ТЕОРИЯ ОПТИМАЛЬНОГО ПРИЕМА СИГНАЛОВ

**1 Основные положения теории оптимального приема сигналов**

Прием сигналов – одна из наиболее сложных теоретических и инженерных задач передачи сообщений. Сложность состоит в том, что в пункте приема сообщения необходимо извлекать из модулированных сигналов-переносчиков, которые в процессе прохождения по линии связи не только ослабляются, но и подвергаются воздействиям различных искажающих факторов и помех.

Весьма желательно располагать методами приема, которые были бы наилучшими (оптимальными) в данных конкретных условиях. Направление, связанное с отысканием таких методов, называется теорией оптимального приема.

Теоретической основой решения задач оптимального приема является теория Байеса.

Пусть некоторая случайная физическая величина, которую назовем причиной, может принимать множество значений(исходов) П с плотностью вероятностей р(П), которая считается априорной(заранее известной). Пусть причина вызывает появление другой случайной величины – следствия С, которое также может принимать множество значений. Плотность вероятностей этих значений зависит от конкретных исходов причины. Поэтому ситуация описывается множеством условных плотностей вероятностей р(С/П).

Статистическим решением называют процедуру, которая состоит в том, чтобы, наблюдая конкретное следствие , указывать вызвавшую его причину . Так как наблюдаемое следствие может быть вызвано любым исходом причины П, то можно определить плотность вероятностей всех возможных исходов, которые могли вызвать данное следствие, т.е. определить функцию р(П/). Эта функция называется апостериорной (послеопытной, установленной на основе имевшего место опыта или наблюдения) плотностью вероятностей причин.

Основой для принятия статистического решения является теорема Байеса

(1)

где р(С/П) – условная плотность распределения следствий;

р(С) – безусловная плотность распределения следствий С, определяемая как

.

Значение этого интеграла не зависит от П, поскольку интегрирование по этой переменной ведется по всей области ее существования Г.

Из (1) следует, что апостериорная плотность вероятностей причины р(П/С) зависит от априорной плотности вероятностей причины р(П) и условной плотности вероятностей следствий р(С/П). плотность р(С/П) является функцией П, ее называют функцией правдоподобия.

В теории статистических решений показано, что при принятии решения о конкретном значении действовавшей причины , вызвавшей наблюдаемое (или заданное) следствие , наименьшую ошибку можно совершить, если выносить решение в пользу того значения причины, при которой условное распределение р(П/) имеет наибольшее значение. Такое правило принятия решения называется байесовским.

Если априорная плотность р(П) неизвестна, то самое большее, что можно сделать – предположить равномерность ее распределения. Тогда решение будет выноситься в пользу того значения причины , при котором функция правдоподобия р(С/П) для наблюдаемого следствия принимает наибольшее значение. Это означает, что такое значение причины считается наиболее правдоподобным среди других возможных значений. Подобная процедура принятия решения называется правилом максимального правдоподобия.

Применим изложенный подход к решению задачи оптимального приема сигналов.

Суть процедуры оптимального приема. Установлено, что между колебаниями и векторами можно установить взаимно-однозначное соответствие. Поэтому вместо колебаний можно рассматривать соответствующие векторы. Исходя из этого, будем считать причиной П случайный вектор х, соответствующий передаваемым сообщениям (или однозначно связанный с ним вектор сигналов s, переносящих эти сообщения), а следствием С – случайный вектор у, соответствующий смеси сигнала шума на входе приемника. С учетом сказанного (1) можно записать либо в виде

(2)

либо в эквивалентном выражению (2) виде

(3)

где x,s,y – векторы в многомерных пространствах, соответствующие сообщениям x(t), сигналам s(t)=s[x(t),t] и входным реализациям y(t)=s(t)+n(t).

При передаче дискретных сообщений множество сообщений x(t) может принимать только конечное число дискретных значений, которому однозначно соответствует конечное число различающихся сигналов

Оптимальная процедура приема состоит в определении величин р(s/ y) для всех М значений , сравнения этих величин между собой и выборе наибольшей из них. Значение , которому соответствует максимальная величина р(/y)

считается переданным сигналом и в соответствии с этим на выходе приемника воспроизводится сообщение .

Основная трудность при решении такой задачи связана с нахождением апостериорного распределения р(s/ y). Наиболее детально задача решена для помехи типа гауссовского белого шума и набора сигналов, заранее известных в точке приема. Если при этом все сообщения равновероятны и независимы, то выражение для р(s/y) можно привести к виду

(4)

где - односторонняя спектральная плотность мощности белого гауссовского шума;

А – некоторая константа.

Нахождение сигнала , максимизирующего величину(4) при наблюдении на входе приемника некоторой реализации y(t), эквивалентно минимизации показателя экспоненты. Следовательно, оптимальный приемник должен выносить решение о приеме того сигнала, при котором функция р(/ y) достигает максимума, а величина

(5)

соответственно становится минимальной.

Учитывая свойства векторного представления функций времени, от выражения(5), можно перейти к эквивалентному ему выражении.

(6)

Выражение(5) или (6) представляет собой алгоритм работы оптимального приемника дискретных сообщений. Работая по этому алгоритму, оптимальный приемник должен вычислить значения величины для всех М, используемых в системе сигналов (где j-1,2,…,М), сравнить их между собой, выбрать наименьшее значение и воспроизвести на выходе соответствующее ему дискретное сообщение.

Иными словами, оптимальный приемник всегда воспроизводит на выходе сообщение, переносимое тем сигналом, к которому наиболее близка входная реализация y(t). В геометрической интерпретации это означает, что оптимальный приемник всегда относит вектор входной реализации y к ближайшему вектору сигнала.

Очевидно, что прием сигналов в присутствии шума может приводить к ошибкам, поскольку вектор входной реализации случаен и с некоторой вероятностью может попасть в любую точку пространства. Допустим, что вектор y, образованный из переданного сигнала и шума n, попал в точку, наиболее близко расположенную к вектору сигнала .

Если i=j, то приемник примет правильное решение, если же , то решение приемника окажется ошибочным и вместо переданного сообщения он ошибочно воспроизведет сообщение .

Несмотря на то, что оптимальный приемник дискретных сообщений может допускать ошибочные решения, их вероятность у этого приемника минимальна по сравнению с любыми реальными приемниками таких сообщений.

Исследования показывают, что алгоритм может быть представлен в более удобном для схемной реализации виде и позволяет получить структурные схемы оптимальных приемников и выражения для расчета помехоустойчивости.

**2 Оптимальный когерентный прием дискретных сигналов и его помехоустойчивость**

В задаче распознавания сигналов, не содержащих случайных параметров(т.е. точно известных), «причинами» являются поступающие на вход сигналы , вероятности которых равны, очевидно, вероятности появления соответствующих элементов . «Следствиями» являются реализации суммы сигнала и помехи.

Количественно описание ситуации удобно производить с помощью рассмотрения векторов соответствующих колебаний. Вместо сигналов будем оперировать однозначно соответствующими им векторами , а вместо реализаций y(t) – векторами , координаты которых определяются выражением, которое в нашем случае запишем так:

(1)

В соответствии с теоремой Байеса

(2)

Как было отмечено, решение обычно выносится в пользу сигнала, имеющего наибольшую апостериорную вероятность. Так как знаменатель не зависит от номера I, то решающее правило(алгоритм решения) определяется так:

(3)

Следует обратить внимание на то, что в этих выражениях -- плотности вероятностей, так как компоненты вектора y, как видно из (1), являются непрерывными случайными величинами.

В выражении (3) априорные вероятности передачи элементов должны быть заданы. Следовательно, необходимо определить только правдоподобия . Это можно сделать исходя из того, что помеха аддитивна. Так как

,

то плотность вероятности некоторого значения вектора равна плотности вероятности, что вектор помехи n примет значение . Отсюда следует, что если- известная нам плотность вероятности вектора помехи, то

(4)

Последний переход справедлив потому, что сигнал и помехи – независимые процессы.

Для дальнейшей конкретизации алгоритма необходимо задать определенный вид помехи. В большинстве случаев имеют место нормальные (гауссовские) или близкие к ним помехи. Вычисления в этом случае оказываются наиболее простыми. При гауссовских помехах каждая компонента вектора распределена по нормальному закону

(5)

В ряде случаев, в частности, при равномерном распределении энергии помехи по полосе рассматриваемых частот, компоненты вектора являются независимыми случайными величинами. Тогда, как известно,

(6)

При зависимых компонентах выражение для существенно усложняется и этот случай здесь рассматривать не будем.

Отметим, что ,т.е. является квадратом длины(нормы) вектора помехи.

Следовательно,

 (7)

Отбросив множители, не зависящие от номера сигнала i, решающее правило(3) можно представить в виде

 (8)

Приемник, работающий по алгоритму(8), называется байесовским или приемником максимальной апостериорной вероятности. Если апостериорные вероятности элементов одинаковы, то решающее правило упрощается:

 (9)

Соответствующий приемник называется приемником максимального правдоподобия. Правило(9) раскрывает механизм работы оптимального приемника.

Получив вектор y, с помощью обработки реализации y(t) необходимо вычислить расстояние от его конца до концов векторов всех возможных сигналов и вынести решение в пользу того сигнала, для которого величина будет минимальной, так как именно в этом случае функция (9) достигнет максимума. Коротко можно сказать, что оптимальный приемник выносит решение в пользу сигнала «ближайшего» к y(t).

Выражение(9) достигает максимума при минимуме показателя экспоненты. Следовательно, правило (9) можно записать в ином виде:

или, учитывая векторное представление

(10)

Здесь первый член в скобках не зависит от номера i. Последний член – есть энергия i-того сигнала. Если энергии всех сигналов одинаковы, что обычно имеет место, то этот член также не зависит от номера i. Таким образом, решающее правило можно записать так:

(11)

Справедливость такого перехода обусловлена тем, что второй член в (10) имеет знак минус и выражение (10) минимизируется, если этот член достигает максимума. Выражение(11) уже позволяет определить структуру оптимального приемника. Однако удобнее это выражение представить в другом виде. Действительно, учтем, что

(12)

Тогда окончательно получим

(13)

Эта структура называется оптимальным корреляционным приемником, так как основная операция, лежащая в его основе, это операция корреляции y(t) со всеми возможными сигналами .

Из проведенного рассмотрения следует, что в состав оптимального приемника должны входить генераторы, вырабатывающие образцы сигналов, тождественные тем, которые используются на передатчике. Кроме того, между работой генераторов передатчика и приемника должна соблюдаться синхронность и синфазность, т.е. обеспечиваться идеальная синхронизация.

**3 Оптимальный некогерентный прием дискретных сигналов и его помехоустойчивость**

Ранее было показано, что если импульсный отклик линии представляет собой -функцию, то такая линия только ослабляет передаваемый сигнал, не изменяя его формы. Пусть ослабление сигнала а — медленно изменяющаяся случайная величина, практически постоянная на интервалах длительностью Тс. Если бы а была постоянной и известной величиной, то осуществлялся бы прием точно известных сигналов с решающим правилом

(1)

При случайном значении а следует усреднить результат по закону распределения р(а); тогда при равновероятностных сигналах решающее правило примет вид

(2)

Из соотношения (2) следует, что при таком подходе структура оптимального приемника останется прежней (инвариантной к случайным значениям а). Вероятность же ошибок (при прочих равных условиях) возрастает. При случайном значении а эти выражения необходимо усреднить по р(а). В частности, для противоположных сигналов усредненное значение вероятности ошибки Р0ш должно определяться в соответствии с выражением

(3)

Для распределения р(а), подчиняющегося закону Рэлея можно показать, что

(4)

где . Нетрудно видеть, что при одинаковых значениях а вероятность ошибок, рассчитанная по формуле (4), значительно превышает вероятность ошибок. Физическая причина увеличения вероятности ошибок ясна: возрастание а приводит к некоторому уменьшению вероятности ошибок, однако падение а приводит к значительному возрастанию этой вероятности вследствие отмеченного выше «порогового эффекта».

Рассмотрим далее случай, когда линия вносит в сигналы только случайный сдвиг начальной фазы, имеющий место в подавляющем большинстве реальных ситуаций. При этом, если

то сигналы на выходе линии (входе приемника)

(5)

Выходные сигналы (5) можно представить в виде двух составах со случайными амплитудами, но постоянными фазами:

(6)

где а и Ь могут, в отличие от предыдущего случая, принимать и положительные и отрицательные значения.

Из (6) видно, что действие линии можно свести к появлению в точке приема двух составляющих сигнала: косинусоидальной и синусоидальной. Анализ этого случая, связанный с выполнением усреднения по обоим случайным параметрам а и Ь, довольно громоздок.

Приведем конечное выражение для решающего правила:

(7)

Из него следует, что оптимальный приемник производит корреляцию принятой реализации у(t) с образцами обоих слагаемых сигнала. Возведение результатов в квадраты перед сложением и выбором максимума вызвано тем, что величины а и Ь могут быть как положительными, так и отрицательными.

Этот алгоритм можно реализовать и с помощью согласованных фильтров. Здесь содержатся детекторы огибающих выходных колебаний согласованных фильтров, после которых и производится отсчет. Физика процессов также ясна: если на вход согласованного с сигналом фильтра подать сдвинутый по фазе сигнал, то в силу линейности фильтра произойдет запаздывание колебания и на выходе фильтра. Поэтому отсчет в момент t= TС не совпадет с максимумом напряжения. В силу случайности этого сдвига наилучшей стратегией оказывается отсчет огибающей, а не мгновенного значения колебания.

Сравним случай приема сигналов при отсутствии случайной фазы (т. е. точно известных по форме сигналов) и при наличии случайной фазы. Первый случай принято называть когерентным, а второй — некогерентным приемом (именно этот случай чаще всего имеет место на практике).

(8)

Сравнивая выражения для когерентного и некогерентного приема при одинаковом значении вероятности ошибки, можно установить, какой энергетический проигрыш дает применение некогерентного приема по сравнению с когерентным. Расчеты показывают, что для обеспечения при некогерентном приеме требуется увеличение энергии сигнала на 15-30% по сравнению с когерентным, т. е. проигрыш невелик.

В более общем случае неидеальность линии обусловливает случайные изменения амплитуды и фазы. Вероятность ошибок от этого увеличивается, так как независимо действуют оба рассмотренных фактора. Можно показать, что в этом случае вероятность ошибок при распознавании бинарных ортогональных сигналов равна

(9)

где - среднее значение энергии принимаемых сигналов.

**4 Оптимальный и квазиоптимальный прием непрерывных сигналов и его помехоустойчивость**

Перейдем к рассмотрению особенностей оптимального приема при передаче непрерывных сообщений. В этом случае передаваемое сообщение х(t) может иметь очень большое (практически бесконечное) число возможных реализаций, каждая из которых представляет собой непрерывную функцию времени. Поэтому в геометрической интерпретации сообщениям и сигналам соответствуют не отдельные точки (или векторы с фиксированной длиной) в многомерных пространствах(как это было при передаче дискретных сообщений), а континуум линий сообщений и сигналов, описываемых концами векторов х и s. Исследования показывают, что в этой ситуации оптимальный прием связан с формированием на приемной стороне такого сигнала s(t), который бы обеспечивал максимум максиморум апостериорной плотности вероятности, определяемой выражением.

Применительно к каналу с гауссовским белым шумом и равновероятными сообщениями указанное условие сводится к минимизации величины

(1)

Чтобы сформировать сигнал s(t), на приемной стороне нужно использовать принятое сообщение х(t), которое представляет собой результат обработки входной реализации у(t) приемником. Сообщение х(t) называют оценкой переданного сообщения х(t). Формирование сигнала s(t) представляет собой модуляцию несущей сигнала колебанием х(t) по тому же закону и с теми же параметрами, что и на передающей стороне.

Сформированный в приемнике сигнал s(t) используется при обработке входной реализации у(t) и последующем формировании оценки сообщения х(t), которая, в свою очередь, необходима для создания сигнала s(t). Нетрудно понять, что указанная процедура может быть выполнена только в устройстве следящего типа, с использованием обратной связи по формируемой оценке сообщения х(t).

В геометрической интерпретации минимизация выражения означает, что оптимальный приемник всегда относит входную текущую реализацию у к ближайшей линии сигналов и в соответствии с этим формирует на выходе оценку сообщения х(t). Из-за влияния шума оценка х(t) отличается от переданного сообщения х(t). Это отличие обычно характеризуют величиной среднеквадратической ошибки (см. л. 1.3). Оптимальный прием обеспечивает минимальное значение этой ошибки по сравнению с любым другим способом приема.

Теория оптимального приема непрерывных сообщений, часто называемая также теорией оптимальной демодуляции аналоговых видов модуляции, или теорией нелинейной фильтрации, представляет важный раздел общей теории связи, основы которой были заложены в работах А.Н. Колмогорова, В.А. Котельникова, Н. Винера, К. Шеннона и ряда других отечественных и зарубежных ученых.

Задачей приемного устройства являются извлечение переданного сообщения х(t) из входного колебания у(t). Однако из-за помех и искажений эта процедура не может быть выполнена точно, и восстановить сообщение на выходе приемника можно только приближенно. Такое приближенное сообщение называют оценкой и обозначают х(t).

Критерием близости х(t) и х(t) в теории и технике связи принята СКО, в соответствии с которой

(2)

где скобки <.> означают операцию усреднения реализации по времени.

Оптимальный приемник непрерывных сообщений обеспечивает наименьшую возможную в заданных условиях величину СКО. Определим эту ошибку.

Основываясь на теории ортогональных разложений передачу любого непрерывного сообщения можно заменить передачей совокупности числовых коэффициентов (параметров). Пусть непрерывное сообщение х(t) представлено рядом

(3)

При известной системе базисных функций передача сообщений x(t) эквивалентна передаче п значений коэффициентов Следовательно, передаваемый сигнал можно рассматривать как функцию времени и коэффициентов , т. е.

(4)

Влияние помех приведет к тому, что каждый коэффициент , будет принят с некоторой погрешностью. В результате оценка сообщения примет вид

(5)

где колебание нужно рассматривать как помеху на выходе приемника.

Если единственной причиной появления этой помехи является белый гауссовский шум на входе приемника, то нетрудно убедиться в том, что помеха имеет нормальное распределение. В.А.Котельников показал, что в режиме надпорогового оптимального приема спектральная плотность такой помехи определяется выражением

(6)

Средний квадрат ошибки при оптимальном приеме непрерывных сообщений с учетом (2) можно найти по формуле

(7)

Для выбранного (или заданного) вида модулированных сигналов помехоустойчивость оптимального приема будет наиболее высокой по сравнению с любым возможным реальным способом приема этих же сигналов. Поэтому такую помехоустойчивость часто называют потенциальной (предельно возможной для данного вида сигналов).

При анализе потенциальной помехоустойчивости полезно различать прямые виды модуляции, у которых передаваемое сообщение x(t) непосредственно входит в выражение для сигнала и интегральные, у которых сигнал — функция интеграла от передаваемого сообщения, т.е. .

Рассмотрим особенности расчета потенциальной помехоустойчивости для некоторых случаев.

Помехоустойчивость сигналов с амплитудной модуляцией. Пусть для передачи непрерывных сообщений используется АМ сигнал. В этом случае

(8)

(9)

(10)

В (8) учтено, что соs2а = 0,5(1+соs2а) и интеграл распадается на две составляющих, одна из которых (с частотой 2w0) близка к нулю и отброшена.

Из (9) следует, что при АМ сигнале спектральная плотность помехи на выходе оптимального приемника постоянна. Эта особенность характерна не только для АМ, но и всех других сигналов с прямыми видами модуляции.

Приняв во внимание, что средние мощности сигнала и шума на входе приемника

где - ширина спектра АМ сигнала, определяющая полосу пропускания приемника, имеем

(11)

В соответствии с (11) потенциальная помехоустойчивость АМ сигналов в основном определяется отношением сигнала к шуму на входе приемника. Для получения малых значений ошибки это отношение должно быть весьма большим.

Помехоустойчивость сигналов с угловой модуляцией. Пусть для передачи непрерывных сообщений используются сигналы с угловой модуляцией. Сначала рассмотрим случай фазовой модуляции.

Из(13) следует, что при ФМ сигнале, как и при АМ, спектральная плотность помехи на выходе постоянна, поскольку ФМ принадлежит к сигналам с прямой модуляцией.

При ЧМ сигнале спектральная плотность помехи на выходе имеет квадратичную зависимость от частоты. Такая зависимость характерна для всех интегральных видов модуляции. В этом случае

(12)

(13)

Из (13) следует, что при ФМ сигнале, как и при АМ, спектральная плотность помехи на выходе постоянна, поскольку ФМ принадлежит к сигналам с прямой модуляцией.

(14)

где и -- средние мощности шума и сигнала на входе приемника; -- полоса частот, занимаемая спектром ФМ сигнала.

Проведем теперь рассмотрение для ЧМ сигнала. Он относится к интегральному виду модуляции.

(15)

где - текущая частота, принимающая значения в интервале

Спектральная плотность помехи на выходе оптимального приемника ЧМ сигналов равна

(16)

Эта формула показывает, что при ЧМ сигнале спектральная плотность помехи на выходе имеет квадратичную зависимость от частоты. Такая зависимость характерна для всех интегральных видов модуляции.

Средний квадрат ошибки при приеме ЧМ сигналов можно записать так:

(17)

где - индекс частотной модуляции.

Проанализируем полученные результаты. Из (14) и (17) следует, что при ФМ и ЧМ помехоустойчивость приема можно повысить только за счет увеличения индекса модуляции (не увеличивая при этом среднюю мощность сигнала Рс). Однако увеличение приводит к расширению спектра ФМ и ЧМ сигналов и соответственно к необходимости использовать более широкую полосу частот. Это уменьшает отношение сигнала к шуму на входе приемникаПри некотором значении индекса величина qс снизится до пороговой величины, , при которой условия надпорогового приема нарушаются и начинает резко возрастать вероятность аномальных ошибок . В этом случае формулами (14) и (17) пользоваться уже нельзя.

