

Министерство образования и науки Российской Федерации
Ярославский государственный университет им. П. Г. Демидова

Ю. А. БРЮХАНОВ, А. Л. ПРИОРОВ

ОБЩАЯ ТЕОРИЯ СВЯЗИ

Учебное пособие

Рекомендовано

*Научно-методическим советом университета
для студентов, обучающихся по направлению
Инфокоммуникационные технологии и системы связи*

Ярославль
ЯрГУ
2014

УДК 621.391(075)
ББК 3811я73
Б89

*Рекомендовано
Редакционно-издательским советом университета
в качестве учебного издания. План 2014 года*

Рецензенты:

Витязев В. В., доктор технических наук;
кафедра основ радиотехники
Московского энергетического института
(технического университета)

Б89 **Брюханов, Юрий Александрович.**
Общая теория связи : учебное пособие / Ю. А. Брюханов,
А. Л. Приоров ; Яросл. гос. ун-т им. П. Г. Демидова. – Ярославль :
ЯрГУ, 2014. – 216 с.

ISBN 978-5-8397-0993-5

Описаны статистические методы современной теории связи.
Пособие предназначено для студентов, обучающихся по
направлению 210700.62 Инфокоммуникационные технологии и
системы связи (дисциплина «Общая теория связи. Часть II»,
цикл Б3), очной формы обучения.

Материал может быть использован также при подготовке
студентами курсовых и выпускных квалификационных работ.

Ил. 81. Табл. 3. Библиогр. : 13 назв.

УДК 621.391(075)
ББК 3811я73

© ЯрГУ, 2014

ISBN 978-5-8397-0993-5

ОГЛАВЛЕНИЕ

ПРИНЯТЫЕ СОКРАЩЕНИЯ.....	5
ВВЕДЕНИЕ.....	7
1. ТЕОРИЯ СЛУЧАЙНЫХ ПРОЦЕССОВ.....	9
1.1. Случайные процессы и их вероятностные характеристики	9
1.2. Числовые характеристики случайных процессов.....	14
1.3. Свойства корреляционной функции случайного процесса	17
1.4. Спектральный анализ случайных процессов	21
1.5. Примеры энергетических спектров стационарных случайных процессов	24
Примеры	29
Контрольные вопросы	35
Литература.....	36
2. ТЕОРИЯ ОЦЕНОК.....	38
2.1. Оценка параметра.....	38
2.2. Свойства точечных оценок.....	39
2.3. Метод максимального правдоподобия	40
2.4. Оценка Байеса.....	41
2.5. Среднеквадратичная оценка.....	42
2.6. Линейная среднеквадратичная оценка	43
Примеры	45
Контрольные вопросы	47
Литература.....	47
3. ТЕОРИЯ ПРИНЯТИЯ РЕШЕНИЙ	49
3.1. Проверка гипотез.....	49
3.2. Критерии проверки гипотез.....	51
Примеры	56
Контрольные вопросы	62
Литература.....	63
4. ПРОХОЖДЕНИЕ СЛУЧАЙНЫХ СИГНАЛОВ ЧЕРЕЗ ЭЛЕКТРИЧЕСКИЕ ЦЕПИ.....	65
4.1. Прохождение случайных сигналов через безынерционные цепи	65
4.2. Прохождение случайных сигналов через линейные цепи	71
4.3. Узкополосные случайные процессы.....	73
Примеры	83
Контрольные вопросы	88
Литература.....	89

5. ОПТИМАЛЬНЫЙ ПРИЕМ	91
5.1. Постановка задачи при оптимальном приеме дискретных сообщений	91
5.2. Критерии качества приема дискретных сообщений	93
5.3. Синтез оптимального демодулятора при известном ансамбле сигналов	100
5.4. Потенциальная помехоустойчивость когерентного приема...	120
5.5. Сравнительный анализ потенциальной помехоустойчивости основных видов цифровой модуляции	125
5.6. Синтез оптимального демодулятора в канале с неопределенной фазой.....	130
5.7. Потенциальная помехоустойчивость некогерентного приема в двоичной системе связи	136
5.8. Оптимальная фильтрация непрерывных сигналов.....	141
Примеры	147
Контрольные вопросы	150
Литература.....	154
6. ОСНОВЫ ТЕОРИИ ИНФОРМАЦИИ.....	156
6.1. Энтропия источника, количество информации	156
6.2. Информационные характеристики источников сообщений	160
Примеры	168
Контрольные вопросы	171
Литература.....	171
7. ОСНОВЫ ТЕОРИИ КОДИРОВАНИЯ.....	173
7.1. Теорема о блочном кодировании (теорема Шеннона для канала с шумами).....	174
7.2. Линейные блочные коды	175
7.3. Способность кода к обнаружению и исправлению ошибок....	176
7.4. Специальные циклические коды. Код Голея	186
Примеры	201
Контрольные вопросы	203
Литература.....	205
ЗАКЛЮЧЕНИЕ	207
ПРЕДМЕТНЫЙ УКАЗАТЕЛЬ	208
ГЛОССАРИЙ	211

ПРИНЯТЫЕ СОКРАЩЕНИЯ

АБГШ – аддитивный белый гауссовский шум
АД – амплитудный детектор
АКФ – автокорреляционная функция
АЧХ – амплитудно-частотная характеристика
АЦП – аналогово-цифровой преобразователь
БГШ – белый гауссовский шум
БИХ – бесконечная импульсная характеристика
БПФ – быстрое преобразование Фурье
БФУ – безынерционный функциональный узел
БШ – белый шум
ВИМ – время-импульсная модуляция
ВК – весовые коэффициенты
ВКФ – взаимная корреляционная функция
ДО – детектор огибающей
ДПФ – дискретное преобразование Фурье
ЗУ – запоминающее устройство
ИХ – импульсная характеристика
КБШ – квазibelый шум
КИН – критерий идеального наблюдателя (критерий ИН, критерий Котельникова)
КИХ – конечная импульсная характеристика
КМП – критерий максимального правдоподобия (критерий МП)
КМСР – критерий минимального среднего риска
КНП – критерий Неймана-Пирсона (критерий НП)
КФ – корреляционная функция
ЛЗ – линия задержки
ЛПП – линейная (цепь) с постоянными параметрами
ЛИС – линейная инвариантная к сдвигу (цепь)
ЛС – линейная стационарная (цепь)
ЛТ – ложная тревога
МП – максимальное правдоподобие
МСИ – межсимвольная интерференция
НПФ – непрерывное преобразование Фурье
ОДПФ – обратное дискретное преобразование Фурье
ОП – отношение правдоподобия

ОПФ – оконное преобразование Фурье
ОС – обратная связь
ОФМ – относительная фазовая манипуляция
ПЗ – полоса задерживания фильтра
ПП – полоса пропускания фильтра
ПВ – плотность вероятности
ПРВ – плотность распределения вероятности
ПФ – полосовой фильтр
ПХ – переходная характеристика
РУ – решающее устройство
СКК – сигнально-кодовая конструкция
СКО – среднеквадратическое отклонение
СВ – случайная величина
СП – случайный процесс
ССП – стационарный случайный процесс
СТС – синхронный телеграфный сигнал
СФ – согласованный фильтр
УГО – условное графическое обозначение
УПР – устройство принятия решений
ФВЧ – фильтр верхних частот
ФД – фазовый детектор
ФНЧ – фильтр нижних частот
ФП – функция правдоподобия
ФР – функция распределения
ФУ – функциональный узел
ФЧХ – фазочастотная характеристика
ЧХ – частотная характеристика
ЦАП – цифроаналоговый преобразователь
ЦОС – цифровая обработка сигналов
ЦФ – цифровой фильтр
ЭП – эргодический процесс (эргодический стационарный случайный процесс)
ЭС – энергетический спектр

ВВЕДЕНИЕ

*Чтобы теорию познать,
придется много потрудиться.*

В начале XXI века состояние и перспективы развития информационных технологий характеризуются широким практическим использованием статистических методов теории связи. Эта область является одной из динамично развивающихся областей в мире инфокоммуникаций. Ее методы вызывают повышенный интерес ученых и специалистов, работающих в различных областях, таких как связь, радиотехника, радиовещание и телевидение, системы управления, измерительная техника и приборостроение. Ежегодно в мире проводятся десятки международных научно-технических конференций и семинаров, посвященных решению актуальных проблем статистической теории связи.

Крупный вклад в статистическую теорию связи внесли американские ученые К. Шеннон и Н. Винер, а также отечественные ученые В. А. Котельников и В. И. Тихонов. Заметный вклад в развитие указанной теории внесли также и другие зарубежные и отечественные ученые, труды которых во многом актуальны до настоящего времени.

Дисциплина «Общая теория связи» является базовой для изучения связистами. Здесь рассматриваются основы теории случайных процессов, теории оценивания, теории принятия решений, спектральная теория случайных сигналов, прохождение их через линейные и нелинейные безынерционные цепи, оптимальный когерентный и некогерентный прием, оптимальная фильтрация Винера, методы помехоустойчивого кодирования. Большую часть этого материала можно найти в известных книгах Д. Д. Кловского и А. П. Сальникова.

Пособие состоит из семи разделов. Разделы 1–3 по сути являются математическим введением в курс. Они приведены для повторения некоторых разделов теории вероятностей и математической статистики, которые дальше активно

используются. Раздел 6 по основам теории информации приведен здесь только для лучшего понимания раздела 7 по теории кодирования. Этот материал уже изучался в первой части данного курса.

Учебное пособие обобщает многолетний опыт авторов и их коллег по преподаванию данного материала студентам старших курсов связанных специальностей и направлений физического факультета Ярославского государственного университета им. П. Г. Демидова. Материал может быть использован при решении задач курсового и дипломного проектирования, а также для выполнения студенческих научных работ.

Выражаем глубокую благодарность нашим коллегам, которые оказали значительное влияние на формирование взглядов авторов в данном научном направлении. Среди них особенно хочется выделить члена-корреспондента РАН Ю. Б. Зубарева, профессоров В. П. Дворковича, А. В. Дворковича, В. В. Витязева, А. А. Петровского, Б. А. Алпатотова, Ю. С. Бехтина, В. Г. Карташева, М. А. Щербакова, А. К. Бернюкова, кандидата технических наук А. М. Лаврентьева. Взгляды авторов формировались также в совместной работе с нашими коллегами с кафедры динамики электронных систем ЯрГУ, среди которых отметим В. В. Хрящева, А. Н. Тараканова, И. В. Апалькова, И. С. Мочалова, М. А. Дубова, А. И. Топникова, С. А. Новоселова, Е. В. Сергеева. Всем им авторы очень признательны и за совместную исследовательскую работу, и за помощь в обеспечении учебного процесса.

Отдельные слова благодарности старшему преподавателю, кандидату технических наук В. А. Волохову за предоставленные материалы и консультации, ведущему электронике Ю. А. Лукашевичу, электроникам И. Трапезникову, А. Лебедеву, Д. Горшкову, ассистенту кафедры ДЭС А. Носкову, аспиранту этой кафедры Е. Аминовой, студенту В. Антипову за активное участие в подготовке материалов пособия.

1. ТЕОРИЯ СЛУЧАЙНЫХ ПРОЦЕССОВ

В современных системах связи как используемые для передачи информации сигналы, так действующие в них шумы и помехи по своей природе являются случайными процессами. Поэтому для их описания удобно использовать математический аппарат теории вероятностей, случайных процессов и математической статистики.

1.1. Случайные процессы и их вероятностные характеристики

Случайный процесс (СП) $X(t)$ является функцией времени, значения которой в любой фиксированный момент времени t_i представляют собой случайную величину $X(t_i)$. Здесь и в дальнейшем случайные величины и функции будем обозначать заглавными буквами, а детерминированные (неслучайные) – строчными.

На рис. 1.1 изображены возможные реализации $x_1(t)$, $x_2(t)$, ... $x_k(t)$ случайного процесса $X(t)$. Важно отметить, что они являются детерминированными функциями времени.

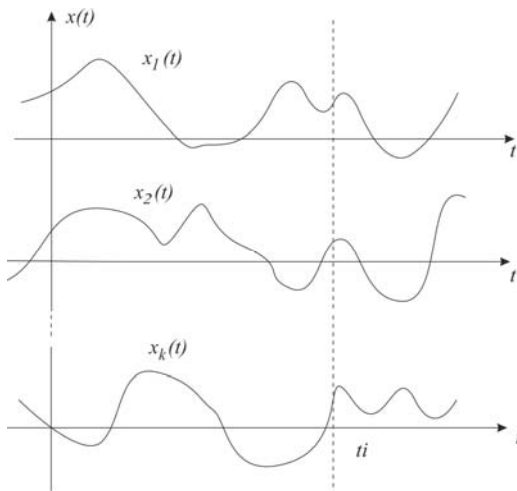


Рис. 1.1. Возможные реализации случайного процесса

Сам процесс можно трактовать как множество (в том числе, и несчетное) подобных реализаций $\{x_k(t)\}$ с соответствующей вероятностной мерой.

Для полного описания сечений $X(t_i)$ СП необходимо указать законы распределения значений СП в этих сечениях. Они могут быть заданы в интегральной (функция распределения) или дифференциальной (плотность вероятности) формах. В табл. 1.1 приведены основные сведения об этих законах и их свойствах.

Таблица 1.1

**Основные свойства функции распределения
и плотности вероятности**

Название и обозначение	Функция распределения $F(x)$	Плотность вероятности $w(x)$
Определение	$F(x) = P(X \leq x)$	$w(x) = \lim_{\Delta x \rightarrow 0} \frac{P(x \leq X \leq x + \Delta x)}{ \Delta x }$
Расчет вероятности $P(x_1 \leq X \leq x_2)$	$P(x_1 \leq X \leq x_2) = F(x_2) - F(x_1)$	$P(x_1 \leq X \leq x_2) = \int_{x_1}^{x_2} w(x) dx$
Взаимосвязь	$F(x) = \int_{-\infty}^x w(y) dy$	$w(x) = \frac{dF(x)}{dx}$
Особенности функции	$F(x_2) \geq F(x_1)$ при $x_2 > x_1$ (неубывающая)	$w(x) \geq 0$ (неотрицательная)
Свойство нормировки	$F(-\infty) = 0,$ $F(\infty) = 1$	$\int_{-\infty}^{\infty} w(x) dx = 1$
Физическая размерность	безразмерная	размерность $\frac{1}{X}$

Примеры распределений случайных величин

Равномерное распределение

На рис. 1.2 приведены плотность вероятности и функция распределения равномерного распределения

$$w(x) = \begin{cases} 0 & x \leq x_1, \\ \frac{1}{x_2 - x_1} & x_1 \leq x \leq x_2, \\ 0 & x \geq x_2, \end{cases} \quad F(x) = \begin{cases} 0 & x \leq x_1, \\ \frac{x - x_1}{x_2 - x_1} & x_1 \leq x \leq x_2, \\ 1 & x \geq x_2. \end{cases}$$

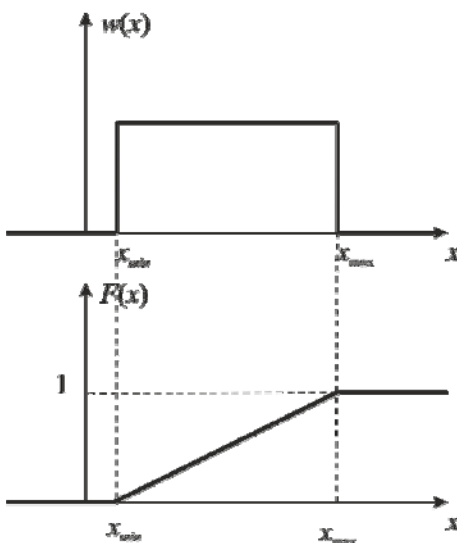


Рис. 1.2. Плотность вероятности и функция распределения равномерного распределения ($x_{\min}=x_1$, $x_{\max}=x_2$)

Нормальное распределение

На рис. 1.3 приведены плотность вероятности и функция распределения нормального (гауссовского) распределения

$$w(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-m)^2}{2\sigma^2}}, \quad F(x) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^x e^{-\frac{(y-m)^2}{2\sigma^2}} dy.$$

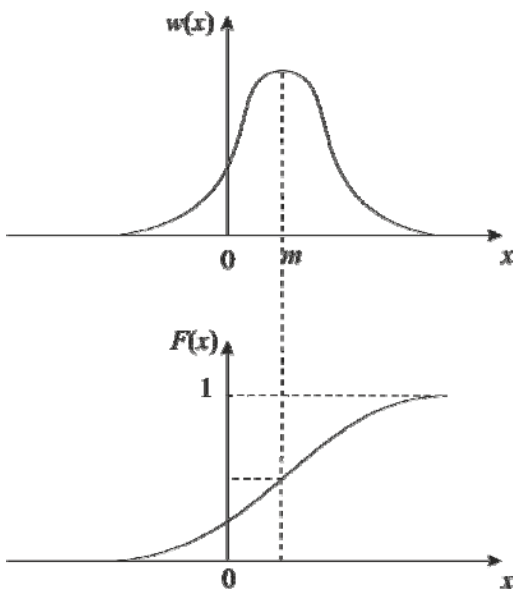


Рис. 1.3. Плотность вероятности и функция распределения нормального (гауссовского) закона распределения

Распределение дискретной случайной величины

На рис. 1.4 приведены закон распределения дискретной случайной величины и ее функция распределения

$$F(x) = \sum_{i=1}^4 P(x_i) \cdot 1(x - x_i), \quad w(x) = \sum_{i=1}^4 P(x_i) \cdot \delta(x - x_i).$$

Информация о сечениях СП не является достаточной для описания самого СП, так как не содержит сведений о зависимостях сечений между собой.

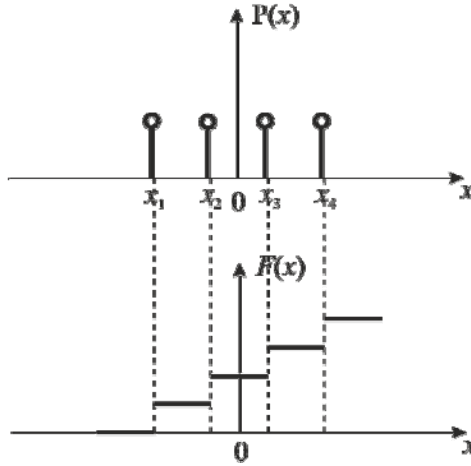


Рис. 1.4. Закон распределения дискретной случайной величины и ее функция распределения

Исчерпывающее описание СП осуществляется с помощью n -мерной функции распределения

$$F(x_1, x_2, \dots, x_n; t_1, t_2, \dots, t_n) = P[X(t_1) \leq x_1, X(t_2) \leq x_2, \dots, X(t_n) \leq x_n]$$

или n -мерной плотности вероятности

$$w(x_1, x_2, \dots, x_n; t_1, t_2, \dots, t_n) = \frac{\partial^n F(x_1, x_2, \dots, x_n; t_1, t_2, \dots, t_n)}{\partial x_1 \partial x_2 \cdots \partial x_n},$$

где x_1, x_2, \dots, x_n – аргументы, t_1, t_2, \dots, t_n – параметры этих функций, а n – любое целое число.

Если n -мерная плотность вероятности или функция распределения СП не меняется при сдвиге всех моментов t_k ($k = 1, 2, \dots, n$) на один и тот же интервал Δt , то такой случайный процесс называют стационарным в узком смысле.

1.2. Числовые характеристики случайных процессов

Полное описание случайных процессов не всегда возможно, а во многих случаях и не требуется. Тогда достаточно знать их основные характеристики. В качестве них часто используются математическое ожидание, дисперсия и корреляционная функция. Их удобно ввести через моменты k -го порядка (рис. 1.5), которые также используются для описания СП.



Рис. 1.5. Классификация моментов случайных процессов

Начальные моменты k -го порядка можно записать в виде

$$m_k[X(t)] = \begin{cases} \int_{-\infty}^{\infty} x^k w(x, t) dx & \text{- для непрерывного СП,} \\ \sum_{i=1}^m x_i^k P(x_i, t) & \text{- для дискретного СП,} \end{cases}$$

а центральные моменты k -го порядка – в виде

$$\mu_k[X(t)] = \overline{X^k(t)} = \begin{cases} \int_{-\infty}^{\infty} [x - \overline{X(t)}]^k w(x, t) dx & \text{- для непрерывного СП,} \\ \sum_{i=1}^m [x_i - \overline{X(t)}]^k P(x_i, t) & \text{- для дискретного СП.} \end{cases}$$

При описании центрального момента использовано понятие **центрированного** случайного процесса

$$\overset{\circ}{X}(t) = X(t) - \overline{X(t)}.$$

1. **Математическое ожидание** СП – начальный момент первого порядка

$$M[X(t)] = \overline{X(t)} = \begin{cases} \int_{-\infty}^{\infty} xw(x,t)dx & \text{- для непрерывного СП,} \\ \sum_{i=1}^m x_i P(x_i,t) & \text{- для дискретного СП.} \end{cases}$$

Математическое ожидание показывает относительное расположение распределения на числовой прямой (рис. 1.6).

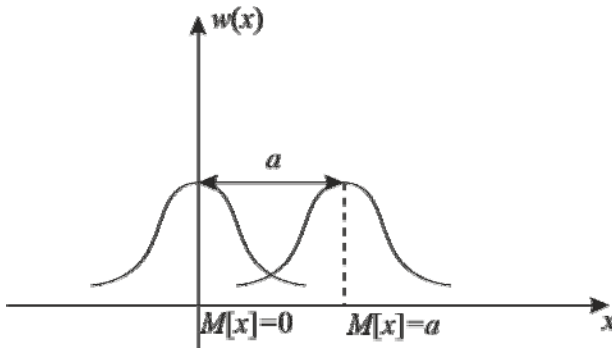


Рис. 1.6. Случайный процесс с нулевым и ненулевым математическим ожиданием

2. **Дисперсия СП** – центральный момент второго порядка

$$D[X(t)] = \overline{[X(t) - \overline{X(t)}]^2} =$$

$$= \overset{\circ}{X}^2(t) = \begin{cases} \int_{-\infty}^{\infty} [x - \overline{X(t)}]^2 w(x,t)dx & \text{- непр. СП,} \\ \sum_{i=1}^m [x_i - \overline{X(t)}]^2 P(x_i,t) & \text{- дискр. СП.} \end{cases}$$

Дисперсия распределения показывает его разброс вокруг среднего значения (рис. 1.7). Приблизительно с 0,9973 вероятностью значение нормально распределённой случайной величины лежит в интервале $(\bar{x} - 3\sigma; \bar{x} + 3\sigma)$. Это свойство известно как «правило трех сигм».

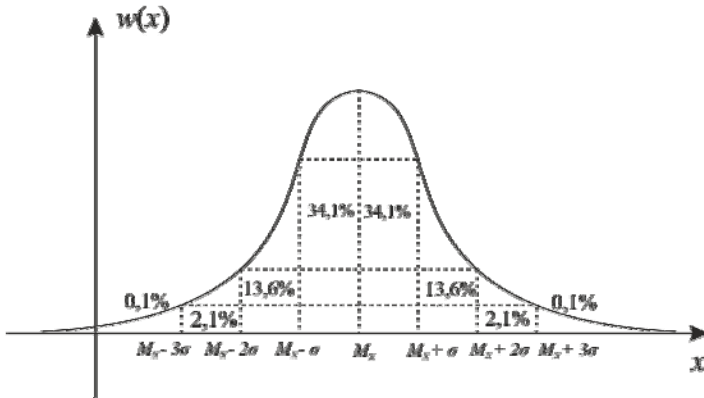


Рис. 1.7. Иллюстрация к правилу трех сигм

Видно, что указанные моменты полностью определяются одномерным распределением и в общем случае произвольного СП являются детерминированными функциями времени. Важно подчеркнуть, что для стационарных в узком смысле СП моменты от времени не зависят.

3. Корреляционная (автокорреляционная) функция (КФ, АКФ, используется также термин «функция корреляции») – центральный смешанный момент второго порядка

$$B_X(t_1, t_2) = \overline{\dot{X}(t_1) \dot{X}(t_2)} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left[x_1 - \overline{X(t_1)} \right] \left[x_2 - \overline{X(t_2)} \right] w(x_1, x_2; t_1, t_2) dx_1 dx_2$$

Случайные процессы называют стационарными в широком смысле (ССП), если выполняются следующие условия:

$$M[X(t)] = \text{const},$$

$$D[X(t)] = \text{const},$$

$$B_X(t_1, t_2) = B_X(\tau), \text{ где } \tau = t_2 - t_1.$$

Стационарность СП в узком смысле влечет его стационарность в широком смысле. Обратное утверждение не выполняется.

1.3. Свойства корреляционной функции случайного процесса

$$1. \overline{\overset{\circ}{X}(t) \overset{\circ}{X}(t)} = \overline{\overset{\circ}{X}^2(t)} = D[X(t)].$$

$$2. |B_X(\tau)| \leq B_X(0) = D[X(t)].$$

3. Корреляционная функция характеризует статистическую связь сечений СП (внутри процесса). Если связи между сечениями $X(t_1)$ и $X(t_2)$ нет (сечения статистически независимы), то $B_X(t_1, t_2) = 0$.

Важно отметить, что отсутствие связи влечет отсутствие корреляции, но не наоборот (рис. 1.8). Обратное утверждение справедливо лишь в случае нормального СП.

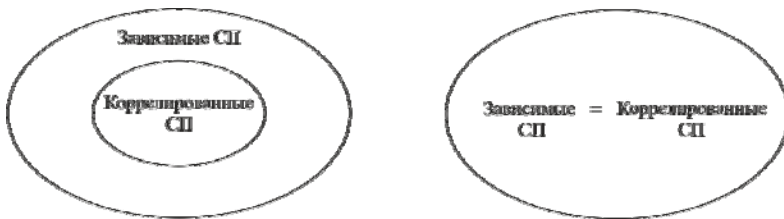


Рис. 1.8. Множества коррелированных и зависимых СП: а) в общем случае; б) для нормального СП

Нормальным (гауссовским) называют случайный процесс, у которого одномерная плотность вероятности имеет вид

$$w(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x-m)^2}{2\sigma^2}},$$

где $m = M[X(t)] = \overline{X(t)}$, $\sigma^2 = D[X(t)]$, а любая n -мерная плотность вероятности описывается выражением

$$w(x_1, x_2, \dots, x_n; t_1, t_2, \dots, t_n) = A_n e^{-\sum_{i=1}^n \sum_{j=1}^n c_{ij} (x_i - a_i)(x_j - a_j)},$$

где A_n , c_{ij} , a_i , a_j – константы, определяемые выбором сечений t_1, t_2, \dots, t_n .

4. Корреляционная функция стационарного случайного процесса является четной $B_X(\tau) = B_X(-\tau)$.

5. Коэффициент корреляции. Чтобы абстрагироваться от дисперсии и учитывать только связи внутри СП, удобно пользоваться нормированной функцией корреляции, которая называется также коэффициентом корреляции –

$$R_X(\tau) = \frac{B_X(\tau)}{D[X(t)]} = \frac{B_X(\tau)}{B_X(0)}.$$

Видно, что $R_X(0) = 1$, $-1 \leq R_X(\tau) \leq 1$.

Интервал корреляции – это грубая числовая оценка связи внутри случайного процесса. Ее часто определяют методом равновеликого прямоугольника

$$\tau_{\text{кор}} = \int_0^{\infty} |R(\tau)| d\tau.$$

Взаимная корреляционная функция двух случайных процессов $X(t)$ и $Y(t)$

$$\begin{aligned} B_{XY}(t_1, t_2) &= \frac{1}{2} \left[\overline{\dot{X}(t_1) \dot{Y}(t_2)} + \overline{\dot{X}(t_2) \dot{Y}(t_1)} \right] = \\ &= \overline{\dot{X}(t_1) \dot{Y}(t_2)} = \overline{\dot{X}(t_2) \dot{Y}(t_1)} = B_{YX}(t_1, t_2). \end{aligned}$$

Корреляционная функция суммы независимых случайных процессов $Z(t) = X(t) + Y(t)$ есть сумма

корреляционных функций каждого из слагаемых СП в отдельности:

$$B_Z(t_1, t_2) = B_X(t_1, t_2) + B_Y(t_1, t_2).$$

Часто удобно вместо усреднения по множеству реализаций (по ансамблю) случайного процесса ввести его усреднение по времени, определяя:

$$\begin{aligned} \widetilde{\widetilde{X}}(t) &= \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} X(t) dt && \text{– постоянную} \\ &&& \text{составляющую СП,} \\ \widetilde{X}(t) &= X(t) - \widetilde{\widetilde{X}}(t) && \text{– переменную} \\ &&& \text{составляющую СП,} \\ \widetilde{\widetilde{X}}^2(t) &= \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} \widetilde{X}^2(t) dt = P_X && \text{– мощность} \\ &&& \text{переменной} \\ &&& \text{составляющей СП.} \end{aligned}$$

Видно, что эти характеристики являются случайными величинами, не зависящими от времени.

Стационарные случайные процессы называют эргодическими, если их усреднение по множеству и по времени приводит к одинаковым результатам:

$$\begin{aligned} \overline{\widetilde{\widetilde{X}}(t)} &= \widetilde{\widetilde{X}}(t) = \text{const}, \\ \overline{\widetilde{X}^2(t)} &= D[X(t)] = \widetilde{\widetilde{X}}^2(t) = P_X = \text{const}. \end{aligned}$$

Связь между этими типами случайных процессов показана на рис. 1.9.

Можно считать, что свойство эргодичности случайного процесса заключается в том, что все его реализации в некотором смысле похожи друг на друга. Отсюда следует возможность получения указанных характеристик эргодического СП усреднением по времени единственной его реализации $x(t)$. Это существенно облегчает построение аппаратуры для измерений этих характеристик. В частности, КФ эргодического СП можно

вычислить по одной реализации с помощью следующего выражения:

$$B_x(\tau) = \overbrace{x_n(t)x_n(t-\tau)} = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} x_n(t)x_n(t-\tau)dt$$

из которого вытекает схема коррелометра, приведенная на рис. 1.10.

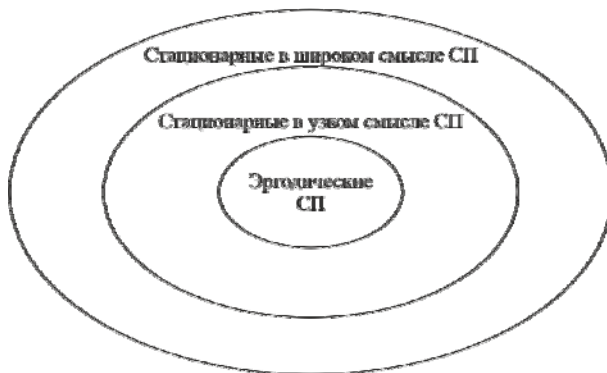


Рис. 1.9. Множества классов случайных процессов

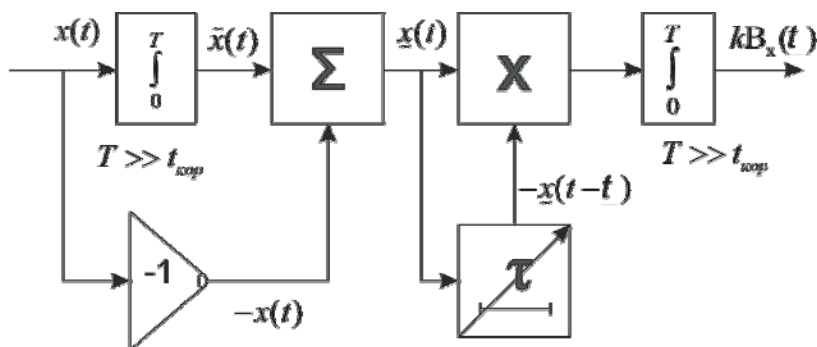


Рис. 1.10. Схема коррелометра

1.4. Спектральный анализ случайных процессов

Спектральный анализ детерминированных сигналов $x(t)$ предполагает использование прямого преобразования Фурье

$$S_x(j\omega) = \int_{-\infty}^{\infty} x(t)e^{-j\omega t} dt.$$

Распространение этого подхода на случайные процессы наталкивается на ряд серьезных проблем.

1. Спектральная функция $S_x(j\omega)$ существует только для функций времени $x(t)$, удовлетворяющих условию абсолютной интегрируемости

$$\int_{-\infty}^{\infty} |x(t)| dt < \infty$$

или хотя бы интегрируемости в квадрате

$$\int_{-\infty}^{\infty} x^2(t) dt < \infty,$$

т. е. для сигналов с ограниченной энергией. Однако реализации ССП с вероятностью 1 имеют бесконечную энергию, так как по определению существуют на бесконечной оси времени и, следовательно, этим требованиям не отвечают. Эту трудность можно обойти, если рассматривать отношение спектральной функции $S_x(j\omega)$ к длительности сигнала T . Тогда достаточным будет требование ограниченной мощности сигнала $x(t)$

$$\lim_{T \rightarrow \infty} \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} x^2(t) dt < \infty.$$

2. Спектральная функция $S_x(j\omega)$ характеризует отдельные реализации $x(t)$ случайного процесса $X(t)$, а не сам процесс целиком. Попытка перейти, как обычно, к усреднению по ансамблю оказывается несостоятельной. Действительно, если определить математическое ожидание случайной спектральной функции $S_X(j\omega)$

$$\overline{S_X(j\omega)} = \overline{S_X(\omega)e^{j\Psi(\omega)}},$$

где $S_X(\omega)$ – амплитудный, а $\Psi(\omega)$ – фазовый спектры случайного процесса $X(t)$, то для независимых $S_X(\omega)$ и $\Psi(\omega)$ при равномерном распределении $\Psi(\omega)$ в интервале $[-\pi, +\pi]$ будет нулевой результат усреднения для ненулевых процессов.

Выход из этой ситуации состоит в отбрасывании фазового и усреднении только амплитудного спектра $S_X(\omega) = |S_X(j\omega)|$ или $|S_X(j\omega)|^2$.

Для реализаций случайных процессов $X(t)$ с ограниченной энергией E_x (нестационарных) по теореме Парсеваля имеем

$$E_x = \int_{-\infty}^{\infty} x^2(t)dt = \int_{-\infty}^{\infty} |S_x(j2\pi f)|^2 df < \infty,$$

где $|S_x(j2\pi f)|^2$ – спектральная плотность энергии реализации $x(t)$.

Усредняя по ансамблю реализаций, получим $\overline{|S_x(j2\pi f)|^2}$ – спектральную плотность энергии случайного процесса $X(t)$ с размерностью $\frac{B^2 \text{ (или } A^2 \text{)}}{\Gamma_{\text{ц}}^2} = \frac{B^2 \text{ (или } A^2 \text{)} \text{с}}{\Gamma_{\text{ц}}}$, что соответствует размерности $\frac{\text{энергия}}{\Gamma_{\text{ц}}}$, если иметь в виду действие $X(t)$ на сопротивлении 1 Ом.

Для ССП на интервале T рассмотрим функцию $G_{X,T}(f) = \frac{|S_{X,T}(j2\pi f)|^2}{T}$, имеющую размерность $\frac{\text{мощность}}{\Gamma_{\text{ц}}}$.

Переходя к пределу при $T \rightarrow \infty$, получим спектральную плотность мощности (СПМ)

$$G_X(f) = \lim_{T \rightarrow \infty} \frac{1}{T} \overline{|S_{X,T}(j2\pi f)|^2}, \quad (1.1)$$

называемую также энергетическим спектром (ЭС) процесса $X(t)$.

Установлено, что СПМ стационарного случайного процесса и его КФ связаны между собой интегральными преобразованиями Фурье. Это было строго доказано советским ученым А. Я. Хинчиным и американцем Н. Винером (теорема Винера-Хинчина)

$$G_X(f) = \int_{-\infty}^{\infty} B_X(\tau) e^{-j2\pi f\tau} d\tau, \quad (1.2)$$

$$B_X(\tau) = \int_{-\infty}^{\infty} G_X(f) e^{j2\pi f\tau} df. \quad (1.3)$$

Свойства энергетических спектров случайных процессов

1. $G_X(f) \geq 0$, что непосредственно следует из его определения (1.1). Из этого факта и соотношения (1.3) вытекает важное следствие для корреляционной функции $B_X(\tau)$ – она является положительно определенной, т. е. имеет неотрицательное преобразование Фурье.

2. $G_X(f)$ – четная функция:

$$\begin{aligned} G_X(f) &= \int_{-\infty}^{\infty} B_X(\tau) e^{-j2\pi f\tau} d\tau = \int_{-\infty}^{\infty} B_X(\tau) [\cos 2\pi f\tau - j\sin 2\pi f\tau] d\tau = \\ &= \int_{-\infty}^{\infty} \underbrace{B_X(\tau) \cos 2\pi f\tau}_{\text{четная функция}} d\tau - j \int_{-\infty}^{\infty} \underbrace{B_X(\tau) \sin 2\pi f\tau}_{\text{нечетная функция}} d\tau = \\ &= 2 \int_0^{\infty} B_X(\tau) \cos 2\pi f\tau d\tau. \end{aligned}$$

На этом свойстве основано понятие одностороннего ЭС, существующего только в области положительных частот

$$G_o(f) = 2G(f), \quad f > 0.$$

$$3. G_X(0) = \int_{-\infty}^{\infty} B_X(\tau) d\tau.$$

$$4. B_X(0) = \int_{-\infty}^{\infty} G_X(f) df = D[X(t)] = P_X.$$

5. Нормированный энергетический спектр

$$\gamma_X(f) = \frac{G_X(f)}{D[X(t)]} = \int_{-\infty}^{\infty} R_X(\tau) e^{-j2\pi f\tau} d\tau,$$

$$\int_{-\infty}^{\infty} \gamma_X(f) df = 1.$$

1.5. Примеры энергетических спектров стационарных случайных процессов

1. Белый шум

Энергетический спектр белого шума (БШ) $N(t)$ равномерен в бесконечной полосе частот ($N = \frac{N_0}{2}$) (рис. 1.11).

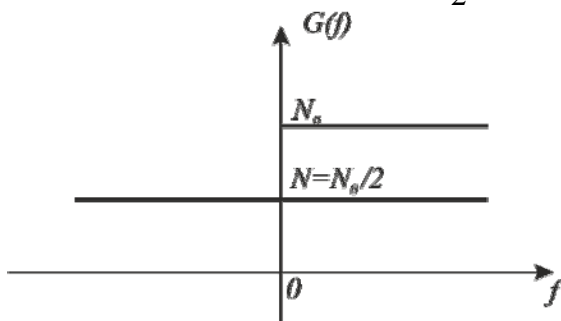


Рис. 1.11. Энергетический спектр белого шума

Корреляционная функция белого шума (рис. 1.12)

$$B_N(\tau) = \int_{-\infty}^{\infty} G_N(f) e^{j2\pi f\tau} df = \int_{-\infty}^{\infty} \frac{N_0}{2} e^{j2\pi f\tau} df = \frac{N_0}{2} \delta(\tau).$$

Здесь использовано одно из определений дельта-функции

$$\delta(\tau) = \int_{-\infty}^{\infty} e^{j2\pi f\tau} df .$$

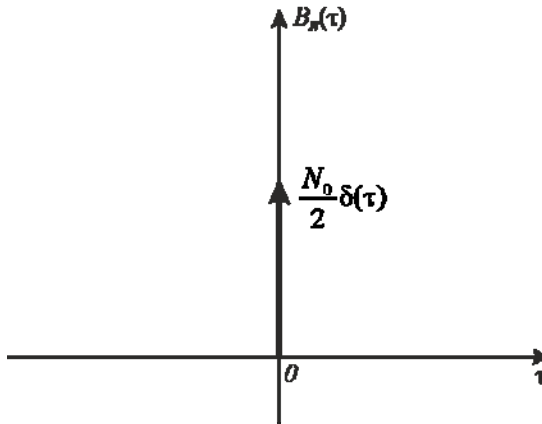


Рис. 1.12. Корреляционная функция белого шума

Из этих результатов вытекает статистическая независимость любых сколь угодно близких сечений такого случайного процесса и его неограниченная дисперсия (мощность)

$$D[N(t)] = \sigma^2 = P_N = N\delta(0) = \frac{N_0}{2} \delta(0) \rightarrow \infty .$$

2. Квазибелый шум

Энергетический спектр квазибелого шума (КБШ) ($N = \frac{N_0}{2}$) равномерен только в ограниченной полосе частот $(-F, +F)$ (рис. 1.13).

Корреляционная функция КБШ имеет вид (рис. 1.14)

$$\begin{aligned}
 B_{N_F}(\tau) &= \int_{-\infty}^{\infty} G_N(f) e^{j2\pi f\tau} df = \int_{-F}^F \frac{N_o}{2} \cos 2\pi f\tau df = \frac{N_o}{2} \frac{\sin 2\pi f\tau}{2\pi\tau} \Big|_{-F}^F = \\
 &= N_o F \frac{\sin 2\pi F\tau}{2\pi F\tau}.
 \end{aligned}$$

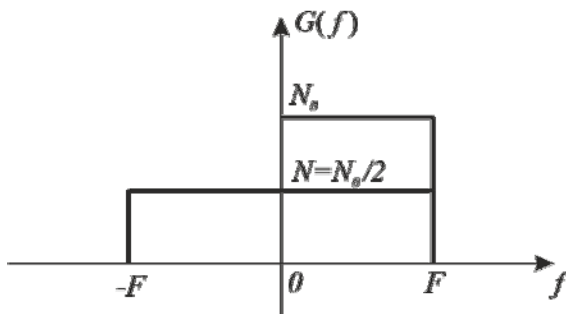


Рис. 1.13. Энергетический спектр квазибелого шума

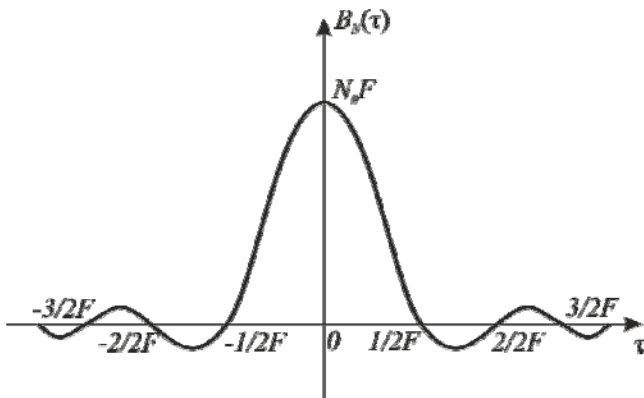


Рис. 1.14. Корреляционная функция квазибелого шума

Из полученного результата вытекает некоррелированность отсчетов КБШ, взятых через интервалы времени $k/2F$. Для нормального СП эти отсчеты оказываются еще и независимыми.

3. Синхронный телеграфный сигнал

Синхронный телеграфный сигнал (СТС) $X(t)$ представляет собой стационарный дискретный СП, принимающий на тактовых интервалах длительностью T значения $+h$ с вероятностью $P(0)$ или $-h$ с вероятностью $P(1)$. Возможная реализация такого процесса показана на рис. 1.15.

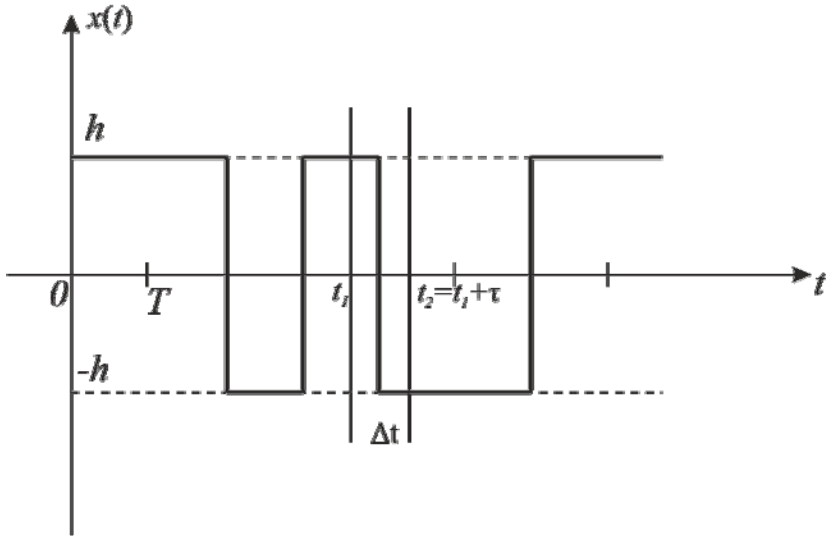


Рис. 1.15. К расчету корреляционной функции телеграфного сигнала

Вычислим корреляционную функцию СТС, исходя из ее определения

$$B_X(t_1, t_2) = \overline{\dot{X}(t_1) \dot{X}(t_2)} = \overline{[X(t_1) - \overline{X(t_1)}][X(t_2) - \overline{X(t_2)}]},$$

где

$$\overline{X(t_1)} = hP(0) + (-h)P(1).$$

В силу стационарности и при $P(0) = P(1) = 0,5$ имеем $\overline{X(t_1)} = \overline{X(t_2)} = 0$ и

$$B_X(t_1, t_2) = B_X(\tau) = \overline{X(t_1)X(t_1 + \tau)}.$$

Далее учтем, что произведение $X(t_1)X(t_1 + \tau) = h^2$, если $\tau < \Delta t$, где Δt – временной интервал от сечения t_1 до ближайшей границы такта (сечения принадлежат одному тактовому интервалу). В противном случае (при $\Delta t < \tau < T$)

$\overline{X(t_1)X(t_1 + \tau)} = h^2 [P(0/0) + P(1/1)] - h^2 [P(0/1) + P(1/0)] = 0$, где $P(0/0)$, $P(0/1)$, $P(1/0)$ и $P(1/1)$ – переходные вероятности передачи символов в соседних тактовых интервалах, которые считаются одинаковыми.

Таким образом,

$$\begin{aligned} B_X(\tau) &= \overline{X(t_1)X(t_1 + \tau)} = h^2 P(T > \Delta t > \tau) = \\ &= h^2 \int_{\tau}^T w(\Delta t) d(\Delta t) = h^2 \left(1 - \frac{\tau}{T} \right), \end{aligned}$$

где $w(\Delta t) = \frac{1}{T}$ – плотность вероятности временного интервала

Δt . Окончательно, учитывая свойство четности корреляционной функции СП, получим

$$B_X(\tau) = h^2 \left(1 - \frac{|\tau|}{T} \right).$$

По полученной КФ можно рассчитать энергетический спектр СТС

$$\begin{aligned} G_X(f) &= \int_{-\infty}^{\infty} B_X(\tau) e^{-j2\pi f\tau} d\tau = 2h^2 \int_0^T \left(1 - \frac{\tau}{T} \right) \cos(2\pi f\tau) d\tau = \\ &= 2h^2 \int_0^T \cos(2\pi f\tau) d\tau - \frac{2h^2}{T} \int_0^T \tau \cos(2\pi f\tau) d\tau = \\ &= 2h^2 \frac{\sin 2\pi fT}{2\pi f} - \frac{2h^2}{T} \underbrace{\left[\frac{\tau \sin(2\pi f\tau)}{2\pi f} \right]_0^T - \int_0^T \frac{\sin(2\pi f\tau)}{2\pi f} d\tau}_{\text{интегрирование по частям}} = \end{aligned}$$

$$= \frac{2h^2}{T} \frac{\cos(2\pi f\tau)}{(2\pi f)^2} \bigg|_0^T = \frac{2h^2}{T(2\pi f)^2} \underbrace{[1 - \cos(2\pi fT)]}_{2\sin^2(\pi fT)} = h^2 T \frac{\sin^2(\pi fT)}{(\pi fT)^2}.$$

Графики корреляционной функции и энергетического спектра СТС приведены на рис. 1.16.

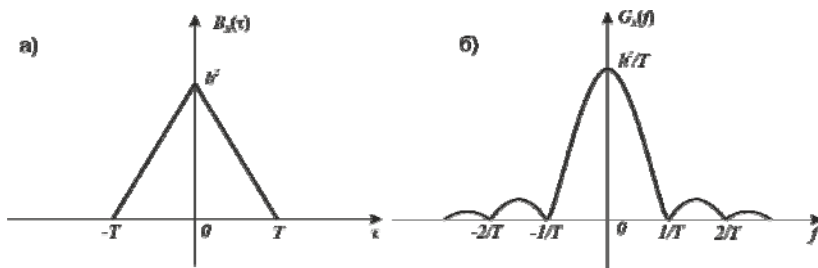


Рис. 1.16. Характеристики синхронного телеграфного сигнала:
а) корреляционная функция; б) энергетический спектр

Примеры

Пример 1

Пусть случайный процесс $X(t)$ задан следующим соотношением:

$$X(t) = A \cos(\omega t + \Theta),$$

где A и ω – константы, а Θ – случайная величина, имеющая равномерное распределение из диапазона $[-\pi, \pi]$. Докажите, что $X(t)$ является стационарным в широком смысле СП.

Решение

Для равномерно распределённой случайной величины плотность вероятности будет иметь вид

$$f_{\Theta}(\Theta) = \begin{cases} \frac{1}{2\pi} & -\pi \leq \Theta \leq \pi, \\ 0 & \text{иначе.} \end{cases}$$

Это означает, что математическое ожидание

$$M[X(t)] = \int_{-\infty}^{\infty} A \cos(\omega t + \Theta) f_{\Theta}(\Theta) d\Theta = \frac{A}{2\pi} \int_{-\pi}^{\pi} \cos(\omega t + \Theta) d\Theta = 0,$$

$$\begin{aligned} D[X(t)] &= \int_{-\infty}^{\infty} (A \cos(\omega t + \Theta) - M[X(t)])^2 w(\Theta) d\Theta = \\ &= \frac{A^2}{2\pi} \int_{-\pi}^{\pi} \cos^2(\omega t + \Theta) d\Theta = \frac{A^2}{2}. \end{aligned}$$

Корреляционная функция будет равна

$$\begin{aligned} B_X(t, t + \tau) &= M[X(t)X(t + \tau)] = \\ &= \frac{A^2}{2\pi} \int_{-\pi}^{\pi} \cos(\omega t + \Theta) \cos[\omega(t + \tau) + \Theta] d\Theta = \\ &= \frac{A^2}{2\pi} \int_{-\pi}^{\pi} \frac{1}{2} [\cos \omega t + \cos(2\omega t + 2\Theta + \omega \tau)] d\Theta = \frac{A^2}{2} \cos \omega \tau. \end{aligned}$$

Так как математическое ожидание случайного процесса является константой, а КФ зависит только от временного сдвига, то можно сделать вывод, что случайный процесс $X(t)$ стационарный в широком смысле.

Примечание. Корреляционная функция $B_X(\tau)$ является периодической с периодом $T_0 = \frac{2\pi}{\omega}$. Стационарный в широком смысле случайный процесс в таком случае называется периодическим.

Пример 2

Доказать, что случайный процесс $X(t)$ из предыдущего примера эргодический одновременно по математическому ожиданию и автокорреляционной функции.

Решение

$$M[X(t)] = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} A \cos(\omega t + \Theta) dt = \frac{A}{T_0} \int_{-\frac{T_0}{2}}^{\frac{T_0}{2}} \cos(\omega t + \Theta) dt = 0,$$

где $T_0 = \frac{2\pi}{\omega}$.

Автокорреляционная функция имеет следующий вид:

$$\begin{aligned} B_X(\tau) &= \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} A^2 \cos(\omega t + \Theta) \cos[\omega(t + \tau) + \Theta] dt = \\ &= \frac{A^2}{T_0} \int_{-\frac{T_0}{2}}^{\frac{T_0}{2}} \frac{1}{2} [\cos \omega \tau + \cos(2\omega t + 2\Theta + \omega \tau)] dt = \frac{A^2}{2} \cos \omega \tau. \end{aligned}$$

Исходя из этого, усреднение по времени дает аналогичный результат, усреднению по ансамблю реализаций СП.

Таким образом, случайный процесс $X(t)$ является эргодическим одновременно по математическому ожиданию и автокорреляционной функции.

Пример 3

Докажите, что спектральная плотность мощности действительного (реального) случайного процесса $X(t)$ является действительной функцией частоты и выполняется условие

$$G_X(-\omega) = G_X(\omega) .$$

Решение

$$\begin{aligned} G_X(\omega) &= \int_{-\infty}^{\infty} B_X(\tau) e^{-j\omega\tau} d\tau = \int_{-\infty}^{\infty} B_X(\tau) (\cos(\omega\tau) - j \sin(\omega\tau)) d\tau = \\ &= \int_{-\infty}^{\infty} B_X(\tau) \cos(\omega\tau) d\tau - j \int_{-\infty}^{\infty} B_X(\tau) \sin(\omega\tau) d\tau. \end{aligned}$$

Из свойства КФ $B_X(-\tau) = B_X(\tau)$ мнимая часть последнего выражения становится равной 0. В итоге для СПМ получается следующее выражение:

$$G_X(\omega) = \int_{-\infty}^{\infty} G_X(\tau) \cos \omega\tau d\tau .$$

Это означает, что $G_X(\omega)$ является действительной функцией частоты.

Кроме того, в силу свойств четности функции $\cos(-\omega\tau) = \cos(\omega\tau)$, получается, что

$$G_X(-\omega) = G_X(\omega) .$$

Таким образом, спектральная плотность мощности действительного случайного процесса $X(t)$ является четной функцией частоты.

Пример 4

Пусть задан модулированный случайный сигнал $Y(t)$

$$Y(t) = AX(t) \cos(\omega_c t + \Theta) ,$$

где $X(t)$ – случайный информационный сигнал, а $A \cos(\omega_c t + \Theta)$ – несущая. Случайный информационный сигнал

$X(t)$ имеет нулевое матожидание, КФ $B_X(\tau)$ и энергетический спектр $G_X(\omega)$.

В несущем сигнале амплитуда A и частота ω_c являются константами, а начальная фаза Θ – случайная величина, равномерно распределенная в диапазоне $[0, 2\pi]$. Предполагая, что $X(t)$ и Θ – независимые величины, найдите матожидание, КФ и энергетический спектр (СПМ) $Y(t)$.

Решение

$$\begin{aligned} M[Y(t)] &= M[AX(t)\cos(\omega_c t + \Theta)] = \\ &= AM[X(t)]M[\cos(\omega_c t + \Theta)] = 0, \end{aligned}$$

т. к. $X(t)$ и Θ – независимы, и $M[X(t)] = 0$.

$$\begin{aligned} B_Y(t, t + \tau) &= M[Y(t)Y(t + \tau)] = \\ &= M[A^2 X(t)X(t + \tau)\cos(\omega_c t + \Theta)\cos[\omega_c(t + \tau) + \Theta]] = \\ &= \frac{A^2}{2} M[X(t)X(t + \tau)]M[\cos \omega_c \tau + \cos(2\omega_c t + \omega_c \tau + 2\Theta)] = \\ &= \frac{A^2}{2} B_X(\tau) \cos \omega_c \tau = B_Y(\tau). \end{aligned}$$

Из того, что матожидание $Y(t)$ является константой и КФ $Y(t)$ зависит только от временного сдвига τ , следует, что $Y(t)$ является стационарной в широком смысле СП.

$$G_Y(\omega) = F[B_Y(\tau)] = \frac{A^2}{2} F[B_X(\tau) \cos \omega_c \tau],$$

$$F[B_X(\tau)] = G_X(\omega),$$

$$F[\cos \omega_c \tau] = \pi \delta(\omega - \omega_c) + \pi \delta(\omega + \omega_c),$$

где $F[\cdot]$ – оператор прямого преобразования Фурье.

Используя соотношение частот и свойство свертки, получается следующее соотношение:

$$\begin{aligned}
 G_Y(\omega) &= \frac{A^2}{4\pi} G_X(\omega) [\pi\delta(\omega - \omega_c) + \pi\delta(\omega + \omega_c)] = \\
 &= \frac{A^2}{4} [G_X(\omega - \omega_c) + G_X(\omega + \omega_c)].
 \end{aligned}$$

Пример 5

Пусть имеется случайный процесс $X(t)$, заданный следующим равенством:

$$X(t) = A \cos \omega t + B \sin \omega t,$$

где ω – константа, а A и B – случайные величины.

1. Докажите, что условие

$$M[A] = M[B] = 0$$

необходимо, чтобы $X(t)$ являлся стационарным в широком смысле случайным процессом.

2. Докажите, что $X(t)$ стационарный в широком смысле СП тогда и только тогда, когда случайные величины A и B не взаимно коррелированы $M[AB] = 0$ и имеют дисперсии, равные $D[A] = D[B] = \sigma^2$.

Решение

1. Математическое ожидание

$$M[X(t)] = M[A] \cos \omega t + M[B] \sin \omega t$$

должно быть независимо от t для того, чтобы $X(t)$ был стационарным в широком смысле случайным процессом. Это возможно только при выполнении условия $M[X(t)] = 0$, а значит,

$$M[A] = M[B] = 0.$$

2. Если $X(t)$ стационарный в широком смысле случайный процесс, значит

$$B\left[X^2(0)\right] = B\left[X^2\left(\frac{\pi}{2\omega}\right)\right] = B_X(0) = \sigma_x^2.$$

Но $X(0) = A$, и $X(\frac{\pi}{2\omega}) = B$, а значит, $D[A] = D[B] = \sigma^2$.

Используя полученный результат, автокорреляционную функцию можно получить в виде

$$\begin{aligned} B_X(t, t + \tau) &= M[X(t)X(t + \tau)] = \\ &= M[(A \cos \omega t + B \sin \omega t)[A \cos \omega(t + \tau) + B \sin \omega(t + \tau)]] = \\ &= \sigma^2 \cos \omega \tau + M[AB] \sin(2\omega t + \omega \tau). \end{aligned}$$

Автокорреляционная функция может быть функцией от τ только, если $M[AB] = 0$, и наоборот: если $M[AB] = 0$ и $D[A] = D[B] = \sigma^2$, тогда $M[X(t)] = 0$ и $B_X(t, t + \tau) = \sigma^2 \cos \omega \tau = B_X(\tau)$, а значит, $X(t)$ стационарный в широком смысле случайный процесс.

Контрольные вопросы

1. Почему для описания сигналов и помех в системах связи используют случайные процессы?
2. Что такое сечение случайного процесса?
3. Каков смысл и размерность одномерной функции распределения СП?
4. Каков смысл и размерность одномерной плотности вероятности СП?
5. Как связаны функция распределения и плотность вероятности СП между собой?
6. Каков смысл n -мерной функции распределения СП?
7. Каков смысл n -мерной плотности вероятности СП?
8. Дайте определение математического ожидания СП
9. Укажите размерность и сущность математического ожидания как математического объекта.
10. Дайте определение дисперсии СП.
11. Укажите размерность и сущность дисперсии как математического объекта.
12. Дайте определение центрированного СП?

13. Дайте определение корреляционной функции (автокорреляционной функции) СП.
14. Какие СП называют стационарными в узком смысле?
15. Какие СП называют стационарными в широком смысле?
16. Какие стационарные СП называют эргодическими?
17. Дайте определение постоянной составляющей СП, укажите ее размерность и сущность как математического объекта.
18. Дайте определение мощности СП, укажите ее размерность и сущность как математического объекта.
19. Какие СП называют нормальными (гуссовскими)?
20. Что понимают под временем корреляции СП?
21. Укажите основные свойства корреляционной функции стационарных СП?
22. Дайте определение спектральной плотности энергии СП и укажите ее размерность.
23. Дайте определение спектральной плотности мощности (энергетическому спектру) СП и укажите ее размерность.
24. Как связаны корреляционная функция и спектральная плотность мощности стационарных СП?
25. Укажите основные свойства энергетического спектра стационарных СП.
26. Какой СП называют белым шумом? Укажите его основные свойства.
27. Какой СП называют квазибелым шумом?
28. Укажите его основные свойства КБШ.
29. Какой СП называют синхронным телеграфным сигналом?
30. Какова корреляционная функция СТС?
31. Как выглядит СПМ синхронного телеграфного сигнала?

Литература

1. Акулиничев, Ю. П. Теория электрической связи : учеб. пособие для вузов / Ю. П. Акулиничев. – СПб. : Лань, 2010. – 233 с.
2. Сальников, А. П. Теория электрической связи : конспект лекций. Ч. 2 / А. П. Сальников. – СПб. : 2002. – 87 с.

3. Теория электрической связи : учебник для вузов / под ред. Д. Д. Кловского. – М. : Радио и связь, 1999. – 432 с.
4. Крук, Б. И. Телекоммуникационные системы и сети : учеб. пособие : в 3 т. Т. 1 : Современные технологии / Б. И. Крук, В. Н. Попантопуло, В. П. Шувалов; под ред. проф. В. П. Шувалова. – М. : Горячая линия – Телеком, 2003. – 647 с.
5. Прокис, Д. Цифровая связь / Д. Прокис ; под ред. Д. Д. Кловского. – М. : Радио и связь, 2000. – 800 с.
6. Основы цифровой обработки сигналов : учеб. пособие / Ю. А. Брюханов, А. Л. Приоров, В. И. Джиган, В. В. Хрящев ; Яросл. гос. ун-т им. П. Г. Демидова. – Ярославль : ЯрГУ, 2013. – 344 с.
7. Гмурман, В. Е. Теория вероятностей и математическая статистика : учеб. пособие для бакалавров / В. Е. Гмурман ; М-во образования и науки РФ – 12-е изд. – М. : Юрайт, 2013. – 479 с.
8. Кремер, Н. Ш. Теория вероятностей и математическая статистика : учебник для вузов. / Н. Ш. Кремер ; М-во образования РФ – 3-е изд., перераб. и доп. – М. : ЮНИТИ, 2009. – 551 с.
9. Теория вероятностей и математическая статистика / сост. Л. П. Бестужева, Н. Л. Майорова; Яросл. гос. ун-т им. П. Г. Демидова, Науч.-метод. совет ун-та. Ч. 1 : практикум. – Б.м. : Б.и., 2012. – 47 с.
10. Hsu, H. Analog and Digital Communications. Schaum's outline series / H. Hsu – McGraw Hill, 2003. – 330 p.
11. Hsu, H. Probability, Random Variables and Random Processes. Schaum's outline series / H. Hsu – McGraw Hill, 1997. – 311 p.

2. ТЕОРИЯ ОЦЕНОК

В данном разделе рассматриваются начальные положения классической теории оценок. Существуют две основных задачи теории оценок: первая заключается в оценивании параметров одной или нескольких случайных величин, вторая – в оценивании значения случайной величины Y , которую невозможно наблюдать, в терминах некоторой другой наблюдаемой случайной величины X .

2.1. Оценка параметра

Пусть X – случайная величина с плотностью вероятности $f(x)$, а X_1, \dots, X_n – набор независимых случайных величин также с плотностями вероятности $f(x)$. Этот набор случайных величин называется случайной выборкой или выборкой размера n случайной величины X . Любая вещественная функция случайной выборки $s(X_1, \dots, X_n)$ называется статистикой.

Пусть X – случайная величина, плотность вероятности $f(\mathbf{x}; \theta)$ которой зависит от некоторого неизвестного параметра θ . Предположим, что X_1, \dots, X_n – выборка случайной величины X . Таким образом, совместная плотность вероятности выборки X_1, \dots, X_n определяется выражением

$$f(\mathbf{x}; \theta) = f(x_1, \dots, x_n; \theta) = \prod_{i=1}^n f(x_i; \theta), \quad (2.1)$$

где x_1, \dots, x_n – набор наблюдаемых значений, полученных из случайной выборки.

Функция оценки параметра θ – это любая статистика $s(X_1, \dots, X_n)$, определяемая как

$$\Theta = s(X_1, \dots, X_n). \quad (2.2)$$

Значение функции $s(X_1, \dots, X_n)$ для частного набора наблюдений $X_1 = x_1, \dots, X_n = x_n$ называется частной

реализацией оценки параметра θ и обозначается $\hat{\theta}$. Оценка является случайной величиной. Наличие одного значения у оцениваемого параметра не является обязательным условием, поэтому оценка, в свою очередь, может иметь интервал значений. Следовательно, оценки, имеющие конкретные значения, называются точечными, а оценки, представляемые интервалом значений, – интервальными.

2.2. Свойства точечных оценок

Несмещенность оценки

Оценка $\Theta = s(X_1, \dots, X_n)$ называется несмещенной оценкой параметра θ , если

$$E(\Theta) = \theta \quad (2.3)$$

для всевозможного набора значений параметра θ . Если Θ является несмещенной оценкой, то ее среднеквадратичная ошибка определяется как

$$E[(\Theta - \theta)^2] = E\{[\Theta - E(\Theta)]^2\} = D(\Theta). \quad (2.4)$$

Из этого выражения следует, что среднеквадратичная ошибка равна дисперсии.

Эффективность оценки

Оценка Θ_1 является более эффективной оценкой параметра θ , чем оценка Θ_2 , если:

1. Для обеих оценок Θ_1 и Θ_2 выполняется условие несмещенности.
2. $\text{Var}(\Theta_1) < \text{Var}(\Theta_2)$.

Оценка $\Theta_{MV} = s(X_1, \dots, X_n)$ называется наиболее эффективной (с минимальной дисперсией) несмещенной оценкой параметра θ , если:

1. Она является несмещенной оценкой параметра θ ;

2. $\text{Var}(\Theta_{MV}) \leq \text{Var}(\Theta)$ для всех Θ .

Состоятельность оценки

Оценка Θ_n параметра θ , полученная на основе случайной выборки размера n , называется состоятельной, если для любого малого значения $\varepsilon > 0$,

$$\lim_{n \rightarrow \infty} P(|\Theta_n - \theta| < \varepsilon) = 1 \quad (2.5)$$

или

$$\lim_{n \rightarrow \infty} P(|\Theta_n - \theta| \geq \varepsilon) = 0. \quad (2.6)$$

Следующие два выражения являются достаточными условиями для определения состоятельности:

$$\lim_{n \rightarrow \infty} E(\Theta_n) = \theta, \quad (2.7)$$

$$\lim_{n \rightarrow \infty} \text{Var}(\Theta_n) = 0. \quad (2.8)$$

2.3. Метод максимального правдоподобия

Пусть $f(\mathbf{x}; \theta) = f(x_1, \dots, x_n; \theta)$ обозначает закон распределения случайной величины X_1, \dots, X_n в дискретном случае и совместную плотность вероятности в непрерывном случае. Пусть

$$L(\theta) = f(\mathbf{x}; \theta) = f(x_1, \dots, x_n; \theta). \quad (2.9)$$

Теперь $L(\theta)$ представляет собой меру правдоподобности того, что значения x_1, \dots, x_n наблюдаемы, когда параметр θ имеет истинное значение. Таким образом, $L(\theta)$ имеет смысл функции правдоподобия (ФП) случайной выборки значений случайной величины. Пусть $\theta_{ML} = s(x_1, \dots, x_n)$ является максимальным значением $L(\theta)$ и вычисляется по формуле:

$$L(\theta_{ML}) = \max_{\theta} L(\theta). \quad (2.10)$$

Следовательно, оценка параметра θ по критерию максимального правдоподобия (МП) определяется как

$$\Theta_{ML} = s(X_1, \dots, X_n), \quad (2.11)$$

где Θ_{ML} – оценка параметра θ по критерию максимального правдоподобия.

Поскольку $L(\theta)$ всегда зависит либо от плотности распределения вероятности, либо от закона распределения дискретной случайной величины, то $L(\theta)$ является всегда положительной для всевозможных значений параметра θ . Таким образом, $\ln L(\theta)$ может быть всегда определенным. Функции $\ln L(\theta)$ и $L(\theta)$ имеют максимум в одной и той же точке. Поэтому $\ln L(\theta)$ можно использовать для решения задачи максимизации с целью упрощения процедуры поиска θ_{ML} .

2.4. Оценка Байеса

Предположим, что неизвестный параметр θ – это некоторая случайная величина, имеющая некоторое фиксированное распределение или априорную плотность вероятности $f(\mathbf{x}; \theta)$. В этом случае $f(\mathbf{x}; \theta)$ является условной плотностью вероятности и обозначается как $f(\mathbf{x}|\theta)$. Следовательно, используя данное обозначение, можно выразить совместную плотность вероятности случайной выборки (X_1, \dots, X_n) и параметра θ как

$$f(x_1, \dots, x_n; \theta) = f(x_1, \dots, x_n | \theta) f(\theta), \quad (2.12)$$

а безусловную плотность вероятности случайной выборки определить следующим выражением

$$f(x_1, \dots, x_n) = \int_{R_\theta} f(x_1, \dots, x_n; \theta) d\theta, \quad (2.13)$$

где R_θ – интервал всевозможных значений параметра θ .

Плотность вероятности, определяемая как

$$f(\theta | x_1, \dots, x_n) = \frac{f(x_1, \dots, x_n; \theta)}{f(x_1, \dots, x_n)} = \frac{f(\theta | x_1, \dots, x_n) f(\theta)}{f(x_1, \dots, x_n)} \quad (2.14),$$

называется апостериорной плотностью вероятности параметра θ . Таким образом, можно сделать вывод, что априорная плотность вероятности $f(\theta)$ дает представления о параметре θ до получения исходов наблюдения X_1, \dots, X_n , в то время как апостериорная плотность вероятности $f(\theta | x_1, \dots, x_n)$ предоставляет информацию о параметре θ после получения случайной выборки.

Математическое ожидание параметра θ в непрерывном случае, определяемое формулой

$$\theta_g = E(\theta | x_1, \dots, x_n) = \int_{R_\theta} \theta f(\theta | x_1, \dots, x_n; \theta) d\theta, \quad (2.15)$$

называется частной реализацией оценки Байеса параметра θ , а

$$\Theta_B = E(\theta | x_1, \dots, x_n) \quad (2.16)$$

называется оценкой Байеса параметра θ .

2.5. Среднеквадратичная оценка

Данный подраздел посвящен рассмотрению второго типа задач оценки – оцениванию значения случайной величины Y , которую невозможно наблюдать, относительно некоторой

другой наблюдаемой случайной величины X . Как правило, оценка \hat{Y} случайной величины Y определяется функцией $g(X)$, зависящей от наблюдаемой случайной величины X . Таким образом, $Y - \hat{Y} = Y - g(X)$ называется ошибкой оценки, а $C[Y - g(X)]$ является стоимостью (стоимостной функцией), связанной с этой ошибкой. Необходимо найти функцию $g(X)$, минимизирующую эту стоимость. В случае если X и Y – непрерывные случайные величины, то часто используемой стоимостной функцией является среднеквадратичная ошибка, определяемая следующим выражением

$$C[Y - g(X)] = E\{[Y - g(X)]^2\}. \quad (2.17)$$

Из (2.17) следует, что оценка случайной величины Y , выражаемая как

$$\hat{Y} = g(X) = E(Y|X), \quad (2.18)$$

является наилучшей в том смысле, что среднеквадратичная ошибка, определяемая выражением (2.17), в данном случае является минимальной.

2.6. Линейная среднеквадратичная оценка

Пусть оценка \hat{Y} случайной величины Y определяется выражением

$$\hat{Y} = g(X) = aX + b. \quad (2.19)$$

Необходимо подобрать такие значения параметров a и b , чтобы среднеквадратичная ошибка, определяемая как

$$e = E\left[(Y - \hat{Y})^2\right] = E\left\{[Y - (aX + b)]^2\right\}, \quad (2.20)$$

имела минимальное значение. При расчете параметров необходимо, чтобы выполнялось условие:

$$E \left\{ \left[Y - (aX + b) \right]^2 \right\} = 0, \quad (2.21)$$

и для a и b выполнялись соотношения

$$a = \frac{\sigma_{XY}}{\sigma_X^2} = \frac{\sigma_Y}{\sigma_X} \rho_{XY}, \quad b = \mu_Y - a\mu_X, \quad (2.22)$$

а минимум среднеквадратичной ошибки e_m определялся выражением

$$e_m = \sigma_Y^2 (1 - \rho_{XY}^2), \quad (2.23)$$

где $\sigma_{XY} = \text{Cov}(X, Y)$, а ρ_{XY} – коэффициент корреляции между случайными величинами X и Y .

Следует отметить, что выражение (2.21) предполагает такую оптимальную линейную среднеквадратичную оценку $\hat{Y} = aX + b$ случайной величины Y , чтобы оценка ошибки $Y - \hat{Y} = Y - (aX + b)$ была ортогональна исходу наблюдения случайной величины X . Данное положение называется условием ортогональности, а линейная функция $y = aX + b$ – линией регрессии.

Теперь предположим, что оценка \hat{Y} случайной величины Y , имеющей линейную комбинацию значений случайной выборки X_1, \dots, X_n , определяется выражением

$$\hat{Y} = \sum_{i=1}^n a_i X_i. \quad (2.24)$$

Так же как и в предыдущем случае, полагается, что линейная оценка имеет минимум среднеквадратичной ошибки, а для параметра a_i должно выполняться условие ортогональности:

$$E \left[\left(Y - \sum_{i=1}^n a_i X_i \right) X_j \right] = 0, \quad j = 1, \dots, n. \quad (2.25)$$

Решая данное выражение относительно параметра a_i , получаем

$$\mathbf{a} = R^{-1}\mathbf{b}, \quad (2.25)$$

где

$$\mathbf{a} = \begin{bmatrix} a_i \\ \vdots \\ a_n \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} b_i \\ \vdots \\ b_n \end{bmatrix}, \quad b_j = E(X Y_j),$$

$$R = \begin{bmatrix} R_{11} & \cdots & R_{1n} \\ \vdots & \ddots & \vdots \\ R_{n1} & \cdots & R_{nn} \end{bmatrix}, \quad R_{ij} = E(X_i X_j),$$

а R^{-1} является транспонированной матрицей R .

Примеры

Пример 1

Пусть дана случайная выборка (X_1, \dots, X_n) случайной величины X , имеющей неизвестное математическое ожидание μ .

Докажите, что оценка μ определяется как

$$M = \frac{1}{n} \sum_{i=1}^n X_i = \bar{X}$$

и является несмещенной оценкой от μ .

Решение

Учитывая, что \bar{X} – выборочное среднее (ансамблю или множеству реализаций),

$$E(M) = E\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n} \sum_{i=1}^n E(X_i) = \frac{1}{n} \sum_{i=1}^n \mu = \frac{1}{n} (n\mu) = \mu.$$

Таким образом, выборочное среднее M равняется несмещенной оценке μ .

Пример 2

Пусть дана случайная выборка (X_1, \dots, X_n) случайной величины X , имеющей неизвестное математическое ожидание μ и дисперсию σ^2 . Докажите, что оценка дисперсии ансамбля реализаций определяется как

$$S^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2,$$

где \bar{X} является выборочным средним, и равна смещенной оценке дисперсии σ^2 .

Решение

$$\sigma^2 = E[(X_i - \mu)^2],$$

$$\begin{aligned} E(S^2) &= E\left[\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2\right] = E\left\{\frac{1}{n} \sum_{i=1}^n [(X_i - \mu) - (\bar{X} - \mu)]^2\right\} = \\ &= E\left\{\frac{1}{n} \sum_{i=1}^n [(X_i - \mu)^2 - 2(X_i - \mu)(\bar{X} - \mu) + (\bar{X} - \mu)^2]\right\} = \\ &= E\left\{\frac{1}{n} \left[\sum_{i=1}^n (X_i - \mu)^2 - n(\bar{X} - \mu)^2\right]\right\} = \\ &= \frac{1}{n} \sum_{i=1}^n E[(X_i - \mu)^2] - E[(\bar{X} - \mu)^2] = \sigma^2 - \sigma_{\bar{X}}^2. \end{aligned}$$

Можно рассчитать, что

$$\begin{aligned} \sigma_{\bar{X}}^2 &= D(\bar{X}) = \sum_{i=1}^n \frac{1}{n^2} \sigma^2 = \frac{1}{n} \sigma^2, \\ E(S^2) &= \sigma^2 - \frac{1}{n} \sigma^2 = \frac{n-1}{n} \sigma^2. \end{aligned}$$

Это означает, что S^2 является смещенной оценкой σ^2 .

Контрольные вопросы

1. Дайте определение точечной и интервальной оценок.
2. Объясните основные отличия и области применения этих оценок.
3. Перечислите основные свойства точечных оценок.
4. Сформулируйте критерий максимального правдоподобия.
5. Приведите примеры ситуаций, в которых применим критерий максимального правдоподобия.
6. Сформулируйте определение байесовской оценки.
7. Приведите примеры ситуаций, в которых применима байесовская оценка.
8. Сформулируйте определение среднеквадратичной оценки.
9. Приведите примеры ситуаций, в которых применима среднеквадратичная оценка.
10. Сформулируйте определение линейной среднеквадратичной оценки.
11. Приведите примеры ситуаций, в которых применима линейная среднеквадратичная оценка.

Литература

1. Акулиничев, Ю. П. Теория электрической связи : учеб. пособие для вузов / Ю. П. Акулиничев. – СПб. : Лань, 2010. – 233 с.
2. Теория электрической связи : учебник для вузов / под ред. Д. Д. Кловского. – М. : Радио и связь, 1999. – 432 с.
3. Крук, Б. И. Телекоммуникационные системы и сети : учеб. пособие : в 3 т. Т. 1 : Современные технологии / Б. И. Крук, В. Н. Попантопуло, В. П. Шувалов; под ред. проф. В. П. Шувалова. – М. : Горячая линия – Телеком, 2003. – 647 с.
4. Прокис, Д. Цифровая связь / Д. Прокис; под ред. Д. Д. Кловского. – М. : Радио и связь, 2000. – 800 с.
5. Основы цифровой обработки сигналов : учеб. пособие / Ю. А. Брюханов, А. Л. Приоров, В. И. Джиган, В. В. Хрящев; Яросл. гос. ун-т им. П. Г. Демидова. – Ярославль : ЯрГУ, 2013. – 344 с.

6. Гмурман, В. Е. Теория вероятностей и математическая статистика : учеб. пособие для бакалавров / В. Е. Гмурман ; М-во образования и науки РФ – 12-е изд. – М. : Юрайт, 2013. – 479 с.
7. Кремер, Н. Ш. Теория вероятностей и математическая статистика : учебник для вузов. / Н. Ш. Кремер ; М-во образования РФ – 3-е изд., перераб. и доп. – М. : ЮНИТИ, 2009. – 551 с.
8. Теория вероятностей и математическая статистика / сост. Л. П. Бестужева, Н. Л. Майорова; Яросл. гос. ун-т им. П. Г. Демидова, Науч.-метод. совет ун-та. Ч. 1 : практикум. – Б.м. : Б.и., 2012. – 47 с.
9. Hsu, H. Probability, Random Variables and Random Processes. Schaum's outline series / H. Hsu – McGraw Hill, 1997. – 311 p.

3. ТЕОРИЯ ПРИНЯТИЯ РЕШЕНИЙ

Существует множество ситуаций, в которых необходимо принимать решение, основываясь на наблюдениях или данных, которые представляют собой случайные переменные. Теория, описывающая подобные ситуации и их решения, называется теорией принятия решений. В связи с радиолокацией она известна также как теория детектирования. В данном разделе рассматриваются начальные положения теории принятия двоичных решений и различные способы проверки гипотез.

3.1. Проверка гипотез

Определения

Статистическая гипотеза – это предположение о законе распределения или вероятности случайной величины. Пусть имеется случайная выборка $(X_1 \dots X_n)$ случайной величины X , чья плотность вероятности $f(\mathbf{x}; \theta) = f(x_1, \dots, x_n; \theta)$ зависит от параметра θ . Требуется проверить, какая из гипотез $\theta = \theta_0$ или $\theta = \theta_1$ является верной. Предположение $\theta = \theta_0$ называется основной (нулевой) гипотезой, а предположение $\theta = \theta_1$ – альтернативной гипотезой:

$$H_0 : \theta = \theta_0,$$

$$H_1 : \theta = \theta_1.$$

Гипотеза называется *простой*, если все параметры определяются точно. Иначе гипотеза называется *сложной*. Таким образом, если $H_0 : \theta = \theta_0$ и $H_1 : \theta \neq \theta_0$, тогда H_0 – простая, а H_1 – составная гипотеза.

Проверка гипотез и типы ошибок

Проверкой гипотезы называется процесс установления ее достоверности. Можно представлять это как процесс разделения пространства наблюдаемых параметров R_n (евклидово n -мерное пространство) на две области R_0 и R_1 . Пусть $\mathbf{x} = (x_1, \dots, x_n)$ – рассматриваемый вектор. Тогда, если $\mathbf{x} \in R_0$, то верной считается гипотеза H_0 , если $\mathbf{x} \in R_1$, то – гипотеза H_1 . Регион R_0

называется *область принятия гипотезы*, R_1 – *область отказа от гипотезы*. Таким образом, можно выделить 4 возможных варианта:

1. H_0 – верна, принята гипотеза H_0 ;
2. H_0 – верна, принята гипотеза H_1 ;
3. H_1 – верна, принята гипотеза H_0 ;
4. H_1 – верна, принята гипотеза H_1 .

Первое и четвертое решения соответствуют правильным решениям, второе и третье – неправильным (ошибочным). Ошибки при принятии решений классифицируются на два рода.

1. *Ошибки первого рода*: отклонение гипотезы H_0 , если она верна (вариант 2).
2. *Ошибки второго рода*: отклонение гипотезы H_1 , если она верна (вариант 3).

Пусть P_1 и P_2 – вероятности ошибок первого и второго рода, тогда можно записать:

$$P_1 = P(D_1|H_0) = P(\mathbf{x} \in R_1; H_0),$$

$$P_2 = P(D_0|H_1) = P(\mathbf{x} \in R_0; H_1),$$

где D_i – решение, принятое в пользу гипотезы H_i . P_1 обычно обозначают α и называют *уровнем значимости критерия*, а P_2 обозначают β , при этом величину $1-\beta$ называют *мощностью критерия*.

Отметим, что α и β представляют вероятности событий одного решения, они не являются независимыми друг от друга и от размера выборки n . Желательно, чтобы в процессе принятия решений оба числа были как можно меньше. Однако зачастую уменьшение одного из этих чисел ведет к увеличению другого при постоянном объеме выборки. Единственным способом уменьшения вероятности ошибок и первого, и второго рода является увеличение объема выборки. Также можно присвоить каждому возможному решению некоторую стоимость и решать задачу по минимизации общей стоимости решения.

Вероятности верных решений можно записать как:

$$P(D_0|H_0) = P(\mathbf{x} \in R_0; H_0),$$

$$P(D_1|H_1) = P(\mathbf{x} \in R_1; H_1).$$

В радиолокации применяются две гипотезы:

1. H_0 – цели нет;

2. H_1 – цель есть.

В этом случае вероятность ошибки первого рода $P_1 = P(D_1|H_0)$ часто называют *вероятностью ложной тревоги*, обозначаемой как P_F . Вероятность ошибки второго рода $P_2 = P(D_0|H_1)$ называют *вероятностью пропуска цели* и обозначают P_M . Вероятность $P(D_1|H_1)$ обозначают P_D и называют *вероятностью обнаружения*. Стоимость ошибки в данном случае обычно вычисляется сложным образом.

3.2. Критерии проверки гипотез

Критерий максимального правдоподобия

Пусть \mathbf{x} – наблюдаемый вектор и $P(\mathbf{x}|H_i)$, $i = 0, 1$ – вероятность того, что при данном \mathbf{x} принятая гипотеза H_i верна. В *критерии максимального правдоподобия* (МП) области принятия решений R_0 и R_1 выбираются следующим образом:

$$R_0 = \{ \mathbf{x} : P(\mathbf{x}|H_0) > P(\mathbf{x}|H_1) \},$$

$$R_1 = \{ \mathbf{x} : P(\mathbf{x}|H_0) < P(\mathbf{x}|H_1) \}.$$

Таким образом, критерий максимального правдоподобия может быть записан как:

$$d(\mathbf{x}) = \begin{cases} H_0, & \text{если } P(\mathbf{x}|H_0) > P(\mathbf{x}|H_1), \\ H_1, & \text{если } P(\mathbf{x}|H_0) < P(\mathbf{x}|H_1). \end{cases}$$

Данное выражение можно переписать в виде:

$$\frac{P(\mathbf{x}|H_1)}{P(\mathbf{x}|H_0)} > 1.$$

Если определить *отношение правдоподобия* $\Lambda(\mathbf{x})$ как

$$\Lambda(x) = \frac{P(\mathbf{x}|H_1)}{P(\mathbf{x}|H_0)},$$

то критерий максимального правдоподобия может быть записан в виде:

$$\Lambda(\mathbf{x}) \underset{H_0}{\overset{H_1}{>}} 1.$$

Единица здесь называется пороговым значением критерия максимального правдоподобия. Часто критерий МП записывают в виде

$$\Lambda(\mathbf{x}) = \frac{f(\mathbf{x}|H_1)}{f(\mathbf{x}|H_0)}.$$

Критерий максимума апостериорной вероятности

Пусть $P(H_i|\mathbf{x})$, $i = 0, 1$ – вероятность того, что H_i верна для конкретного значения \mathbf{x} . Условную вероятность $P(H_i|\mathbf{x})$ называют апостериорной (послеопытной) вероятностью, так как она вычисляется после того как сделано наблюдение. Вероятность $P(H_i)$, $i = 0, 1$ называют априорной (доопытной) вероятностью. В критерии максимума апостериорной вероятности регионы R_0 и R_1 разделены следующим образом:

$$R_0 = \{\mathbf{x} : P(H_0|\mathbf{x}) > P(H_1|\mathbf{x})\},$$

$$R_1 = \{\mathbf{x} : P(H_0|\mathbf{x}) < P(H_1|\mathbf{x})\}.$$

Таким образом, критерий можно переписать так:

$$d(\mathbf{x}) = \begin{cases} H_0, & \text{если } P(H_0|\mathbf{x}) > P(H_1|\mathbf{x}) \\ H_1, & \text{если } P(H_0|\mathbf{x}) < P(H_1|\mathbf{x}) \end{cases}$$

или

$$\frac{P(H_1|\mathbf{x})}{P(H_0|\mathbf{x})} \underset{H_0}{\overset{H_1}{>}} 1.$$

Используя правило Байеса, это выражение перепишем в виде:

$$\frac{P(\mathbf{x}|H_1)P(H_1)}{P(\mathbf{x}|H_0)P(H_0)} > \underset{H_2}{\overset{H_1}{1}}.$$

Используя отношение максимального правдоподобия, запишем

$$\Lambda(\mathbf{x}) \underset{H_0}{\overset{H_1}{>}} \eta = \frac{P(H_1)}{P(H_0)},$$

где $\eta = \frac{P(H_1)}{P(H_0)}$ – пороговое значение в критерии максимума апостериорной вероятности. Если $P(H_1|\mathbf{x}) = P(H_0|\mathbf{x})$, то критерий равносителен критерию максимального правдоподобия.

Критерий Неймана-Пирсона

Как было отмечено ранее, невозможно одновременно уменьшать значения вероятности ошибок первого и второго рода. Критерий Неймана-Пирсона дает работоспособное решение для данной проблемы. В нем минимизируется β при фиксированном значении α . В критерии область R_1 выбирается так, что $1-\beta = 1 - P(D_0|H_1) = P(D_1|H_1)$ является максимальным при фиксированном $\alpha = P(D_1|H_0) = \alpha_0$. Это классическая задача оптимизации с ограничениями, которая может быть решена при помощи метода множителей Лагранжа.

Можно построить функцию

$$J = (1-\beta) - \lambda(\alpha - \alpha_0),$$

где $\lambda \geq 0$ – множитель Лагранжа. Тогда регион R_1 выбирается таким образом, чтобы максимизировать J . Критерий Неймана-Пирсона можно записать в терминах отношения правдоподобия

$$\Lambda(\mathbf{x}) \underset{H_0}{\overset{H_1}{>}} \eta = \lambda,$$

где λ – пороговое значение. Оно равно множителю Лагранжа и выбрано, удовлетворяющим ограничению $\alpha = \alpha_0$.

Критерий Байеса

Пусть C_{ij} – стоимость, соответствующая паре (D_i, H_j) , описывающей событие, при котором выбрана гипотеза D_i , при верной H_j . Тогда средняя стоимость, называемая *байесовским риском*, может быть записана как

$$\begin{aligned} \bar{C} = & C_{00}P(D_0, H_0) + C_{10}P(D_1, H_0) + \\ & + C_{01}P(D_0, H_1) + C_{11}P(D_1, H_1), \end{aligned}$$

где $P(D_i, H_j)$ – вероятность, соответствующая выбору D_i при верном H_j . Используя правило Байеса, имеем

$$\begin{aligned} \bar{C} = & C_{00}P(D_0|H_0)P(H_0) + C_{10}P(D_1|H_0)P(H_0) + \\ & + C_{01}P(D_0|H_1)P(H_1) + C_{11}P(D_1|H_1)P(H_1). \end{aligned}$$

В общем случае предполагается, что $C_{10} > C_{00}$ и $C_{01} > C_{11}$, так как более естественно считать, что стоимость неверного решения больше, чем стоимость верного. Критерий, в котором минимизируется общая цена \bar{C} , называется *критерием Байеса*. Он может быть записан в терминах отношения правдоподобия:

$$\Lambda(\mathbf{x}) \underset{H_0}{\overset{H_1}{>}} \eta = \frac{(C_{10} - C_{00})P(H_0)}{(C_{01} - C_{11})P(H_1)}.$$

Отметим, что если $C_{10} - C_{00} = C_{01} - C_{11}$, то критерии Байеса и максимальной апостериорной вероятности равносильны.

Критерий минимума вероятности ошибки

Если положить $C_{10} = C_{01} = 1$ и $C_{11} = C_{00} = 0$, то получим

$$\bar{C} = P(D_1, H_0) + P(D_0, H_1) = P_e$$

– вероятность принятия неверного решения. Таким образом, из критерия Байеса получаем критерий минимума вероятности ошибки:

$$\Lambda(\mathbf{x}) \underset{H_0}{\overset{H_1}{>}} \eta = \frac{P(H_0)}{P(H_1)}.$$

Как видно, критерий минимума ошибки и критерий максимума апостериорной вероятности идентичны.

Минимаксный критерий

Критерий Байеса требует знания априорных вероятностей $P(H_i)$, $i = 0, 1$. Зачастую эти значения неизвестны. В таких случаях критерий Байеса не может быть применен, вместо него используют минимаксный критерий. В минимаксном критерии используется критерий Байеса с наименее благоприятной априорной вероятности $P(H_0)$. В критерии область R_1^* определяется как

$$\begin{aligned} \max_{P(H_0)} \bar{C}[P(H_0), R_1^*] = \\ = \min_{R_1} \max_{P(H_0)} \bar{C}[P(H_0), R_1] < \max_{P(H_0)} \bar{C}[P(H_0), R_1] \end{aligned}$$

для всех $R_1 \neq R_1^*$. Другими словами, R_1^* – область отказа от гипотезы, соответствующая байесовскому риску при наименее благоприятной априорной вероятности $P(H_0)$. Используя свойство коммутативности операций взятия минимума и максимума, запишем

$$\min_{R_1} \max_{P(H_0)} \bar{C}[P(H_0), R_1] = \max_{P(H_0)} \min_{R_1} \bar{C}[P(H_0), R_1].$$

Минимизация $\bar{C}[P(H_0), R_1]$ с соответствующим R_1 упрощает критерий Байеса, следовательно,

$$\bar{C}[P(H_0), R_1] = C^*[P(H_0)],$$

где $C^*[P(H_0)]$ – минимум байесовского риска, соответствующего априорной вероятности $P(H_0)$. Таким образом, минимаксный критерий вычисляется как байесовский критерий при наихудшей $P(H_0)$ и наибольшей соответствующей ей $\bar{C}[P(H_0)]$.

Примеры

Пример 1

Предположим, что производитель микросхем памяти отмечает, что вероятность отказа микросхемы $p=0.05$. Пусть предложена новая технология с целью улучшения конструкции микросхем. Для ее тестирования было произведено 200 микросхем. Пусть случайная величина X отражает число микросхем, не прошедших тестирование. Введем правило тестирования, состоящее в том, что новая технология может быть принята, если $X \leq 5$.

Пусть

$$H_0 : p = 0.05 \text{ (отказаться от новой технологии),}$$

$$H_1 : p < 0.05 \text{ (принять новую технологию).}$$

Найти вероятность ошибки первого рода.

Решение. Если предполагается, что тесты новой технологии производства независимы и имеют одинаковую вероятность отказа на каждом тесте, тогда X является биномиальной случайной величиной с параметрами $(n, p) = (200, p)$. Ошибка первого рода возникает, если $X \leq 0.5$, тогда как на самом деле $p = 0.05$. С использованием формулы

$$F_x(x) = \sum_{k=0}^n \binom{n}{k} p^k (1-p)^{n-k}, \quad n \leq x < n+1,$$

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

получим

$$P_1 = P(D_1|H_0) = P(X \leq 5; p = 0.05) = \sum_{k=0}^5 \binom{200}{k} (0.05)^k (0.95)^{200-k}.$$

Поскольку n достаточно велико, а p мало, это биномиальное распределение может быть аппроксимировано распределением Пуассона с параметром $\lambda = np = 200(0.05) = 10$.

С использованием формулы

$$p^k (1-p)^{n-k} \approx e^{-\lambda} \frac{\lambda^k}{k!}, \quad np = \lambda$$

получим

$$P_1 \approx \sum_{k=0}^5 e^{-10} \frac{10^k}{k!} = 0.067.$$

Необходимо отметить, что в данном случае H_0 – простая гипотеза, а H_1 – сложная.

Пример 2

Рассмотрим пример 1, но с другими гипотезами:

$H_0 : p = 0.05$ (отказаться от новой технологии),

$H_1 : p = 0.02$ (улучшенная гипотеза).

Решающее правило, при котором новая технология будет отвергнута, $X > 5$. Найти ошибку второго рода.

Решение. Ошибка второго рода возникает, если $X > 5$, тогда как на самом деле $p = 0.02$. С использованием формулы

$$F_x(x) = \sum_{k=0}^n \binom{n}{k} p^k (1-p)^{n-k}, \quad n \leq x < n+1,$$

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

получим

$$P_2 = P(D_0|H_1) = P(X > 5; p = 0.02) =$$

$$= \sum_{k=6}^{\infty} \binom{200}{k} (0.02)^k (0.98)^{200-k}.$$

Используя аппроксимацию P_2 на основе распределения Пуассона с параметром $\lambda = np = 200(0.02) = 4$, получим

$$P_2 \approx 1 - \sum_{k=0}^5 e^{-4} \frac{4^k}{k!} = 0.215.$$

Пример 3

Пусть (X_1, \dots, X_n) является случайной выборкой нормально распределенной случайной величины X со средним значением μ и дисперсией 100. Пусть

$$H_0 : \mu = 50,$$

$$H_1 : \mu = \mu_1 (> 50),$$

и объем выборки $n = 25$. В процедуре принятия решения, используется правило: отклонить H_0 , если $\bar{x} \geq 52$, где \bar{x} – величина выборочного среднего $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$.

Найти:

а) вероятность отклонения гипотезы $H_0 : \mu = 50$ как функцию от $\mu (> 50)$;

б) вероятность ошибки первого рода α ;

в) вероятность ошибки второго рода β , когда $\mu = 52$ и когда $\mu = 55$.

Решение

а) Поскольку отклонение гипотезы $H_0 : \mu = 50$ происходит тогда, когда $\bar{x} \geq 52$, то вероятность отклонения H_0 задается следующим выражением

$$g(\mu) = P(\bar{X} \geq 52; \mu),$$

$$D(\bar{X}) = \sigma_x^2 = \frac{1}{n} \sigma^2 = \frac{100}{25} = 4.$$

Таким образом, \bar{X} является нормально распределенной и имеет математическое ожидание μ и дисперсию 4. Используя формулу

$$F_x(x) = \Phi\left(\frac{x-\mu}{\sigma}\right), \quad \Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-\xi^2/2} d\xi,$$

получаем

$$g(\mu) = P\left(\frac{\bar{X}-\mu}{2} \geq \frac{52-\mu}{2}; \mu\right) = 1 - \Phi\left(\frac{52-\mu}{2}\right), \quad \mu \geq 50.$$

Функция $g(\mu)$ называется функцией мощности критерия, а значение $g(\mu)$ при $\mu = \mu_1$ называется мощностью критерия при μ_1 .

б) Функция мощности критерия при $\mu = 50$ представляет собой вероятность отклонения гипотезы $H_0: \mu = 50$, когда H_0 верна, то есть $g(50)$ является ошибкой первого рода. Таким образом, используя таблицу для функции Φ , получаем

$$\alpha = P_1 = g(50) = 1 - \Phi\left(\frac{52-50}{2}\right) = 1 - \Phi(1) = 0.1587.$$

в) Функция мощности критерия при $\mu = \mu_1$ представляет собой вероятность отклонения гипотезы $H_0: \mu = 50$, когда $\mu = \mu_1$. Таким образом, $1 - g(\mu_1)$ является вероятностью принятия H_0 , когда $\mu = \mu_1$, то есть $1 - g(\mu_1)$ является вероятностью ошибки второго рода β .

Подставляя $\mu = \mu_1 = 53$ в уравнение

$$g(\mu) = P\left(\frac{\bar{X}-\mu}{2} \geq \frac{52-\mu}{2}; \mu\right) = 1 - \Phi\left(\frac{52-\mu}{2}\right)$$

и используя таблицу для функции Φ , получаем

$$\begin{aligned}\beta = P_{11} &= 1 - g(53) = 1 - \Phi\left(\frac{52-53}{2}\right) = \\ &= \Phi\left(-\frac{1}{2}\right) = 1 - \Phi\left(\frac{1}{2}\right) = 0.3085.\end{aligned}$$

Аналогично, при $\mu = \mu_1 = 55$, получаем

$$\begin{aligned}\beta = P_{11} &= 1 - g(55) = 1 - \Phi\left(\frac{52-55}{2}\right) = \\ &= \Phi\left(-\frac{3}{2}\right) = 1 - \Phi\left(\frac{3}{2}\right) = 0.0668.\end{aligned}$$

Необходимо отметить, что вероятность ошибки второго рода зависит от величины μ_1 .

Пример 4

Рассмотрим бинарную задачу принятия решения из примера 3. Изменим решающее правило следующим образом: отклоняем гипотезу H_0 , если $x \geq c$.

а) Найти значения c , при которых вероятность ошибки первого рода $\alpha = 0.05$;

б) Найти вероятность ошибки второго рода β при $\mu = 55$ с учетом модифицированного решающего правила.

Решение

а) Воспользуемся результатом части б) из примера 3. Величина c выбрана такой, что $g(\mu) = P(\bar{X} \geq 52; \mu)$

$$\alpha = g(50) = P(\bar{X} \geq c; \mu = 50) = 0.05.$$

Однако когда $\mu = 50$ и \bar{X} является нормально распределенной с математическим ожиданием 50 и дисперсией 4, получим

$$g(50) = P\left(\frac{\bar{X} - 50}{2} \geq \frac{c - 50}{2}; \mu = 50\right) = 1 - \Phi\left(\frac{c - 50}{2}\right) = 0.05.$$

Из таблицы для функции Φ имеем $\Phi(1.645)=0.95$.
Таким образом,

$$\frac{c-50}{2}=1.645 \text{ и } c=50+2(1.645)=53.29.$$

б) Функция мощности критерия $g(\mu)$ с модифицированным решающим правилом имеет вид

$$\begin{aligned} g(\mu) &= P(\bar{X} \geq 53.29; \mu) = P\left(\frac{\bar{X} - \mu}{2} \geq \frac{53.29 - \mu}{2}; \mu\right) = \\ &= 1 - \Phi\left(\frac{53.29 - \mu}{2}\right). \end{aligned}$$

Полагая $\mu = \mu_1 = 55$ и используя таблицу для функции Φ , получаем

$$\begin{aligned} \beta &= P_2 = 1 - g(55) = 1 - \Phi\left(\frac{53.29 - 55}{2}\right) = \\ &= \Phi(-0.855) = 1 - \Phi(0.855) = 0.1963. \end{aligned}$$

Сравнивая полученный результат с результатами примера 3, необходимо отметить, что с изменением решающего правила, α уменьшается от 0.1587 до 0.05, тогда как β увеличивается от 0.0668 до 0.1963.

Пример 5

Повторить задачу из примера 4 для случая, когда объем выборки $n = 100$.

Решение

а) При $n=100$, имеем

$$D(\bar{X}) = \sigma_x^2 = \frac{1}{n} \sigma^2 = \frac{100}{100} = 1.$$

Как и в части а) примера 4, величина c выбирается так, чтобы

$$\alpha = g(50) = P(\bar{X} \geq c; \mu = 50) = 0.05.$$

Так как \bar{X} является нормально распределенной с математическим ожиданием 50 и дисперсией 1, получим

$$g(50) = P\left(\frac{\bar{X} - 50}{1} \geq \frac{c - 50}{1}; \mu = 50\right) = 1 - \Phi(c - 50) = 0.05.$$

Таким образом,

$$c - 50 = 1.645 \text{ и } c = 51.645.$$

б) Функция мощности критерия

$$\begin{aligned} g(\mu) &= P(\bar{X} \geq 51.645; \mu) = P\left(\frac{\bar{X} - \mu}{1} \geq \frac{51.645 - \mu}{1}; \mu\right) = \\ &= 1 - \Phi(51.645 - \mu). \end{aligned}$$

Полагая $\mu = \mu_1 = 55$ и используя таблицу для функции Φ , получаем

$$\begin{aligned} \beta &= P_2 = 1 - g(55) = 1 - \Phi(51.645 - 55) = \\ &= \Phi(-3.355) \approx 0.0004. \end{aligned}$$

Необходимо отметить, что при объеме выборки $n=100$, α и β уменьшились относительно первоначальных значений 0.1587 и 0.0668, полученных при объеме выборки $n=25$.

Контрольные вопросы

1. Дайте определение статистической гипотезы.
2. Дайте определение ошибок первого и второго рода.
3. Объясните физический смысл ошибок первого и второго рода.
4. Дайте определение критерия максимального правдоподобия.
5. Приведите примеры ситуаций, в которых можно использовать критерий максимального правдоподобия.
6. Дайте определение отношения правдоподобия.
7. Объясните физический смысл отношения правдоподобия.

8. Дайте определение критерия максимума апостериорной вероятности.
9. Объясните физический смысл порогового значения в критерии максимума апостериорной вероятности.
10. Приведите примеры ситуаций, в которых можно использовать критерий максимума апостериорной вероятности.
11. Дайте определение критерия Неймана-Пирсона.
12. Объясните физический смысл порогового значения в критерии Неймана-Пирсона.
13. Приведите примеры ситуаций, в которых можно использовать критерий Неймана-Пирсона.
14. Дайте определение критерия Байеса.
15. Объясните физический смысл порогового значения в критерии Байеса.
16. Объясните физический смысл байесовского риска.
17. Дайте определение критерия минимума вероятности ошибки.
18. Приведите примеры ситуаций, в которых можно использовать критерий минимума вероятности ошибки.
19. Дайте определение минимаксного критерия.
20. Объясните связь минимаксного критерия с критерием Байеса.
21. Приведите примеры ситуаций, в которых можно использовать минимаксный критерий.

Литература

1. Акулиничев, Ю. П. Теория электрической связи : учеб. пособие для вузов / Ю. П. Акулиничев. – СПб. : Лань, 2010. – 233 с.
2. Теория электрической связи: учебник для вузов / под ред. Д. Д. Кловского. – М.: Радио и связь, 1999. – 432 с.
3. Крук, Б. И. Телекоммуникационные системы и сети: учеб. пособие: в 3 т. Т. 1 : Современные технологии / Б. И. Крук, В. Н. Попантопуло, В. П. Шувалов; под ред. проф. В. П. Шувалова. – М. : Горячая линия – Телеком, 2003. – 647 с.
4. Прокис, Д. Цифровая связь / Д. Прокис ; под ред. Д. Д. Кловского. – М. : Радио и связь, 2000. – 800 с.

5. Основы цифровой обработки сигналов : учеб. пособие / Ю. А. Брюханов, А. Л. Приоров, В. И. Джиган, В. В. Хрящев ; Яросл. гос. ун-т им. П. Г. Демидова. – Ярославль : ЯрГУ, 2013. – 344 с.
6. Гмурман, В. Е. Теория вероятностей и математическая статистика : учеб. пособие для бакалавров / В. Е. Гмурман ; М-во образования и науки РФ – 12-е изд. – М. : Юрайт, 2013. – 479 с.
7. Кремер, Н. Ш. Теория вероятностей и математическая статистика : учебник для вузов. / Н. Ш. Кремер ; М-во образования РФ – 3-е изд., перераб. и доп. – М. : ЮНИТИ, 2009. – 551 с.
8. Теория вероятностей и математическая статистика / сост. Л. П. Бестужева, Н. Л. Майорова; Яросл. гос. ун-т им. П. Г. Демидова, Науч.-метод. совет ун-та. Ч. 1 : практикум. – Б.м. : Б.и., 2012. – 47 с.
9. Hsu, H. Probability, Random Variables and Random Processes. Schaum's outline series / H. Hsu – McGraw Hill, 1997. – 311 p.

4. ПРОХОЖДЕНИЕ СЛУЧАЙНЫХ СИГНАЛОВ ЧЕРЕЗ ЭЛЕКТРИЧЕСКИЕ ЦЕПИ

В общем случае решение задачи прохождения заданного СП через конкретную электрическую цепь – функциональный узел (ФУ) – произвольной сложности предполагает определение n -мерной плотности вероятности (или n -мерной функции распределения) реакции цепи $Y(t)$ на заданное случайное воздействие $X(t)$ (рис. 4.1). Однако общего метода решения такой задачи не существует. Поэтому ограничимся рассмотрением некоторых частных случаев.

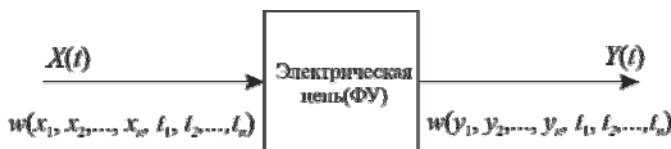


Рис. 4.1. К постановке задачи прохождения случайного процесса через электрическую цепь

4.1. Прохождение случайных сигналов через безынерционные цепи

Безынерционная цепь (безынерционный функциональный узел – БФУ) полностью описывается функциональной зависимостью $y = f(x)$, связывающей мгновенные значения воздействия (входного сигнала цепи) $x(t)$ и реакции (выходного сигнала цепи) $y(t)$ в совпадающие моменты времени. В результате имеем дело с функциональным преобразованием случайного процесса $Y(t) = f[X(t)]$.

Для вычисления одномерной плотности вероятности реакции $w(y)$ по известной плотности вероятности воздействия $w(x)$ рассмотрим рис. 4.2, на котором изображены функциональная характеристика БФУ $y = f(x)$, заданная плотность вероятности воздействия $w(x)$ и искомая плотность вероятности реакции БФУ $w(y)$.

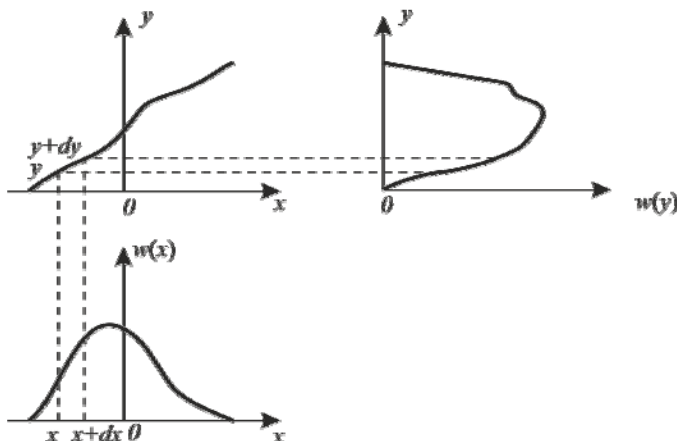


Рис. 4.2. К расчету $w(y)$ на выходе безынерционной цепи

Учитывая, что при попадании случайной величины X в интервал $(x, x+dx)$ случайная величина Y с вероятностью единица попадает в соответствующий ему интервал $(y, y+dy)$, можно написать следующее соотношение

$$P(x \leq X \leq x + dx) = w_X(x) |dx| = P(y \leq Y \leq y + dy) = w_Y(y) |dy|,$$

из которого получаем

$$w_Y(y) = w_X(x) \left| \frac{dx}{dy} \right| = w_X \left[f^{-1}(y) \right] \underbrace{\left| \frac{d[f^{-1}(y)]}{dy} \right|}_{\text{обратная крутизна}}, \quad (4.1)$$

где $f^{-1}(y)$ – обратная функция ($x = x(y) = f^{-1}(y)$).

Дифференциалы dx , dy и производная обратной функции в полученном выражении взяты по модулю в силу свойства положительности плотности вероятности.

Примеры преобразований

1. Линейное безынерционное преобразование

$$y = f(x) = ax + b.$$

Обратная функция

$$x = f^{-1}(y) = \frac{y - b}{a},$$

$$w_Y(y) = w_X\left(\frac{y-b}{a}\right) \frac{1}{|a|}.$$

Таким образом, при линейном преобразовании случайной величины ее кривая плотности распределения смещается на величину b , а масштаб по координатным осям изменяется в $|a|$ раз.

2. Кусочно-линейное преобразование $y = f(x)$

Задачу решим графически, определяя вид кривой $w_Y(y)$ на отдельных интервалах оси y (рис. 4.3).

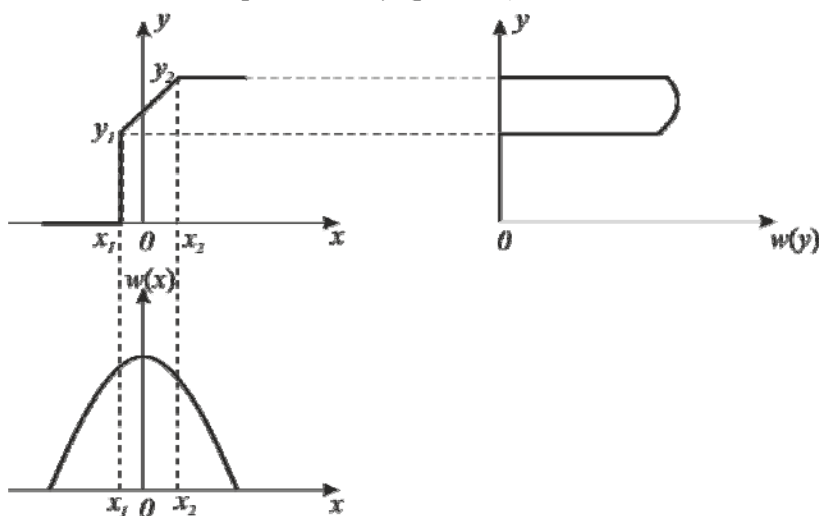


Рис. 4.3. Кусочно-линейное преобразование случайной величины

Из рассмотрения функциональной характеристики $y = f(x)$ вытекает, что

а) при $y < 0$ и $y > y_2$ $w_Y(y) = 0$, т. к. значения реакции y не могут выйти за пределы уровней отсечки ($y = 0$) и насыщения ($y = y_2$);

б) при $0 < y < y_1$ $w_Y(y) = 0$, т. к. в этот интервал (протяженностью y_1) значения реакции попадают при

единственном значении воздействия $x = x_1$, вероятность которого $w_x(x_1)dx \rightarrow 0$;

$$\text{в) при } y_1 \leq y < y_2 \quad w_Y(y) = w_X\left(\frac{y-b}{a}\right) \frac{1}{|a|}, \text{ где } b = y_1,$$

$$a = \frac{y_2 - y_1}{x_2 - x_1} \text{ (см. пример 1);}$$

$$\text{г) при } y = 0 \quad w_Y(y) = \delta(y) \int_{-\infty}^{x_1} w_X(x) dx \rightarrow \infty, \text{ т. к. } y = 0 \text{ для}$$

всех $x < x_1$;

$$\text{д) при } y = y_2 \quad w_Y(y) = \delta(y - y_2) \int_{x_2}^{\infty} w_X(x) dx, \text{ т. к. } y = y_2 \text{ для}$$

всех $x > x_2$.

3. Преобразование при неоднозначной обратной функции $x = f^{-1}(y)$

На практике встречаются ситуации, когда обратная функциональная характеристика является многозначной (рис. 4.4). Рассуждая аналогично тому, как это делали при выводе выражения (4.1), можно убедиться в том, что в этом случае для интервала $y_n < y < y_b$

$$\begin{aligned} w_Y(y) &= \sum_{i=1}^3 w_X[x_i(y)] \left| \frac{d[x_i(y)]}{dy} \right| = \\ &= \sum_{i=1}^3 w_X[f_i^{-1}(y)] \left| \frac{d[f_i^{-1}(y)]}{dy} \right| \end{aligned}$$

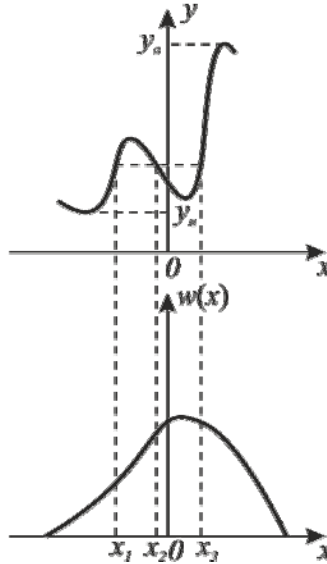


Рис. 4.4. К расчету распределения реакции безынерционной цепи с неоднозначной обратной функциональной характеристикой

Если при анализе прохождения СП через БФУ достаточно знать только основные характеристики распределения реакции, то их можно найти, не определяя $w_Y(y)$. В частности:

математическое ожидание

$$M[Y(t)] = \overline{Y(t)} = \int_{-\infty}^{\infty} y w_Y(y, t) dy = \overline{f[X(t)]} = \int_{-\infty}^{\infty} f(x) w_X(x, t) dx,$$

дисперсия

$$\begin{aligned} D[Y(t)] &= \overline{[Y(t) - \overline{Y(t)}]^2} = \overline{[f[X(t)] - \overline{f[X(t)}]^2} = \\ &= \int_{-\infty}^{\infty} [f(x) - \overline{f[X(t)}]^2 w_X(x, t) dx, \end{aligned}$$

функция корреляции

$$\begin{aligned}
B_Y(t_1, t_2) &= \overline{\dot{Y}(t_1) \dot{Y}(t_2)} = \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [y_1 - \overline{Y(t_1)}][y_2 - \overline{Y(t_2)}] w(y_1, y_2; t_1, t_2) dy_1 dy_2 = \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [f(x_1) - \overline{f[X(t_1)]}] \cdot [f(x_2) - \overline{f[X(t_2)]}] w(x_1, x_2; t_1, t_2) dx_1 dx_2.
\end{aligned}$$

Функциональное преобразование двух случайных процессов

Заданы два случайных процесса $X_1(t)$ и $X_2(t)$ с известной совместной плотностью вероятности их значений в совпадающие моменты времени $w(x_1, x_2; t)$. С этими процессами связаны два других СП $Y_1(t)$ и $Y_2(t)$ с известными функциональными зависимостями

$$y_1 = f_1(x_1, x_2),$$

$$y_2 = f_2(x_1, x_2).$$

Требуется определить $w(y_1, y_2; t)$ – совместную плотность вероятности процессов $Y_1(t)$ и $Y_2(t)$ в совпадающие моменты времени.

Решение

По аналогии с (4.1) можно написать следующее соотношение

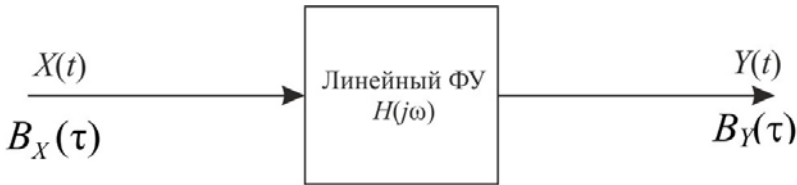
$$w(y_1, y_2; t) = w(x_1, x_2; t) \left| \frac{\partial(x_1, x_2)}{\partial(y_1, y_2)} \right| = w[\phi_1(y_1, y_2), \phi_2(y_1, y_2); t] |J|,$$

где J – якобиан преобразования переменных x_1, x_2 в y_1, y_2

$$J = \frac{\partial(x_1, x_2)}{\partial(y_1, y_2)} = \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} \end{vmatrix}. \quad (4.2)$$

4.2. Прохождение случайных сигналов через линейные цепи

Общей процедуры определения закона распределения реакции линейной цепи на произвольное случайное воздействие также не существует. Однако возможен корреляционный анализ, т. е. расчет корреляционной функции реакции по заданной КФ воздействия, который удобно проводить спектральным методом по схеме, показанной на рис. 4.5.



$$G_X(f) = \int_{-\infty}^{\infty} B_X(\tau) e^{-j2\pi f\tau} d\tau \rightarrow G_Y(f)$$

Рис. 4.5. Схема вычисления корреляционной функции реакции линейной цепи

Для вычисления энергетического спектра $G_Y(f)$ реакции линейного ФУ с передаточной функцией (частотной характеристикой) $H(j\omega)$ воспользуемся его определением (4.1)

$$\begin{aligned} G_Y(f) &= \lim_{T \rightarrow \infty} \frac{1}{T} |S_{Y,T}(jf)|^2 = \lim_{T \rightarrow \infty} \frac{1}{T} S_{Y,T}(jf) S_{Y,T}^*(jf) = \\ &= \lim_{T \rightarrow \infty} \frac{1}{T} S_{X,T}(jf) H(jf) S_{X,T}^*(jf) H^*(jf) = G_X(f) |H(jf)|^2. \end{aligned}$$

Корреляционную функцию $B_Y(\tau)$ определим преобразованием Фурье энергетического спектра $G_Y(f)$

$$B_Y(\tau) = \int_{-\infty}^{\infty} G_Y(f) e^{j2\pi f\tau} df.$$

Вернемся к определению закона распределения реакции линейной цепи в отдельных частных случаях.

Линейное преобразование нормального СП порождает также нормальный СП. Измениться могут только параметры его распределения.

Сумма нормальных СП есть также нормальный СП.

При прохождении СП с произвольным распределением через узкополосный фильтр наблюдается явление нормализации распределения реакции $Y(t)$. Оно заключается в том, что закон распределения реакции приближается к нормальному. Степень этого приближения тем больше, чем сильнее неравенство $\Delta F \ll \Delta f_X$ (т. е. при ширине полосы пропускания фильтра ΔF существенно меньшей ширины энергетического спектра воздействия Δf_X) (рис. 4.6).

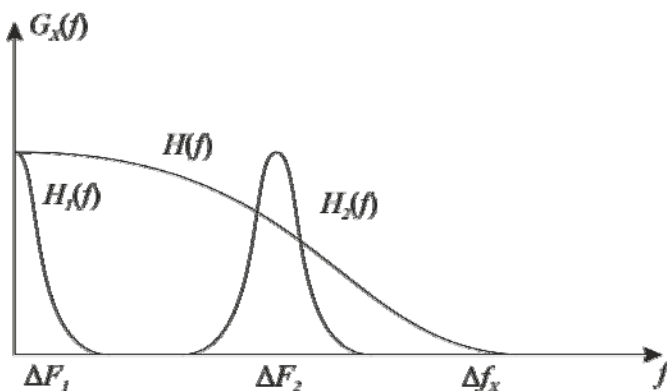


Рис. 4.6. Спектр широкополосного СП и АЧХ узкополосных ФНЧ и ПФ

Объяснить это можно следующим образом. В результате прохождения СП через узкополосный фильтр происходит существенное уменьшение ширины его энергетического спектра (с Δf_X до ΔF) и, соответственно, увеличение времени корреляции (с τ_X до τ_Y). В результате между некоррелированными отсчетами реакции фильтра $Y(k\tau_Y)$ располагается примерно $\Delta f_X / \Delta F$ некоррелированных отсчетов воздействия $X(l\tau_X)$, каждый из которых дает вклад в формирование единственного отсчета реакции с весом, определяемым видом импульсной характеристики фильтра, т. е.

$$Y(k\tau_Y) = \int_0^{k\tau_Y} g(\tau)X(k\tau_Y - \tau)d\tau \cong \sum_{l=1}^{\tau_Y/\tau_X} g(l\tau_X)X(k\tau_Y - l\tau_X).$$

Таким образом, в некоррелированных сечениях $Y(k\tau_Y)$ происходит суммирование большого числа также некоррелированных случайных величин $X(l\tau_X)$ с ограниченными математическими ожиданиями и дисперсиями, что в соответствии с центральной предельной теоремой А. М. Ляпунова обеспечивает приближение распределения их суммы к нормальному с увеличением числа слагаемых.

4.3. Узкополосные случайные процессы

Случайный процесс $X(t)$ с относительно узким энергетическим спектром ($\Delta f_X \ll f_c$), как и узкополосные детерминированные сигналы, удобно представлять в квазигармонической форме

$$X(t) = A(t) \cos \Psi(t) = A(t) \cos [\omega_c t + \phi(t)],$$

где огибающая $A(t)$, фаза $\Psi(t)$ и начальная фаза $\phi(t)$ являются случайными процессами, а ω_c – круговая частота, выбираемая произвольно (обычно как средняя частота его спектра).

Для определения огибающей $A(t)$ и фазы $\Psi(t)$ целесообразно воспользоваться аналитическим СП

$$\dot{X}(t) = X(t) + j\tilde{X}(t) = A(t)e^{j\Psi(t)}.$$

Тогда

$$A(t) = |\dot{X}(t)| = \sqrt{X^2(t) + \tilde{X}^2(t)}, \quad (4.3)$$

$$\Psi(t) = \arctg \frac{\tilde{X}(t)}{X(t)} + \frac{\pi}{2} [1 - \text{sign } X(t)] \text{sign } \tilde{X}(t), \quad (4.4)$$

$$X(t) = \text{Re} [\dot{X}(t)] = A(t) \cos \Psi(t), \quad (4.5)$$

$$\tilde{X}(t) = \text{Im} [\dot{X}(t)] = H[X(t)] = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{X(\tau)}{t - \tau} d\tau = A(t) \sin \Psi(t). \quad (4.6)$$

Основные моментные функции аналитического случайного процесса $\dot{X}(t)$

Математическое ожидание

$$M[\dot{X}(t)] = \overline{\dot{X}(t)} = \overline{X(t)} + j \overline{\tilde{X}(t)}.$$

Дисперсия

$$D[\dot{X}(t)] = M\left[\left|\dot{X}(t) - \overline{\dot{X}(t)}\right|^2\right] = \overline{\left|\dot{X}(t)\right|^2} = \overline{\dot{X}(t) \dot{X}^*(t)}.$$

Корреляционная функция

$$B_{\dot{X}}(t_1, t_2) = \overline{\dot{X}(t_1) \dot{X}^*(t_2)},$$

при этом

$$B_{\dot{X}}(t_1, t_1) = \overline{\dot{X}(t_1) \dot{X}^*(t_1)} = D[\dot{X}(t)],$$

$$B_{\dot{X}}(t_2, t_1) = B_{\dot{X}}^*(t_1, t_2).$$

Аналитический случайный процесс называют стационарным, если

$$M[\dot{X}(t)] = \overline{\dot{X}(t)} = \text{const},$$

$$B_{\dot{X}}(t_1, t_2) = B_{\dot{X}}(\tau),$$

откуда $D[\dot{X}(t)] = B_{\dot{X}}(0) = \text{const}$.

Прохождение нормального случайного процесса через полосовой фильтр, амплитудный и фазовый детекторы

Рассмотрим типичную для систем связи задачу прохождения нормального СП через полосовой фильтр (ПФ), амплитудный (АД) и фазовый (ФД) детекторы (рис. 4.7). Сигнал на выходе ПФ становится узкополосным $X(t) = A(t) \cos[\omega_0 t + \phi(t)]$. Это означает, что его огибающая $A(t)$ и начальная фаза $\phi(t)$ будут медленно меняющимися функциями времени по сравнению с гармонической функцией $\cos \omega_0 t$, где ω_0 – средняя частота полосы пропускания ПФ. По

определению, сигнал на выходе АД будет пропорционален огибающей входного сигнала $A(t)$, а на выходе ФД – его начальной фазе $\varphi(t)$. Таким образом, для решения задачи достаточно вычислить распределение огибающей $A(t)$ и фазы $\Psi(t)$, поскольку распределение начальной фазы $\varphi(t) = \Psi(t) - \omega_0 t$ отличается от распределения $\Psi(t)$ только математическим ожиданием $\omega_0 t$.

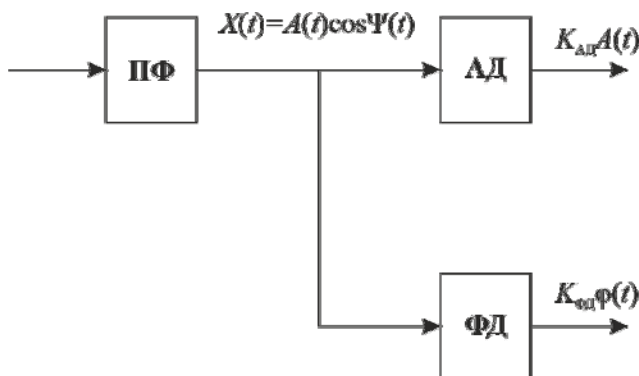


Рис. 4.7. Иллюстрация к задаче прохождения случайного процесса через полосовой фильтр, амплитудный и фазовый детекторы

Таким образом, при решении данной задачи считаются заданными следующие положения.

1. $X(t) = A(t)\cos\Psi(t)$ – узкополосный центрированный стационарный нормальный СП (на выходе ПФ),

2. $w(x) = \frac{1}{\sqrt{2\pi}\sigma_x} e^{-\frac{x^2}{2\sigma_x^2}}.$

Требуется определить следующие характеристики:

1. $w(A)$ – одномерную плотность вероятности огибающей,
2. $w(\Psi)$ – одномерную плотность вероятности фазы.

Решение задачи проведем в три этапа.

1. Переход к аналитическому СП $\dot{X}(t) = X(t) + j\tilde{X}(t) = A(t)e^{j\Psi(t)}$ и определение совместной плотности вероятности $w(x, \tilde{x})$.

2. Расчет совместной плотности вероятности $w(A, \Psi)$ по вычисленной совместной ПВ $w(x, \tilde{x})$ и связям $A(t)$, $\Psi(t)$ с $X(t)$, $\tilde{X}(t)$ (формулы (4.3)–(4.6)).

3. Определение одномерных плотностей вероятности $w(A)$ и $w(\Psi)$ по вычисленной совместной ПВ $w(A, \Psi)$.

Решение

I этап. Найдем одномерную плотность вероятности $w(\tilde{x})$ процесса $\tilde{X}(t)$. На основе линейности преобразования Гильберта $\tilde{X}(t) = H[X(t)]$ делаем вывод о том, что $\tilde{X}(t)$ – нормальный СП. Далее, учитывая, что $G_{\tilde{X}}(f) = G_X(f)$, получаем $B_{\tilde{X}}(\tau) = B_X(\tau)$, а, следовательно,

$$\sigma_{\tilde{X}}^2 = B_{\tilde{X}}(0) = B_X(0) = \sigma_X^2.$$

Таким образом, имеем

$$w(\tilde{x}) = \frac{1}{\sqrt{2\pi\sigma_X}} e^{-\frac{\tilde{x}^2}{2\sigma_X^2}}.$$

Докажем некоррелированность $X(t)$ и $\tilde{X}(t)$ в совпадающие моменты времени, т. е. что $B_{X, \tilde{X}}(0) = 0$:

$$B_{X, \tilde{X}}(0) = \overline{X(t)\tilde{X}(t)} = X(t) \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{X(\tau)}{t - \tau} d\tau.$$

После подстановки $t - \tau = y$, $\tau = t - y$, $d\tau = -dy$, учитывая, что при $\tau \rightarrow \infty$ $y \rightarrow -\infty$, получим

$$B_{X, \tilde{X}}(0) = -X(t) \frac{1}{\pi} \int_{\infty}^{-\infty} \frac{X(t-y)}{y} dy = X(t) \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{X(t-y)}{y} dy =$$

$$\begin{aligned}
&= \frac{1}{\pi} X(t) \left[\int_0^\infty \frac{X(t-y)}{y} dy + \int_{-\infty}^0 \frac{X(t-y)}{y} dy \right] = \\
&= \frac{1}{\pi} X(t) \left[\int_0^\infty \frac{X(t-y)}{y} dy - \int_0^\infty \frac{X(t+y)}{y} dy \right] = \\
&= \frac{1}{\pi} \left[X(t) \int_0^\infty \frac{X(t-y) - X(t+y)}{y} dy \right] = \frac{1}{\pi} \int_0^\infty \frac{\overbrace{B_X(y) - B_X(-y)}^0}{y} dy = 0.
\end{aligned}$$

Как известно, некоррелированность сечений нормальных случайных процессов влечет их независимость, следовательно,

$$w(x, \tilde{x}) = w(x)w(\tilde{x}) = \frac{1}{2\pi\sigma_X^2} e^{-\frac{x^2 + \tilde{x}^2}{2\sigma_X^2}}.$$

II этап. Расчет совместной плотности вероятности

$$w(A, \Psi) = w(x, \tilde{x}) |J|,$$

где, согласно (4.2), (4.5) и (4.6),

$$J = \begin{vmatrix} \frac{\partial x}{\partial A} & \frac{\partial x}{\partial \Psi} \\ \frac{\partial \tilde{x}}{\partial A} & \frac{\partial \tilde{x}}{\partial \Psi} \end{vmatrix} = \begin{vmatrix} \cos \Psi & -A \sin \Psi \\ \sin \Psi & A \cos \Psi \end{vmatrix} = A.$$

Следовательно, с учетом (4.3), имеем

$$w(A, \Psi) = \frac{A}{2\pi\sigma_X^2} e^{-\frac{x^2 + \tilde{x}^2}{2\sigma_X^2}} = \frac{A}{2\pi\sigma_X^2} e^{-\frac{A^2}{2\sigma_X^2}}. \quad (4.7)$$

III этап. Определение одномерных плотностей вероятности

$$w(A) = \int_0^{2\pi} w(A, \Psi) d\Psi = \int_0^{2\pi} \frac{A}{2\pi\sigma_X^2} e^{-\frac{A^2}{2\sigma_X^2}} d\Psi = \frac{A}{\sigma_X^2} e^{-\frac{A^2}{2\sigma_X^2}},$$

$$\begin{aligned}
 w(\Psi) &= \int_0^{\infty} w(A, \Psi) dA = \int_0^{\infty} \frac{A}{2\pi\sigma_X^2} e^{-\frac{A^2}{2\sigma_X^2}} dA = \\
 &= \frac{1}{2\pi} \underbrace{\int_0^{\infty} \frac{A}{\sigma_X^2} e^{-\frac{A^2}{2\sigma_X^2}} dA}_{w(A)} = \frac{1}{2\pi}.
 \end{aligned}$$

Окончательно

$$w(A) = \frac{A}{\sigma_X^2} e^{-\frac{A^2}{2\sigma_X^2}} \quad (A \geq 0), \quad (4.8)$$

$$w(\Psi) = \frac{1}{2\pi} \quad (0 \leq \Psi \leq 2\pi). \quad (4.9)$$

Выражение (4.8) известно как распределение Рэлея (Релея), его график приведен на рис. 4.8. Впервые оно введено в 1880 г. лордом Рэлеем (Джоном Уильямом Стреттом) в связи с задачей сложения гармонических колебаний со случайными фазами.

На рис. 4.9 приведен график равномерного распределения фазы (4.9).

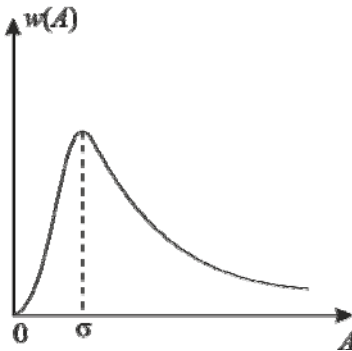


Рис. 4.8. Распределение огибающей $A(t)$

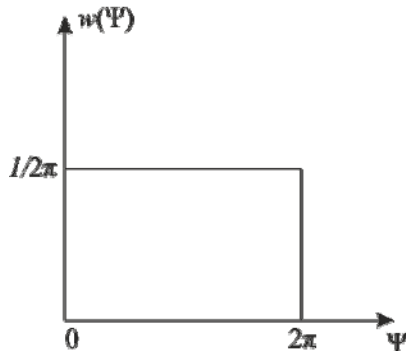


Рис. 4.9. Распределение фазы $\Psi(t)$

Выражение (4.7) можно представить в виде произведения (4.8) и (4.9)

$$w(A, \Psi) = w(A)w(\Psi),$$

из чего следует независимость огибающей $A(t)$ и фазы $w(\Psi)$ нормального СП.

Прохождение аддитивной смеси нормального случайного процесса с гармоническим сигналом через полосовой фильтр, амплитудный и фазовый детекторы

Рассмотрим теперь несколько усложненную задачу прохождения аддитивной смеси рассмотренного выше нормального СП с гармоническим сигналом через амплитудный и фазовый детекторы. Постановка задачи сохраняется прежней за исключением исходного процесса $Y(t)$, который приобретает вид:

$$Y(t) = X(t) + U \cos \omega_c t,$$

где $X(t)$ – центрированный нормальный СП.

Поскольку

$$\overline{Y(t)} = \overline{X(t) + U \cos \omega_c t} = U \cos \omega_c t,$$

то

$$w(y) = \frac{1}{\sqrt{2\pi}\sigma_X} e^{-\frac{(y - U \cos \omega_c t)^2}{2\sigma_X^2}}.$$

Запишем $Y(t)$ в квазигармонической форме

$$Y(t) = A(t) \cos \Psi(t) = A(t) \cos [\omega_c t + \phi(t)]$$

и будем решать задачу определения плотностей вероятности $w(A)$ и $w(\phi)$ по выше приведенному плану.

Предварительно запишем $X(t)$ в квазигармонической форме и через его квадратурные компоненты

$$\begin{aligned} X(t) &= B(t) \cos [\omega_c t + \theta(t)] = \\ &= B(t) \cos \theta(t) \cos \omega_c t - B(t) \sin \theta(t) \sin \omega_c t = \\ &= B_c(t) \cos \omega_c t - B_s(t) \sin \omega_c t. \end{aligned}$$

Тогда

$$\begin{aligned} Y(t) &= X(t) + U \cos \omega_c t = B_c(t) \cos \omega_c t - B_s(t) \sin \omega_c t + U \cos \omega_c t = \\ &= [B_c(t) + U] \cos \omega_c t - B_s(t) \sin \omega_c t = A(t) \cos [\omega_c t + \phi(t)] = \\ &= A(t) \cos \phi(t) \cos \omega_c t - A(t) \sin \phi(t) \sin \omega_c t, \end{aligned}$$

где

$$B_c(t) + U = A(t) \cos \phi(t), \quad A = \sqrt{[B_c(t) + U]^2 + B_s^2(t)},$$

$$B_s(t) = A(t) \sin \phi(t),$$

$$\phi(t) = \operatorname{arctg} \frac{B_s(t)}{B_c(t) + U} + \frac{\pi}{2} [1 - \operatorname{sign}(B_c(t) + U)] \operatorname{sign} B_s(t).$$

Отсюда

$$w(A, \phi) = w(B_c, B_s) |J|, \quad (4.10)$$

$$J = \begin{vmatrix} \frac{\partial [B_c(t) + U]}{\partial A} & \frac{\partial [B_c(t) + U]}{\partial \phi} \\ \frac{\partial [B_s(t)]}{\partial A} & \frac{\partial [B_s(t)]}{\partial \phi} \end{vmatrix} = \begin{vmatrix} \cos \phi & -A \sin \phi \\ \sin \phi & A \cos \phi \end{vmatrix} = A. \quad (4.11)$$

Для нахождения $w(B_c, B_s)$ обратимся к аналитическому случайному процессу

$$\begin{aligned} \dot{X}(t) &= B(t) e^{j[\omega_c t + \theta(t)]} = B(t) e^{j\theta(t)} e^{j\omega_c t} = \\ &= [B(t) \cos \theta(t) + jB(t) \sin \theta(t)] e^{j\omega_c t} = [B_c(t) + jB_s(t)] e^{j\omega_c t}. \end{aligned}$$

Видно, что $B_c(t)$ и $B_s(t)$ являются линейными преобразованиями центрированного нормального СП $X(t)$:

$$\dot{B}(t) = B_c(t) + jB_s(t) = X(t) e^{-j\omega_c t},$$

$$B_c(t) = \operatorname{Re} [X(t) e^{-j\omega_c t}],$$

$$B_s(t) = \operatorname{Im} [X(t) e^{-j\omega_c t}]$$

и, следовательно, имеют нормальное распределение с дисперсиями

$$\sigma_{B_c}^2 = \sigma_{B_s}^2 = \sigma_X^2.$$

Далее можно записать

$$\begin{aligned} w(B_c, B_s) &= w(B_c)w(B_s) = \frac{1}{2\pi\sigma_X^2} e^{-\frac{B_c^2 + B_s^2}{2\sigma_X^2}} = \\ &= \frac{1}{2\pi\sigma_X^2} e^{-\frac{(A\cos\phi - U)^2 + A^2\sin^2\phi}{2\sigma_X^2}} \end{aligned}$$

и с учетом (4.10) и (4.11) получаем

$$w(A, \phi) = \frac{A}{2\pi\sigma_X^2} e^{-\frac{A^2 + U^2 - 2AU\cos\phi}{2\sigma_X^2}}. \quad (4.12)$$

Поскольку выражение (4.12) невозможно представить в виде произведения одномерных функций $w(A, \phi) \neq w(A)w(\phi)$, то можно сделать вывод о зависимости процессов $A(t)$ и $\phi(t)$.

Для нахождения распределения огибающей суммы центрированного нормального СП с гармоническим сигналом проинтегрируем (4.12) по всем возможным значениям случайной фазы $\phi(t)$

$$w(A) = \int_0^{2\pi} w(A, \phi) d\phi = \frac{A}{2\pi\sigma_X^2} e^{-\frac{A^2 + U^2}{2\sigma_X^2}} \int_0^{2\pi} e^{-\frac{AU\cos\phi}{\sigma_X^2}} d\phi.$$

Интеграл вида

$$\frac{1}{2\pi} \int_0^{2\pi} e^{a\cos\phi} d\phi = I_0(a)$$

известен в математике как модифицированная функция Бесселя нулевого порядка. С учетом этого окончательно имеем

$$w(A) = \frac{A}{\sigma_X^2} e^{-\frac{A^2 + U^2}{2\sigma_X^2}} I_0\left(\frac{AU}{\sigma_X^2}\right) \quad (A \geq 0). \quad (4.13)$$

Выражение (4.13) называют распределением Райса или обобщенным распределением Рэлея. Графики этого выражения приведены на рис. 4.10 для следующих частных случаев:

1. $U=0$ $w(A) = \frac{A}{\sigma_X^2} e^{-\frac{A^2}{2\sigma_X^2}}$ – распределение Рэлея,
2. $\sigma_X=0$ $w(A) = \delta(A-U)$ – отсутствие во входной смеси $Y(t)$ случайного процесса $X(t)$,
3. $\sigma_X = U = 5\sigma_X$ – распределение Райса.

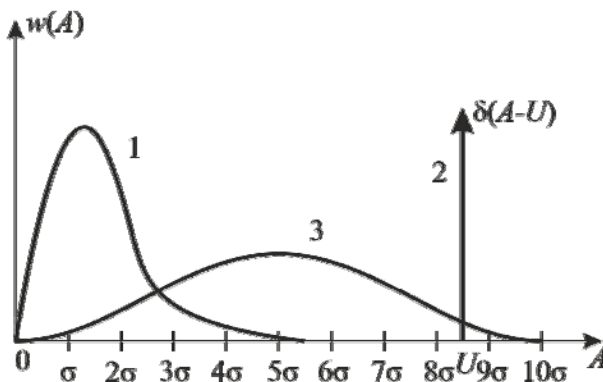


Рис. 4.10. Распределение Райса для трех частных случаев

Из графиков видно, что чем больше отношение сигнал/шум (U / σ_X) , тем правее смещен максимум плотности вероятности и тем симметричнее (ближе к нормальному распределению) кривая $w(A)$.

Таким образом, по рассмотренным задачам можно сделать следующие выводы.

1. Если мгновенные значения централизованного СП $X(t)$ имеют нормальное распределение, то его огибающая $A(t)$ распределена по закону Рэлея

$$w(A) = \frac{A}{\sigma_X^2} e^{-\frac{A^2}{2\sigma_X^2}},$$

а фаза $\Psi(t)$ равномерно

$$w(\Psi) = \frac{1}{2\pi}.$$

2. Распределение огибающей аддитивной смеси центрированного нормального СП и гармонического сигнала подчиняется обобщенному распределению Райса Рэлея (оно же обобщенное распределение Рэлея)

$$w(A) = \frac{A}{\sigma_X^2} e^{-\frac{A^2 + U^2}{2\sigma_X^2}} I_0\left(\frac{AU}{\sigma_X^2}\right) \quad (A \geq 0).$$

Примеры

Пример 1

Стационарный в широком смысле случайный процесс $X(t)$ подан на вход линейной цепи с импульсной характеристикой $h(t) = 3e^{-2t}u(t)$. Найти математическое ожидание на выходе цепи $Y(t)$, если $M[X(t)] = 2$.

Решение. Частотная характеристика линейной цепи есть преобразование Фурье $F[\cdot]$ от ее импульсной характеристики. Она имеет вид

$$H(i\omega) = F[h(t)] = 3 \frac{1}{i\omega + 2}.$$

Математическое ожидание $Y(t)$ можно вычислить следующим образом:

$$M[Y(t)] = M[X(t)]H(0) = 2\left(\frac{3}{2}\right) = 3.$$

Пример 2

Стационарный в широком смысле случайный процесс $X(t)$ с корреляционной функцией

$$B_X(\tau) = Ae^{-a|\tau|},$$

где A и a – реальные положительные константы, подается на вход линейной цепи с импульсной характеристикой

$$h(t) = e^{-bt} u(t),$$

где b – реальная (действительная) положительная константа. Требуется найти корреляционную функцию для $Y(t)$ на выходе системы.

Решение. Сначала найдем частотную характеристику $H(i\omega)$

$$H(i\omega) = F[h(t)] = \frac{1}{i\omega + b}.$$

И амплитудно-частотная характеристика равна

$$|H(i\omega)|^2 = \frac{1}{\omega^2 + b^2}.$$

Тогда спектральная плотность мощности для СП $X(t)$

$$G_X(\omega) = F[B_X(\tau)] = A \frac{2a}{\omega^2 + a^2},$$

а спектральная плотность мощности для случайного процесса $Y(t)$

$$\begin{aligned} G_Y(\omega) &= |H(i\omega)|^2 G_X(\omega) = \left(\frac{1}{\omega^2 + b^2} \right) \left(\frac{2aA}{\omega^2 + a^2} \right) = \\ &= \frac{aA}{(a^2 - b^2)b} \left(\frac{2b}{\omega^2 + b^2} \right) - \frac{A}{a^2 - b^2} \left(\frac{2a}{\omega^2 + a^2} \right). \end{aligned}$$

Выполнив обратное преобразование Фурье для обеих частей этого равенства, получим следующее выражение:

$$B_Y(\tau) = \frac{aA}{(a^2 - b^2)b} e^{-b|\tau|} - \frac{A}{a^2 - b^2} e^{-a|\tau|}.$$

Пример 3

Проверьте, что спектральная плотность мощности любого стационарного в широком смысле случайного процесса $X(t)$ является вещественной функцией и $G_X(\omega) \geq 0$ для всех ω .

Решение. Реальность (вещественность) спектральной плотности мощности $X(t)$ можно рассмотреть на примере идеального полосового фильтра с амлитудно-частотной характеристикой (рис. 4.11)

$$|H(i\omega)| = \begin{cases} 1, & \omega_1 \leq |\omega| \leq \omega_2 \\ 0, & \text{иначе} \end{cases}$$

и со случайным процессом $X(t)$ на входе. Спектральная плотность мощности $G_Y(\omega)$ итогового сигнала на выходе $Y(t)$ будет равна

$$G_Y(\omega) = \begin{cases} G_X, & \omega_1 \leq |\omega| \leq \omega_2 \\ 0, & \text{иначе} \end{cases}.$$

Из этого следует, что

$$E[Y^2(t)] = \frac{1}{2\pi} \int_{-\infty}^{\infty} G_Y(\omega) d\omega = 2 \left(\frac{1}{2\pi} \right) \int_{\omega_1}^{\omega_2} G_X(\omega) d\omega \geq 0.$$

Это означает, что площадь под кривой $G_X(\omega)$ на любом интервале неотрицательна, а это возможно только при $G_X(\omega) \geq 0$ для любой ω .

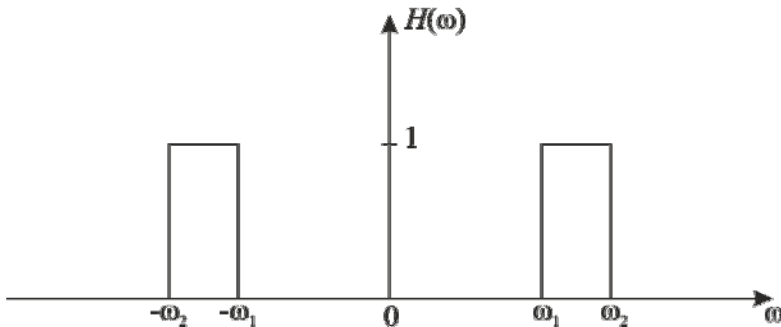


Рис. 4.11. АЧХ идеального полосового фильтра

Пример 4

Пусть имеется стационарный в широком смысле случайный процесс $X(t)$ с корреляционной функцией $B_X(\tau)$ и спектральной плотностью мощности $G_X(\omega)$. Предположим, что

$$X'(t) = \frac{dX(t)}{dt}.$$

Показать что:

$$1. \quad B_{XX'}(\tau) = \frac{dB_{XX}(\tau)}{d\tau},$$

$$2. \quad B_{XX'}(\tau) = -\frac{d^2 B_{XX}(\tau)}{d\tau^2},$$

$$3. \quad G_{XX'}(\omega) = \omega^2 G_{XX}(\omega).$$

Решение. Система с частотной характеристикой $H(i\omega) = i\omega$ называется дифференциатором (рис. 4.12). Так, имея на входе случайный процесс $X(t)$, на выходе будет случайный процесс $Y(t) = X'(t)$.

$$1. \quad G_{XX'}(\omega) = H(i\omega)G_{XX}(\omega) = i\omega G_{XX}(\omega).$$

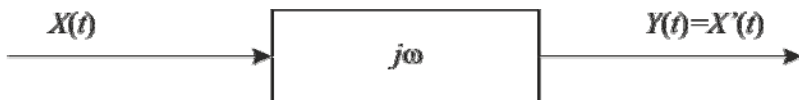


Рис. 4.12. Схема дифференциатора

Вычислив обратное преобразование Фурье,

$$B_{XX'}(\tau) = \frac{dB_{XX}(\tau)}{d\tau}.$$

$$2. \quad G_{XX'}(\omega) = H^*(i\omega)G_{XX'}(\omega) = -i\omega G_{XX'}(\omega)$$

Применив обратное преобразование Фурье к обеим частям равенства, используя ранее полученный результат, получаем что

$$B_{XX'}(\tau) = -\frac{dB_{XX'}(\tau)}{d\tau} = -\frac{d^2 B_{XX}(\tau)}{d\tau^2}.$$

3. И соответственно

$$G_{XX'}(\omega) = |H(i\omega)|^2 G_{XX}(\omega) = |i\omega|^2 G_{XX}(\omega) = \omega^2 G_{XX}(\omega).$$

Пример 5

Пусть на выходе линейной цепи с импульсной характеристикой $h(t)$ имеется случайный процесс $Y(t)$, если на вход подается СП $X(t)$.

Докажите, что:

$$1. \quad B_{XY}(t_1, t_2) = \int_{-\infty}^{\infty} h(\beta) B_{XX}(t_1, t_2 - \beta) d\beta,$$

$$2. \quad B_{YY}(t_1, t_2) = \int_{-\infty}^{\infty} h(\alpha) B_{XY}(t_1 - \alpha, t_2) d\alpha.$$

Решение

$$1. \quad B_{XY}(t_1, t_2) = M[X(t_1)Y(t_2)] = M\left[X(t_1) \int_{-\infty}^{\infty} h(\beta) X(t_2 - \beta) d\beta\right] = \\ = \int_{-\infty}^{\infty} h(\beta) M[X(t_1)X(t_2 - \beta)] d\beta = \int_{-\infty}^{\infty} h(\beta) B_{XX}(t_1, t_2 - \beta) d\beta.$$

2. Аналогично

$$B_{YY}(t_1, t_2) = M[Y(t_1)Y(t_2)] = M\left[\int_{-\infty}^{\infty} h(\alpha) X(t_1 - \alpha) d\alpha Y(t_2)\right] = \\ = \int_{-\infty}^{\infty} h(\alpha) M[X(t_1 - \alpha)Y(t_2)] d\alpha = \int_{-\infty}^{\infty} h(\alpha) B_{XY}(t_1 - \alpha, t_2) d\alpha.$$

Контрольные вопросы

1. Сформулируйте задачу анализа прохождения СП через заданный функциональный узел.
2. Как вычисляют плотность вероятности $w(y)$ реакции безынерционной цепи по известной плотности вероятности $w(x)$ воздействия?
3. Как вычисляют математическое ожидание реакции безынерционной цепи на случайное воздействие $X(t)$?
4. Как вычисляют дисперсию реакции безынерционной цепи на случайное воздействие $X(t)$?
5. Как вычисляют функцию корреляции реакции безынерционной цепи на случайное воздействие $X(t)$?
6. Как вычисляют совместную плотность вероятности $w(y_1, y_2; t)$ двух СП $Y_1(t)$ и $Y_2(t)$, связанных известными функциональными зависимостями $y_1 = f_1(x_1, x_2)$ и $y_2 = f_2(x_1, x_2)$ с двумя другими СП $X_1(t)$ и $X_2(t)$?
7. Как меняется распределение нормального СП при его прохождении через линейную цепь?
8. Как меняется произвольное распределение СП при его прохождении через узкополосный фильтр?
9. В чем суть явления нормализации широкополосного процесса при его прохождении через узкополосный фильтр? Дайте математическое обоснование этому явлению.
10. Опишите процедуру корреляционного анализа прохождения СП через линейную цепь.
11. Дайте определение огибающей и фазы СП.
12. Дайте определения аналитическому СП, его математическому ожиданию, дисперсии и функции корреляции.
13. Каким условиям удовлетворяет стационарный аналитический СП?
14. Каково распределение огибающей центрированного нормального СП?

15. Каково распределение фазы центрированного нормального СП?
16. Каково распределение огибающей суммы центрированного нормального СП и гармонического сигнала?
17. Приведите аналитическое выражение закона Рэлея.
18. Распределение какого СП он характеризует?
19. Приведите аналитическое выражение закона Райса (обобщенного закона Рэлея).
20. Распределение какого СП характеризует закон Райса?

Литература

1. Акулиничев, Ю. П. Теория электрической связи : учеб. пособие для вузов / Ю. П. Акулиничев. – СПб. : Лань, 2010. – 233 с.
2. Сальников, А. П. Теория электрической связи: конспект лекций. Ч. 2 / А. П. Сальников. – СПб. : 2002. – 87 с.
3. Теория электрической связи: учебник для вузов / под ред. Д. Д. Кловского. – М. : Радио и связь, 1999. – 432 с.
4. Крук, Б. И. Телекоммуникационные системы и сети : учеб. пособие : в 3 т. Т. 1 : Современные технологии / Б. И. Крук, В. Н. Попантопуло, В. П. Шувалов ; под ред. проф. В. П. Шувалова. – М. : Горячая линия – Телеком, 2003. – 647 с.
5. Прокис, Д. Цифровая связь / Д. Прокис ; под ред. Д. Д. Кловского. – М.: Радио и связь, 2000. – 800 с.
6. Основы цифровой обработки сигналов : учеб. пособие / Ю. А. Брюханов, А. Л. Приоров, В. И. Джиган, В. В. Хрящев ; Яросл. гос. ун-т им. П. Г. Демидова. – Ярославль : ЯрГУ, 2013. – 344 с.
7. Гмурман, В. Е. Теория вероятностей и математическая статистика : учеб. пособие для бакалавров / В. Е. Гмурман ; М-во образования и науки РФ – 12-е изд. – М. : Юрайт, 2013. – 479 с.
8. Кремер, Н. Ш. Теория вероятностей и математическая статистика : учебник для вузов. / Н. Ш. Кремер ; М-во образования РФ – 3-е изд., перераб. и доп. – М. : ЮНИТИ, 2009. – 551 с.

9. Теория вероятностей и математическая статистика / сост. Л. П. Бестужева, Н. Л. Майорова; Яросл. гос. ун-т им. П. Г. Демидова, Науч.-метод. совет ун-та. Ч. 1 : практикум. – Б.м. : Б.и., 2012. – 47 с.
10. Hsu, H. Analog and Digital Communications. Schaum's outline series / H. Hsu – McGraw Hill, 2003. – 330 p.
11. Hsu, H. Probability, Random Variables and Random Processes. Schaum's outline series / H. Hsu – McGraw Hill, 1997. – 311 p.

5. ОПТИМАЛЬНЫЙ ПРИЕМ

5.1. Постановка задачи при оптимальном приеме дискретных сообщений

При решении данной задачи считаются заданными следующие положения.

1. Источник дискретных сообщений. Это значит, что известен ансамбль передаваемых сообщений

$$B = \{b_i\}_m \quad i = 1, 2, \dots, m,$$

где m – объем алфавита источника, и их статистика (распределение вероятностей) $P(b_i)$.

2. Модулятор. Это значит, что известны правила преобразования каждого сообщения в непрерывный сигнал и длительность сигнала T

$$b_i \rightarrow s_i(t); \quad i = 1, 2, \dots, m; \quad t \in (0, T).$$

3. Непрерывный канал. Канал задается своей математической моделью, описывающей связь его реакции $Z(t)$ с воздействием $s_i(t)$ и канальными помехами $N(t)$, например

$$Z(t) = s_i(t) + N(t); \quad i = 1, 2, \dots, m.$$

4. Тактовая синхронизация осуществляется идеально. Вопросы синхронизации не рассматриваются в рамках курса ОТС, поэтому здесь и в дальнейшем всегда будем считать, что границы между сигналами $s_i(t)$ в приемнике определяются точно, иначе говоря, в нем осуществляется дискретизация времени функцией $\delta(t-kT)$, при которой границы тактов совпадают с границами сигналов.

При решении задачи оптимального приема дискретных сообщений требуется определить правило решения вида

$$z(t) \rightarrow \hat{b}_j \quad t \in T, \quad j = 1, 2, \dots, m,$$

т.е. указать, каким образом на основе анализа принятой реализации $z(t)$ СП $Z(t)$ на каждом интервале T следует принимать решение \hat{b}_j о переданном символе b_i (при $j = i$ имеет место правильный прием, при $j \neq i$ – ошибочный).

Такой постановке задачи можно дать следующую геометрическую трактовку (рис. 5.1).

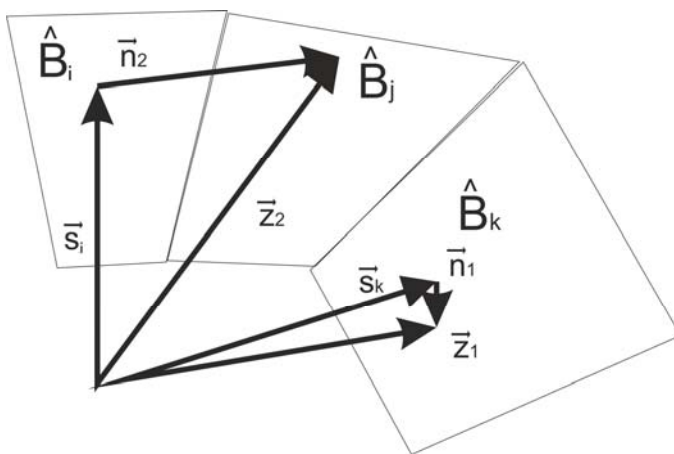


Рис. 5.1. Геометрическая трактовка постановки задачи

Совокупность всех возможных реализаций $z(t)$ образует пространство принимаемых колебаний (обычно это бесконечномерное пространство Гильберта $L_2(T)$), в котором присутствуют m различных векторов \vec{s}_i передаваемых сигналов $s_i(t)$ ($i = 1, 2, \dots, m$). Таким образом, выбор правила решения сводится к разбиению этого пространства на m непересекающихся областей \hat{B}_i , каждая из которых соответствует принятию решения о передаче конкретного сообщения b_i (сигналом $s_i(t)$). На рис. 5.1 показаны две ситуации: 1) конец вектора колебания $z_1(t) = s_k(t) + n_1(t)$ попадает в область \hat{B}_k , отведенную под решение о передаче сообщения b_k сигналом $s_k(t)$, что соответствует правильному приему; 2) конец вектора колебания $z_2(t) = s_i(t) + n_2(t)$ попадает в область \hat{B}_j , отведенную под решение о передаче сообщения b_j сигналом $s_j(t)$, что соответствует ошибочному приему.

Разные правила решения (разные приемные устройства) различаются способом разбиения пространства принимаемых колебаний на области \hat{B}_i . В этой связи возникает задача наилучшего разбиения, которое всегда существует в определенном смысле. Например, если сообщение b_i передается чаще сообщения b_j и важно, чтобы как можно меньше передаваемых символов принимались ошибочно, то следует область \hat{B}_i расширить за счет области \hat{B}_j . Наилучшее разбиение пространства принимаемых сигналов (оптимизация решающей схемы) может быть найдено на основе критерия качества приема, разработка которого требует отдельного рассмотрения на основе теории статистических решений.

В такой постановке задача приема дискретных сообщений в канале с аддитивной, нормальной помехой была решена советским ученым В. А. Котельниковым (1946 г.), который заложил основы теории потенциальной помехоустойчивости. Приемник, реализующий наилучшее разбиение пространства принимаемых сигналов по выбранному критерию качества приема, Котельников назвал идеальным, а достигаемую им помехоустойчивость, при которой обеспечивается максимум средней вероятности правильного приема при заданной модуляции – потенциальной помехоустойчивостью. В дальнейшем такой идеальный приемник будем называть оптимальным демодулятором, как это часто принято в современной теории связи.

Теория потенциальной помехоустойчивости конструктивна, т. к. не только позволяет определить пределы достигаемой помехоустойчивости, но и указывает пути реализации соответствующих демодуляторов.

5.2. Критерии качества приема дискретных сообщений

В разделе, посвященном теории принятия решений, были рассмотрены основные критерии проверки гипотез. В данном подразделе рассматривается применение некоторых из них в качестве критериев качества приема дискретных сообщений, а

также вводятся необходимые для дальнейшего использования обозначения.

5.2.1. Критерий идеального наблюдателя

Критерий идеального наблюдателя (КИН), называемый также критерием Котельникова, требует обеспечения минимума средней вероятности ошибочного приема.

Для двоичной системы

$$\min P = \min [P(0)P(1/0) + P(1)P(0/1)],$$

для m -ичной системы

$$\min P = \min \sum_{i=1}^m P(b_i) \sum_{j=1}^m \underbrace{P(\hat{b}_j / b_i)}_{j \neq i},$$

где $P(\hat{b}_j / b_i)$ – условная вероятность j -й ошибки при передаче i -го сообщения,

$$\sum_{j=1}^m \underbrace{P(\hat{b}_j / b_i)}_{j \neq i} - \text{условная вероятность любой ошибки при}$$

передаче i -го сообщения, P – безусловная вероятность любой ошибки.

Вычислим условную вероятность конкретной ошибки

$$P(\hat{b}_j / b_i) = \int_{\hat{B}_j} w(\bar{\mathbf{z}} / b_i) d\bar{\mathbf{z}},$$

где $w(\bar{\mathbf{z}} / b_i) = w(z_1, z_2, \dots, z_n; t_1, t_2, \dots, t_n / b_i)$ – n -мерная условная плотность вероятности (при разложении $\bar{\mathbf{z}}$ в n -мерном евклидовом пространстве по любому базису), а интеграл, вычисляемый по векторной переменной $\bar{\mathbf{z}}$, n -кратный. Таким образом, критерий Котельникова приобретает вид

$$\min P = \min_{\hat{B}_j} \sum_{i=1}^m P(b_i) \sum_{j=1}^m \underbrace{\int_{\hat{B}_j} w(\bar{\mathbf{z}} / b_i) d\bar{\mathbf{z}}}_{j \neq i}, \quad (5.1)$$

где $\min P$ находится варьированием областей \hat{B}_j .

Минимуму средней вероятности ошибок соответствует максимум средней вероятности правильного приема (иная эквивалентная форма записи КИН)

$$\max (1 - P) = \underbrace{\max_{\hat{B}_i}}_{\hat{B}_i} \sum_{i=1}^m P(b_i) \int_{\hat{B}_i} w(\vec{z} / b_i) d\vec{z}. \quad (5.2)$$

Учитывая, что демодулятор должен реализовать критерий (5.1) или (5.2), принимая решение \hat{b}_j на основе анализа единственной реализации \vec{z} на интервале $0 - T$, рассмотрим апостериорную вероятность вида $P(b_i / \vec{z})$, т. е. вероятность того, что при приеме сигнала $z(t) \Rightarrow \vec{z}$ передавалось сообщение b_i .

Максимум средней вероятности правильного приема будет достигнут, если каждую реализацию принятого колебания $z(t)$ относить к той области \hat{B}_i , для которой апостериорная вероятность $P(b_i / \vec{z})$ максимальна. Таким образом, решение в пользу \hat{b}_i принимается при совместном выполнении совокупности неравенств

$$P(b_i / \vec{z}) > P(b_j / \vec{z}), \quad j = 1, 2, \dots, m \quad j \neq i.$$

Иначе говоря, КИН требует максимизации апостериорной (обратной) вероятности и его можно записать в виде

$$\hat{i} = \arg \left[\underbrace{\max_i}_{\hat{i}} [P(b_i / \vec{z})] \right]. \quad (5.3)$$

Для выполнения анализа (5.3) воспользуемся известной формулой Байеса

$$P(b_i / \vec{z}) w(\vec{z}) = P(b_i) w(\vec{z} / b_i).$$

Тогда

$$\max P(b_i / \vec{z}) = \max P(b_i) \frac{w(\vec{z} / b_i)}{w(\vec{z})},$$

а выражение (5.3) принимает вид

$$\hat{i} = \arg \left\{ \max_i [P(b_i)w(\vec{z} / b_i)] \right\}. \quad (5.4)$$

Важно отметить, что безусловная плотность вероятности $w(\vec{z})$ здесь исключена, т. к. она не зависит от i и поэтому не влияет на решение.

В развернутом виде критерий (5.4) можно записать в виде системы из $m-1$ неравенств

$$P(b_i)w(\vec{z} / b_i) > \underbrace{P(b_j)w(\vec{z} / b_j)}_{j=1,2,\dots,m; j \neq i} \rightarrow \hat{b}_i,$$

или

$$\frac{w(\vec{z} / b_i)}{w(\vec{z} / b_j)} > \frac{P(b_j)}{P(b_i)} \quad (j = 1, 2, \dots, m; j \neq i) \rightarrow \hat{b}_i.$$

Как уже отмечалось в третьем разделе, условную плотность вероятности $w(\vec{z} / b_i)$, рассматриваемую при известном после приема векторе \vec{z} как функцию аргумента b_i , называют функцией правдоподобия гипотезы о передаче сообщения b_i , а $\Lambda_{i,j} = \frac{w(\vec{z} / b_i)}{w(\vec{z} / b_j)}$ – отношением правдоподобия

двух гипотез о передаче сообщений b_i и b_j . С учетом этого КИН можно записать в виде:

$$\text{если } \Lambda_{i,j} > \frac{P(b_j)}{P(b_i)} \quad (j = 1, 2, \dots, m; j \neq i), \text{ то решение } \hat{b}_i. \quad (5.5)$$

Критерий идеального наблюдателя обладает следующими особенностями, которые во многих случаях являются его недостатками.

1. Требуется знания априорных вероятностей отдельных сообщений $P(b_i)$. В реальных системах связи такая информация может отсутствовать.

2. Безразличен к виду ошибок $P(\hat{b}_j / b_i)$. Все виды ошибок при его использовании одинаково нежелательны. Это приводит

к росту ошибок при приеме менее вероятных сообщений, которые несут больше информации.

5.2.2. Критерий максимального правдоподобия

Данный критерий был описан в третьем разделе. Здесь рассмотрим его несколько подробнее и введем необходимые для дальнейшего использования обозначения.

Будем полагать, что все передаваемые сообщения равновероятны

$$P(b_1) = P(b_2) = \dots = P(b_m) = \frac{1}{m}.$$

Тогда из (5.5) получим:

если $\Lambda_{i,j} > 1$ ($j = 1, 2, \dots, m; j \neq i$), то решение \hat{b}_i .

Во многих случаях удобно, кроме гипотез о передаче сообщений b_i ($i = 1, 2, \dots, m$), ввести еще одну «нулевую» гипотезу о том, что никакое сообщение не передавалось. В этом случае принятое колебание является реализацией только помехи $z(t) = n(t)$ ($\vec{z} = \vec{n}$).

Обозначим отношение правдоподобия

$$\frac{w(\vec{z} / b_i)}{w(\vec{z} / 0)} = \Lambda_i.$$

В этом случае правило решения (5.5) можно записать в следующем виде:

если $\Lambda_i > \Lambda_j$, при всех $j \neq i$, то решение \hat{b}_i

или

$$\hat{i} = \arg \left[\underbrace{\max}_i (\Lambda_i) \right]. \quad (5.6)$$

Далее будем использовать критерием максимального правдоподобия в таком виде. Отметим, что он совпадает с КИН при равных вероятностях передаваемых сообщений.

5.2.3. Критерий минимального среднего риска (критерий Байеса)

Критерий идеального наблюдателя нужно обобщить для учета разных последствий ошибок передачи сообщений, минимизируя сумму условных вероятностей $P(\hat{b}_j / b_i)$ в его выражении (5.1) с заранее назначенными весами (ценой, платой) L_{ij} . Средневзвешенная сумма условных вероятностей $P(\hat{b}_j / b_i)$ при передаче сообщения b_i (обычно называемая условным риском) имеет вид

$$R_i = \sum_{j=1}^m L_{i,j} P(\hat{b}_j / b_i),$$

а сам критерий требует

$$\min R = \min \sum_{i=1}^m P(b_i) \underbrace{\sum_{j=1}^m L_{i,j} P(\hat{b}_j / b_i)}_{L_{i,j}=0}, \quad (5.7)$$

где R – средний риск.

При использовании этого критерия оптимальной считается решающая схема, обеспечивающая наименьшее значение среднего риска (5.7).

Из критерия минимального среднего риска (КМСП), как наиболее общего, вытекают оба рассмотренных выше критерия:

критерий ИН при $L_{i,i} = 0$ и $L_{i,j} = \text{const}$ ($i \neq j$),

критерий максимального правдоподобия при $L_{i,i} = 0$ и

$$L_{i,j} = L_i = \frac{1}{P(b_i)}.$$

Использовать критерий Байеса на практике часто мешают необходимость знать вероятности передачи сообщений $P(b_i)$ и сложность объективного определения весовых коэффициентов $L_{i,j}$.

5.2.4. Критерий Неймана-Пирсона

Данный критерий, также рассмотренный в третьем разделе, обычно применяется в двоичных системах в ситуациях,

когда неизвестны априорные вероятности отдельных сообщений, а последствия ошибок разного рода неодинаковы. Особенно характерна такая ситуация для радиолокации. Там производится зондирование пространства узким радиолучом и прием отраженного от цели сигнала. При этом имеют место две стандартные ситуации: 1) наличие цели – колебание на входе приемника содержит сигнал в аддитивной смеси с помехой (с неизвестной априорной вероятностью $P(b_1)$); 2) отсутствие цели – на входе приемника действует одна помеха (с вероятностью $P(b_0) = 1 - P(b_1)$).

Задача приема состоит в обнаружении сигнала на фоне помех. При ее реализации возможны два вида ошибок.

1. Ложная тревога с условной вероятностью $P(\hat{b}_1 / b_0)$. В этом случае цель отсутствует, но принято решение о наличии отраженного от нее сигнала.
2. Пропуск цели с условной вероятностью $P(\hat{b}_0 / b_1)$. Здесь цель имеется, но отраженный от нее сигнал не обнаружен.

В реальных ситуациях последствия этих ошибок могут сильно различаться. Тогда целесообразно стремиться к уменьшению условной вероятности ошибки, вызывающей особо тяжелые последствия. В данной ситуации это пропуск цели. Такое уменьшение можно сделать только за счет увеличения вероятности ошибки другого вида (ложной тревоги). Это можно делать до определенной степени, т. к. слишком большая вероятность ложной тревоги приведет к ощутимым экономическим потерям и к подрыву доверия к системе в целом.

Стандартный подход заключается в том, чтобы зафиксировать вероятность ложной тревоги на выбранном уровне ε

$$P(b_1 / b_0) = \int_{\hat{B}_1} w(\vec{z} / b_0) d\vec{z} = \varepsilon, \quad (5.8)$$

а затем минимизировать вероятность пропуска цели

$$\min P(b_0 / b_1) = \min \int_{\hat{B}_0} w(\vec{z} / b_1) d\vec{z} = \min \left[1 - \int_{\hat{B}_1} w(\vec{z} / b_1) d\vec{z} \right]. \quad (5.9)$$

Минимизация (5.9) при заданной величине (5.8) достигается, если решение о наличии цели принимается при выполнении неравенства

$$\Lambda_1 = \frac{w(\bar{\mathbf{z}} / b_1)}{w(\bar{\mathbf{z}} / 0)} > \lambda(\varepsilon),$$

где $\lambda(\varepsilon)$ – порог, определяемый заданной вероятностью ЛТ.

5.3. Синтез оптимального демодулятора при известном ансамбле сигналов

5.3.1. Задача когерентного приема

Синтез оптимального демодулятора при известном ансамбле сигналов называется также когерентным приемом. При решении данной задачи считаются заданными следующие положения.

1. Ансамбль сигналов на выходе модулятора

$$\{s_i(t)\}; i = 1, 2, \dots, m; t \in (0, T).$$

2. Непрерывный канал

$$Z(t) = s_i(t) + N(t); \quad i = 1, 2, \dots, m,$$

где $N(t)$ – квазибелый нормальный шум, энергетический спектр которого равен

$$G_N(f) = \begin{cases} N_0 = \text{const} & f \in (0, F) \\ 0 & f \notin (0, F) \end{cases}.$$

3. В качестве критерия качества приема задан критерий максимального правдоподобия (5.6)

$$\hat{i} = \arg \left\{ \max_i [\Lambda_i] \right\} = \arg \left\{ \max_i \left[\frac{w(\bar{\mathbf{z}} / b_i)}{w(\bar{\mathbf{z}} / 0)} \right] \right\}.$$

В задаче требуется синтезировать оптимальный демодулятор, т. е. найти алгоритм оптимальной обработки входного сигнала и принятия решения о передаваемом сообщении.

В основу решения положим выражение заданного критерия качества приема, для чего рассмотрим входящие в него функции правдоподобия гипотез:

- 1) о наличии во входном колебании $z(t)$ i -го сигнала $[z(t) = s_i(t) + n(t)]$

$$w(\vec{z} / b_i) = w(z_1, z_2, \dots, z_n; t_1, t_2, \dots, t_n / b_i),$$

- 2) об отсутствии в нем какого-либо сигнала $[z(t) = n(t)]$

$$w(\vec{z} / 0) = w(z_1, z_2, \dots, z_n; t_1, t_2, \dots, t_n / 0),$$

где $n = \frac{T}{\Delta t}$, $\Delta t = \frac{1}{2F}$, $n = 2FT$.

Начнем с последней. Учитывая, что сечения квазибелого шума, разделенные интервалами $\Delta t = \frac{1}{2F}$, не коррелированы, а в силу нормального распределения шума и независимы, получим

$$w(\vec{z} / 0) = \prod_{k=1}^n w(z_k / 0) = \frac{1}{(\sqrt{2\pi}\sigma)^n} \exp\left(-\frac{1}{2\sigma^2} \sum_{k=1}^n z_k^2\right).$$

Поскольку СП $Z(t) = s_i(t) + N(t)$ отличается от шума $N(t)$ только известным, а потому неслучайным сигналом $s_i(t)$, играющим роль математического ожидания $Z(t)$, то

$$w(\vec{z} / b_i) = \frac{1}{(\sqrt{2\pi}\sigma)^n} \exp\left[-\frac{1}{2\sigma^2} \sum_{k=1}^n (z_k - s_{i,k})^2\right],$$

где использовано обозначение $s_{i,k} = s_i(t_k)$.

В итоге отношение правдоподобия гипотез о наличии и отсутствии сигнала принимает вид

$$\Lambda_i = \frac{w(\vec{z} / b_i)}{w(\vec{z} / 0)} = \exp\left[\frac{1}{2\sigma^2} \sum_{k=1}^n z_k^2 - \frac{1}{2\sigma^2} \sum_{k=1}^n (z_k - s_{i,k})^2\right],$$

или с учетом того, что $2\sigma^2 = 2N_0 F = \frac{N_0}{\Delta t}$,

$$\Lambda_i = \exp\left[\frac{1}{N_0} \sum_{k=1}^n z_k^2 \Delta t - \frac{1}{N_0} \sum_{k=1}^n (z_k - s_{i,k})^2 \Delta t\right].$$

Перейдем к белому шуму, сняв ограничение на ширину его спектра ($F \rightarrow \infty$). Иначе говоря, от евклидова пространства перейдем к гильбертовому. При этом

$$\Delta t = \frac{1}{2F} \rightarrow dt, \quad s_{i,k} = s_i(t_k) \rightarrow s_i(t),$$

$$z_k = z(t_k) \rightarrow z(t), \quad \sum \rightarrow \int$$

и

$$\Lambda_i = \exp \left\{ \frac{1}{N_0} \int_0^T z^2(t) dt - \frac{1}{N_0} \int_0^T [z(t) - s_i(t)]^2 dt \right\}. \quad (5.10)$$

Синтезируемый демодулятор должен принимать решение в пользу \hat{b}_i , обеспечивающего максимум выражения (5.10), или, что эквивалентно, максимум показателя экспоненты в нем

$$\underbrace{\max_i}_{i} \Lambda_i \rightarrow \underbrace{\max_i}_{i} \left\{ \int_0^T z^2(t) dt - \int_0^T [z(t) - s_i(t)]^2 dt \right\} \rightarrow \hat{b}_i. \quad (5.11)$$

Видно, что максимум (5.11) достигается при минимуме вычитаемого

$$\underbrace{\max_i}_{i} \Lambda_i \rightarrow \underbrace{\min_i}_{i} \left\{ \int_0^T [z(t) - s_i(t)]^2 dt \right\} = \underbrace{\min_i}_{i} d^2(\vec{z}, \vec{s}_i) \rightarrow \hat{b}_i. \quad (5.12)$$

Демодулятор оптимальный по критерию максимального правдоподобия принимает решение в пользу того символа \hat{b}_i , сигнал $s_i(t)$ которого отстоит от принятого колебания $z(t)$ на меньшее расстояние.

Рассматривая выражение (5.12) как алгоритм обработки принятого колебания $z(t)$, приходим к схеме демодулятора, представленной на рис. 5.2.

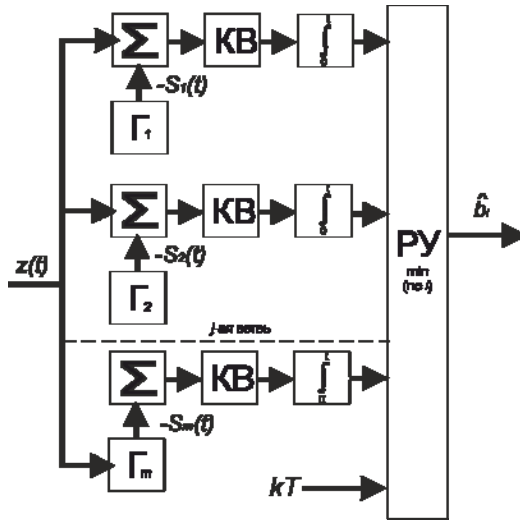


Рис. 5.2. Схема демодулятора, оптимального по критерию максимального правдоподобия

Другую форму алгоритма можно получить из выражения (5.11)

$$\begin{aligned}
 & \max_i \left\{ \int_0^T z^2(t) dt - \int_0^T [z(t) - s_i(t)]^2 dt \right\} = \\
 & \max_i \left\{ \int_0^T [z^2(t) - z^2(t) + 2z(t)s_i(t) - s_i^2(t)] dt \right\} \rightarrow \\
 & \rightarrow \max_i \left[\int_0^T z(t)s_i(t) dt - \frac{1}{2} \int_0^T s_i^2(t) dt \right] = \max_i \left[(\vec{z}, \vec{s}_i) - \frac{1}{2} \underbrace{\|\vec{s}_i\|^2}_{E_i} \right] \rightarrow \hat{b}_i,
 \end{aligned}$$

или

$$\hat{i} = \arg \left\{ \max_i \left[(\vec{z}, \vec{s}_i) - \frac{1}{2} E_i \right] \right\}, \quad (5.13)$$

где E_i – энергия i -го сигнала.

Схема оптимального демодулятора, реализующего алгоритм (5.13), приведена на рис. 5.3.

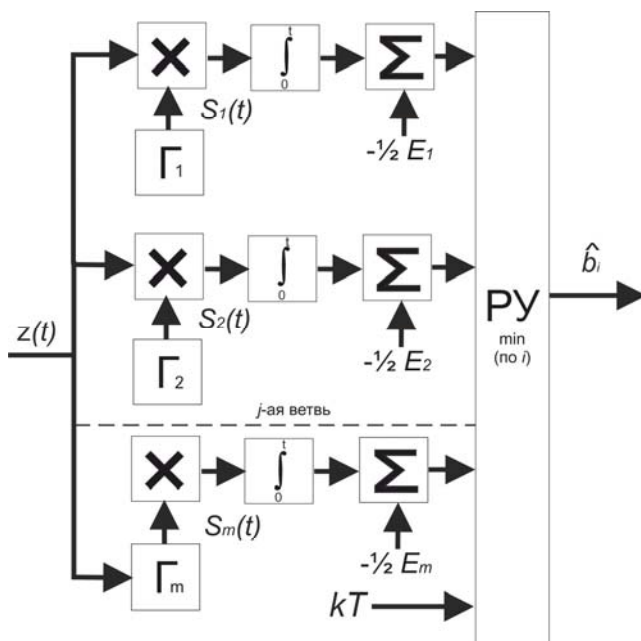


Рис. 5.3. Оптимальный демодулятор на корреляторах (энергии сигналов не равны)

Поскольку в каждой ветви такого демодулятора присутствует вычислитель скалярного произведения (\vec{z}, \vec{s}_i) или коррелятор, то его называют демодулятором на корреляторах (активных фильтрах). Если использовать сигналы равных энергий, то алгоритм (5.13) и схема демодулятора (рис. 5.3) существенно упрощаются (рис. 5.4)

$$\hat{i} = \arg \left\{ \max_i (\vec{z}, \vec{s}_i) \right\} = \arg \left\{ \max_i \int_0^T z(t) s_i(t) dt \right\}. \quad (5.14)$$

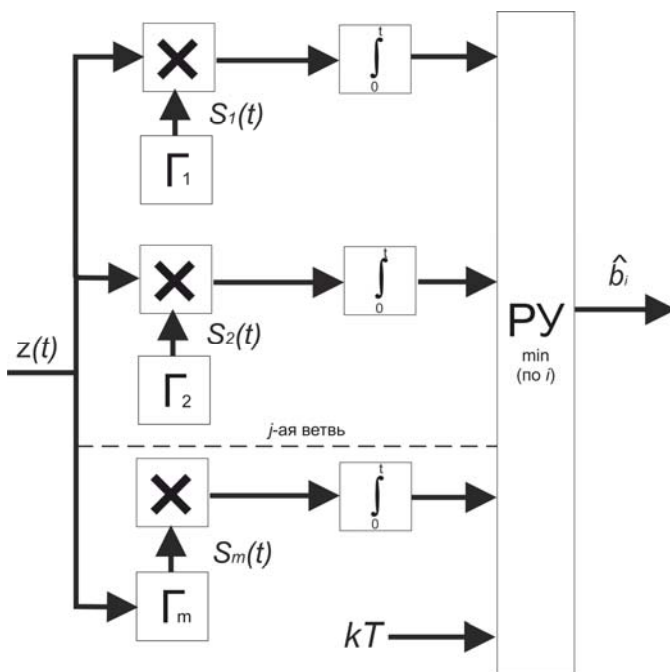


Рис. 5.4. Оптимальный демодулятор на корреляторах (энергии сигналов равны)

Все рассмотренные демодуляторы используют всю информацию о форме сигналов $s_i(t)$, включая начальную фазу. В каждой их ветви содержатся генераторы, генерирующие синфазные образцы этих сигналов, поэтому их называют когерентными демодуляторами.

5.3.2. Синтез оптимального когерентного демодулятора на согласованных фильтрах

Сохраняя постановку задачи синтеза демодулятора из предыдущего раздела и опираясь на алгоритмы (5.13) и (5.14), попробуем заменить коррелятор (активный фильтр), вычисляющий скалярные произведения приходящего колебания и образцов сигналов, на пассивный линейный фильтр, реализующий ту же операцию.

Известно, что реакция линейного аналогового фильтра на воздействие $z(t)$ вычисляется с помощью интеграла Дюамеля (интеграл наложения, интеграл свертки)

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

Потребуем, чтобы в заранее выбранный момент времени t_0 значение этой реакции $y(t_0)$ с точностью до коэффициента совпало со скалярным произведением (5.14)

$$y_i(t_0) = \int_0^{t_0} z(\tau)g_i(t_0 - \tau)d\tau = a \int_0^T z(\tau)s_i(\tau)d\tau.$$

Как видно, это достигается при $g_i(t_0 - \tau) = as_i(\tau)$ и $t_0 \geq T$. После замены переменных $t_0 - \tau = t$ получаем

$$g_i(t) = as_i(t_0 - t). \quad (5.15)$$

Фильтры, обладающие такими импульсными характеристиками, называют согласованными (СФ) с соответствующими сигналами.

На рис. 5.5 изображены сигнал длительностью T и импульсные характеристики согласованных с ним фильтров для $t_0 = T$ и $t_0 > T$, из которых видно, что импульсная характеристика согласованного фильтра является «зеркальным отражением» сигнала относительно момента времени $0,5t_0$.

Таким образом, фильтры с импульсными характеристиками (5.15) могут заменить корреляторы в ветвях оптимального демодулятора (рис. 5.3 и рис. 5.4), если решения принимать по отсчетам их реакции $y_i(kT)$ (рис. 5.6).

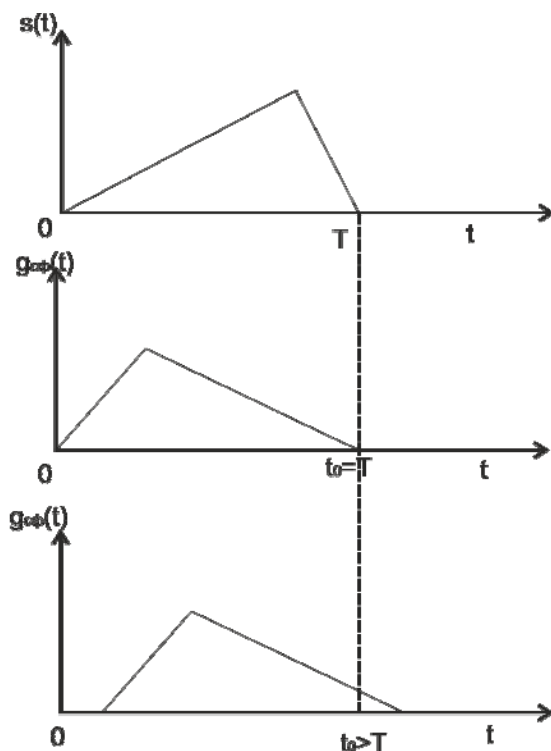


Рис. 5.5. Сигнал и импульсные характеристики согласованных фильтров

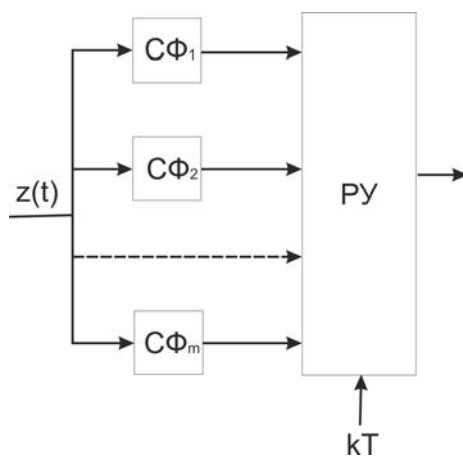


Рис. 5.6. Демодулятор на согласованных фильтрах

Основные свойства согласованных фильтров

1. Импульсная характеристика согласованного фильтра является «зеркальным отражением» сигнала, с которым он согласован, относительно момента времени $0,5t_0$ (с точностью до постоянного коэффициента)

$$g_{\text{сф}}(t) = as_i(t_0 - t).$$

Это свойство было положено в основу определения СФ (5.15).

2. Передаточная функция (частотная характеристика) согласованного фильтра

$$H_{\text{сф}}(j\omega) = \int_{-\infty}^{\infty} g_{\text{сф}}(t) e^{-j\omega t} dt = a \int_{-\infty}^{\infty} s(t_0 - t) e^{-j\omega t} dt.$$

После замены $t_0 - t = \tau$, $t = t_0 - \tau$, $dt = -d\tau$, при $t \rightarrow \infty$, $\tau \rightarrow -\infty$

$$\begin{aligned} H_{\text{сф}}(j\omega) &= -a \int_{\infty}^{-\infty} s(\tau) e^{-j\omega(t_0 - \tau)} d\tau = \\ &= ae^{-j\omega t_0} \int_{-\infty}^{\infty} s(\tau) e^{j\omega \tau} d\tau = ae^{-j\omega t_0} S^*(j\omega) \end{aligned}$$

Таким образом, передаточная функция СФ с точностью до множителя $ae^{-j\omega t_0}$ совпадает с сопряженной спектральной функцией сигнала, с которым он согласован

$$H_{\text{сф}}(j\omega) = ae^{-j\omega t_0} S^*(j\omega).$$

Амплитудно-частотная характеристика согласованного фильтра

$$H_{\text{сф}}(\omega) = |H_{\text{сф}}(j\omega)| = a |S(j\omega)| = aS(\omega)$$

с точностью до коэффициента a повторяет амплитудный спектр сигнала, с которым он согласован.

Фазочастотная характеристика согласованного фильтра $\phi_{\text{сф}}(\omega) = \arg[H_{\text{сф}}(j\omega)] = \arg[S^*(j\omega)] - \omega t_0 = -\arg[S(j\omega)] - \omega t_0$ отличается знаком от фазового спектра сигнала, с которым он согласован (без учета слагаемого $-\omega t_0$).

3. Форма отклика согласованного фильтра на сигнал, с которым он согласован

$$y_{\text{СФ}}(t) = \int_{-\infty}^{\infty} s(\tau) g_{\text{СФ}}(t - \tau) d\tau.$$

Учитывая (5.15) можно записать, что

$$y_{\text{СФ}}(t) = a \int_{-\infty}^{\infty} s(\tau) s(\tau - t + t_0) d\tau = a B_s(t - t_0).$$

Таким образом, отклик СФ на сигнал, с которым он согласован, с точностью до коэффициента совпадает с его корреляционной функцией, смещенной по оси времени на интервал t_0 (рис. 5.7)

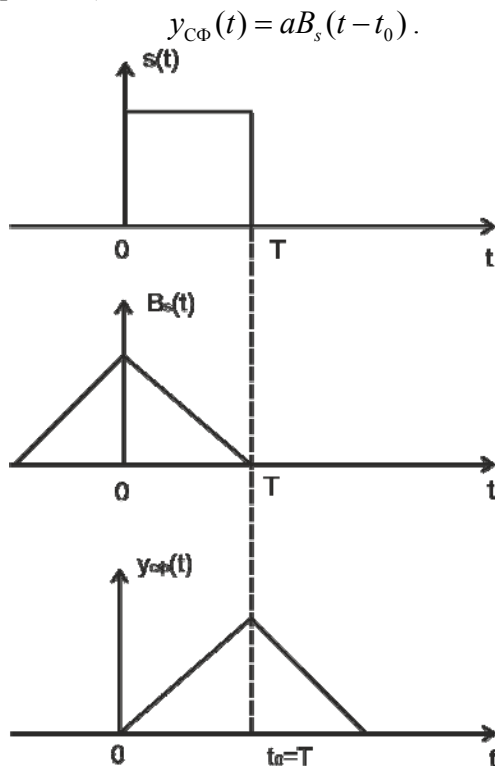


Рис. 5.7. Сигнал $s(t)$, его корреляционная функция $B_s(t)$ и отклик СФ $y_{\text{СФ}}(t)$

Из полученного результата вытекают следующие выводы.

- Отклик согласованного фильтра на сигнал, с которым он согласован, с точностью до постоянного коэффициента совпадает с его корреляционной функцией.
- Длительность отклика на сигнал, с которым он согласован, всегда равна $2T$.
- Согласованный фильтр не восстанавливает форму сигнала, искаженного шумом. Он создает один отсчет $y(t_0)$, по которому можно наилучшим образом судить о присутствии на входе сигнала, с которым он согласован.

4. Согласованный фильтр обеспечивает наибольшее отношение сигнал/шум (с/ш) на своем выходе при действии на входе аддитивной смеси сигнала, с которым он согласован, и центрированного нормального белого шума со спектральной плотностью мощности $N=N_0/2$.

Важно отметить, что под отношением с/ш на выходе СФ часто понимают отношение математического ожидания отсчета случайной реакции СФ $Y(t)$ в момент времени $t_0 = T$ к корню из ее дисперсии

$$\left(\frac{с}{ш} \right)_{\text{ВЫХ}} = \frac{\overline{Y(T)}}{\sigma_Y}. \quad (5.16)$$

Это важное свойство СФ иногда закладывают в основу его определения.

Найдем величину отношения с/ш на выходе СФ при действии на его входе сигнала, с которым он согласован,

$$\left(\frac{с}{ш} \right)_{\text{ВЫХ}} = \frac{\sqrt{E}}{\sqrt{N}} = \sqrt{\frac{2E}{N_0}} = \sqrt{2}h, \quad (5.17)$$

где E – энергия сигнала, с которым согласован СФ, N_0 – односторонняя спектральная плотность мощности шума,

$$h = \sqrt{\frac{E}{N_0}}.$$

Таким образом, максимальное отношение с/ш на выходе согласованного фильтра определяется энергией сигнала, с которым он согласован, независимо от формы этого сигнала.

Определим отношение с/ш по мощности

$$\left(\frac{P_s}{P_N}\right)_{\text{ВЫХ}} = \left(\frac{с}{ш}\right)_{\text{ВЫХ}}^2 = \frac{2E}{N_O} = \frac{2E \cdot TF_K}{N_O \cdot TF_K} =$$

$$= \frac{E}{T} \frac{1}{N_O F_K} 2F_K T = \left(\frac{P_s}{P_N}\right)_{\text{ВХ}} 2F_K T,$$

где F_K – ширина полосы пропускания канала.

При совпадении ширины полосы пропускания канала с шириной спектра сигнала $F_K = F_s$ имеем

$$\left(\frac{P_s}{P_N}\right)_{\text{ВЫХ}} = \left(\frac{P_s}{P_N}\right)_{\text{ВХ}} \cdot \underbrace{2F_s T}_{\text{база } s(t)}.$$

Отсюда вытекает целесообразность выбора сигналов с большой базой $2F_s T$ для передачи дискретных сообщений, что позволяет увеличить отношение с/ш при согласованной фильтрации.

5.3.3. Согласованная фильтрация и корреляционный прием некоторых типичных сигналов

Рассмотрим особенности когерентного приема некоторых сигналов и реализации соответствующих согласованных фильтров.

Прямоугольный видеоимпульс

Сигнал в виде прямоугольного видеоимпульса $s(t)$ (рис. 5.8а) и импульсная характеристика $g_{\text{СФ}}(t)$ согласованного с ним фильтра (рис. 5.8б) описываются выражениями

$$s(t) = \begin{cases} A & t \in (0, T) \\ 0 & t \notin (0, T) \end{cases}, \quad g_{\text{СФ}}(t) = as(T-t) = \begin{cases} aA & t \in (0, T) \\ 0 & t \notin (0, T) \end{cases}.$$

Вычислим передаточную функцию СФ

$$H_{\text{СФ}}(j\omega) = \int_{-\infty}^{\infty} g_{\text{СФ}}(t) e^{-j\omega t} dt = aA \int_0^T e^{-j\omega t} dt = \frac{aA}{j\omega} (1 - e^{-j\omega T}).$$

Сомножитель $\frac{1}{j\omega}$ представляет собой передаточную

функцию интегратора, вычитаемое $e^{-j\omega T}$ в скобках $(1 - e^{-j\omega T})$ – передаточная функция элемента задержки на время T , а сама скобка соответствует алгебраическому сумматору. В итоге приходим к схеме СФ, показанной на рис. 5.9.

Реакция согласованного фильтра на прямоугольный импульс показана на рис. 5.8в. Для сравнения на рис. 5.8г показана реакция на тот же сигнал коррелятора (рис. 5.10).

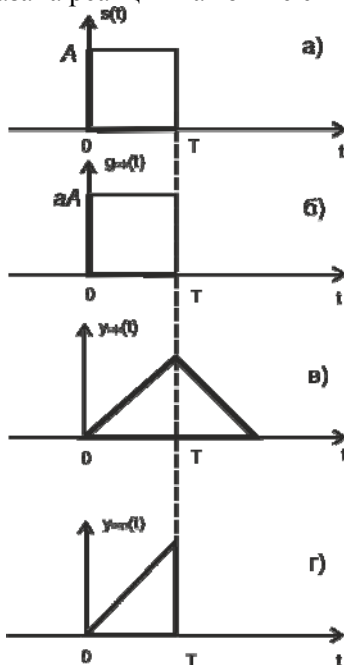


Рис. 5.8. Сигнал, импульсная характеристика СФ, реакции СФ и коррелятора

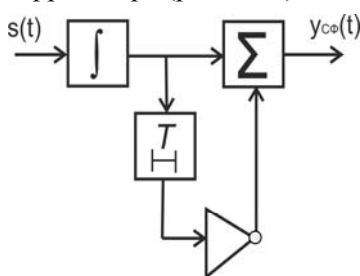


Рис. 5.9. Схема СФ для прямоугольного видеоимпульса

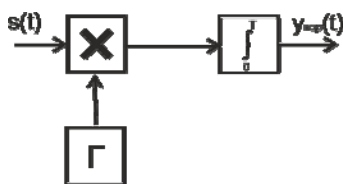


Рис. 5.10. Схема коррелятора

Прямоугольный радиоимпульс

Сигнал в виде прямоугольного радиоимпульса $s(t)$ описывается выражением

$$s(t) = \begin{cases} A \sin(\omega_c t + \phi_0) & t \in (0, T) \\ 0 & t \notin (0, T) \end{cases}.$$

Импульсная характеристика $g_{\text{сф}}(t)$ согласованного с ним фильтра на интервале $t \in (0, T)$

$$\begin{aligned} g_{\text{сф}}(t) &= as(T-t) = aA \sin[\omega_c(T-t) + \phi_0] = \\ &= aA \sin(\omega_c t - \underbrace{\omega_c T - \phi_0 + \pi}_{\psi}) = aA \sin(\omega_c t + \psi). \end{aligned}$$

Такого рода импульсной характеристикой обладает колебательный контур с добротностью $Q \rightarrow \infty$, однако у него она продолжается во времени неограниченно. Для «гашения» импульсной характеристики (реакции контура на воздействие $\delta(t)$) в момент $t=T$ можно воспользоваться соответствующей коммутацией контура (рис. 5.11а) или вычитанием самой задержанной на T реакции (рис. 5.11б).

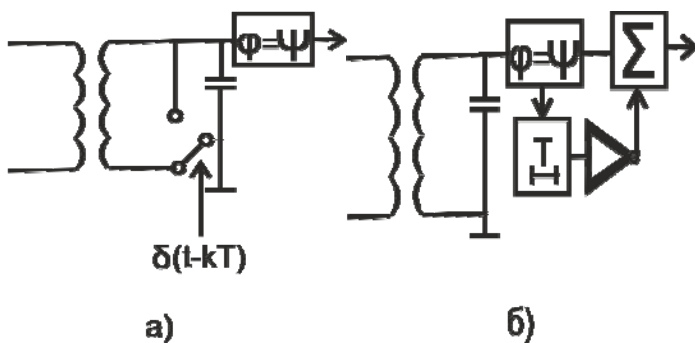


Рис. 5.11. Согласованные фильтры для прямоугольных радиоимпульсов

Прямоугольные радиоимпульсы и реакции на них СФ и коррелятора приведены на рис. 5.12.

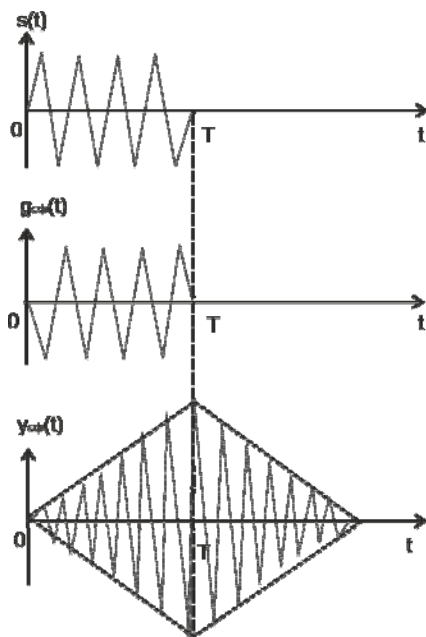


Рис. 5.12. Прямоугольный радиоимпульс, импульсная характеристика СФ и его реакция на сигнал, с которым он согласован

Сложные двоичные сигналы

Рассмотрим сигналы в виде n -последовательностей импульсов прямоугольной формы положительной и отрицательной полярности с фиксированным размахом. Возможный вид такого сигнала при $n = 7$ показан на рис. 5.13а. Усложнение сигнала объясняется желанием получить определенную (острую) форму отклика на выходе согласованного с ним фильтра и повысить отношение с/ш. Поскольку $y_{\text{сф}}(t) = aB_s(t - t_0)$, то чем острее (короче) $B_s(t - t_0)$, тем шире спектр сигнала и больше его база. Сигналы такого рода удобно использовать в радиолокационных и в асинхронно адресных телекоммуникационных системах.

Синтез СФ для сложного двоичного сигнала произведем, отталкиваясь от его ИХ (рис. 5.13б). Видно, что требуемую

форму $g_{\text{сф}}(t)$ можно получить суммированием прямоугольных импульсов длительностью $\Delta t = T / n$, сдвинутых на кратные Δt интервалы времени с соответствующими полярностями. Такие импульсы можно получить «размножением» единственного исходного П-импульса длительностью Δt с помощью линии задержки (ЛЗ) с n отводами (через Δt), а сам П-импульс в качестве импульсной характеристики – на выходе фильтра (СФ_П), согласованного с ним по форме. Эти рассуждения приводят к схеме фильтра, называемого трансверсальным (ТФ) (рис. 5.14). В цепях отводов ЛЗ включены повторители или инверторы сигналов с коэффициентами передачи +1 или -1 соответственно.

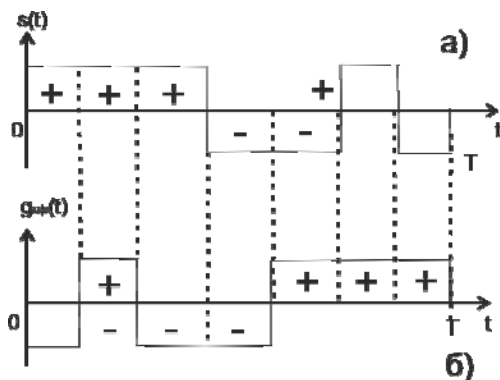


Рис. 5.13. Сложный двоичный сигнал и импульсная характеристика СФ

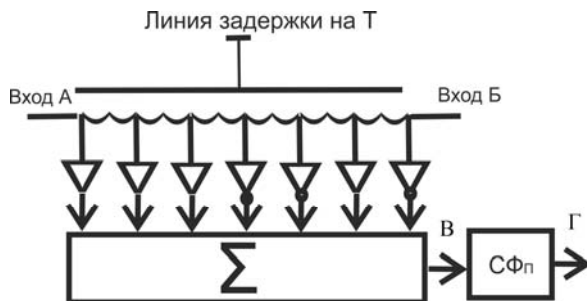


Рис. 5.14. Трансверсальный фильтр для сложного двоичного сигнала

Проанализируем импульсную характеристику ТФ со стороны входа А, как его реакцию на воздействие в виде δ -функции. Поданная на этот вход δ -функция (рис. 5.15а) появится на отводах ЛЗ с соответствующими задержками и после суммирования в сумматоре (с учетом полярности) создаст последовательность, показанную на рис. 5.15б.

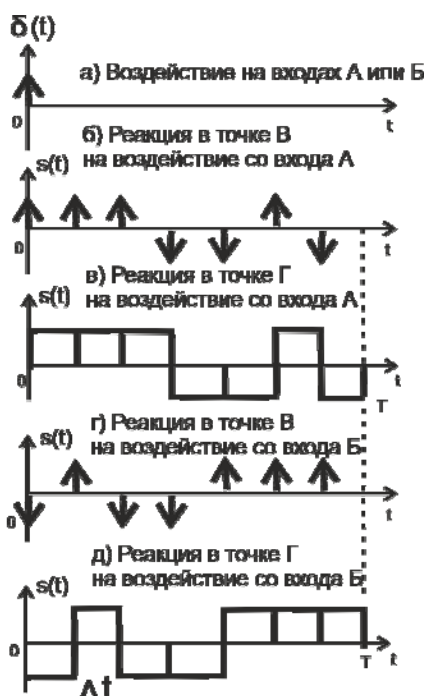


Рис. 5.15. Сигналы в отдельных точках transversального фильтра

На выходе СФ_П, согласованного с одиночным П-импульсом длительностью Δt , каждая из этих δ -функций вызовет реакцию (по определению, импульсную характеристику) в виде самого этого импульса. В результате получим на выходе ТФ n -последовательность, изображенную на рис. 5.15в, т. е. рассмотренная схема работает как формирователь сложного двоичного сигнала (рис. 5.15а).

Интересно, что при подаче δ -функции на вход Б, на выходе ТФ получим реакцию (см. графики на рис. 5.15г и рис. 5.15д), совпадающую с показанной на рис. 5.13б. Таким образом, один и тот же ТФ можно использовать в качестве формирователя сложного двоичного сигнала (вход А) и в качестве согласованного с этим сигналом фильтра (вход Б).

Из двоичных n -последовательностей наибольший интерес представляют собой последовательности (коды) Баркера. Они обладают важным свойством

$$|B_{\text{БОК}}| \leq \frac{B(0)}{n},$$

где $B_{\text{БОК}}$ – величина боковых лепестков корреляционной функции, $B(0)$ – начальное значение КФ.

Показанная на рис. 5.15 двоичная последовательность как раз и является кодом Баркера при $n = 7$.

Произвольные F -финитные сигналы

Согласно теореме отсчетов (теорема Котельникова, теорема Найквиста, теорема дискретизации), сигналы с ограниченным частотой F спектром точно передаются последовательностью своих отсчетов, взятых через интервалы

$$\Delta t = \frac{1}{2F}. \quad \text{Следовательно, они могут быть заменены}$$

ступенчатой функцией (рис. 5.16а), которая отличается от двоичного сигнала (рис. 5.13а) только размахами отдельных П-импульсов длительностью Δt . Отсюда вытекает возможность формирования такого рода сигналов и их согласованной фильтрации с помощью аналогового трансверсального фильтра (рис. 5.17). Его схема отличается от схемы двоичного ТФ (рис. 5.14) только заменой фильтра, согласованного с П-импульсом (СФ_П) на сглаживающий фильтр (ФНЧ с частотой среза F) и использованием в цепях отводов ЛЗ усилителей с коэффициентами усиления, пропорциональным отсчетам F -финитного сигнала.

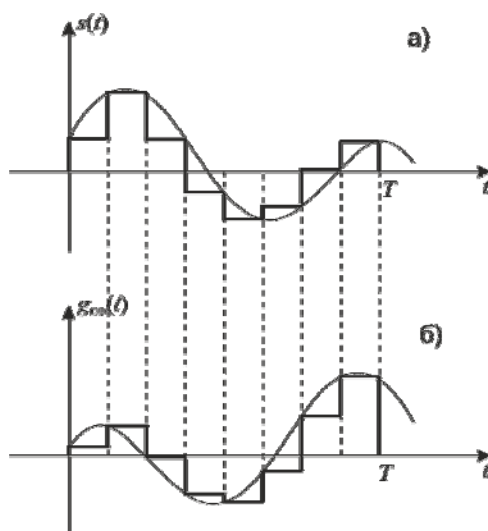


Рис. 5.16. F -финитный сигнал и импульсная характеристика СФ

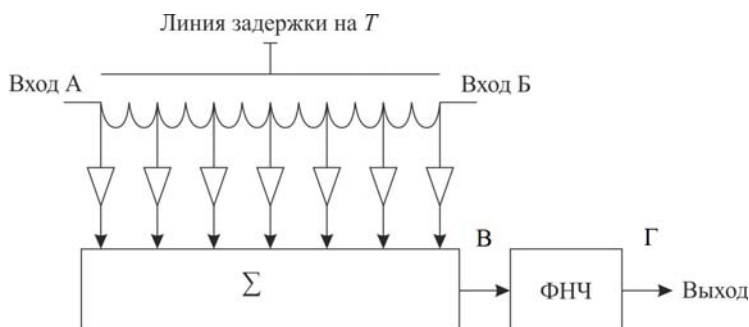


Рис. 5.17. Трансверсальный фильтр для F -финитного сигнала

5.3.4. Оптимальный когерентный прием при окрашенном шуме

Рассмотрим задачу синтеза согласованного фильтра, обеспечивающего максимальное отношение с/ш на своем выходе для случая, когда на его входе действует аддитивная

смесь известного сигнала $s(t)$ и нормального окрашенного (небелого) шума ($G(\omega) \neq N = \text{const}$). Эту задачу можно решить, разделив синтезируемый СФ на два последовательно включенных фильтра (рис. 5.18). От первого из них потребуем выравнивания энергетического спектра шума, т. е. превращения его в белый. По этой причине этот фильтр называют обесцвечивающим. Его передаточную функцию $H_1(j\omega)$ можно определить из условия

$$G(\omega) |H_1(j\omega)|^2 = N = \text{const}.$$

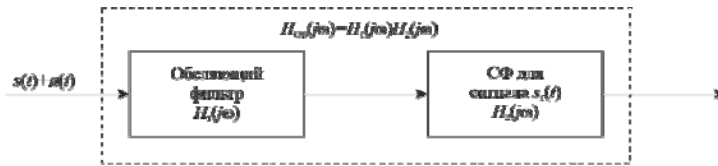


Рис. 5.18. Согласованный фильтр при окрашенном шуме

Передаточную функцию второго фильтра определим из условия его согласования с сигналом $s_1(t)$. Для этого предварительно найдем его спектральную функцию

$$S_{s_1}(j\omega) = S_s(j\omega)H_1(j\omega).$$

Тогда

$$H_2(j\omega) = aS_{s_1}^*(j\omega)e^{-j\omega T} = aS_s^*(j\omega)H_1^*(j\omega)e^{-j\omega T}.$$

Вычислим передаточную функцию искомого согласованного фильтра

$$\begin{aligned} H_{\text{сф}}(j\omega) &= H_1(j\omega)H_2(j\omega) = aH_1(j\omega)S_s^*(j\omega)H_1^*(j\omega)e^{-j\omega T} = \\ &= aS_s^*(j\omega) |H_1(j\omega)|^2 e^{-j\omega T} = aN \frac{S_s^*(j\omega)}{G(\omega)} e^{-j\omega T}. \end{aligned}$$

Аналогично поступают при реализации оптимального когерентного демодулятора на корреляторах.

5.4. Потенциальная помехоустойчивость когерентного приема

При решении данной задачи считаются заданными следующие положения.

1. Ансамбль сигналов на выходе модулятора

$$\{s_i(t)\}_m; i = 1, 2, \dots, m; t \in (0, T).$$

2. Непрерывный канал

$$Z(t) = s_i(t) + N(t); \quad i = 1, 2, \dots, m,$$

где $N(t)$ – квазигбелый нормальный шум, т. е.

$$G_N(f) = \begin{cases} N_0 = \text{const} & f \in (0, F) \\ 0 & f \notin (0, F) \end{cases}.$$

3. Алгоритм работы демодулятора. Предполагается, что это оптимальный когерентный по критерию максимального правдоподобия, демодулятор, работающий по правилу 5.13

$$\hat{i} = \arg \left\{ \max_i \left[(\bar{z}, \bar{s}_i) - \frac{1}{2} E_i \right] \right\}.$$

Нужно определить среднюю вероятность ошибочного приема P .

Ограничимся случаем двоичной системы ($m = 2$), когда

$$\{s_i(t)\}_{m=2} = \begin{cases} s_0(t) \\ s_1(t) \end{cases} \quad t \in (0, T).$$

Перепишем алгоритм (5.13) в развернутом виде

$$(\bar{z}, \bar{s}_0) - \frac{1}{2} E_0 \overset{\hat{b}_0}{>} \overset{\hat{b}_1}{<} (\bar{z}, \bar{s}_1) - \frac{1}{2} E_1,$$

или

$$\int_0^T z(t) s_0(t) dt - \frac{1}{2} E_0 \overset{\hat{b}_0}{>} \overset{\hat{b}_1}{<} \int_0^T z(t) s_1(t) dt - \frac{1}{2} E_1.$$

Из иной записи того же алгоритма

$$\int_0^T z(t) [s_0(t) - s_1(t)] dt \begin{matrix} \hat{b}_0 \\ > \\ \hat{b}_1 \end{matrix} \frac{1}{2} (E_0 - E_1) = \lambda_{\text{опт}}$$

вытекает достаточность одной ветви в оптимальном демодуляторе, которая должна содержать либо коррелятор с опорным генератором разностного сигнала, либо согласованный с этим разностным сигналом фильтр (рис. 5.19).

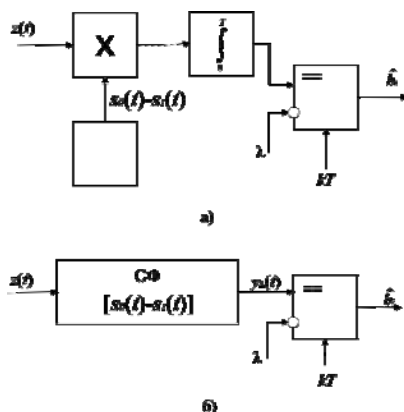


Рис. 5.19. Демодуляторы двоичной системы:
а) на корреляторе; б) на согласованном фильтре

В этих демодуляторах в качестве решающих устройств используются компараторы со стробированием.

Компаратор представляет собой дифференциальный усилитель с цифровым выходом и коэффициентом усиления $K \rightarrow \infty$. Напряжение на выходе компаратора может принимать одно из двух значений: высокое (уровень логической «1»), если напряжение на его прямом входе больше, чем на инверсном, и низкое (уровень логического «0») в противном случае. В данном случае производится сравнение выходного напряжения коррелятора или СФ с пороговым в моменты kT поступления коротких стробирующих импульсов. Символом « \equiv » в

условном графическом обозначении (УГО) компаратора обозначена операция сравнения, а кружком – инверсный вход.

Для решения поставленной задачи рассмотрим случайную величину $Y_{\Delta}(T)$ – отсчеты реакции СФ в конце каждого сигнала на входной СП $Z(t) = s_i(t) + N(t)$. Случайная величина $Y_{\Delta}(T)$ имеет нормальное распределение с двумя возможными математическими ожиданиями $\overline{Y_{\Delta}(T)}$:

y_0 – при передаче сообщения b_0 ,

y_1 – при передаче сообщения b_1 ,

$$y_0 = \int_{-\infty}^{\infty} yw(y/b_0)dy = S_0,$$

$$y_1 = \int_{-\infty}^{\infty} yw(y/b_1)dy = S_1.$$

Условные распределения величины $Y_{\Delta}(T)$ показаны на рис. 5.20.

В двоичных системах имеют место ошибки двух типов. Определим их вероятности

$$P(\hat{b}_1/b_0) = \int_{\lambda}^{\infty} yw(y/b_0)dy = S_0, \quad P(\hat{b}_0/b_1) = \int_{-\infty}^{\lambda} yw(y/b_1)dy = S_1.$$

Средняя вероятность ошибочного приема

$$P = P(b_0)P(\hat{b}_1/b_0) + P(b_1)P(\hat{b}_0/b_1).$$

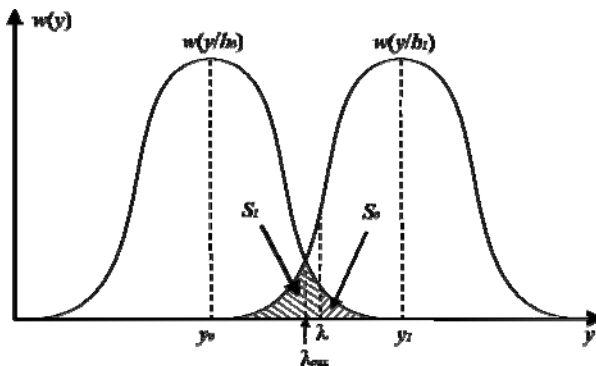


Рис. 5.20. Условные распределения отсчетов $Y_{\Delta}(T)$

При равных вероятностях передаваемых сообщений $P(b_0) = P(b_1) = 0,5$

$$P = 0,5[P(\hat{b}_1 / b_0) + P(\hat{b}_0 / b_1)] = 0,5[S_0 + S_1].$$

Минимизация вероятности P означает минимизацию суммы $S_0 + S_1$, что достигается при выборе оптимального порога $\lambda_{\text{опт}}$, определяемого из условия $w(y / b_0) = w(y / b_1)$ (рис. 5.20)

$$\lambda_{\text{опт}} = \frac{y_0 + y_1}{2}.$$

При таком выборе порога

$$P(\hat{b}_1 / b_0) = P(\hat{b}_0 / b_1) = P,$$

поэтому для вычисления средней вероятности ошибочного приема P достаточно определить любую условную вероятность ошибок, например $P(\hat{b}_1 / b_0)$

$$P = P(\hat{b}_1 / b_0) = \int_{\lambda_{\text{опт}}}^{\infty} y w(y / b_0) dy = \int_{\lambda_{\text{опт}}}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y-y_0)^2}{2\sigma^2}} dy.$$

Произведем замену переменных

$$\begin{aligned} \frac{y - y_0}{\sigma} &= v, & \frac{dy}{\sigma} &= dv, & v_{\text{опт}} &= \frac{\lambda_{\text{опт}} - y_0}{\sigma} = \\ &= \frac{\frac{y_0 + y_1}{2} - y_0}{\sigma} = \frac{y_1 - y_0}{2\sigma} \end{aligned}$$

и получим

$$\begin{aligned} P &= \frac{1}{\sqrt{2\pi}} \int_{v_{\text{опт}}}^{\infty} e^{-\frac{v^2}{2}} dv = Q(v_{\text{опт}}) = \\ &= 1 - F(v_{\text{опт}}) = \frac{1}{2} [1 - \Phi(v_{\text{опт}})], \end{aligned} \quad (5.18)$$

где $Q(v_{\text{опт}})$ – дополнительная функция ошибок, $F(v_{\text{опт}})$ – функция ошибок, $\Phi(v_{\text{опт}})$ – функция Крампа.

Все эти функции табулированы, их можно найти в соответствующих математических справочниках.

Полученный результат свидетельствует, что для любой двоичной системы при когерентном приеме вероятность ошибок определяется исключительно величиной $v_{\text{опт}}$, на которой сосредоточим свое внимание. Из рассмотренного вытекает

$$v_{\text{опт}} = \frac{y_1 - y_0}{2\sigma} = \frac{1}{2} \frac{\overline{Y_{\Delta}(T)}}{\sigma} = \frac{1}{2} \left(\frac{c}{\text{ш}} \right)_{\text{ВЫХ}},$$

где $\overline{Y_{\Delta}(T)} = y_1 - y_0$ – математическое ожидание отклика фильтра, согласованного с разностным сигналом $s_{\Delta}(t) = s_1(t) - s_0(t)$, на «свой» сигнал в момент $t = T$, а σ – квадратный корень из дисперсии этого отклика.

Используя ранее вычисленное значение отношения с/ш на выходе согласованного фильтра (5.17), получаем

$$v_{\text{опт}} = \frac{1}{2} \left(\frac{c}{\text{ш}} \right)_{\text{ВЫХ}} = \frac{1}{2} \sqrt{\frac{2E_{\Delta}}{N_0}} = \frac{1}{\sqrt{2}} h_{\Delta}, \quad (5.18)$$

где E_{Δ} – энергия разностного (эквивалентного) сигнала $s_{\Delta}(t)$, N_0 – спектральная плотность мощности шума,

$$h_{\Delta} = \sqrt{\frac{E_{\Delta}}{N_0}}.$$

Учитывая геометрический смысл энергии сигнала $E_{\Delta} = \|\vec{s}_1 - \vec{s}_0\|^2$, выражение (5.18) можно переписать в виде

$$v_{\text{опт}} = \frac{d(\vec{s}_1, \vec{s}_0)}{\sqrt{2N_0}}.$$

Таким образом, можно утверждать следующее.

1. Помехоустойчивость когерентного приема в двоичных системах определяется исключительно соотношением энергии E_{Δ} разностного сигнала (расстоянием между сигналами) и спектральной плотности мощности N_0 нормального белого шума

$$h_{\Delta} = \sqrt{\frac{E_{\Delta}}{N_0}} = \frac{d(\vec{s}_1, \vec{s}_0)}{\sqrt{N_0}}. \quad (5.19)$$

2. Средняя вероятность ошибочного приема для этого случая вычисляется с помощью дополнительной функции ошибок по формуле

$$P = Q\left(\frac{1}{\sqrt{2}}h_3\right) = Q\left[\frac{d(\vec{s}_1, \vec{s}_0)}{\sqrt{2N_0}}\right]. \quad (5.20)$$

5.5. Сравнительный анализ потенциальной помехоустойчивости основных видов цифровой модуляции

Для сравнения помехоустойчивости основных видов цифровой модуляции АМ, ЧМ (при использовании ортогональных сигналов) и ФМ достаточно для каждого из них определить эквивалентную энергию E_3 разностного сигнала $s_3(t) = s_1(t) - s_0(t)$ или расстояние между этими сигналами и воспользоваться выражением (5.20). Сравнение удобно выполнять на энергетической основе, т. е. определять соотношение энергий сигналов с разными видами модуляции, при котором обеспечиваются равные вероятности ошибочного приема. На рис. 5.21 в двумерном пространстве показаны векторы сигналов $s_0(t)$, $s_1(t)$ с равными энергиями и $s_3(t)$ для следующих случаев: а) АМ (при $s_0(t) = 0$), б) ЧМ и в) ФМ.

Из этих рисунков и выражения 5.20 следует:

$$P_{\text{АМ}} = Q\left(\frac{1}{\sqrt{2}}\sqrt{\frac{E}{N_0}}\right) = Q\left(\frac{1}{\sqrt{2}}h\right), \quad (5.21)$$

$$P_{\text{ЧМ}} = Q\left(\frac{1}{\sqrt{2}}\sqrt{\frac{2E}{N_0}}\right) = Q(h), \quad (5.22)$$

$$P_{\text{ФМ}} = Q\left(\frac{1}{\sqrt{2}}\sqrt{\frac{4E}{N_0}}\right) = Q(\sqrt{2}h), \quad (5.23)$$

где $h = \sqrt{\frac{E}{N_0}}$.

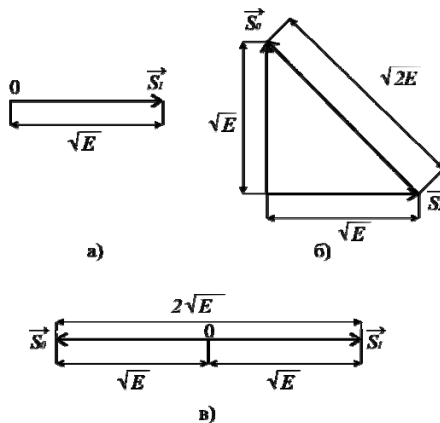


Рис. 5.21. К определению эквивалентной энергии
АМ, ЧМ и ФМ сигналов

Для достижения одинаковой помехоустойчивости ($P_{\text{АМ}} = P_{\text{ЧМ}} = P_{\text{ФМ}}$) энергия сигналов E при ЧМ должна быть в 2 раза, а при ФМ – в 4 раза меньше чем при АМ. Это означает, что по пиковой мощности ЧМ обеспечивает двукратный, а ФМ четырехкратный энергетический выигрыш по сравнению с АМ. По средней мощности выигрыши ЧМ и ФМ уменьшаются в 2 раза за счет пассивной паузы при АМ.

Таким образом, при равных энергиях сигналов наибольшей помехоустойчивостью обладает система с ФМ (использующая противоположные сигналы), наименьшей – система с АМ (с пассивной паузой). Система с ЧМ с ортогональными сигналами занимает промежуточное положение.

Следует отметить, что оптимальный порог в демодуляторе при использовании АМ не равен нулю, как при ЧМ и ФМ (при использовании сигналов с равными энергиями). Он зависит от энергии E (мощности) сигнала, которая может быть неизвестной или изменяться в процессе передачи, что затрудняет практическую реализацию оптимального приема.

Полученные результаты имеют общий характер и относятся не столько к конкретному виду модуляции при использовании гармонического переносчика, сколько к выбору

сигналов. В частности, формулы расчета средней вероятности ошибочного приема применимы для любых двоичных систем:

(5.21) – с пассивной паузой,

(5.22) – с ортогональными сигналами,

(5.23) – с противоположными сигналами.

Практическая реализация оптимального приема сигналов с наиболее помехоустойчивой модуляцией – ФМ является весьма проблематичной из-за чрезмерных требований к точности работы системы синхронизации:

При использовании согласованной фильтрации требуется высокая временная точность взятия отсчета реакции на выходе СФ (погрешность не должна превышать малой доли периода несущей частоты).

При использовании активного фильтра (коррелятора) столь же высокие требования предъявляются к фазовой погрешности опорного колебания.

Использование автономного опорного генератора в демодуляторе по этой причине исключается. Использование систем автоподстройки его частоты и фазы к соответствующим параметрам несущего колебания невозможно по причине его отсутствия в спектре ФМ сигнала (при равновероятных сообщениях). Возможный выход из этой ситуации состоит в использовании различных схем восстановления несущего колебания из принимаемого сигнала, например, схемы Пистолькорса А.А., использующей последовательно включенные умножитель и делитель частоты в два раза (рис. 5.22). Однако все схемы такого рода обладают существенным недостатком – неоднозначностью фазы (0 или π) восстановленного колебания несущей частоты. Это может приводить к так называемой «обратной работе», когда принимаемые сообщения инвертируются, т. е. вместо 0 регистрируются 1 и наоборот.

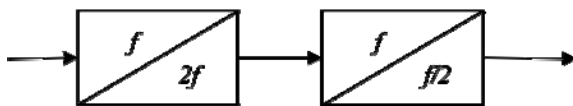


Рис. 5.22. Схема Пистолькорса

Эффективный способ решения этих проблем был предложен советским ученым Н. Т. Петровичем путем перехода к относительной фазовой модуляции (ОФМ). При ОФМ сообщение («0» или «1») передается не абсолютным значением фазы несущего колебания (0 или π), а разностью фаз текущего и предшествующего сигналов. В этом случае «0» передается сохранением фазы колебания, а «1» ее изменением на π .

Систему с ОФМ можно рассматривать как систему с ФМ со специальным перекодированием кодовых символов b_k в c_k на входе фазового модулятора по правилу $c_k = b_k \oplus c_{k-1}$. Символ \oplus означает суммирование по модулю 2 (логическую операцию «исключающее ИЛИ»). Принимать сигналы с ОФМ можно с помощью фазовых демодуляторов (рис. 5.18) с последующим обратным перекодированием выходных символов (рис. 5.23). В этой схеме обратное перекодирование осуществляется логическим элементом «исключающее ИЛИ» (символ «=1» на УГО) совместно с элементом задержки на T .

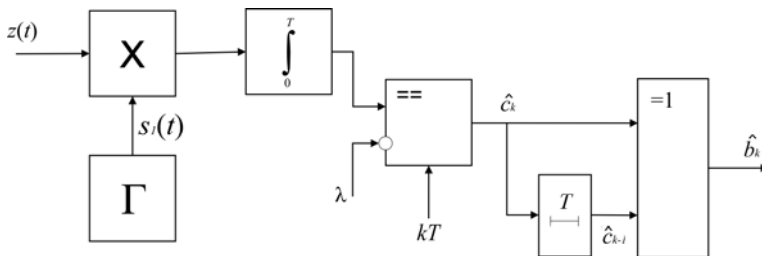


Рис. 5.23. Демодулятор двоичной системы с ОФМ

Определим вероятность ошибочного приема в системе с ОФМ при когерентном приеме. Поскольку в формировании выходного символа \hat{b}_k участвуют символы \hat{c}_k и \hat{c}_{k-1} , ошибочный прием имеет место при выполнении одного из двух условий:

- символ \hat{c}_k принят верно, а символ \hat{c}_{k-1} ошибочно,
- символ \hat{c}_k принят ошибочно, а символ \hat{c}_{k-1} верно.

Каждое из этих условий реализуется с вероятностью $P_{\text{ФМ}}(1-P_{\text{ФМ}})$. Таким образом, получаем $P_{\text{ОФМ}} = 2P_{\text{ФМ}}(1-P_{\text{ФМ}})$.

Поскольку требуется обеспечивать $P_{\text{ФМ}} \ll 1$, то

$$P_{\text{ОФМ}} \cong 2P_{\text{ФМ}} = 2Q(\sqrt{2}h).$$

Таким образом, своеобразной платой за переход от ФМ к ОФМ для устранения «обратной работы» является удвоение средней вероятности ошибочного приема.

На рис. 5.24 приведены кривые помехоустойчивости когерентного приема в двоичных системах, рассчитанные по указанным формулам.

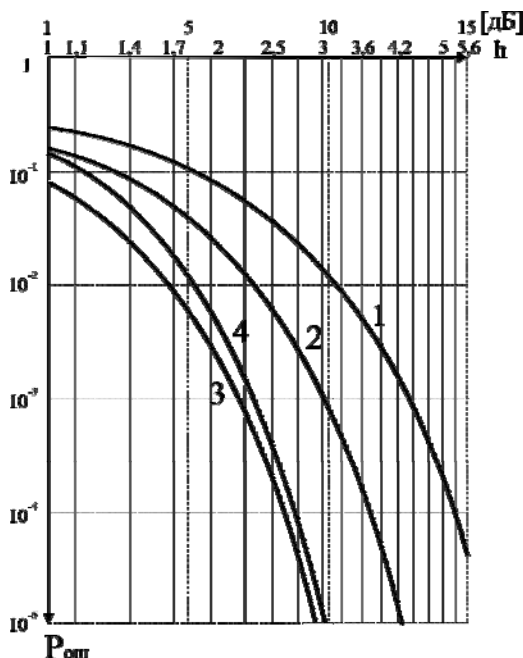


Рис. 5.24. Кривые помехоустойчивости оптимального когерентного приема:
1 – АМ; 2 – ЧМ; 3 – ФМ; 4 – ОФМ

5.6. Синтез оптимального демодулятора в канале с неопределенной фазой

Синтез оптимального демодулятора в канале с неопределенной фазой называется также некогерентным приемом. При решении данной задачи считаются заданными следующие положения.

Ансамбль сигналов на выходе модулятора

$$\{s_i(t)\}_m; i = 1, 2, \dots, m; t \in (0, T).$$

Непрерывный канал с неопределенной фазой

$$Z(t) = s_i(t - \tau) + N(t); \quad i = 1, 2, \dots, m,$$

где τ – случайная задержка сигнала в канале,

$$s_i(t - \tau) = A_i \cos(\omega_i t + \phi_i - \Theta_k) = s_{i, \Theta_k}; \quad i = 1, 2, \dots, m,$$

$\Theta_k = \omega_i \tau$ – случайная фаза с равномерным распределением

$$w(\Theta_k) = \frac{1}{2\pi}, \quad \Theta_k \in (0, 2\pi),$$

$N(t)$ – квазигибельный нормальный шум, т. е.

$$G_N(f) = \begin{cases} N_0 = \text{const} & f \in (0, F) \\ 0 & f \notin (0, F) \end{cases}.$$

В качестве критерия качества приема будем использовать критерий максимального правдоподобия (5.6). В нем отношение правдоподобия $\Lambda_i = \Lambda_i(\Theta_k)$, зависящее от Θ_k , является случайной величиной, поэтому потребуем максимизации его математического ожидания

$$\hat{i} = \arg \left\{ \max_i \left[\overline{\Lambda_i(\Theta_k)} \right] \right\} = \arg \left\{ \max_i \int_0^{2\pi} \Lambda_i(\Theta_k) w(\Theta_k) d\Theta_k \right\}. \quad (5.24)$$

В задаче требуется синтезировать оптимальный демодулятор, т. е. найти алгоритм оптимальной обработки входного сигнала и принятия решения о передаваемом сообщении.

Исходя из выражения для Λ_i (5.10), с учетом (5.13) можно записать

$$\Lambda_i(\Theta_k) = \exp \left\{ \frac{2}{N_o} \left[(\vec{z}, \vec{s}_{i, \Theta_k}) - \frac{1}{2} E_i \right] \right\}.$$

Для дальнейшего удобно разложить сигнал на квадратурные составляющие по углу Θ_k

$$\begin{aligned} s_{i, \Theta_k} &= A_i \cos(\omega_i t + \phi_i - \Theta_k) = \\ &= \underbrace{A_i \cos(\omega_i t + \phi_i)}_{s_i(t)} \cos \Theta_k + \underbrace{A_i \sin(\omega_i t + \phi_i)}_{\tilde{s}_i(t)} \sin \Theta_k = \\ &= s_i(t) \cos \Theta_k + \tilde{s}_i(t) \sin \Theta_k. \end{aligned}$$

Тогда можно записать

$$\begin{aligned} (\vec{z}, \vec{s}_{i, \Theta_k}) &= \int_0^T z(t) [s_i(t) \cos \Theta_k + \tilde{s}_i(t) \sin \Theta_k] dt = \\ &= \cos \Theta_k \int_0^T z(t) s_i(t) dt + \sin \Theta_k \int_0^T z(t) \tilde{s}_i(t) dt = \\ &= (\vec{z}, \vec{s}_i) \cos \Theta_k + (\vec{z}, \vec{\tilde{s}}_i) \sin \Theta_k = V_i \cos(\Theta_k - \Theta_i), \end{aligned}$$

где

$$V_i = \sqrt{(\vec{z}, \vec{s}_i)^2 + (\vec{z}, \vec{\tilde{s}}_i)^2}, \quad (5.25)$$

$$\Theta_i = \arctg \frac{(\vec{z}, \vec{\tilde{s}}_i)}{(\vec{z}, \vec{s}_i)} + \frac{\pi}{2} \text{sign}(\vec{z}, \vec{\tilde{s}}_i) [1 - \text{sign}(\vec{z}, \vec{s}_i)].$$

Вернемся к отношению правдоподобия

$$\begin{aligned} \Lambda_i(\Theta_k) &= \exp \left\{ \frac{2}{N_o} \left[(\vec{z}, \vec{s}_{i, \Theta_k}) - \frac{1}{2} E_i \right] \right\} = \\ &= \exp \left\{ \frac{2}{N_o} \left[V_i \cos(\Theta_k - \Theta_i) - \frac{1}{2} E_i \right] \right\} = \\ &= \exp \left(-\frac{E_i}{N_o} \right) \exp \left[\frac{2V_i}{N_o} \cos(\Theta_k - \Theta_i) \right] = \end{aligned}$$

$$= \exp(-h_i^2) \exp\left[\frac{2V_i}{N_o} \cos(\Theta_k - \Theta_i)\right].$$

Далее найдем математическое ожидание отношения правдоподобия (5.24)

$$\begin{aligned} \overline{\Lambda_i(\Theta_k)} &= \int_0^{2\pi} \exp(-h_i^2) \exp\left[\frac{2V_i}{N_o} \cos(\Theta_k - \Theta_i)\right] w(\Theta_k) d\Theta_k = \\ &= \frac{1}{2\pi} \exp(-h_i^2) \int_0^{2\pi} \exp\left[\frac{2V_i}{N_o} \cos(\Theta_k - \Theta_i)\right] d\Theta_k. \end{aligned}$$

Учитывая, что $\frac{1}{2\pi} \int_0^{2\pi} e^{a \cos \phi} d\phi = I_0(a)$ – модифицированная

функция Бесселя 0-го порядка, получим

$$\overline{\Lambda_i(\Theta_k)} = \exp(-h_i^2) I_0\left(\frac{2V_i}{N_o}\right).$$

Таким образом, окончательно искомый алгоритм можно записать в виде

$$\hat{i} = \arg \left\{ \max_i \left[\exp(-h_i^2) I_0\left(\frac{2V_i}{N_o}\right) \right] \right\}.$$

В таком виде алгоритм сложен для реализации. Для его упрощения можно применить любую монотонную функцию к выражению, стоящему в прямоугольных скобках $[x]$, например $\ln[x]$, что не изменит его суть

$$\begin{aligned} \hat{i} &= \arg \left\{ \max_i \ln \left[\exp(-h_i^2) I_0\left(\frac{2V_i}{N_o}\right) \right] \right\} = \\ &= \arg \left\{ \max_i \left[-h_i^2 + \ln I_0\left(\frac{2V_i}{N_o}\right) \right] \right\}. \end{aligned} \tag{5.26}$$

Из алгоритма (5.26) вытекает схема оптимального демодулятора, показанная на рис. 5.25.

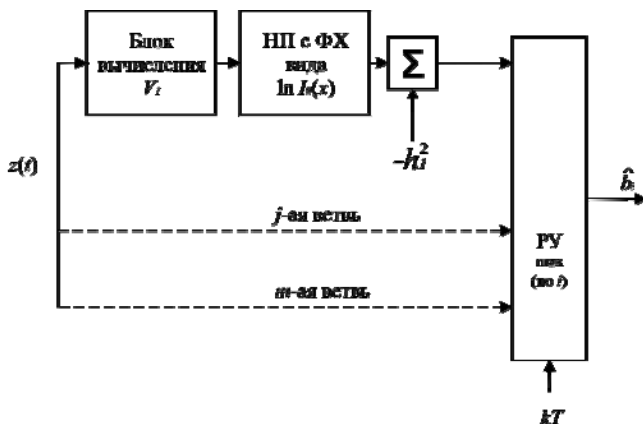


Рис. 5.25. Схема оптимального демодулятора

Такая схема сложна для реализации, а сам алгоритм чувствителен к h_i^2 . Снятие этой проблемы и упрощение схемы демодулятора возможно при выборе сигналов равных энергий $E_1 = E_2 = \dots = E_m$, что обеспечивает равенство $h_1 = h_2 = \dots = h_m$. Это позволяет исключить в ветвях демодулятора сумматоры и нелинейные преобразователи со сложной монотонной функциональной характеристикой вида $\ln[I_0(x)]$ (рис. 5.26), а алгоритм (5.26) принимает вид

$$\hat{i} = \arg \left\{ \max_i V_i \right\} = \arg \left\{ \max_i \left[(\vec{z}, \vec{s}_i)^2 + (\vec{z}, \vec{\tilde{s}}_i)^2 \right] \right\}. \quad (5.27)$$

Способ приема сигналов, при котором не используется информация о его фазе, называют некогерентным, как и соответствующие демодуляторы. Его алгоритм был впервые получен советским ученым Л. М. Финком.

Введенная функция V_i , как это следует из выражения (5.25), представляет собой огибающую реакции согласованного фильтра для соответствующего сигнала $s_i(t)$. Отсюда вытекает возможность реализации оптимального демодулятора, содержащего в каждой своей ветви СФ и детектор огибающей

(ДО) (рис. 5.27). Решение о переданном символе принимается по максимуму огибающей в моменты kT .

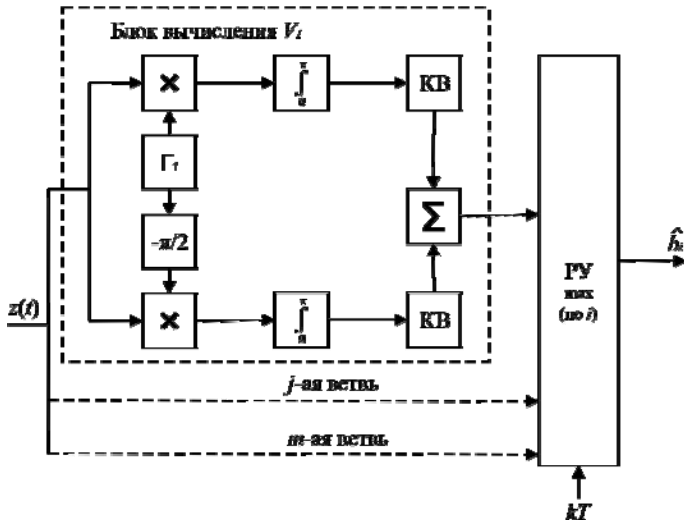


Рис. 5.26. Оптимальный некогерентный демодулятор

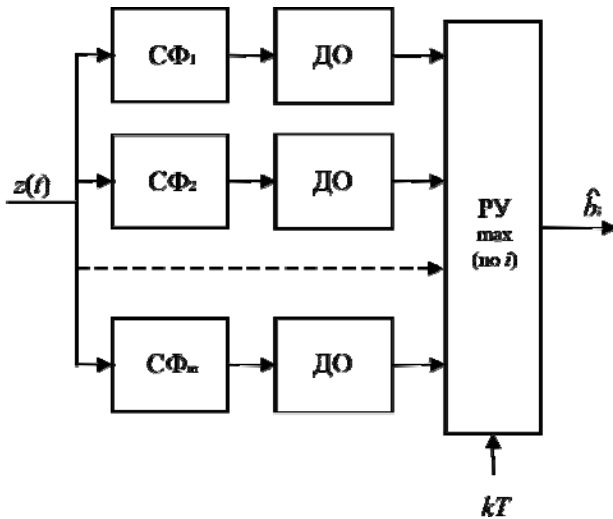


Рис. 5.27. Схема оптимального демодулятора на согласованных фильтрах

Из выражения (5.25) видно, что максимальная помехоустойчивость некогерентного приема достигается при минимальном (нулевом) значении огибающей V_j (в моменты отсчетов) на выходах ветвей $j \neq i$ при передаче сигнала $s_i(t)$. Для этого необходимо выбирать сигналы равных энергий, удовлетворяющие требованию ортогональности в усиленном смысле.

Математически это требование можно записать в виде

$$(\vec{s}_i, \vec{s}_j) = \int_0^T s_i(t) s_j(t) dt = 0; \quad i, j = 1, 2, \dots, m; \quad j \neq i,$$

$$(\vec{s}_i, \vec{\tilde{s}}_j) = \int_0^T s_i(t) \tilde{s}_j(t) dt = 0; \quad i, j = 1, 2, \dots, m.$$

Приведем несколько примеров ортогональных в усиленном смысле сигналов.

1. Сигналы с ЧМ при соответствующем выборе частот

$$s_i(t) = A \cos \left[\frac{2\pi(l+i)}{T} + \phi_i \right]; \quad i, l = 1, 2, 3, \dots; \quad l \gg i.$$

2. Сигналы с время-импульсной модуляцией (ВИМ) (рис. 5.28а)

$$s_i(t) = \begin{cases} A \sin \omega t, & t \in \left[i \frac{T}{m}, (i+1) \frac{T}{m} \right] \\ 0, & t \notin \left[i \frac{T}{m}, (i+1) \frac{T}{m} \right] \end{cases} \quad i = 0, 1, \dots, (m-1).$$

3. Сигналы с ОФМ обладают ортогональностью в усиленном смысле на интервале $-T \div T$ (рис. 5.28б). На этом интервале сообщения «0» и «1» передаются сигналами:

$$"0" \rightarrow s_0(t) = A \sin \omega t, \quad t \in (-T, T)$$

$$"1" \rightarrow s_1(t) = \begin{cases} A \sin \omega t, & t \in (-T, 0) \\ -A \sin \omega t, & t \in (0, T) \end{cases}.$$

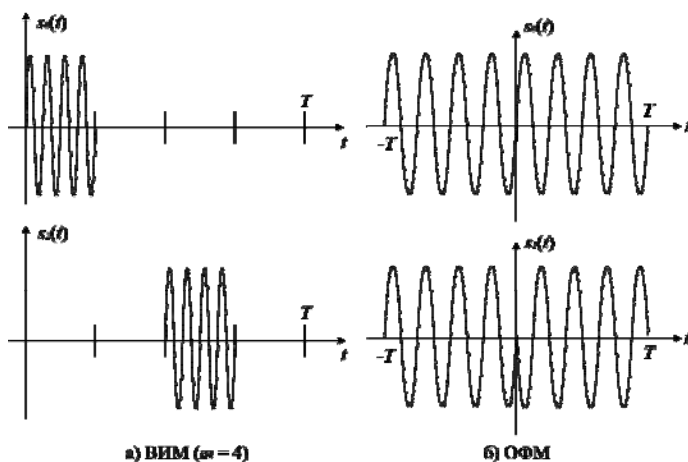


Рис. 5.28. Примеры ортогональных в усиленном смысле сигналов

5.7. Потенциальная помехоустойчивость некогерентного приема в двоичной системе связи

Для определения средней вероятности ошибки оптимального некогерентного приема в двоичной системе при равных вероятностях передаваемых сообщений $P(b_0) = P(b_1)$ достаточно вычислить условную вероятность ошибки любого типа

$$P = P(\hat{b}_0 / b_1) = P(\hat{b}_1 / b_0).$$

Вычислим $P(\hat{b}_0 / b_1)$, ориентируясь на схему некогерентного демодулятора на СФ (рис. 5.29). Ошибка вида \hat{b}_0 / b_1 (при передаче сообщения b_1 принимается решение в пользу \hat{b}_0) возникает, если для отсчетов огибающих на выходах

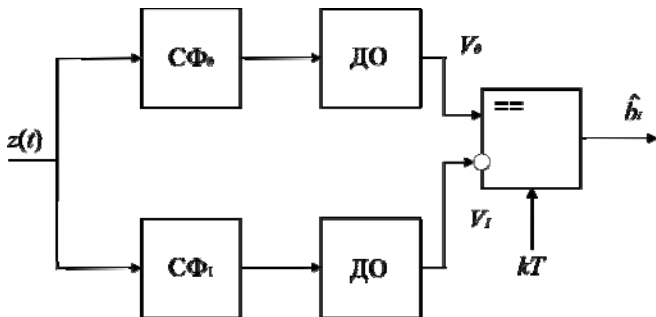


Рис. 5.29. Некогерентный демодулятор

ветвей демодулятора выполняется неравенство $V_0 > V_1 / b_1$. Для определения его вероятности

$$\begin{aligned} \overline{P(V_0 > V_1 / b_1)} &= \int_0^{\infty} P(V_0 > V_1) w(V_1 / b_1) dV_1 = \\ &= \int_0^{\infty} w(V_1 / b_1) dV_1 \int_{V_1}^{\infty} w(V_0 / b_1) dV_0 \end{aligned}$$

надо знать условные плотности вероятности $w(V_0 / b_1)$ и $w(V_1 / b_1)$. Вычислим их, используя выражение

$$V_i = \sqrt{(\vec{Z}, \vec{s}_i)^2 + (\vec{Z}, \vec{\tilde{s}}_i)^2},$$

где i – номер ветви (индекс сигнала, на который настроен СФ)

$$\vec{Z} = \vec{N} + \vec{s}_j, j - \text{индекс передаваемого сообщения.}$$

Теперь вычислим огибающую

$$\begin{aligned} V_i &= \sqrt{(\vec{N} + \vec{s}_j, \vec{s}_i)^2 + (\vec{N} + \vec{s}_j, \vec{\tilde{s}}_i)^2} = \\ &= \sqrt{[(\vec{N}, \vec{s}_i) + (\vec{s}_j, \vec{s}_i)]^2 + [(\vec{N}, \vec{\tilde{s}}_i) + (\vec{s}_j, \vec{\tilde{s}}_i)]^2} = \\ &= \sqrt{[(\vec{N}, \vec{s}_i) + (\vec{s}_j, \vec{s}_i)]^2 + (\vec{N}, \vec{\tilde{s}}_i)^2}. \end{aligned}$$

Запишем отсчеты огибающих V_0 и V_1 на выходах соответствующих ветвей демодулятора при передаче сигнала $s_1(t)$ ($j = 1$)

$$i = 0$$

$$V_0 = \sqrt{\left[\left(\vec{N}, \vec{s}_0 \right) + \left(\vec{s}_1, \vec{s}_0 \right) \right]^2 + \left(\vec{N}, \vec{\tilde{s}}_0 \right)^2} = \sqrt{\left(\vec{N}, \vec{s}_0 \right)^2 + \left(\vec{N}, \vec{\tilde{s}}_0 \right)^2}$$

$$i = 1$$

$$V_1 = \sqrt{\left[\left(\vec{N}, \vec{s}_1 \right) + \left(\vec{s}_1, \vec{s}_1 \right) \right]^2 + \left(\vec{N}, \vec{\tilde{s}}_1 \right)^2} = \sqrt{\left[\left(\vec{N}, \vec{s}_1 \right) + E \right]^2 + \left(\vec{N}, \vec{\tilde{s}}_1 \right)^2}.$$

Входящие в эти выражения скалярные произведения

$$\left(\vec{N}, \vec{s}_i \right) = \int_0^T N(t) s_i(t) dt, \quad \left(\vec{N}, \vec{\tilde{s}}_i \right) = \int_0^T N(t) \tilde{s}_i(t) dt$$

представляют собой нормально распределенные случайные величины с нулевым математическим ожиданием, следовательно, огибающие V_i будут иметь распределение Рэлея на выходе СФ₀ ($i = 0$)

$$w(V_0 / b_1) = \frac{V_0}{\sigma^2} e^{-\frac{V_0^2}{2\sigma^2}}$$

и распределение Райса на выходе СФ₁ ($i = 1$)

$$w(V_1 / b_1) = \frac{V_1}{\sigma^2} e^{-\frac{V_1^2 + E^2}{2\sigma^2}} I_0 \left(\frac{V_1 E}{\sigma^2} \right).$$

Теперь вернемся к вычислению средней вероятности ошибки при оптимальном некогерентном приеме

$$P = P(\hat{b}_0 / b_1) = \overline{P(V_0 > V_1 / b_1)} = \int_0^\infty w(V_1 / b_1) dV_1 \int_{V_1}^\infty w(V_0 / b_1) dV_0 =$$

$$\int_0^\infty w(V_1 / b_1) dV_1 \int_{V_1}^\infty w(V_0 / b_1) dV_0 =$$

$$= \int_0^\infty \frac{V_1}{\sigma^2} e^{-\frac{V_1^2 + E^2}{2\sigma^2}} I_0 \left(\frac{V_1 E}{\sigma^2} \right) dV_1 \int_{V_1}^\infty \frac{V_0}{\sigma^2} e^{-\frac{V_0^2}{2\sigma^2}} dV_0.$$

Произведя замену переменных

$$\frac{V_0^2}{2\sigma^2} = x, \quad dx = \frac{V_0 dV_0}{\sigma^2}, \quad V_1 \rightarrow x_1 = \frac{V_1^2}{2\sigma^2},$$

получим

$$\begin{aligned} P &= \int_0^\infty \frac{V_1}{\sigma^2} e^{-\frac{V_1^2 + E^2}{2\sigma^2}} I_0\left(\frac{V_1 E}{\sigma^2}\right) dV_1 \int_{x_1}^\infty e^{-x} dx = \\ &= \int_0^\infty \frac{V_1}{\sigma^2} e^{-\frac{V_1^2 + E^2}{2\sigma^2}} I_0\left(\frac{V_1 E}{\sigma^2}\right) dV_1 e^{-x} \Big|_\infty^{x_1} = \\ &= \int_0^\infty \frac{V_1}{\sigma^2} e^{-\frac{V_1^2 + E^2}{2\sigma^2}} I_0\left(\frac{V_1 E}{\sigma^2}\right) dV_1 e^{-\frac{V_1^2}{2\sigma^2}} = \\ &= \int_0^\infty \frac{V_1}{\sigma^2} e^{-\frac{2V_1^2 + E^2}{2\sigma^2}} I_0\left(\frac{V_1 E}{\sigma^2}\right) dV_1 = \\ &= \int_0^\infty \frac{V_1}{\sigma^2} e^{-\frac{4V_1^2 + 2E^2}{4\sigma^2}} I_0\left(\frac{V_1 E}{\sigma^2}\right) dV_1 = \\ &= \frac{1}{2} e^{-\frac{E^2}{4\sigma^2}} \int_0^\infty \frac{2V_1}{\sigma^2} e^{-\frac{4V_1^2 + E^2}{4\sigma^2}} I_0\left(\frac{V_1 E}{\sigma^2}\right) dV_1. \end{aligned}$$

Обозначим $2V_1 = V$, $2\sigma^2 = \gamma^2$, тогда $2dV_1 = dV$, $\frac{V_1}{\sigma^2} = \frac{V}{\gamma^2}$ и

$$P = \frac{1}{2} e^{-\frac{E^2}{4\sigma^2}} \underbrace{\int_0^\infty \frac{V}{\gamma^2} e^{-\frac{V^2 + E^2}{2\gamma^2}} I_0\left(\frac{VE}{\gamma^2}\right) dV}_{w(V)} = \frac{1}{2} e^{-\frac{E^2}{4\sigma^2}}.$$

Здесь учтено то, что подынтегральное выражение можно трактовать как плотность вероятности распределения Райса.

Вычислим мощность шума σ^2 на выходе согласованного фильтра

$$\begin{aligned}\sigma^2 &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{N_0}{2} |H_{\text{сф}}(j\omega)|^2 d\omega = \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{N_0}{2} S(j\omega) S^*(j\omega) d\omega = \frac{N_0}{2} E\end{aligned}$$

Тогда окончательно получим

$$P = \frac{1}{2} e^{-\frac{E^2 2}{4N_0 E}} = \frac{1}{2} e^{-\frac{1}{2} h^2},$$

где, как и ранее, $h = \sqrt{\frac{E}{N_0}}$.

Важно отметить, что полученный результат относится к любым двоичным системам, использующим ортогональные в усиленном смысле сигналы.

Для некогерентного приема в двоичных системах с пассивной паузой (AM)

$$P_{\text{AM}} = \frac{1}{2} e^{-\frac{1}{4} h^2}.$$

Некогерентный прием сигналов с ФМ в данном случае исключается, т. к. огибающие противоположных сигналов неразличимы, однако возможна реализация оптимального некогерентного демодулятора для системы с ОФМ (рис 5.30).

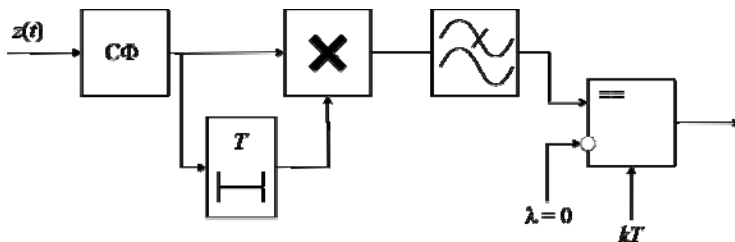


Рис. 5.30. Некогерентный демодулятор системы с ОФМ

Ее сигналы ортогональны в усиленном смысле на двойном интервале $2T$ и, следовательно, имеют на этом интервале удвоенную энергию. По этой причине

$$P_{\text{ОФМ}} = \frac{1}{2} e^{-h^2}.$$

Кривые помехоустойчивости оптимального некогерентного приема сигналов с АМ, ЧМ и ОФМ приведены в учебниках из списка литературы.

На практике используют также квазиоптимальный прием ЧМ сигналов, применяя в схеме (рис. 5.29) вместо СФ полосовые фильтры (ПФ) с прямоугольной АЧХ, согласуя их с шириной спектра сигналов. Советским ученым В. И. Сифоровым было установлено, что максимальное отношение с/ш на выходе ПФ с прямоугольной АЧХ

достигается при ширине полосы пропускания $\Delta f_{\text{ПФ}} = \frac{1,37}{T}$. Оно

несколько меньше, чем для СФ (энергетический проигрыш 0,86 дБ). Дополнительное снижение помехоустойчивости при использовании ПФ вызывается межсимвольной интерференцией (МСИ), возникающей из-за наложения переходных процессов от предшествующих посылок на последующие (что не имеет места при согласованной фильтрации).

5.8. Оптимальная фильтрация непрерывных сигналов

Рассмотрим задачу оптимальной фильтрации непрерывных сигналов, а точнее оптимальный фильтр Колмогорова-Винера. При решении данной задачи считаются заданными следующие положения.

На вход линейной цепи (в данном случае – фильтра) (рис. 5.31) поступает аддитивная смесь $Z(t)$ непрерывного случайного сигнала $S(t)$ и помехи $N(t)$.

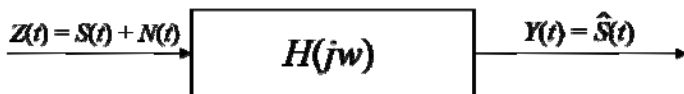


Рис. 5.31. К постановке задачи оптимальной фильтрации непрерывных сигналов

На рис. 5.32 приведены реализации сигнала $s(t)$ на входе фильтра и отклика фильтра $y(t) = \hat{s}(t)$. Различие между ними $\varepsilon(t) = \hat{s}(t) - s(t)$ – реализация случайной ошибки $E(t)$ воспроизведения переданного сигнала $s(t)$.

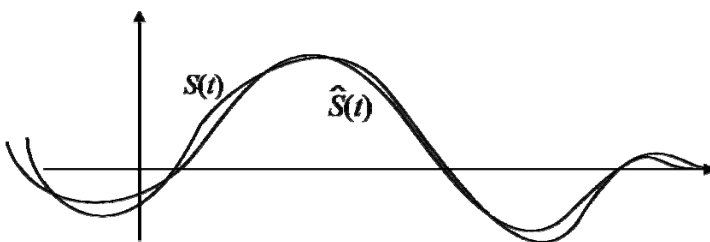


Рис. 5.32. К определению ошибки воспроизведения непрерывного сигнала

Требуется определить передаточную функцию (частотную характеристику) $H(j\omega)$ фильтра, обеспечивающего воспроизведение на своем выходе переданного сигнала $S(t)$ с минимальной погрешностью.

В данной постановке требуется уточнить, что понимать под погрешностью воспроизведения непрерывного сигнала и, следовательно, что минимизировать.

Следует отметить, что здесь не ставится задача выделения сигнала $s(t)$ из его смеси с помехой $s(t) + n(t)$ (это в принципе невозможно). Правильно говорить о таком преобразовании смеси сигнала с помехой, при котором отклик фильтра на эту смесь приближался бы к сигналу $s(t)$.

Советский ученый А. Н. Колмогоров (в 1939 г.) и независимо от него американский ученый Н. Винер (в 1942 г.) решили эту задачу, используя в качестве критерия качества

воспроизведения непрерывных сигналов среднеквадратичную погрешность $\sqrt{\overline{E^2(t)}}$.

Далее несколько уточним постановку задачи оптимальной фильтрации непрерывных сигналов. При ее решении считаются заданными следующие положения.

$S(t)$ и $N(t)$ – стационарные и взаимно независимые случайные процессы.

Энергетические спектры сигнала $G_S(\omega)$ и помехи $G_N(\omega)$ считаются известными.

В качестве критерия оптимизации нужно использовать $\overline{\min E^2(t)}$.

Требуется определить передаточную функцию $H(j\omega)$ оптимального фильтра.

Учитывая, что для стационарных процессов

$$\overline{E^2(t)} = \frac{1}{2\pi} \int_{-\infty}^{\infty} G_E(\omega) d\omega,$$

вычисляется энергетический спектр погрешности $G_E(\omega)$.

При вычислении погрешности будем считать допустимой задержку в фильтре на время τ :

$$E(t) = \hat{S}(t) - S(t - \tau) = S_{\text{вых}}(t) + N_{\text{вых}}(t) - S(t - \tau),$$

где $S_{\text{вых}}(t)$ – реакция фильтра на сигнал, $N_{\text{вых}}(t)$ – реакция фильтра на помеху.

Полученный результат можно трактовать как сумму двух независимых процессов $\Delta(t) = S_{\text{вых}}(t) - S(t - \tau)$, $N_{\text{вых}}(t)$.

Применительно к отдельным реализациям процесса $\Delta(t)$ имеем

$$S_{\Delta}(j\omega) = S_s(j\omega)H(j\omega) - S_s(j\omega)e^{-j\omega\tau} = S_s(j\omega)[H(j\omega) - e^{-j\omega\tau}].$$

Отсюда следует, что

$$G_{\Delta}(\omega) = G_S(\omega) \left| H(j\omega) - e^{-j\omega\tau} \right|^2,$$

$$G_{N_{\text{вых}}}(\omega) = G_N(\omega) \left| H(j\omega) \right|^2 = G_N(\omega) H^2(\omega).$$

Теперь можно определить энергетический спектр погрешности:

$$G_E(\omega) = G_\Delta(\omega) + G_{N_{\text{блх}}}(\omega) = G_S(\omega) \left| H(j\omega) - e^{-j\omega\tau} \right|^2 + G_N(\omega) H^2(\omega)$$

и саму среднеквадратичную погрешность:

$$\overline{E^2} = \frac{1}{2\pi} \int_{-\infty}^{\infty} [G_S(\omega) |H(j\omega) - e^{-j\omega\tau}|^2 + G_N(\omega) H^2(\omega)] d\omega. \quad (5.28)$$

Осталось минимизировать найденную погрешность (или, что то же самое, найти минимум подынтегрального выражения), оптимизируя передаточную функцию фильтра $H(j\omega)$. Оптимизацию

$$H(j\omega) = H(\omega) e^{j\varphi(\omega)}$$

произведем в два этапа.

1) Оптимизация по ФЧХ [$\varphi(\omega) = \text{var}$].

Для минимизации разности $H(j\omega) - e^{-j\omega\tau}$ необходимо совпадение векторов $H(j\omega)$ и $e^{-j\omega\tau}$ по направлению (рис. 5.33), что достигается при выборе $\varphi_{\text{опт}}(\omega) = -\omega\tau$.

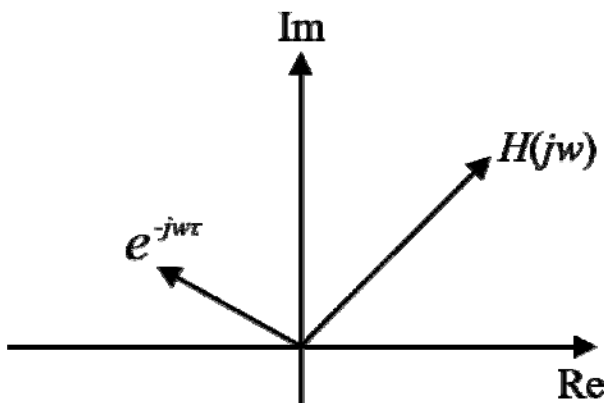


Рис. 5.33. Оптимизация ФЧХ

2) Оптимизация по АЧХ [$H(\omega) = \text{var}$].

Для нахождения минимума подынтегрального выражения (7.1) варьированием $H(\omega)$ следует производную этого выражения по $H(\omega)$ приравнять к нулю:

$$\frac{d \left\{ G_s(\omega) [H(\omega) - 1]^2 |e^{-j\omega\tau}|^2 + G_N(\omega) H^2(\omega) \right\}}{dH(\omega)} =$$

$$= 2H(\omega)G_s(\omega) - 2G_s(\omega) + 2H(\omega)G_N(\omega) = 0,$$

и решить полученное уравнение относительно $H(\omega)$:

$$H_{\text{опт}}(\omega) = \frac{G_s(\omega)}{G_s(\omega) + G_N(\omega)}.$$

Окончательно имеем

$$H_{\text{опт}}(j\omega) = \frac{G_s(\omega)}{G_s(\omega) + G_N(\omega)} e^{-j\omega\tau}.$$

При этом

$$\min \overline{E^2} = \frac{1}{2\pi} \int_{-\infty}^{\infty} \left\{ G_s(\omega) \left[\frac{G_s(\omega)}{G_s(\omega) + G_N(\omega)} - 1 \right]^2 + \right.$$

$$\left. + \frac{G_N(\omega) G_s^2(\omega)}{[G_s(\omega) + G_N(\omega)]^2} \right\} d\omega =$$

$$= \frac{1}{2\pi} \int_{-\infty}^{\infty} \left\{ \frac{G_N^2(\omega) G_s(\omega)}{[G_s(\omega) + G_N(\omega)]^2} + \frac{G_N(\omega) G_s^2(\omega)}{[G_s(\omega) + G_N(\omega)]^2} \right\} d\omega =$$

$$= \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{G_s(\omega) G_N(\omega)}{G_s(\omega) + G_N(\omega)} d\omega.$$

Рассмотрим возможности практической реализации фильтров с вычисленными выше характеристиками.

1. При отсутствии задержки сигнала в фильтре ($\tau = 0$).

В этом случае передаточная функция фильтра

$$H_{\text{опт}}(j\omega) = \frac{G_s(\omega)}{G_s(\omega) + G_N(\omega)}.$$

Является вещественной, а, следовательно, его импульсная характеристика

$$g(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{G_S(\omega) G_N(\omega)}{G_S(\omega) + G_N(\omega)} e^{j\omega\tau} d\omega$$

оказывается четной функцией частоты (рис. 5.34, а), из чего следует вывод о невозможности физической реализации такого фильтра.

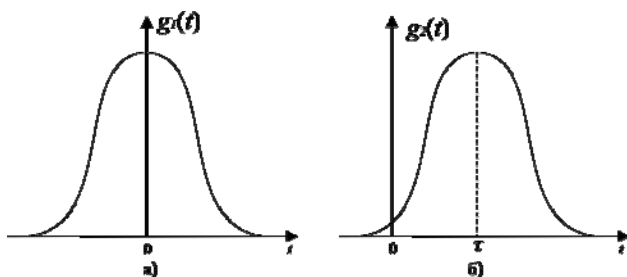


Рис. 5.34. Импульсные характеристики фильтров: а) $\tau = 0$; б) $\tau > 0$

2. При допустимости задержки сигнала в фильтре ($\tau > 0$).

Теперь, как это видно из рис. 5.34б, возможна реализация фильтра с импульсной характеристикой, близкой к требуемой (при отбрасывании ее части, соответствующей $t < 0$). Связанная с этим обстоятельством погрешность тем меньше, чем больше допустимая задержка сигнала в фильтре. Установлено, что эту погрешность нельзя дополнительно уменьшить, изменяя форму импульсной характеристики в области $t > 0$. Поэтому импульсная характеристика физического реализуемого оптимального фильтра Колмогорова-Винера при заданной задержке τ должна совпадать с идеальной в области $t > 0$.

Результаты оптимальной фильтрации сигналов неизвестной формы можно существенно улучшить, используя на передающей стороне предсказывающий фильтр с такой АЧХ, уменьшающей относительную ширину перекрытия энергетических спектров сигнала и помех. В общем случае следует стремиться к тому, чтобы сумма энергетических

спектров предсказанного сигнала $G_{\text{сиг}}(\omega)$ и канальных помех $G_N(\omega)$ приблизилась к константе:

$$G_{\text{сиг}}(\omega) + G_N(\omega) = \text{const.}$$

Изменения формы сигнала, вызванные предсказывающим фильтром на передающей стороне, следует компенсировать в приемнике с помощью корректирующего фильтра с обратной АЧХ, как это делают, например, в системах вещания с использованием аналоговой частотной модуляции.

Примеры

Пример 1

Пусть принимаемый сигнал имеет вид $z(t) = s_i(t) + n_0(t)$, $i = 1, 2$, где сигнал $s_i(t)$ принимает в момент передачи два возможных уровня $s_1(t) = a_1$ и $s_2(t) = a_2$, а $n_0(t)$ – нормальный случайный процесс с нулевым математическим ожиданием и дисперсией $\sigma_{n_0}^2$. Доказать, что при детектировании принимаемого сигнала с использованием критерия максимального правдоподобия порог принятия решения будет определяться с использованием следующего выражения:

$$\lambda_0 = \frac{a_1 + a_2}{2}.$$

Решение

Используя условные плотности вероятности, определяющие наблюдение z на приемной стороне при условии, что передаваемый сигнал был либо $s_1(t) = a_1$, либо $s_2(t) = a_2$, в виде

$$f(z | s_1) = \frac{1}{\sqrt{2\pi}\sigma_{n_0}} e^{-(z-a_1)^2 / (2\sigma_{n_0}^2)},$$

$$f(z | s_2) = \frac{1}{\sqrt{2\pi}\sigma_{n_0}} e^{-(z-a_2)^2 / (2\sigma_{n_0}^2)},$$

получим отношение правдоподобия:

$$\Lambda(z) = \frac{f(z | s_1)}{f(z | s_2)} = \frac{e^{-(z-a_1)^2 / (2\sigma_{n_0}^2)}}{e^{-(z-a_2)^2 / (2\sigma_{n_0}^2)}} = e^{z(a_1-a_2)/\sigma_{n_0}^2 - (a_1^2-a_2^2)/(2\sigma_{n_0}^2)}.$$

Подставляя $\Lambda(z)$ в выражение для критерия максимального правдоподобия, получим искомое выражение для порога

$$z \underset{H_2}{>} \frac{a_1^2 - a_2^2}{2(a_1 - a_2)} = \frac{a_1 + a_2}{2} = \lambda_0.$$

Пример 2

Найдите сигнал на выходе согласованного фильтра и определите максимальное отношение с/ш, если входной сигнал $s(t)$ является прямоугольным импульсом с амплитудой A и длительностью T (рис. 5.35).

Решение

Для данного сигнала $s(t)$ импульсная характеристика $h(t)$ согласованного фильтра

$$h(t) = s(T-t) = s(t).$$

Она совпадает с входным сигналом $s(t)$. Таким образом, сигнал на выходе $z(t)$, изображенный на рис. 5.35в, имеет вид:

$$z(t) = s(t) * h(t) = \begin{cases} A^2 t & \text{при } 0 \leq t \leq T, \\ -A^2 t + 2A^2 T & \text{при } T \leq t \leq 2T, \\ 0 & \text{при других } t. \end{cases}$$

Здесь $z(t) = A^2 T$ – максимальное значение $z(t)$.

Максимальное отношение с/ш на выходе согласованного фильтра можно определить следующим образом:

$$\left(\frac{с}{ш} \right)_{\max} = \frac{2E}{\eta},$$

где $\eta/2$ – спектральная плотность мощности АБГШ,

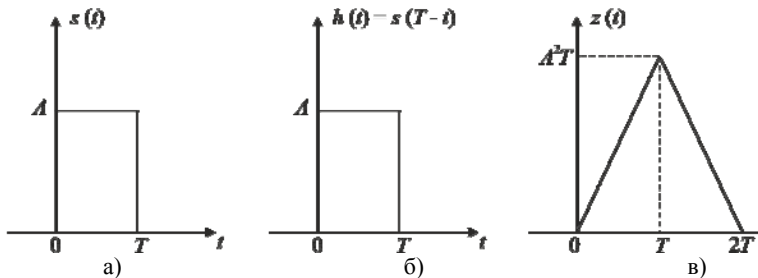


Рис. 5.35. Сигнал $s(t)$, импульсная характеристика согласованного фильтра $h(t)$ и результат свертки $s(t)$ и $h(t)$

E — энергия сигнала,
$$E = \int_{-\infty}^{\infty} s^2(t) dt = \int_0^T A^2 dt = A^2 T.$$

Следовательно, итоговое выражение для отношения с/ш имеет вид:

$$\left(\frac{с}{ш} \right)_{\max} = \frac{2A^2 T}{\eta}.$$

Пример 3

Биполярный двоичный сигнал $s_i(t)$ — это импульс от +1 В до -1 В на интервале $(0, T)$. Аддитивный белый шум со спектральной плотностью мощностью $\eta/2 = 10^{-5}$ Вт/Гц суммируется с сигналом. Определите максимальную скорость передачи данных при вероятности возникновения битовых ошибок $P_e \leq 10^{-4}$.

Решение

Вероятность возникновения битовых ошибок для биполярного двоичного сигнала можно определить с использованием следующего выражения:

$$P_e = Q\left(\sqrt{\frac{2A^2 T}{\eta}}\right),$$

где $Q(z)$ – комплементарная функция ошибки, значение которой можно определить по таблице, $Q(z) = \frac{1}{2} - \text{erf}(z)$,

$\text{erf}(z) = \frac{1}{\sqrt{2\pi}} \int_0^z e^{-\lambda^2/2} d\lambda$. В итоге получим, что

$$P_e = Q\left(\sqrt{\frac{2A^2T}{\eta}}\right) \leq 10^{-4}.$$

Используя таблицу для $Q(z)$, определим, что если $Q(z) = 10^{-4}$, то $z = 3.71$, следовательно,

$$\sqrt{\frac{2A^2T}{\eta}} = \sqrt{\frac{2(1)^2T}{2(10^{-5})}} = 3.71.$$

Таким образом, можно получить, что

$$T = (3.71)^2(10^{-5}) = 13.76(10^{-5}) \text{ с},$$

а максимальная скорость передачи данных

$$R = \frac{1}{T} = 7.26(10^3) \text{ бит/с} = 7.26 \text{ кбит/с}.$$

Контрольные вопросы

1. Сформулируйте задачу оптимального приема дискретных сообщений.
2. Дайте геометрическую трактовку задаче оптимального приема дискретных сообщений.
3. Что называют правилом решения (решающей схемой) демодулятора?
4. Что такое идеальный (оптимальный) приемник дискретных сообщений?
5. Что понимают под потенциальной помехоустойчивостью приема дискретных сообщений?
6. В чем суть теории потенциальной помехоустойчивости? Когда и кем были заложены ее основы?

7. Какой смысл вкладывают в понятие критерия качества приема дискретных сообщений? Перечислите известные вам критерии.
8. В чем суть критерия идеального наблюдателя?
9. Укажите особенности КИН.
10. Что представляет собой критерий максимального правдоподобия?
11. Как критерий максимального правдоподобия соотносится с критерием идеального наблюдателя?
12. В чем сущность критерия минимального среднего риска.
13. В чем общность критерия минимального среднего риска?
14. При каких условиях критерий минимального среднего риска совпадает с критерием идеального наблюдателя?
15. При каких условиях критерий минимального среднего риска совпадает с критерием максимального правдоподобия?
16. В чем сущность критерия Неймана-Пирсона?
17. Когда целесообразно использование критерия Неймана-Пирсона?
18. Сформулируйте задачу синтеза оптимального когерентного демодулятора.
19. Выведите алгоритм работы оптимального когерентного демодулятора.
20. Нарисуйте схему оптимального когерентного демодулятора на корреляторах.
21. В чем проявляется упрощение алгоритма (схемы) оптимального когерентного демодулятора при выборе ансамбля сигналов с равными энергиями?
22. Какие фильтры называют согласованными с сигналами?
23. Как импульсная характеристика согласованного фильтра связана с сигналом, с которым фильтр согласован?
24. Каковы передаточная функция, АЧХ и ФЧХ согласованного фильтра?
25. Какова форма отклика согласованного фильтра на «свой» сигнал?
26. Какова длительность отклика согласованного фильтра на «свой» сигнал?

27. Чему равно отношение с/ш на выходе согласованного фильтра?
28. В какой степени изменяется отношение с/ш при согласованной фильтрации аддитивной смеси сигнала с нормальным белым шумом?
29. Нарисуйте схему оптимального когерентного демодулятора на согласованных фильтрах.
30. Нарисуйте схему реализации согласованного фильтра для прямоугольного видеоимпульса.
31. Нарисуйте импульсную характеристику фильтра, согласованного с прямоугольным видеоимпульсом, и его реакцию на «свой» сигнал.
32. Нарисуйте схему реализации согласованного фильтра для прямоугольного радиоимпульса.
33. Нарисуйте импульсную характеристику фильтра, согласованного с прямоугольным радиоимпульсом, и его реакцию на «свой» сигнал.
34. Нарисуйте импульсную характеристику фильтра, согласованного со сложным двоичным сигналом, и его реакцию на «свой» сигнал.
35. Какие двоичные n -последовательности относятся к кодам Баркера? Каким свойством обладают их корреляционные функции? В чем полезность этих свойств для связи и радиолокации?
36. Нарисуйте схему трансверсального фильтра для формирования и согласованной фильтрации сложного двоичного сигнала.
37. Нарисуйте схему трансверсального фильтра для формирования и согласованной фильтрации произвольного F -финитного сигнала.
38. Какова передаточная функция, АЧХ и ФЧХ согласованного фильтра при небелом шуме на его входе?
39. Как количественно оценивают помехоустойчивость систем передачи дискретных сообщений (СПДС)?
40. Сформулируйте задачу расчета потенциальной помехоустойчивости СПДС.

41. Напишите алгоритм оптимального когерентного демодулятора двоичной системы связи.
42. Нарисуйте схему оптимального когерентного демодулятора АМ сигналов.
43. Нарисуйте схему оптимального когерентного демодулятора ЧМ сигналов.
44. Нарисуйте схему оптимального когерентного демодулятора ФМ сигналов.
45. Изложите методологию расчета средней вероятности ошибочного приема в двоичной системе связи.
46. От чего зависит помехоустойчивость двоичной системы связи?
47. Приведите формулу расчета средней вероятности ошибочного приема АМ сигналов в двоичной СПДС.
48. Приведите формулу расчета средней вероятности ошибочного приема ЧМ сигналов в двоичной СПДС.
49. Приведите формулу расчета средней вероятности ошибочного приема ФМ сигналов в двоичной СПДС.
50. В каком соотношении находятся энергии (мощности) сигналов с разными видами цифровой модуляции, обеспечивающие одинаковую помехоустойчивость? Дайте геометрическую трактовку этим соотношениям.
51. Перечислите проблемы практического использования ФМ в СПДС.
52. Что такое «обратная работа» и по каким причинам она возникает?
53. В чем сущность ОФМ?
54. Как формируют сигналы с ОФМ?
55. Как осуществляют оптимальный когерентный прием с ОФМ?
56. Как вычисляется средняя вероятность ошибочного приема в системах с ОФМ?
57. Расположите системы с АМ, ЧМ, ФМ и ОФМ в порядке убывания помехоустойчивости при равных энергиях сигналов.
58. Сформулируйте задачу синтеза оптимального некогерентного демодулятора.

59. Напишите алгоритм оптимального приема дискретных сообщений в канале с неопределенной фазой.
60. Как упрощается алгоритм некогерентного приема при условии равенства энергий используемых сигналов?
61. Нарисуйте схему оптимального некогерентного демодулятора для системы сигналов с разными энергиями.
62. Нарисуйте схему оптимального некогерентного демодулятора для системы сигналов с равными энергиями.
63. Как вычисляется огибающая V_i .
64. Нарисуйте схему блока определения огибающей V_i .
65. Нарисуйте схему оптимального некогерентного демодулятора на согласованных фильтрах.
66. Что означает ортогональность сигналов в усиленном смысле?
67. Почему для достижения максимальной помехоустойчивости некогерентного приема требуется использование ортогональных в усиленном смысле сигналов?
68. Приведите примеры систем, ортогональных в усиленном смысле сигналов.
69. При каких видах цифровой модуляции возможен некогерентный прием сигналов?
70. В чем состоит методология расчета средней вероятности ошибок при оптимальном некогерентном приеме.
71. Почему возможен некогерентный прием сигналов с ОФМ, а с ФМ он не возможен?
72. Приведите формулы для вычисления средней вероятности ошибочного некогерентного приема АМ, ЧМ и ОФМ сигналов.
73. Что называют квазиоптимальным некогерентным приемом? В чем его достоинства и недостатки?

Литература

1. Акулиничев, Ю. П. Теория электрической связи : учеб. пособие для вузов / Ю. П. Акулиничев. – СПб. : Лань, 2010. – 233 с.

2. Сальников, А. П. Теория электрической связи : конспект лекций. Ч. 2 / А. П. Сальников. – СПб. : 2002. – 87 с.
3. Теория электрической связи : учебник для вузов / под ред. Д. Д. Кловского. – М. : Радио и связь, 1999. – 432 с.
4. Крук, Б. И. Телекоммуникационные системы и сети : учеб. пособие : в 3 т. Т. 1 : Современные технологии / Б. И. Крук, В. Н. Попантопуло, В. П. Шувалов; под ред. проф. В. П. Шувалова. – М. : Горячая линия – Телеком, 2003. – 647 с.
5. Прокис, Д. Цифровая связь / Д. Прокис; под ред. Д. Д. Кловского. – М. : Радио и связь, 2000. – 800 с.
6. Сальников, А. П. Теория электрической связи: конспект лекций / А. П. Сальников. – СПб. : Изд-во «Линк», 2007. – 272 с.
7. Основы цифровой обработки сигналов : учеб. пособие / Ю. А. Брюханов, А. Л. Приоров, В. И. Джиган, В. В. Хрящев; Яросл. гос. ун-т им. П. Г. Демидова. – Ярославль : ЯрГУ, 2013. – 344 с.
8. Гмурман, В. Е. Теория вероятностей и математическая статистика : учеб. пособие для бакалавров / В. Е. Гмурман; М-во образования и науки РФ – 12-е изд. – М. : Юрайт, 2013. – 479 с.
9. Кремер, Н. Ш. Теория вероятностей и математическая статистика : учебник для вузов. / Н. Ш. Кремер; М-во образования РФ – 3-е изд., перераб. и доп. – М. : ЮНИТИ, 2009. – 551 с.
10. Теория вероятностей и математическая статистика / сост. Л. П. Бестужева, Н. Л. Майорова; Яросл. гос. ун-т им. П. Г. Демидова, Науч.-метод. совет ун-та. Ч. 1 : практикум. – Б.м. : Б.и., 2012. – 47 с.
11. Hsu, H. Analog and Digital Communications. Schaum's outline series / H. Hsu – McGraw Hill, 2003. – 330 p.
12. Hsu, H. Probability, Random Variables and Random Processes. Schaum's outline series / H. Hsu – McGraw Hill, 1997. – 311 p.

6. ОСНОВЫ ТЕОРИИ ИНФОРМАЦИИ

6.1. Энтропия источника, количество информации

В теории инфокоммуникаций под информацией понимают любые сведения о состоянии или поведении некоторого объекта (системы) либо о каких-то событиях, явлениях, предметах, подлежащие передаче от отправителя к получателю. Информация – это объект передачи по системе связи, она не зависит от физических представлений и формы сигналов.

На рис. 6.1 приведена структурная схема системы передачи информации. При этом сообщение, вырабатываемое источником, переносит собственную информацию, а сигналы на выходе передатчика, линии связи и приемника – взаимную информацию, т. е. информацию, содержащуюся в этих сигналах относительно передаваемого сообщения.

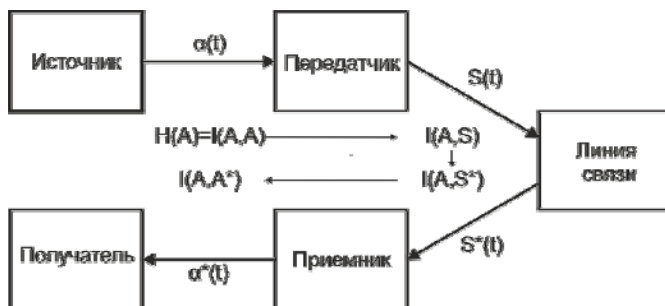


Рис. 6.1. Структурная схема системы передачи информации

Если получатель априори (заранее) точно (достоверно) знает, что передает отправитель, то количество получаемой им информации нулевое. Такая передача (связь) нецелесообразна. Информация, воспринимаемая получателем, отлична от нуля только в том случае, если передаваемые сведения являются для него новыми, непредвиденными. Другими словами, чем больше неожиданности (неизвестности) в полученном сообщении, тем больше информации в нем содержится.

Введем следующие определения.

Взаимная информация $I(A, Y)$ – это информация в наблюдаемом процессе $Y(t)$ относительно интересующего нас процесса $A(t)$.

Собственная информация $I(A, A) = I(A)$ – это информация, которая содержится в процессе $A(t)$ относительно себя самого. Величина $I(A)$ выражается через энтропию источника, обычно обозначаемую как $H(A)$.

Энтропия характеризует меру неопределенности. Действительно, если сообщение источника может быть предсказано заранее, то неопределенности относительно него не существует. В этом случае никакой процесс $Y(t)$ информации относительно $A(t)$ не содержит. Таким образом, количество информации в процессе или относительно процесса зависит от его априорной неопределенности (степени случайности).

Взаимная информация характеризует меру среднего уменьшения неопределенности относительно реализации $A(t)$ в процессе наблюдения за процессом $Y(t)$ и определяется следующим соотношением

$$\begin{array}{ccccc}
 I(A, Y) & = & H(A) & - & H(A/Y) \\
 \downarrow & & \downarrow & & \downarrow \\
 \text{Получена} & & \text{Была} & & \text{Осталась} \\
 \text{информация} & & \text{неопределенность} & & \text{неопределенность}
 \end{array}$$

где $H(A)$ – безусловная энтропия, характеризующая априорную (до передачи) неопределенность относительно $A(t)$; $H(A/Y)$ – условная энтропия, характеризующая апостериорную (после передачи) неопределенность относительно $A(t)$ при условии наблюдения $Y(t)$.

Количественные меры и единицы измерения информации

Во многих случаях нужно уметь измерять количество информации. При этом важным является вопрос выбора количественной меры. К ней предъявляются определенные требования, чтобы мера обладала необходимыми свойствами.

Свойства количественной меры информации

Информационная содержательность события должна определяться вероятностью его появления, а именно: количество информации должно быть обратно пропорционально вероятности наступления события.

Количество информации – величина неотрицательная.

Количество информации, содержащееся в достоверном событии, равно нулю.

Количество информации, заключенное в совокупности независимых сообщений, равно сумме количеств информации, заключенных в каждом сообщении в отдельности. Другими словами, количество информации целого складывается из количества информации о его частях. Это свойство называется свойством аддитивности (от addition – сложение).

Исторически первой количественной мерой информации является мера Хартли:

$$I(\alpha_i) = \log_b m^n = n \log_b m,$$

где m – объем алфавита, n – длина кодового слова.

Недостаток данной меры в том, что она не учитывает вероятностную (статистическую) структуру наблюдаемых сообщений и сигналов.

Количественную меру информации, учитывающую статистику сообщений и сигналов, ввел Клод Шеннон. Её соответственно и называют мерой Шеннона:

$$I(a_i) = \log_b \frac{1}{P_i},$$

где $I(a_i)$ – количество информации, заключенное в некотором сообщении a_i , P_i – вероятность появления сообщения a_i .

Единицы измерения количества информации определяются величиной основания логарифма b :

$$b = 10 \rightarrow \lg \rightarrow \text{dit},$$

$$b = e \rightarrow \ln \rightarrow \text{nat},$$

$$b = 2 \rightarrow \log_2 \rightarrow \text{bit}.$$

В основном в теории информации основание логарифма принимают равным 2, т. е. количество информации измеряют в двоичных единицах, или битах (1 бит = 1 дв. ед.).

Одна двоичная единица информации (или один бит) – это количество информации, которое получается, если произошло событие, вероятность появления которого равна 0,5:

$$I(a_i) = \log_2 \frac{1}{1/2} = \log_2 2 = 1 \text{ дв.ед. (бит)}.$$

Часто основание логарифма не обозначают, считая его равным 2 по умолчанию.

Количество информации, которое заключено в некотором многомерном сообщении, параметры которого принимают i, j, \dots, k -е значения с вероятностью $p(i, j, \dots, k)$, равно:

$$I(i, j, \dots, k) = \log_2 \frac{1}{P(i, j, \dots, k)}.$$

Чем меньше вероятность реализации случайного процесса при условии знания процесса за предыдущее время, тем больше неожиданность, а значит, и большее количество информации будет получено.

Классификация методов кодирования

По влиянию на избыточность различают три вида кодирования: примитивное, экономное и помехоустойчивое.

Примитивное кодирование характеризуется тем, что не влияет на избыточность сообщений в процессе их кодирования.

Чаще всего примитивные коды применяются для согласования объемов алфавитов источников сообщений и каналов связи. Например, равномерные двоичные коды: 5-разрядный код МТК-2 (код Бодо), 7-разрядный код ASCII, 8-разрядный код КОИ-8 и др. Примитивное кодирование используется также для засекречивания передаваемой информации, для получения желаемых временных и спектральных характеристик кодированных сигналов, для повышения устойчивости систем синхронизации. В последнем случае правило кодирования выбирается так, чтобы вероятность появления на выходе кодера длинной последовательности из

одних 0 (или 1) была бы минимальной. Такой код еще называют скремблером.

Экономное (эффективное) кодирование (сжатие данных) используется для уменьшения избыточности источников дискретных сообщений с целью уменьшения времени их передачи или требуемого объема памяти при хранении информации в запоминающем устройстве (ЗУ).

Экономное кодирование базируется преимущественно на теореме кодирования Шеннона (для канала без помех). Оно реализуется методами неравномерного кодирования и укрупнения алфавита источника. В качестве примера можно привести префиксные неравномерные коды Шеннона-Фано, Хаффмена, Зива-Лемпеля.

Экономное кодирование широко применяется в вычислительной технике. Современные версии операционных систем ЭВМ содержат в своем составе программы сжатия данных (архиваторы). Кроме того, протоколы связи, используемые в компьютерных модемах для связи между ЭВМ по телефонным сетям общего пользования, включают сжатие в число процедур обработки данных.

Помехоустойчивое (корректирующее) кодирование применяется для повышения помехоустойчивости связи путем обнаружения и (или) исправления ошибок, возникающих при передаче дискретных сообщений по каналам связи с помехами. Отличительной особенностью помехоустойчивого кодирования является увеличение избыточности сообщений в процессе его осуществления. Фундаментальное значение для помехоустойчивого кодирования имеет основная теорема кодирования Шеннона для канала связи с помехами.

6.2. Информационные характеристики источников сообщений

Энтропия источника независимых дискретных сообщений

Символы источника дискретных сообщений принимают различные значения с разными вероятностями, т. е. переносят различное количество информации. Для характеристики такого

источника информации в целом вводят среднюю меру, которая называется энтропией.

Энтропия источника информации представляет собой среднее количество информации, приходящееся на один символ дискретного источника.

Рассмотрим дискретный источник информации, который производит последовательность независимых кодовых символов a_i с соответствующими вероятностями P_i . Этот источник задается одномерным распределением $\{P_i\}$. Если задан алфавит (m – объем алфавита) и распределение, говорят, что задан ансамбль. Он описывается множеством

$$\text{Ансамбль} = \left\{ \begin{matrix} a_i \\ p_i \end{matrix} \right\}.$$

Символы не зависят друг от друга, поэтому такой источник называется источником без памяти. Согласно количественной мере Шеннона, количество информации, содержащееся в символе a_i , определяется как:

$$I(a_i) = \log_b \frac{1}{p_i}.$$

Поскольку появление символа a_i случайно, то случайно и количество информации, которое он несет. Поэтому проводят усреднение по всему ансамблю, чтобы найти среднее количество информации, которое переносит один символ дискретного источника, т. е. энтропию источника:

$$H(A) = I(A) = \sum_{i=0}^{m-1} p_i I(a_i) = \sum_{i=0}^{m-1} p_i \log_b \frac{1}{p_i} = - \sum_{i=0}^{m-1} p_i \log_b p_i.$$

Свойства энтропии

$H(A) \geq 0$, т. к. $p_i \leq 1$. Энтропия – величина неотрицательная.

$H(A) = 0$, если $p_i = 1, p_j = 0, i \neq j$. Энтропия равна нулю, если одно событие достоверно, а остальные имеют нулевую вероятность.

$$H_n(A_1 \cap A_2 \cap \dots A_n) = H(A_1) + H(A_2) + \dots H(A_n),$$

т. е. энтропия совместного наступления независимых событий есть сумма их энтропий. Например, энтропия слова складывается из энтропий его букв.

$$H_n(A) = nH(A),$$

т. е. энтропия последовательности из n независимых символов в n раз больше энтропии одного символа.

$$H_{\max}(A) = \log_b m, \text{ если } p_0 = p_1 = p_2 = \dots = p_i = \dots = p_{m-1} = \frac{1}{m},$$

т. е. энтропия максимальна, или, другими словами, источник создает максимальное количество информации, если символы независимы и равновероятны.

Допустим $p_i = \frac{1}{m}$, тогда

$$\begin{aligned} H(A) &= -\sum_{i=0}^{m-1} \frac{1}{m} \log_b \frac{1}{m} = -\frac{1}{m} \log_b \frac{1}{m} \sum_{i=0}^{m-1} 1 = \\ &= -\frac{1}{m} \left(\log_b \frac{1}{m} \right) m = -\log_b \frac{1}{m} = \log_b m = H_{\max}(A). \end{aligned}$$

Отсюда $H_{\max}(A) = \log_b m$.

Рассмотрим двоичный источник ($m = 2$) без памяти, т. е. источник, выдающий независимые символы $a_0 = 0$ и $a_1 = 1$ с вероятностями соответственно p_1 и p_2 (причем $p_1 + p_2 = 1$). Тогда энтропия такого двоичного источника независимых сообщений рассчитывается по формуле:

$$H(A) = -p_0 \log_2 p_0 - p_1 \log_2 p_1.$$

На рис. 6.2 показан график зависимости энтропии такого двоичного источника от вероятности символов.

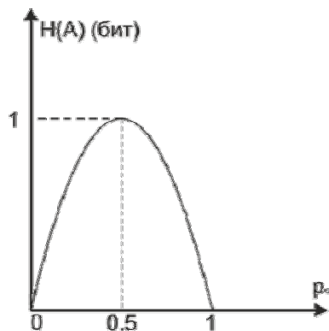


Рис. 6.2. Энтропия двоичного источника независимых событий

Энтропия источника зависимых дискретных сообщений

Сообщения (символы) называются зависимыми, если вероятность появления одного символа зависит от того, какие символы и с какими вероятностями ему предшествуют. В этом случае говорят, что источник имеет память и необходимо учитывать корреляционные связи между символами.

Источник с памятью описывается помимо ансамбля условными вероятностями. Любой стационарный дискретный источник с памятью можно рассматривать как однородную цепь Маркова связности n (n – память источника). На рис. 6.3 показан граф взаимосвязи между символами такого источника.

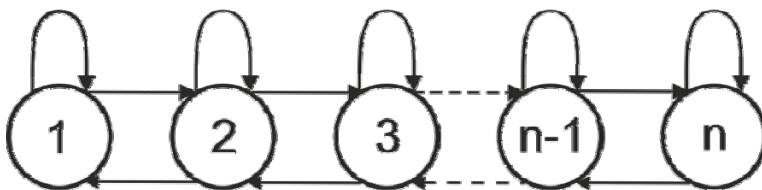


Рис. 6.3. Граф взаимосвязи символов дискретного источника
(n -связная цепь Маркова)

Рассмотрим простой марковский источник (рис. 6.4), который характеризуется односвязной цепью Маркова (т. е. памятью на одну позицию).

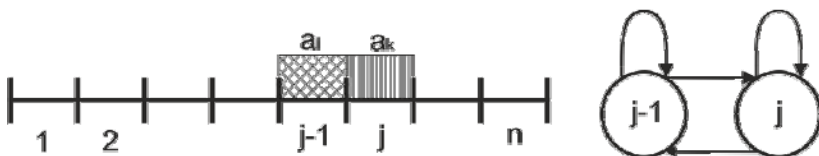


Рис. 6.4. Простой марковский источник (односвязная цепь Маркова)

Можно показать, что среднее количество информации, которое переносит символ на j -й позиции при влиянии предыдущего $(j-1)$ -го символа меньше, чем без учета этого влияния, т. е. условная энтропия меньше безусловной:

$$H(A_j / A_{j-1}) < H(A_j).$$

Отсюда следует важный вывод – корреляционные связи между символами уменьшают энтропию источника.

Избыточность источника

Избыточность источника характеризует степень недоиспользования его предельных характеристик и определяется следующим коэффициентом:

$$\rho = \frac{H_{\max} - H(A)}{H_{\max}} = 1 - \frac{H(A)}{H_{\max}},$$

где $H(A)$ – энтропия реального источника; H_{\max} – максимальная энтропия источника, если бы все символы были независимы и равновероятны.

Коэффициент избыточности ρ показывает, насколько можно сократить время передачи сообщений, если сократить избыточность $\rho = 1 - \mu$, где

$$\mu = \frac{H(A)}{H_{\max}},$$

а μ – информационная насыщенность источника сообщений. Величина $H_{\max}/H(A)$ определяет коэффициент сжатия текста.

Пример избыточности источника. Рассмотрим передачу по системе связи текста на русском языке. Практически объем алфавита $m = 32$. Энтропия реального источника, вырабатывающего текст на русском языке, составляет величину порядка $H(A) = 1.5$ бит.

Максимум энтропии можно было бы достичь, если бы символы были равновероятны, т. е. при

$$p_i = \frac{1}{32} H_{max} = \log_2 32 = 5 \text{ бит}.$$

Тогда коэффициент избыточности

$$\rho = 1 - \frac{H(A)}{H_{max}} = 1 - \frac{1.5}{5} = 0.7.$$

Таким образом, тексты на русском языке обладают значительной избыточностью. Например, в данном случае она составляет 70 %.

Каждый из известных естественных языков также обладает определенной избыточностью. Среди европейских языков русский обладает одним из самых высоких уровней избыточности. Об этом можно судить по размерам русского перевода английского текста. Обычно он примерно на 30 % больше. Если речь идет о стихотворном тексте, избыточность может быть даже в два раза выше.

Производительность источника

Производительность источника определяется количеством собственной информации, вырабатываемой им в единицу времени:

$$H'(A) = \frac{I(A, A)}{T} = \frac{H(A)}{T}.$$

Сравним два двоичных источника, причем оба выдают независимые символы, но отличаются распределением их

вероятностей. Численные значения такого сравнения приведены в табл. 6.1.

Таблица 6.1

Сравнение двух независимых двоичных источников

Характеристики источника	1-й источник	2-й источник
	$p_0 = 0,25; p_1 = 0,75$	$p_0 = 0,5; p_1 = 0,5$
Энтропия $H(A)$	0,8 (бит)	1 (бит)
Производительность $H'(A)$	0,8(бит)/Т(бит/сек)	1(бит)/Т(бит/сек)
Избыточность ρ	20%	0

Причины статистической избыточности сообщений и способы ее уменьшения

Сообщения отправителя обладают статистической избыточностью. С технической точки зрения это означает, что нужная получателю информация поступает от источника в неоптимальной форме. Отдельные сегменты сообщения источника повторяют друг друга, наиболее важные части сообщения смешаны с несущественными частями и т. п.

Основные причины малой энтропии и большой статистической избыточности источников сообщений заключаются в следующем.

Наличие корреляционных связей между символами

Для увеличения энтропии и уменьшения избыточности осуществляют ослабление корреляционных связей. Один из способов декорреляции символов – укрупнение сообщений, т. е. символами кода будут не отдельные буквы, а целые слова. Корреляционные связи между словами гораздо меньше, чем между символами. Следовательно, укрупнение сообщений увеличит энтропию. Например, сообщение «удовлетворительно», состоящее из 17 букв, можно закодировать одной цифрой «3». Таким образом, скорость передачи информации для этого частного случая увеличится в 17 раз, так как одно и то же количество информации будет передано не семнадцатью символами, а только одним.

Неравновероятность символов

Для увеличения энтропии и уменьшения избыточности надо перекодировать сообщения с учетом их вероятностей так, чтобы символы нового кода были бы практически равновероятны. Такое кодирование, учитывающее статистику передачи символов, носит название статистического (или эффективного) кодирования.

Информационные характеристики каналов связи

Рассмотрим основные информационные характеристики канала связи.

Количество информации, передаваемой по каналу связи, определяется как взаимная информация, содержащаяся в сигнале на выходе канала связи относительно сигнала на его входе.

Количество информации равно разности энтропии в сигнале, наблюдаемом на выходе канала, и энтропии шума в канале.

Скорость (R) передачи информации по каналу связи определяется как количество информации, передаваемой по каналу связи в единицу времени.

Пропускная способность (C) определяется как максимальная скорость передачи информации по каналу связи:

$$C = R_{\max} = \max_{\{p_i\}} R.$$

Максимум ищется по всем возможным вероятностным распределениям источника и всем возможным способам передачи и приема. Поэтому пропускная способность не зависит от свойств источника, а зависит только от помеховой (шумовой) ситуации в канале связи.

Примечание 1. Дискретный канал связи.

Пропускная способность двоичного симметричного канала связи зависит только от вероятности ошибки p в канале и, соответственно, равна:

$$C = 1 + p_e \log_2 p_e + (1 - p_e) \log_2 (1 - p_e) = 1 - H(E),$$

где $H(E)$ – энтропия шума E в канале связи.

Примечание 2. Непрерывный канал связи.

Шеннон показал, что шум определяет вероятность ошибки при передаче сигнала по каналу связи и в конечном итоге пропускную способность непрерывного канала связи. Он вывел формулу для пропускной способности непрерывного канала связи C :

$$C = F \log_2 \left(1 + \frac{P_s}{P_{ш}} \right),$$

где F – полоса пропускания канала, в котором действует сигнал с мощностью P_s и нормальный белый шум с мощностью $P_{ш}$.

Информационная эффективность канала связи определяется отношением скорости передачи информации к его пропускной способности:

$$\eta = R / C.$$

Эффективность η использования канала связи не может быть максимальной, т. е. равной 1, если скорость R передачи информации по каналу мала (из-за статистической избыточности информации) по сравнению с пропускной способностью C канала связи. Поэтому справедливо неравенство $0 < \eta < 1$.

Для повышения эффективности использования каналов связи применяют различные методы сокращения избыточности сообщений источника путем их эффективного кодирования.

Примеры

Пример 1

В течение 5 секунд было передано сообщение, объем которого составил 375 байт. Каков максимальный размер алфавита, с помощью которого записано сообщение, если скорость передачи составила 200 символов в секунду.

Решение. За 1 секунду передано $\frac{375 \cdot 8}{5} = 600$ бит.

Количество бит, приходящихся на символ, равно $\frac{600}{200} = 3$

бит/символ. Следовательно, размер алфавита $2^3 = 8$ символов.

Пример 2

В урне находятся 8 белых и 24 черных шара. Какое количество информации несет сообщение о том, что из урны достали белый шар? А черный шар? Объясните полученный результат.

Решение

Общее количество шаров в урне $24 + 8 = 32$. Вероятность достать белый шар $\frac{8}{32} = \frac{1}{4}$. Количество информации

$\log_2 4 = 2$ бита. Для черного шара аналогично количество информации $\log_2\left(\frac{24}{32}\right) = \log_2\left(\frac{4}{3}\right) = \log_2 4 - \log_2 3 = 2 - \log_2 3$.

Сообщение о том, что достали черный шар менее информативно, так как черных шаров больше и вероятность этого события выше.

Пример 3

Дан алфавит, состоящий из 4 символов, вероятности появления букв равны соответственно $p_1 = p_2 = 0,25$, $p_3 = 0,34$, $p_4 = 0,16$. Определить количество информации на символ сообщения, составленного из такого алфавита.

Решение

Количество информации на символ алфавита есть энтропия данного алфавита. Так как символы алфавита неравновероятны, то энтропия равна

$$H = - \sum_{i=1}^m p_i \log_2 p_i = -(2 \cdot 0.25 \log_2 0.25 + 0.34 \log_2 0.34 + 0.16 \log_2 0.16) = 1.9521 \text{ бит/символ.}$$

Пример 4

Имеется 27 монет, одна из которых фальшивая и легче всех остальных. Сколько потребуется взвешиваний на двухчашевых весах, чтобы однозначно найти фальшивую монету?

Решение

В этой задаче неудобно использовать бит в качестве основной единицы измерения информации. Двухчашевые весы могут принимать три состояния: перевешивает левая чаша, перевешивает правая чаша, весы в равновесии. Таким образом, одно взвешивание может уменьшить неопределенность в 3 раза, следовательно, будем использовать в качестве основной единицы измерения количества информации трит. По формуле Хартли $H = \log_3 27 = 3$ трита. Таким образом, в данном примере требуется три взвешивания.

Замечание: если в результате вычисления по формуле Хартли получилось нецелое число, а в задаче требуется указать целое, то результат следует округлить в большую сторону.

Пример 5

В некоторой стране автомобильный номер состоит из 7 символов. В качестве символов используют 18 различных букв и десятичные цифры в любом порядке. Каждый такой номер в компьютерной программе записывается минимально возможным и одинаковым целым количеством байтов, при этом используют посимвольное кодирование и все символы кодируются одинаковым и минимально возможным количеством битов. Определите объем памяти, отводимый этой программой для записи 60 номеров.

Решение

Установим количество бит, приходящихся на один символ. Всего используется $18 + 10 = 28$ символов. По формуле Хартли $H = \log_2 28 = 4,80$ бит, а после округления 5 бит.

Далее вычислим, какой объем памяти нужен на хранение одного номера. Поскольку номер состоит из 7 символов, то требуется $7 \cdot 5 = 35$ бит для его хранения. Однако, по условию, номер должен храниться в целом числе байт, следовательно, необходимо найти ближайшее большее кратное 8 число. Это число 40. Следовательно, каждый номер записывается $\frac{40}{8} = 5$ байтами. Для хранения 60 номеров требуется $60 \cdot 5 = 300$ байт памяти.

Контрольные вопросы

1. Что понимается под термином «количество информации».
2. Дайте определение энтропии.
3. Как определяется производительность источника?
4. Объясните назначение и способы кодирования источника.
5. Каковы причины статистической избыточности сообщений?
6. Какие существуют способы уменьшения статистической избыточности сообщений?

Литература

1. Акулиничев, Ю. П. Теория электрической связи : учеб. пособие для вузов / Ю. П. Акулиничев. – СПб. : Лань, 2010. – 233 с.
2. Добаткина, Н. В. Основы теории телекоммуникаций / Н. В. Добаткина. – М. : Издательский центр «Интерэкомс», 2010. – 176 с.
3. Теория электрической связи : учебник для вузов / под ред. Д. Д. Кловского. – М. : Радио и связь, 1999. – 432 с.
4. Крук, Б. И. Телекоммуникационные системы и сети : учеб. пособие : в 3 т. Т. 1 : Современные технологии / Б. И. Крук, В. Н. Попантопуло, В. П. Шувалов; под ред. проф.

- В. П. Шувалова. – М. : Горячая линия – Телеком, 2003. – 647 с.
5. Прокис, Д. Цифровая связь / Д. Прокис ; под ред. Д. Д. Кловского. – М. : Радио и связь, 2000. – 800 с.
 6. Основы цифровой обработки сигналов: учеб. пособие / Ю. А. Брюханов и др. – Ярославль: ЯрГУ, 2013. – 344 с.
 7. Гмурман, В. Е. Теория вероятностей и математическая статистика : учеб. пособие для бакалавров / В. Е. Гмурман ; М-во образования и науки РФ – 12-е изд. – М. : Юрайт, 2013. – 479 с.
 8. Кремер, Н. Ш. Теория вероятностей и математическая статистика : учебник для вузов. / Н. Ш. Кремер ; М-во образования РФ – 3-е изд., перераб. и доп. – М. : ЮНИТИ, 2009. – 551 с.
 9. Теория вероятностей и математическая статистика / сост. Л. П. Бестужева, Н. Л. Майорова; Яросл. гос. ун-т им. П. Г. Демидова, Науч.-метод. совет ун-та. Ч. 1 : практикум. – Б.м. : Б.и., 2012. – 47 с.
 10. Hsu, H. Analog and Digital Communications. Schaum's outline series / H. Hsu – McGraw Hill, 2003. – 330 p.
 11. Hsu, H. Probability, Random Variables and Random Processes. Schaum's outline series / H. Hsu – McGraw Hill, 1997. – 311 p.

7. ОСНОВЫ ТЕОРИИ КОДИРОВАНИЯ

Кодирование с контролем ошибок

В данном разделе рассматривается составление кодов, позволяющих реализовать передачу цифровой информации по зашумленному каналу связи. В зависимости от избыточности, содержащейся в коде, он может детектировать (обнаруживать) ошибки и исправлять их. Коды, которые могут только обнаруживать ошибки, называются **кодами с обнаружением ошибок**. Коды, способные исправлять обнаруженные ошибки, называются **корректирующими**. В настоящее время существует много видов таких кодов, которые широко используются в современных инфокоммуникационных системах для повышения их помехоустойчивости. В теории помехоустойчивого кодирования различают **блочные** и **сверточные** коды. В данном разделе описываются **двоичные** или **бинарные** коды. Они применяются к алфавиту, состоящему всего из двух элементов (0 и 1). Множество $\{0, 1\}$ далее обозначается как K .

Канальное кодирование

Простейшая схема канального кодирования представлена на рис. 7.1.



Рис. 7.1. Схема канального кодирования

Исходная двоичная последовательность, поступающая на канальный кодер, может быть получена как из кодера источника, так и непосредственно из источника сообщения. Канальный кодер вносит систематичную избыточность в поток цифровых данных путем добавления бит в сообщение, которые помогут обнаружить и исправить ошибки в сообщении на приемной стороне. Канальный декодер в приемнике использует

эти биты, чтобы принять решение о том, какое сообщение было на самом деле передано в канал. Совместное использование канальных кодера и декодера позволяет минимизировать воздействие шума, неизбежно присутствующего в канале связи.

7.1. Теорема о блочном кодировании (теорема Шеннона для канала с шумами)

Пусть дан цифровой сигнал X с энтропией $H(X)$ бит/символ и цифровой канал с пропускной способностью C_s . Тогда, если $H(X) \leq C_s$, то существует такой код, что исходное сообщение, закодированное с его использованием, может быть передано через канал со сколь угодно малой вероятностью ошибки.

Обратно: если $H(X) > C_s$, то не существует кода, способного передать сообщение через канал с произвольно малой вероятностью ошибки.

Замечание: данная теорема позволяет судить только о возможности построения кода, но не о способе его создания.

Блочные коды

Блочным называется код, все слова которого имеют одинаковую длину. В блочном коде двоичное сообщение сначала разбивается на блоки (части), длиной по k бит каждый. Затем каждый такой блок преобразуется в блоки, длиной n бит, при этом $n > k$. Получившийся блочный код называют (n, k) -блочным кодом. Любая из возможных 2^n последовательностей в каждом блоке называется n -кортежем. Множество всех возможных кортежей, определенных на множестве $K = \{0, 1\}$, записывают как K^n . Таким образом, канальный кодер выполняет преобразование

$$T: U \rightarrow V,$$

где U – множество двоичных слов, длиной k , V – множество двоичных слов, длиной n , где $n > k$. Каждое из 2^k исходных слов преобразуется в уникальное кодовое слово. Отношение k / n называют коэффициентом кодирования.

7.2. Линейные блочные коды

Двоичное поле

Множество $K = \{0, 1\}$ является двоичным полем. На нем определены две операции: сложение и умножение. Результат каждой из этих операций также принадлежит полю K . Правила выполнения этих операций следующие.

1. Сложение:

$$\begin{aligned}0 \oplus 0 &= 0, \\1 \oplus 1 &= 0, \\0 \oplus 1 &= 1 \oplus 0 = 1.\end{aligned}$$

2. Умножение:

$$\begin{aligned}0 * 0 &= 0, \\1 * 1 &= 1, \\0 * 1 &= 1 * 0 = 0.\end{aligned}$$

Линейные коды

Пусть $a = (a_1, a_2, \dots, a_n)$ и $b = (b_1, b_2, \dots, b_n)$ – два кодовых слова в некотором коде C . Сумма a и b , записанная как $a \oplus b$, определяется как $(a_1 \oplus b_1, a_2 \oplus b_2, \dots, a_n \oplus b_n)$. Код C называют *линейным*, если сумма двух слов этого кода также является словом кода C . Линейный код C должен содержать нулевое слово $0 = (0, 0, \dots, 0)$, так как $a \oplus a = 0$.

Вес Хэмминга и расстояние Хэмминга

Пусть c – кодовое слово длины n . Тогда *весом Хэмминга* кодового слова c , записываемого как $w(c)$, называют количество единичных бит в слове c .

Пусть a и b – кодовые слова длины n . *Расстояние Хэмминга* между словами a и b , записываемое как $d(a, b)$, равно числу позиций, в которых a и b различны.

Таким образом, вес Хэмминга слова c может быть определен как расстояние Хэмминга между словом c и нулевым словом:

$$w(c) = d(c, 0).$$

В свою очередь, расстояние Хэмминга может быть выражено в терминах веса:

$$d(a, b) = w(a \oplus b).$$

Минимальное расстояние

Минимальное расстояние d_{min} линейного кода C определяется как наименьшее расстояние между всевозможными парами кодовых слов кода C .

Исходя из определения линейного кода, можно вывести следующее утверждение.

Теорема 1

Минимальное расстояние в коде C равно наименьшему из весов Хэмминга среди всех кодовых слов этого кода.

7.3. Способность кода к обнаружению и исправлению ошибок

Минимальное расстояние d_{min} кода C определяет его способность к обнаружению и исправлению ошибок.

Теорема 2

Линейный код C с минимальным расстоянием d_{min} способен обнаружить до t ошибок тогда и только тогда, когда выполнено условие:

$$d_{min} \geq t + 1.$$

Теорема 3

Линейный код C с минимальным расстоянием d_{min} способен исправить до t ошибок тогда и только тогда, когда выполнено условие:

$$d_{min} \geq 2 * t + 1.$$

Данное выражение имеет геометрическую интерпретацию, приведенную на рис. 7.2. Здесь представлены две сферы Хэмминга радиуса t , с центрами в точках, соответствующих словам c_i и c_j . На рис. 7.2а описан случай, когда эти две сферы не пересекаются, что соответствует

условию $d(c_i, c_j) \geq 2 * t + 1$. Тогда передаваемое слово c_i , принятое как r , будет однозначно декодировано как c_i , так как r ближе всего к нему. В другом случае (рис. 7.2б) две сферы пересекаются, что соответствует условию $d(c_i, c_j) < 2 * t$, и декодер может принять ошибочное решение в пользу c_j .

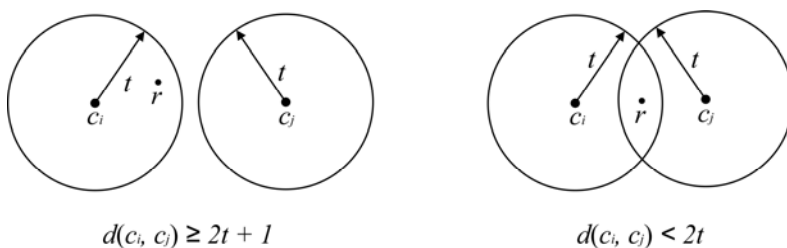


Рис. 7.2. Иллюстрация расстояния Хэмминга

Порождающая матрица

Пусть при рассмотрении блочного кода C кодовый вектор c и кодируемое сообщение d были определены следующим образом:

$$c = [c_1, c_2, \dots, c_n],$$

$$d = [d_1, d_2, \dots, d_k].$$

Если биты, присутствующие в сообщении d , появляются в некотором фрагменте кода c , то такой код называется систематическим, иначе – несистематическим. Таким образом, первые k бит из c соответствуют первым k битам сообщения d , оставшиеся $(n - k)$ бит – биты проверки, формируемые как линейная комбинация бит данных:

$$\begin{aligned} c_1 &= d_1 \\ c_2 &= d_2 \\ c_{k+1} &= p_{11}d_1 \oplus p_{12}d_2 \oplus \dots \oplus p_{1k}d_k \\ c_{k+2} &= p_{21}d_1 \oplus p_{22}d_2 \oplus \dots \oplus p_{2k}d_k \\ c_{k+m} &= p_{m1}d_1 \oplus p_{m2}d_2 \oplus \dots \oplus p_{mk}d_k, \end{aligned} \tag{7.1}$$

где $m = n - k$. Выражения (7.1) могут быть записаны в матричной форме:

$$c = d * G =$$

$$= \begin{bmatrix} d_1 & d_2 & \dots & d_k \end{bmatrix} \begin{bmatrix} 1 & 0 & \dots & 0 & p_{11} & p_{21} & \dots & p_{m1} \\ 0 & 1 & \dots & 0 & p_{12} & p_{22} & \dots & p_{m2} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 & p_{1k} & p_{2k} & \dots & p_{mk} \end{bmatrix}, \quad (7.2)$$

где

$$G = \begin{bmatrix} I_k & P^T \end{bmatrix}, \quad (7.3)$$

здесь I_k – матрица k -го порядка, P^T – транспонированная матрица, полученная из P ,

$$P = \begin{bmatrix} p_{11} & p_{21} & \dots & p_{m1} \\ p_{12} & p_{22} & \dots & p_{m2} \\ \dots & \dots & \dots & \dots \\ p_{1k} & p_{2k} & \dots & p_{mk} \end{bmatrix}.$$

Матрица G с размерами $k \times n$ называется порождающей матрицей. Отметим, что матрица G должна иметь k строк, n столбцов и ранг k , так как k строк из G линейно независимы.

Матрица проверки

Запишем матрицу H , имеющую размеры $k \times n$, определенную как

$$H = \begin{bmatrix} P & I_m \end{bmatrix},$$

где $m = n - k$ и I_m – матрица m -го порядка. Тогда

$$H^T = \begin{bmatrix} P^T \\ I_m \end{bmatrix}. \quad (7.4)$$

Используя выражения (7.3) и (7.4) имеем

$$GH^T = \begin{bmatrix} I_k & P^T \end{bmatrix} \begin{bmatrix} P^T \\ I_m \end{bmatrix} = P^T + P^T = O, \quad (7.5)$$

где O – матрица размера $k \times m$, заполненная нулями. Из выражений (7.2) и (7.5) имеем

$$cH^T = dGH^T = 0, \quad (7.6)$$

где 0 – матрица размерами $l \times m$, заполненная нулями.

Матрица H называется *матрицей проверки* кода C . Отметим, что ранг матрицы H равен $m = n - k$, и ее строки линейно независимы.

Минимальное расстояние d_{\min} линейного кода C сильно связано со структурой его матрицы проверки. Это утверждение можно сформулировать при помощи следующей теоремы.

Теорема 4

Минимальное расстояние d_{\min} линейного блочного кода C равно минимальному числу строк в H^T , сумма которых равна 0, здесь H^T – транспонированная матрица проверки H кода C .

Синдромное декодирование

Пусть r – принятое слово длины n , а c – кодовое слово длины n , переданное через зашумленный канал связи. Тогда $r = c \oplus e$, где e – вектор ошибок. Отметим, что $e = r + c$.

Будем считать, что допущена одна ошибка в i -й позиции, тогда вектор e может быть представлен в виде:

$$e = \begin{bmatrix} 0 & \dots & 0 & 1 & 0 & \dots & 0 \end{bmatrix}.$$

i -я позиция

Следующим шагом вычислим rH^T и получим

$$rH^T = (c \oplus e)H^T = cH^T \oplus eH^T = eH^T = s,$$

т. е. частный случай выражения (7.6), s называется синдромом принятого слова r .

Таким образом, используя s и учитывая, что eH^T является i -й строкой H^T , становится возможным определение позиции ошибки путем сравнения s и строк H^T . Такой способ декодирования посредством сравнения называется *синдромным декодированием*. Следует отметить, что не все вектора ошибок могут быть корректно декодированы синдромным декодированием. Нулевой вектор ошибок означает, что принятое слово является словом кода и является безошибочным.

При помощи синдромного декодирования (n, k) -линейного блочного кода можно исправить до t ошибок в каждом принятом кодовом слове только при выполнении условия Хэмминга:

$$2^{n-k} \geq \sum_{i=0}^t \binom{n}{i},$$

где

$$\binom{n}{i} = \frac{n!}{(n-i)!i!}.$$

Блочный код, для которого выполнено это условие, называется *совершенным кодом*. Совершенный код, способный исправить одну ошибку, называется кодом Хэмминга. Выполнение условия Хэмминга является необходимым, но не достаточным для построения линейного блочного кода, корректирующего t ошибок.

Циклические коды

Пусть $c = (c_0, c_1, \dots, c_{n-1})$ является кодовым словом кода C . Определим *циклический сдвиг* $\sigma(c)$ как

$$\sigma(c) = c^{(1)} = (c_{n-1}, c_0, c_1, \dots, c_{n-2}). \quad (7.7)$$

Для следующего циклического сдвига получим:

$$\sigma^2(c) = \sigma(\sigma(c)) = c^{(2)} = (c_{n-2}, c_{n-1}, c_0, c_1, \dots, c_{n-3}),$$

отметим, что $\sigma^n(c) = c$.

Код C называется *циклическим*, если циклический сдвиг каждого кодового слова также является кодовым словом кода C . Циклические коды – частный случай линейных блочных кодов.

Кодовые полиномы

Определим *кодový полином* $c(x)$, соответствующий слову c , следующим образом:

$$c(x) = c_0 + c_1x + \dots + c_{n-1}x^{n-1}. \quad (7.8)$$

Кодовый полином $c(x)$ является полиномом степени $(n-1)$, определенным над полем $K = \{0, 1\}$. Коэффициенты c_0, c_1, \dots, c_{n-1} являются элементами K . Таким образом, код C длины n может быть представлен набором полиномов над полем K степени не выше $n-1$. Полиномы над полем K складываются и умножаются обычным способом, за исключением случая $1 \oplus 1 = 0$, вместо которого имеется $x^k \oplus x^k = 0$. Множество полиномов над K записывается как $K[x]$.

Пусть $f(x)$ и $h(x)$ принадлежат $K[x]$ и $h(x) \neq 0$. Тогда существуют единственные полиномы $q(x)$ и $r(x)$, принадлежащие $K[x]$, такие что

$$f(x) = q(x)h(x) + r(x),$$

с $r(x) = 0$ либо порядком $r(x)$ меньшим, чем порядок $h(x)$. Полином $q(x)$ называется *частным*, а полином $r(x)$ – *остатком*. Процедура нахождения $q(x)$ и $r(x)$ представляет собой деление столбиком коэффициентов полиномов по модулю 2.

Если $r(x)$ является остатком при делении $f(x)$ по модулю $h(x)$, то можно записать:

$$r(x) = f(x) \bmod h(x).$$

Если $c(x) \bmod h(x) = r(x) = b(x) \bmod h(x)$, то говорят, что $c(x)$ и $b(x)$ равны по модулю $h(x)$, что записывается в виде

$$c(x) = b(x) \bmod h(x).$$

Также

$$x^n = 1 \bmod (1 + x^n). \quad (7.9)$$

Порождающий полином

Кодовый полином $c^{(l)}(x)$ в соответствии с $\sigma(c)$ (выражение (7.7)) можно записать в виде

$$c^{(l)}(x) = c_{n-1} + c_0 x + c_1 x^2 + \dots + c_{n-2} x^{n-1}.$$

Из выражения (7.8) следует, что

$$xc(x) = c_0 x + c_1 x^2 + \dots + c_{n-1} x^n.$$

Используя (7.9), получим

$$c^{(l)}(x) = xc(x) \bmod (1 + x^n).$$

Аналогично

$$c^{(i)}(x) = x^i c(x) \bmod (1 + x^n).$$

Теорема 5

Если $g(x)$ полином степени $(n-k)$ с фактором $1 + x^n$, то $g(x)$ порождает (n, k) циклический код C с кодовым полиномом $c(x)$ для информационного слова $d = (d_0, d_1, \dots, d_{k-1})$, полученного как:

$$c(x) = d(x)g(x),$$

где $d(x) = d_0 + d_1 x + \dots + d_{k-1} x^{k-1}$ – полином, соответствующий информационному слову.

Полином $g(x)$ называют *порождающим полиномом*. Отметим, что результирующий код C не является систематическим.

Полином проверки

Пусть $h(x)$ – полином степени k с фактором $1 + x^n$. Тогда $h(x)$ является полиномом проверки для (n, k) циклического кода C . Полином проверки $h(x)$ и порождающий полином $g(x)$ кода C связаны соотношением

$$g(x)h(x) = 0 \bmod (1 + x^n)$$

или

$$g(x)h(x) = 1 + x^n.$$

Синдромный полином

Пусть $r(x)$ является кодовым полиномом принятого слова, а $c(x)$ – кодовый полином переданного слова. Тогда

$$r(x) = c(x) + e(x),$$

где $e(x)$ – вероятный полином ошибки. Синдромный полином $s(x)$ определяется как

$$s(x) = r(x) \bmod g(x).$$

Предполагая, что $g(x)$ имеет степень $n-k$, $s(x)$ будет иметь степень меньше, чем $n-k$ и будет соответствовать двоичному слову s длиной $n-k$. Так как $c(x) = d(x)g(x)$, имеем

$$s(x) = e(x) \bmod g(x). \quad (7.10)$$

Таким образом, синдромный полином зависит только от ошибки. Отметим, что если степень полинома $e(x)$ меньше степени полинома $g(x)$, то $e(x) = e(x) \bmod g(x)$ и тогда $s(x) = e(x)$.

Следующая теорема определяет условия существования единственного синдромного полинома.

Теорема 6

Пусть C – некоторый циклический код с минимальным расстоянием d_{min} . Каждый полином ошибки с весом, меньше

чем $\frac{1}{2}d_{min}$, имеет уникальный синдромный полином.

Отсюда следует, что проблема исправления ошибок сводится к нахождению уникального полинома $e(x)$ с наименьшим весом, удовлетворяющим условию (7.10). Эта задача может быть решена следующим образом. Для каждого корректного $e(x)$ вычислено и занесено в табл. 7.1 соответствующее $s(x)$. Эта таблица носит название *таблица вычисления синдромов*. Декодер находит полином ошибки $e(x)$ путем вычисления $s(x)$ через $r(x)$, затем находит $s(x)$ в таблице вычисления синдромов и ставит в соответствие нужный $e(x)$.

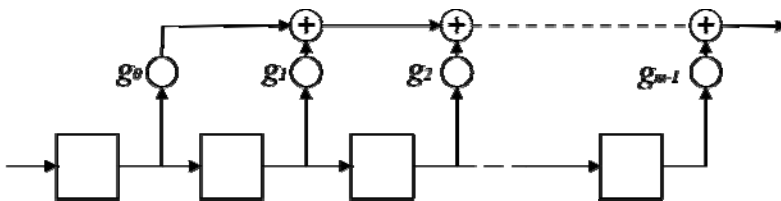
Таблица 7.1

Таблица вычисления синдромов

$e(x)$	$s(x)$
1	$1 \bmod g(x)$
x	$x \bmod g(x)$
x^2	$x^2 \bmod g(x)$
\vdots	\vdots
$1+x$	$1+x \bmod g(x)$
$1+x^2$	$1+x^2 \bmod g(x)$
\vdots	\vdots

Реализация

Одним из преимуществ циклических кодов является их простая реализация. Циклические кодеры могут быть реализованы на сдвиговых регистрах. Эти устройства состоят из m регистров (элементов задержки) и часов, которые контролируют сдвиг данных в регистрах. Например, m -разрядный сдвиговый регистр, представленный на рис. 7.3, состоит из m регистров.

Рис. 7.3. m -разрядный сдвиговый регистр

Выходом m -разрядного регистра является линейная комбинация, состоящая из содержимого регистров. Она может быть описана с использованием коэффициентов g_0, g_2, \dots, g_{m-1} , каждый из которых принадлежит полю $K = \{0, 1\}$. Пусть c_t – выход в момент времени t , тогда

$$c_i = g_0 X_0(t) + g_1 X_1(t) + \dots + g_{m-1} X_{m-1}(t),$$

где $X_i(t)$ – значение, содержащееся в регистре i в момент времени t .

Имея фиксированный порождающий полином $g(x)$ степени $n-k$ для (n, k) линейного циклического кода C , можно построить $(n-k+1)$ -разрядный сдвиговый регистр с порождающим полиномом $g(x)$ для реализации кодирования информационного кодового полинома $d(x)$.

Полиномиальное деление и, как следствие, декодирование циклического кода можно реализовать при помощи устройства, называемого сдвиговый регистр с обратной связью (ОС). Это сдвиговый регистр с выходом, направленным на вход, как показано на рис. 7.4.

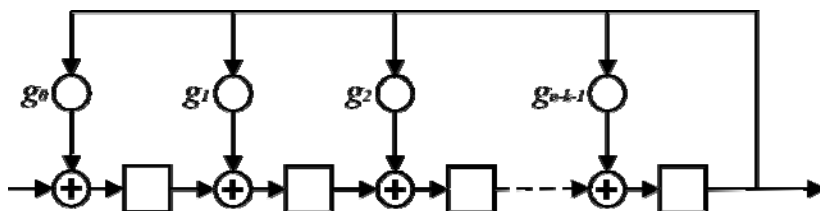


Рис. 7.4. m -разрядный сдвиговый регистр с обратной связью

Подача полинома $r(x) = r_0 + r_1x + \dots + r_{n-1}x^{n-1}$ на вход сдвигового регистра с ОС с порождающей матрицей $g(x)$, начиная с коэффициентов высших порядков, эквивалентна делению $r(x)$ на $g(x)$. Выход после n тактов будет частным от деления (начиная с коэффициентов высших порядков), а содержимое регистров – частным от деления (коэффициенты высших порядков в конце).

Порождающая матрица

Существует множество порождающих матриц G для линейных циклических кодов. Строки простейшей порождающей матрицы представляют собой коэффициенты порождающего полинома $g(x)$ и их первые $k-1$ циклических сдвигов

$$G = \begin{bmatrix} g(x) \\ xg(x) \\ \vdots \\ x^{k-1}g(x) \end{bmatrix}. \quad (7.11)$$

Получившаяся порождающая матрица G , скорее всего, будет иметь вид, отличный от выражения (7.3). Обычно требуется провести некоторые действия со строками (7.11), чтобы привести ее к указанному виду.

7.4. Специальные циклические коды. Код Голея

Единственный известный совершенный код, исправляющий 3 ошибки, это (23, 12) – код Голея. Минимальное расстояние в коде равно 7. Код порождается полиномами:

$$g_1(x) = 1 + x + x^5 + x^6 + x^7 + x^9 + x^{11}$$

или

$$g_2(x) = 1 + x^2 + x^4 + x^5 + x^6 + x^{10} + x^{11}.$$

Заметим, что

$$1 + x^{23} = (1 + x)g_1(x)g_2(x).$$

Коды Боуза-Чоудхури-Хоквингема

Коды Боуза-Чоудхури-Хоквингема (БЧХ-коды) – наиболее эффективные из всех известных циклических кодов с коррекцией ошибок. Они обладают гибкостью в выборе параметров, таких как длина блока и коэффициент кодирования. Большинство применяемых кодов БЧХ характеризуется следующими параметрами: для любого целого $m \geq 3$ и t существует БЧХ код со следующими параметрами:

Длина блока	$n = 2^m - 1$
Длина сообщения	$n - k \leq mt$
Минимальное расстояние	$d_{\min} \geq 2t + 1$

Отметим, что код Хэмминга, исправляющий одну ошибку, может быть описан кодом БЧХ с $t = 1$ и $n - k = m$.

Коды Рида-Соломона

Коды Рида-Соломона (РС коды) – важный подкласс недвоичных БЧХ кодов. Кодеры, основанные на РС кодах, работают с несколькими битами. Коды имеют большие расстояния и полезны в ситуациях, когда происходят всплески интенсивности ошибок.

Сверточные коды

Два основных используемых типа кодов – это блочные и сверточные коды. В блочном (n, k) коде информационная последовательность группируется в k -битовые блоки и кодируется в n -битовые кодовые слова с добавлением $(n - k)$ бит проверки. Каждому n -битовому блоку ставится в строгое соответствие k -битовый блок. В (n, k, m) сверточных кодах кодер также преобразует k -битовую последовательность в n -битовую, однако при этом учитывает не только k бит, но и $m - 1$ предыдущих k -битовых блоков.

Таким образом, важной особенностью сверточных кодов, отличающей их от блочных, является кодер с памятью. Целый параметр m называется *длиной кодового ограничения*. На практике n и k являются небольшими целыми числами, а m контролирует избыточность.

Общая схема кодера для (n, k, m) сверточного кода приведена на рис. 7.5. Он состоит из mk -разрядного сдвигового регистра и n двоичных сумматоров. В каждый момент времени k бит сдвигаются в первые k разрядов в регистрах, все биты в регистрах сдвигаются вправо, а выходы с n сумматоров последовательно считываются, формируя закодированную выходную последовательность.

Важным частным случаем является случай, когда $k = 1$. Тогда входная последовательность не делится на блоки и может обрабатываться непрерывно. В дальнейшем будет рассмотрен только этот случай.

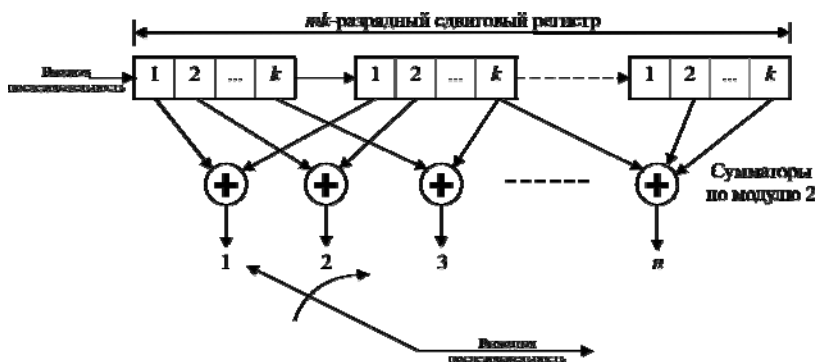


Рис. 7.5. Схема (n, k, m) сверточного кодера

Сверточные коды часто используются на практике. Существует несколько способов их представления: таблица истинности, полиномиальное представление, диаграмма состояний, дерево, решетчатая диаграмма. На рис. 7.6 приведена иллюстрация простого $(2, 1, 3)$ сверточного кодера с $n = 2$, $k = 1$ и $m = 3$. В каждый момент времени биты данных сдвигаются в регистр, а два кодированных бита записываются в выходную последовательность.

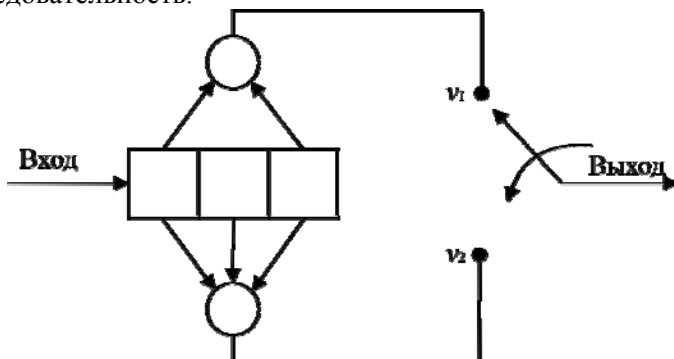


Рис. 7.6. Схема $(2, 1, 3)$ сверточного кодера

В дальнейшем будем использовать этот кодер для описания различных представлений сверточных кодов.

Таблица истинности

Обозначим содержимое каждого из разрядов сверточного кодера с рис. 7.6 как X_0 , X_1 , X_2 и выходные значения как v_1 и v_2 . Тогда

$$v_1 = X_0 \oplus X_2,$$

$$v_2 = X_0 \oplus X_1 \oplus X_2.$$

Такая схема осуществляет кодирование все время, пока биты данных проходят через разряды регистра. Положим входную последовательность равной $d = (1\ 0\ 1)$. Тогда для кодера на рис. 7.6 таблица истинности будет иметь следующий вид:

Время	Вход	X_0	X_1	X_2	v_1	v_2
1	1	1	0	0	1	1
2	0	0	1	0	0	1
3	1	1	0	1	0	0
4	0	0	1	0	0	1
5	0	0	0	1	1	1
6	0	0	0	0		

Таким образом,

Входная последовательность: 1 0 1

Выходная последовательность: 11 01 00 01 11

Импульсный отклик кодера

Выходная последовательность кодера при подаче на него одного бита, равного «1», называется импульсным откликом кодера. Пусть такой сигнал подается на вход сверточного кодера с рис. 7.6. Тогда таблица истинности выглядит следующим образом:

Вход	X_0	X_1	X_2	v_1	v_2
1	1	0	0	1	1
0	0	1	0	0	1
0	0	0	1	1	1

Входная последовательность:	1 0 0
Выходная последовательность:	11 01 11

Отметим, что импульсный отклик v_1 равен (1 0 1), а v_2 – (1 1 1). Для входной последовательности $d = (1\ 0\ 1)$ выходная последовательность может быть найдена как линейная сумма сдвинутых импульсных откликов:

Вход d	Выход				
1	11	01	11		
0		00	00	00	
1			11	01	11
Сумматоры:	11	01	00	01	11

Результирующая последовательность уже была получена выше. Это показывает, что сверточные коды являются линейными. Отметим, что выходные последовательности на v_1 и v_2 являются свертками входной последовательности и соответствующего импульсного отклика. Именно по этой причине данный тип кодов называется сверточным.

Полиномиальное представление

Сравнивая рис. 7.6 и рис. 7.3, заметим, что возможно записать порождающий полином $g_1(x)$ для верхнего сумматора и порождающий полином $g_2(x)$ для нижнего сумматора следующим образом:

$$g_1(x) = 1 + x,$$

$$g_2(x) = 1 + x + x^2.$$

Для выходной последовательности имеем:

$$c(x) = d(x)g_1(x) \text{ с чередованием } d(x)g_2(x).$$

Для $d = (1\ 0\ 1)$ имеем $d(x) = 1 + x^2$ и

$$\begin{aligned} d(x)g_1(x) &= (1 + x^2)(1 + x) = \\ &= 1 + x^4 = 1 + 0x + 0x^2 + 0x^3 + 1x^4, \end{aligned}$$

$$d(x)g_2(x) = (1+x^2)(1+x+x^2) =$$

$$= 1 + x + x^3 + x^4 = 1 + 1x + 0x^2 + 1x^3 + 1x^4$$

и $c(x) = (1,1) + (0,1)x + (0,0)x^2 + (0,1)x^3 + (1,1)x^4$.

Таким образом, получился тот же результат $c = (1\ 1\ 0\ 1\ 0\ 0\ 0\ 1\ 1\ 1)$.

Представление состояний и диаграмма состояний

В состоянии $(n, 1, m)$ сверточный кодер определяет содержимое первых $m-1$ регистров. Таким образом, кодер представим в виде конечного автомата с $m-1$ состояниями. Зная состояние в текущий момент и следующий бит на входе, можно определить, какой будет следующий бит на выходе. Нулевым состоянием является состояние, в котором каждый из всех сдвиговых регистров содержит 0. Переход между состояниями определяется следующим входным битом (0 или 1). Кодер может иметь 2^{m-1} возможных состояний. Диаграмма состояний, соответствующая кодеру с рис. 7.6, приведена на рис. 7.7. Так как $m = 3$, то существует 4 возможных состояния, обозначенных как $a = 0\ 0$, $b = 1\ 0$, $c = 0\ 1$, $d = 1\ 1$.

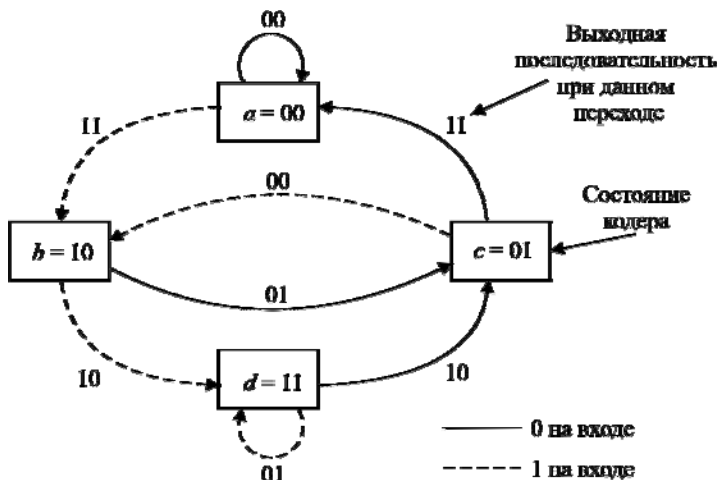


Рис. 7.7. Диаграмма состояний, соответствующая сверточному кодеру

Возможно только два перехода из каждого состояния, соответствующих 0-му и 1-му входным битам. Стрелки показывают направление перехода. Переход из одного состояния в другое, соответствующий нулевому биту, представлен сплошной стрелкой, первому биту – пунктирной. Следующие биты определяют дальнейшее перемещение, а путь, по которому оно происходит, определяет выходные биты.

Кодирование начинается из положения a (все регистры в 0). Кодовое слово, обусловленное входными битами, определяет дальнейший путь через диаграмму состояний и составляет выходную последовательность. После последнего ненулевого элемента входной последовательности кодер переходит в нулевое состояние через последовательность длины $m-1$, состоящей из нулей. Например, для исходной последовательности $d = (1\ 0\ 1)$ путь будет $a\ b\ c\ b\ c\ a$, и, соответственно, выход – $(1\ 1\ 0\ 1\ 0\ 0\ 0\ 1\ 1\ 1)$.

Дерево состояний

Дерево состояний добавляет временное измерений к диаграмме состояний. Дерево состояний, соответствующее сверточному кодеру с рис. 7.6, приведено на рис. 7.8. Вертикальные линии называются узлами дерева (или нодами), горизонтальные – ветвями дерева. Выходная последовательность для каждого входа показана на ветвях дерева.

Структура дерева подчиняется следующему правилу: переход вверх соответствует нулю на входе, переход вниз – единице. Следуя этому правилу, видим, что путь, выделенный жирным, соответствует входной последовательности $1\ 0\ 1\ 0\ 0$. Заметим, что структура дерева повторяется за третьим узлом. Это связано с тем, что выход зависит только от двух предыдущих состояний и текущего входа. Кроме того, отметим, что количество возможных путей растет экспоненциально с ростом времени.

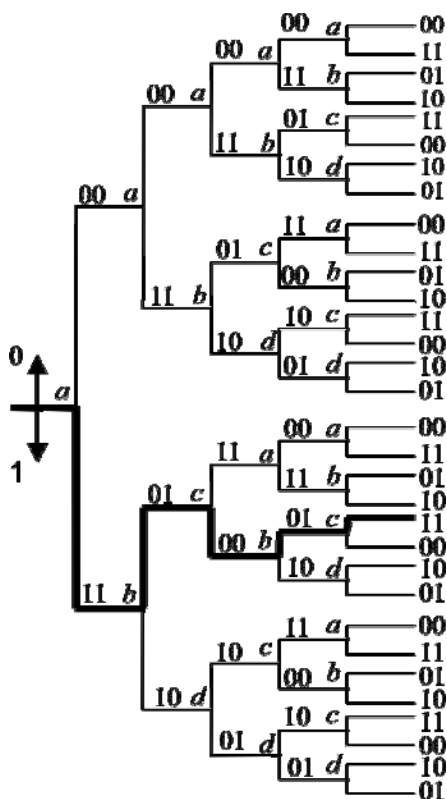


Рис. 7.8. Дерево состояний

Решетчатая диаграмма

Повторение структуры дерева состояний приводит к возможности иначе изобразить возможные состояния кодера. Это можно сделать в виде решетчатой диаграммы, приведенной на рис. 7.9. Четыре состояния расположены вдоль вертикальной оси, переходы между состояниями с течением времени представлены вдоль горизонтальной оси. При изображении решетчатой диаграммы использованы те же правила, что и при построении диаграммы состояний: сплошной линией показан переход при поступлении нуля на вход, пунктирной – при поступлении единицы. Решетчатая диаграмма в этом примере

демонстрирует повторяющийся характер структуры решетки после глубины решетки $m = 3$ (в момент t_4).

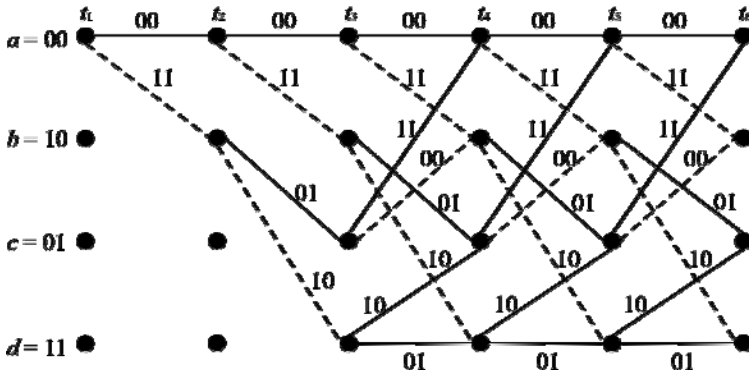


Рис. 7.9. Решетчатая диаграмма

Повторяющаяся структура решетчатой диаграммы еще более упрощает представление кодера. Таким образом, дерево состояний, диаграмма состояний и решётчатая диаграмма представляют отношение вход-выход в кодере сверточного кода и используются для нахождения наиболее правдоподобного пути, с помощью которого код был создан. Отметим, что решетчатая диаграмма наиболее полезна в этом подходе.

Декодирование сверточных кодов

Как в блочных кодах, так и в сверточных расстояние является важной характеристикой, определяющей способность кода к коррекции ошибок. Наиболее важной характеристикой расстояния в сверточных кодах является свободное расстояние, обозначаемое как d_{free} . Свободное расстояние сверточного кода определяется как минимальное расстояние между двумя словами в коде. Сверточный код со свободным расстоянием d_{free} может исправлять до t ошибок тогда и только тогда, когда выполнено условие

$$d_{free} \geq 2t + 1.$$

Так как сверточный код является линейным, то свободное расстояние является минимальным весом кодового слова,

порожденного ненулевой последовательностью. Таким образом, свободное расстояние может быть получено из решетчатой диаграммы сверточного кода.

Для примера рассмотрим решетчатую диаграмму, приведенную на рис. 7.9, которая соответствует кодеру с рис. 7.6. Для начала преобразуем решетчатую диаграмму так, как показано на рис. 7.10, переобозначив ветви диаграммы их весами вместо кодовых символов. Положим, что все пути начинаются и заканчиваются в нулевом состоянии и никогда не проходят через него насквозь. Видно, что существует только один путь $a b c a$ длиной 5. Аналогично существует два пути – $a b d c a$ и $a b d d c a$, имеющих вес 6. Таким образом, свободное расстояние в данном примере равно 5.

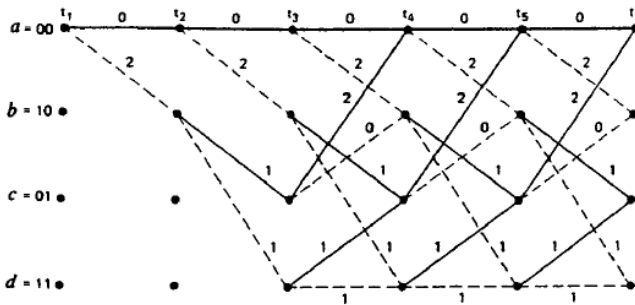


Рис. 7.10. Решетчатая диаграмма с весами

Декодирование на основе максимального правдоподобия

Если все входные последовательности равновероятны, то декодер выберет значение \hat{c} , если

$$P(r|\hat{c}) = \max_{c_i} P(r|c_i),$$

где r – принятое сообщение, а c_i – один из возможных вариантов переданного сообщения.

Такой декодер называется *декодером максимального правдоподобия*. Условная вероятность $P(r|c_i)$ называется *функцией правдоподобия*. Отметим, что для двоичного

симметричного канала декодер максимального правдоподобия сводится к декодеру минимальной дистанции. Решающее правило для такого декодера: выбрать такое \hat{c} , которое обладает минимальным расстоянием между принятой последовательностью r и переданной последовательностью c .

Алгоритм декодирования Витерби

Декодер максимального правдоподобия включает в себя поиск по всему кодовому пространству и в основном не эффективен ввиду высокой вычислительной стоимости. Однако алгоритм, предложенный Витерби, повышает вычислительную эффективность декодера максимального правдоподобия для кодов с малой длиной кодового ограничения. При помощи его можно декодировать сверточный код посредством выбора пути на решетчатой диаграмме.

Алгоритм Витерби дает возможность декодировать сверточные коды и фактически является алгоритмом динамического программирования.

Решетчатые диаграммы декодера Витерби подобны соответствующим диаграммам для кодера, но дополнены парой чисел-метрик для каждого узла (состояния) декодера. Метрика представляет собой расстояние между последовательностью битов на входе декодера и каждой из последовательностей, соответствующих возможным путям, входящим в узлы. Верхнее число – это метрика верхнего входящего в узел пути, нижнее – метрика нижнего входящего пути.

Декодер работает следующим образом. На каждом шаге (такте) для каждого узла рассчитываются метрики двух входящих в этот узел путей и путь с большей метрикой отбрасывается. Число выживших путей на предшествующих тактах уменьшается по мере их удаления от текущего. Эта процедура продолжается до тех пор, пока на очередном шаге не останется единственный общий фрагмент выживших путей в предшествующих тактах. По полученному таким образом фрагменту восстанавливается последовательность кодированных битов, а уже по ней – последовательность информационных битов. Установлено, что декодированная

таким образом последовательность битов по критерию максимального правдоподобия оказывается наиболее близкой к переданной последовательности.

Декодирование с жестким и мягким решениями

При рассмотрении предыдущих способов декодирования по умолчанию предполагалось, что на вход декодера поступает последовательность битов с выхода демодулятора. Эта двоичная последовательность имеет место благодаря наличию в выходной части демодулятора решающего устройства, осуществляющего двоичное квантование отсчетов реакции коррелятора или согласованного фильтра, входящих в схему демодулятора. Отметим, что двоичное квантование по своей сути является одноразрядным двоичным аналого-цифровым преобразованием отсчетов сигнала. Такую схему принятия решения называют жесткой, а последующее декодирование – декодированием с жестким решением.

Демодулятор и декодер в жесткой схеме принятия решения работают независимо друг от друга, а выход демодулятора одновременно является и выходом двоичного канала. Демодулятор формирует из принимаемой последовательности модулированных сигналов с помехами оценки каждого переданного бита, а декодер – оценки кодовых последовательностей. При использовании жесткого решения теряется информация о его надежности, т. е. о степени близости (удаленности) аналоговых отсчетов по отношению к порогу в РУ.

Идея декодирования с мягким решением состоит в том, чтобы в процессе декодирования использовать информацию о надежности решения демодулятора для повышения помехоустойчивости системы связи. В этом случае выход демодулятора оставляют аналоговым (канал с АБГШ). Аналоговые отсчеты с него вводят непосредственно в декодер для формирования оценок кодовых последовательностей без промежуточных оценок отдельных битов.

В общем случае мягкое решение более трудоемко, чем жесткое, поскольку требует выполнения операций не с

логическими величинами, а с действительными числами. На практике при цифровой реализации аналоговые отсчеты квантуются с конечной точностью и представляются двоичными числами конечной разрядности. Установлено, что достаточно иметь трехбитное квантование для достижения эффективности системы, близкой к предельной, которая достигается при использовании вычислений с неограниченной точностью.

Отметим, что в декодере Витерби переход к декодированию с мягким решением достигается заменой хэмминговой метрики на евклидовую в процедуре определения выживших путей на решетчатой диаграмме декодера.

Кодированная модуляция

В основе данного вида модуляции лежит введенная в 1974 году Мэсси концепция, рассматривающая кодирование и цифровую модуляцию как единый процесс обработки сигналов. Суть ее в том, чтобы найти наилучшее отображение двоичных данных, подаваемых на вход модулятора, в сигналы на его выходе. Множество сигналов модулятора графически можно изобразить в виде набора сигнальных точек на комплексной плоскости. Обычно они называются сигнальными созвездиями. Ключевым моментом является способ связи (нумерации) сигнальных точек с двоичными данными на входе модулятора. Установлено, что это можно сделать таким образом, чтобы повысить помехоустойчивость системы без увеличения средней мощности сигналов и расширения их спектра. Один из таких подходов был предложен Ангербеком в виде решетчатой кодированной модуляции. В отечественной литературе для ее обозначения чаще используется термин сигнально-кодовые конструкции (СКК). Обобщенная структура кодера Ангербека приведена на рис. 7.11.

Рассмотрим конкретный пример реализации указанного кодера с использованием сверточного кода $(1/2)$ и модулятора ФМ-8 (рис. 7.12).



Рис. 7.11. Кодер Ангербока

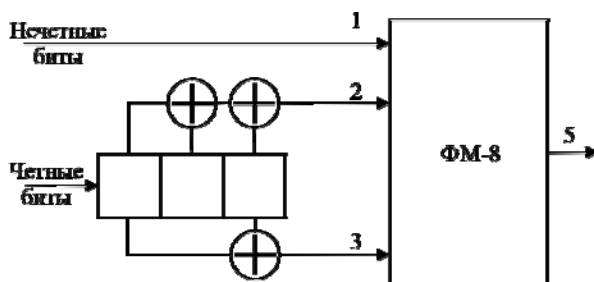


Рис. 7.12. Кодированная модуляция

Модулятор формирует сигнальное созвездие из восьми равномерно распределенных по окружности точек (рис. 7.13). Каждый модулированный сигнал служит для одновременной передачи трех битов – двух кодированных с выхода сверточного кодера и одного некодированного. Действию формирователя СКК подвергаются четные биты информационной последовательности. Нечетные биты при этом остаются некодированными. Они более уязвимы к искажениям в канале по сравнению с кодированными, и, следовательно, целесообразно присваивать им модулированные сигналы с наибольшими расстояниями между ними. Исходя из этого, сигнальное созвездие для случая ФМ-8 можно разбить на 4 разных созвездия, каждое из которых содержит по две

противоположно расположенные сигнальные точки (000 – 100, 001 – 101, 010 – 110, 011 – 111). Далее, конкретный выбор созвездий из этой четверки можно связать с двумя кодированными битами (дибитами) на выходе сверточного кодера (00, 01, 10, 11), а сигнальную точку в каждом из них будет определять некодированный бит. Указанная нумерация сигнальных точек выбрана в данном формирователе сигнально-кодовой конструкции (рис. 7.13).

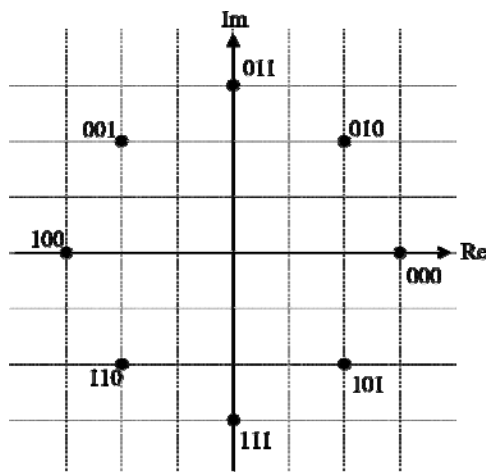


Рис. 7.13. Формирование сигнально-кодовой конструкции

Комбинированные коды

Установлено, что более мощные коды с длинными блоками и большими кодовыми расстояниями при относительно простой процедуре их декодирования можно строить следующим образом. Можно объединить несколько коротких кодов (часто, с перемежителями) в единую конструкцию путем их суммирования, перемножения, последовательного и параллельного соединения, каскадирования и других операций. Такой подход используется при разработке итеративных, каскадных и турбокодов, которые широко используются в современных системах связи.

Примеры

Пример 1

Докажите, что код $C = \{000, 111\}$ является линейным.

Решение. Рассмотрим все возможные варианты сумм, которые можно получить в коде C :

$$(000) \oplus (000) = (000); (111) \oplus (000) = (111);$$

$$(000) \oplus (111) = (111); (111) \oplus (111) = (000).$$

Так как все возможные комбинации также являются словами кода C , то код C – линейный.

Пример 2

Докажите, что минимальное расстояние в блочном линейном коде равно минимальному весу Хэмминга среди ненулевых кодовых слов.

Решение. Так как

$$d(c_i, c_j) = w(c_i \oplus c_j),$$

то

$$d_{\min} = \min_{c_i \neq c_j} d(c_i, c_j) = \min_{c_i \neq c_j} w(c_i \oplus c_j).$$

Тогда из определения сверточного кода следует

$$d_{\min} = \min_{c \neq 0} w(c).$$

Пример 3

Пусть код $C = \{000, 111\}$. Покажите, что он способен исправить одну ошибку, а также не способен исправить вектор ошибки (110).

Решение. Так как $d_{\min} = \min_{c \neq 0} w(c)$, то $d_{\min} = 3$, и $\frac{1}{2}(d_{\min} - 1) = 1$, то получаем, что C – код, способный корректировать одну ошибку. Пусть вектор ошибки получен при передаче слова $c = (000)$, тогда

$$d(c, c \oplus e) = d(000, 110) = 2,$$

$$d(111, c \oplus e) = d(111, 110) = 1.$$

Таким образом, $c \oplus e$ ближе к слову (111), чем к (000), следовательно, код C не способен исправить вектор ошибки (110).

Пример 4

Покажите, что код $C = (000, 100, 011, 111)$ не является циклическим.

Решение. Так как циклический сдвиг $\sigma(c)$ кодового слова $c = (100)$ равен (010) и не является кодовым словом кода C , то код C – нециклический.

Пример 5

Пусть $C - (2,1,2)$ сверточный код.

Для кода C :

- изобразите схему кодера;
- найдите импульсный отклик кодера;
- используя импульсный отклик кодера, определите выходную битовую последовательность для входных данных $d = (1 \ 0 \ 1)$.

Решение

Для данного кода схема кодера будет иметь следующий вид (рис. 7.14.)

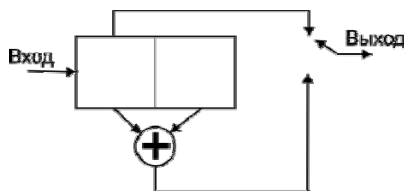


Рис. 7.14. Схема (2,1,2)-кодера

Импульсный отклик:

Вход	X_0	X_1	v_1	v_2
1	1	0	1	1
0	0	1	0	1

Таким образом, импульсный отклик равен 1101.

Найдем выходную последовательность для $d = (1\ 0\ 1)$

Вход	Выход			
1	11	01		
0		00	00	
1			11	01
Итого	11	01	11	01

Выходная последовательность равна (1 1 0 1 1 1 0 1).

Контрольные вопросы

1. Дайте определение канального кодирования.
2. Объясните назначение и способы канального кодирования.
3. Дайте определение блочного (n,k) -кода.
4. Дайте определение линейного блочного (n,k) -кода.
5. Дайте определения веса Хэмминга, расстояния Хэмминга, минимального расстояния кода. Какова связь между ними?
6. Сформулируйте условия на обнаружение и коррекцию ошибок кодом в терминах минимального расстояния.
7. Дайте определение порождающей матрицы.
8. Дайте определение матрицы проверки.
9. Дайте определение синдромного декодирования.
10. Объясните схему синдромного декодирования.
11. Сформулируйте условие Хэмминга на возможность исправления кодом ошибок при синдромном кодировании.
12. Дайте определение циклического сдвига и циклического кода.
13. Дайте определение кодового полинома, полинома проверки, порождающего полинома и синдромного полинома. Какова связь между ними?

14. Приведите схему реализации кодера и декодера циклического кода на основе сдвиговых регистров.
15. Дайте определение сверточного (n, k, m)-кода.
16. Объясните представление сверточного кода в виде таблицы истинности.
17. Объясните представление сверточного кода в полиномиальном представлении.
18. Объясните представление сверточного кода в виде диаграммы состояний.
19. Объясните представление сверточного кода в виде дерева состояний.
20. Объясните представление сверточного кода в виде решетчатой диаграммы.
21. Объясните декодирование сверточных кодов на основе максимального правдоподобия.
22. Объясните декодирование сверточных кодов на основе алгоритма Витерби.
23. Дайте определения кодированию в широком и узком смыслах.
24. Что такое примитивное, экономное (эффективное) и помехоустойчивое (корректирующее) кодирование?
25. Сформулируйте теорему кодирования Шеннона для канала с помехами. В чем ее практическая значимость?
26. Дайте геометрическую трактовку декодированию с обнаружением ошибок.
27. Дайте геометрическую трактовку декодированию с исправлением ошибок.
28. Опишите универсальный (табличный) метод кодирования и декодирования. В чем заключается проблематичность его практического использования?
29. В чем сущность конструктивной теории кодирования?
30. Перечислите и охарактеризуйте признаки, по которым классифицируют коды.
31. Какие коды относят к линейным?
32. Как осуществляется процесс кодирования и декодирования линейных кодов на матричной основе?
33. Какие коды относят к циклическим?

34. Как осуществляется процесс кодирования и декодирования циклических кодов на полиномиальной основе?
35. Дайте определение сверточным кодам. Какими основными параметрами они характеризуются?
36. Что такое относительная скорость и длина кодового ограничения сверточного кода?
37. Прокомментируйте диаграмму состояний сверточного кода (1/2) при длине кодового ограничения 3.
38. Как строится решетчатая диаграмма сверточного кода? Проследите на ней (рис. 7.4) путь, соответствующий входной последовательности битов 00101011. Определите по этому пути выходную последовательность дибитов кодера.
39. Как может осуществляться декодирование сверточных кодов?
40. Поясните различие декодирования с мягким и жестким решениями.
41. В чем сущность решетчатой кодированной модуляции? Как она реализуется?
42. За счет чего достигается повышение помехоустойчивости системы с использованием сигнально-кодовых конструкций?

Литература

1. Акулиничев, Ю. П. Теория электрической связи : учеб. пособие для вузов / Ю. П. Акулиничев. – СПб. : Лань, 2010. – 233 с.
2. Сальников, А. П. Теория электрической связи: конспект лекций / А. П. Сальников. – СПб. : Изд-во «Линк», 2007. – 272 с.
3. Теория электрической связи : учебник для вузов / под ред. Д. Д. Кловского. – М. : Радио и связь, 1999. – 432 с.
4. Крук, Б. И. Телекоммуникационные системы и сети : учеб. пособие : в 3 т. Т. 1 : Современные технологии / Б. И. Крук, В. Н. Попантопуло, В. П. Шувалов; под ред. проф. В. П. Шувалова. – М. : Горячая линия – Телеком, 2003. – 647 с.

5. Прокис, Д. Цифровая связь / Д. Прокис ; под ред. Д. Д. Кловского. – М. : Радио и связь, 2000. – 800 с.
6. Основы цифровой обработки сигналов : учебное пособие / Ю. А. Брюханов, А. Л. Приоров, В. И. Джиган, В. В. Хрящев ; Яросл. гос. ун-т им. П. Г. Демидова. – Ярославль : ЯрГУ, 2013. – 344 с.
7. Гмурман, В. Е. Теория вероятностей и математическая статистика : учеб. пособие для бакалавров / В. Е. Гмурман ; М-во образования и науки РФ – 12-е изд. – М. : Юрайт, 2013. – 479 с.
8. Кремер, Н. Ш. Теория вероятностей и математическая статистика : учебник для вузов. / Н. Ш. Кремер ; М-во образования РФ – 3-е изд., перераб. и доп. – М. : ЮНИТИ, 2009. – 551 с.
9. Теория вероятностей и математическая статистика / сост. Л. П. Бестужева, Н. Л. Майорова; Яросл. гос. ун-т им. П. Г. Демидова, Науч.-метод. совет ун-та. Ч. 1 : практикум. – Б.м. : Б.и., 2012. – 47 с.
10. Hsu, H. Analog and Digital Communications. Schaum's outline series / H. Hsu – McGraw Hill, 2003. – 330 p.
11. Hsu, H. Probability, Random Variables and Random Processes. Schaum's outline series / H. Hsu – McGraw Hill, 1997. – 311 p.

ЗАКЛЮЧЕНИЕ

В учебном пособии рассмотрены базовые вопросы общей теории связи, основанные на методах теории вероятностей и математической статистики, включая оптимальные методы обработки, основы теории информации и кодирования, которые вот уже много десятилетий широко используются в инфокоммуникационных и радиотехнических системах различного назначения.

Рассмотрены также оптимальные методы приема, включая когерентный и некогерентный прием на корреляторах и согласованных фильтрах.

Дано общее представление об оптимальных фильтрах Колмогорова-Винера, определены требования к ним.

Материал пособия может помочь при изучении других разделов современных инфокоммуникаций. Традиционно статистические методы общей теории связи применяются в системах радио- и телевизионного вещания, сотовой и космической связи, которые в последнее время быстро развиваются и широко используются.

ПРЕДМЕТНЫЙ УКАЗАТЕЛЬ

активный фильтр	128	дисперсия	15
алгебраический		интервал корреляции.....	18
сумматор.....	113	интервальные оценки	39
алгоритм Витерби.....	196	источник дискретных	
алфавит источника	91	сообщений	91
амплитудный детектор.....	74	канальный декодер	173
аналоговый		канальный кодер	173
трансверсальный		квазибелый шум	25
фильтр.....	116	квазигармоническая	
апостериорная		форма	73
вероятность	52	квазиоптимальный	
апостериорная		прием	142
плотность вероятности.....	42	когерентные	
априорная вероятность.....	52	демодуляторы	105
априорная		код Баркера	118
плотность вероятности.....	41	код Боуза-Чоудхури-	
Байеса оценка.....	42	Хоквингема	186
безынерционная цепь	65	код Голея	186
белый шум.....	24	код Рида-Соломона	187
Блочный код.....	174	коды с обнаружением	
вероятность ложной		ошибок.....	173
тревоги.....	51	компаратор	122
вероятность		корректирующие коды ...	173
обнаружения	51	коррелятор.....	104
взаимная корреляционная		корреляционная	
функция	18	функция	16
выборка случайной		кортеж.....	174
величины	38	коэффициент	
двоичное поле	175	корреляции	18
декодер максимального		Критерий Байеса	54
правдоподобия	195	критерий идеального	
декодирование сверточных		наблюдателя	94, 96
кодов	194	критерий максимального	
дерево состояний	192	правдоподобия	41, 51
диаграмма состояний	191		

критерий минимального среднего риска	98	нормированный энергетический спектр	24
критерий минимума ошибки.....	54	обеляющий фильтр	120
критерий Неймана- Пирсона	98	обратная работа	128
кусочно-линейное преобразование	67	обратная функциональная характеристика.....	68
линейная среднеквадратичная оценка	44	оптимальный демодулятор	93
линейное безынерционное преобразование	66	ортогональность в усиленном смысле	136
линейные блочные коды	175	отношение правдоподобия	51
линейные цепи	71	отношение сигнал/шум	82
линия задержки	116	Ошибки второго рода.....	50
ложная тревога.....	99	ошибки первого рода.....	50
Математическое ожидание	15	передаточная функция интегратора	113
матрица проверки	178	передаточная функция элемента задержки.....	113
Минимаксный критерий ..	55	плотность вероятности.....	10
Минимальное расстояние линейного кода	176	полином проверки	182
множитель Лагранжа.....	53	Полиномиальное представление	190
модифицированная функция Бесселя нулевого порядка	82	полосовой фильтр.....	72
модулятор	91	порождающая матрица...	177
мощность теста	50	потенциальная помехоустойчивость.....	93
непрерывный канал	91	пропуск цели	99
неравенство Коши- Буняковского-Шварца....	111	простая гипотеза	49
несмещенность оценки	39	пространство принимаемых колебаний	92
Нормальное (гауссовское) распределение	11	процесс стационарный в широком смысле.....	16
		прямое преобразование Фурье	21
		прямоугольный видеоимпульс	112

прямоугольный		тактовый интервал.....	27
радиоимпульс.....	114	теорема Винера-Хинчина	23
Равномерное		теорема Парсеваля.....	22
распределение.....	11	теорема Шеннона.....	174
распределение дискретной		точечные оценки.....	39
случайной величины.....	12	узкополосные случайные	
распределение Райса.....	82	процессы.....	73
регион отказа от		уровень значимости.....	50
гипотезы.....	50	условие	
регион принятия		ортогональности.....	44
гипотезы.....	50	фазовый детектор.....	74
решетчатая диграмма.....	188	фильтр Колмогорова-	
сверточные коды.....	187	Винера.....	142
сдвиговый регистр.....	184	функциональный узел.....	65
синдром.....	179	функция Байеса.....	42
синдромное		функция правдоподобия..	40
декодирование.....	179	функция распределения...10	
синхронный		Хэмминга вес.....	175
телеграфный сигнал.....	27	центральная предельная	
сложные двоичные		теорема.....	73
сигналы.....	115	центральные моменты.....	14
совершенный код.....	180	центрированный СП.....	15
состоятельность оценки...40		циклический код.....	180
спектральная плотность		циклический сдвиг.....	180
мощности.....	22	цифровая модуляция.....	126
спектральная плотность		энергетический спектр.....	23
энергии.....	22	энтропия.....	174
стационарность в узком		эргодический случайный	
смысле.....	13	стационарный процесс.....	19
сферы Хэмминга.....	176	эффективность оценки.....	39
схема Пистолькорса.....	128	якобиан преобразования..	70
таблица истинности.....	188		

ГЛОССАРИЙ

Аналоговый сигнал – непрерывный сигнал непрерывного времени

ВОС (взаимосвязь открытых систем) – взаимодействие между реальными системами, которое организовано в соответствии со стандартами Международной организации по стандартизации

ВОЛС – волоконно-оптическая линия связи

Временная коммутация – однокоординатная коммутация цифрового сигнала, при которой в качестве разделительного используется временной признак

Временное разделение сигналов (каналов) – принцип разделения передаваемых сигналов, базирующийся на периодическом выделении каждому сигналу промежутка времени для передачи в отведенном интервале времени

Демодуляция – процесс, заключающийся в извлечении из меняющихся во времени параметров высокочастотного модулированного сигнала исходного низкочастотного сигнала

Дискретизация – замена непрерывной функции ее дискретными значениями, отсчитываемыми через определенные интервалы времени

Дискретный сигнал – дискретный сигнал дискретного времени

ДОФМ – двухкратная относительная фазовая модуляция

Единичный интервал – минимальный интервал времени, которому равны значащие интервалы времени сигнала

Единичный элемент – элемент сигнала, имеющий длительность, равную единичному интервалу времени

Защита информации – действия, направленные на предотвращение ущерба от определенной угрозы безопасности информации

ИКМ – импульсно-кодовая модуляция

Информация – сведения, являющиеся объектом передачи, распределения, преобразования, хранения или непосредственного использования

Квантование – замена отсчетных значений сигнала значением линейного разрешенного уровня

Код Хемминга – систематический код с кодовым расстоянием $d_0 = 3$ или $d_0 = 4$

Кодек – кодер и декодер, объединенные в одном устройстве преобразования сигналов

Кодовое расстояние – минимальное для данного кода расстояние Хемминга

Конфиденциальная информация – информация, не являющаяся общедоступной

Коэффициент ошибок по кодовым комбинациям – отношение числа ошибочно принятых кодовых комбинаций к числу переданных на заданном интервале времени (интервале анализа)

Коэффициент ошибок по элементам – отношение числа ошибочно принятых элементов к общему числу переданных за интервал анализа

МККТТ – международный консультативный комитет по телефонии и телеграфии

Модем – модулятор и демодулятор, объединенные в одном устройстве преобразования сигналов

Модуляция – процесс, заключающийся в изменении во времени параметров высокочастотного гармонического колебания (амплитуды, частоты, фазы) по закону изменения во времени низкочастотного сигнала, подлежащего передаче

МОС – Международная организация по стандартизации

Нарушение связи – недопущение связи или задержки срочных сообщений

Обеспечение доступности – обеспечение доступа к информации со стороны пользователей, имеющих на то надлежащие полномочия
Обеспечение целостности информации – обеспечение полноты, точности, достоверности

Основание кода – характеризует возможное число различных значащих позиций поступающего от источника сообщений сигнала

Основной цифровой канал – канал со скоростью передачи двоичных цифр в секунду, или 64 Кбит/с

ОФМ – относительная фазовая модуляция

Пакет – часть сообщения, представленная в виде блока с заголовком, имеющего установленный формат (структуру данных) и ограниченную длину, и передаваемая по сети как единое целое

ПД – передача данных

ПДС – передача дискретных сообщений

Перехват данных – обзор данных несанкционированным пользователем

ПЗС – прибор с зарядовой связью

ПК – персональный компьютер

Представляющий (информационный) параметр сигнала данных – параметр сигнала данных, изменение которого отображает изменение сообщения

Производительность источника – среднее количество информации, выдаваемое источником в единицу времени

ПРК – пространственное разделение каналов, способ, при котором в качестве признака используется пространственный признак

Пропускная способность канала связи – максимально возможное значение скорости передачи информации по каналу связи

Разделимый код – код, в кодовых комбинациях которого элементы можно разделить на информационные и проверочные

Расстояние Хэмминга – степень различия комбинаций кода. Определяется как число позиций, в которых соответствующие символы двух слов одинаковой длины различны

Регенерация – восстановление формы двоичного цифрового сигнала после передачи его на определенное расстояние под действием помех и искажений

РОС – решающая обратная связь

Сеть связи – совокупность узлов и соединительных трактов, предусматривающих соединение между двумя или более точками обеспечения связи между ними

Сигнал – форма представления сообщения с помощью физической величины, изменения одного или нескольких параметров которой отображает изменение самого сообщения

Синхронизация – процесс установления и поддержания определенных временных соотношений между двумя или несколькими процессами

Скорость передачи информации – число бит, передаваемых в секунду

Скорость телеграфирования (скорость модуляции) – число единичных элементов, которое можно передать в секунду

Скремблирование (от англ. Scramble – переменчивость) – сложение цифрового сигнала по правилам двоичной арифметики с псевдослучайной двоичной последовательностью с целью исключения из сигнала длинных последовательностей нулей, изменения спектра сигнала и т. п.

СЛ – соединительная линия

СМО – система массового обслуживания

Сообщение – форма представления информации

СП – система передачи

Спектр – совокупность гармоник, образующих периодический сигнал сложной формы

Спектр амплитуд (фаз) – совокупность амплитуд (фаз) гармоник, входящих в состав сигнала

Стартостопная последовательность – последовательность элементов, содержащая стартовый и стоповый элементы. Наличие этих элементов позволяет отделить одну кодовую комбинацию от другой и обеспечить к моменту прихода следующей стартостопной последовательности накопившееся за время стартостопного цикла расхождение по фазе передатчика и приемника

ТГ – тактовый генератор

УПС – устройство преобразования сигналов

Факсимильная связь – область электросвязи, которая занимается передачей неподвижных изображений

Циклический код – разновидность систематического кода, основное свойство которого заключается в следующем. Если комбинация a_0, a_1, \dots, a_{n-1} разрешенная, то комбинация, получаемая из нее путем циклической перестановки элементов, т. е. комбинация $a_{n-1}, a_0, a_1, \dots, a_{n-2}$ также является разрешенной

Цифровой сигнал – последовательность, состоящая из чередующихся случайным образом импульсов и пауз одинаковой длительности; при этом импульс обозначают 1-й, а паузу – 0-м

ЦСД – цифровой сигнал данных. Сигнал, описываемый функцией дискретного времени и конечным множеством возможных значений

Частотное разделение каналов (сигналов) – принцип разделения каналов, базирующийся на отведении сигналу каждого канала своей полосы частот в заданном частотном диапазоне

ЧЗ – чувствительность к задержке

ЧИ – чувствительность к искажениям

ЧРК – частотное разделение каналов

Энтропия – мера неопределенности в поведении источника дискретных сообщений

Учебное издание

Брюханов Юрий Александрович
Приоров Андрей Леонидович

ОБЩАЯ ТЕОРИЯ СВЯЗИ

Учебное пособие

Редактор, корректор М. В. Никулина
Правка, верстка Ю. А. Лукашевич

Подписано в печать 06.05.2014. Формат 60×84¹/₁₆.
Усл. печ. л. 12,55. Уч.-изд. л. 7,5.
Тираж 60 экз. Заказ № 1010.

Оригинал-макет подготовлен
в редакционно-издательском отделе ЯрГУ.

Отпечатано в типографии ООО «Филигрань».
г. Ярославль, ул. Свободы, 91,
Тел. (4852) 98-27-05,
pechataet@bk.ru