_1, t_1; \xi_2, t_2) d\xi_1 d\xi_2.$$

n -мерный начальный момент порядка $(\nu_1 + \nu_2 + \dots + \nu_n)$ рассчитывается как

$$\begin{aligned} \alpha_{\nu_1 \nu_2 \dots \nu_n}(t_1, t_2, \dots, t_n) &= M[x^{\nu_1}(t_1) \cdot x^{\nu_2}(t_2) \cdot \dots \cdot x^{\nu_n}(t_n)] = \\ &= \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \xi_1^{\nu_1} \xi_2^{\nu_2} \dots \xi_n^{\nu_n} f_n(\xi_1, t_1; \xi_2, t_2; \dots; \xi_n, t_n) d\xi_1 d\xi_2 \dots d\xi_n. \end{aligned}$$

2.4.2 Центральные моментные функции

Центральные моментные функции определяются по аналогии с начальными моментными функциями, но вместо $x(t_n)^{\nu_n}$ используется центрированное значение $(x(t_n) - m_x(t_n))^{\nu_n}$, где $m_x(t)$ – математическое ожидание СП (см. раздел 2.4.3).

2.4.3 Математическое ожидание СП

Математическое ожидание $m_x(t)$ СП есть неслучайная функция времени, получаемая осреднением в каждый момент времени СВ, получаемой в сечении СП, определяемая выражением

$$m_x(t) = M[x_t] = \int_{-\infty}^{\infty} \xi f(\xi, t) d\xi, \quad (4)$$

характеризующая среднее течение процесса.

Математическое ожидание СП обладает следующими свойствами:

- математическое ожидание детерминированной функции $g(t)$ равно самой функции:

$$M[g(t)] = g(t);$$

- математическое ожидание суммы (разности) двух взаимно независимых СП x_t и y_t равно сумме (разности) математических ожиданий этих СП:

$$M[x_t \pm y_t] = M[x_t] \pm M[y_t];$$

- математическое ожидание произведения СП x_t и детерминированной функции $g(t)$ равно произведению детерминированной функции на математическое ожидание СП x_t :

$$M[g(t)x_t] = g(t)M[x_t];$$

- Размерность математического ожидания СП x_t совпадает с размерностью самого СП x_t .

2.4.4 Ковариационная и корреляционная функции СП

Центрированная СВ определяется выражением

$$\overset{\circ}{x}(t) = x_t - m_x(t).$$

Ковариационная функция СП вычисляется по формуле

$$K_x(t_1, t_2) = M[\overset{\circ}{x}(t_1)\overset{\circ}{x}(t_2)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (\xi_1 - m_x(t_1))(\xi_2 - m_x(t_2)) f_2(\xi_1, t_1; \xi_2, t_2) d\xi_1 d\xi_2. \quad (5)$$

Свойства ковариации:

- ковариация симметрична;
- дисперсия СВ есть ковариация СВ с самой собой;
- ковариация взаимно независимых величин равна нулю;
- $D(x + -y) = Dx + Dy + -2Kxy$;
- множитель выносится за знак ковариации;
- можно прибавить к СВ постоянную;
- $|Kxy| < \sigma_x \sigma_y$.

Корреляционная функция СП вычисляется по формуле

$$R_x = \frac{K_x(t_1, t_2)}{\sigma_x(t_1)\sigma_x(t_2)}. \quad (6)$$

Свойства корреляции:

- $|R_x| < 1$;

- корреляция двух взаимно независимых СВ равна нулю;
- если X и Y связаны линейно $Y = aX + b$ то
 - при $a > 0$ $R_{xy} = 1$;
 - при $a < 0$ $R_{xy} = -1$;
- если $R_{xy} = 1$ то X и Y связаны линейной зависимостью.

2.4.5 Взаимные функции распределения и плотности вероятностей СП

Часто приходится рассматривать два и более СП одновременно:

$$F_2(\xi_1, \eta_1, t_1) = P\{x(t_1) < \xi_1; y(t_1) < \eta_1\} = F_2(\xi_1, t_1; \eta_1, t_1).$$

Если процессы статистически связаны, то

$$F_2(\xi_1, \eta_1, t_1) = P\{x(t_1) < \xi_1\} P\{y(t_1) < \eta_1 | x(t_1) < \xi_1\} = F_2(\xi_1, t_1; \eta_1, t_1).$$

$$f_2(\xi_1, \eta_1, t_1) d\xi_1 d\eta_1 = P\{\xi_1 < x(t_1) < \xi_1 + d\xi_1; \eta_1 < y(t_1) < \eta_1 + d\eta_1\}.$$

Для отражения взаимной статистической связи используют взаимные функции распределения и плотности вероятностей.

2.5 Некоторые классы СП

Общая математическая модель случайных явлений, развивающихся во времени. Среди случайных явлений можно выделить явления со специфическими свойствами. Для их описания используются СП частного вида.

Последовательность n -мерных плотностей вероятности достаточно полно описывает СП. С увеличением n сложность его аналитического описания очень быстро растет. Однако, существуют классы СП, для которых при $n \geq 3$ n -мерные плотности вероятности выражаются через одно- и двумерные:

- белый шум;
- СП, описываемые корреляционной теорией;
- марковские СП;
- гауссовские СП.

2.5.1 СП с независимыми значениями

СП называется СП с независимыми значениями, если при любом n и любых $t_1, t_2, \dots, t_n$ СВ $x(t_1), x(t_2), \dots, x(t_n)$ статистически независимы в совокупности.

$$F_n(\xi_1, t_1; \xi_2, t_2; \dots; \xi_n, t_n) = \prod_{i=1}^n F_1(\xi_i, t_i).$$
$$f_n(\xi_1, t_1; \xi_2, t_2; \dots; \xi_n, t_n) = \prod_{i=1}^n f_1(\xi_i, t_i).$$

Является простейшим классом СП.

2.5.2 СП с независимыми приращениями

Это СП такой, что при любом n и любых $t_1, t_2, \dots, t_n$ СВ $x(t_n) - x(t_{n-1}), x(t_{n-1}) - x(t_{n-2}), \dots, x(t_2) - x(t_1)$ статистически независимы в совокупности. Если они только некоррелированы, то СП x_t называют СП с некоррелированными приращениями.

2.5.3 Марковские СП

t_n – настоящее, до него – прошлое, после – будущее.

Марковское свойство: СП x_t называется марковским, если его статистические характеристики в момент времени t полностью определяются значением процесса в момент времени t_n и не зависят от того, какие значения процесс принимал в предшествующие t_n моменты времени.

$$P\{x(t) < \xi/x(t_1), x(t_2), \dots, x(t_n)\} = P\{x(t) < \xi/x(t_n)\}.$$

Статистический смысл марковского свойства. Из всех траекторий СП выберем лишь те, которые в момент времени t_n прошли через точку ξ_n, t_n . Из этого множества выберем те траектории СП, которые в момент времени t_1 прошли через точку ξ_1, t_1 и т. д. до ξ_n, t_n . Статистические характеристики СП, определяемые для $t > t_n$ по траекториям из первого множества, и статистические характеристики, определяемые по второму более узкому множеству, примерно совпадают.

$f_n(\xi_1, t_1; \xi_2, t_2; \dots; \xi_n, t_n) = f(\xi_n, t_n/\xi_{n-1}, t_{n-1}) f(\xi_{n-1}, t_{n-1}/\xi_{n-2}, t_{n-2}) \dots f(\xi_2, t_2/\xi_1, t_1) f(\xi_1, t_1)$, где $f(\xi_1, t_1)$ задает начальное распределение значений СП, а $f(\xi_n, t_n/\xi_{n-1}, t_{n-1})$ – условные (переходные) плотности вероятности марковского СП. Для марковского СП условная плотность вероятности зависит лишь от значения СП в предыдущий момент времени.

2.5.4 Стационарные СП

СП называют стационарным, если его статистические характеристики не зависят от времени, т. е. статистические характеристики $x_{t+\tau}$ совпадают с соответствующими статистическими характеристиками x_t при любом τ . Различают СП, стационарные в широком смысле (слабая стационарность) и в узком смысле (сильная стационарность).

СП стационарен в узком смысле, если его n -мерная функция распределения и плотность вероятности при любых n не зависят от временного сдвига τ для любых $t_1, t_2, \dots, t_n$, т. е.

$$F_n(\xi_1, t_1; \xi_2, t_2; \dots; \xi_n, t_n) = F_n(\xi_1, t_1 + \tau; \xi_2, t_2 + \tau; \dots; \xi_n, t_n + \tau),$$

$$f_n(\xi_1, t_1; \xi_2, t_2; \dots; \xi_n, t_n) = f_n(\xi_1, t_1 + \tau; \xi_2, t_2 + \tau; \dots; \xi_n, t_n + \tau).$$

Вероятностные характеристики стационарных СП не зависят от выбора момента начала отсчета.

СП стационарен в широком смысле, если он удовлетворяет следующим требованиям:

- математическое ожидание СП постоянно: $M[x_t] = m_x(t) = \text{const}$;
- ковариационная функция СП зависит только от $\tau = t_2 - t_1$:

$$\begin{aligned} K_{xx}(t_1, t_2) &= K_{xx}(t_1, t_1 + \tau) = M\left[\overset{\circ}{x}(t_1) \overset{\circ}{x}(t_1 + \tau)\right] = \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (\xi_1(t_1) - m_x)(\xi_2(t_1 + \tau) - m_x) f_2(\xi_1, t_1; \xi_2, t_1 + \tau) d\xi_1 d\xi_2. \end{aligned}$$

2.5.5 СП с некоррелированными приращениями

СП с некоррелированными приращениями определяется распределением приращений $x(t) - x(s)$ для любых t и s и распределением $x(t_1)$. Если распределение $x(t) - x(s)$ зависит только от $(t - s)$, то x_t называют СП со стационарными приращениями, или однородным СП. Если случайные величины $x(t) - x(s)$ имеют нормальное распределение, то x_t называют СП с независимыми нормальными приращениями. СП с независимыми приращениями относится к марковским СП.

2.5.6 Гауссовские (нормальные) СП

СП называется нормальным или гауссовским, если при любом n и при любых $t_1, t_2, \dots, t_n$ случайные величины $x(t_1), x(t_2), \dots, x(t_n)$ образуют нормально распределенный случайный вектор.

n -мерная плотность вероятности гауссовского СП определяется выражением

$$f_n(\xi_1, t_1; \xi_2, t_2; \dots; \xi_n, t_n) = \frac{1}{\sqrt{(2\pi)^n \det K}} e^{-\frac{1}{2}(\xi-m)^T K^{-1}(\xi-m)}, \quad (7)$$

где K – матрица ковариаций СП, $\xi = [\xi_1 \ \xi_2 \ \dots \ \xi_n]^T$ и $m = [m_1 \ m_2 \ \dots \ m_n]^T$.

Гауссовские СП обладают следующими свойствами:

- любые статистические характеристики гауссовского процесса могут быть рассчитаны по известному математическому ожиданию $m_x(t)$ и ковариационной функции $K_{xx}(t_1, t_2)$, т. е. эти две характеристики содержат полную информацию о гауссовском СП;
- для гауссовского СП условия стационарности в широком и в узком смысле совпадают;
- многомерные плотности вероятности и функции распределения гауссовского СП являются следствием двумерных;
- при линейных преобразованиях характер закона распределения гауссовского СП сохраняется, меняются лишь его параметры;
- некоррелированность гауссовского процесса означает его статистическую независимость;
- линейные инерционные элементы систем обладают свойством нормализации. Если интервал корреляции гауссовского СП много меньше характерной постоянной времени элемента, то, независимо от статистических свойств входного воздействия, выходной СП элемента будет близок к гауссовскому.

2.5.7 Эргодические СП

СП называется эргодическим, если его вероятностные характеристики, полученные усреднением по ансамблю реализаций, равны характеристикам, получаемым усреднением по времени при обработке одной реализации. Это означает, что поведение каждой реализации эргодического СП статистически эквивалентно статистике всего ансамбля реализаций. Эргодические СП относятся

к классу стационарных в узком смысле, более того, стационарность является необходимым условием эргодичности СП.

Большим преимуществом эргодических СП является то, что все его статистические характеристики могут быть получены по одной реализации. Многие СП, встречающиеся на практике, могут быть отнесены к классу эргодических, однако установить этот факт для реального СП трудно. В теоретических исследованиях свойство эргодичности СП, как правило, принимается гипотетически.

Вопросы для самоконтроля

1. Что такое случайный процесс? По каким критериям процесс с неопределенным поведением можно отнести к случайным?
2. Что такое функции распределения вероятностей и плотности вероятности случайного процесса? Как они задаются?
3. Какие существуют моментные характеристики случайных процессов?
4. Что такое математическое ожидание случайного процесса и каковы его свойства?
5. Что такое ковариационная функция случайного процесса и каковы ее свойства?
6. Что такое корреляционная функция случайного процесса и каковы ее свойства?
7. Какие существуют классы случайных процессов?

3 Стохастические системы

Важным классом стохастических динамических систем являются системы со случайно изменяющейся в процессе функционирования структурой. Структуры стохастических систем очень разнообразны. Ниже приводятся несколько примеров типовых функциональных конфигураций систем со случайной структурой.

3.1 Функциональные схемы стохастических систем управления

Функциональная схема – это условное графическое изображение реальной системы с указанием ее функциональных узлов и функциональных связей между узлами.

Структурная схема системы есть условное графическое изображение математической модели системы (элементов математического описания системы) с указанием аналитических зависимостей (связей) между элементами описания.

Элементы – типовые динамические звенья.

В любой СУ прежде всего выделяют ОУ.

Под ОУ понимается физический объект либо техническая система, оснащенная органами управления – исполнительными устройствами и, возможно, датчиками для измерения контролируемых величин.

УУ служит для формирования управляющих воздействий на ОУ. В нем могут быть различные функциональные блоки, но обязательно есть регулятор (Р).

Выделим отдельно систему измерений. Текущий контроль осуществляется датчиками. При измерениях возникают ошибки измерений.

Задающее устройство (ЗУ).

Из величин, с помощью которых достигается описание системы, прежде всего выделим регулируемые величины (управляемые переменные) – это переменные ОУ, по отношению к которым формируется цель управления.

y – вектор регулируемых величин (по отношению к ним формулируется цель управления);

x – вектор состояния ОУ;

$x(0)$ – начальные условия вектора состояния ОУ (компоненты – случайные величины);

w_t – вектор случайных возмущающих воздействий, действующих на ОУ;

v_t – вектор помех измерений;

$\hat{x}$ – вектор оценок значений переменных вектора состояния ОУ;

z – вектор состояния генератора задающего воздействия;

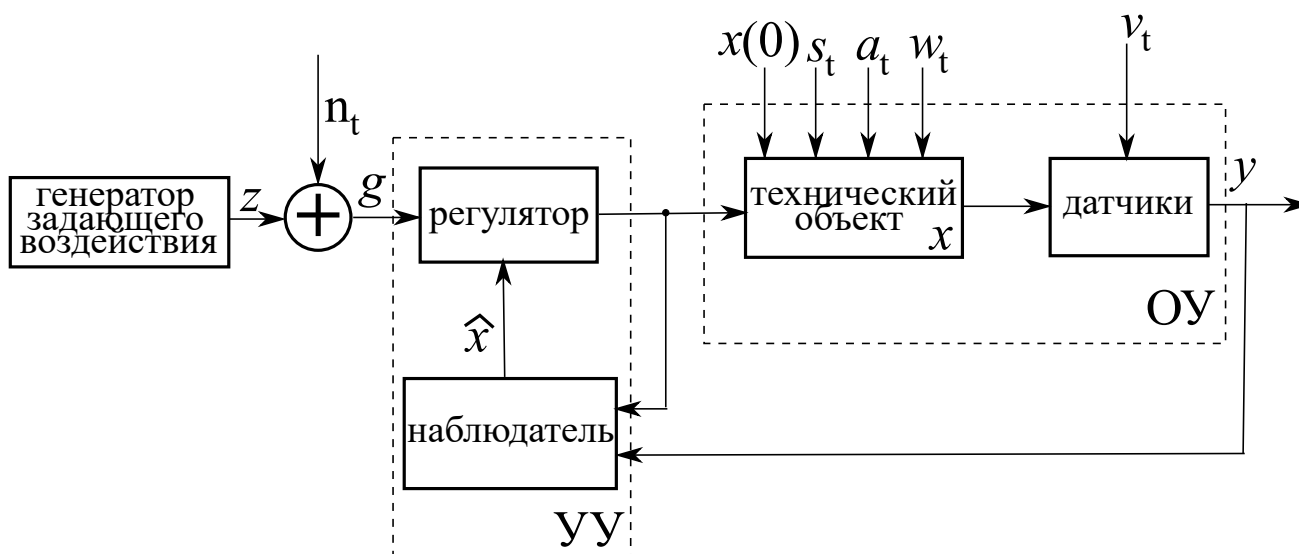


Рис. 2: Схема системы

g – вектор задающего воздействия (входной сигнал);

n_t – случайная составляющая задающего воздействия;

α_t – вектор параметров математической модели процесса функционирования ОУ;

s_t – переменная, отражающая возможность случайного изменения структуры системы в ходе ее функционирования.

Системы с изменяющейся структурой регулятора

В таких системах устройство управления имеет несколько различных регуляторов (см. рисунок 3). На неперекрывающихся интервалах времени (на каждом из таких интервалов) выработка управляющего воздействия осуществляется одним из регуляторов. Выходной сигнал одного из регуляторов подается на вход объекта управления при помощи устройства переключения (УП).

Системы со скачкообразно изменяющимися случайными возмущениями

На такие системы действует внешнее возмущение w_t с резко изменяющимися на случайных интервалах времени статистическими характеристиками. Такую систему при построении ее математической модели целесообразно представить как систему с изменяющейся в случайные моменты времени структурой генератора возмущающих воздействий, где УП осуществляет переключение между различными моделями возмущений (МВ) (см. рисунок 4).

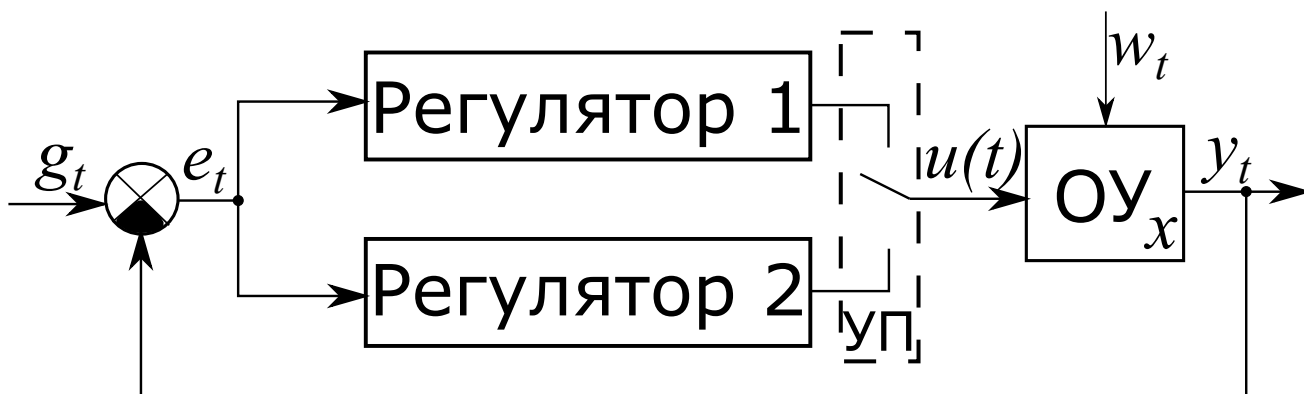


Рис. 3: Система с изменяемой структурой регулятора

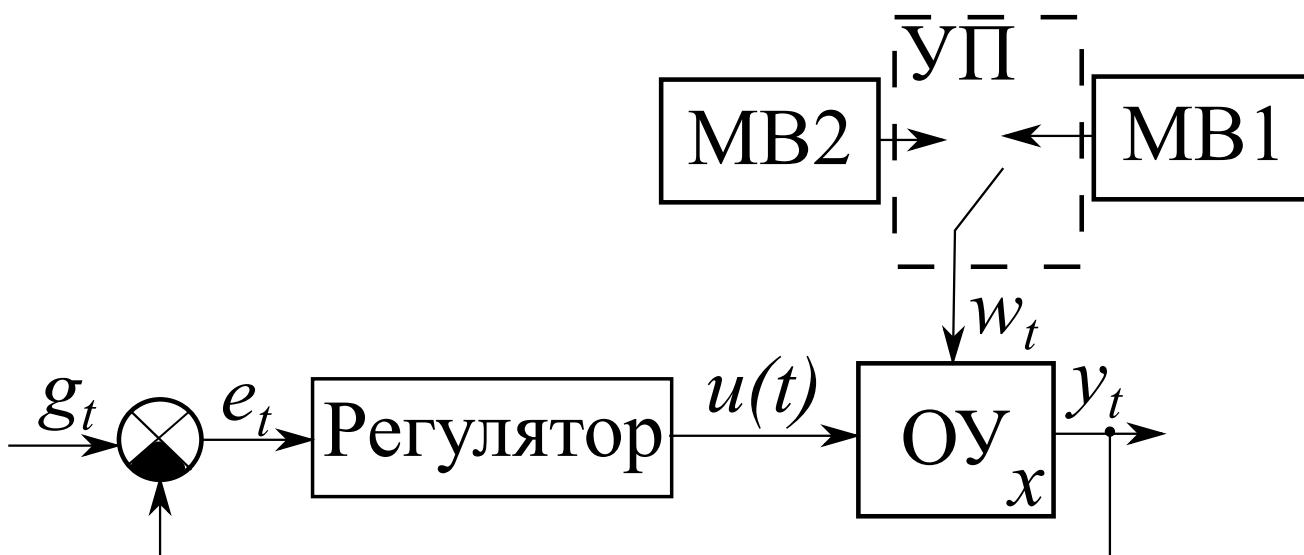


Рис. 4: Система со скачкообразно изменяющимся случайным возмущением

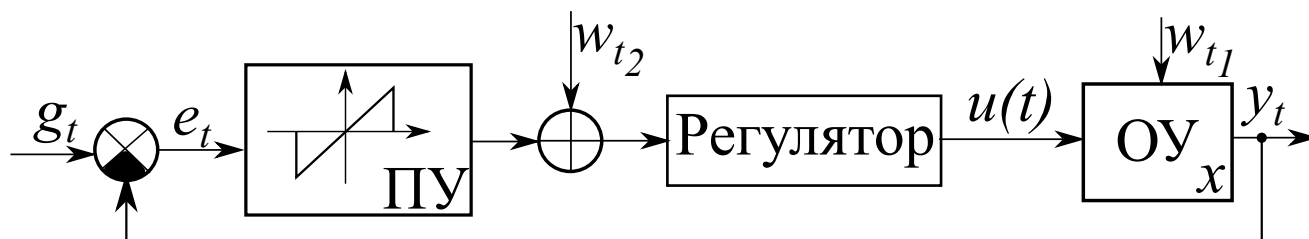


Рис. 5: Система пространственного слежения

В таких системах в явном виде устройство переключения реально отсутствует. Однако в условной схеме системы (и в ее математической модели) удобно ввести фиктивное устройство переключения.

Системы пространственного слежения

Системы пространственного слежения, называемые также системами поиска и захвата, имеют ограниченную апертуру пеленгационных устройств. При сопровождении объекта (в режиме слежения) такая система работает как замкнутая следящая система (см. рисунок 5).

На рисунке 5 ПУ – это пеленгационное устройство, w_{t1} – помехи, наведенные на выходной сигнал пеленгационного устройства, w_{t2} – случайное возмущение, действующее на объект управления.

При срыве слежения (при потере объекта), т.е. при выходе объекта за ограничения апертуры, обратные связи по сигналам рассогласования в системе размыкаются, т.е. система в любой случайный момент времени может скачкообразно изменить свою структуру, а вместе с ней – цели и режимы функционирования.

Системы слежения за маневрирующей целью

В системах пространственного слежения для оценки текущего состояния объекта управления и модели маневрирования цели широко применяются специальные наблюдающие устройства – фильтры Калмана (ФК).

При синтезе фильтра Калмана должна быть известна (и "фиксирована") модель движения цели. Если цель способна применить интенсивный противозахватный маневр, при организации управления системой слежения эту ситуацию необходимо учесть. Для этого строят модель интенсивного маневра цели и синтезируют для нее отдельный фильтр Калмана (см. рисунок 6).

На рисунке 5 ПУ – это пеленгационное устройство, w_{t1} – помехи, наведенные на выходной сигнал пеленгационного устройства, w_{t2} – случайное возмущение, действующее на объект управления.

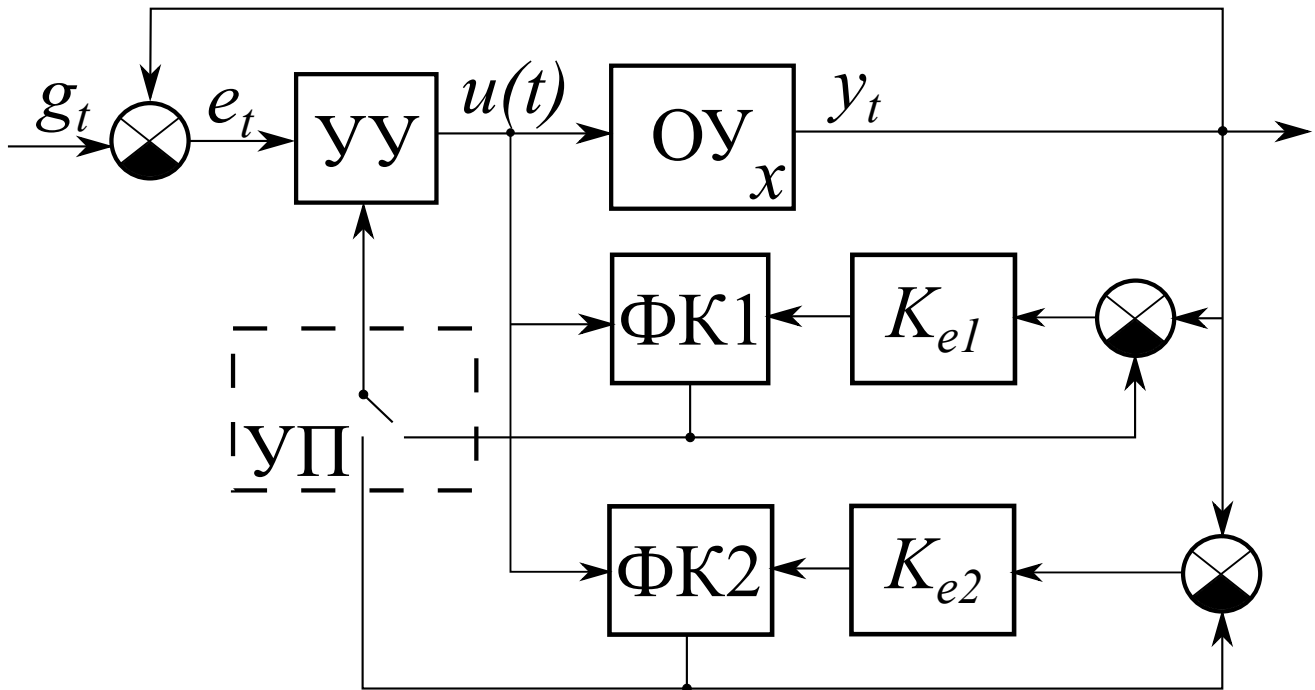


Рис. 6: Система слежения за маневрирующей целью

В режиме нормального функционирования отслеживание движения цели производится устройством управления с фильтром Калмана 1, а при интенсивном маневрировании цели – с фильтром Калмана 2. В силу того, что момент начала маневра цели непредсказуем, структура системы изменяется в случайные моменты времени.

Вопросы для самоконтроля

1. Какова обобщенная структурная схема стохастической системы и какие элементы и сигналы в нее входят?
2. В каких элементах структурной схемы стохастической системы возможно случайное переключение структуры?
3. Как с точки зрения стохастических систем могут быть построены системы слежения?

4 Математические модели стохастических динамических систем

Стохастическими динамическими системами называются системы, функционирующие в условиях наличия внешних и внутренних случайных возмущений.

К внешним случайным возмущениям относят случайные возмущающие воздействия, приложенные к объекту управления, а также случайные составляющие задающих воздействий.

Внутренние случайные возмущения порождаются непредсказуемыми изменениями физических, химических и/или иных свойств элементов системы, процессами "старения" элементов, а также нестабильностью функционирования элементов системы. Также к внутренним случайным возмущениям относят случайные распределения начальных условий системы.

При анализе поведения системы следует четко отличать реальную систему (технический объект) от ее модели. Воздействия на реальную систему могут быть отражены в модели различными способами:

- случайные аддитивные величины;
- мультипликативные случайные возмущения (чаще всего отражаются в параметрах системы);
- случайное распределение начальных условий системы;
- случайное изменение структуры системы (отражают, например, случайные изменения режимов и/или условий функционирования).

4.1 Модели непрерывных стохастических динамических систем

Непрерывная стохастическая динамическая система может быть описана математической моделью вида

$$\dot{x} = f(x, u, w, \alpha), \quad x(0) = \bar{x}_0 + \overset{\circ}{x}_0, \quad (8)$$

$$y = g(x, v, \beta), \quad (9)$$

где $x = [x_1 \ x_2 \ \dots \ x_n]^T \in R^n$ – вектор состояния системы, $x(0)$ – вектор начальных условий переменных вектора состояния системы, $\bar{x}_0 = M[x(0)]$ – математическое ожидание вектора начальных условий $x(0)$, $\overset{\circ}{x}_0 = x(0) - \bar{x}_0$ – центрированное значение вектора начальных условий $x(0)$, $u = [u_1 \ u_2 \ \dots \ u_k]^T$

– вектор управляющих воздействий, $y = [y_1 \ y_2 \ \dots \ y_m]^T$ – вектор выходных (регулируемых) величин, по отношению к которым формулируется задача управления, $w = [w_1 \ w_2 \ \dots \ w_l]^T$ – вектор внешних возмущающих воздействий, $v = [v_1 \ v_2 \ \dots \ v_q]^T$ – вектор ошибок измерений и помех в измерительных каналах связи, $\alpha = [\alpha_1 \ \alpha_2 \ \dots \ \alpha_\gamma]^T$ и $\beta = [\beta_1 \ \beta_2 \ \dots \ \beta_p]^T$ – векторы параметров модели.

Случайным процессом будем называть процесс, изменяющийся во времени, такой, что значение этого процесса в любой фиксированный момент времени представляет собой случайную величину.

Основными задачами, решаемыми в теории управления, являются:

- задача анализа поведения системы с заданным описанием;
- задача синтеза управляющего устройства для системы с заданным описанием объекта управления, условий и цели функционирования системы.

Решение этих задач для стохастических систем с учетом всех случайных факторов представляет собой нетривиальную комплексную проблему, поэтому при описании функционирования таких систем часто вводятся различные упрощающие предположения.

4.2 Линейные стохастические системы

Динамическая система с аддитивными случайными воздействиями на входе и в измерительном канале связи, осуществляющая линейное преобразование входных сигналов, называется линейной стохастической системой. Математическая модель такой системы описывается уравнениями

$$\dot{x} = A(\alpha_t) + Bu + Gw_t, \quad x(0) = \bar{x}_0 + \overset{\circ}{x}_0, \quad (10)$$

$$y = C(\beta_t)x + Dv_t, \quad (11)$$

где векторные переменные x , y и u и случайные воздействия v_t и w_t определяются по аналогии с математической моделью (8)-(9), а A , B , G , C и D – матрицы соответствующих размерностей.

Если компоненты векторов α_t и β_t – постоянные числа, то линейная стохастическая система осуществляет детерминированное преобразование случайных векторов v_t и w_t , если же они представляют собой случайные величины или случайные процессы, то линейная стохастическая система осуществляет стохастическое преобразование случайных векторов v_t и w_t .

Если параметры линейной стохастической системы зависят от времени, система является нестационарной, в противном случае – система стационарная.

Линейная стохастическая система – это приближенная модель реальной системы, полученная посредством линеаризации непрерывной (в общем случае – нелинейной) модели стохастической системы.

Принимаются некоторые упрощающие предположения. Для моделирования воздействия на систему случайных возмущений в качестве моделей выбирают случайные процессы w_t и v_t с хорошо изученными свойствами, имеющие статистические характеристики в удобном и наглядном аналитическом представлении.

Стремятся так выбрать модели возмущений и помех, чтобы:

- процессы аналитически отражались минимальным числом статистических характеристик;
- в этих характеристиках (функциях) должно быть минимально возможное число параметров, имеющих практически содержательную трактовку и смысл.

Чаще всего в качестве моделей w_t и v_t принимают гауссовские белые шумы с нулевыми средними и матричными ковариационными функциями вида

$$M [w(t), w^T(t)] = Q\delta(t - \tau),$$

$$M [v(t), v^T(t)] = R\delta(t - \tau),$$

где Q и R – положительно определенные матрицы, а $\delta(t - \tau)$ – дельта-функция.

Если возмущения w_t и v_t имеют ненулевые математические ожидания (средние значения), эти математические ожидания включают в детерминированную составляющую модели.

Исследование даже линейных стохастических систем в общих условиях приводит к сложным и громоздким математическим выкладкам, поэтому часто принимаются дополнительные упрощающие предположения:

- возмущения w_t и v_t считают некоррелированными, т.е.

$$M [w(t)v^T(t)] = 0 \quad \forall t, \tau;$$

- начальные условия $x(0)$ представлены случайным гауссовским вектором с математическим ожиданием $\bar{x}(0) = M [x(0)]$ и положительно определенной матрицей ковариаций $P = M [(x(0) - \bar{x}(0)) ((x(0) - \bar{x}(0))^T)]$.

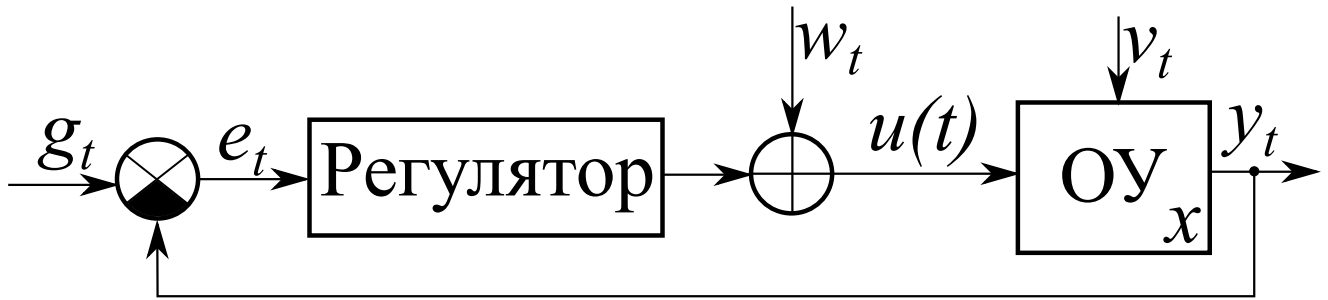


Рис. 7: Система со случайно переключающейся структурой

4.3 Линейные непрерывные стохастические системы со случайной структурой

Системы со случайной структурой отличаются тем свойством, что в каждый момент времени может произойти изменение их структуры. Будем считать, что система может иметь несколько (небольшое число, в общем случае – счетное множество) структурных состояний.

Описание динамики поведения системы в каждом структурном состоянии будем описывать линейной моделью вида

$$\dot{x} = F(s_t)x + G(s_t)w_t, \quad x_{s_t}(0) = \bar{x}_{s_t0} + \overset{\circ}{x}_{s_t0}, \quad (12)$$

$$y = C(s_t)x + D(s_t)v_t, \quad (13)$$

где $F(s_t) = A(s_t) - B(s_t)K(s_t)$ – матрица динамики замкнутой системы, s_t – случайный процесс, описывающий изменения (переключений) структуры системы в процессе ее функционирования, $I = \{1, 2, \dots, \nu\}$ – число структурных состояний.

В качестве примера систем со случайно изменяющейся в процессе функционирования структурой (см. рисунок 7) можно привести следующие системы:

- системы поиска-захвата-сопровождения цели;
- системы с возможным нарушением и восстановлением структуры.

В математической модели системы со случайной структурой должны быть заданы:

- уравнения описания поведения системы в каждом из структурных состояний;
- статистические характеристики случайных возмущений и помех в каждом структурном состоянии;

- условия переключения режимов, т.е. условия, при которых система переходит из одного структурного состояния в другое;
- модель чередования структурных состояний.

Стоит отметить, что матрицы F , G , C и D различны для каждого структурного состояния. Их размерности согласованы с размерностью вектора состояния в каждом конкретном структурном состоянии и могут отличаться друг от друга для различных структурных состояний.

При синтезе регуляторов для систем со случайно изменяющейся структурой процедура проводится для каждого структурного состояния отдельно.

В качестве математических моделей процесса переключения структурных состояний системы можно выбрать:

- марковские процессы с дискретным временем;
- марковские процессы с непрерывным временем;
- полумарковские процессы.

Многофункциональные системы – это системы, выполняющие в процессе работы несколько функций.

Многорежимные системы – это системы, имеющие несколько режимов функционирования.

Мультиструктурные системы – это системы, структура которых может изменяться при их функционировании.

4.4 Дискретные модели стохастических систем

Дискретные модели находят широкое применение в связи с тем, что системы управления реализуются на различных ЭВМ, функционирующих в дискретном времени. Линейная дискретная стохастическая система может быть описана уравнениями

$$x(m+1) = A(\alpha_m)x(m) + Bu(m) + Gw(m), \quad x(0) = \bar{x}_0 + \overset{\circ}{x}_0, \quad (14)$$

$$y(m) = C(\beta_m)x(m) + Dv(m). \quad (15)$$

В модели (14)-(15) векторы x , u и y имеют те же размерности и смысл, что и в модели (10)-(11). Элементы матриц $A(\alpha_m)$ и $C(\beta_m)$ выражены через векторы параметров α_m и β_m , компоненты которых могут быть как постоянными числами, так и случайными величинами или случайными процессами с дискретным временем. В общем случае от параметров могут зависеть элементы каждой из матриц данной системы.

Для дискретных стохастических систем вводятся следующие упрощающие предположения:

- возмущающие воздействия $w(m)$ и $v(m)$ представляют собой гауссовские белые шумы с нулевыми средними либо с ненулевыми математическими ожиданиями $M[w(m)] = \bar{w}(m)$ и $M[v(m)] = \bar{v}(m)$ и ковариационными матрицами Q и R неотрицательно определенными) соответственно:

$$M \left[(w(m) - \bar{w}(m)) (w(k) - \bar{w}(k))^T \right] = Q \sigma_{km},$$

$$M \left[(v(m) - \bar{v}(m)) (v(k) - \bar{v}(k))^T \right] = R \sigma_{km},$$

где σ_{km} – символ Кронекера;

- возмущающие воздействия $w(m)$ и $v(m)$ полагаются некоррелированными (взаимно независимыми), т.е. $M[w(m)v^T(k)] = 0 \forall m, k$;
- вектор начальных условий для переменных вектора состояния $x(0)$ – гауссовский случайный вектор со средним $\bar{x}(0) = M[x(0)]$ и неотрицательно определенной матрицей ковариаций

$$P(0) = M \left[(x(0) - \bar{x}(0)) (x(0) - \bar{x}(0))^T \right];$$

- возмущающие воздействия $w(m)$ и $v(m)$ при любых значениях m не зависят от вектора начальных условий $x(0)$, т.е.

$$M[x(0)w^T(m)] = 0, M[x(0)v^T(m)] = 0.$$

4.5 Марковская модель чередования структурных состояний системы

Будем предполагать, что описываемая система имеет конечное число возможных структурных состояний. Тогда марковский процесс (конечная цепь Маркова) переключения структурных состояний может быть описан в виде макро-вектора

$$s_t = \{S, P, T, \Pi, K\}, \quad (16)$$

где $S = \{s_i; i = \overline{1, \nu}\}$ – счетное множество состояний случайного процесса переключения структурных состояний, $P = \{p_i; i = \overline{1, \nu}\}$ – счетное множество вероятностей пребывания системы в структурном состоянии s_i , обладающее стохастическим свойством $\sum_{i=1}^{\nu} p_i = 1$, $T = \{\Delta t_i; i = \overline{1, \nu}\}$ – счетное множество

длительностей пребывания случайного процесса в состояниях s_i , Π – матрица вероятностей P_{ij} перехода (переходных вероятностей) системы из структурного состояния s_i в структурное состояние s_j , $K = \{k : 0 \leq k < \infty\}$ – счетное множество дискретных моментов времени, с помощью которых случайный процесс параметризуется так, что для вектора вероятностей $P(k+1)$ пребывания в состоянии $s_t(k+1)$ при условии, что случайный процесс находился с вероятностью $P(k)$ в исходном (текущем) состоянии $s_t(k)$, оказывается справедливой рекуррентная процедура

$$P(k+1) = P(k)\Pi(k), \quad P(0) = P\{s_t(0)\}, \quad (17)$$

где $P(0)$ – распределение вероятностей начальных структурных состояний.

4.6 Полумарковская модель чередования режимов в системах со случайной структурой

При моделировании процесса чередования режимов в системах со случайно изменяющейся структурой с помощью полумарковских процессов считается, что переход из одного структурного состояния системы в другое может быть совершен в произвольный случайный момент времени.

В процессе чередования режимов, если система попадает в некоторое структурное состояние и задерживается в этом состоянии, распределение времени задержки в этом состоянии зависит от того, каким окажется следующее состояние процесса. В полумарковской модели можно контролировать число шагов переключения структурных состояний. С учетом этого говорят о вложенном в полумарковский процесс марковском процессе. Приступая к анализу динамики чередования режимов системы со случайной структурой, можно вычислить статистические характеристики вложенного в полумарковский процесс марковского процесса.

Рассмотрим в качестве примера модель функционирования лазерной измерительной системы. Лазерная измерительная система предназначена для контроля профиля зеркала большого полноповоротного радиотелескопа во время его эксплуатации. Система реализует оптико-локационный способ измерения координат контрольных точек на поверхности антенны радиотелескопа. На зеркале антенны организуется система контрольных точек, в которых наклеиваются отражатели. При контроле профиля осуществляется последовательный обход лучом профилометра контрольных точек, в каждой контрольной точке производится измерение двух угловых координат положения точки и расстояния до контрольной точки.

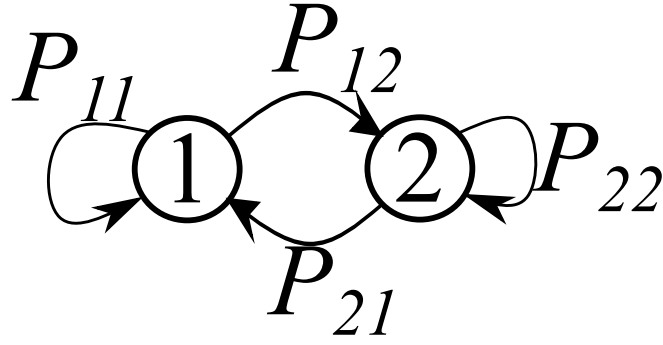


Рис. 8: Граф состояний и переходов профиломера

По результатам профилактических измерений оцениваются параметры модели профиля антенны радиотелескопа, а также определяются угловые смещения оси зеркала радиотелескопа и угловые и линейные перемещения контррефлектора радиотелескопа. Оценки указанных смещений, возникших под действием возмущений, вводятся в систему наведения радиотелескопа.

До начала измерений известны координаты предполагаемого расположения контрольных точек. Луч профиломера наводится в эту область, в ней осуществляется поиск и захват отражателя, после чего производятся измерения координат этой точки.

Упрощенная модель процесса чередования режимов работы профиломера содержит два состояния: наведение и измерение (см. рисунок 8).

$$I = \{1, 2\}$$

$$\Pi = \begin{bmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{bmatrix}$$

$$N(t) = \begin{bmatrix} n_{11}(t) & n_{12}(t) \\ n_{21}(t) & n_{22}(t) \end{bmatrix}$$

$$P(k+1) = P(k)\Pi, \quad P(0) = [p_{10} \ p_{20}]$$

Для модели с двумя состояниями возможно определение основных статистических характеристик процесса чередования режимов в аналитическом виде.

4.7 Модель динамики технических состояний системы с возможными нарушениями

Марковские и полумарковские модели технических систем могут быть использованы не только для описания процессов случайного чередования структур-

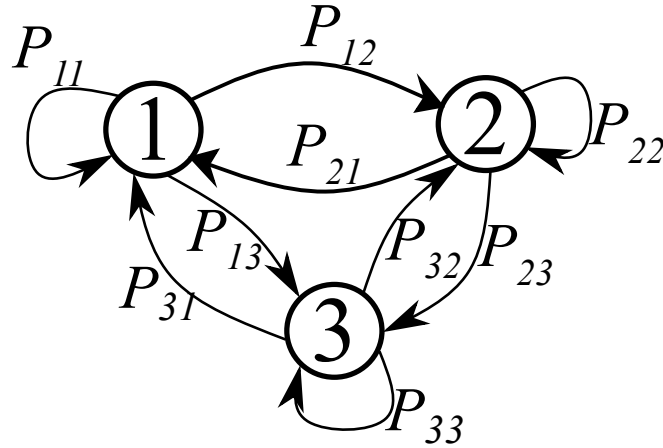


Рис. 9: Граф состояний и переходов процесса переключения состояний системы

ных состояний системы, но также и для описания динамики изменения технического состояния системы, чередования режимов функционирования системы, для оптимизации стратегий обслуживания систем и т.д.

Для количественной характеристики технического состояния системы вводят параметры технического состояния. Они отражаются признаками, описанными в технической документации. Важно, чтобы они были связаны с количественными характеристиками.

Функции системы условно делятся на основные и вспомогательные. Для многофункциональной системы вводятся следующие состояния:

- исправное (все параметры технического состояния – основные и вспомогательные в пределах допусков);
- работоспособное (все основные параметры технического состояния в пределах допусков);
- неработоспособное (по крайней мере один из основных параметров вне допуска).

Граф состояний и переходов полумарковской модели процесса переключения между перечисленными состояниями системы приведен на рисунке 9.

Математическая модель функционирования описанной системы будет иметь вид

$$\Pi = \begin{bmatrix} P_{11} & P_{12} & P_{13} \\ P_{21} & P_{22} & P_{23} \\ P_{31} & P_{32} & P_{33} \end{bmatrix},$$

$$P(k+1) = P(k)\Pi, \quad P(0) = [p_{10} \ p_{20} \ p_{30}],$$

$$N(t) = \begin{bmatrix} n_{11}(t) & n_{12}(t) & n_{13}(t) \\ n_{21}(t) & n_{22}(t) & n_{23}(t) \\ n_{31}(t) & n_{32}(t) & n_{33}(t) \end{bmatrix}.$$

Вопросы для самоконтроля

1. Как в математическую модель стохастической динамической системы могут входить случайные величины или процессы?
2. Как строятся математические модели линейных стохастических систем и каковы рекомендации к выбору моделей случайных процессов в них?
3. Как описываются линейные непрерывные стохастические системы со случайной структурой?
4. Как строятся дискретные модели линейных стохастических систем и каковы рекомендации к выбору моделей случайных процессов в них?
5. Как строятся математические модели чередования структурных состояний стохастических систем?

5 Марковские процессы

Случайный процесс $X(t)$ называется марковским [3], если для любых n моментов времени $t_1 < t_2 < \dots < t_n$ из отрезка $[0, T]$ условная функция распределения “последнего” значения $X(t_n)$ при фиксированных значениях $X(t_1), X(t_2), \dots, X(t_{n-1})$ зависит только от $X(t_{n-1})$, т. е. при заданных значениях $x_1, x_2, \dots, x_n$ справедливо соотношение

$$P(X(t_n) \leq x_n | X(t_1) = x_1, \dots, X(t_{n-1}) = x_{n-1}) = P(X(t_n) \leq x_n | X(t_{n-1}) = x_{n-1}). \quad (18)$$

Для трех моментов времени $t_i > t_j > t_k$ выражение (18) принимает вид

$$P(X(t_i) \leq x_i | X(t_k) = x_k, X(t_j) = x_j) = P(X(t_i) \leq x_i | X(t_j) = x_j). \quad (19)$$

Поэтому часто говорят, что характерное свойство марковских процессов состоит в следующем: если точно известно состояние марковского процесса в настоящий момент времени t_j , то будущее состояние в момент времени t_i не зависит от прошлого состояния в момент времени t_k .

В качестве определения марковского процесса можно принять также следующее симметричное относительно времени соотношение:

$$\begin{aligned} P(X(t_i) \leq x_i, X(t_k) \leq x_k | X(t_j) = x_j) = \\ = P(X(t_i) \leq x_i | X(t_j) = x_j) P(X(t_k) \leq x_k | X(t_j) = x_j). \end{aligned} \quad (20)$$

Такая запись означает, что при фиксированном состоянии процесса в настоящий момент времени t_j будущее в момент времени t_i и прошлое в момент времени t_k состояния марковского процесса независимы. Выражения (19) и (20) по существу эквивалентны, однако в правой части выражения (19) фигурируют не три, а лишь два момента времени. Ввиду этого упрощения для установления характера марковского процесса в дальнейшем будем опираться на выражение (19).

Укажем еще одно общее и важное свойство марковских процессов: для них эволюция вероятности перехода $P = P(X(t) \leq x | X(t_0) = x_0)$ описывается уравнением вида

$$\frac{d}{dt}P = \mathcal{A}P,$$

где $\mathcal{A}$ – некоторый линейный оператор (матрица, дифференциальный оператор, интегродифференциальный оператор и т. п.). Это свойство позволяет исследовать поведение марковских процессов при помощи хорошо разработанных методов решения соответствующих дифференциальных уравнений.

В дальнейшем применительно к марковским процессам будем использовать следующие обозначения: $X(t) = \lambda(t)$, если пространство состояний (фазовое пространство) процесса непрерывно, и $X(t) = \theta(t)$, если пространство состояний дискретно.

5.1 Цепи Маркова

Рассмотрим некоторую физическую систему, имеющую конечное число K всех возможных фазовых состояний $\vartheta_1, \vartheta_2, \dots, \vartheta_k, \dots, \vartheta_K$ [3]. Пусть в зависимости от вмешательства случая система шаг за шагом (в моменты времени $t_0 < t_1 < t_2 < \dots$) скачкообразно меняет свое фазовое состояние, т. е. имеют место переходы $\theta_0 \rightarrow \theta_1 \rightarrow \theta_2 \rightarrow \dots$, где $\theta_n = \theta(t_n)$ – состояние системы через n шагов, а $\theta_0 = \theta(t_0)$ – начальное состояние системы.

Полное вероятностное описание поведения рассматриваемой системы дается совместными конечномерными вероятностями $P(\theta_0, \theta_1, \dots, \theta_n)$ при разных n . На основании теоремы умножения вероятностей можно записать

$$\begin{aligned} P(\vartheta_{k_0}, \vartheta_{k_1}, \dots, \vartheta_{k_n}) &= P(\theta_0 = \vartheta_{k_0}, \theta_1 = \vartheta_{k_1}, \dots, \theta_n = \vartheta_{k_n}) = \\ &= P(\theta_0 = \vartheta_{k_0}, \theta_1 = \vartheta_{k_1}, \dots, \theta_{n-1} = \vartheta_{k_{n-1}}) P(\theta_n = \vartheta_{k_n} | \theta_0 = \vartheta_{k_0}, \theta_1 = \vartheta_{k_1}, \dots, \theta_{n-1} = \vartheta_{k_{n-1}}), \end{aligned} \quad (21)$$

где ϑ_{k_μ} – одно из возможных фазовых состояний θ_μ .

Расписывая аналогично вероятность $P(\theta_0 = \vartheta_{k_0}, \theta_1 = \vartheta_{k_1}, \dots, \theta_{n-1} = \vartheta_{k_{n-1}})$, получим

$$P(\theta_0, \theta_1, \theta_2, \dots, \theta_{n-1}, \theta_n) = P_0(\theta_0) \prod_{\mu=1}^n \pi_\mu(\theta_\mu | \theta_0, \dots, \theta_{\mu-1}), \quad (22)$$

где $P_0(\theta_0)$ – вероятность начального состояния (в момент времени t_0), а $\pi_\mu(\theta_\mu | \theta_0, \dots, \theta_{\mu-1})$ – условная вероятность, то есть вероятность перехода на μ -м шаге, определяемая выражением

$$\pi_\mu(\theta_\mu | \theta_0, \dots, \theta_{\mu-1}) = P(\theta_\mu = \vartheta_{k_\mu} | \theta_0 = \vartheta_{k_0}, \dots, \theta_{\mu-1} = \vartheta_{k_{\mu-1}}), \quad 1 \leq \mu \leq n. \quad (23)$$

Следовательно, для вычисления совместных вероятностей $P(\theta_0, \theta_1, \dots, \theta_n)$ нужно задать начальное состояние системы и указать физический механизм осуществления смены состояний, позволяющий вычислить вероятности перехода π_μ . Исходя из возможной простоты решения задачи, можно указать следующие частные случаи.

1. Смена всех состояний происходит независимо, т. е. вероятность какого-либо состояния на μ -м шаге не зависит от того, в каких состояниях находилась

система в предыдущие моменты времени. В данном случае

$$\pi_n(\theta_n|\theta_0, \theta_1, \dots, \theta_{n-1}) = P_n(\theta_n),$$

и выражение (22) упрощается до

$$P(\theta_0, \theta_1, \dots, \theta_n) = \prod_{\mu=0}^n P_\mu(\theta_\mu).$$

Такое выражение описывает последовательность независимых испытаний, и это есть частный (вырожденный) случай цепи Маркова.

2. К определению общей простой цепи Маркова приводит более общая возможность, когда вероятность фазового состояния параметра θ_n в момент времени t_n зависит лишь от того, в каком состоянии находилась система в непосредственно предшествующий ему момент времени t_{n-1} и не зависит от того, в каких состояниях находилась система в более ранние моменты времени $t_0, t_1, \dots, t_{n-1}$. Это свойство характерно для всех марковских процессов (поэтому их также называют процессами без последствия).

Таким образом, последовательность состояний или совокупность случайных величин образует простую цепь Маркова, если для всех $n = 1, 2, 3, \dots$ и всех возможных значений случайных величин θ_n выполняется соотношение

$$\pi_n(\theta_n|\theta_0, \theta_1, \dots, \theta_{n-1}) = \pi_n(\theta_n|\theta_{n-1}).$$

Условные вероятности $\pi_n(\theta_n|\theta_{n-1})$ называют вероятностями перехода или переходными вероятностями.

Для простой цепи Маркова выражение (22) принимает вид

$$P(\theta_0, \theta_1, \dots, \theta_n) = P_0(\theta_0) \prod_{\mu=1}^n \pi_\mu(\theta_\mu|\theta_{\mu-1}). \quad (24)$$

Из этого следует, что полное вероятностное описание простой цепи Маркова достигается заданием вероятностей начального состояния и вероятностей перехода

$$\pi_\mu(\theta_\mu|\theta_{\mu-1}) = P(\theta_\mu = \vartheta_{k_\mu} | \theta_{\mu-1} = \vartheta_{k_{\mu-1}}).$$

Стоит отметить, что для двух случайных величин соотношение (24) следует из теоремы умножения вероятностей и может быть написано всегда. Но уже для трех случайных величин выражение (24) накладывает существенные ограничения на вид возможных совместных вероятностей.

3. Можно также ввести определение цепи Маркова порядка m ($m \geq 1$), если вероятность нового состояния зависит только от m состояний системы, непосредственно ему предшествующих:

$$\pi_n(\theta_n|\theta_0, \theta_1, \dots, \theta_{n-1}) = \pi_n(\theta_n|\theta_{n-m}, \dots, \theta_{n-1}), \quad n > m. \quad (25)$$

Значение $m = 1$ соответствует простой цепи Маркова, при $m > 1$ имеем сложную цепь Маркова порядка m .

Сложная цепь Маркова порядка m может быть сведена к простой цепи Маркова для m -мерного вектора [3]. Для этого используем то обстоятельство, что члены конечной случайной последовательности длины m можно рассматривать как составляющие случайного вектора. Пусть задан случайный вектор

$$\theta_n = [\theta_{i,n}, i = \overline{1, m}] = [\theta_{n-m+1} \dots \theta_n].$$

На предыдущем шаге m -мерный вектор равен

$$\theta_{n-1} = [\theta_{i,n-1}, i = \overline{1, m}] = [\theta_{n-m} \dots \theta_{n-1}],$$

причем между компонентами векторов θ_n и θ_{n-1} выполняется соотношение

$$\theta_{i,n} = \theta_{i+1,n-1} = \theta_{n-m+i}, i = \overline{1, m-1}; \quad \theta_{im,n} = \theta_n, \quad m < n. \quad (26)$$

Следовательно, сложная цепь Маркова порядка m представляет собой частный случай простой m -мерной цепи Маркова, если положить вероятность перехода $\pi_n(\theta_n|\theta_{n-1})$ равной

$$\pi_n(\theta_n|\theta_{n-1}) = \pi_n(\theta_{m,n}|\theta_{1,n-1}, \dots, \theta_{m,n-1}) \prod_{i=1}^{m-1} \delta(\theta_{i,n}, \theta_{i+1,n-1}), \quad (27)$$

где $\delta(\mu, \nu) = \delta_{\mu\nu}$ — символ Кронекера, принимающий единичное значение при равенстве индексов μ и ν и нулевое значение во всех остальных случаях. Стоит отметить, что вероятности перехода (27) отличны от нуля и совпадают с соответствующими вероятностями (25) только для векторов, компоненты которых удовлетворяют соотношениям (26).

Случайные цепи, определенные соотношением (24), а также более общего вида, называют марковскими потому, что они впервые систематически изучались известным русским математиком А. А. Марковым.

5.2 Полумарковские процессы

Полумарковские процессы широко используются в теории надежности и теории массового обслуживания. Они объединяют теорию цепей Маркова, разрывных марковских процессов и теорию восстановления. Приведем сначала определение полумарковского процесса [3].

Пусть поведение некоторой системы описывается следующим образом. В каждый момент времени система может находиться в одном из K возможных фазовых состояний $\vartheta_1, \vartheta_2, \dots, \vartheta_K$, причем известны начальное состояние системы (в начальный момент времени t_0 она находится в состоянии $\vartheta_0 = \vartheta_i$) и одношаговые вероятности перехода $\pi_{jk} = P\{\theta_m = \vartheta_k | \theta_{m-1} = \vartheta_j\}$, $j, k = \overline{1, K}$. Следовательно, процесс $\theta_m = \theta(t_m)$ есть однородная цепь Маркова.

Сопоставим каждому ненулевому элементу π_{jk} матрицы вероятностей перехода случайную величину T_{jk} с функцией распределения $F_{jk}(t) = P\{\tau_{jk} \leq t\}$. Величине T_{jk} , зависящей как от состояния ϑ_j , так и от состояния ϑ_k , можно придавать разный физический смысл. Она может быть непрерывной или дискретной. В теории надежности и теории массового обслуживания случайную величину T_{jk} обычно рассматривают как время пребывания системы в состоянии ϑ_j при условии, что следующим состоянием, в которое перейдет система, будет ϑ_k . При этом величина T_{jk} считается неотрицательной и непрерывной с плотностью вероятности $f_{jk}(t)$. При такой интерпретации величину T_{jk} можно назвать временем ожидания в состоянии ϑ_j до перехода в состояние ϑ_k .

Можно себе представить, что точка, отображающая поведение системы на фазовой плоскости, остается в состоянии ϑ_j в течение времени T_{jk} , прежде чем она перейдет в состояние ϑ_k . По достижении ϑ_k “мгновенно” (в соответствии с матрицей вероятностей перехода Π) выбирается следующее состояние ϑ_i , и после того, как состояние ϑ_i выбрано, время ожидания в ϑ_k полагается равным T_{ki} с функцией распределения $F_{ki}(t)$ или плотностью вероятности $f_{ki}(t)$. Этот процесс затем следует неограниченно продолжать, выбирая каждый раз независимо следующее состояние и время ожидания. Если через $\theta(t)$ обозначить состояние системы, занятое в момент времени t , то полученный процесс принято называть полумарковским.

Из приведенного определения следует, что если игнорировать случайный характер времени ожидания и интересоваться только моментами перехода, то процесс $\theta(t)$ будет представлять собой однородную цепь Маркова (в литературе ее часто называют вложенной цепью Маркова, а в общем случае – вложенным марковским процессом). Однако при учете пребывания процесса в разных состояниях в течение случайного отрезка времени процесс $\theta(t)$ не будет удовлетворять уравнению Маркова (если не все времена ожидания распределены экспоненциально). Следовательно, процесс является марковским только в мо-

менты перехода. Сказанное оправдывает название “полумарковский процесс” или “полумарковская цепь”.

При заданном начальном состоянии дальнейшее поведение полумарковского процесса (полумарковской цепи) полностью определяется матрицей вероятностей перехода Π и матрицей функций распределения или (для непрерывных случайных величин T_{jk}) матрицей плотностей вероятности.

Предположим, что в некоторый момент времени, принимаемый за начальный ($t_0 = 0$), система находится в состоянии ϑ_i . Пусть в качестве следующего выбрано (с вероятностью π_{ij}) состояние ϑ_j . Тогда по теоремам умножения и сложения вероятностей находим безусловную функцию распределения полного времени ожидания системы в состоянии ϑ_i :

$$F_i(t) = P\{t_i < t\} = \sum_{i=1}^K \pi_{ij} F_{ij}(t), \quad i, j = \overline{1, K}. \quad (28)$$

Соответственно, для безусловной плотности вероятности полного времени ожидания в состоянии ϑ_i можно написать

$$f_i(t) = \sum_{i=1}^K \pi_{ij} f_{ij}(t), \quad i, k = \overline{1, K}. \quad (29)$$

Среднее значение безусловного времени ожидания в состоянии ϑ_i равно

$$\langle T_i \rangle = \int_0^{\infty} t f_i(t) dt = \sum_{i=1}^K \pi_{ij} \langle T_{ij} \rangle. \quad (30)$$

Для дальнейших рассуждений введем обозначение

$$\Psi_i(t) = 1 - F_i(t) = \int_t^{\infty} f_i(t) dt = P\{T_i > t\}. \quad (31)$$

Перейдем теперь к решению основной задачи, возникающей при анализе полумарковских процессов – вычислению вероятностей состояний. Пусть $\Phi_{ij}(t)$ есть условная вероятность того, что в момент времени t система находится в состоянии ϑ_j , если в момент времени $t_0 = 0$ она была в состоянии ϑ_i . Вероятность $\Phi_{ij}(t)$ можно назвать интервально-переходной вероятностью. Система, стартуя из начального состояния ϑ_i , может попасть в состояние ϑ_j в момент времени t разными путями. Во-первых, если $\vartheta_i = \vartheta_j$, она может не покидать состояние ϑ_i в течение всего промежутка времени или, выйдя из состояния ϑ_i

по меньшей мере однажды, она все-таки возвращается в состояние $\vartheta_j = \vartheta_i$ к моменту времени t . Во-вторых, система может попасть в произвольное состояние ϑ_j , занимая в момент времени τ некоторое промежуточное состояние ϑ_k . Вероятности этих двух взаимоисключающих возможностей должны складываться, что приводит к выражению

$$\Phi_{ij}(t) = \delta_{ij}\Psi_i(t) + \sum_{k=1}^K \pi_{ik} \int_0^t f_{ik}(\tau) \Phi_{kj}(t-\tau) d\tau, \quad 1 \leq i, j \leq K, \quad (32)$$

где δ_{ik} – символ Кронекера: $\delta_{ij} = 1$ при $i = j$ и $\delta_{ij} = 0$ при $i \neq j$.

Первый член справа в выражении (32) учитывает вероятность события $\vartheta_i = \vartheta_j$, так как $\Psi_i(t)$ есть вероятность того, что система не покинет состояние ϑ_i до момента времени t . Второй член представляет вероятность последовательности событий, когда система совершает переход из ϑ_i в ϑ_k (может быть даже в себя) к моменту τ и затем переходит из состояния ϑ_k в состояние ϑ_j за оставшееся время $t - \tau$. Вероятности частных переходов суммируются по всем промежуточным состояниям ϑ_k , в которые возможны переходы из начального состояния ϑ_i , и интегрируются по всевозможным временам перехода τ между 0 и t .

Система линейных интегральных уравнений (32) является основной. Она дает выражение интервально-переходных вероятностей через основные характеристики полумарковского процесса. Получить аналитическое решение такой системы – задача нетривиальная. Некоторое упрощение дает применение метода преобразования Лапласа. Обозначим одностороннее преобразование Лапласа от функции $f(x)$ через

$$f^*(s) = \int_0^\infty e^{-sx} f(x) dx. \quad (33)$$

Применив преобразование Лапласа к выражению (32), получим

$$\Phi_{ij}^*(s) = \delta_{ij}\Psi_i^*(s) + \sum_{k=1}^K \pi_{ik} f_{ik}^*(s) \Phi_{kj}^*(s), \quad 1 \leq i, j \leq K. \quad (34)$$

Из (31) имеем:

$$\Psi_i^*(s) = \frac{1}{s} (1 - f_i^*(s)). \quad (35)$$

Полученная система алгебраических уравнений связывает преобразование Лапласа от интервально-переходных вероятностей с основными характеристиками процесса.

Чтобы воспользоваться известными результатами матричного исчисления, запишем систему уравнений (34) в матричном виде. Для этого введем матрицы $\Phi(t)$, Π , f , а также диагональные матрицы $W(t) : w_{ij} = \delta_{ij} f_i(t)$, $F(t) : f_{ij} = \delta_{ij} F_i(t)$, $\Psi(t) : \psi_{ij} = \delta_{ij} \Psi_i(t)$. Для учета специфического вида суммирования в (34) определим специальный вид умножения матриц, обозначив эту операцию символом $\times$. Если A , B и C – квадратные матрицы размерности $K \times K$, то $C = A \times B$ означает, что $c_{ij} = a_{ij} b_{ij}$, $1 \leq i, j \leq K$. Иными словами, операция $\times$ обозначает лишь перемножение соответствующих элементов матриц.

Теперь выражение (34) можно переписать следующим образом:

$$\Phi^*(s) = \Psi^*(s) + (\Pi \times f^*(s)) \Phi^*(s) \quad (36)$$

или

$$\Phi^*(s) = (I - \Pi \times f^*(s))^{-1} \Psi^*(s), \quad (37)$$

где I – единичная матрица. Предполагается, что обратная матрица в выражении (37) существует.

Из уравнения (37) следует, что интервально-переходные вероятности зависят только от произведения π_{ij} и f_{ij} , а не от каждого из сомножителей в отдельности. Уравнение (37) определяет интервально-переходные вероятности через основные характеристики полумарковского процесса, которые считаются заданными. Решение этого уравнения позволяет исследовать как режим установления, так и стационарное состояние. Если же интересоваться только стационарным состоянием, в уравнении (37) возможны упрощения.

В теории преобразований Лапласа хорошо известен следующий асимптотический результат:

$$\Phi = \lim_{t \rightarrow \infty} \Phi(t) = \lim_{s \rightarrow 0} s \Phi^*(s),$$

если существует хотя бы один из этих пределов.

На основании выражения (37) можно записать:

$$\Phi = \lim_{s \rightarrow 0} s \Phi^*(s) = \lim_{s \rightarrow 0} s (I - \Pi \times f^*(s))^{-1} \lim_{s \rightarrow 0} \Psi^*(s). \quad (38)$$

Рассмотрим каждый из пределов в правой части выражения (38) отдельно.

На основании (35) можно записать:

$$\lim_{s \rightarrow 0} \Psi^*(s) = \lim_{s \rightarrow 0} \frac{1}{s} (I - W^*(s)). \quad (39)$$

Учитывая условие нормировки плотностей вероятности $f_i(t)$, нетрудно убедиться, что $W^*(0) = I$. Поэтому значение предела в правой части (39) оказы-

вается неопределенным. Применив правило Лопиталя для раскрытия неопределенности, находим

$$\lim_{s \rightarrow 0} \Psi^*(s) = -\frac{d}{ds} W^*(s) \Big|_{s=0} = \int_0^{\infty} t W(t) dt = T, \quad (40)$$

где T – диагональная матрица размерности $K \times K$ средних безусловных времен ожидания в каждом из состояний (30).

Найдем теперь предел первого сомножителя в правой части выражения (38). Введя обозначение

$$H(s) = s(I - \Pi \times f^*(s))^{-1},$$

можно записать

$$H(s) - H(s) \Pi \times f^*(s) = sI. \quad (41)$$

Из условия нормировки плотностей вероятности следует, что

$$\lim_{s \rightarrow 0} f_{ij}^*(s) = \lim_{s \rightarrow 0} \int_0^{\infty} e^{-st} f_{ij}(t) dt = 1.$$

Поэтому при $s \rightarrow 0$ все элементы матрицы $f^*(s)$ равны единице. С учетом определенной выше операции перемножения матриц из (41) получим

$$\lim_{s \rightarrow 0} H(s) = H(0) = H(0) \Pi. \quad (42)$$

Покажем, что строки матрицы $H(0)$ должны быть пропорциональны финальным вероятностям состояний вложенной цепи Маркова, которая описывает характер смены состояний без учета их длительности. Предположим, что вложенная цепь Маркова имеет однозначные финальные вероятности, не зависящие от начального состояния. Эти финальные вероятности являются решением системы алгебраических уравнений (2.35) и (2.38):

$$p_j = \sum_{i=1}^K p_i \pi_{ij}, \quad \sum_{i=1}^K p_i = 1. \quad (43)$$

Учитывая предположенную единственность решения этой системы уравнений и сравнивая (42) с (43) или (2.34), можно прийти к заключению, что в качестве решения матричного уравнения (41) можно взять квадратную матрицу

$H(0)$ размерности $K \times K$, i -ая строка которой $H_i = [h_{i1} \ h_{i2} \ \dots \ h_{iK}]$ пропорциональна вектору-строке финальных вероятностей $P_f = [p_1 \ p_2 \ \dots \ p_K]$, т. е.

$$H_i = k_i P_f,$$

где k_i – некоторый коэффициент пропорциональности.

Следовательно, для стационарного состояния уравнение (38) принимает вид

$$\Phi = H(0) T.$$

Отсюда следует, что элементы матрицы Φ должны удовлетворять соотношению

$$\Phi_{ij} = h_{ij}(0) \langle T_j \rangle = k_i p_j \langle T_j \rangle. \quad (44)$$

Для определения коэффициентов k_i воспользуемся условием, что сумма финальных интервально-переходных вероятностей должна равняться единице:

$$\sum_{i=1}^K \Phi_{ij} = k_i \sum_{i=1}^K p_i \langle T_i \rangle = 1.$$

Отсюда следуем, что

$$k_i = \left(\sum_{i=1}^K p_i \langle T_i \rangle \right)^{-1}. \quad (45)$$

Коэффициенты пропорциональности оказываются одинаковыми для всех состояний.

На основании (44) и (45) можно записать:

$$\Phi_{ij} = p_j \langle T_j \rangle \left(\sum_{i=1}^K p_i \langle T_i \rangle \right)^{-1} = \Phi_j.$$

Видно, что финальные интервально-переходные вероятности Φ_{ij} не зависят от начального состояния ϑ_i и для их обозначения можно использовать только второй индекс. Эти вероятности равны финальным вероятностям состояний вложенного марковского процесса, взвешенным с относительными средними безусловными временами ожидания в каждом состоянии. Следовательно, единственной статистической характеристикой каждого времени ожидания, которая влияет на поведение процесса в стационарном состоянии, является его среднее значение.

Рассмотрим иллюстративный пример, показывающий, что из теории полумарковских процессов как частный случай следуют результаты теории цепей Маркова.

Пусть полумарковский процесс с дискретным временем задан матрицей одношаговых вероятностей перехода Π и времена ожидания в каждом состоянии одинаковы и равны единице, т. е.

$$f_{ij} = \delta(t - 1), \quad f_{ij}^* = e^{-s}, \quad 1 \leq i, j \leq K.$$

В данном случае

$$f_i^*(s) = \sum_{j=1}^K \pi_{ij} e^{-s} = e^{-s},$$

$$\Psi_i^*(s) = \frac{1}{s} (1 - f_i^*(s)) = \frac{1}{s} (1 - e^{-s}), \quad 1 \leq i \leq K.$$

Согласно (37) можно записать

$$\Phi^*(s) = (I - e^{-s}\Pi)^{-1} \frac{1}{s} (1 - e^{-s})$$

или

$$\Phi^*(s) = R^*(s) q(s),$$

где

$$R^*(s) = (I - e^{-s}\Pi)^{-1}, \quad q(s) = \frac{1 - e^{-s}}{s}.$$

В теории матриц доказывается, что справедливо следующее соотношение:

$$(I - A)^{-1} = \sum_{n=0}^{\infty} A^n.$$

Отсюда следует, что

$$R^*(s) = \sum_{n=0}^{\infty} e^{-ns} \Pi^n.$$

Из обратного преобразования Лапласа получаем

$$R(t) = \sum_{n=0}^{\infty} \Pi^n \delta(t - n), \quad q(t) = \begin{cases} 1, & 0 \leq t \leq 1 \\ 0, & t > 1 \end{cases}.$$

На основании этих выражений найдем матрицу интервально-переходных вероятностей

$$\Phi(t) = \int_0^t R(t - \tau) q(\tau) d\tau = \Pi^n, \quad n \leq t \leq n + 1.$$

Это выражение показывает, что для рассматриваемого полумарковского процесса, когда переходы возможны только в дискретные, равноотстоящие моменты времени, вероятность перехода из первоначального состояния ϑ_i в состояние ϑ_j за n шагов равна элементу матрицы Π^n с индексом ij . Этот результат свидетельствует о том, что частная конкретизация полумарковского процесса с дискретным временем позволяет получить корректные результаты для цепей Маркова.

Вопросы для самоконтроля

1. Какие случайные процессы относятся к марковским?
2. Что такое цепь Маркова? Как задается простая цепь Маркова?
3. Что такое полумарковский процесс? Как он задается? В чем отличия полумарковского процесса от цепи Маркова?
4. Какие существуют характеристики полумарковского процесса и как они рассчитываются?

6 Датчики физических величин в системах управления

При анализе и проектировании систем управления систему разделяют на элементы, которые стараются разбить на две группы: линейные инерционные звенья и звенья, описываемые безынерционными нелинейными зависимостями.

Линейные инерционные элементы называют типовыми динамическими звеньями.

Нелинейное безынерционное звено вырабатывает на своем выходе сигнал, значение которого определяется значением сигнала на входе в данный момент времени и не зависит от значений входного сигнала в предшествующие моменты времени.

Функция, связывающая значения выходной переменной со значениями входной переменной в данный момент времени, называется функцией преобразования:

$$y = \varphi(x, v, \beta), \quad (46)$$

где $y \in R^m$ – вектор выходных переменных, $x \in R^n$ – вектор измеряемых величин, $v \in R^q$ – вектор влияющих величин (помех, шумов), $\beta \in R^p$ – вектор параметров модели датчика.

Датчики в системах управления как технические устройства характеризуются рядом показателей качества, которые делятся на:

- метрологические характеристики;
- характеристики датчика как элемента системы управления;
- показатели надежности (аппаратной, алгоритмической, информационной).

Основными метрологическими характеристиками датчиков являются:

- статическая погрешность, определяемая как составляющая совокупной погрешности датчика, оцениваемая в условиях статических измерений;
- диапазон измерений;
- порог чувствительности датчика, определяемый как минимальное изменение измеряемой величины, вызывающее уверенно фиксируемое изменение выходного сигнала датчика;
- чувствительность.

В стохастических моделях датчиков изменения влияющих величин описываются как случайные процессы. Для моделирования поведения параметров β_i используют математический аппарат описания случайных векторов для медленного случайного изменения во времени параметров (для описания дрейфа параметров) используются модели случайных процессов с хорошо известным аналитическим описанием.

6.1 Методы статистического оценивания в исследовании датчиков систем управления

Рассмотрим задачу оценивания параметров и статистических характеристик выходного сигнала датчика системы управления по результатам градуировочного эксперимента.

По результатам градуировочного эксперимента, проведенного с испытываемым датчиком системы управления, с использованием методов статистического оценивания определяются оценки значений параметров преобразовательной характеристики датчика и статистические характеристики распределения значений оценок параметров и оценок значений выходного сигнала датчика.

Модель преобразовательной характеристики датчика представим в виде линейной регрессии

$$y = y(x, \beta) = \varphi^T(x)\beta + v, \quad (47)$$

где $y \in R^1$ – выходной (скалярный) сигнал датчика, $x \in R^m$ – вектор входных (измеряемых) величин, $\varphi^T(x) = [\varphi_1(x) \ \varphi_2(x) \ \dots \ \varphi_p(x)]$ – векторная функция, компоненты которой устанавливаются из физических принципов работы датчика либо подбираются при моделировании его преобразовательной характеристики, $\beta \in R^p$ – вектор неизвестных (оцениваемых) параметров преобразовательной характеристики датчика, $v \in R^1$ – случайная составляющая результата измерения выходного сигнала датчика.

Будем считать v случайной величиной с независимыми отсчетами (при фиксации выходного сигнала y в градуировочном эксперименте) с нулевым средним ($M(v) = 0$), подчиняющейся нормальному распределению, т.е. $v \sim N(0, \delta^2)$.

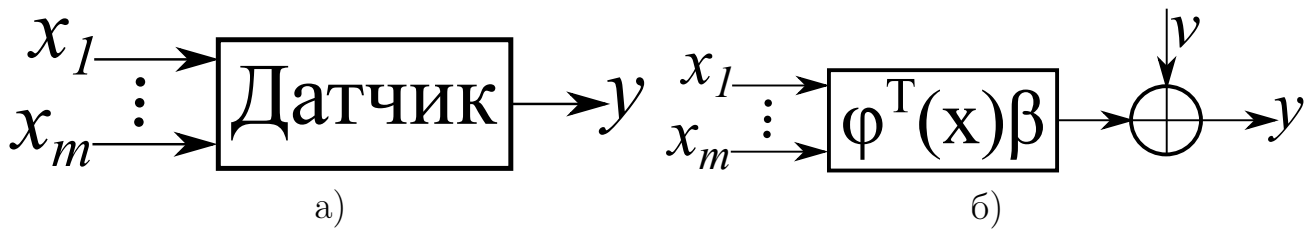


Рис. 10: Функциональная (а) и структурная (б) схемы датчика.

Для оценки статистических характеристик параметров преобразовательной характеристики и выходного сигнала датчика проводится градуировочный эксперимент, в результате которого выполняется n измерений выходного сигнала датчика ($n \geq p$) при различных значениях вектора измеряемых величин x .

Варианты значений измеряемых величин, фиксируемых в градуировочном эксперименте (план эксперимента), представляются в матричном виде:

$$X = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1m} \\ x_{21} & x_{22} & \dots & x_{2m} \\ \vdots & \vdots & \vdots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{nm} \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}.$$

Матрицу X называют матрицей регрессоров.

Используя установленную (фиксированную) совокупность векторов x_i ($i = \overline{1, n}$), вычисляют расширенную матрицу регрессоров, которая имеет вид

$$H = \begin{bmatrix} \varphi_1(x_1) & \varphi_2(x_1) & \dots & \varphi_p(x_1) \\ \varphi_1(x_2) & \varphi_2(x_2) & \dots & \varphi_p(x_2) \\ \vdots & \vdots & \vdots & \vdots \\ \varphi_1(x_n) & \varphi_2(x_n) & \dots & \varphi_p(x_n) \end{bmatrix}.$$

Элементы $\varphi_j(x_i)$ ($i = \overline{1, n}, j = \overline{1, p}$) расширенной матрицы регрессоров при выбранном плане эксперимента могут быть вычислены до проведения градуировочного эксперимента, тогда функции $\varphi_j(x_i)$ известны априори.

Если функции $\varphi_j(x_i)$ подбираются при обработке результатов градуировочного эксперимента, то матрица H формируется после выбора функций $\varphi_j(x_i)$ и вычисления значений этих функций при фиксированных (выбранных) значениях вектора x_i .

По результатам измерений значений выходного сигнала датчика при фиксированных (выбранных) значениях входных величин формируется вектор измеряемых значений выхода датчика $Y \in R^n$.

Уравнение преобразовательной характеристики датчика (47) справедливо для каждого отдельного опыта в градуировочном эксперименте, т.е.

$$y_i = \varphi^T(x_i) \beta + v_i, \quad i = \overline{1, n}. \quad (48)$$

Совокупный результат всех опытов градуировочного эксперимента представим в векторно-матричном виде:

$$Y = H\beta + V, \quad (49)$$

где $V \in R^n$ – значения случайной составляющей результата измерения выходного сигнала датчика в каждом опыте градуировочного эксперимента.

Целью градуировочного эксперимента является определение с использованием уравнения (49) вектора $\hat{\beta} \in R^p$ оценок значений параметров преобразовательной характеристики датчика (47) на основании измерительной информации, полученной в ходе эксперимента.

Оценки $\hat{\beta}_i$ ($i = \overline{1, p}$), которые будут вычислены, являются случайными величинами с некоторыми вероятностными распределениями. Для оценок $\hat{\beta}_i$ установим их важнейшие статистические характеристики, а для выходного сигнала датчика при фиксированном значении вектора измеряемых величин x_i определим предсказанное значение отклика датчика $\hat{y}$ и оценим значение дисперсии отклика $d(\hat{y}(x_i))$.

В математической статистике развито несколько методов оценивания параметров регрессионных зависимостей. Рассмотрим один из них – метод максимального правдоподобия. В этом методе в качестве оценок параметров $\hat{\beta}_i$ принимаются значения, доставляющие максимум функции правдоподобия $L(x, \beta)$, определяющей плотность вероятности того, что появится некоторая точка в пространстве выборок, т.е.

$$\hat{\beta} = \underset{\beta}{\operatorname{argmax}} L(x, \beta).$$

Для вычисления вектора оценок $\hat{\beta}$, доставляющего максимум функции правдоподобия, используют необходимое условие вида

$$\frac{\partial L(x, \beta)}{\partial \beta} = 0. \quad (50)$$

Для распределений экспоненциального типа часто применяют логарифмическую функцию правдоподобия $\ln L(x, \beta)$. Максимумы функций правдоподобия $L(x, \beta)$ и $\ln L(x, \beta)$ совпадают, т.к. логарифм является монотонной функцией.

При оценке векторного параметра необходимое условие правдоподобия приводит к системе уравнений относительно частных производных по компонентам искомого вектора.

В рассматриваемом случае функция плотности распределения измеренного значения отклика датчика y_i имеет вид

$$W(y_i, \beta) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(y_i - \varphi^T(x_i)\beta)^2}. \quad (51)$$

Представим функцию правдоподобия измеренных значений отклика датчика в виде

$$L(y, \beta) = (2\pi\sigma^2)^{-\frac{n}{2}} e^{-\sum_{i=1}^n \frac{1}{2\pi\sigma^2}(y_i - \varphi^T(x_i)\beta)^2}. \quad (52)$$

В матричных обозначениях функция правдоподобия (52) представляется выражением

$$L(y, \beta) = (2\pi\sigma^2)^{-\frac{n}{2}} e^{\frac{1}{2\sigma^2}(Y-H\beta)^T(Y-H\beta)}. \quad (53)$$

Для вычисления оценки вектора неизвестных параметров преобразовательной характеристики датчика $\hat{\beta}$ воспользуемся методом максимального правдоподобия. Составим логарифмическую функцию правдоподобия

$$\ln L(y, \beta) = \ln (2\pi\sigma^2)^{-\frac{n}{2}} - \frac{1}{2\sigma^2} (Y - H\beta)^T (Y - H\beta). \quad (54)$$

Найдем производную функции правдоподобия (54):

$$\frac{\partial \ln L(y, \beta)}{\partial \beta} = -\frac{2}{2\sigma^2} H^T (Y - H\beta).$$

Приравняв производную функции правдоподобия к нулю, получим линейную относительно вектора β систему уравнений

$$(H^T H) \beta = H^T Y. \quad (55)$$

Единственное (при невырожденной матрице $H^T H$) решение уравнения (55) доставляет максимум функции правдоподобия (54) и дает следующую оценку вектору параметров преобразовательной характеристики датчика:

$$\hat{\beta} = (H^T H)^{-1} H^T Y. \quad (56)$$

Оценка, определяемая выражением (56), является несмещенной, т.е. выполняется соотношение

$$M(\hat{\beta}) = (H^T H)^{-1} H^T M(Y) = (H^T H)^{-1} H^T (H\beta + M(V)) = \beta,$$

в силу чего можно сделать вывод о ее состоятельности и эффективности.

Ковариационная матрица $K(\hat{\beta})$ оценки вектора параметров преобразовательной характеристики датчика определяется выражением

$$K(\hat{\beta}) = \sigma^2 (H^T H)^{-1},$$

в котором матрицу $H^T H$ называют информационной матрицей Фишера.

Нормальность случайной составляющей результата измерения выходного сигнала v определяет нормальность распределения оценки $\hat{\beta}$, т.е.

$$\hat{\beta} \sim N(\hat{\beta}, \sigma^2 (H^T H)^{-1}).$$

Таблица 1: Результаты градуировочного эксперимента

x	-5	-4	-3	-2	-1	0	1	2	3	4	5
y	-20.1	-15.7	-11.6	-8	-4.2	0.7	4.7	7.4	11.3	15.2	19.6

В результате вычисления оценки $\hat{\beta}$ параметров преобразовательной характеристики датчика получена несмещенная и эффективная оценка прогноза значений выходного сигнала датчика при измерениях в статическом режиме (при фиксированном значении входных величин x).

Для оценки значений выходного сигнала датчика используется выражение

$$\hat{y}(x) = \varphi^T(x) \hat{\beta}. \quad (57)$$

Оценка выходного сигнала датчика (57) также имеет нормальное распределение, дисперсия которого определяется соотношением

$$d(\hat{y}(x)) = \varphi^T(x) K(\hat{\beta}) \varphi(x) = \sigma^2 \varphi^T(x) (H^T H)^{-1} \varphi(x). \quad (58)$$

Рассмотрим в качестве примера задачу оценки параметров линейной регрессионной модели преобразовательной характеристики датчика на основе данных градуировочного эксперимента, приведенных в таблице 1.

Графически результаты градуировочного эксперимента из таблицы 1 представлены на рисунке 11. Необходимо определить функции регрессора $\varphi(x)$ и рассчитать значения оценок вектора параметров $\hat{\beta}$, которые бы обеспечили правдоподобную оценку выходного сигнала датчика в виде (57).

Рассмотрим для начала одномерную регрессионную модель вида (47) с функцией регрессора $\varphi_1(x) = 4x$. Рассчитаем регрессор $H_1(x)$ для данных градуировочного эксперимента:

$$H_1(x) = \begin{bmatrix} -20 \\ -16 \\ -12 \\ -8 \\ -4 \\ 0 \\ 4 \\ 8 \\ 12 \\ 16 \\ 20 \end{bmatrix}.$$

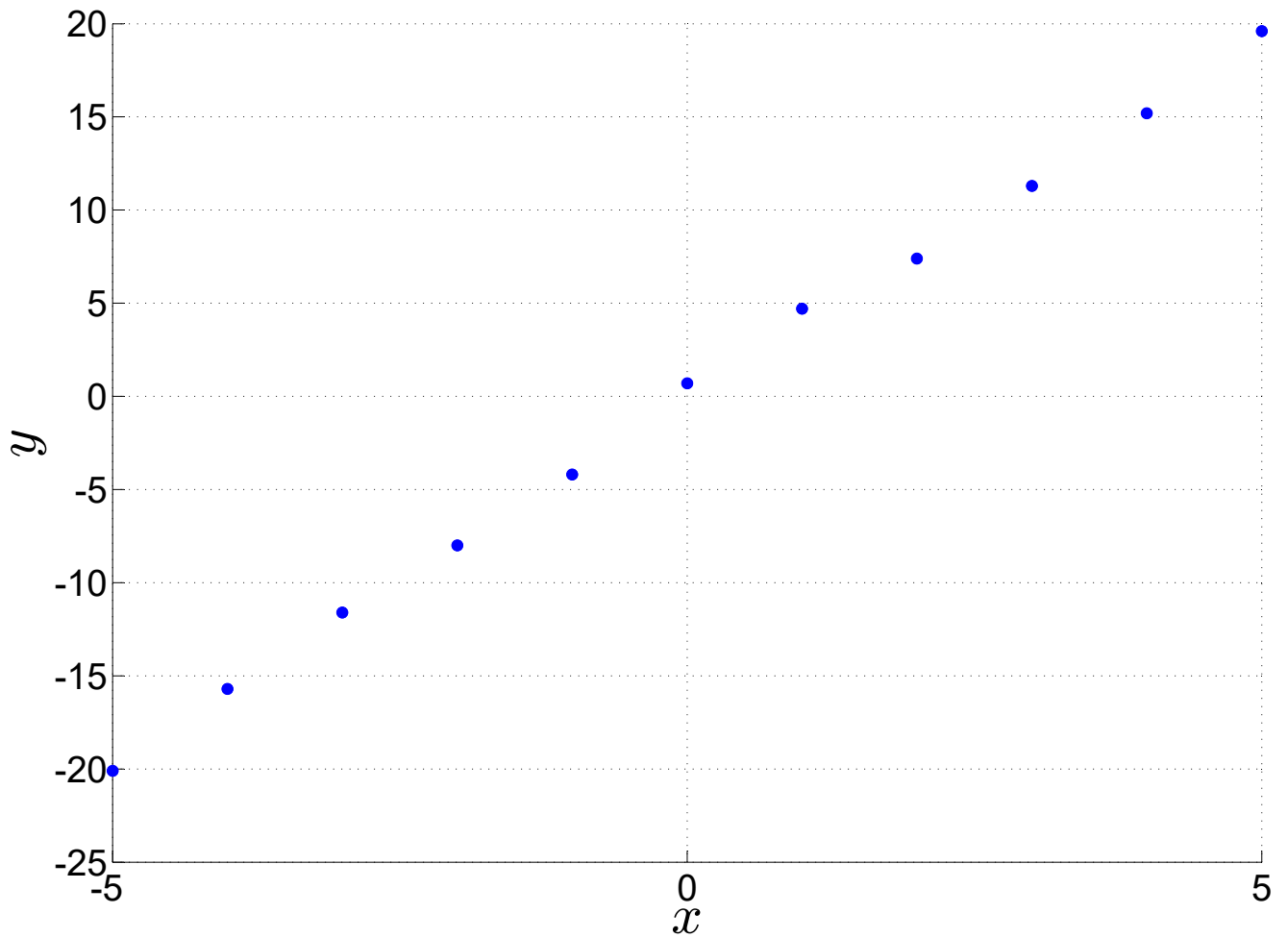


Рис. 11: Результаты градуировочного эксперимента.

Оценку $\hat{\beta}_1$ параметра β регрессионной модели преобразовательной характеристики вычислим в силу соотношения (56):

$$\hat{\beta}_1 = (H_1^T H_1)^{-1} H_1^T Y = 0.97.$$

Тогда оценка (прогноз) значений выходного сигнала датчика будет иметь вид

$$\hat{y}_1(x) = H_1(x) \hat{\beta} = \begin{bmatrix} -19.568 \\ -15.655 \\ -11.741 \\ -7.827 \\ -3.914 \\ 0 \\ 3.914 \\ 7.827 \\ 11.741 \\ 15.655 \\ 19.568 \end{bmatrix}.$$

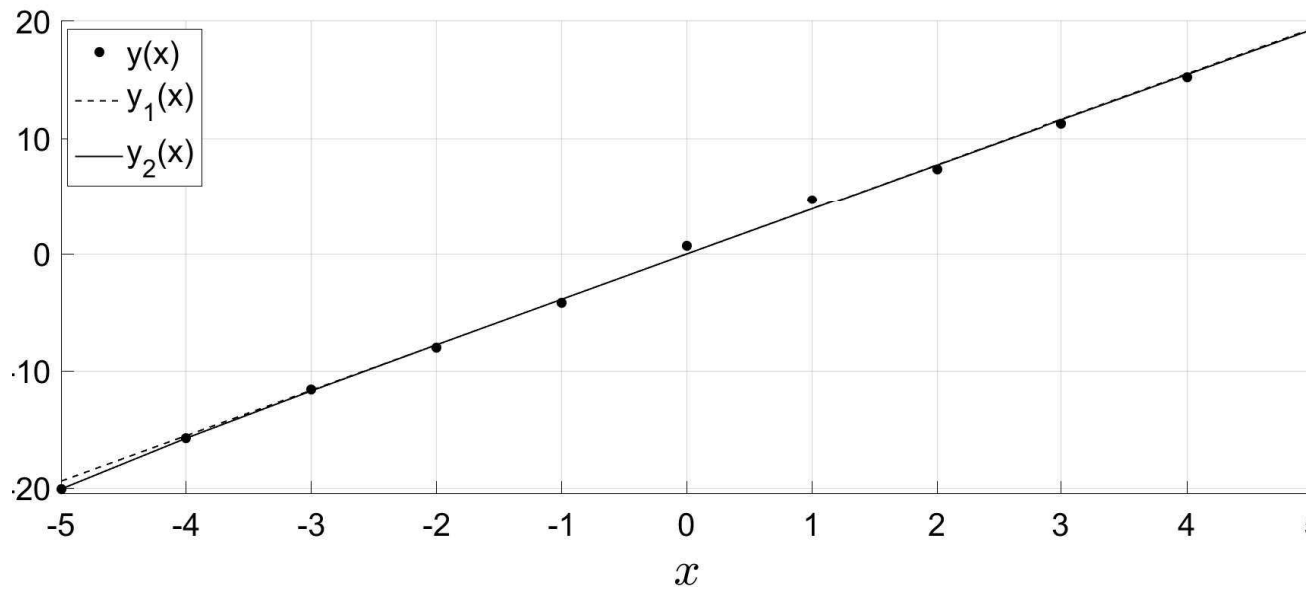


Рис. 12: Результаты градуировочного эксперимента.

Графическое сравнение полученной оценки $\hat{y}_1$ с данными градуировочного эксперимента представлено на рисунке 12. Видно, что оценка близка к результатам эксперимента. Степень этой близости оценим при помощи среднеквадратического отклонения оценки от данных градуировочного эксперимента, приняв за математическое ожидание оценки данные градуировочного эксперимента:

$$\sigma_1 = \sum_{i=1}^{10} \frac{\sqrt{(\hat{y}_{1i} - M[y_i])^2}}{10} = 0.365.$$

Улучшать точность можно путем расширения и/или корректировки функций регрессора. Дополним регрессор функцией $\phi_2(x) = \exp^{-x}$. Тогда регрессор

$H_2(x)$ примет вид

$$H_2(x) = \begin{bmatrix} -20 & 148.413 \\ -16 & 54.598 \\ -12 & 20.086 \\ -8 & 7.389 \\ -4 & 2.718 \\ 0 & 1 \\ 4 & 0.368 \\ 8 & 0.135 \\ 12 & 0.05 \\ 16 & 0.018 \\ 20 & 0.006738 \end{bmatrix}.$$

Оценку $\hat{\beta}_2$ вектора параметров β регрессионной модели преобразовательной характеристики вычислим в силу соотношения (56):

$$\hat{\beta}_2 = (H_2^T H_2)^{-1} H_2^T Y = \begin{bmatrix} 0.966 \\ 0.005081 \end{bmatrix}.$$

Тогда оценка (прогноз) значений выходного сигнала датчика будет иметь вид

$$\hat{y}_2(x) = H_2(x) \hat{\beta} = \begin{bmatrix} -20.083 \\ -15.74 \\ -11.699 \\ -7.769 \\ -3.88 \\ 0.005081 \\ 3.864 \\ 7.731 \\ 11.597 \\ 15.463 \\ 19.329 \end{bmatrix}.$$

Графическое сравнение полученной оценки $\hat{y}_2$ с данными градуировочного эксперимента представлено на рисунке 12. Видно, что оценка близка к результатам эксперимента. Степень этой близости оценим при помощи среднеквадратического отклонения оценки от данных градуировочного эксперимента, приняв за математическое ожидание оценки данные градуировочного эксперимента:

$$\sigma_2 = \sum_{i=1}^{10} \frac{\sqrt{(\hat{y}_{2i} - M[y_i])^2}}{10} = 0.31.$$

Можно увидеть, что прибавка в точности в 5% обеспечена за счет того, что оценка с регрессором $H_2(x)$ имеет большую точность в интервале от -5 до -3 .

Вопросы для самоконтроля

1. Какие показатели качества выделяют у датчиков физических величин в системах управления? На какие из них могут оказать влияние случайные возмущения?
2. Как может быть задана модель преобразовательной характеристики датчика?
3. Что такое градуировочный эксперимент и каковы его цель и задачи?
4. Что такое метод максимального правдоподобия и как он применяется в задаче оценки параметров преобразовательной характеристики датчика?

7 Введение в теорию эксперимента

При изучении сложных технических систем в настоящее время наиболее широко применяется системный подход, в основе которого лежит стремление принять во внимание максимально полный набор входных и выходных параметров объекта исследования, а также применение проблемно-ориентированного подхода. Одним из ярких примеров применения системного подхода может послужить математическая теория планирования эксперимента, методы которой имеют в своей основе развитие принципов многофакторного анализа. Благодаря своей надежности, эффективности и универсальности эти методы получили широкое распространение в самых разнообразных областях исследований, предполагающих планирование и проведение экспериментальных исследований [4].

Хотя математические методы планирования наиболее эффективно работают при исследовании физических объектов путем постановки опытов с этими объектами в реальных условиях их работы, планирование эксперимента, если под экспериментом иметь в виду любое действие с системой или ее моделью, позволяет существенно поднять эффективность решений на математических моделях как в линейной, так и нелинейной постановке задач. И в этом плане применение математической теории эксперимента при моделировании задач на вычислительных машинах позволяет поднять производительность научного труда на всех стадиях исследования, разработки и проектирования сложных систем.

Математические методы планирования эксперимента имеют в своей основе представление об объекте исследования как о так называемом «черном ящике». "Черный ящик" представляет собой систему, в которой известны только входные сигналы, которые задаются исследователем по определенным правилам, и выходные сигналы, которые регистрируются измерительной аппаратурой. О параметрах или структуре самого "черного ящика" при этом ничего не известно. В такой постановке задачи входные сигналы интерпретируются как независимые переменные или управляемые факторы и обозначаются $x_1, x_2, x_3, \dots, x_k$, а выходы (параметры процесса или количественные характеристики исследуемого объекта) интерпретируются как параметры (критерии) оптимизации $\eta_1, \eta_2, \eta_3, \dots, \eta_m$.

Для каждого фактора предусмотрен набор из нескольких значений или уровней, которые он может принимать в ходе эксперимента. Каждая уникальная комбинация уровней нескольких факторов характеризует одно из возможных состояний исследуемой системы, при этом в факторном пространстве каждой такой комбинации соответствует определенная точка (точка O на рисунке 13) и соответствующий отклик исследуемой системы. Таким образом, между

значениями факторов и реакцией исследуемой системы существует некоторая функциональная зависимость, называемая функцией отклика

$$\eta_l = f(\chi_1, \chi_2, \chi_3, \dots, \chi_k), \quad l = \overline{1, m}. \quad (59)$$

Геометрически функциональную зависимость (59) можно интерпретировать как некоторую поверхность в пространстве факторов, которую называют поверхностью отклика (см. рисунок 13).

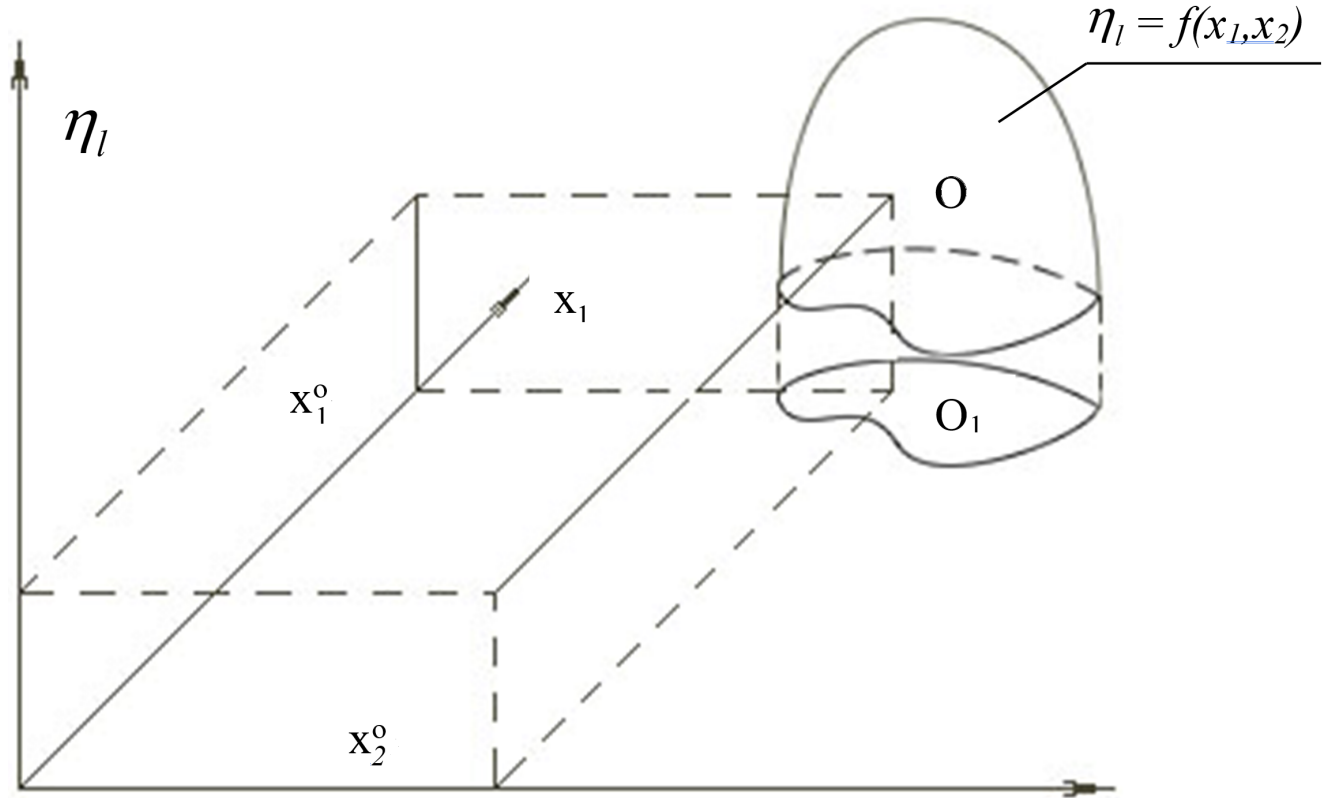


Рис. 13: Поверхность отклика.

Функция отклика может быть интерпретирована (с определенной долей условности) как математическая модель исследуемой системы. В общем случае вид функции (59) неизвестен, поэтому для описания поверхности отклика используют разложение этой функции в ряд вида

$$\eta = \beta_0 + \sum_{i=1}^K \beta_i x_i + \sum_{i \neq j} \beta_{ij} x_i x_j + \dots + \sum_{i \neq j \neq q} \beta_{ijk} x_i x_j x_q, \quad (60)$$

где $\beta_0, \beta_i, \beta_{ij}, \dots, \beta_{ijk}$ – коэффициенты регрессии, значения которых определяют форму поверхности отклика в изучаемой области. Такое уравнение обычно называют уравнением регрессии.

Таким образом, задача отыскания функции отклика может быть сведена к получению по экспериментальным данным выборочной оценки для η в виде

$$y = b_0 + \sum_{i=1}^K b_i x_i + \sum_{i \neq j} b_{ij} x_i x_j + \dots + \sum_{i \neq j \neq q} b_{ijk} x_i x_j x_q, \quad (61)$$

где $b_0, b_i, b_{ij}, \dots, b_{ijk}$ – оценки для теоретических коэффициентов регрессии.

В рамках поставленной задачи нахождения оценки функции отклика задача планирования эксперимента может быть поставлена как определение такого плана эксперимента, при котором за наименьшее число опытов можно было бы, используя составленные специальным образом сочетания уровней факторов, определить оценку функции отклика и оптимальные значения оценок коэффициентов регрессии.

Отметим основные возможности, которые дает исследователям теория эксперимента:

1. В план эксперимента при помощи методов математической статистики стало возможно вводить аналоги случайных величин путем рандомизации детерминированных факторов, которые слишком сложно либо вообще невозможно контролировать полностью, что дает возможность учитывать их влияние при помощи статистических методов.
2. Планирование эксперимента позволяет резко повысить эффективность эксперимента за счет оптимального использования пространства независимых переменных. Это означает, что параметры, интересующие исследователя, могут быть определены со значительно меньшей ошибкой, чем при традиционных методах исследования.
3. Методы математического планирования эксперимента позволяют построить такой план эксперимента, в котором содержание последующих опытов корректируется или даже существенно изменяется результатами предыдущих, что позволяет существенно сократить общее время исследования в случае, например, эксперимента по поиску оптимальных параметров функционирования системы (вместо перебора всех сочетаний уровней факторов результаты каждого опыта сужают область оптимальных параметров).
4. Использование методов математического планирования эксперимента позволяет представить результаты эксперимента в компактной, легко читаемой форме, а также предоставляет критерии точности и достоверности результатов эксперимента. Это позволяет не только удобно представлять результаты экспериментов при публикации, но и дает возможность сопоставлять и сравнивать результаты экспериментов, проведенных другими

исследователями в данной области, что также повышает эффективность исследований.

Факторами называются переменные величины, принимающие в некоторый момент времени определенное значение и соответствующие способам воздействия на объект. Они определяют как сам объект, так и его состояние.

Факторы могут быть количественными и качественными. Количественными факторами являются переменные величины, которые можно оценить количественно. Качественные факторы, если это нужно экспериментатору, всегда можно сделать количественными, построив условную порядковую шкалу для кодирования уровня качества числами натурального ряда. Наиболее простой пример такого кодирования – оценка качества по пятибалльной системе.

К факторам обычно предъявляют ряд требований. Так, факторы должны быть управляемыми, то есть для каждого фактора должна быть обеспечена возможность поддержания требуемого уровня в течение заданного периода времени в рамках эксперимента с заданной точностью. Также факторы должны воздействовать непосредственно на объект исследования или, иными словами, не быть функциями других переменных (особенно неконтролируемых). Отсюда вытекает также требование некоррелированности факторов, то есть уровень каждого фактора должен устанавливаться независимо от других факторов. Кроме того, запланированные для эксперимента сочетания факторов должны быть физически реализуемы.

Не менее важной, чем определение факторов, является задача выбора параметра оптимизации, которая определяется целью исследований. Цель исследований должна быть сформулирована очень четко и допускать количественную оценку. К параметру оптимизации, как и к факторам, обычно предъявляют ряд требований. Так, выбранный параметр оптимизации должен оценивать именно выход исследуемой системы, а не каких-либо ее подсистем, то есть быть эффективным, а также универсальным. Кроме того, параметр оптимизации должен допускать однозначное количественное выражение. Если последнее невозможно, качественное выражение параметра оптимизации кодируется численно (как правило, последовательностью натуральных чисел). Наконец, параметр оптимизации должен нести некоторый физический смысл, что позволит более корректно сформулировать результаты эксперимента.

Свойства и поведение исследуемой системы могут быть описаны различными в определенной степени эквивалентными моделями. Выбор конкретной модели должен быть согласован с целью исследования, а выбранная модель соответствовать ряду критериев. Главный из них – адекватность модели, которая отражает способность модели предсказать поведение исследуемой системы на необходимый для эксперимента промежуток времени с требуемой точностью.

Также модель должна быть содержательной и таким образом обеспечивать возможность не только описать уже известные свойства объекта исследования, но и выявить ранее неизвестные. Наконец, сложность модели следует подбирать исходя из условий эксперимента и имеющихся вычислительных мощностей. В теории математического планирования эксперимента наибольшее распространение получили полиномиальные модели.

7.1 Полный факторный эксперимент

Традиционным способом испытаний является однопараметрический эксперимент, при котором все переменные, кроме исследуемой, фиксируются на постоянных уровнях. Поочередная смена уровней всех переменных на выходе дает зависимости, не отражающие взаимодействия параметров. Факторное планирование позволяет устранить этот недостаток.

При факторном планировании число уровней для каждого из факторов не может быть меньше двух [4]. Количество уровней факторов и, соответственно, их уникальных комбинаций задает число возможных откликов исследуемой системы, в связи с чем число уровней редко выбирают большее трех. Экспериментальные планы, в которых все факторы варьируются на двух уровнях, называются планами типа 2^k , где k – количество факторов. Эксперимент, в ходе которого реализуются все возможные комбинации уровней всех факторов, называют полным факторным экспериментом (ПФЭ). Очевидно, что для того, чтобы исчерпать все возможные комбинации для факторов на двух уровнях, следует провести 2^k опытов.

Для облегчения расчета коэффициентов регрессии по результатам полного факторного эксперимента производят кодирование переменных согласно выражению

$$x_i = \frac{\bar{x}_i - \bar{x}_{i0}}{\bar{\lambda}_i}; \quad i = \overline{1, k}, \quad (62)$$

где x_i – кодированное значение i -го фактора, $\bar{x}_i$ – значение i -го фактора на одном из уровней в натуральных переменных, $\bar{x}_{i0}$ – значение i -го фактора на основном уровне, который определяется как среднее арифметическое уровней фактора (см. рисунок 14а), $\bar{\lambda}_i$ – натуральное значение интервала варьирования, величина которого равна разности между любым из уровней и основным уровнем фактора.

Путем выбора интервалов варьирования в пространстве факторов исследователем задается некоторая область, в рамках которой будет изучаться поведение системы. Такая область будет симметрична относительно своего центра, который соответствует нулевому уровню. Стоит отметить, что для качественных факторов нулевой уровень не имеет физического смысла.

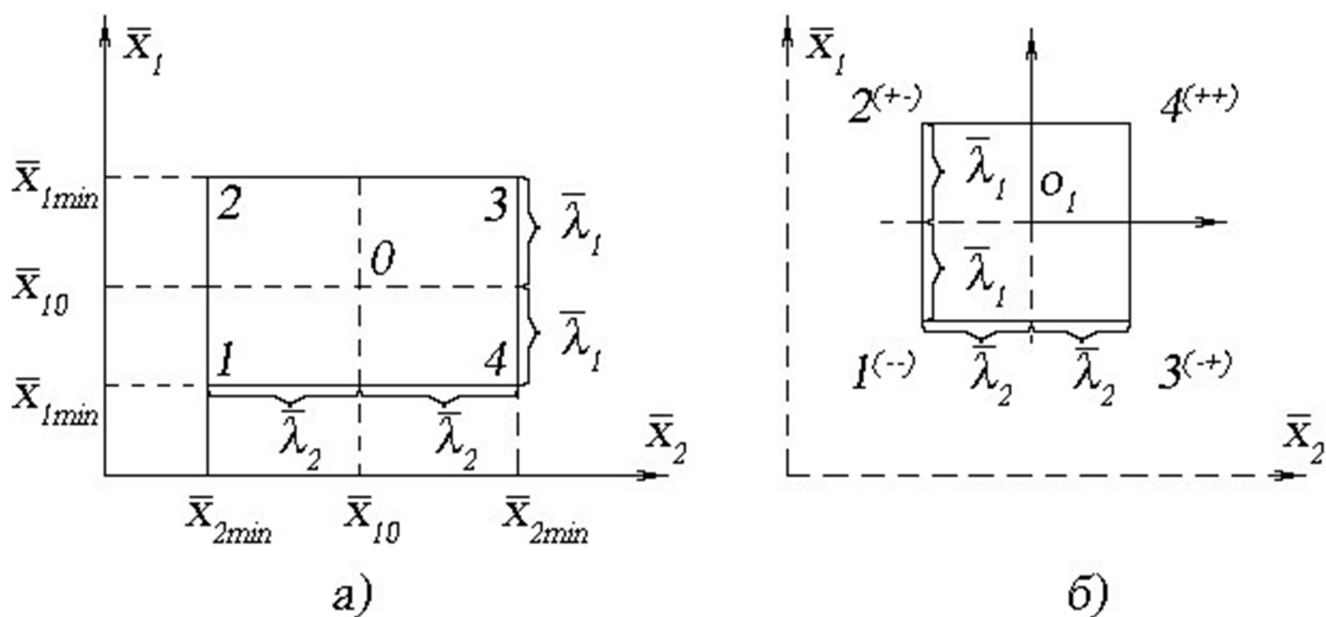


Рис. 14: Геометрическая интерпретация двухфакторного эксперимента в натуральных (а) и кодированных (б) переменных.

Кодирование переменных представляет собой линейное преобразование факторного пространства, заключающееся в переносе начала координат в нулевую точку (см. рисунок 14б) и выбор масштабов по осям в единицах интервалов варьирования. В относительных единицах значения факторов на верхнем и нижнем уровнях при таком преобразовании будут соответственно $+1$ и -1 , пределы варьирования кодированных переменных $\lambda_1 = \lambda_2 = 1$.

Запись значений границ интервалов в относительных единицах ± 1 приводит к стандартной форме записи матрицы планирования, использующей только знаки. План, например, полного факторного эксперимента ПФЭ- 2^2 показан в таблице 2, а геометрическое изображение плана в виде квадрата, вершины которого соответствуют номерам и условиям плана – на рисунке 14б. Для реализаций этого плана необходимо поставить $N = 2^k$ т.е. $N = 2^2 = 4$ опыта, а математическая модель будет иметь вид

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

и требует определения $N = 2^2 = 4$ коэффициентов.

План эксперимента ПФЭ- 2^2 , показанный в таблице 2, имеет один столбец с фиктивной переменной X_0 , которой во всех опытах присвоено значение $+1$, и столбец произведения X_1X_2 . При этом столбцы матрицы X_1 и X_2 используются непосредственно для планирования и определяют условия проведения опытов, а столбцы X_0 и X_1X_2 используются только для расчетов коэффициентов b_0 и b_{12} (коэффициент взаимодействия факторов X_1 и X_2 модели).

Таблица 2: План эксперимента ПФЭ-2²

№ опыта	Переменные				Выход
	x_0	x_1	x_2	x_3	
1	+	−	−	+	y_1
2	+	+	−	−	y_2
3	+	−	+	−	y_3
4	+	+	+	+	y_4

Таблица 3: План эксперимента ПФЭ-2³

№ опыта	Переменные			Выход
	x_1	x_2	x_3	
1	−	−	−	y_1
2	+	−	−	y_2
3	−	+	−	y_3
4	+	+	−	y_4
5	−	−	+	y_5
6	+	−	+	y_6
7	−	+	+	y_7
8	+	+	+	y_8

Матрица планирования для произвольного числа независимых переменных, для которых заданы два уровня, могут быть сформированы, например чередования знаков. Так, в таблице 3 приведен план эксперимента ПФЭ-2³, в котором знаки в столбце первого фактора чередуются, в столбце второго фактора меняются через две строки (аналогично со столбцами факторов X_1 и X_2 в таблице 2), а в столбце третьего фактора меняются через четыре строки. Данную тенденцию можно экстраполировать на любое число факторов, сменяя знаки через число строк, соответствующее степеням двойки. Геометрическая интерпретация плана эксперимента ПФЭ-2³ приведена на рисунке 15.

Плану ПФЭ присущи критерии оптимальности – ортогональность и ротатабельность. Ортогональность плана обеспечивает получение независимых оценок коэффициентов регрессии. Ротатабельность плана позволяет получить одинаковую дисперсию предсказанных значений функции отклика во всех равноудаленных от центра эксперимента точках.

Приведенные выше планы полных факторных экспериментов обладают рядом важных свойств. Во-первых, эти планы симметричны, что выражается в том, что все сочетания уровней факторов симметричны относительно центра плана (что наглядно показано на графических интерпретациях планов). Ма-

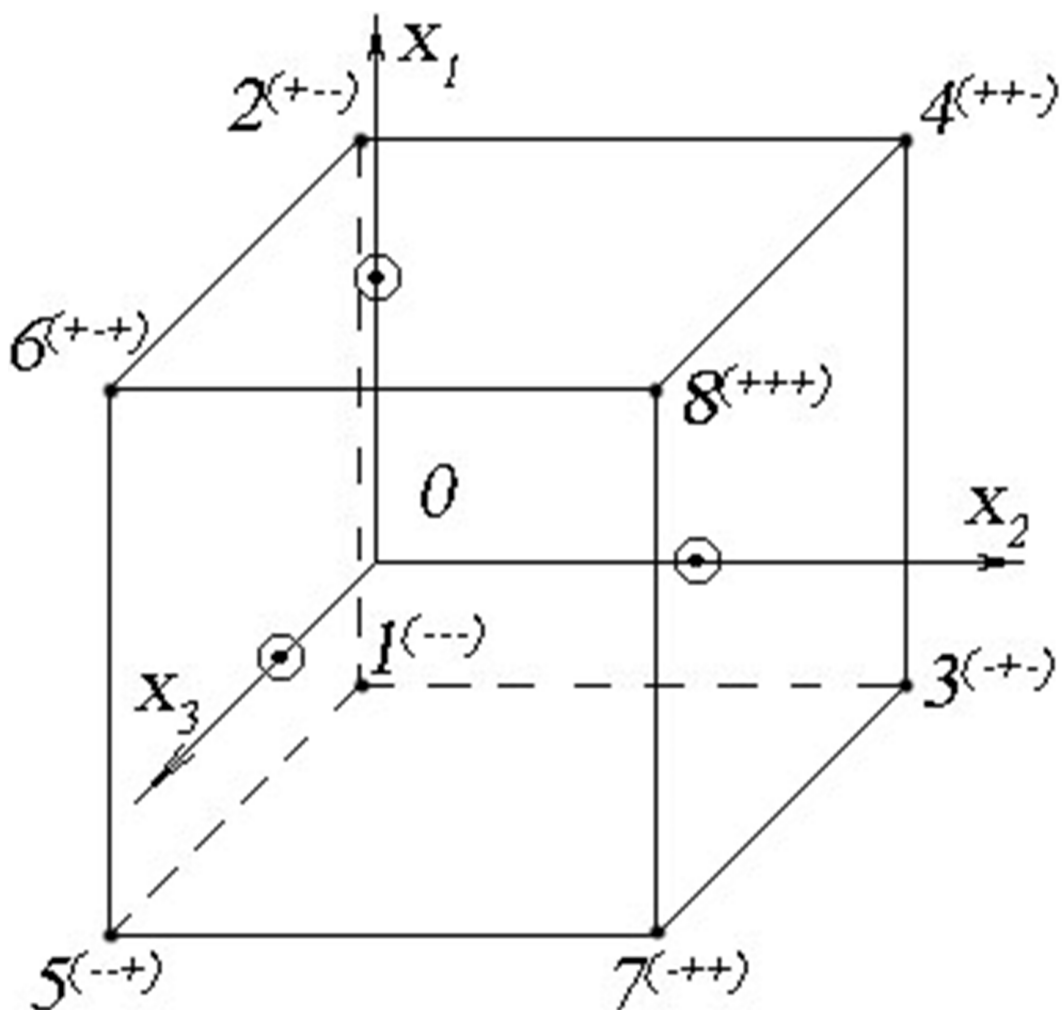


Рис. 15: Геометрическая интерпретация двухфакторного эксперимента в натуральных (а) и кодированных (б) переменных.

тематически это выражается в равенстве нулю сумм элементов каждого из столбцов матрицы планирования:

$$\sum_{u=1}^N x_{iu} = 0, \quad i = \overline{1, k}. \quad (63)$$

Во-вторых, для приведенных планов выполняется свойство нормировки, что выражается в том, что сумма квадратов элементов каждого столбца равно количеству строк в матрице планирования:

$$\sum_{u=1}^N x_{iu}^2 = N, \quad i = \overline{1, k}. \quad (64)$$

Наконец, для приведенных выше планов выполняется свойство ортогональности, что означает, что сумма попарных произведений элементов любых двух

столбцов матрицы планирования равна нулю:

$$\sum_{u=1}^N x_{iu}x_{ju} = 0, \quad i \neq j, \quad i, j = \overline{1, k}. \quad (65)$$

Матрицы планирования показывают, в каких точках факторного пространства надо произвести измерения выхода (отклика), т.е. произвести опыты. А столбцы X_i ($i = 1, 2, 3, \dots$) определяют условия проведения опытов. Например (см. таблицу 2): опыт по третьей строке ($U = 3$) проводится при условии, когда фактор X_1 , находится на нижнем уровне, а фактор X_2 – на верхнем. В целях «рандомизации» опыты зачастую проводятся в произвольном порядке, а не в том, в котором они следуют в таблице плана эксперимента.

Многофакторные эксперименты чаще всего проводятся на реальном оборудовании. С технической точки зрения фактор удобно представить параметрами пассивных элементов R, L, C , токами и напряжениями, а условия окружающей среды температурой, давлением, влажностью и т.д. В большинстве случаев факторы варьируются на двух уровнях $\bar{x}_{iH}$, $\bar{x}_{iO}$, и в целях калибровки часто необходимо измерять отклик в центре локальной области исследований, зафиксировав значения всех факторов на основном уровне $\bar{x}_i = \bar{x}_{iO}$ или в кодированных переменных $x_i = 0$. Для этого в экспериментальном оборудовании должна быть предусмотрена возможность варьирования каждого из факторов на трех уровнях. Данная задача может быть решена, например, с помощью набора отдельных элементов, которые переключаются в зависимости от требуемого уровня в соответствии с условиями проведения опытов по строкам план-матрицы эксперимента.

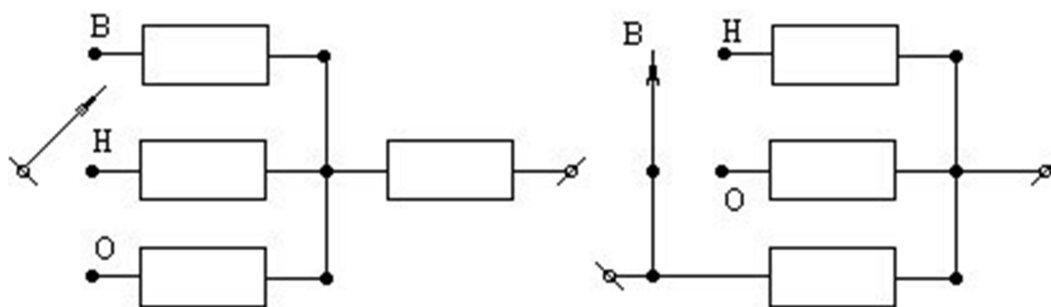


Рис. 16: Примеры схем варьирования трех уровней факторов при помощи переключения сопротивлений.

На рисунке 16 приведены примеры схем реализации верхнего (В), нижнего (Н) и основного (О) уровней с помощью последовательного и параллельного соединений. А само переключение в простейшем случае осуществляется с помощью шагового искателя.

7.2 Дробный факторный эксперимент

В полном факторном эксперименте опыты реализуются при всех возможных наборах уровней факторов. Например, для двухуровневого факторного эксперимента опыты проводятся в $N = 2^K$ точках факторного пространства. При этом очевидно, что с введением каждого нового фактора число опытов в матрице ПФЭ- 2^K удваивается (при $K = 2$, $N = 4$, при $K = 3$, $N = 8$, при $K = 4$, $N = 16$ и т.д.) [4].

На основе опытов в N точках факторного пространства можно найти $N = 2^K$ коэффициентов уравнения регрессии. Если число факторов $k \geq 4$, эффекты взаимодействия высокого порядка становятся статистически незначимыми. Практика эксперимента показывает, что при $k = q \geq 4$ влияние сомножителей $x_1, x_2, \dots, x_4$ на отклик взаимно компенсируется, и соответствующий коэффициент $b_{123\dots q}$ становится статистически незначимым. Это означает, что для экспериментов со значительным количеством факторов возникает потребность в построении таких планов экспериментов, которые бы позволяли оценивать в основном линейные эффекты факторов, в меньшей степени эффекты их парных взаимодействий и, при необходимости, эффекты тройных взаимодействий факторов, исключая из оценки функции отклика взаимодействия факторов более высоких порядков. При таком подходе сокращается количество необходимых опытов, а значит и временные и трудозатраты.

Рассмотрим пример эксперимента с 10 факторами и будем исходить из того, что необходимо оценить только линейные эффекты факторов и их парные взаимодействия, а остальными взаимодействиями в силу статистической незначительности будем пренебрегать. Тогда количество опытов, которые необходимо произвести (а значит, и количество коэффициентов, которые нужно оценить), будет равно

$$N_{\text{оп}} = C_{10}^0 + C_{10}^1 + C_{10}^2 = 1 + 10 + 45 = 56.$$

При этом в полном факторном эксперименте требуется оценить $N = 2^{10} = 1024$ коэффициентов. Можно сделать вывод, что при использовании введенных предположений затраты времени и трудозатраты сократятся в $1024/56 = 17.3$ раза по сравнению с полным факторным экспериментом.

Исходя из вышесказанного можно заключить, что в экспериментах с большим количеством факторов целесообразно сокращать количество оцениваемых коэффициентов регрессии по сравнению с полным факторным экспериментом. Тогда возникает задача построения таких планов экспериментов, в которых число опытов было бы равно либо несущественно превышало количество оцениваемых коэффициентов регрессии. Такие эксперименты называются дробными факторными экспериментами (ДФЭ) 2^{K-p} . Количество опытных точек в дробном факторном эксперименте ДФЭ- 2^{K-p} в 2^p раз меньше, чем в пол-

Таблица 4: План эксперимента ДФЭ-2³⁻¹

№ опыта	Переменные				Выход
	x_0	x_1	x_2	$x_3 = x_1x_2$	
1	+	−	−	+	y_1
2	+	+	−	−	y_2
3	+	−	+	−	y_3
4	+	+	+	+	y_4

ном факторном эксперименте ПФЭ-2^K. Здесь p – число факторов, введенных путем замены статистически незначимых взаимодействий. В силу того, что дробный факторный эксперимент ДФЭ-2^{K-p} есть часть полного факторного эксперимента ПФЭ-2^K, то ДФЭ называют также дробной репликой факторного эксперимента. Например, дробный факторный эксперимент, образующий половину ПФЭ-2^K, обозначают ДФЭ-2^{K-1} и называют полурепликой ПФЭ-2^K, ДФЭ-2^{K-2} называют 1/4-репликой ПФЭ-2^K и т. д.

Построение плана дробного факторного эксперимента рассмотрим на основе описанных выше полных факторных экспериментов ПФЭ-2² и ПФЭ-2³. По аналогии могут быть построены и другие ДФЭ. Для начала поставим задачу построения плана ДФЭ-2³⁻¹. В таком эксперименте план строится на четырех опытных точках $N = 2^{3-1} = 2^2 = 4$. Ядром плана является ПФЭ-2², что позволяет принять модель функции отклика в виде

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

При условии, что взаимодействие x_1x_2 является статистически незначительным, соответствующий столбец в плане ПФЭ-2² (см. таблицу 2) можно использовать для добавления некоторого нового фактора X_3 и, таким образом, построить план ДФЭ-2³⁻¹. ДФЭ-2³⁻¹ является полурепликой ПФЭ-2² и позволяет получить трехфакторную модель

$$y = b_0 + b_1x_1 + b_2x_2 + b_3x_3$$

по результатам четырех опытов (см. таблицу 4).

Далее рассмотрим задачу построения плана дробного факторного эксперимента ДФЭ-2⁴⁻¹. План ДФЭ-2⁴⁻¹ строится на основе полного факторного эксперимента ПФЭ-2³, $N = 2^3 = 8$. При условии, что тройное взаимодействие $X_1X_2X_3$ является статистически незначительным, те же восемь опытов можно использовать для построения четырехфакторной модели

$$y = b_0 + b_1x_1 + b_2x_2 + b_{12}x_1x_2 + b_3x_3 + b_{23}x_2x_3 + b_4x_4.$$

Таблица 5: План эксперимента ДФЭ-2⁴⁻¹

№ опыта	Переменные								Выход
	x_0	x_1	x_2	x_1x_2	x_3	x_1x_3	x_2x_3	$x_4 = x_1x_2x_3$	
1	+	-	-	+	+	+	-	-	y_1
2	+	+	-	-	-	+	+	+	y_2
3	+	-	+	-	+	-	+	+	y_3
4	+	+	+	+	-	-	-	-	y_4
5	+	-	-	+	-	-	+	+	y_5
6	+	+	-	-	+	-	-	-	y_6
7	+	-	+	-	-	+	-	-	y_7
8	+	+	+	+	+	+	+	+	y_8

Таблица 6: План эксперимента ДФЭ-2⁵⁻²

№ опыта	Переменные								Выход
	x_0	x_1	x_2	x_1x_2	x_3	x_1x_3	x_2x_3	$x_5 = x_1x_2x_3$	
1	+	-	-	+	-	+	+	-	y_1
2	+	+	-	-	-	-	+	+	y_2
3	+	-	+	-	-	+	-	+	y_3
4	+	+	+	+	-	-	-	-	y_4
5	+	-	-	+	+	-	-	+	y_5
6	+	+	-	-	+	+	-	-	y_6
7	+	-	+	-	+	-	+	-	y_7
8	+	+	+	+	+	+	+	+	y_8

Полную план-матрицу эксперимента ДФЭ-2⁴⁻¹ иллюстрирует таблица 5. ДФЭ-2⁴⁻¹ является полурепликой ПФЭ-2³.

На базе ПФЭ-2³ можно получить и более насыщенные планы, например, ДФЭ-2⁵⁻², т.е. 1/4 реплики ПФЭ-2⁵ этом плане выполняется $N = 8$ вместо $N = 32$ ПФЭ-2⁵. Модель ДФЭ-2⁵⁻² будет иметь вид

$$y = b_0 + b_1x_1 + b_2x_2 + b_{12}x_1x_2 + b_3x_3 + b_4x_4 + b_{23}x_2x_3 + b_5x_5.$$

План ДФЭ-2⁵⁻² иллюстрируется таблицей 6.

Дробные факторные планы ДФЭ-2^{K-p} позволяют существенно сократить количество опытов. Однако при построении таких планов надо иметь в виду, что оценки коэффициентов регрессии будут смешанными. Так происходит потому, что разные эффекты в матрице планирования дробного факторного эксперимента отражают одни и те же столбцы. Так, например, в матрице ДФЭ-2⁴⁻¹ (см. таблицу 5) столбец для x_1 полностью повторяет столбец для взаимо-

действия $x_2x_3x_4$, а столбец для x_3 совпадает со столбцом для взаимодействия $x_1x_2x_4$. Аналогичным образом столбец для взаимодействия x_1x_2 будет совпадать со столбцом x_3x_4 . Это означает, что оценка, полученная в результате ДФЭ- 2^{4-1} , на самом деле является смешанной оценкой

$$b_1 \rightarrow \beta_1 + \beta_{234},$$

где β_1 – истинное значение коэффициента регрессии в уравнении модели. Аналогичным образом

$$b_2 \rightarrow \beta_2 + \beta_{134},$$

$$b_3 \rightarrow \beta_3 + \beta_{124},$$

$$b_4 \rightarrow \beta_4 + \beta_{123}.$$

Смешивание оценок в матрице ДФЭ- 2^{K-p} задается определяющим контрастом, который определяется из соотношения, описывающего, с каким взаимодействием в матрице полного факторного эксперимента ПФЭ- 2^K смешан дополнительно введенный линейный эффект. В рассматриваемом случае ДФЭ- 2^{4-1} (см. таблицу 5) это соотношение (называемое генерирующим соотношением) было $x_4 = x_1x_2x_3$. Умножая обе части равенства на x_4 , получим $x_4^2 = x_1x_2x_3x_4$. Результат описанной операции называется определяющим контрастом. С помощью определяющего контраста можно определить, с каким эффектом смешан любой эффект в выбранной матрице. Для этого достаточно умножить обе части равенства на интересующий фактор. Например, для фактора x_2 получим

$$x_2 = x_1x_2^2x_3x_4 = x_1x_2x_3,$$

а для взаимодействия x_1x_2 получим

$$x_1x_2 = x_1^2x_2^2x_3x_4 = x_3x_4.$$

Если включить в эксперимент еще один фактор, то получим четверть-реплику ПФЭ- 2^5 или ДФЭ- 2^{5-2} (см. таблицу 6). Здесь $J_1 = X_4 = X_1X_3$ и $J_2 = X_5 = X_1X_2X_3$. Тогда определяющий контраст $X_4 = X_1X_3$ или $1 = X_1X_2X_4$, а также $X_5 = X_1X_2X_3$ или $1 = X_1X_2X_3X_5$. Перемножив два определяющих контраста, получим

$$1 = X_1X_2X_4 = X_1X_2X_3X_5. \quad (66)$$

При наличии двух и более генераторов наряду с понятием контраста используется понятие обобщенного определяющего контраста. Обобщенный определяющий контраст имеет вид $J_{оп} = J_1 = J_2 = J_1J_2$.

$$1 = X_1X_2X_4 = X_1X_2X_3X_5 = X_2X_4X_5. \quad (67)$$

В соответствии с контрастом (66) в результате опытов, поставленных по плану ДФЭ-2⁵⁻² (см. таблицу 6), будут получены оценки коэффициентов регрессии, смешанные следующим образом:

$$b_2 \rightarrow \beta_2 + \beta_{135} + \beta_{45} + \beta_{34},$$

$$b_3 \rightarrow \beta_3 + \beta_{125} + \beta_{14} + \beta_{245},$$

$$b_4 \rightarrow \beta_4 + \beta_{1235} + \beta_{25} + \beta_{13},$$

$$b_5 \rightarrow \beta_5 + \beta_{123} + \beta_{134} + \beta_{24}.$$

С учетом того, что в значительном числе случаев тройные взаимодействия статистически незначимы, а также того, что некоторые двойные взаимодействия также априори являются статистически незначимыми (например, β_{14} , где x_1 – время регулирования и x_2 – давление окружающей температуры; x_1 – время успокоения колебаний стрелки указательного прибора и x_2 – толщина стенки корпуса прибора и т. д.), оценки подучаются "чистыми". А в приведенном примере объем эксперимента сокращается в четыре раза. Таким образом, исследователь должен максимально использовать априорные сведения, чтобы перейти от ПФЭ-2^K к ДФЭ-2^{K-p} для повышения эффективности своей работы.

7.3 Обработка результатов многофакторного эксперимента

Результаты многофакторного эксперимента являются выборкой из некоторой генеральной совокупности, т. е. являются случайными величинами. Основная задача обработки результатов многофакторного эксперимента заключается в оценке параметров принятой модели и проверке гипотезы об адекватности модели (её соответствии) результатам эксперимента. В случае, если использованная в ходе эксперимента модель не прошла проверку на адекватность, следует использовать иную модель либо провести новый эксперимент, например, с другими параметрами варьирования факторов. Процедура обработки результатов эксперимента хорошо формализована и может осуществляться при помощи вычислительных средств, однако на некоторых этапах обработки необходимо принимать не формальные, а творческие решения. При этом стоит помнить, что ни один метод обработки не поможет, если данные эксперимента содержат грубые ошибки или необоснованные решения. Кроме того, нарушение установленной процедуры обработки данных эксперимента может также привести к ложным выводам.

Основной задаче обработки данных многофакторного исследования подчинен ряд следующих частных задач:

Таблица 7: Результаты многофакторного эксперимента

№ опыта	X_{1u}	X_{2u}	...	X_{iu}	...	X_{ku}	Y_{u1}	Y_{u2}	...	Y_{ui}	...	Y_{um}
	2	3	4	5	6	7	8	9	10	11	12	13
1	—	—	...	—	...	—	Y_{11}	Y_{12}	...	Y_{1i}	...	Y_{1m}
2	+	—	...	—	...	—	Y_{21}	Y_{22}	...	Y_{2i}	...	Y_{2m}
3	...	...	.	.	.	.	.	.	.	.	.	.
...	...	...	.	.	.	.	.	.	.	.	.	.
...	...	...	.	.	.	.	.	.	.	.	.	.
...	—	+	...	+	...	+	Y_{u1}	Y_{u2}	...	Y_{ui}	...	Y_{um}
...	...	...	.	.	.	.	.	.	.	.	.	.
...	...	...	.	.	.	.	.	.	.	.	.	.
...	...	...	.	.	.	.	.	.	.	.	.	.
...	+	+	...	+	...	+	Y_{N1}	Y_{N2}	...	Y_{Ni}	...	Y_{Nm}

1. Оценка математических ожиданий и дисперсий выхода в отдельных точках факторного пространства или строках плана $U = 1, 2, \dots, N$. Чаще эти оценки называются построчными средними $\bar{y}_u$, построчными дисперсиями S_u^2 (далее под термином дисперсия будем понимать оценку дисперсии).
2. Проверка однородности статистического материала в целях исключения грубых ошибок.
3. Проверка однородности статистических дисперсий.
4. Определение дисперсии воспроизводимости.
5. Проверка информативности эксперимента.
6. Определение коэффициентов модели.
7. Проверка значимости bi-коэффициентов модели.
8. Проверка адекватности модели.

Результаты полного и дробного факторного эксперимента удобно оформлять в виде таблицы 7 [4]. Таблица в графах 2-7 содержит план эксперимента X_n , в остальных графах записываются результаты опытов. Предполагается, что во всех точках $U = 1, 2, \dots, N$ делается m повторов, причем $m = 3$, или в одной из строк, а лучше в центре локальной области исследований делается не менее десяти повторов.

Еще одним инструментом обработки результатов многофакторного эксперимента является вычисление построчных средних и дисперсий. Построчное среднее вычисляется в силу соотношения

$$\bar{y}_u = \frac{1}{m} \sum_{j=1}^m y_{uj}; \quad (68)$$

Дисперсия вычисляется в силу соотношения

$$S_u^2 = \frac{1}{m-1} \sum_{j=1}^m (y_{uj}^2 - m\bar{y}_u^2), \quad (69)$$

где m – число повторов. Полученные в результате приведенных выше расчетов данные заносятся в таблицу 7 и используются в дальнейшей обработке данных.

Следующим шагом в обработке результатов эксперимента является исключение грубых ошибок. Грубые ошибки могут внести существенные искажения в результаты эксперимента. Существует ряд методов исключения грубых ошибок, но чаще используется так называемый r -критерий. В соответствии с этим критерием данные строки, в которой подозревается промах, располагаются в порядке возрастания $y_{u1} < y_{u3} < y_{u2} < \dots < y_{um}$, где m – число параллельных опытов. Одно из крайних значений, например, y_{u1} считается подозрительным, если оно сильно отличается от всех остальных. В этом случае необходимо проверить основную статистическую гипотезу H_0 : y_{u1} принадлежит остальной совокупности. Альтернативной гипотезой является гипотеза H_1 : y_{u1} не принадлежит остальной совокупности, т. е. его резкое отличие объясняется грубой ошибкой. Доля проверки гипотезы H_0 определяет расчетное значение критерия:

$$r_{\min} = \frac{\bar{y}_u - y_{u1}}{s_u \sqrt{m(m-1)}}, \quad (70)$$

если сомнительным считается наименьшее в строке значение y_{u1} , и

$$r_{\max} = \frac{y_{um} - \bar{y}_u}{s_u \sqrt{m(m-1)}}, \quad (71)$$

если сомнительным считается наибольшее в строке y_{um} . Величина r_{kp} по числу степеней $f_u = m - 1$ уровню значимости берется из справочной таблицы r -распределения. Если расчетное значение $r_{\min}$ или $r_{\max}$ окажется меньше r_{kp} , гипотеза H_0 принимается с надежностью вывода $p = 1 - \alpha$, т. е. сомнительное крайнее значение отклика считается принадлежащим остальному ансамблю y_{uj} . В противном случае при $r_{\min}$ или $r_{\max} > r_{kp}$ гипотеза H_0 в отклоняется, т. е. резко выделяющиеся выборки y_{u1} или y_{um} из дальнейшей обработки

исключаются, и в данной строке остается $m - 1$ наблюдений. Это приводит к нарушению ортогональности эксперимента, что затрудняет использование метода наименьших квадратов. Поэтому для применения стандартных процедур обработки данных надо исключить наблюдение, признанное промахом и провести в этой точке дополнительное измерение.

Ещё одним этапом обработки результатов эксперимента служит проверка однородности построчных дисперсий. Цель этой проверки состоит в том, чтобы установить, являются ли измерения отклика во всех точках (строчках) плана одинаково точными. Если разброс y_{uj} в некоторой u -й строчке объясняется не только ошибкой измерения, но и шумом самого объекта исследований, проверяется однородность рассеивания, обусловленная обеими причинами. Понятие однородности нескольких оценок дисперсий S_1^2, S_2^2 означает, что все величины S_u^2 являются оценками одной и той же дисперсии σ_u^2 , которая называется дисперсией воспроизводимости. В этом случае различия между оценками S_u^2 , $u = 1, 2, 3, \dots, N$ объясняются их случайным характером.

Для проверки однородности оценок S_u^2 применяют f -критерий, критерий Кохрена G и реже критерий Бартлетта. Наиболее строгим является критерий Кохрена, когда число поворотов во всех строках постоянно: $m = \text{const}$, $u = 1, 2, \dots, N$:

$$G_p = \frac{S_{u\max}^2}{\sum_{u=1}^N S_u^2}, \quad (72)$$

где $S_{u\max}^2$ – наибольшая построчная дисперсия.

С величиной G связаны два параметра: $f_u = m - 1$ – число степеней свободы дисперсии $S_{u\max}^2$ и $f_\Sigma = N$ – число степеней свободы суммы, стоящей в знаменателе. По числам f_u и $f_\Sigma = N$ и уровню значимости α (обычно 0.05) из распределения Кохрена находят G_{kp} . Гипотеза об однородности дисперсий принимается, если $G_p \leq G_{kp}$. При $G_p > G_{kp}$ дисперсии признаются неоднородными и принимаются меры для достижения их однородности, например, стабилизацией условий эксперимента, использованием одного и того же прибора для измерения отклика и т. д.

Выполнение требования однородности построчных дисперсий позволяет определить дисперсию воспроизводимости как среднеарифметическое построчных дисперсий:

$$S_y^2 = \frac{1}{N} \sum_{u=1}^N S_u^2 = \frac{1}{N(m-1)} \sum_{u=1}^N \sum_{j=1}^m (y_{uj}^2 - m\bar{y}_u^2). \quad (73)$$

Величина $S_y = \sqrt{S_y^2}$ является оценкой среднеквадратического отклонения и называется ошибкой опытов. Выражение (73) может применяться, если в

каждой строке проведено $m > 1$ параллельных опытов. Если $m = 1$, т. е. в каждой строке проведен один опыт, то дисперсию воспроизводимости можно определить по ряду параллельных опытов в одной и той же области эксперимента, например, в центре локальной области исследования.

В некоторых лабораторных экспериментах (например, при компьютерной симуляции) повторные измерения в отдельных точках плана дают один и тот же результат. Тогда для расчета дисперсии воспроизводимости можно воспользоваться метрологическими характеристиками измерительных приборов и устройств. В паспортных данных приборов указывается класс точности (K , % от предела измерения $A_{\text{шк}}$). Это позволяет определить максимальную ошибку измерения:

$$\Delta_{\text{max}} = A_{\text{шк}} \frac{k}{100}. \quad (74)$$

Случайная ошибка прибора подчиняется нормальному закону распределения. В машиностроении считается, что $\Delta_{\text{max}} = 3\sigma_y$, при этом вероятность попадания в интервал $-3\sigma_y \leq y \leq 3\sigma_y$ $P = 0.9973$ и является технической единицей.

В радиоэлектронной аппаратуре стабильность активных и пассивных элементов значительно ниже, и надежность $= 0.95$ вполне приемлема. Для данной вероятности $\Delta_{\text{max}} = 2\sigma_y$. Используя величину Δ_{max} , можно получить

$$\sigma_y^2 = \left(\frac{A_{\text{шк}} k}{300} \right)^2, \quad (75)$$

или

$$\sigma_y^2 = \left(\frac{A_{\text{шк}} k}{200} \right)^2. \quad (76)$$

По метрологическим характеристикам вычисляется не оценка S_y^2 , а дисперсия σ_y^2 .

Еще одним этапом оценки результатов эксперимента является вычисление b_i коэффициентов и проверка их значимости. Определение b_i -коэффициентов состоит в отыскании точечных оценок β_i -коэффициентов параметров теоретической модели (60). В результате расчета b_i -коэффициентов и проверки их статистической значимости получается эмпирическое уравнение регрессии – математическая модель (61).

Для расчета b_i -коэффициентов используется общая процедура метода наименьших квадратов (МНК), а также формализованные алгоритмы, например, метод Иэтса на основе метода наименьших квадратов

$$B = (X^T X)^{-1} X^T Y. \quad (77)$$

Раскрывая выражение (77) на примере двухфакторного эксперимента, получим

$$(X^T X)^{-1} = \left(\begin{bmatrix} +1 & +1 & +1 & +1 \\ -1 & +1 & -1 & +1 \\ -1 & -1 & +1 & +1 \\ +1 & -1 & -1 & +1 \end{bmatrix} \begin{bmatrix} +1 & -1 & -1 & +1 \\ +1 & +1 & -1 & -1 \\ +1 & -1 & +1 & -1 \\ +1 & +1 & +1 & +1 \end{bmatrix} \right)^{-1} = \begin{bmatrix} 1/4 & 0 & 0 & 0 \\ 0 & 1/4 & 0 & 0 \\ 0 & 0 & 1/4 & 0 \\ 0 & 0 & 0 & 1/4 \end{bmatrix};$$

$$X^T Y = \begin{bmatrix} +1 & +1 & +1 & +1 \\ -1 & +1 & -1 & +1 \\ -1 & -1 & +1 & +1 \\ +1 & -1 & -1 & +1 \end{bmatrix} \begin{bmatrix} y \\ y \\ y \\ y \end{bmatrix} = \begin{bmatrix} +y & +y & +y & +y \\ -y & +y & -y & +y \\ -y & -y & +y & +y \\ +y & -y & -y & +y \end{bmatrix}$$

Отсюда

$$b_0 = \frac{y_1 + y_2 + y_3 + y_4}{4}; \quad b_1 = \frac{-y_1 - y_2 + y_3 + y_4}{4};$$

$$b_2 = \frac{-y_1 + y_2 - y_3 + y_4}{4}; \quad b_3 = \frac{y_1 - y_2 - y_3 + y_4}{4}.$$

В общем виде для ортогональных планов имеем

$$b_0 = \frac{1}{\sum_{u=1}^N x_{iu}^2} \sum_{u=1}^N x_{0u} \bar{y}_u = \frac{1}{N} \sum_{u=1}^N x_{0u} \bar{y}_u; \quad (78)$$

$$b_i = \frac{1}{N} \sum_{u=1}^N x_{iu} \bar{y}_u; \quad b_{ij} = \frac{1}{N} \sum_{u=1}^N x_{iu} x_{ju} \bar{y}_u, \quad i, j = \overline{1, N}, i \neq j, \quad (79)$$

где x_{iu} – величина, расположенная на пересечении i -го столбца и u -й строки матрицы планирования.

Если опыты ставятся в соответствии с планом ПФЭ- 2^K , то для расчета коэффициентов регрессии удобно использовать формализованную схему, предложенную Иэйтсом. Согласно этой схеме, вычисляют сначала парные суммы $y_1 + y_2$; $y_3 + y_4$ и так далее для всех y_u , затем разности $y_2 - y_1$; $y_4 - y_3$ и так далее также для всех y_u , располагая результаты в такой же вертикальный столбец рядом с вектором-столбцом y_u (1-й шаг в таблице 8). С полученными таким образом двумя значениями повторяют операции попарного сложения и вычитания и строят следующий столбец. Общее число таких столбцов должно равняться $+1$, включая исходный вектор-столбец результатов y_u , результаты последнего столбца, деленные на число опытов в матрице, и будут оценками для коэффициентов регрессии b_i .

Таблица 8: Таблица для расчета коэффициентов регрессии по методу Иэйтса для ПФЭ-2³

№	$X_1 X_2 X_3$	y_u	1 шаг	2 шаг	3 шаг	Обозначения строк
1	— — —	Y_1	$Y_1 + Y_2$	$Y_1 + Y_2 + Y_3 + Y_4$	$Y_1 + Y_2 + Y_3 + Y_4 + Y_5 + Y_6 + Y_7 + Y_8$	"1"
2	+ — —	Y_2	$Y_3 + Y_4$	$Y_5 + Y_6 + Y_7 + Y_8$	$Y_2 - Y_1 + Y_4 - Y_3 + Y_6 - Y_5 + Y_8 - Y_7$	X_1
3	— + —	Y_3	$Y_5 + Y_6$	$Y_2 - Y_1 + Y_4 - Y_3$	$Y_3 + Y_4 - Y_1 - Y_2 + Y_7 + Y_8 - Y_5 - Y_6$	X_2
4	+ + —	Y_4	$Y_7 + Y_8$	$Y_6 - Y_5 + Y_8 - Y_7$	$Y_4 - Y_3 - Y_2 + Y_1 + Y_8 - Y_7 - Y_6 + Y_5$	X_{12}
5	— — +	Y_5	$Y_2 - Y_1$	$Y_3 + Y_4 - Y_1 - Y_2$	$Y_5 + Y_6 + Y_7 + Y_8 - Y_1 - Y_2 - Y_3 - Y_4$	X_3
6	+ — +	Y_6	$Y_4 - Y_3$	$Y_7 + Y_8 - Y_5 - Y_6$	$Y_6 - Y_5 + Y_8 - Y_7 - Y_2 + Y_1 - Y_4 + Y_3$	X_{31}
7	— + +	Y_7	$Y_6 - Y_5$	$Y_4 - Y_3 - Y_2 + Y_1$	$Y_7 + Y_8 - Y_5 - Y_6 - Y_3 - Y_4 + Y_1 + Y_2$	X_{32}
8	+ + +	Y_8	$Y_8 - Y_7$	$Y_8 - Y_7 - Y_6 + Y_5$	$Y_8 - Y_7 - Y_6 + Y_5 - Y_4 + Y_3 + Y_2 - Y_1$	X_{321}

Полученные оценки коэффициентов модели являются независимыми друг от друга. Их численные значения и знаки указывают на силу и характер влияния факторов. Чем больше величина коэффициента, тем больше влияние оказывает фактор на параметр оптимизации. Если коэффициент имеет знак плюс, то с увеличением значения фактора параметр оптимизации увеличивается, а если минус, то уменьшается.

Проверка правильности численного расчета b_i коэффициентов проводится по формуле

$$\sum_{u=1}^N y_u^2 = N \sum_{i=0}^{N=1} b_i^2. \quad (80)$$

Эффект взаимодействия двух факторов называется эффектом взаимодействия первого порядка, взаимодействие трех факторов – взаимодействие второго порядка и т.д.

Из полного факторного эксперимента нельзя извлечь информацию о квадратичных членах модели, т.к. соответствующие оценки смешаны с b_0 . Действительно, в матрице, показанной в таблице 2, столбцы $X_1^2 X_2^2$ (если их дописать) совпадают друг с другом и со столбцом X_0 . Поэтому нельзя сказать, за счет чего получена величина b_0 . Она включает значение свободного члена и вклада квадратичных членов. Для получения квадратичных членов моделей используют композиционные планы, так называемые центральные композиционные планы.

Статистическая значимость b_i -коэффициентов определяется по t -критерию Стьюдента. Для этого необходимо знать дисперсию коэффициентов $S_{b_i}^2$. Дисперсия коэффициентов, если известна дисперсия воспроизводимости S_y^2 (значения отклика во всех точках $u = 1, 2, \dots, N$; $S_1^2 = S_2^2 = S_N^2$), может быть по теореме о дисперсии суммы независимых случайных величин определена как

$$S_{b_i}^2 = \frac{1}{N^2} \sum_{u=1}^N x_{iu}^2 S_{yu}^2 = \frac{1}{N^2} \frac{S_y^2}{m} \sum_{u=1}^N x_{iu}^2 = \frac{S_y^2}{Nm}. \quad (81)$$

Находя значимость b_i -коэффициентов, проверяем основную статистическую гипотезу $H_0: \beta_I = M[b_i] = 0$ против альтернативы $H_1: \beta_I = M[b_i] \neq 0$. Значение t -критерия вычисляется по формуле

$$t_{pi} = \frac{|b_i - \beta_i|}{S_{b_i}} = \frac{|b_i|}{S_y \sqrt{Nm}}. \quad (82)$$

Число степеней свободы t статистики t_p равно числу степеней свободы дисперсии воспроизводимости S_y^2 т.е. $f_p - f_y = N(m - 1)$. Критическое значение $t_{кр} = t_{табл}$ берется из таблиц распределения Стьюдента по уровню значимости α и числу степеней свободы f_y . Условие принятия гипотезы 0 :

$$b_i \leq b_{кр} = S_{b_i} t_{табл}. \quad (83)$$

Другими словами, коэффициенты b_i можно считать отличными от нуля (или, то же самое, значимыми), если

$$|b_i| \geq t \sqrt{S_{b_i}}. \quad (84)$$

Уровень значимости α характеризует степень достоверности выводов, сделанных на основании неравенства (84). Например, если принимаем достаточным 5% уровень значимости, то наше утверждение об отличии рассчитанного значения b_i , от нуля будет справедливо с вероятностью 95%.

Наконец, рассмотрим задачу проверки адекватности модели. Понятие «адекватность» является относительным и зависит от принятого критерия адекватности. В технике широкое применение имеет при оценке адекватности модели F -критерий Фишера. В основе этого критерия лежит сравнение двух видов рассеивания – рассеивания экспериментальных точек y_{ui} относительно построчных средних $\bar{y}_u$ и рассеивания построчных средних $\bar{y}_u$ относительно предсказанной по уравнению регрессии y_u . Первое характеризуется формулой S_y^2 или

$$S_y^2 = \frac{1}{N(m-1)} \sum_{u=1}^N \sum_{j=1}^m (y_{uj} - \bar{y}_u)^2. \quad (85)$$

Рассеивание построчных средних относительно предсказанной по регрессии

$$S_{ад}^2 = \frac{m}{N-1} \sum_{u=N-1}^N (\bar{y}_u - \hat{y}_u)^2, \quad (86)$$

где l – число коэффициентов модели после оценки их значимости, $\hat{y}$ – значение отклика, предсказанное по уравнению регрессии поверхности отклика (по модели), m – число параллельных опытов по строкам плана.

Расчетное значение критерия вычисляется по формуле:

$$F_{\text{расч.}} = \frac{S_{\text{ад}}^2}{S_y^2}. \quad (87)$$

Критическое или табличное значение берется из таблиц F -распределения для уровня значимости α и степеней свободы $S_y = N(m - 1)$ и $S_{\text{ад}} = N - l$.

Гипотеза адекватности принимается, если $F_{\text{расч.}} < F_{\text{табл.}}$. Модель, удовлетворяющая условию адекватности, может использоваться для дальнейшей работы. Если модель не удовлетворяет условию адекватности, то для уменьшения $S_{\text{ад}}^2$ уменьшают пределы варьирования или переходят к планированию второго порядка.

Адекватность на основе F -критерия устанавливается при выполнении двух обязательных условий:

1. Количество параллельных опытов $m > 1$, т.к. при $m = 0$ не может быть использована формула (81)
2. Статистически значимых коэффициентов модели должно быть меньше, чем количество опытных точек ($l < N$), т.к. при $l = N$ формула (86) не работает.

Проверка адекватности модели при $m > 1$ и $l = N$. В этом случае предсказанное значение выхода в центре локальной области $\hat{y}_{\text{цп}} = b_0$. Действительное значение выхода в центре плана определяют, поставив m_0 параллельных опытов в центре. Далее находят среднее $\bar{y}_{\text{цп}}$ и дисперсию отклика $S_{\text{цп}}^2$ (последняя имеет степень свободы $S_{\text{цп}} = m_0 - 1$), после чего по критерию Стьюдента определяется статистическая значимость отклонения

$$\varepsilon_{\text{ц.п.}} = \bar{y}_{\text{ц.п.}} - \hat{y}_{\text{ц.п.}} = \bar{y}_{\text{ц.п.}} - b_0. \quad (88)$$

Если отклонение $\Sigma_{\text{цп}}$ статистически незначимо, модель признается адекватной.

Проверка адекватности при $m = 1$ и $l < N$. В этом случае формула (85) не работает. Поэтому в центре плана ставят m_0 опытов, по которым вычисляют оценки $\bar{y}_{\text{цп}}$ и $S_{\text{цп}}^2$. Адекватность проверяется по F -критерию по формуле (86), где вместо оценки S_y^2 берется S_y^2 с числом степеней свободы $f_{\text{цп}} = m_0 - 1$.

В технике также широко применяется так называемый допусковый контроль, когда изделие или система считаются годными, если его определяющий (выходной) параметр находится в заранее установленных пределах. Такой контроль можно использовать при проверке адекватности модели.

При заданном допуске δ модель считается адекватной, если $|y_u - \hat{y}_u| \leq \delta$, т.е. для адекватности модели необходимо, чтобы абсолютная величина разности между фактическим и предсказанным значениями отклика во всех опытных точках $u = 1, 2, \dots, N$ и в центре плана была меньше δ . Допусковый контроль, как правило, неизбежен при цифровом моделировании, когда повторные опыты дают один и тот же результат.

Если величина δ заранее неизвестна, её можно приближенно определить по паспортным данным на элементы и характеристики объекта исследования.

Вопросы для самоконтроля

1. Что такое факторы эксперимента? Как формируется факторное пространство? Какие требования предъявляются к факторам?
2. Что такое функция отклика и как она формируется?
3. Что такое полный факторный эксперимент и как строится его план?
4. Что такое дробный факторный эксперимент и как строится его план?
5. Какие задачи решаются в рамках обработки результатов многофакторного эксперимента?
6. Как рассчитываются b_i коэффициенты функции отклика? Как проверить значимость этих коэффициентов в модели эксперимента?

8 Методы прогнозирования. Фильтр Калмана

8.1 Методы и системы прогнозирования

В данном разделе рассматривается решение задачи построения алгоритмов прогнозирования поведения и динамики изменения состояния различных технических систем в рамках определенного временного промежутка [5]. Разработка таких алгоритмов является достаточно сложной задачей и подразумевает как выбор подходящего под конкретную задачу метода прогнозирования из широкого числа разнообразных имеющихся методов, так и последующую валидацию полученных результатов с целью оценки адекватности разработанного алгоритма прогнозирования.

В общем случае к задачам оценки состояния системы на определенный момент в будущем можно отнести изучение не только технических, но также социальных, экономических и политических процессов. В упомянутых выше системах можно выделить общие характеристики, которые являются важнейшими при выработке методов прогнозирования. Выявление и анализ таких характеристик позволит выделить единые принципы построения системы прогнозирования для задач такого класса. В то же время для построения оптимального в некотором смысле алгоритма необходимо четко представлять специфику процессов в данной предметной области. В рамках данного раздела под прогнозированием понимается процесс получения численного прогноза значения случайного скалярного процесса, параметризованного временем, на некоторый заданный момент в будущем, который обеспечил бы возможность совершить упреждающие действия для предотвращения или смягчения возможных нарушений функционирования (наступления критических режимов) рассматриваемого процесса.

Прогнозы можно формально разделить на два различных типа: поисковые и нормативные. Под поисковым прогнозом понимают прогнозирование потенциально возможных состояний объекта прогнозирования на некоторый период времени вперед, основанное на некотором условном продлении в будущее тенденций развития рассматриваемого процесса в настоящем времени и в прошлом. Такой прогноз отвечает на вопрос: что вероятнее всего произойдет при условии сохранения существующих тенденций? Под нормативным прогнозом понимают процесс оценки сроков и способов достижения процессом некоторых принятых в качестве целевых состояний. Такой прогноз отвечает на вопрос: какими путями можно достичь желаемого?

Методы прогнозирования можно разбить на качественные и количественные. Качественные методы прогнозирования, главным образом, основаны на логическом анализе общих закономерностей поведения объекта прогнозирова-

ния и рассматривают, как правило, долгосрочные прогнозы. Подобные методы во многом основываются на анализе и обработке экспертной информации (в том числе с использованием автоматизированных экспертных систем). Количественные методы прогнозирования являются чисто формальными и основываются на аппарате математического анализа и статистики имеющихся данных о функционировании технического объекта (например, массивов входных и выходных данных) за некоторый период времени. Также подобные методы могут быть основаны на эконометрических или социометрических моделях соответствующих процессов.

8.2 Классификация методов прогнозирования

Все методы прогнозирования глобально можно разделить на два класса: эвристические и фактографические [5]. Эвристические методы прогнозирования в основном основываются на информации, получаемой при помощи логических правил и методических приемов, которые математически описывают процесс функционирования исследуемого процесса. С другой стороны, фактографические методы прогнозирования результаты прямых или косвенных измерений параметров течения прогнозируемого процесса и анализе массивов данных, полученных в результате этих измерений.

Эвристические методы прогнозирования можно также разделить на два класса, а именно интуитивные и аналитические (см. рисунок 17) [5]. Интуитивные методы при составлении прогноза поведения исследуемого процесса опираются на экспертные оценки, которые позволяют определить тенденции развития оцениваемого процесса. На таких методах, в частности, строятся автоматизированные экспертные системы. Подобные методы полагаются во многом на опыт и интуицию специалистов в соответствующих областях знаний. Аналитические же методы прогнозирования имеют в своей основе логический анализ модели (например, математической) протекания исследуемого процесса.

Интуитивные методы прогнозирования получили широкое распространение в практике прогнозирования, поскольку сбор и обобщение мнений экспертов, обладающих необходимыми знаниями об объекте прогноза, позволяет получить довольно надежную, а в ряде случаев и единственную возможную информацию. Такие методы применяют при прогнозировании поведения процессов, формализация описания поведения которых в виде, например, математической модели, представляет собой существенно трудоемкую задачу, либо осуществима лишь частично, или вовсе невозможна известными методами.

Интуитивные методы прогнозирования можно естественным образом разделить на две группы: индивидуальные экспертные оценки и коллективные экс-

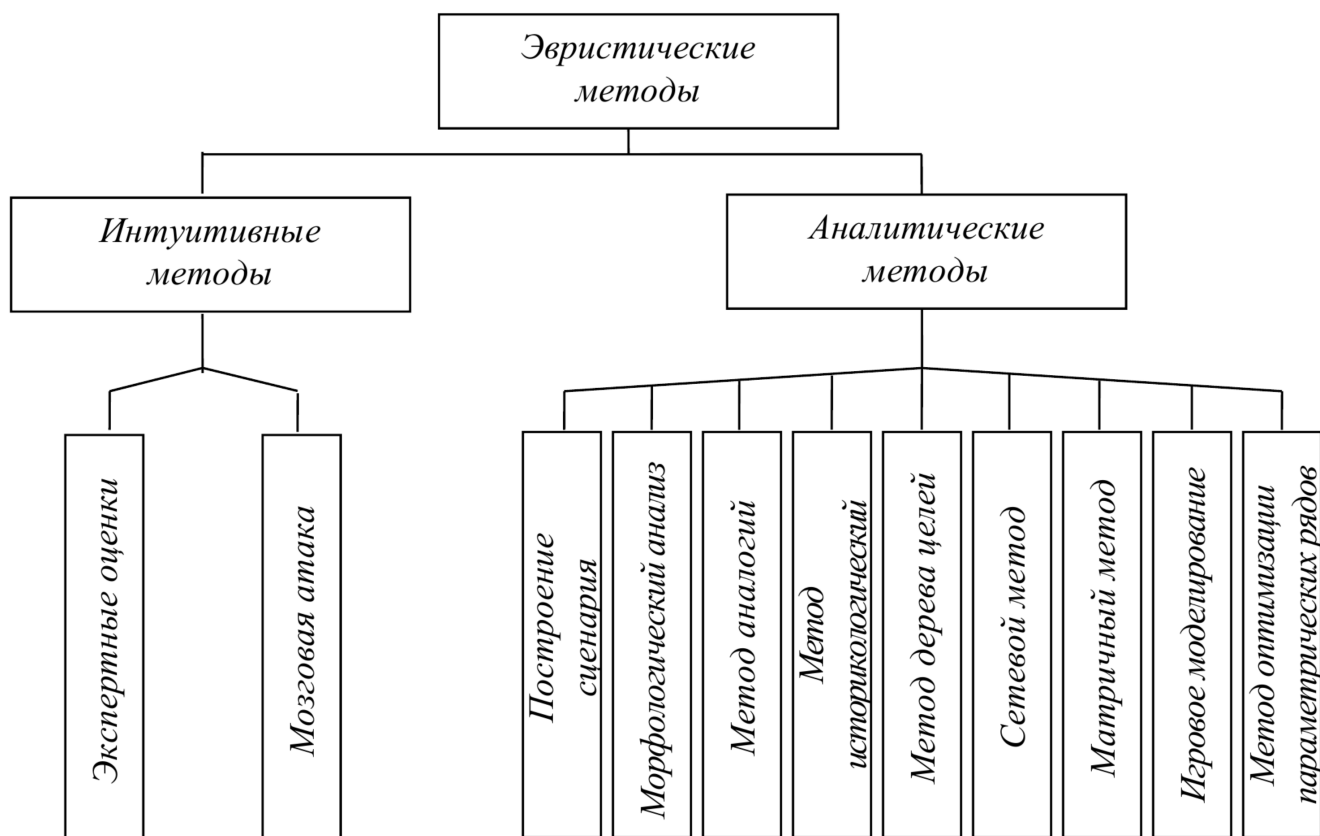


Рис. 17: Классификация эвристических методов прогнозирования.

пертные оценки. В силу того, что экспертные оценки могут быть и ошибочны, индивидуальные оценки не обязательно используют оценки и прогнозы только одного эксперта, но каждый из экспертов в этом случае вырабатывает свои оценки независимо. Коллективные же экспертные оценки вырабатываются при непосредственном взаимодействии коллектива соответствующим образом подобранных экспертов в ходе различного рода совещаний, "мозговых штурмов" и т. п. Как и в случае с индивидуальными экспертными оценками, часто собирают несколько таких групп экспертов, которые вырабатывают свои оценки и прогнозы независимо.

Основная задача при использовании экспертных оценок состоит в корректной процедуре сбора экспертной информации. Для этого существуют различные методики. Так, одним из наиболее распространенных методов сбора и обработки экспертной информации является метод "Дельфи" (название образовано от дельфийского оракула), в основе которого лежит идея о том, что некоторое количество независимых экспертов или небольших экспертных коллективов лучше оценивает и предсказывает течение исследуемого процесса, нежели один большой структурированный коллектив экспертов.

В случаях, когда имеется возможность с допустимой точностью осуще-

ствить формализацию описания оцениваемого процесса (например, в форме математической модели), находят свое применение аналитические методы прогнозирования. Подобные методы дают возможность установить количественные связи между различными параметрами исследуемого процесса. Аналитические методы прогнозирования в общем случае дают лучшие результаты, чем интуитивные, поскольку применяемые в последних экспертные оценки базируются на некоторых абстрактных представлениях о течении исследуемого процесса и могут не принимать во внимание второстепенные аспекты функционирования процесса, которые оказывают влияние на его течение, и результаты этого влияния могут быть неверно интерпретированы. В то же время аналитические методы прогнозирования предоставляют возможность учесть такие аспекты в математической модели процесса и получить в результате более точные и надежные прогнозы.

Одним из примеров аналитических методов прогнозирования может послужить морфологический подход. Данный метод основывается на анализе структурных взаимосвязей между объектами, явлениями и концепциями. Морфологический подход представляет собой упорядоченный способ рассмотрения предметов и получения систематизированной информации по всем возможным решениям изучаемой проблемы. Такой подход позволяет не только упорядочить имеющуюся информацию, но также предоставляет инструментарий для получения качественно новой информации.

Морфологический подход основан на приеме систематизированного охвата, который состоит в том, что исследование начинают с данного уровня знаний и систематически просматривают одну за другой все возможные области знаний, пока не будет достигнут новый уровень знаний. Такой прием позволяет накапливать данные для последующих исследований по методу “морфологического ящика”, который включает следующие этапы:

1. Точная формальная постановка решаемой задачи.
2. Анализ всех параметров, важных с точки зрения решения поставленной задачи.
3. Построение “морфологического ящика”, потенциально содержащего все решения поставленной задачи. Данная процедура реализуется в форме построения дерева или матрицы, в узлах которых помещены соответствующие параметры. Последовательное соединение одного какого-либо параметра первого уровня с одним из параметров последующих уровней представляет собой одно из возможных решений поставленной задачи. Общее количество возможных решений равно произведению числа всех параметров, учитываемых в “ящике” построчно.

4. Анализ “морфологического ящика”, состоящий в изучении полученных решений с точки зрения их сообразности решению поставленной задачи. Для этого с учетом сформулированных критериев ценность полученных решений представляется графически в форме так называемых топологических характеристических карт. При этом способ определения ценности каждого решения из “ящика” может варьироваться от задачи к задаче.
5. Выбор наилучших решений и их применение в рамках поставленной задачи.

Сильной стороной морфологического метода является то, что он может быть применен в условиях ограниченной информированности по изучаемой задаче, и при этом для оценки эффективности решений могут быть применены самые общие критерии.

Далее необходимо написать сценарий прогноза. Написание сценария – это метод, с помощью которого устанавливается логическая последовательность событий с целью показать, как, исходя из существующей ситуации, может шаг за шагом разворачиваться будущее состояние объекта исследования. Описание обычно совершается в явно выраженных временных координатах. Задача сценария состоит в определении основной цели развития объекта прогнозирования, выявлении основных факторов среды и формулировании критериев для оценки верхних уровней “дерева целей”. При разработке сценария, поскольку в ней принимает участие группа специалистов, всегда возникает неопределенность, связанная с субъективностью суждений.

“Деревом целей” называется граф, выражающий отношения между вершинами – этапами или проблемами достижения некоторой цели. “Дерево целей”, вершины которого ранжированы, т. е. выражены количественными оценками их важности, широко используется для количественной оценки приоритета различных направлений развития. Построение “дерева целей” требует решения многих прогнозных задач, в частности: прогноза развития объекта в целом; формулировки сценария прогнозируемой цели, уровней и вершин “дерева целей”, критериев и их весов в ранжировании вершин. Каждая из этих прогнозных задач по необходимости решается методом экспертных оценок.

Специальный матричный метод позволяет произвести сравнение различных прогнозных трендов по степени важности для решения поставленной задачи. Поскольку развитие объекта прогноза обычно зависит от значительного числа взаимосвязанных факторов, то применение матричного метода требует все множество различных факторов разбить на ряд кластеров, в каждый из которых входят в определенной мере однородные факторы. Далее осуществляется оценка влияния сформированных кластеров друг на друга и на достижение конечных целей на основе использования операций с матрицами для решения

задачи выбора и обоснования оптимального размещения ресурсов. Это достигается путем ранжирования факторов и определения их относительных весов внутри кластера.

Взаимное влияние двух кластеров может быть выражено в форме матрицы влияния $A[m, n]$, элементами которой a_{ij} являются оценки, отражающие влияние i -го фактора кластера M на j -й фактор кластера N .

Матричный метод относится к классу нормативных методов прогнозирования, в котором задаются конечные цели и в процессе прогнозирования определяются пути и средства их достижения. Прогностическая функция матричного метода заключается в оценке влияния различных вариантов происходящих сдвигов на достижение конечных целей объекта прогноза.

Алгоритм применения матричного метода прогнозирования можно сформулировать следующим образом:

1. идентификация факторов, влияющих на достижение поставленных целей;
2. выделение однородных кластеров из факторов путем группировки этих факторов по характеру их влияния;
3. формирование матриц влияния кластеров друг на друга и на достижение целей;
4. определение влияния факторов на достижение комплекса целей путем выполнения операций над матрицами влияний (умножение, сложение, вычитание) в соответствии со схемой направления влияний одних факторов на другие (графом влияний), определение относительных весов факторов и ранжирование их.

Исходной информацией для прогнозирования по матричному методу с использованием экспертных оценок являются: перечень целей объекта прогнозирования и коэффициенты их относительной важности; перечень факторов, влияющих на достижение целей объекта прогноза, сгруппированных в однородные кластеры; коэффициенты матриц, определяющих влияние одного кластера на другой или на достижение целей; показатели относительной самооценки компетентности экспертов, принимающих участие в работе по прогнозированию развития объекта; данные о группах, участвовавших в работе экспертов, необходимые для определения компетентности экспертов.

Еще одним распространенным методом прогнозирования является метод когнитивного моделирования. Данный метод применяется для построения как нормативных, так и поисковых прогнозов.

В основе метода когнитивного моделирования лежит построение когнитивных карт. Когнитивные карты представляют из себя умственные представления среды, которые подвластны внутреннему взору человека. Под когнитивными элементами (концептами) понимаются произвольные социальные объекты, определенные в виде упорядоченной пары, со своей внутренней и внешней структурой. Внутренняя структура объекта определяется как множество интересов и множество отношений между ними; внешняя структура представляется в виде множества элементов поведения объекта, объединенное с множеством отношений между элементами этого поведения. Внутренняя структура подсистемы предполагается неизвестной: иначе говоря, она оценивается лишь с точки зрения наблюдаемого поведения, т. е. внешней структуры данного объекта.

Традиционное понятие когнитивной карты возникло и развивалось в виде знаковых орграфов. Нечеткая когнитивная карта – это причинно-следственная сеть, отражающая какую-либо область знания посредством дуг и узлов положительноотрицательной сети (нечеткой нейронной сети). В такой сети дугам соответствуют отношения (связи) из интервала $[-1, 1]$, а узлами соответствуют атрибуты (концепты). Нечеткие семантические сети используются для анализа взаимовлияния концептов друг на друга. Процедура построения когнитивной карты называется когнитивным отображением. Данный процесс может быть основан на использовании прямых экспертных процедур опроса.

В ходе составления когнитивной карты очень важно правильно определить, какие из рассматриваемых концептов являются причинами, какие следствиями и как они связаны между собой. Кроме того, важно установить, какое именно воздействие на концепт B окажет усиление концепта A , например, усиливающее или ослабляющее. Термины "усиление" и "ослабление" здесь носят довольно абстрактный характер и нуждаются в дополнительной интерпретации, чтобы установить четкую причинную связь.

Управляемые когнитивные процессы в основном ориентированы на качественный анализ систем (предметных областей или когнитивных моделей). Нечеткая логика, лежащая в основе представления когнитивных карт, позволяет проводить "нечеткий" анализ причин и следствий в предметной области, а также она используется для оценки консонанса. Если иметь в виду каузальное преобразование, то когнитивная матрица может использоваться для прогноза.

Альтернативой эвристическим методам являются фактографические методы прогнозирования. Класс фактографических методов прогнозирования подразделяется на статистические и опережающие. Статистические основываются на построении и анализе эмпирических временных рядов характеристик объекта, а опережающие – на обработке информации, относящейся непосредственно ко времени упреждения.

Одним из распространенных статистических методов является метод экс-

траполяции. Этот метод является одним из основных в прогнозировании развития больших систем. На основе анализа статистических данных, характеризующих объект прогнозирования за предшествующий период, устанавливают закономерности изменения и соответственно тенденции развития этих данных. Затем определяют значения прогнозируемых величин за пределами имеющихся эмпирических временных, или динамических, рядов. Применение метода экстраполяции возможно тогда, когда более или менее точно установлена зависимость во временных рядах, а также определена область, на которую распространяется экстраполяция.

Данный метод прогнозирования дает хорошие результаты, если правильно определена форма кривой, отражающей установленную закономерность изменения эмпирических данных.

Временной ряд $y(t)$, может быть представлен в следующем виде:

$$y(t) = x(t) + \varepsilon(t), \quad (89)$$

где $x(t)$ – детерминированная (неслучайная) компонента процесса; $\varepsilon(t)$ – стохастическая (случайная) компонента процесса.

Задача прогноза состоит в определении вида экстраполирующих функций $x(t)$ и $\varepsilon(t)$ на основе исходных эмпирических данных.

Наиболее распространенными методами оценки параметров зависимостей являются метод наименьших квадратов и его модификации, метод экспоненциального сглаживания, метод вероятностного моделирования и метод адаптивного сглаживания.

Сущность метода наименьших квадратов состоит в отыскании параметров модели тренда, обеспечивающих минимум отклонения ее выходной величины от точек исходного временного ряда, т.е.

$$S = \sum_{i=1}^n (\tilde{y}_i - y_i)^2 \rightarrow \min, \quad (90)$$

где $\tilde{y}_i$ – модельные значения временного ряда; y_i – фактические значения исходного временного ряда; n – число значений.

Классический метод наименьших квадратов предполагает равноценность исходной информации в модели. В реальной же практике будущее поведение процесса в большей степени определяется поздними наблюдениями, чем ранними. Это обстоятельство породило так называемое дисконтирование, т.е. уменьшение ценности более ранней информации.

Дисконтирование можно ввести путем добавления в (90) некоторых весов

$\beta_i < 1$. Тогда

$$S = \sum_{i=1}^n \beta_i (\tilde{y}_i - y_i)^2 \rightarrow \min. \quad (91)$$

Коэффициенты β_i могут задаваться заранее в числовом виде или в виде функциональной зависимости таким образом, чтобы по мере продвижения в прошлое веса убывали, например $\beta_i = a^i$, где $a < 1$. К сожалению, формальных процедур выбора параметра не разработано, и он выбирается исследователем произвольно.

Метод наименьших квадратов широко применяется для получения конкретных прогнозов, что объясняется его простотой и легкостью реализации на ЭВМ. Недостаток метода состоит в том, что модель тренда жестко фиксируется, и с помощью метода можно получить надежный прогноз на небольшой период упреждения. Поэтому метод наименьших квадратов относится главным образом к методам краткосрочного прогнозирования. Кроме того, существенной трудностью метода наименьших квадратов является правильный выбор вида модели, а также обоснование и выбор весовых коэффициентов во взвешенном методе наименьших квадратов.

Другим весьма эффективным и надежным методом прогнозирования является метод экспоненциального сглаживания. Основные достоинства метода состоят в возможности учета весов исходной информации, в простоте вычислительных операций, в гибкости описания различных динамик процессов. Метод экспоненциального сглаживания дает возможность получить оценку параметров тренда, характеризующих не средний уровень процесса, а тенденцию, сложившуюся к моменту последнего наблюдения. Наибольшее применение метод нашел для реализации среднесрочных прогнозов. Для метода экспоненциального сглаживания основным и наиболее трудным моментом является выбор параметра сглаживания α , начальных условий и степени прогнозирующего полинома.

Пусть исходный динамический ряд описывается уравнением

$$y(t) = a_0 + a_1 t + \frac{a_2}{2} t^2 + \dots + \frac{a_p}{p!} t^p + \varepsilon(t). \quad (92)$$

Метод экспоненциального сглаживания, являющийся обобщением метода скользящего среднего, позволяет построить такое описание процесса, при котором более поздним наблюдениям придаются большие веса по сравнению с ранними наблюдениями, причем веса наблюдений убывают по экспоненте. Выражение

$$S_i^{[k]}(y) = \alpha \sum_{i=0}^n (1 - \alpha)^i S_{i-1}^{[k+1]}(y) \quad (93)$$

называется экспоненциальной средней k -го порядка для ряда (t) , где α – параметр сглаживания.

Важную роль в методе экспоненциального сглаживания играет выбор оптимального параметра сглаживания α , так как именно он определяет оценки коэффициентов модели, а следовательно, и результатов прогноза.

В зависимости от величины параметра прогнозные оценки по-разному учитывают влияние исходного ряда наблюдений: чем больше α , тем больше вклад последних наблюдений в формирование тренда, а влияние начальных условий быстро убывает. При малом α прогнозные оценки учитывают все наблюдения, при этом уменьшение влияния более "старой" информации происходит медленно.

Известно несколько основных соотношений, позволяющих найти приближенную оценку α [1]. Одно из них – соотношение Р.Брауна, выведенное из условия равенства скользящей и экспоненциальной средних

$$\alpha = \frac{2}{N + 1}, \quad (94)$$

где N – число точек ряда, для которых динамика считается однородной и устойчивой (период сглаживания).

Выбор параметра α целесообразно связывать с точностью прогноза, поэтому для более обоснованного выбора α можно использовать процедуру обобщенного сглаживания, которая позволяет получить соотношения, связывающие дисперсию прогноза и параметр сглаживания. Существуют подходы к оценке точности прогноза для случая, когда автокорреляция исходного ряда не равна нулю. Эти подходы дают возможность эффективно решать задачу определения оптимального значения параметра сглаживания α , что существенно повышает точность метода.

Весьма важным для практического использования является вопрос о выборе порядка прогнозирующего полинома, что во многом определяет качество прогноза. В некоторых работах показано, что использование полинома выше второго порядка не приводит к существенному увеличению точности прогноза, но значительно усложняет процедуру расчета.

Метод экспоненциального сглаживания является одним из наиболее надежных и применяемых в практике прогнозирования.

В методе вероятностного моделирования прогнозирование с использованием вероятностных моделей базируется на методе экспоненциального сглаживания. Вероятностные модели по своей сути отличны от экстраполяционных моделей временных рядов, в которых основой является описание изменения во времени процесса. Во временных рядах модели представляют собой некоторую функцию времени с коэффициентами, значения которых оцениваются по

наблюдениям. В вероятностных моделях оцениваются вероятности, а не коэффициенты. При достаточном количестве исходной информации вероятностная модель может дать вполне надежный прогноз. Недостатками этой модели является требование большого количества наблюдений и знание начального распределения. Тем не менее, при определении начального распределения можно использовать вероятностную модель как эффективный метод прогноза.

Более простым является метод прогноза на основе вычисления автокорреляционных функций. Разлагая временной ряд $y(t) = x(t) + \varepsilon(t)$ на две составляющие $x(t)$ и $\varepsilon(t)$, можно применить математико-статистический метод прогнозирования, основанный на использовании автокорреляционных функций. Автокорреляция, как известно, выражает взаимную связь между соседними членами временного ряда.

Исходный пункт использования данного метода – неизменность статистических характеристик рассматриваемого временного ряда. Как правило, временные ряды, например, экономических процессов содержат в себе нестационарный тренд. Однако, используя некоторые преобразования, можно свести задачу к анализу стационарного ряда. Так, уместно воспользоваться весьма распространенным и довольно простым способом: образованием разностей соответствующего порядка (например, для линейно изменяющейся функции разность первого порядка является константой).

Степень изменчивости случайной функции $y(t)$ во времени, т.е. ее автокорреляционная функция, характеризует связь между ее значениями при различных значениях аргумента и выражается так:

$$K_y(t, t') = M \left[(y(t) - m_y(t)) (y(t') - m_y(t')) \right], \quad (95)$$

где $m_y(t)$ и $m_y(t')$ – математические ожидания.

Прогнозирование с использованием корреляционных методов состоит в нахождении математических формул, отражающих статистическую связь одного показателя с другим (парная корреляция) либо с группой других (множественная корреляция). Корреляционные методы дают возможность на основе анализа поведения исследуемых параметров в прошлом, представленных в форме временных рядов, строить формулы и затем рассчитывать по ним наиболее вероятные значения одних показателей (функций) в зависимости от других показателей, принятых за факторы (аргументы).

Применению корреляционного метода к исследованию временных рядов непременно предшествует логический анализ статистических данных. Исследуемых данных должно быть много, чтобы надежность корреляционных формул, используемых при расчете, была достаточна, и они должны быть качественно однородны.

В решение корреляционных и регрессионных уравнений входит нахождение численных значений параметров исходных зависимостей, которые могут быть найдены способом наименьших квадратов.

Часто в процессах прогнозирования заранее неизвестен характер функциональной взаимосвязи, и он оказывается более сложным, чем свойственный степенному или тригонометрическому многочлену, или некоторой другой определенной функции. Поэтому при прогнозировании целесообразно использовать функции с гибкой структурой, т.е. функции, структура которых может изменяться, автоматически приспособливаясь к изучаемому процессу.

Эта идея предложена и впервые практически реализована проф. Н.К. Куликовым при помощи установленной им обобщенной формулы для представления функций.

Установлены функции с гибкой структурой для нескольких аргументов. На основе этих функций можно разработать метод прогнозирования многофакторных процессов.

Факторный анализ в современном виде представляет собой определенный раздел математической статистики. Появление его в начале XX в. обычно связывают с именами психологов Ч. Спирмена, С. Барту, Л. Терстоуна и др. Первоначальная его цель состояла в построении математических моделей способностей и поведения человека.

Имеется большое разнообразие методов факторного анализа и его модификаций, которые известны в настоящее время. Рассмотрим кратко сущность одного из них.

Пусть X – n -мерный случайный вектор, представляющий случайную выборку измерений совокупности взаимосвязанных параметров $x(i)$; F – k -мерный вектор, компонентами которого являются непосредственно не наблюдаемые переменные (факторы); $\bar{X}$ – математическое ожидание вектора X ; U – вектор сумм ненаблюдаемых ошибок и специфических факторов. Согласно основному предположению факторного анализа, каждое конкретное измерение $x(i)$ вектора X может рассматриваться как сумма воздействий некоторого небольшого числа групповых факторов $f(j)$ (взятых с определенными весами $a(i, j)$), специфического фактора $s(i)$, воздействующего только на данную переменную, и ошибки измерения $e(i)$. Поскольку $s(i)$ и $e(i)$ в факторном анализе неразличимы, их обычно рассматривают как сумму $u(i) = s(i) + e(i)$; A – матрица размера $n \times k$ ($n > k$), элементы которой – факторные веса $a(i, j)$, определяющие нагрузку i -й переменной на j -й фактор; N – число наблюдений над вектором X , по которым производится оценка. Запишем основное положение факторного анализа в матричной форме:

$$X = AF + \bar{X} + U. \quad (96)$$

Приравняем для простоты все средние нулю: $\bar{X} = 0$, т.е. будем рассматривать далее несмещенные распределения $x(i)$.

Обозначим произведение $AF = Q$, тогда

$$X = Q + U, \quad (97)$$

где Q принято называть общей частью, а U – специфической частью.

Предполагается, что U не зависит от Q , и все $u(i)$ не коррелируют между собой. При этом матрица $E(QU) = 0$, а матрица $E(UU^T)$ – диагональная (E – оператор взятия математического ожидания). Тогда

$$\begin{aligned} E(XX^T) &= E[(Q + U)(Q^T + U^T)] = \\ &= E(QQ^T) + E(UU^T) + E(QU^T) + E(UQ^T) = E(QQ^T) + E(UU^T). \end{aligned} \quad (98)$$

Если нормировать компоненты вектора X по величинам стандартных отклонений $\sigma(i)$ ($z(i) = \frac{x(i)_p}{\sigma(i)}$, где $x(i)_p$ – компонента вектора X , p – порядковый номер единичного наблюдения над вектором X), то $E(ZZ^T) = R$ есть корреляционная матрица.

В соответствии с формулой (98) ее можно представить в виде

$$R = R_0 + U^{(2)} = R_0 + I + H^2, \quad (99)$$

где R – исходная матрица с единицами на главной диагонали; R_0 – так называемая редуцированная матрица; U – диагональная матрица из квадратов сумм специфических факторных весов и ошибок; H^2 – диагональная матрица так называемых коммунальностей; I – единичная матрица.

Разложению в факторном анализе подвергается обычно матрица R_0 , определяемая выражением (99). При этом R_0 аппроксимируется произведением $R_0 = A_0A_0^T$, где A_0 принимается за матрицу факторных весов размера $n \times k$. Факторы $f(1), f(2), \dots, f(k)$ предполагаются некоррелированными. В этом случае факторные веса можно рассматривать как коэффициенты в линейном уравнении регрессии для оценки переменных по факторам.

После определения факторных нагрузок, соответствующих совокупности наблюдаемых переменных (факторов), обычно следует попытка их интерпретации, т.е. полезного и общедоступного толкования сущности различных сторон сложного явления, отражаемых выделенными факторами. В связи с тем, что процедура получения нагрузок в факторном анализе не приводит к однозначному результату (при числе факторов, большем единицы), можно получать эквивалентные множества нагрузок путем ортогонального преобразования. Геометрически это будет соответствовать дополнительному вращению факторов в пространстве измерений.

Основным преимуществом прогнозирования развития факторов является то, что в процессе прогнозирования автоматически решается задача синтеза и увязки прогнозов отдельных показателей.

Наряду с процедурой прогнозирования, использующей непосредственную экстраполяцию главных факторов с последующим их разложением на совокупность прогнозных значений исходных переменных, факторный анализ может использоваться в сочетании с моделями регрессионного анализа. В этом случае строят регрессию некоторого выходного показателя исследуемого комплекса, но не на совокупности входных показателей, как обычно, а на главных компонентах, полученных в процессе факторного анализа.

Преимущества регрессионных моделей, построенных на главных компонентах факторной структуры, сводятся к следующему: во-первых, размерность модели существенно меньше, чем регрессии на исходных переменных; во-вторых, отбор наиболее существенных компонент проводится по четкому количественному критерию; в-третьих, факторная модель исключает мультиколлинеарность за счет ортогональности главных факторов.

Другими весьма перспективными в настоящее время методами являются методы, основанные на использовании для прогнозирования методов теории распознавания образов. С использованием этой теории решается комплекс задач, имеющих важное значение в прогнозировании.

Обычно при прогнозировании временных рядов предполагает однородность динамики. Действительно, исходный временной ряд, являющийся основой прогнозирования какого-либо процесса, может содержать в себе интервалы, внутри которых динамика характеризуется определенными отличными от других интервалов условиями. Естественно, эти интервалы на перспективу искажают полученный прогнозный результат. В этой связи возникает необходимость четкого выделения тех интервалов, для которых характерна однородная динамика. Решение этого вопроса эффективно реализуется с помощью методов теории распознавания образов. Другим важным приложением теории распознавания образов для получения прогнозов является описание и прогнозирование поведения какого-либо объекта по набору признаков, определяющих поведение этого объекта.

При решении целого ряда задач исследователь сталкивается по существу с одной проблемой – разделением всего изучаемого множества объектов по обнаруженному сходству и различию на отдельные группы или множества. При этом часто невозможно провести одновременный анализ всего комплекса признаков, и приходится рассматривать его "по частям".

Существует несколько типов задач распознавания образов. Основные из них – задача обучения распознаванию образов, задача сокращения (минимизации) описания, задача таксономии.

Для первой задачи характерно по некоторому набору признаков с помощью выбранного решающего правила определить, к какому классу относится рассматриваемый объект. Первоначально существует некоторое количество объектов, образующих так называемые обучающие выборки, для которых указываются классы, содержащие эти объекты. По мере рассмотрения признаков для каждого объекта вырабатывается некоторый критерий, называемый решающим правилом, который и позволяет определить принадлежность каждого нового объекта тому или иному классу с ошибкой, не превышающей заранее заданную.

Задача минимизации описания объекта совокупностью признаков, характеризующих каждый рассматриваемый объект, позволяет выбрать те признаки, которые являются наиболее информативными с точки зрения распознавания. Иначе эта задача формулируется следующим образом: необходимо построить такое преобразование пространства признаков в некоторое другое, чтобы размерность нового пространства была меньше исходной, а функция потерь при его использовании существенно не увеличилась.

Задача таксономии (самообучения) заключается в том, чтобы из некоторого множества объектов выделить с помощью заданного правила классы однородных ("одинаковых") объектов.

Широкое распространение для получения прогнозных результатов в практических исследованиях используются методы, основанные на применении спектрального анализа. Данные методы позволяют достаточно точно описывать процессы, динамика которых содержит колебательные или гармонические составляющие. Так, очень часто в этих процессах рассматриваются четыре компоненты временного ряда: вековой уровень, описываемый гладкими апериодическими функциями; сезонные колебания с двенадцатимесячным периодом; колебания с периодом большим, чем сезонные, но меньшим, чем у соответствующих колебаний векового уровня; случайные колебания с широкими периодами, но с небольшой интенсивностью. Таким образом, исследуемый процесс можно представить в виде

$$x(t) = x_1(t) + x_2(t) + x_3(t) + \sigma(t). \quad (100)$$

Для описания и прогноза первой составляющей $x_1(t)$ эффективно используются такие методы, как скользящее среднее, экспоненциальное сглаживание, гармонические веса. Использование одного из этих методов позволяет выделить негармоническую составляющую процесса. Для оставшихся гармонических составляющих процесса используется представление в виде тригонометрического полинома.

Выполняется анализ периодограммы $I(w)$, имеющей для нечетного числа

наблюдений $N = 2m + 1$ вид

$$I_{2m+1}(w) = \frac{2m+1}{2\pi} \left(\alpha^2(w) + (\alpha')^2(w) \right), \quad (101)$$

где

$$a(w) = \sum_{I=1}^M (b_i u_i(w) + c_i v_i(w)) + \gamma(w), \quad (102)$$

$$u_i(w) = \frac{1}{2m+1} \sum_{t=1}^{2m+1} \cos(w_i t) \cos(wt), \quad (103)$$

$$v_i(t) = \frac{1}{2m+1} \sum_{t=1}^{2m+1} \sin(w_i t) \cos(wt), \quad (104)$$

$$\gamma(w) = \frac{1}{2m+1} \sum_{t=1}^{2m+1} \sigma_i \cos(wt). \quad (105)$$

Аналогично

$$a'(w) = \sum_{i=1}^m (b_i u'_i(w) + c_i v'_i(w)) + \gamma'(w), \quad (106)$$

где $u'_i(w)$, $v'_i(w)$ и $\gamma'(w)$ получаются из $u_i(w)$, $v_i(w)$ и $\gamma(w)$ заменой $\cos(wt)$ на $\sin(wt)$. Полученные величины $a(w)$ и $a'(w)$ являются аналогами ковариации рассматриваемого временного ряда $x(t)$ с косинусами и синусами исследуемых частот

$$w_i = \frac{2\pi i}{2m+1}, \quad i = \overline{1, m}. \quad (107)$$

Данный метод основан на сходимости по вероятности величин $a(w)$ и $a'(w)$, поэтому его следует применять при работе с длинными временными рядами, для которых может быть использован тест проверки значимых решений.

Вычислительная процедура получения прогнозов зависит от того, исследуется ли просто динамика ряда или структура временного ряда. В первом случае реализуется метод оценки взвешенных средних (метод согласованных фильтров). Схема расчетов при исследовании динамики такова. На первом шаге по критерию Стьюдента проводится оценка параметров согласованных фильтров. На втором шаге определяется статистическое качество полученных результатов, т.е. по критерию Дарбина-Уотсона проводится проверка случайной составляющей на автокорреляцию и вычисляются значения коэффициентов вариации оценок параметров. На третьем шаге согласований фильтр используется для прогноза.

В случае, когда целью исследования является определение структуры временного ряда, применяется метод оценки детерминированной составляющей в виде явной функции времени. На первом шаге расчетов решается характеристическое уравнение и восстанавливаются искомые функции времени. На втором шаге производится оценка линейных параметров найденных функций времени и оценивается покомпонентное разбиение суммарной детерминированной составляющей. На третьем шаге – оцененная (в виде явной функции времени) суммарная порождающая функция применяется для экстраполяции исследуемого показателя.

Специфика определенных задач прогноза требует привлечения различного математического аппарата. На практике часто возникает следующая задача: известны начальное и конечное состояния рассматриваемого объекта и затраты на изменение данного объекта в каждой точке пространства состояний. Необходимо определить оптимальную в смысле заданного критерия траекторию изменения объекта для достижения некоторого конечного состояния. Решение подобной задачи возможно при использовании элементов вариационного исчисления.

Пусть необходимо определить такую кривую, выраженную зависимостью $y = y(x)$, которая, соединяя две заданные точки А и Б, имела бы минимальную длину. Решение данной задачи выполняется минимизацией функционала

$$\int_a^b \sqrt{1 + (y'_x)^2} dx. \quad (108)$$

В общем случае задача формулируется следующим образом: необходимо найти такую функциональную зависимость $y = y(x)$, которая обеспечивала бы экстремум функционалу

$$J(y) = \int_a^b F(x, y(x), y'_x(x)) dx \quad (109)$$

при заданных условиях $y(a) = A$, $y(b) = B$, $y'_x(x)$ является производной от функции $y = y(x)$ по переменной x .

В качестве переменной x может выступать любой показатель, определяющий функционирование рассматриваемого объекта, в том числе и время.

Решение задач данного типа реализуется с помощью метода вариационного исчисления на основании теоремы Эйлера, которая гласит, что минимум функционала на множестве функций $y = y(x)$, имеющих первую производную

и подчиняющихся заданным условиям, достигается для функции, удовлетворяющей уравнению

$$F_y - \frac{d}{dx} F_{y'} = 0. \quad (110)$$

Это уравнение называется уравнением Эйлера, где

$$F_y = \frac{\partial F(x, y, y')}{\partial y}; \quad F_{y'} = \frac{\partial F(x, y, y')}{\partial y'}. \quad (111)$$

Рассмотрим еще одну задачу, которая в вариационном исчислении носит название задачи с подвижными концами. Непосредственный ее смысл заключается в том, что определяется оптимальная траектория не между точками, а между двумя фиксированными кривыми, т.е. рассматривается переход от одного режима функционирования к другому.

Таким образом, для функционала

$$J(y) = \int_{x_0}^{x_1} F(x, y, y') dx \quad (112)$$

находится такая функция $y = y(x)$, которая обеспечивает экстремум этому функционалу, и концы этой функции лежат на фиксированных кривых $y = \phi(x)$ и $y = \psi(x)$. В этом случае искомая функциональная зависимость $y = y(x)$ находится из уравнения Эйлера с граничными условиями

$$F + F_{y'} (\psi' - y') \Big|_{x=x_1} = 0, \quad (113)$$

$$F + F_{y'} (\phi' - y') \Big|_{x=x_0} = 0. \quad (114)$$

Эти граничные условия называются условиями трансверсальности.

Рассмотренные выше задачи позволяют определить наиболее эффективные (оптимальные) изменения объектов в пространстве выбранных показателей. Очевидно, данное изменение является наиболее вероятным, что дает основание использовать рассмотренный метод для решения целого класса задач прогнозирования.

Для целого ряда важных практических приложений используется метод прогнозирования на основе теории цепей Маркова.

Метод цепей Маркова может быть использован для прогноза множества показателей, которые меняются из года в год одновременно, но когда функциональные связи между ними не установлены ввиду отсутствия информации или крайней сложности этих связей.

В основе прогноза, построенного на основе простых цепей Маркова, лежит вычисление матрицы перехода, элементами которой являются вероятности перехода прогнозируемых параметров из одного состояния в другое, от одного значения к другому.

Реализация прогнозов с помощью цепей Маркова позволяет по мере поступления новой информации регулярно корректировать ошибки, что существенно повышает надежность получаемых результатов.

В ряде случаев эффективным является получение параметров прогноза в терминах теории игр. При этом под прогнозированием следует понимать получение статистического вывода о поведении в будущем некоторой случайной величины Y . Предсказывающая функция дает возможность по прошлым значениям случайной величины Y описать ее поведение на перспективу. Прогнозом являются сами значения предсказывающей функции.

Пусть реализуется прогноз величины Y , возможные значения которой принадлежат множеству A , т.е. $y \in A$. Случайный вектор X является k -мерной случайной переменной наблюдений, по которым возможна реализация прогноза. Значения случайного вектора X образуют k -мерное пространство R^k , т.е. $x \in R^k$. Пусть также существуют функции условного распределения $G(y/x)$ для данного вектора X и $F(x/y)$ для данного значения y .

Выбирая определенный вид предсказывающей функции $d(x)$, можем получить значение прогноза, например $a = d(x)$. Очевидно, что прогноз $a \in A$ и множество A состоит из возможных решений a и состояний рассматриваемой величины y .

Функция $d(x)$ преобразует множество R^k в множество A и является функцией решения. Введем также функцию потерь, которые возникают в результате неточного прогноза $L(a, y)$, определяемую на произведении $A \times A$ и показывающую, какие потери возникают, когда действительным значением рассматриваемой переменной Y будет y , а полученный на основании функции решения $d(x)$ прогноз равен a , причем, возможно, $a \neq y$. Поэтому оптимальная функция решения определяется исходя из минимизации математического ожидания функции потерь $L(a, y)$. В такой постановке задачи наилучшим прогнозом является такой, который минимизирует средние потери в условном распределении $G(y/x)$, т.е.

$$R^*(a_0) = \inf R^*(a), \quad a \in A; \quad R^*(a) = \int_A L(y, a) dG(y/x), \quad (115)$$

где a_0 – оптимальный (наилучший) прогноз.

Можно определить, следовательно, функцию риска $R(y, d)$ для предска-

вающей функции $d(x)$

$$R(y, d) = E_y L(y, a) = \int_{R^*} L(y, a) dF(x/y). \quad (116)$$

Таким образом, сформулирована задача статистической игры (A, D, R) , моделирующей задачу прогнозирования, где A – совокупность стратегий поведения рассматриваемой случайной величины Y , т.е. первого игрока; D – совокупность стратегий поведения второго игрока, т.е. прогнозиста, принимающего определенное прогнозное решение, и R – платежная функция, принимающая конечные числовые значения и определенная для каждой пары $L(y, a)$.

Выбор оптимальной предсказывающей функции $d(x)$ осуществляется в соответствии с общей теорией статистических игр с использованием байесовской или минимаксной функции решения.

Согласно такому подходу, наилучшей функцией решения будет та, при которой риск минимален, т.е.

$$r(G, d_0) = \min r(G, d), \quad d(x) \in D, \quad (117)$$

где r – байесовский риск; G – распределение переменной Y .

Задавая различные виды функции потерь, можно получить различные прогнозные результаты. Наиболее известными являются квадратичная функция потерь, кусочно-линейная несимметричная, симметричная кусочно-линейная, ограниченная функция потерь и др. Естественно, задание определенного вида функции потерь должно обуславливаться смыслом исследуемого процесса Y .

С помощью теории игр можно сформулировать и так называемые минимаксные прогнозы, смысл которых состоит в выборе такой прогнозирующей функции $d(x) \in D$, которая минимизирует максимальный риск $R(y, d)$, т.е. для которой

$$\sup R(y, d_0) = \inf \sup R(y, d), \quad d \in D, \quad y \in A. \quad (118)$$

Кроме того, с помощью теории игр можно реализовывать и качественные прогнозы, суть которых сводится к тому, чтобы оценить, произойдет ли в будущем некоторое случайное событие.

Прогноз может быть выполнен с использованием нейросети. В основе нейросети лежит понятие «искусственного нейрона», который позволяет реализовать на практике нелинейную функцию многих переменных. Она отображает совокупность входных переменных в вещественное число из отрезка $[0, 1]$ и зависит от набора G числовых коэффициентов (весов), рассматриваемого в качестве

параметра этой функции. Функция имеет вполне определенный вид и реализуется на базе только двух элементов – сумматора и нелинейного преобразователя. Первый вычисляет взвешенную сумму входных значений, т.е. компонент вектора признаков, второй выполняет определенное нелинейное преобразование $F(x, G)$, где G – набор весов на входе нейрона. В результате получается выходное значение нейрона $z = F(x, G)$. Так как обычно несколько нейронов «принимают сигналы» из предыдущего слоя, то набор весов на входе такого j -го нейрона обозначим G_j .

Из искусственных нейронов строится сеть, имеющая определенную архитектуру, которая включает несколько десятков, а иногда сотен или тысяч нейронов, соединенных своими входами и выходами. Веса нейронов будут определять веса соединений («ребер» сетевой структуры), влияющих на уровень сигнала, распространяющегося по нейросети в виде информационного потока от ее входов к выходу.

Наиболее популярной является такая архитектура сети, в которой нейроны каждого слоя непосредственно не взаимодействуют друг с другом и могут быть соединены своими входами-выходами только с расположенными на двух соседних с ними слоях. При этом одна часть нейронов, имеющих лишь по одному входу и выходу в первом слое, используется только для непосредственной обработки входных признаков, а другая – для обработки сигналов, полученных от нейронов из предыдущего (в частности, их первого).

Выходной слой нейронов используется для формирования результата. Если он содержит один нейрон, то в результате вычислений с помощью нейросети будет получено вещественное число, если же в него входят два нейрона (или более) – то вектор с двумя (или более) компонентами.

На этапе создания модели проводится обучение, или настройка, нейросети. В ее ходе встроенный в нейропакет алгоритм глобальной минимизации находит числовые веса, описывающие силу межнейронных связей. Критерием минимизации является уменьшение квадрата невязки между результатом расчетов по модели и реально наблюдаемой величиной. Веса фиксируются после завершения работы алгоритма минимизации и на этапе прогноза уже не изменяются.

На этапе настройки, обычно градиентными методами, получают числовые веса всех нейронов, принимающих сигналы от нейронов предыдущего слоя. По их набору $(G_1, \dots, G_n)$ минимизируется квадратичная функция потерь, которая задает штраф на значения отклика, действительно имевшими место при фиксированных весах и формируемой нейросетью, причем это выполняется для всех примеров, вошедших в материал обучения.

В результате настройки получается нелинейная модель связи входов и отклика $t(x) = W(x(G_1, \dots, G_n))$ в виде аппроксимации реальной их связи $T(x)$. Здесь через $(G_1, \dots, G_n)$ обозначена совокупность всех весовых коэффи-

циентов в нейросети некоторой фиксированной архитектуры. По работе нейросети на независимом наборе данных, который не использовался при ее настройке, оценивается качество последней.

Нейросети являются универсальным средством аппроксимации функции многих переменных $T(x)$, поэтому описанный подход позволяет наряду с применением традиционных моделей многомерной статистики решать типичные прикладные задачи классификации, распознавания и прогнозирования.

Основные достоинства нейросетей – адаптируемость (легкая переобучаемость при больших потоках входных данных), а также возможность потенциального распараллеливания вычислений. Нейросети будут предпочтительны там, где имеется очень много входных данных, в которых скрыты закономерности. В этом случае почти автоматически можно учесть различные нелинейные взаимодействия между показателями-признаками, характеризующими такие данные.

К недостаткам нейросети относится «непрозрачность» процесса их работы и трудности интерпретации результатов. Кроме того, правильно настроенная нейросеть, хотя и может адекватно оценивать сходные ситуации, обычно плохо проводит анализ принципиально новых ситуаций, не представленных ранее примерами в материале обучения.

Следует помнить, что для получения устойчивых нелинейных нейросетевых моделей с хорошими прогностическими свойствами требуется, чтобы объемом поступающей на их обучение информации был достаточно велик (т.е. число заданных для обучения примеров существенно превышало число нейронов ее входного слоя). При статистическом подходе, не в пример нейросетевому, в подобном случае можно построить доверительные интервалы.

Предложенная классификационная схема методов прогнозирования относится только к сингулярным методам, т.е. использующим лишь один принцип работы. Однако особенности сложных предметных областей вызывают трудности прогнозирования их развития с помощью какого-либо одного такого метода. Поэтому при разработке прогнозов для таких областей необходимо применять комплексные методы (или системы) прогнозирования, синтезирующие определенным образом алгоритмы сингулярных методов и эвристические знания о свойствах предметных областей.

8.3 Прогнозирование как оценка состояния динамической системы. Теория фильтрации

При разработке систем линейного предсказания и оценки состояния технических объектов широкое применение нашла теория фильтрации [5]. Для линей-

ных систем с линейными изменениями при гауссовских шумах теория фильтрации хорошо разработана и имеются точные алгоритмы оптимального решения.

С точки зрения общей классификации методов прогнозирования теория фильтрации относится к статистическим методам, так как в основе многих процедур оптимальной фильтрации лежит известный критерий Байеса.

Отправной точкой для теории фильтров явились работы, выполненные в 40-х годах Колмогоровым и Винером. Рассмотрим общую постановку задачи, из которой они исходили (постановка в терминах теории информации) [6].

Пусть полезный сигнал $s(t, \theta)$ зависит от скалярного параметра θ . Наблюдению доступен сигнал $x(t) = s(t, \theta) + n(t)$, где $n(t)$ – гауссовский шум. Требуется найти оценку $\hat{\theta}$ параметра θ .

В общем случае для решения такой задачи предлагается следующая процедура:

1. Задать множество значений параметров Θ , состоящее из всех возможных состояний параметра θ ;
2. Задать распределение случайной величины X , которая характеризуется условной плотностью распределения вероятностей $W(x, \theta)$, с математическим ожиданием $s(t, \theta)$ и остальными параметрами, совпадающими с распределением вероятности шума $n(t)$;
3. Задать множество оценок $\hat{\Theta}$;
4. Задать функцию $L(\theta, \hat{\theta})$, называемую функцией потерь, которая каждой паре $\theta, \hat{\theta}$ ставит в соответствие действительное число, называемое потерями. На практике наиболее часто употребляются две функции потерь: квадратичная ($L = (\theta - \hat{\theta})^2$) и модульная ($L = |\theta - \hat{\theta}|$);
5. Задать множество D , называемое пространством решений и состоящее из множества отображений случайной величины X в пространство оценок параметров $\hat{\Theta}$;
6. Задать функцию риска от принятия решения d при состоянии θ_1 в виде условного математического ожидания функции потерь

$$R(\theta_1, d) = E \{L(\theta, X | \theta = \theta_1)\};$$

7. Найти оптимальное решение из условия минимума байесовского риска:

$$r(d) = \int_{\Theta} R(\theta_1, d(X)) W(\theta_1) d\theta, \quad r(d^*) = \min_{d \in D} r(d). \quad (119)$$

На основании общего алгоритма решения разработано большое количество простых алгоритмов оптимальной фильтрации. Наиболее известные них:

- согласованный фильтр – максимизация отношения сигнал/шум;
- оптимальный фильтр Винера;
- оптимальный фильтр Калмана;
- адаптивные и нелинейные оптимальные фильтры (фильтр Бар-Шаломы-Ци, фильтр Бревера, адаптивный фильтр Калмана и т.п.).

После получения текущей оценки состояния динамической системы спрогнозировать состояние в будущем не представляет проблемы, т.к. переменные состояния однозначно определяют поведение системы в будущем.

8.4 Качественные показатели прогноза

О точности прогноза принято судить по величине погрешности (ошибки) прогноза – разности между прогнозируемым и фактическим значением (реализацией) исследуемой переменной [5]. Однако в практической работе проблему точности прогноза надо решать, как правило, тогда, когда период упреждения еще не прошел и истинное значение прогнозируемой переменной неизвестно. В этом случае проблема точности может рассматриваться в плане сопоставления априорных качеств, свойств, присущих альтернативным прогностическим моделям. Так, если прогнозирование осуществляется статистическими методами, то, вероятно, понятие точности прогноза можно сделать более узким, а именно связав априорную точность прогноза с размером доверительного интервала. Модель, дающая более узкий доверительный интервал при одной и той же доверительной вероятности, и является более точной.

Ширина доверительного интервала в значительной мере зависит от принятой доверительной вероятности. Чем меньше эта вероятность, тем уже интервал.

Очевидно, что достоверность прогноза определяется вероятностью реализации соответствующей прогностической оценки. Чем она выше, тем выше и достоверность. Вероятность реализации может быть оценена субъективно (экспертное прогнозирование) или может быть связана с доверительными интервалами прогноза. В этом случае достоверность является характеристикой, сопряженной мере погрешности, если под мерой погрешности понимается размер доверительного интервала. Отсюда чем выше достоверность прогноза, тем ниже его погрешность, и наоборот. В практике такого рода прогностических

расчетов обычно принимают, что доверительная вероятность равна 90% или менее.

Достоверность прогноза тесно связана с доверительной вероятностью прогноза, однако для оценки качества прогноза введем также понятие надежности прогноза.

Под надежностью прогноза будем понимать степень уверенности в том, что оценка элемента модели примет спрогнозированное значение в определенный момент времени в будущем.

Надежность прогноза базируется на анализе ряда оценок, полученного после предварительной обработки динамического ряда.

На надежность прогноза оказывают влияние такие параметры, как:

1. наличие данных для прогнозирования;
2. период времени между двумя последними оценками;
3. монотонность изменения значений оценок;
4. период упреждения прогноза;
5. относительная скорость изменения последней оценки и др.

8.5 Расчет фильтра Калмана

После того, как в 1963 г. Люэнбергер разработал детерминированную теорию наблюдателей, различные исследователи распространили эту теорию на дискретные и непрерывные стохастические системы, в результате чего была создана теория линейных стохастических наблюдателей.

Рассмотрим постановку типовой задачи оценивания состояния для дискретной системы [5]

$$x(k+1) = A(k)x(k) + B(k)w(k), \quad (120)$$

$$y(k) = C(k)x(k) + v(k), \quad (121)$$

где $x \in R^n$ – вектор состояния системы, $y \in R^m$ – вектор измерений (выходов), k – номер интервала дискретности, $w \in R^l$ и $v \in R^m$ – последовательности белых шумов, $A(k)$ – $(n \times n)$ -матрица динамики системы с рангом n , $B(k)$ – $(n \times l)$ -матрица входов системы с рангом l , $C(k)$ – $(m \times n)$ -матрица с рангом m . Статистические характеристики белых шумов полагаются известными:

$$M[w(k)] = 0, \quad M[w(k)w^T(k)] = \delta_{nl}Q(k), \quad (122)$$

$$M[v(k)] = 0, \quad M[v(k)v^T(k)] = \delta_{nm}R(k), \quad (123)$$

$$M [x(k) v^T(k)] = 0, \quad M [w(k) v^T(k)] = 0, \quad (124)$$

$$M [x(0) w^T(0)] = 0, \quad M [x(0)] = M_{x_0}, \quad (125)$$

$$M [(x(0) - M_{x_0})(x(0) - M_{x_0})^T] = D_{x_0}, \quad (126)$$

где δ – дельта-функция Кронекера, начальное состояние $x(0)$ представляет собой независимый случайный вектор с известными математическим ожиданием M_{x_0} и дисперсией D_{x_0} . Функционирование системы рассматривается на интервале времени $k = \overline{0, N}$.

В задаче оценивания состояния необходимо получить оценку $\hat{x}$ вектора состояния x , являющуюся оптимальной в некотором смысле. Если требуется, чтобы оценка была оптимальной в среднеквадратическом смысле (то есть оценка $\hat{x}$ должна минимизировать величину $M [\|\hat{x} - x\|^2]$), то решение задачи оценивания обеспечивает фильтр Калмана.

В основе фильтра Калмана лежит модель объекта управления, поведение которой корректируется обратной связью по невязке оценивания подобно наблюдателю Луэнбергера. Однако в измерительных сигналах объекта управления обычно присутствуют шумы. Из этого следует подход к расчету коэффициентов обратной связи, основанный на обеспечении наилучших точностных показателей.

Будем искать фильтр (наблюдатель состояния) в классе линейных оптимальных фильтров, дающих несмещенную оценку в виде линейной комбинации оценки предсказания $\tilde{x}(k+1) = A(k)\hat{x}(k)$ и вектора наблюдения $y(k+1)$:

$$\hat{x}(k+1) = L(k+1)\tilde{x}(k+1) + K(k+1)y(k+1). \quad (127)$$

Выражение для ошибки оценивания имеет вид

$$e(k+1) = \hat{x}(k+1) - x(k+1) = L(k+1)\tilde{x}(k+1) + K(k+1)y(k+1) - x(k+1). \quad (128)$$

Учитывая уравнение наблюдения и введя в рассмотрение ошибку предсказания

$$\tilde{e}(k+1) = \tilde{x}(k+1) - x(k+1), \quad (129)$$

получим

$$e(k+1) = L(k+1)\tilde{e}(k+1) + (L(k+1) + K(k+1)C(k+1) - I)x(k+1) + K(k+1)v(k+1). \quad (130)$$

Из этого уравнения следует, что если оценка предсказания $\tilde{x}(k+1)$ является несмещенной, т.е. $M[\tilde{e}(k)] = 0$, то и оценка $\hat{x}(k+1)$ будет также несмещенной,

т.е. $M[e(k)] = 0$, в том случае, если выполняется равенство

$$L(k+1) = I - K(k+1)C(k+1). \quad (131)$$

Таким образом, уравнение фильтрации можно записать в виде

$$\begin{aligned} \hat{x}(k+1) &= \tilde{x}(k+1) + K(k+1)(y(k+1) - C(k+1)\tilde{x}(k+1)) = \\ &= A(k)\hat{x}(k) + K(k+1)(y(k+1) - C(k+1)A(k)\hat{x}(k)). \end{aligned} \quad (132)$$

Определим значение матрицы $K(k+1)$ коэффициентов передачи фильтра из условия получения минимальной дисперсии ошибки фильтрации. Уравнение ошибки с учетом (130) и (131) примет вид

$$e(k+1) = (I - K(k+1)C(k+1))\tilde{e}(k+1) + K(k+1)v(k+1). \quad (133)$$

Запишем уравнение относительно матрицы дисперсий $P(k)$ ошибки фильтрации:

$$\begin{aligned} P(k+1) &= M[e(k+1)e^T(k+1)] = \\ &= (I - L(k+1)C(k+1))\tilde{P}(k+1)(I - L(k+1)C(k+1))^T + \\ &\quad + K(k+1)R(k+1)K^T(k+1), \end{aligned} \quad (134)$$

где $\tilde{P}(k) = M[\tilde{e}(k)\tilde{e}^T(k)]$ – априорная матрица дисперсий ошибки предсказания. Здесь учтено, что

$$M[\tilde{e}(k+1)v^T(k+1)] = 0,$$

так как оценка предсказания не зависит от шума наблюдения в следующий момент времени.

Выберем матрицу $K(k+1)$ таким образом, чтобы минимизировать скалярную функцию $trP(k+1)$, представляющую собой сумму дисперсий ошибок оценивания координат вектора состояния. Для этого приведем выражение (134) к виду

$$P(k+1) = \tilde{P} - KC\tilde{P} - \tilde{P}C^TK^T + KC\tilde{P}C^TK^T + KRK^T. \quad (135)$$

Продифференцировав $trP(k+1)$ по матрице $K(k+1)$ и приравняв производную нулевой матрице, получим

$$\begin{aligned} \frac{\partial trP(k+1)}{\partial K(k+1)} &= -2\tilde{P}(k+1)C^T(k+1) + \\ &+ 2K(k+1)C(k+1)\tilde{P}(k+1)C^T(k+1) + 2K(k+1)R(k+1) = 0, \end{aligned} \quad (136)$$

что эквивалентно

$$(I - K(k) C(k)) \tilde{P}(k) C^T(k) - K(k) R(k) = 0. \quad (137)$$

Из этого следует, что

$$K(k+1) = \tilde{P}(k+1) C^T(k+1) \left(R(k+1) + C(k+1) \tilde{P}(k+1) C^T(k+1) \right)^{-1}. \quad (138)$$

Выражение (138) обеспечивает минимальность функции $tr P(k+1)$ в силу положительной определенности матрицы $P(k+1)$. Подставим теперь выражение (138) в соотношение для матрицы дисперсий:

$$\begin{aligned} P(k+1) &= \tilde{P}(k+1) - K(k+1) C(k+1) \tilde{P}(k+1) - \\ &\quad - \tilde{P}(k+1) C^T(k+1) K^T(k+1) + \\ &\quad + K(k+1) \left(C(k+1) \tilde{P}(k+1) C^T(k+1) + R(k+1) \right) K^T(k+1) = \\ &= \tilde{P}(k+1) - \tilde{P}(k+1) C^T(k+1) \left(R(k+1) + C(k+1) \tilde{P}(k+1) C^T(k+1) \right)^{-1} \cdot \\ &\quad \cdot C(k+1) \tilde{P}(k+1) = (I - K(k+1) C(k+1)) \tilde{P}(k+1). \end{aligned} \quad (139)$$

Запишем соотношение для матрицы коэффициентов передачи в виде

$$K(k+1) = R(k+1) + C(k+1) \tilde{P}(k+1) C^T(k+1) = \tilde{P}(k+1) C^T(k+1).$$

Сгруппировав слагаемые с учетом выражения (139), получим

$$\begin{aligned} K(k+1) R(k+1) &= -(K(k+1) C(k+1) - I) \tilde{P}(k+1) C^T(k+1) = \\ &= P(k+1) C^T(k+1). \end{aligned}$$

Отсюда следует, что

$$K(k+1) = P(k+1) C^T(k+1) R^{-1}(k+1). \quad (140)$$

Остается найти соотношение, связывающее априорную $\tilde{P}(k)$ и апостериорную $P(k)$ матрицы. Вычтем из уравнения наблюдения уравнение объекта:

$$\tilde{e}(k+1) = A(k) e(k) - B w(k).$$

С учетом этого уравнение матрицы дисперсий ошибки предсказания принимает вид

$$\tilde{P}(k+1) = M [\tilde{e}(k+1) \tilde{e}^T(k+1)] = A(k) P(k) A^T(k) + B(k) Q(k) B^T(k).$$

В результате уравнение (139) преобразуется к виду

$$P(k) = \tilde{P}(k) - \tilde{P}(k) C^T(k) \left(C(k) \tilde{P}(k) C^T(k) + R(k) \right)^{-1} C(k) \tilde{P}(k), \quad (141)$$

которое является уравнением типа Риккати.

Аналогично получим уравнение Риккати для матрицы $\tilde{P}(k)$ априорных значений

$$\begin{aligned} \tilde{P}(k+1) = & A(k) \tilde{P}(k) A^T(k) + B(k) Q(k) B^T(k) - \\ & - A(k) \left(\tilde{P}(k) C^T(k) \left(C(k) \tilde{P}(k) C^T(k) + R(k) \right)^{-1} + C(k) \tilde{P}(k) \right) A^T(k). \end{aligned} \quad (142)$$

Приведенные уравнения описывают изменение матрицы $K(k)$ коэффициентов передачи фильтра и матриц дисперсий $P(k)$ и $\tilde{P}(k)$ во времени.

Остается доказать справедливость предположения о несмещенности оценки. Для этого подставим в выражение для величины $e(k)$ значение $e(k+1)$:

$$\tilde{e}(k+1) = A(k) (I - K(k) C(k)) \tilde{e}(k) + A(k) K(k) v(k) - B(k) w(k).$$

Усреднив правую и левую части полученного разностного уравнения, можно записать уравнение относительно среднего значения ошибки предсказания:

$$M_{\tilde{e}}(k+1) = A(k) (I - K(k) C(k)) M_{\tilde{e}}(k).$$

Отсюда следует, что $M_{\tilde{e}}(k) = 0$ в случае, если $M_{\tilde{e}}(0) = 0$, т.е.

$$M[\tilde{x}(0)] = M[x(0)] = M_{x_0}.$$

Рассмотрим алгоритм многомерного цифрового фильтра Калмана. Уравнения фильтра Калмана имеют вид

$$\hat{x}(k+1) = A(k) \hat{x}(k) + K(k+1) (y(k+1) - C(k+1) A(k) \hat{x}(k)), \quad (143)$$

$$K(k+1) = \tilde{P}(k+1) C^T(k+1) \left(C(k+1) \tilde{P}(k+1) C^T(k+1) + R(k+1) \right)^{-1}, \quad (144)$$

$$\tilde{P}(k+1) = A(k) P(k) A^T(k) + B(k) Q(k) B^T(k), \quad (145)$$

$$P(k+1) = (I - K(k+1) C(k+1)) \tilde{P}(k+1). \quad (146)$$

Для инициализации алгоритма Калмана в момент времени $k = 0$ принимается

$$\hat{x}(0) = M[x(0)] = M_{x_0},$$

$$P(0) = M \left[(\hat{x}(0) - x(0)) (\hat{x}(0) - x(0))^T \right] = D_{x_0}.$$

Вектор $\hat{x}(k+1)$ представляет собой минимальную среднеквадратичную оценку состояния $x(k+1)$ по измерениям, включающим момент времени k . Оценка формируется на основании прогнозируемой на один шаг оценки $\tilde{x}(k+1) = A(k)\hat{x}(k)$ и обратных связей по невязке оценивания $K(k+1)\tilde{y}(k+1)$. Матрицы $P(k+1)$ и $\tilde{P}(k+1)$ размерности $(n \times n)$ представляют собой ковариационные матрицы фильтрованной (апостериорной) и на один шаг вперед прогнозированной невязки оценивания (априорной) соответственно.

Если рассмотреть установившиеся свойства оптимального наблюдателя, то можно утверждать, что при некоторых ограничениях (матрицы A , B , C , Q , R являются постоянными и не вырожденными) решение для ковариационной матрицы невязки оценивания P сходится к установившемуся решению P_∞ , являющемуся постоянной матрицей, представляющей собой единственное неотрицательно определенное решение алгебраического уравнения типа Риккати:

$$P_\infty = A \left(P_\infty - P_\infty C^T (C P_\infty C^T + R)^{-1} C P_\infty \right) A^T + B Q B^T. \quad (147)$$

При этом матрица коэффициентов обратных связей также будет сходиться к стационарной:

$$K = P_\infty C^T (C P_\infty C^T + R)^{-1}. \quad (148)$$

Рассмотрим пример построения фильтра Калмана для получения текущей оценки состояния некоторой дискретной системы, математическая модель которой в форме вход-состояние-выход имеет следующий вид:

$$x(k+1) = \bar{A}x(k) + \bar{B}u(k), \quad (149)$$

$$y(k) = \bar{C}x(k). \quad (150)$$

На вход модели поступает скалярный белый дискретный шум $u(k) = w(k)$ с нулевым средним и единичной дисперсией:

$$M[w(k)] = 0, \quad M[w(k)w^T(k)] = Q = 1.$$

Дополним модель шумом измерений в виде аддитивного скалярного дискретного белого шума $v(k)$ с нулевым средним и известной дисперсией R :

$$M[v(k)] = 0, \quad M[v(k)v^T(k)] = Q = 0.01.$$

Шум измерений и шум состояния предполагаются некоррелированными:

$$M[w(k)v^T(k)] = 0.$$

Тогда модель (149)-(150) примет вид

$$x(k+1) = \bar{A}x(k) + \bar{B}w(k), \quad (151)$$

$$y(k) = \bar{C}x(k) + v(k). \quad (152)$$

В состав фильтра Калмана входит модель объекта управления без учета шумов и обратные связи по невязке наблюдения:

$$\hat{x}(k+1) = \bar{A}\hat{x}(k) + U_{\text{н}},$$

$$\hat{y}(k) = \bar{C}\hat{x}(k).$$

Вектор обратной связи по невязке наблюдения определяется выражением

$$U_{\text{н}} = K_{\text{н}}\tilde{y} = \begin{bmatrix} k_1^{\text{н}} & k_2^{\text{н}} \end{bmatrix} \times (y - \hat{y}).$$

Модель ошибок $\tilde{x} = x - \hat{x}$ примет вид

$$\tilde{x}(k+1) = (\bar{A} - K_{\text{н}}C) \tilde{x}(k) + \begin{bmatrix} \bar{B} & -K \end{bmatrix} \times \begin{bmatrix} w(k) \\ v(k) \end{bmatrix} = A_{\text{н}}\tilde{x}(k) + B_{\text{н}}\xi(k),$$

$$\tilde{y}(k) = \bar{C}\tilde{x}(k),$$

где

$$A_{\text{н}} = \bar{A} - K_{\text{н}}C, \quad B_{\text{н}} = \begin{bmatrix} \bar{B} & -K \end{bmatrix}, \quad \xi(k) = \begin{bmatrix} w(k) \\ v(k) \end{bmatrix}.$$

Определим ковариационную матрицу ошибки в установившемся режиме в виде решения алгебраического уравнения типа Риккати:

$$P_{\infty} = \bar{A} \left(P_{\infty} - P_{\infty} \bar{C}^T (\bar{C} P_{\infty} \bar{C}^T + R)^{-1} \bar{C} P_{\infty} \right) \bar{A}^T + \bar{B} Q \bar{B}^T.$$

Тогда коэффициенты обратной связи можно вычислить при помощи выражения

$$K = P_{\infty} C^T (C P_{\infty} C^T + R)^{-1}.$$

Рассмотрим теперь в качестве численного примера процедуру синтеза фильтра Калмана (ФК) для системы управления упрощенной моделью летательного аппарата. Положение центра масс летательного аппарата x_1 относительно наземной системы координат измеряется наземной радиолокационной станцией (РЛС). Переменная x_1 является длиной радиус-вектора, проведенного от антенны радиолокатора к центру масс летательного аппарата и измеряется в метрах. Скорость летательного аппарата x_2 определяется с использованием

сигнала линейного акселерометра на борту и измеряется в метрах в секунду. Таким образом, вектор состояния объекта управления имеет вид

$$x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}.$$

Математическая модель объекта управления в форме вход-состояние-выход имеет вид

$$\dot{x} = Ax + Bu - Gw, \quad (153)$$

$$y = Cx + v, \quad (154)$$

где y – показания РЛС (выходной сигнал объекта управления), u – сигнал от акселерометра (входной сигнал объекта управления), w – шумы в показаниях акселерометра, v – шумы в показаниях РЛС, а матрицы A , B , G и C имеют вид

$$A = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \quad B = \begin{bmatrix} 0 \\ b_2 \end{bmatrix}, \quad G = \begin{bmatrix} 0 \\ g_2 \end{bmatrix}, \quad C = [c_1 \ 0].$$

Подставив численные значения в выражения (153)-(154), получим итоговую математическую модель объекта управления:

$$\begin{aligned} \dot{x}_1 &= x_2, \\ \dot{x}_2 &= b_2 u - g_2 w, \\ y &= c_1 x_1 + v. \end{aligned} \quad (155)$$

Так как при помощи фильтра Калмана можно оценить лишь наблюдаемые переменные вектора состояния объекта управления, необходимо проверить его на наблюдаемость. Для этого нужно составить матрицу Q_H наблюдаемости объекта управления и определить ее ранг, который равен количеству наблюдаемых переменных вектора состояния. Кроме того, матрица наблюдаемости должна быть невырождена. Матрица наблюдаемости имеет вид

$$Q_H = \begin{bmatrix} C \\ CA \end{bmatrix} = \begin{bmatrix} c_1 & 0 \\ 0 & c_2 \end{bmatrix}.$$

Так как $\det Q_H = c_1^2 \neq 0$, матрица наблюдаемости является невырожденной. Ранг матрицы наблюдаемости равен 2, то есть равен порядку объекта управления, что говорит о полной наблюдаемости объекта управления, то есть с помощью фильтра Калмана возможно восстановить весь вектор состояния объекта управления.

В качестве моделей шумов w и v примем гауссовские стационарные случайные процессы с нулевыми средними значениями. В качестве ковариационных матриц шумов Q и R примем дисперсии шумов σ_w^2 и σ_v^2 соответственно.

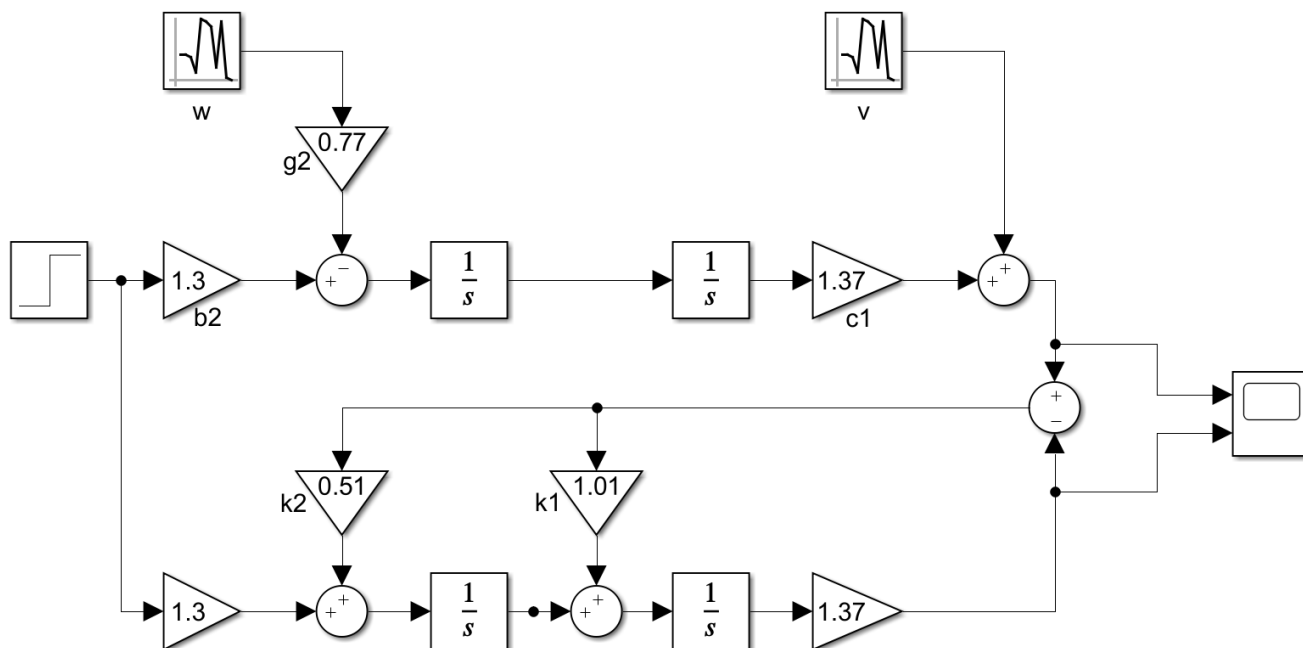


Рис. 18: Структурная схема объекта управления с фильтром Калмана.

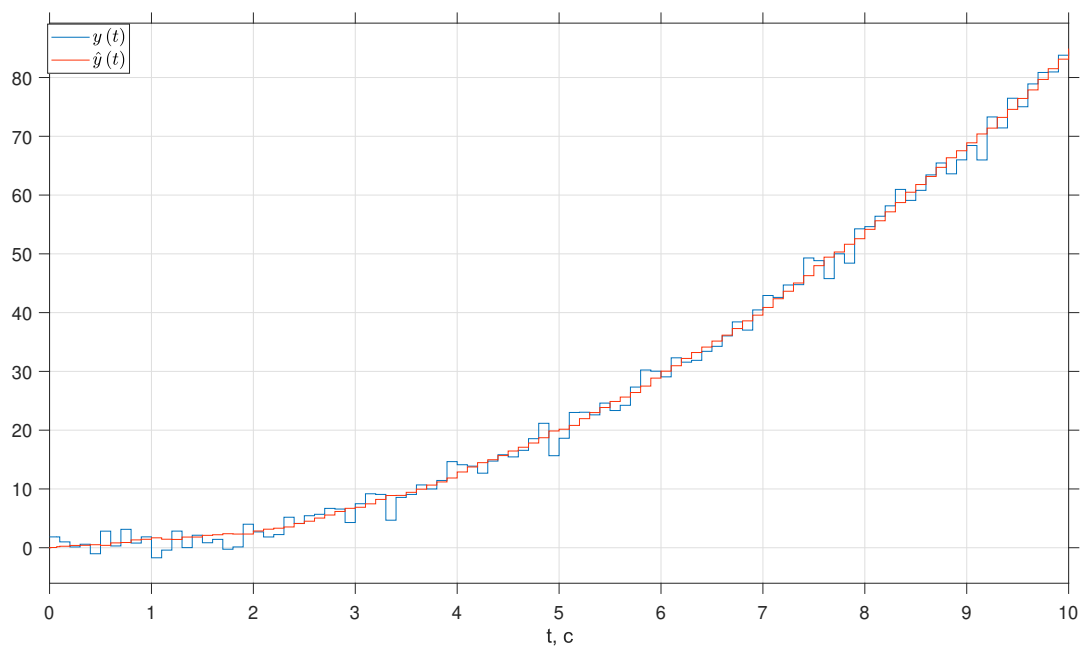


Рис. 19: Результаты моделирования, выход y и его оценка $\hat{y}$.

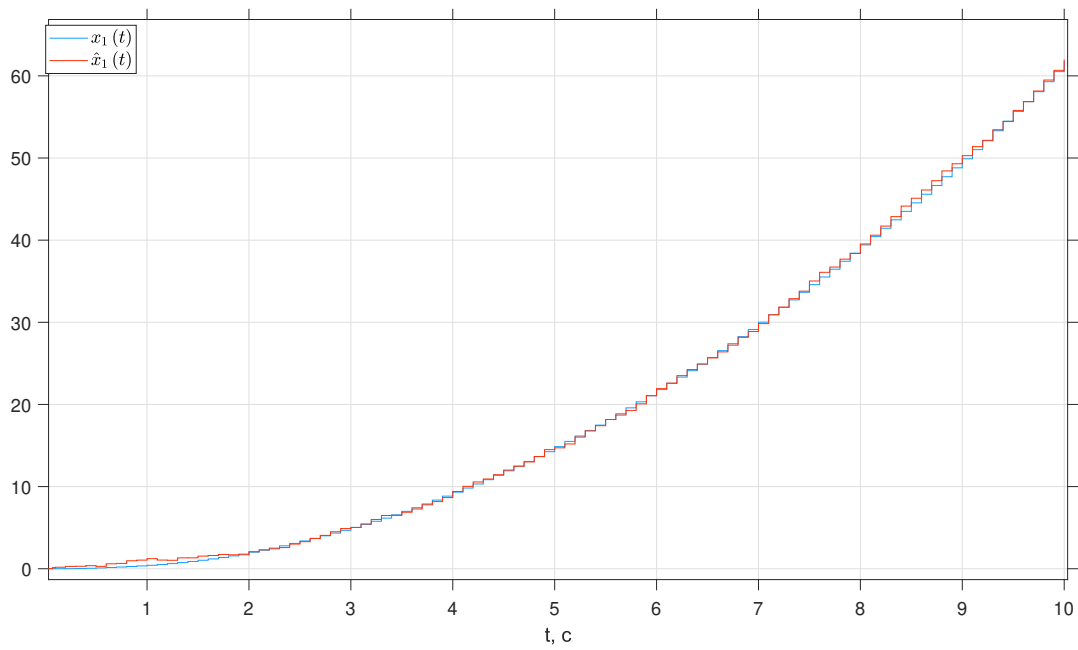


Рис. 20: Результаты моделирования, переменная состояния x_1 и ее оценка $\hat{x}_1$.

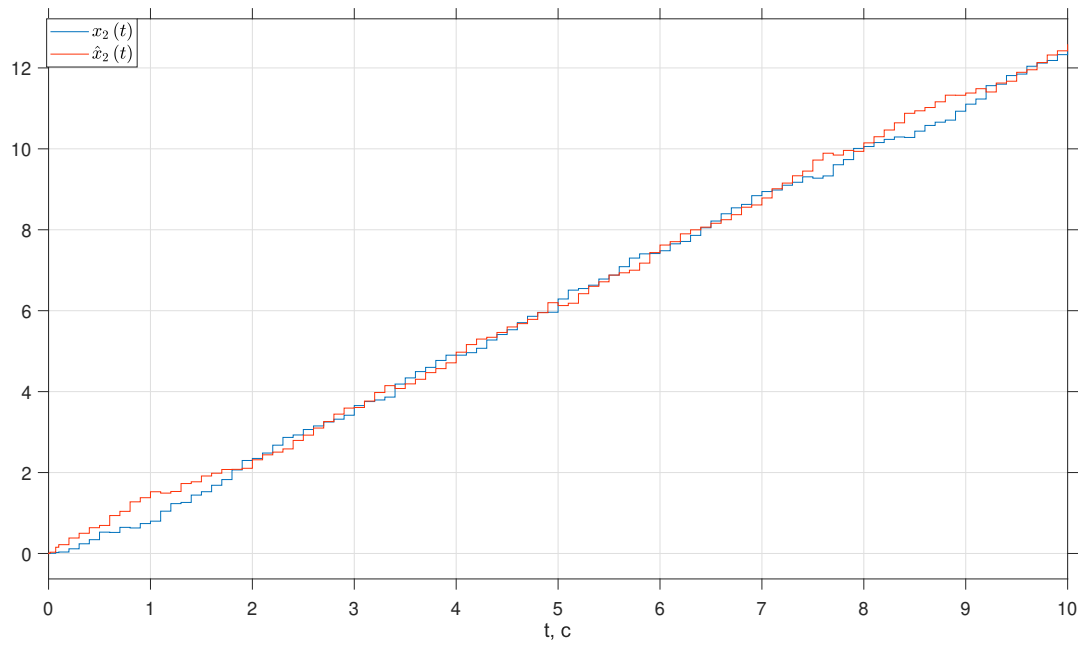


Рис. 21: Результаты моделирования, переменная состояния x_2 и ее оценка $\hat{x}_2$.

Далее необходимо вычислить матрицу ковариаций ошибок оценивания (невязок) P . Для этого необходимо решить уравнение Рикатти

$$\dot{P} = AP + PA^T - PC^T \frac{1}{\sigma_v^2} CP + G\sigma_w^2 G^T. \quad (156)$$

Ограничимся построением фильтра Калмана для стационарного оценивания и положим $\dot{P} = 0$, тогда матрицу P будем вычислять из уравнения

$$AP + PA^T - PC^T \frac{1}{\sigma_v^2} CP + G\sigma_w^2 G^T = 0. \quad (157)$$

Так как матрица P симметричная, примем обозначение для установившегося значения матрицы P

$$P_0 = \begin{bmatrix} p_1 & p_* \\ p_* & p_2 \end{bmatrix}.$$

После подстановки известных матриц в уравнение (157) получим

$$\begin{bmatrix} p_* & p_2 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} p_* & 0 \\ p_2 & 0 \end{bmatrix} - \begin{bmatrix} \sigma_v^{-2} c_1^2 p_1^2 & \sigma_v^{-2} c_1^2 p_1 p_* \\ \sigma_v^{-2} c_1^2 p_1 p_* & \sigma_v^{-2} c_1^2 p_*^2 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & \sigma_w^2 g_2^2 \end{bmatrix} = 0.$$

В силу симметрии для нахождения матрицы P необходимо решить систему из трех уравнений:

$$2p_* - \sigma_v^{-2} c_1^2 p_1^2 = 0, \quad (158)$$

$$p_2 - \sigma_v^{-2} c_1^2 p_1 p_* = 0, \quad (159)$$

$$-\sigma_v^{-2} c_1^2 p_*^2 + \sigma_w^2 g_2^2 = 0. \quad (160)$$

Из уравнения (160) найдем p_* :

$$p_*^2 = \frac{g_2^2 \sigma_w^2 \sigma_v^2}{c_1^2} \rightarrow p_* = \frac{g_2}{c_1} \sigma_w \sigma_v.$$

Подставив найденное значение p_* в (158), найдем p_1 :

$$\frac{2g_2}{c_1} \sigma_w \sigma_v = \frac{c_1^2}{\sigma_v^2} p_1^2 \rightarrow p_1 = \frac{\sqrt{2}}{c_1} \sigma_v^{3/2} \sqrt{g_2 \sigma_w}.$$

Подставив найденные значения p_* и p_1 в (159), найдем p_2 :

$$p_2 = \frac{c_1^2}{\sigma_v^2} p_1 p_* = \sqrt{2\sigma_v} (g_2 \sigma_w)^{3/2}.$$

Сформируем матрицу P_0 установившихся значений матрицы P

$$P_0 = \begin{bmatrix} \frac{\sqrt{2}}{c_1} \sigma_v^{3/2} \sqrt{g_2 \sigma_w} & \frac{g_2}{c_1} \sigma_w \sigma_v \\ \frac{g_2}{c_1} \sigma_w \sigma_v & \sqrt{2 \sigma_v} (g_2 \sigma_w)^{3/2} \end{bmatrix}.$$

Далее вычислим матрицу K коэффициентов усиления фильтра Калмана:

$$K = \begin{bmatrix} k_1 \\ k_2 \end{bmatrix} = P_0 C^T R^{-1} = \begin{bmatrix} \sqrt{2 g_2 \frac{\sigma_w}{\sigma_v}} \\ g_2 \frac{\sigma_w}{\sigma_v} \end{bmatrix}.$$

Структурная схема объекта управления с фильтром Калмана, реализованная в Simulink, приведена на рисунке 18.

Моделирование производилось при следующих значениях параметров: $b_2 = 1.3$, $g_2 = 0.77$, $c_1 = 1.37$, $\sigma_w^2 = 1.1$, $\sigma_v^2 = 2.5$. При этом коэффициенты фильтра Калмана приняли значения $k_1 = 1.0488$ и $k_2 = 1.58113883$. Результаты моделирования приведены на рисунках 19, 20 и 21.

По результатам моделирования видно, что синтезированный фильтр Калмана успешно справляется с задачей фильтрации случайных возмущений.

Вопросы для самоконтроля

1. В чем состоит задача прогнозирования и как можно классифицировать методы прогнозирования?
2. В чем заключается морфологический метод прогнозирования и какие этапы он в себя включает?
3. В чем заключается матричный метод прогнозирования и каковы его основные этапы?
4. Что такое метод когнитивного моделирования? Как составляется когнитивная карта?
5. В чем состоит метод экстраполяции? Какие методы оценки параметров зависимостей используют в рамках этого метода прогнозирования?
6. В чем заключается метод экспоненциального сглаживания? Какие существуют критерии выбора параметра сглаживания и порядка прогнозирующего полинома в рамках этого метода прогнозирования?
7. В чем заключается метод факторного анализа применительно к теории прогнозирования?

8. Как применяется в задачах прогнозирования математический аппарат теории игр?
9. Как может быть решена задача прогнозирования с использованием нейронных сетей?
10. Как решается задача оценки параметров сигнала с точки зрения теории фильтрации?
11. Что такое фильтр Калмана и как он строится? Какие задачи могут быть решены с использованием фильтра Калмана?

Список литературы

- [1] Булинский А. В., Ширяев А. Н. Теория случайных процессов. – М.: Физматлит, 2005. – 400 с.
- [2] Knill O. Probability Theory and Stochastic Processes with Applications. – Overseas Press, 2009. – 373 p.
- [3] Тихонов В. И., Миронов М. А. Марковские процессы. – М.: Советское радио, 1977. – 488 с.
- [4] Чистяков А. И. Статистические методы в инженерных исследованиях: Методические указания к лабораторно-практическим занятиям. – Ульяновск.: УлПИ, 1993. – 24 с.
- [5] Григорьев В.В., Быстров С.В., Бойков В.И., Болтунов Г.И., Коровьяков А.Н., Мансурова О.К., Першин И.М Проектирование регуляторов для стохастических систем и объектов с неопределенными параметрами – Санкт-Петербург: НИУ ИТМО, 2014. – 172 с.
- [6] Калман Р.Е., Бьюси Р.С. Новые результаты в линейной фильтрации и теории предсказания // Труды американского общества инженеров-механиков. – Техническая механика. – 1961,- т. 83. Сер.Д.- №1.

Краснов Александр Юрьевич

Статистические методы в инженерных исследованиях

Учебно-методическое пособие

В авторской редакции

Редакционно-издательский отдел Университета ИТМО

Зав. РИО

Н.Ф. Гусарова

Подписано к печати

Заказ №

Тираж

Отпечатано на ризографе

Редакционно-издательский отдел
Университета ИТМО
197101, Санкт-Петербург, Кронверкский пр., 49, литер А