

ТЕОРИЯ АВТОМАТИЧЕСКОГО УПРАВЛЕНИЯ ДЛЯ «ЧАЙНИКОВ»

Часть II. Управление при случайных возмущениях. Оптимальные линейные системы

К.Ю. Поляков

*Санкт-Петербург
2009*

«В ВУЗе нужно излагать материал на высоком профессиональном уровне. Но поскольку этот уровень проходит значительно выше головы среднего студента, я буду объяснять на пальцах. Это не очень профессионально, зато понятно».

Неизвестный преподаватель

Предисловие

Эта методичка – вторая часть «Теории автоматического управления для чайников». Предполагается, что первая часть уже прочитана и понята. Основное содержание второй части – случайные процессы в системах автоматического управления и оптимальные линейные системы. Задача автора – объяснить «на пальцах» основные понятия теории и подготовить читателя к восприятию профессиональной литературы в этой области. Нужно рассматривать это пособие только как первую ступень в изучении предмета, который может стать очень интересным и увлекательным.

Список существующих учебников и монографий по теории случайных процессов и оптимальным системам управления огромен. Тем не менее, осваивать серьезную литературу современному студенту сложно. Мозг при восприятии новой информации ищет что-то знакомое, за что можно «зацепиться», и на этой основе «привязать» новое к уже известным понятиям. А если такая «зацепка» не обнаруживается – просто отключается.

В большинстве серьезных трудов материал излагается на высоком научном уровне, грамотно, точно и полно. Но читать их очень сложно, потому что такой задачи – написать понятно – не ставилось изначально.

Как правило, в основе любой научной теории лежат достаточно простые и понятные идеи. Однако, со временем они «накрываются» таким математическим аппаратом, что читатель-новичок вязнет в нем, не добравшись до самих идей, которые авторам учебников кажутся очевидными. Автор этого пособия своей основной задачей считал написание понятной книжки, в которой обсуждению идей отводится ведущее место.

При любом улучшении приходится чем-то жертвовать. В данном случае в жертву были принесены строгость и полнота изложения. Математик найдет здесь много недоговоренностей и упущений, поскольку (в соответствии с целями пособия) между строгостью и понятностью выбор всегда делается в пользу понятности. Кроме того, были отброшены все второстепенные (на взгляд автора) результаты, о которых на первых порах можно не говорить без существенного ущерба для результатов инженерной практики.

Сознательно ничего не было сказано о решении рассматриваемых задач с помощью моделей в пространстве состояний. Переход в пространство состояний позволяет исследовать процессы в более общем виде, но одновременно скрывает (и теряет) важные структурные свойства системы. Например, с помощью моделей в пространстве состояний очень сложно выявить многие особенности задачи синтеза оптимальных регуляторов, которые «лежат на поверхности» при использовании классического (частотного) подхода.

Основная часть примеров связана с судовыми системами управления, что определяется личными вкусами автора. Специально для студентов-судостроителей написана глава «Морское волнение», в которой собраны базовые сведения по этой теме.

От читателя требуются небольшие предварительные знания. Нужно иметь представление о некоторых разделах курса высшей математики: теории вероятностей, производных и интегралах, комплексных числах.

Благодарности

Автор выражает глубокую признательность к.т.н. В.Н. Калиниченко, который внимательно прочитал предварительную версию пособия и высказал много ценных замечаний, которые позволили улучшить изложение и сделать его более понятным.

Содержание

1. СЛУЧАЙНЫЕ СОБЫТИЯ	4
1.1. Что такое случайное событие?	4
1.2. Случайные величины	4
1.3. Гистограмма распределения	4
1.4. Плотность распределения вероятностей	5
1.5. Средние значения	7
1.6. Какие бывают распределения?	8
2. СЛУЧАЙНЫЕ ПРОЦЕССЫ	11
2.1. Что такое случайный процесс?	11
2.2. Стационарность	12
2.3. Эргодичность	12
2.4. Корреляционная функция	13
2.5. Спектральная плотность	14
2.6. Гармонический сигнал	16
2.7. Белый шум	17
3. ОЦЕНКА И МОДЕЛИРОВАНИЕ СЛУЧАЙНЫХ ПРОЦЕССОВ	19
3.1. Оценка корреляционной функции	19
3.2. Оценка спектральной плотности	19
3.3. Прохождение случайных сигналов через линейные системы	24
3.4. Моделирование случайных сигналов	26
4. МОРСКОЕ ВОЛНЕНИЕ	31
4.1. Что такое морское волнение?	31
4.2. Кажущиеся спектры	33
4.3. Моделирование действия морского волнения на судно	36
5. ОПТИМИЗАЦИЯ СИСТЕМ ПРИ СЛУЧАЙНЫХ ВОЗМУЩЕНИЯХ	39
5.1. Что такое оптимальная система?	39
5.2. Оптимальная фильтрация	39
5.3. Оптимальное управление в замкнутых системах	44
5.4. Стандартная система	46
5.5. Особенности задачи оптимизации	48
5.6. Кривая качества	50
6. ОПТИМАЛЬНЫЕ СЛЕДЯЩИЕ СИСТЕМЫ	52
6.1. Постановка задачи	52
6.2. Теорема Парсеваля	53
6.3. Эквивалентность двух задач	53
6.4. Разомкнутые системы	54
6.5. Замкнутые системы	56
ЗАКЛЮЧЕНИЕ	58
ЛИТЕРАТУРА ДЛЯ ПОСЛЕДУЮЩЕГО ЧТЕНИЯ	59

1. Случайные события

1.1. Что такое случайное событие?

Случайное событие – это такое событие, которое может произойти или не произойти, причем это можно выяснить только в результате опыта.

Основная характеристика случайного события – это его *вероятность*, то есть, частота появления события в большой серии опытов. Вероятность события – это знания, которые у нас есть до проведения опыта¹.

Если в большой серии из N опытов событие X случилось n_X раз, можно говорить о том, что вероятность появления события $P(X)$ примерно равна

$$P(X) \approx \frac{n_X}{N}.$$

Это приближенное равенство (теоретически) превращается в точное при стремлении числа опытов N к бесконечности. Например, если из партии в 1000 автомобилей 50 имеют дефекты (а остальные – исправны), то вероятность купить дефектный автомобиль, выбрав его наугад из этой партии, составляет 5%.

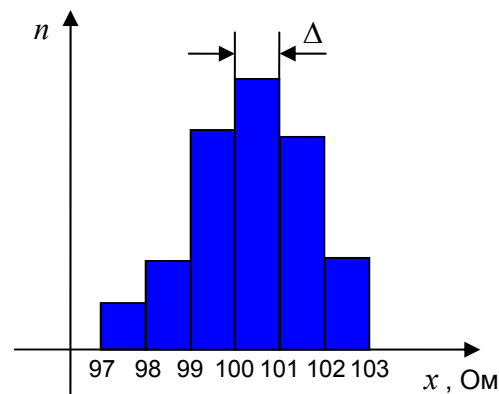
1.2. Случайные величины

Говоря о случайных событиях, мы рассматриваем только два варианта, «случилось» или «не случилось». Однако часто результаты эксперимента можно выразить в виде числа, количественно.

Предположим, что нас интересует сопротивление резисторов, купленных в магазине. Номинальное значение сопротивления, равно, например, 100 Ом. Однако, при изготовлении всегда есть допуски, то есть, разрешенные отклонения от номинала. Например, при допуске $\pm 3\%$ сопротивление взятого наугад резистора может быть *любым числом* в интервале от 97 до 103 Ом. Это – *случайная величина*. В общем случае интервал может быть и бесконечным, например, от 0 до бесконечности.

1.3. Гистограмма распределения

Разобьем интервал на несколько равных частей (подынтервалов), выберем случайным образом N резисторов, измерим их сопротивления и подсчитаем, сколько резисторов «попали» в каждый интервал. Изобразим эти данные на столбчатой диаграмме, где высота каждого столбика – это количество резисторов в данном интервале. Это – *гистограмма распределения* случайной величины. В данном случае по гистограмме мы сразу видим, что больше всего резисторов имеют сопротивление от 100 до 101 Ом.



На гистограмме можно показывать не только количество, но и *долю* резисторов, попавших в данный интервал. Например, если 100 резисторов из 1000 имеют сопротивление от 97 до

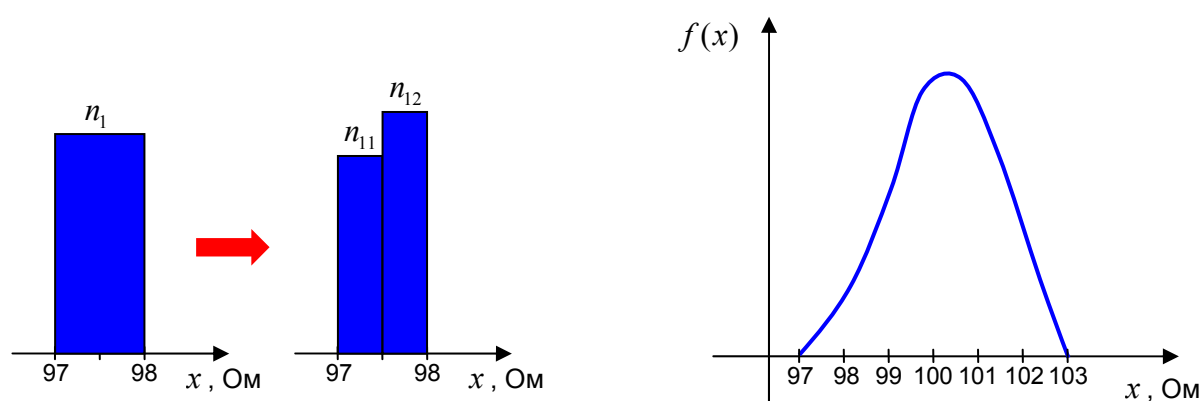
¹ В науке такие сведения принято называть *априорными* (лат. *a priori*, до опыта).

98 Ом, их доля составляет 0,1 от общего числа (10%). Тогда высоту соответствующего (первого) столбика гистограммы можно сделать равной 0,1. При этом сумма высот всех столбиков будет равна 1 (или 100%).

Кроме того, в данном случае мы выбрали ширину интервала $\Delta = 1$, поэтому *площадь* всех столбиков также будет равна 1. В этом случае гистограмма называется *нормированной*. В ней высота столбика с номером i равна $\frac{n_i}{N \cdot \Delta}$, где n_i – число резисторов, «попавших» в i -ый интервал, а N – их общее количество.

1.4. Плотность распределения вероятностей

Теперь уменьшим ширину интервала Δ в 2 раза и увеличим в 2 раза N , так чтобы произведение $N \cdot \Delta$ осталось постоянным. Фактически каждый исходный интервал мы разбили на 2 равных подынтервала (рисунок слева). Через n_{11} и n_{12} обозначим количества резисторов в первом и втором подынтервалах.



Поскольку общее количество резисторов удвоилось, в интервал $[97; 98]$ (то есть в два новых подынтервала) попало примерно в 2 раза больше резисторов, то есть, $n_{11} + n_{12} \approx 2n_1$. Поэтому высота обоих столбиков будет (скорее всего) близка к тому, что было раньше. Мы только уточнили распределение резисторов внутри исходного интервала $[97; 98]$.

Такое деление можно выполнять и для нормированной гистограммы (с единичной площадью). В пределе при $\Delta \rightarrow 0$ (и $N \rightarrow \infty$) мы получаем прямоугольники бесконечно малой ширины. Через их «вершины» можно провести некоторую линию. Она представляет собой график функции, которую называют *плотностью распределения вероятностей* (или просто *плотностью распределения*) случайной величины X и обозначают $f(x)$ (здесь x – одно из допустимых значений случайной величины X).

Так как выполняется условие нормировки, площадь под этой линией равна 1, она может быть вычислена как интеграл от функции $f(x)$ на всем множестве ее допустимых значений. Если заранее известно, что величина x находится в некотором интервале $[a; b]$, получаем

$\int_a^b f(x) dx = 1$. В общем случае (если случайная величина может принимать любые вещественные значения), справедлива формула

$$\int_{-\infty}^{\infty} f(x) dx = 1.$$

Интеграл от $f(x)$ на некотором интервале $[x_1; x_2]$ определяет вероятность того, что случайная величина x при очередном испытании окажется в этом интервале, то есть выполнится неравенство $x_1 \leq x \leq x_2$.

Чему же равна вероятность точного равенства $x = x_1$ для некоторого заданного x_1 ? Чтобы ее найти, нужно взять интеграл $\int_{x_1}^{x_1} f(x) dx$. Поскольку верхний и нижний пределы интегрирования совпадают, для «обычных» функций такой интеграл равен нулю, то есть, в рассмотренном выше примере вероятность выбрать наугад резистор с сопротивлением 100 Ом равна нулю.

Может ли интеграл $\int_{x_1}^{x_1} f(x) dx$ быть ненулевым? Оказывается да, но для этого функция $f(x)$ в точке x_1 должна быть бесконечной. Этим свойством обладает, например, так называемая *дельта-функция* (или функция Дирака) $\delta(x)$, которая определяется так:

$$\delta(x) = \begin{cases} 0, & x \neq 0 \\ \infty, & x = 0 \end{cases} \quad \int_{-\infty}^{\infty} \delta(x) dx = 1.$$

Дельта-функция равна нулю во всех точках, кроме $x = 0$, где она обращается в бесконечность, причем интеграл от нее по всей оси равен 1.

Когда плотность вероятности может содержать дельта-функции? Предположим, что мы измеряем сигнал x на выходе цифрового устройства, который может принимать только два значения, например, 0 или 1. Такой сигнал, принимающий значения только из заранее заданного множества, называется *дискретным*.

Пусть вероятность появления нуля равна 0,4, а вероятность появления единицы – 0,6. Попробуем построить плотность распределения этого сигнала, используя интуитивные соображения («здравый смысл»).

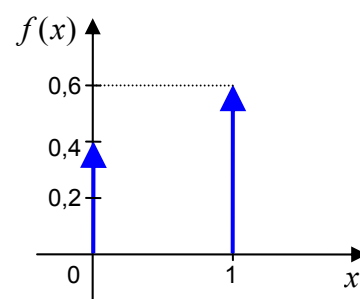
Во-первых, сигнал не может принимать другие значения, кроме 0 и 1, поэтому плотность вероятности равна нулю везде, за исключением этих двух точек. Во-вторых, вероятность того, что $x = 0$ ненулевая (равна 0,4), и вероятность того, что $x = 1$ равна 0,6. Таким образом, имеем (при любом малом ε)

$$\int_{0-\varepsilon}^{0+\varepsilon} f(x) dx = 0,4 \quad \text{и} \quad \int_{1-\varepsilon}^{1+\varepsilon} f(x) dx = 0,6.$$

Отсюда следует, что плотность распределения $f(x)$ содержит дельта-функции в точках $x = 0$ и $x = 1$ (интегралы от которых равны соответственно 0,4 и 0,6) и равна нулю в остальных точках². Иначе говоря,

$$f(x) = 0,4 \cdot \delta(x) + 0,6 \cdot \delta(x-1).$$

Дельта-функцию, имеющую бесконечное значение, на графике обозначают стрелкой, высота которой равна ее площади (см. рисунок справа).



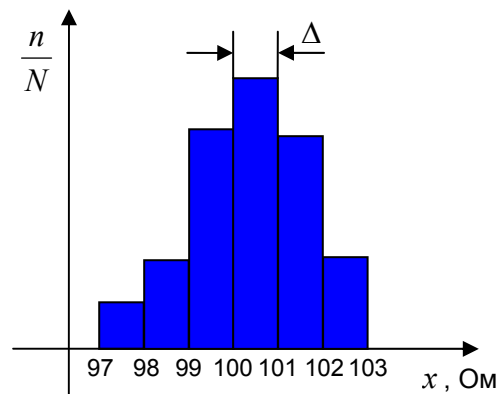
² Может случиться и так, что плотность распределения представляет собой сумму «нормальной» функции и дельта-функций. Например, мы знаем, что в коробке есть 20 резисторов, сопротивление которых точно равно 100 Ом, а сопротивление остальных может быть любым в пределах допуска, от 97 до 103 Ом.

1.5. Средние значения

Плотность распределения вероятности дает полную информацию о свойствах случайной величины. Например, с помощью гистограммы несложно найти среднее значение.

Вернемся к примеру с резисторами. Пусть у нас есть гистограмма, построенная по результатам испытаний, где отложены доли резисторов, сопротивление которых оказалось в данном интервале (см. рисунок). Как (приблизительно) найти среднее значение сопротивления?

Сложность в том, что мы не знаем сопротивлений отдельных резисторов и их количества. Тем не менее, гистограмма дает всю необходимую информацию.



Можно приблизительно считать, что все резисторы, попавшие в интервал $[97;98]$ (их количество равно n_1), имеют сопротивление $x_1 = 97,5$ Ом (середина интервала). Более точной информации все равно нет. Тогда общая сумма сопротивлений всех этих резисторов равна $x_1 \cdot n_1$. Повторяя те же рассуждения для остальных интервалов, получаем, что сумма сопротивлений всех резисторов равна $x_1 \cdot n_1 + x_2 \cdot n_2 + \dots + x_K \cdot n_K$, где K – количество интервалов, а x_i ($i = 1, \dots, K$) – середина каждого из интервалов. Чтобы найти среднее значение, эту сумму нужно разделить на общее количество резисторов N :

$$\bar{x} = \frac{x_1 \cdot n_1 + x_2 \cdot n_2 + \dots + x_K \cdot n_K}{N} = x_1 \cdot \frac{n_1}{N} + x_2 \cdot \frac{n_2}{N} + \dots + x_K \cdot \frac{n_K}{N} = \sum_{i=1}^K x_i \cdot \frac{n_i}{N}.$$

Значения $\frac{n_i}{N}$ ($i = 1, \dots, K$) – это доли от общего количества, то есть, высоты столбцов гистограммы. Таким образом, мы можем (приблизительно) найти среднее значение по гистограмме, не зная ни сопротивлений отдельных резисторов, ни даже их количества.

В теории вероятности среднее значение $\bar{x}$ называется *математическим ожиданием* случайной величины x и обозначается $E\{x\}$. Если известна плотность распределения $f(x)$, сумма заменяется интегралом

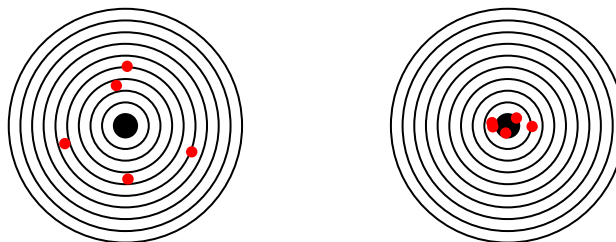
$$\bar{x} = E\{x\} = \int_{-\infty}^{\infty} x \cdot f(x) dx.$$

Аналогично можно найти среднее значение *любой функции*, умножив ее на плотность распределения и проинтегрировав произведение на всей оси. Например, средний квадрат $\overline{x^2}$ вычисляется так

$$\overline{x^2} = E\{x^2\} = \int_{-\infty}^{\infty} x^2 \cdot f(x) dx.$$

Среднее значение (математическое ожидание) не может полностью характеризовать случайную величину. На рисунках показаны мишени, пораженные двумя стрелками (каждый сле-

лал по 5 выстрелов). В обоих случаях математическое ожидание – это центр мишени, то есть «в среднем» они бьют по центру. Однако всем понятно, что второй явно стреляет лучше. Как выразить это в виде числа?



У первого стрелка больше *разброс* точек попадания относительно средней точки. на языке теории вероятности разброс называется *дисперсией* – эта величина равна *среднему квадрату отклонения* от среднего значения $\bar{x}$. То есть, дисперсия вычисляется по формуле:

$$D_x = E\{(x - \bar{x})^2\} = \int_{-\infty}^{\infty} (x - \bar{x})^2 \cdot f(x) dx.$$

Раскрыв скобки в подынтегральном выражении, можно показать, что дисперсия равна разности среднего квадрата и квадрата математического ожидания:

$$D_x = \overline{x^2} - (\bar{x})^2.$$

Если математическое ожидание равно нулю, дисперсия и средний квадрат совпадают.

Использовать дисперсию не очень удобно, поскольку ее единицы измерения не совпадают с единицами измерения исходной величины (если x измеряется в метрах, то дисперсия – в квадратных метрах). Поэтому на практике чаще применяют *среднеквадратическое отклонение (СКВО)* – квадратный корень из дисперсии:

$$\sigma_x = \sqrt{D_x}.$$

В иностранной литературе эту величину называют *стандартное отклонение*.

1.6. Какие бывают распределения?

Существует бесчисленное множество разных распределений, но в технике применяются лишь некоторые из них.

1.6.1. Равномерное распределение

Самое простое – это *равномерное* распределение. Например, снег в безветренную погоду ложится на плоскую поверхность равномерно – ровным слоем, который имеет одинаковую толщину во всех точках. Обычно предполагается, что ошибка квантования непрерывных сигналов в цифровом компьютере имеет равномерное распределение. Равномерное распределение на интервале $[a; b]$ описывается плотностью распределения

$$f(x) = \begin{cases} 0, & x < a \\ \frac{1}{b-a}, & a \leq x \leq b \\ 0, & x > b \end{cases}$$

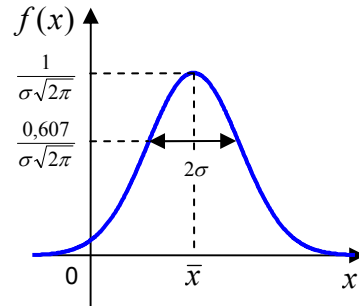
Среднее значение такой случайной величины равно $\bar{x} = (a + b)/2$, а дисперсия

$$D_x = \int_a^b \left(x - \frac{a+b}{2}\right)^2 \frac{1}{b-a} dx = \frac{(b-a)^2}{12}.$$

1.6.2. Нормальное распределение (распределение Гаусса)

Самое важное распределение в практических задачах – *нормальное распределение (распределение Гаусса)*, для которого график плотности распределения имеет форму колокола:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\bar{x})^2}{2\sigma^2}\right)$$



Здесь $\bar{x}$ – среднее значение, а σ – среднеквадратическое отклонение случайной величины. Распределение Гаусса обладает несколькими замечательными свойствами:

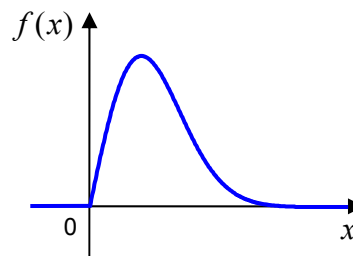
- 1) сумма (и любая линейная комбинация) случайных величин с нормальными распределениями тоже имеет нормальное распределение;
- 2) если на величину действует множество независимых помех, ее плотность вероятности стремится к нормальному закону;
- 3) при прохождении случайного сигнала с нормальным распределением через линейную систему сигнал на выходе тоже имеет нормальное распределение.

Если нет никаких теоретических или экспериментальных данных о распределении случайной величины (например, шума измерений), чаще всего предполагают, что это распределение – нормальное.

1.6.3. Другие распределения

В специальных задачах применяют и другие распределения. Если случайная величина имеет равномерное распределение с центром в нуле, ее модуль распределен по *закону Рэля*:

$$f(x) = \begin{cases} \frac{x}{\sigma^2} \exp\left(-\frac{x^2}{2\sigma^2}\right), & x \geq 0 \\ 0, & x < 0 \end{cases}$$

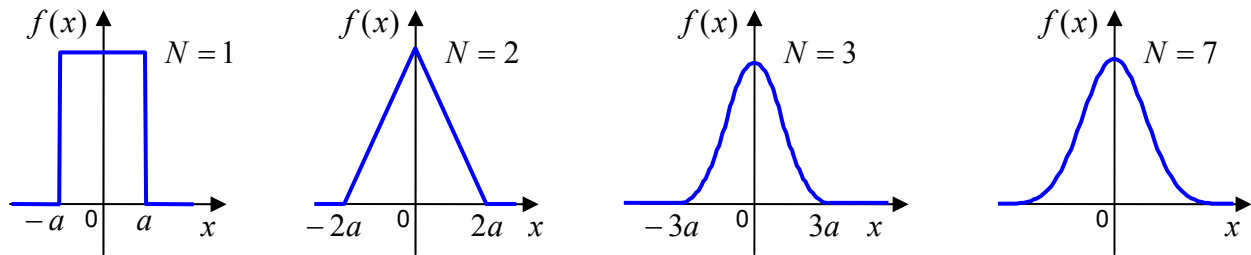


Например, так распределяются высоты волн при морском волнении.

Для моделирования случайных события в компьютерных моделях используют датчики *псевдослучайных* чисел (они похожи на случайные, но каждое следующее вычисляется по некоторой формуле, использующей предыдущие значения). Большинство датчиков «выдают» равномерно распределенные значения, из которых с помощью математических операций можно получить другие распределения.

Например, рассмотрим сумму нескольких *независимых* случайных значений, равномерно распределенных на симметричном интервале $[-a; a]$. Можно показать, что для суммы двух чисел получится треугольное распределение. Таким образом, складывая два случайных числа, по-

лученных со стандартного датчика с равномерным распределением, мы получим числа с треугольным распределением. Для суммы трех чисел график $f(x)$ состоит из кусочков парабол:



При увеличении N график плотности распределения вероятностей становится всё больше похож на «колокол» нормального распределения. Доказано, что при больших N распределение суммы N чисел действительно стремится к нормальному. Более того, это справедливо для суммы большого количества независимых случайных величин с *любым* распределением (не обязательно равномерным).

Можно показать, что дисперсия суммы *независимых* случайных величин равна сумме дисперсий отдельных величин. Поэтому если взять $a = 0,5$, то дисперсия суммы 12 чисел, равномерно распределенных на интервале $[-a; a]$, будет равна

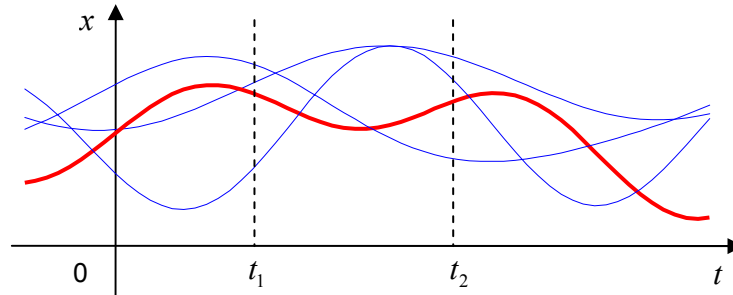
$$D_z = 12 \cdot \frac{(a - (-a))^2}{12} = 1.$$

Этот прием (сложение 12 чисел) используют для получения случайных величин с нормальным распределением и единичной дисперсией.

2. Случайные процессы

2.1. Что такое случайный процесс?

Случайный процесс – это случайная функция времени. Это означает, что наблюдатель «видит» только одну *реализацию* случайного процесса (она выделена на рисунке красным цветом) из множества возможных функций (синие линии).



Полный набор всех возможных реализаций называют *ансамблем*. Случайный процесс – это и есть ансамбль реализаций, а не функция в обычном понимании. Далее будем обозначать весь ансамбль (случайный процесс) через $X(t)$, а отдельную реализацию – через $x(t)$.

Характеристикой случайного процесса (точнее – характеристикой ансамбля реализаций) в каждый фиксированный момент времени $t = t_1$ является *плотность распределения вероятности* $f(X_1)$ случайной величины $X_1 = X(t_1)$. По этим данным можно найти среднее значение (математическое ожидание), дисперсию, СКВО и другие характеристики случайного процесса. Процессы с нулевым средним значением называются *центрированными*.

Для многих (хотя и не для всех) случайных процессов значения в моменты времени t_1 и t_2 как-то связаны. Чтобы оценить связь случайных величин $X_1 = X(t_1)$ и $X_2 = X(t_2)$ используют *корреляцию* – математическое ожидание произведения $X_1 \cdot X_2$:

$$R_{X_1 X_2} = E\{X_1 \cdot X_2\}.$$

Корреляция позволяет выявить *линейную* зависимость между двумя величинами. В случае $R_{X_1 X_2} > 0$ знаки X_1 и X_2 чаще всего совпадают (оба положительные или оба отрицательные), а при $R_{X_1 X_2} < 0$ – больше вероятность того, что знаки разные. Если $R_{X_1 X_2} = 0$, величины X_1 и X_2 называются *некоррелированными*. Важно понимать, что это не означает, что они независимы. С другой стороны, независимые величины всегда некоррелированы. Для случайных величин с нормальным распределением некоррелированность одновременно означает и независимость.

Вспоминая, что X_1 и X_2 – это значения случайного процесса в моменты t_1 и t_2 , можно рассматривать корреляцию как функцию двух аргументов:

$$R_X(t_1, t_2) = E\{X(t_1) \cdot X(t_2)\}.$$

Эта функция называется *корреляционной* (или *автокорреляционной*) функцией случайного процесса $X(t)$. В этой формуле используется *усреднение по ансамблю*, то есть по всем возможным реализациям случайного процесса. Практически эта операция трудновыполнима, так как нужно иметь полную информацию о процессе (распределения вероятностей).

Если случайный процесс – это напряжение в вольтах, то его корреляционная функция измеряется в B^2 , так же, как средний квадрат и дисперсия.

2.2. Стационарность

Если все свойства случайного процесса (плотности распределения вероятностей) не зависят от времени, случайный процесс называется *стационарным* (в узком смысле). Иначе процесс – нестационарный, его свойства со временем изменяются. Строго говоря, все реальные процессы – нестационарные, они когда-то начались и когда-то закончатся. Однако часто на практике можно считать, что на интересующем нас интервале времени (например, во время перехода судна из одного порта в другой) свойства случайных процессов (волнения, ветра) не изменяются. Это допущение позволяет существенно упростить решение многих задач.

Стационарность – это очень сильное допущение. Чтобы доказать его справедливость, нужно знать все плотности распределения³ в любой момент времени, а они чаще всего неизвестны. К счастью, стационарность (в узком смысле) совсем не требуется в инженерных задачах. Вместо этого достаточно рассматривать процессы, *стационарные в широком смысле*, для которых

- 1) математическое ожидание не зависит от времени;
- 2) корреляционная функция $R_X(t_1, t_2)$ зависит только от того, насколько моменты t_1 и t_2 далеки друг от друга, то есть от разности $t_1 - t_2$, поэтому ее часто записывают в виде

$$R_X(\tau) = E\{X(t)X(t+\tau)\}, \text{ где } \tau = t_1 - t_2.$$

Далее, говоря о стационарных процессах, мы будем иметь в виду процессы, стационарные в широком смысле.

2.3. Эргодичность

При первом знакомстве со случайными процессами всегда возникает закономерный вопрос: «Как же изучать случайные процессы на практике?» Дело в том, что во многих случаях мы наблюдаем только одну реализацию из всего ансамбля, и повторить опыт с теми же условиями невозможно.

Исследователи почти всегда предполагают, что длительное наблюдение за *одной реализацией* случайного процесса позволяет изучить свойства *ансамбля*, то есть, один элемент ансамбля содержит информацию обо всех остальных элементах. Случайные процессы, обладающие таким свойством, называют *эргодическими*. Заметим, что только стационарный процесс может быть эргодическим.

С одной стороны, в реальных ситуациях очень сложно доказать эргодичность. С другой – обычно имеет смысл предположить, что процесс эргодический, если нет веских доводов против этого.

Для эргодических процессов по одной реализации можно найти все основные характеристики, заменив усреднение по ансамблю на усреднение по времени. Например, математическое ожидание стационарного случайного процесса можно найти через его плотность распределения:

$$\bar{x} = E\{X\} = \int_{-\infty}^{\infty} x \cdot f(x) dx.$$

³ Строго говоря, нужно учитывать *совместные* плотности распределения (плотности распределения нескольких случайных величин). Подробнее об этом можно почитать в [1,3].

Если мы знаем только одну реализацию, можно попробовать оценить среднее значение на интервале $[-T, T]$, разделив интеграл от функции $x(t)$ на ширину интервала:

$$\hat{x}_T = \frac{1}{2T} \int_{-T}^T x(t) dt.$$

Переходя к пределу при $T \rightarrow \infty$ (применяя усреднение на бесконечном интервале), получаем оценку среднего значения по одной реализации $x(t)$:

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

Для эргодических процессов это значение совпадает с $\bar{x}$, которое получено путем усреднения по ансамблю.

2.4. Корреляционная функция

Корреляционная функция $R_X(\tau)$ стационарного процесса $X(t)$ также может быть вычислена двумя способами, усреднением по ансамблю (через совместную плотность вероятности) и усреднением одной реализации по времени. Для эргодического процесса оба метода дают один и тот же результат.

Далее мы будем рассматривать только эргодические процессы, для которых можно найти корреляционную функцию по одной реализации. Чтобы вычислить $R_X(\tau)$ для некоторого τ , нужно найти среднее значение произведения $x(t)x(t+\tau)$:

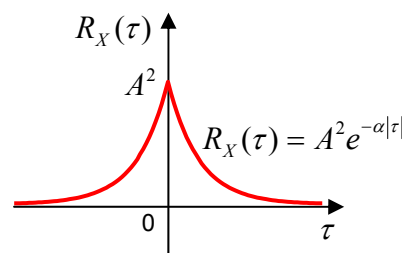
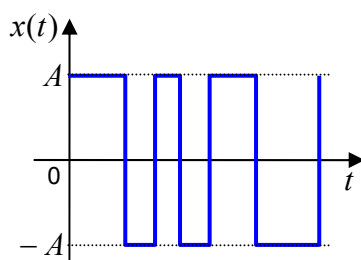
$$R_X(\tau) = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T x(t)x(t+\tau) dt. \quad (1)$$

Построить график функции $R_X(\tau)$ можно по точкам, выполнив такое интегрирование для каждого значения τ из некоторого массива.

Корреляционная функция обладает рядом важных **свойств**:

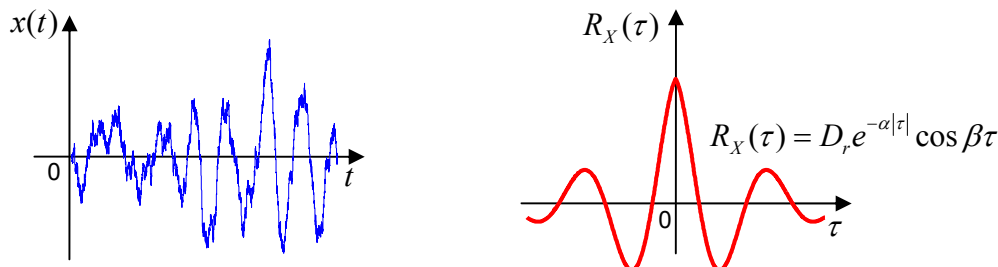
- 1) $R_X(0)$ – это средний квадрат случайного процесса, поэтому всегда $R_X(0) \geq 0$; для центрированных процессов (с нулевым средним) эта величина совпадает с дисперсией;
- 2) при $\tau = 0$ корреляционная функция имеет наибольшее значение, в том числе и наибольшее по модулю, то есть $|R_X(\tau)| \leq R_X(0)$ при всех τ ;
- 3) $R_X(\tau) = R_X(-\tau)$, то есть $R_X(\tau)$ – симметричная (четная) функция, это доказывается подстановкой $-\tau$ вместо τ в интеграл (1); поэтому можно считать корреляционную функцию только для $\tau \geq 0$, а вторую часть строить симметрично.

В качестве примера приведем корреляционную функцию дискретного сигнала, который переключается между значениями A и $-A$ в случайные моменты времени:



Корреляционная функция имеет вид $R_X(\tau) = A^2 e^{-\alpha|\tau|}$, где α – среднее число переключений за 1 с. Заметим, что одна и та же корреляционная функция может соответствовать многим совершенно различным процессам.

Корреляционная функция не всегда положительна. На следующих рисунках показано изменение ординаты поверхности морского волнения и корреляционная функция этого сигнала (одна из теоретических моделей):



Здесь D_r – дисперсия волновой ординаты, α – коэффициент затухания и β – средняя частота волнения.

Заметим также, что чаще всего корреляционная функция убывает по модулю, т.е. чем дальше от нуля, тем меньше значение модуля корреляционной функции (чем больше расстояние между отсчетами, тем меньше связь между ними). Это справедливо не для всех случайных процессов, но для большинства практических ситуаций.

2.5. Спектральная плотность

В теории управления существуют и взаимно дополняют друг друга два подхода:

- 1) *временной* – исследование процессов во времени;
- 2) *частотный* – исследование частотных свойств сигналов и систем (с помощью передаточных функций и частотных характеристик).

Аналогичная ситуация наблюдается и при рассмотрении случайных процессов. Основная временная характеристика стационарного процесса – это корреляционная функция, а частотные свойства описываются *спектральной плотностью*.

Спектральная плотность – это функция, которая показывает распределение *мощности* сигнала по частотам. Такая информация о полезных сигналах, помехах и возмущениях очень важна для разработчика систем управления. Система должна быть спроектирована так, чтобы усиливать сигналы с «полезными» частотами и подавлять «вредные» частоты, характерные для помех и возмущений.

Для перехода от временного описания *детерминированных* (не случайных) процессов к частотному, используют преобразования Фурье и Лапласа. Аналогично спектральная плотность *случайного* процесса может быть найдена как преобразование Фурье от корреляционной функции⁴:

$$S_X(\omega) = \int_{-\infty}^{\infty} R_X(\tau) e^{-j\omega\tau} d\tau = \mathcal{F}\{R_X(\tau)\}.$$

⁴ Эта формула называется *формулой Винера-Хинчина*. Строго говоря, это не определение спектральной плотности, а следствие из него. Математически корректное определение можно найти в литературе [1,3].

Здесь $j = \sqrt{-1}$ – мнимая единица, а ω – угловая частота в рад/с ($\omega = 2\pi f$, где f – «обычная» частота в герцах). Используя формулу Эйлера, можно представить экспоненту в виде сумму вещественной (косинусной) и мнимой (синусной) составляющих: $e^{-j\omega\tau} = \cos \omega\tau - j \sin \omega\tau$. Функция $R_X(\tau) \sin \omega\tau$ – нечетная по τ , поэтому интеграл от нее в симметричных пределах равен нулю. Напротив, функция $R_X(\tau) \cos \omega\tau$ – четная, так что при интегрировании можно взять интервал от 0 до ∞ и удвоить результат:

$$S_X(\omega) = 2 \int_0^{\infty} R_X(\tau) \cos \omega\tau d\tau. \quad (2)$$

Спектральная плотность чем-то похожа на плотность распределения вероятностей, только она характеризует плотность распределения мощности сигнала по частотам. Если случайный процесс – это напряжение в вольтах, то его корреляционная функция измеряется в В², а спектральная плотность – в В²/Гц.

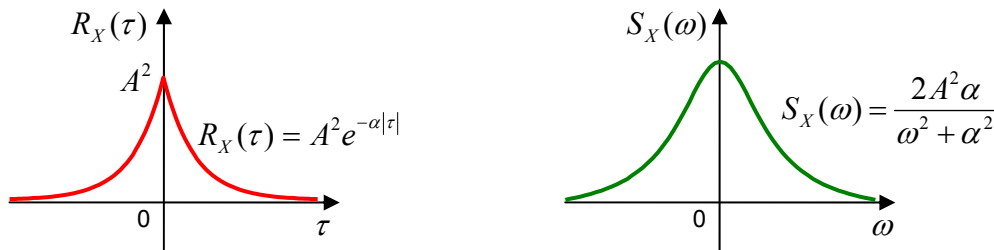
Спектральная плотность случайного процесса, имеющего корреляционную функцию $R_X(\tau) = A^2 e^{-\alpha|\tau|}$, вычисляется как

$$S_X(\omega) = A^2 \int_{-\infty}^{\infty} e^{-\alpha|\tau|} e^{-j\omega\tau} d\tau = A^2 \int_{-\infty}^0 e^{(\alpha-j\omega)\tau} d\tau + A^2 \int_0^{\infty} e^{-(\alpha+j\omega)\tau} d\tau.$$

Интервал интегрирования разбит на две части. При $\tau < 0$ имеем $|\tau| = -\tau$, а при $\tau > 0$ – $|\tau| = \tau$. Выполняя интегрирование, получаем

$$S_X(\omega) = \frac{A^2 e^{(\alpha-j\omega)\tau}}{\alpha-j\omega} \Big|_{-\infty}^0 - \frac{A^2 e^{-(\alpha+j\omega)\tau}}{\alpha+j\omega} \Big|_0^{\infty} = A^2 \left(\frac{1}{\alpha-j\omega} + \frac{1}{\alpha+j\omega} \right) = \frac{2A^2\alpha}{\omega^2 + \alpha^2}.$$

На рисунке слева показана корреляционная функция, а справа – соответствующая ей спектральная плотность мощности:



Свойства спектральной плотности:

- 1) это неотрицательная, четная функция угловой частоты ω (график расположен выше оси абсцисс и симметричен относительно вертикальной оси);
- 2) интеграл от $S_X(\omega)$ на некотором интервале частот $[\omega_1; \omega_2]$ дает мощность, которая связана с этими частотами; поскольку функция $S_X(\omega)$ – четная, результат интегрирования на $[\omega_1; \omega_2]$ нужно удвоить, чтобы учесть также и полосу $[-\omega_2; -\omega_1]$;
- 3) площадь под кривой определяет *средний квадрат* случайного процесса (для центрированного процесса он равен дисперсии):

$$\overline{x^2} = \frac{1}{2\pi} \int_{-\infty}^{\infty} S_X(\omega) d\omega.$$

Множитель $1/2\pi$ нужен для согласования единиц измерения, поскольку угловая частота $\omega = 2\pi f$ измеряется не в герцах, а в рад/с. Учитывая, что функция $S_X(\omega)$ четная, можно интегрировать ее только при $\omega > 0$, а результат удвоить:

$$\overline{x^2} = \frac{1}{\pi} \int_0^{\infty} S_X(\omega) d\omega.$$

В теории управления нередко записывают спектральную плотность как функцию комплексной переменной s , связанной с угловой частотой по формуле $s = j\omega$ (отсюда следует $s^2 = -\omega^2$). Хотя это не совсем корректно с точки зрения математики, мы будем использовать запись $S_X(s)$ для обозначения спектральной плотности $S_X(\omega)$, в которой выполнена замена $s^2 = -\omega^2$:

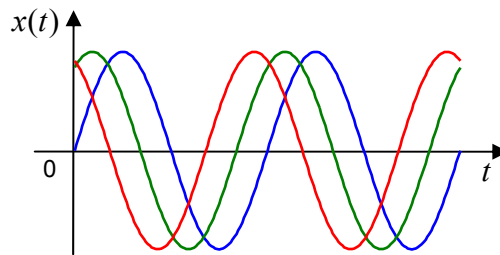
$$S_X(\omega) = \frac{2A^2\alpha}{\omega^2 + \alpha^2} \Leftrightarrow S_X(s) = \frac{2A^2\alpha}{-s^2 + \alpha^2}.$$

2.6. Гармонический сигнал

Рассмотрим гармонический сигнал

$$x(t) = A \sin(\omega_0 t + \theta),$$

где θ – случайная фаза, равномерно распределенная в интервале от 0 до 2π . Три реализации этого процесса (с разными фазами θ) показаны на рисунке:



Это тоже случайный процесс, однако его отличие от «классических» случайных процессов состоит в том, что зная (или определив) случайную фазу θ , мы можем вычислить значение этого сигнала при любом t . Таким процессом называют *квазидетерминированным*. Как только фаза θ определена, процесс становится детерминированным (не случайным).

Использование формулы для усреднения по времени (1) дает

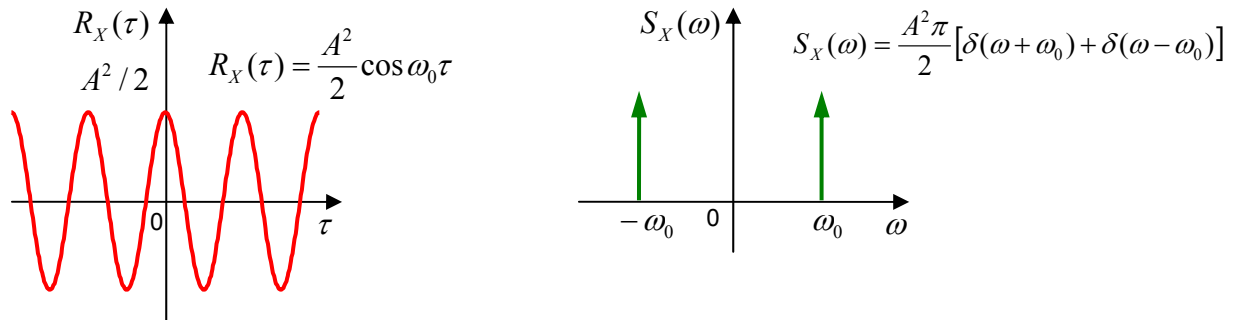
$$R_X(\tau) = A^2 \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T \sin(\omega_0 t + \theta) \sin(\omega_0 t + \theta + \omega_0 \tau) dt.$$

После несложных преобразований (применение формулы произведения синусов и интегрирование), получим $R_X(\tau) = \frac{A^2}{2} \cos \omega_0 \tau$.

Чтобы найти спектр такого сигнала, вычислим преобразование Фурье для корреляционной функции. По таблицам находим $\mathcal{F}\{\cos \omega_0 \tau\} = \pi[\delta(\omega + \omega_0) + \delta(\omega - \omega_0)]$, поэтому

$$S_X(\omega) = \frac{A^2 \pi}{2} [\delta(\omega + \omega_0) + \delta(\omega - \omega_0)].$$

Это значит, что спектральная плотность состоит из двух дельта-функций для частот $-\omega_0$ и ω_0 , а в остальных точках равна нулю. Действительно, с самого начала было легко догадаться, что вся энергия такого сигнала сосредоточена на одной частоте.



2.7. Белый шум

В математике для теоретических исследований иногда удобно использовать математические объекты, которые нереализуемы на практике (например, дельта-функцию). В теории случайных процессов важную роль играет *белый шум*⁵, имеющий равномерную спектральную плотность по всем частотам, то есть, $S_X(\omega) = S_0 = \text{const}$. Очевидно, что при этом площадь под кривой спектральной плотности (определяющая средний квадрат процесса) бесконечна, то есть сигнал имеет бесконечную мощность и не может существовать в природе.

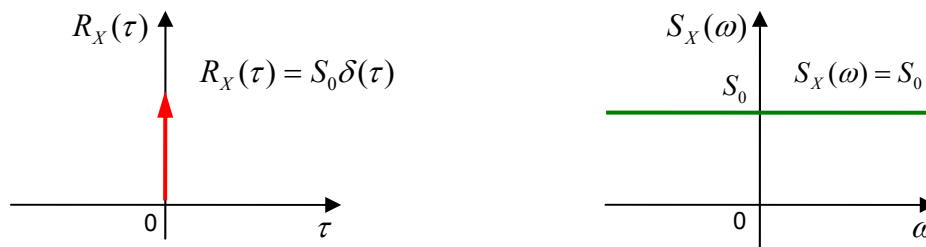
Если нет никакой информации о свойствах случайных возмущений, действующих на системы, часто считают, что они приближенно описываются моделью белого шума. Если мы докажем, что даже в этом (наихудшем) случае характеристики системы останутся удовлетворительными, то они будут не хуже и при любой другой случайной помехе.

Корреляционная функция белого шума равна $R_X(\tau) = S_0\delta(\tau)$. Действительно, преобразование Фурье сразу дает

$$S_X(\omega) = \int_{-\infty}^{\infty} S_0 \delta(\tau) e^{-j\omega\tau} d\tau = S_0.$$

Значения такого сигнала отстоящие по времени на любой, сколь угодно малый интервал, некоррелированы. Это означает, что нет никакой зависимости между соседними, сколь угодно близко расположенными друг к другу, отсчетами такого случайного процесса.

Корреляционная функция и спектральная плотность белого шума показаны на рисунках:



Белый шум, как сигнал с бесконечной энергией, невозможно получить на практике. При моделировании его обычно заменяют на *белый шум с ограниченной полосой*, который имеет равномерный спектр в полосе частот от $-\omega_0$ до ω_0 , и нулевой вне этой полосы:

⁵ Это название связано с белым светом, спектр которого содержит все частоты видимого спектра.

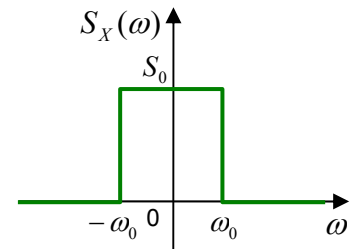
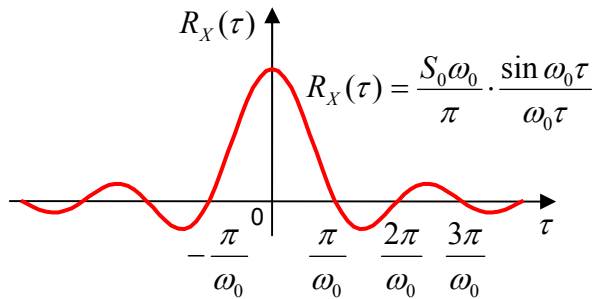
$$S_X(\omega) = \begin{cases} S_0, & |\omega| \leq \omega_0 \\ 0, & |\omega| > \omega_0 \end{cases}.$$

Средний квадрат такого сигнала равен $\overline{x^2} = S_0 \omega_0 / \pi$, а не бесконечности. Корреляционную функцию можно найти с помощью обратного преобразования Фурье:

$$R_X(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} S_X(\omega) d\omega = \frac{S_0 \omega_0}{\pi} \cdot \frac{\sin \omega_0 \tau}{\omega_0 \tau}.$$

Поскольку $\sin \pi k = 0$ при любом целом k , корреляционная функция равна нулю при всех $\tau = \frac{\pi k}{\omega_0}$, где $k \neq 0$ – любое целое число, не равное нулю. Это значит, что значения, взятые из

такого сигнала в моменты времени $0, \frac{\pi}{\omega_0}, \frac{2\pi}{\omega_0}, \frac{3\pi}{\omega_0}, \dots$, (выборка с периодом $\frac{\pi}{\omega_0}$) будут некоррелированы.



3. Оценка и моделирование случайных процессов

3.1. Оценка корреляционной функции

В прикладных задачах часто нужно определить корреляционную функцию и спектральную плотность по экспериментальным данным. При этом мы можем наблюдать и анализировать только «кусоч» реализации на временном интервале от нуля до некоторого T , поэтому для невозможно использовать усреднение по ансамблю. Остается надеяться на то, что процесс эргодический, и применять усреднение по времени.

Пусть известна реализация случайного процесса $x(t)$ на интервале от 0 до T . Для оценки (приближенного вычисления) корреляционной функции при $0 \leq \tau \ll T$ (то есть при положительных τ , достаточно малых по сравнению с T) можно использовать формулу

$$\hat{R}_X(\tau) = \frac{1}{T-\tau} \int_0^{T-\tau} x(t) x(t+\tau) dt. \quad (3)$$

Обратите внимание, что время усреднения равно $T-\tau$, а не T , потому что только интервал $[0; T-\tau]$ содержит как t , так и $t+\tau$. К сожалению, точно вычислить этот интеграл невозможно, потому что мы не знаем математическую формулу для $x(t)$. В реальности обычно известны только значения этой функции (*выборка*) в моменты $0, \Delta, 2\Delta, \dots, N\Delta$, где Δ – интервал между измерениями. Тогда $\hat{R}_X(\tau)$ можно приближенно подсчитать только для $\tau = 0, \Delta, 2\Delta, \dots, M\Delta$ (где $M \ll N$) по формуле

$$\hat{R}_X(i\Delta) = \frac{1}{N-i+1} \sum_{k=0}^{N-i} x(k\Delta) \cdot x(k\Delta + i\Delta), \quad i = 0, 1, \dots, M \ll N,$$

в которой интеграл заменен на сумму. С теоретической точки зрения математическое ожидание такой оценки (при усреднении по ансамблю) совпадает с истинной корреляционной функцией, то есть это – *несмещенная оценка*.

3.2. Оценка спектральной плотности

3.2.1. Использование оценки корреляционной функции

Предположим, что мы исследуем эргодический процесс и знаем одну реализацию $x(t)$ на интервале от 0 до некоторого T . Выше было показано, что по этим данным можно построить оценку корреляционной функции. Если бы мы знали полностью непрерывную корреляционную функцию $R_X(\tau)$, для оценки спектральной плотности можно было бы использовать преобразование Фурье (формулу (2)):

$$\hat{S}_X(\omega) = 2 \int_0^{\infty} \hat{R}_X(\tau) \cos \omega \tau d\tau.$$

В реальности известны лишь значения $\hat{R}_X(i\Delta)$ в отдельных точках, поэтому последнюю формулу нужно перевести в дискретный вид, заменив интеграл на конечную сумму:

$$\hat{S}_X(\omega) = 2\Delta \sum_{i=0}^M \hat{R}_X(i\Delta) \cos \omega i\Delta. \quad (4)$$

Этот метод оценки спектральной плотности называют *методом Блэкмана-Тьюки*.

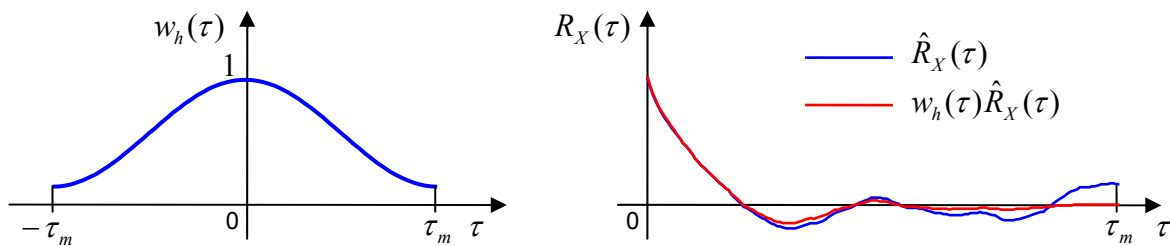
К сожалению, такой подход не всегда дает удовлетворительные результаты. Дело в том, что мы знаем только часть корреляционной функции, для значений τ от 0 до $\tau_m = M\Delta$. Эта неполнота знаний может очень существенно влиять на результаты оценки спектра, вплоть до того, что вычисления по формуле (4) могут дать для некоторых частот отрицательные значения спектральной плотности. Этого не может быть в принципе, потому что мощность сигнала (и любой его составляющей) не может быть отрицательной.

3.2.2. Окна

Чтобы исправить ситуацию, нужно как-то «сгладить» незнание корреляционной функции при больших τ и сделать оценку спектральной плотности более надежной. Для этого используются так называемые «окна» – четные функции, на которые умножается корреляционная функция перед тем, как применить к ней преобразование Фурье. Одно из простейших «окон» – окно Хэмминга:

$$w_h(\tau) = \begin{cases} 0,54 + 0,46 \cos \frac{\pi\tau}{\tau_m}, & \tau \geq \tau_m \\ 0, & \tau < \tau_m \end{cases}.$$

На рисунке слева показано окно Хэмминга, а справа – исходная оценка корреляционной функции $\hat{R}_X(\tau)$ и результат применения к ней окна Хэмминга $w_h(\tau)\hat{R}_X(\tau)$ (красная линия):



Ясно видно, что применение этого окна (и других тоже) практически не изменяет форму корреляционной функции при малых τ , но сглаживает все выбросы при больших τ , которые, скорее всего, вызваны случайными ошибками.

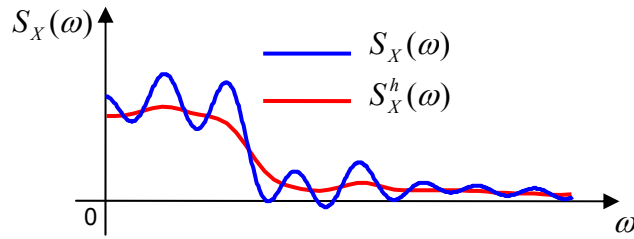
Для оценки спектральной плотности с учетом окна $w(\tau)$ применяют формулу, аналогичную (4):

$$\hat{S}_X(\omega) = 2\Delta \sum_{i=0}^M w(i\Delta) \hat{R}_X(i\Delta) \cos \omega i\Delta. \quad (5)$$

Не стоит печалиться по поводу того, что окно вносит дополнительное искажение. Так или иначе, «окно» используется всегда. Фактически, усекая корреляционную функцию, мы применяем прямоугольное окно:

$$w_r(\tau) = \begin{cases} 1, & \tau \geq \tau_m \\ 0, & \tau < \tau_m \end{cases}.$$

На следующем рисунке показаны оценки спектра сигнала, полученные при использовании прямоугольного окна ($S_X(\omega)$, синяя линия) и окна Хэмминга ($S_X^h(\omega)$, красная линия).



Хорошо видно, что график $S_X(\omega)$ заходит в отрицательную область, что невозможно с физической точки зрения. Применение окна Хэмминга позволило избавиться от этой проблемы и сгладить скачкообразные изменения оценки спектра.

3.2.3. Использование дискретного преобразования Фурье

Главный недостаток классического метода оценки спектральной плотности (метода Блэкмана-Тьюки) — большой объем вычислений. Гораздо меньше операций требуется при использовании прямого метода, основанного на использовании дискретного преобразования Фурье и современных вычислительных алгоритмах быстрого преобразования Фурье. При этом не нужно строить корреляционную функцию, а можно сразу найти спектральную плотность, обработав выборку значений исходного сигнала.

В теории обработки аналоговых сигналов для перехода из временной области в частотную используется преобразование Фурье

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

Оно имеет смысл для любой детерминированной (неслучайной) функции $f(t)$, которая абсолютно интегрируема, то есть интеграл от ее модуля на всей оси сходится:

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

Для стационарного случайного процесса, не равного нулю, это условие никогда не будет выполняться, поэтому использовать преобразование Фурье в обычном смысле для анализа спектра случайных процессов нельзя.

Однако если рассмотреть *усеченный* процесс $x_T(t)$, равный реализации $x(t)$ случайного процесса на интервале $[-T; T]$ и нулю вне этого интервала, для него можно найти преобразование Фурье:

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

Квадрат модуля этой функции, деленный на ширину интервала $2T$, характеризует среднюю мощность сигнала на частоте ω . В пределе, при $T \rightarrow \infty$, мы должны получить спектральную плотность мощности. Так как $x(t)$ — это только одна реализация случайного процесса, в окончательной формуле нужно использовать усреднение по ансамблю (математическое ожидание):

$$S_X(\omega) = \lim_{T \rightarrow \infty} \frac{E\{|F_X(\omega)|^2\}}{2T}.$$

При реальных измерениях мы знаем только одну реализацию случайного процесса $x(t)$ на интервале $[0; T]$, поэтому усреднение по ансамблю чаще всего невозможно. Тогда для оценки спектральной плотности можно использовать предыдущую формулу без усреднения::

$$\hat{S}_x(\omega) = \frac{|F_x(\omega)|^2}{T}, \quad \text{где } F_x(\omega) = \int_0^T x(t) e^{-j\omega t} dt.$$

Теперь остается найти (приближенно) $F_x(\omega)$ по дискретным измерениям процесса $x(t)$. Предположим, что известны его значения $x_k = x(k \cdot \Delta t)$ при $t = k \cdot \Delta t$ для $k = 0, 1, \dots, N-1$, так что интервал $[0; T]$ разделен на N подынтервалов шириной Δt (потому $T = N \cdot \Delta t$). Тогда интегрирование можно приближено заменить суммой:

$$F_x(\omega) \approx \sum_{k=0}^{N-1} x_k e^{-j\omega k \cdot \Delta t} \cdot \Delta t = \Delta t \sum_{k=0}^{N-1} x_k e^{-j\omega k \cdot \Delta t}.$$

Для оценки спектра в теории обработки сигналов обычно используют сетку частот (в герцах)

$$f_m = \frac{m}{N \cdot \Delta t}, \quad m = 0, 1, \dots, N-1$$

с шагом $\Delta f = \frac{1}{N \cdot \Delta t}$. В теории управления принято строить спектры как функции *угловой частоты* (в радианах в секунду), которая получается из «обычной» частоты умножением на 2π :

$$\omega_m = \frac{2\pi}{N \cdot \Delta t} \cdot m, \quad m = 0, 1, \dots, N-1.$$

Для частоты ω_m получаем

$$F_x(\omega_m) = \Delta t \cdot \sum_{k=0}^{N-1} x_k e^{-\frac{2\pi j}{N} mk} = \Delta t \cdot X_m, \quad (6)$$

где через X_m обозначена сумма, называемая *дискретным преобразованием Фурье (ДПФ)*:

$$X_m = \sum_{k=0}^{N-1} x_k e^{-\frac{2\pi j}{N} mk}.$$

Заметим, что эта величина – комплексная, содержащая как вещественную, так и мнимую части.

Легко подсчитать, что при расчете ДПФ по этим формулам для N частот количество операций сложения и умножения будет пропорционально N^2 (обозначается $O(N^2)$). Это значит, что если N увеличивается, скажем, в 10 раз, то количество операций – примерно в 100 раз. Для больших N , особенно при анализе сигналов в реальном времени, такие расчеты выполняются недопустимо долго.

Для быстрого вычисления ДПФ были разработаны специальные алгоритмы, которые называются *быстрым преобразованием Фурье (БПФ)*. Они позволили сократить количество операций с $O(N^2)$ до $O(N \log N)$. В функции **fft** среды МАТЛАВ используется модификация алгоритма БПФ, предложенного Дж. Кули и Дж. Тьюки. Этот алгоритм наиболее эффективен, если число отсчетов N представляет собой степень двойки ($N = 2^p$ при целом $p > 0$). Заметим, что если это не так, всегда можно дополнить ряд нулями до ближайшей степени двойки.

Казалось бы, формула (6) позволяет оценить спектр для всех частот вплоть до $\omega_{N-1} = \frac{2\pi(N-1)}{N \cdot \Delta t}$. Однако нужно учесть, что для анализа мы используем только дискретные из-

мерения с периодом Δt . Остальные значения непрерывного сигнала $x(t)$ (между моментами измерений) теряются, и с ними теряется информация о высокочастотных составляющих.

Согласно теореме Котельникова-Шеннона, по дискретным измерениям с периодом Δt можно восстановить частотные свойства сигнала только до частоты $f_{\max} = \frac{1}{2\Delta t}$ (или до соответствующей угловой частоты $\omega_{\max} = \frac{\pi}{\Delta t}$, которая называется *частотой Найквиста*⁶). Поэтому только оценка спектра на частотах $\omega_0, \dots, \omega_{N/2}$ дает нам практически полезную информацию⁷.

Подведем итог. Для оценки спектра сигнала по N отсчетам x_k ($k = 0, 1, \dots, N-1$) нужно выполнить следующие действия:

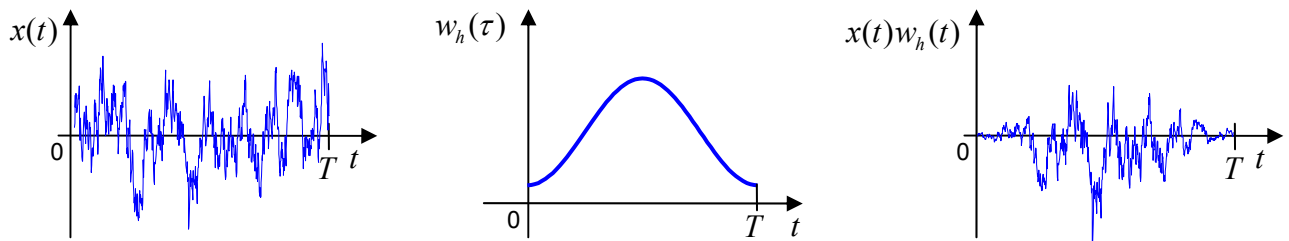
- 1) с помощью БПФ (функция **fft** в МАТЛАВ) найти массив X_m ($m = 0, 1, \dots, N-1$);
- 2) взяв первую половину этого массива, рассчитать соответствующие значения $F_X(\omega_m) = \Delta t \cdot X_m$ ($m = 0, 1, \dots, N/2$) для частот, не превышающих частоту Найквиста

$$\omega_{\max} = \omega_{N/2} = \frac{\pi}{\Delta t};$$

- 3) для каждой частоты ω_m ($m = 0, 1, \dots, N/2$) найти оценку спектральной плотности мощности

$$\text{по формуле: } \hat{S}_X(\omega_m) = \frac{|F_X(\omega_m)|^2}{T} = \frac{|F_X(\omega_m)|^2}{N \cdot \Delta t}.$$

Для сглаживания спектральной плотности так же, как и в методе Блэкмана-Тьюки, используются окна. Только теперь на весовую функцию умножается не оценка корреляционной функции, а сама реализация на интервале $[0; T]$:



Для этого случая окно Хэмминга на интервале $[0; T]$ принимает вид

$$w_h(t) = \begin{cases} 0,54 + 0,46 \cos \frac{2\pi}{T} \left(t - \frac{T}{2} \right), & 0 \leq t \leq T \\ 0, & t < 0, t > T \end{cases}.$$

Далее дискретное преобразование Фурье вычисляется для отсчетов взвешенной функции, то есть, вместо (6) получаем

$$F_X(\omega_m) = \Delta t \cdot \sum_{k=0}^{N-1} x_k w_k e^{-\frac{2\pi j}{N} mk}, \text{ где } w_k = w_h(k \cdot \Delta t).$$

Использование окна для исходного сигнала приводит к уменьшению его энергии и, как следствие, к заниженным оценкам спектральной плотности. Чтобы компенсировать эти потери, весо-

⁶ Иначе говоря, синусоиду нужно измерять более 2 раз за период.

⁷ Можно доказать, что $X_{N-k} = X_k^*$, где звездочка обозначает комплексно-сопряженную величину.

вая функция умножается на дополнительный коэффициент q , который часто выбирают из условия нормировки (сохранения энергии весовой функции окна, которая должна остаться такой же, как для прямоугольного окна):

$$\int_0^T q^2 w_h^2(t) dt = \int_0^T 1 dt = T.$$

Несложно подсчитать, что для окна Хэмминга из этого условия следует

$$q = \frac{1}{\sqrt{0,54^2 + 0,46^2 / 2}} \approx 1,586.$$

3.3. Прохождение случайных сигналов через линейные системы

Существует два подхода к исследованию систем управления при случайных возмущениях:

- 1) *вероятностный* – на основе плотностей распределения вероятностей;
- 2) *статистический* – с помощью усредненных характеристик: математического ожидания, дисперсии, корреляционной функции, спектральной плотности.

Применение вероятностного подхода, как правило, связано со значительными трудностями. С одной стороны, они вызваны недостатком информации о плотностях распределения случайных сигналов. С другой стороны, существующий математический аппарат достаточно сложен. Приведем только один важный факт: если входной сигнал имеет нормальное распределение, то на выходе линейной системы будет также сигнал с нормальным распределением (линейная система не «портит» нормальность).

В прикладных задачах нас чаще всего интересует не плотность распределения вероятностей на выходе системы, а некоторые более осязаемые характеристики – среднее значение, дисперсия и т.д. Поэтому в подавляющем большинстве случаев используется статистический подход. Далее мы будем предполагать, что на вход линейной системы с известной передаточной функцией $F(s)$ действует стационарный случайный процесс с заданной спектральной плотностью $S_x(\omega)$.

Прежде всего, отметим, что при стационарном случайном входе выход $y(t)$ линейной *стационарной* системы (у которой все характеристики не зависят от времени) – тоже стационарный случайный процесс. Для процесса $y(t)$ требуется найти

- математическое ожидание $\bar{y}$;
- дисперсию D_y ;
- корреляционную функцию $R_y(\tau)$;
- спектральную плотность $S_y(\omega)$.

Проще всего решается вопрос с математическим ожиданием: среднее значение выхода равно среднему значению входа, умноженному на статический коэффициент усиления системы (коэффициент усиления постоянного сигнала):

$$\bar{y} = k_F \bar{x}, \quad k_F = \lim_{s \rightarrow 0} F(s).$$

Учитывая, что в линейных системах справедлив *принцип суперпозиции* (реакция на сумму двух сигналов равна сумме реакций на отдельные сигналы), далее мы для простоты будем рассматривать только *центрированные* процессы, имеющие нулевые средние значения.

Остальные характеристики удобнее определять с помощью спектральной плотности выхода. Вспомним, что спектральная плотность – это плотность распределения мощности сигнала по частотам. Сначала рассмотрим, как изменяется мощность гармонического сигнала $x(t) = A \sin \omega_0 t$, когда он проходит через линейную систему. Из классической теории автоматического управления известно, что на выходе устанавливается гармонический процесс с той же частотой, но с другой амплитудой и фазой:

$$y(t) = B \sin(\omega_0 t + \varphi),$$

причем его амплитуда B определяется по частотной характеристике $F(j\omega_0)$:

$$B = A \cdot |F(j\omega_0)|.$$

Мощность гармонического сигнала (средний квадрат) пропорциональна квадрату амплитуды:

$$B^2 = A^2 \cdot |F(j\omega_0)|^2 = A^2 \cdot F(j\omega_0) \cdot F(-j\omega_0).$$

В последнем равенстве использовано свойство комплексного числа $F(j\omega_0)$ – квадрат его модуля равен произведению этого числа на комплексно-сопряженное, то есть на $F(-j\omega_0)$.

Таким образом, мы с помощью простых рассуждений вышли на очень важный результат: при гармоническом входе с частотой ω_0 мощность сигнала на выходе линейной системы равна мощности входного сигнала, умноженной на $F(j\omega_0)F(-j\omega_0)$. Учитывая, что это свойство справедливо на всех частотах, и заменив слово «мощность» на «спектральную плотность», получаем формулу, позволяющую сразу найти спектр процесса на выходе:

$$S_Y(\omega) = S_X(\omega) \cdot F(j\omega) \cdot F(-j\omega).$$

Соответствующая корреляционная функция $R_Y(\tau)$ может быть найдена как обратное преобразование Фурье от $S_Y(\omega)$. Для вычисления среднего квадрата процесса $y(t)$ нужно проинтегрировать $S_Y(\omega)$:

$$\overline{y^2} = \frac{1}{\pi} \int_0^\infty S_Y(\omega) d\omega.$$

Если процесс центрированный, средний квадрат совпадает с дисперсией, то есть $D_y = \overline{y^2}$.

В общем случае для вычисления дисперсии нужно использовать равенство $D_y = \overline{y^2} - (\bar{y})^2$.

Выделим один важный случай, когда входной сигнал – это *единичный* белый шум с постоянной спектральной плотностью $S_X(\omega) = 1$ (белый шум единичной интенсивности). Тогда получаем

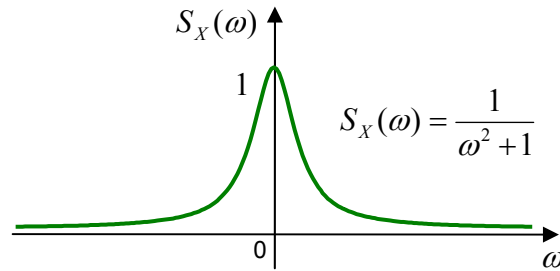
$$S_Y(\omega) = F(j\omega) \cdot F(-j\omega). \quad (7)$$

Таким образом, спектральная плотность выхода системы, на вход которой действует единичный белый шум, равна квадрату ее амплитудной характеристики.

Пусть передаточная функция линейной системы равна $F(s) = \frac{1}{s+1}$. Белый шум, проходя через такое звено, превращается в сигнал, имеющий спектральную плотность

$$S_Y(\omega) = F(j\omega)F(-j\omega) = \frac{1}{\omega^2 + 1},$$

график которой показан на рисунке:



Белый шум «содержит» все частоты, но они по-разному преобразуются. Постоянный сигнал (имеющий частоту $\omega = 0$) передается на выход системы без изменений. На низких частотах искажения достаточно малые, а высокие частоты подавляются (фильтруются) системой. Это типичный *фильтр низких частот* (он пропускает низкочастотные сигналы и блокирует высокочастотные). Отметим очень важный факт: поскольку высокие частоты подавляются, отклонения спектра входного сигнала от равномерного спектра белого шума в этой области не будут существенно влиять на спектр выхода. На этой идее основано компьютерное моделирование случайных процессов (см. далее).

3.4. Моделирование случайных сигналов

К сожалению, анализ системы далеко не всегда можно выполнить теоретически. Это особенно актуально для нелинейных систем. В этом случае единственным методом остается имитационное моделирование. Поэтому важно уметь моделировать случайные процессы, действующие на систему: возмущения (например, влияние ветра и волн на судно) и помехи измерения (погрешности измерительной системы).

3.4.1. Случайные числа

Моделирование случайных процессов на цифровых компьютерах основано на использовании случайной последовательности чисел. Обеспечить подлинную случайность в программе практически очень сложно (иногда для этого используют шум звуковой карты, счетчик тактов процессора). Поэтому обычно применяют генераторы *псевдослучайных* чисел. В последовательности псевдослучайных чисел каждое следующее число x_{k+1} рассчитывается по какой-то математической формуле на основе предыдущего x_k (или нескольких предыдущих). Например, во многих системах программирования используется *линейный конгруэнтный метод*:

$$x_{k+1} = (ax_k + c) \bmod m.$$

Здесь a , c и m – некоторые целые числа. По этой формуле вычисляются псевдослучайные числа, равномерно распределенные⁸ на интервале от 0 до $m-1$. Для краткости будем далее называть псевдослучайные числа просто случайными.

Параметры a , c и m нужно выбрать так, чтобы полученная последовательность содержала как можно меньше закономерностей и как можно дольше не повторялась. Это отдельная математическая проблема, которая до сих пор не имеет однозначного решения. Например, в функции **rand** среды MATLAB используется генератор с параметрами⁹

$$a = 7^5 = 16807, \quad c = 0, \quad m = 2^{31} - 1 = 2147483647.$$

⁸ Теоретически, конечно. Реальное распределение всегда немного неравномерное.

⁹ См. http://www.mathworks.com/company/newsletters/news_notes/pdf/Cleve.pdf.

Последовательность начинается с некоторого начального числа (англ. *seed* – семя, зерно). Чтобы получить в точности такую же последовательность (например, чтобы повторить полученные ранее результаты моделирования), нужно начать последовательность с того же числа.

Как показано в разд. 1.6, при сложении равномерно распределенных чисел распределение их суммы стремится к нормальному. Поэтому простейший способ получения нормального распределения – сложить некоторое количество случайных чисел, полученных с помощью стандартного датчика. Часто используют сумму 12 чисел, равномерно распределенных на интервале $[0;1]$, потому что это сразу дает случайную величину с единичной дисперсией.

В МАТЛАВ есть встроенная функция **randn**, которая генерирует последовательность случайных чисел с *нормальным* распределением (с нулевым средним и единичной дисперсией) на основе более сложного алгоритма¹⁰.

3.4.2. Формирующий фильтр

На практике обычно известна спектральная плотность $S_X(\omega)$ и требуется смоделировать процесс, имеющий такую спектральную плотность.

Вспомним, что спектральная плотность неотрицательна для любой частоты. Тогда функция $S_X(s)$, полученная при подстановке $\omega^2 = -s^2$ в $S_X(\omega)$, неотрицательна на мнимой оси, то есть при $s = j\omega$ для всех ω . Можно доказать, что в этом случае ее можно представить в виде произведения $S_X(s) = F(s)F(-s)$, то есть в форме (7). При этом всегда можно выбрать передаточную функцию $F(s)$ так, чтобы она была устойчивой (не имела полюсов в правой полуплоскости) и минимально-фазовой (не имела нулей в правой полуплоскости). Такой переход от $S_X(s)$ к $F(s)$ называется *факторизацией* (англ. *разложение на множители*).

Как следует из (7), если на вход звена с передаточной функцией $F(s)$ подать единичный белый шум, процесс на выходе будет иметь заданную спектральную плотность $S_X(\omega)$. Функция $F(s)$ называется *передаточной функцией формирующего фильтра*.

Проще всего моделировать процессы с дробно-рациональной спектральной плотностью. Например, одна из моделей морского волнения (подробности см. в главе 4) описывается спектральной плотностью

$$S(\omega) = \frac{4D_r\alpha\omega^2}{\omega^4 + 2(\alpha^2 - \beta^2)\omega^2 + (\alpha^2 + \beta^2)^2},$$

где D_r – дисперсия волновой ординаты, α – коэффициент затухания и β – частота максимума спектра. Заменяя ω^2 на $-s^2$, получаем

$$S(s) = \frac{-4D_r\alpha s^2}{s^4 - 2(\alpha^2 - \beta^2)s^2 + (\alpha^2 + \beta^2)^2}.$$

Очевидно, что формирующий фильтр будет иметь передаточную функцию вида

$$F(s) = \frac{\gamma s}{s^2 + \delta_1 s + \delta_0},$$

так что

¹⁰ См. http://www.mathworks.com/company/newsletters/news_notes/clevescorner/spring01_cleve.html.

$$F(s) F(-s) = \frac{-\gamma^2 s^2}{s^4 - (\delta_1^2 - 2\delta_0)s^2 + \delta_0^2}.$$

Приравнявая коэффициенты числителя и знаменателя $S(s)$ и $F(s)F(-s)$, сразу находим:

$$\gamma = 2\sqrt{D_r\alpha}, \quad \delta_0 = \alpha^2 + \beta^2, \quad \delta_1 = \sqrt{2(\alpha^2 - \beta^2 + \delta_0)} = 2\alpha.$$

В более сложных случаях факторизация выполняется с помощью численных методов. Нужно разложить на простейшие сомножители числитель и знаменатель $S(s)$ и включить в $F(s)$ только те множители, корни которых находятся в левой полуплоскости (в числителе допускаются корни на мнимой оси). Посмотрим, как работает этот метод на том же примере. Числитель спектральной плотности $S(s)$ имеет двойной корень в точке $s = 0$, его можно представить в симметричном виде

$$-4D_r\alpha s^2 = 2\sqrt{D_r\alpha} \cdot s \cdot 2\sqrt{D_r\alpha} \cdot (-s).$$

Знаменатель имеет корни в точках $\alpha \pm j\beta$ и $-\alpha \pm j\beta$. Учитывая, что $\alpha > 0$, определяем, что корни $-\alpha \pm j\beta$ расположены в левой полуплоскости. Поэтому знаменатель $F(s)$ представляет собой произведение $(s + \alpha + j\beta)(s + \alpha - j\beta) = s^2 + 2\alpha s + \alpha^2 + \beta^2$. Таким образом,

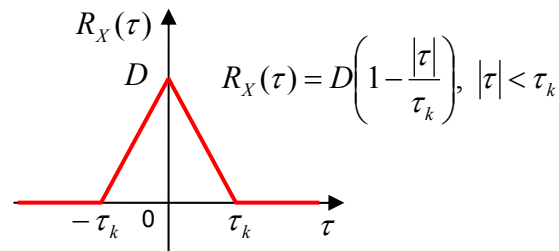
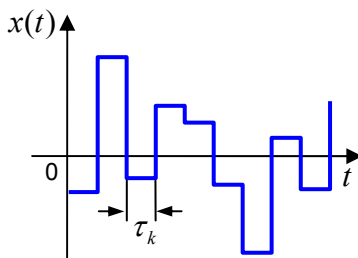
$$S(s) = \frac{-4D_r\alpha s^2}{s^4 - 2(\alpha^2 - \beta^2)s^2 + (\alpha^2 + \beta^2)^2} = F(s)F(-s), \quad \text{где} \quad F(s) = \frac{2\sqrt{D_r\alpha} \cdot s}{s^2 + 2\alpha s + \alpha^2 + \beta^2}.$$

Такой же результат был получен ранее методом неопределенных коэффициентов.

Итак, формирующий фильтр мы построили. Теперь остается один очень практический вопрос: как получить белый шум, который, как известно, является сигналом с бесконечной энергией? Вспомним, что белый шум – это только вспомогательный сигнал, который, проходя через систему с передаточной функцией $F(s)$, генерирует сигнал с заданной спектральной плотностью. Оказывается, можно заменить его на другой сигнал (который просто получить на компьютере), и при этом спектральная плотность выхода оказывается достаточно близка к заданной.

3.4.3. Случайный ступенчатый сигнал вместо белого шума

Как мы видели, на компьютере несложно получить последовательность случайных (точнее – псевдослучайных) чисел с равномерным или нормальным распределением. По этим числам можно построить ступенчатый сигнал, фиксируя каждое значение в течение некоторого времени τ_k :



Теоретически эти числа некоррелированы¹¹; при этом можно показать, что корреляционная функция $R_X(\tau)$ ступенчатого сигнала – треугольная (см. рисунок справа). При $\tau = 0$ она равна дисперсии D последовательности случайных чисел, а при $|\tau| > \tau_k$ обращается в нуль (по-

¹¹ Строго говоря, это не совсем так. Из-за неидеальности компьютерного датчика псевдослучайных чисел эти значения будут коррелированы и корреляционная функция будет немного отличаться от треугольной.

тому что моменты времени t и $t + \tau$ находятся на разных интервалах и, следовательно, соответствующие значения некоррелированы). Число τ_k называют *интервалом корреляции* – так называется интервал, после которого можно считать корреляционную функцию (примерно) равной нулю.

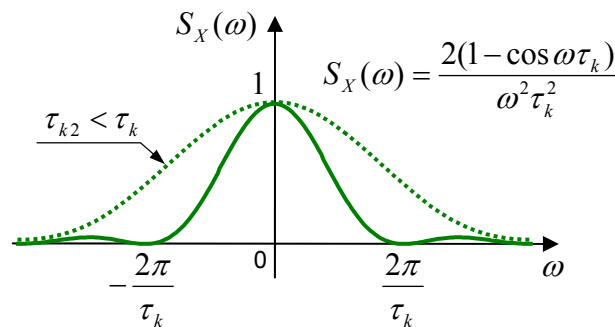
Взяв преобразование Фурье от корреляционной функции

$$R_X(\tau) = \begin{cases} D \left(1 - \frac{|\tau|}{\tau_k} \right), & |\tau| < \tau_k \\ 0, & |\tau| \geq \tau_k \end{cases}$$

получаем спектральную плотность

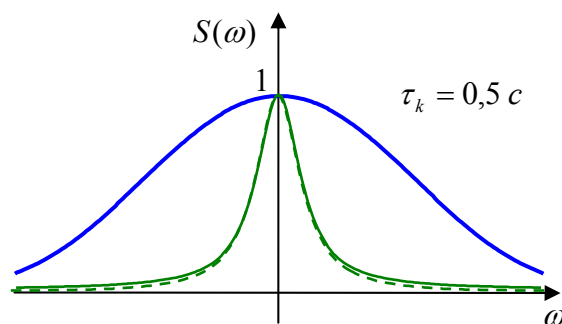
$$S_X(\omega) = \frac{2D(1 - \cos \omega \tau_k)}{\omega^2 \tau_k}.$$

Вычисляя предел этой функции при $\omega \rightarrow 0$, находим $\lim_{\omega \rightarrow 0} S_X(\omega) = D \tau_k$, так что при выборе $D = 1/\tau_k$ это значение равно 1 (как у белого шума). Заметим, что $S_X(\omega) = 0$, когда $\cos \omega \tau_k = 1$, то есть $\omega \tau_k = 2\pi m$ при любом целом m . Форма спектральной плотности показана на рисунке ниже (здесь и далее принимается $D = 1/\tau_k$):

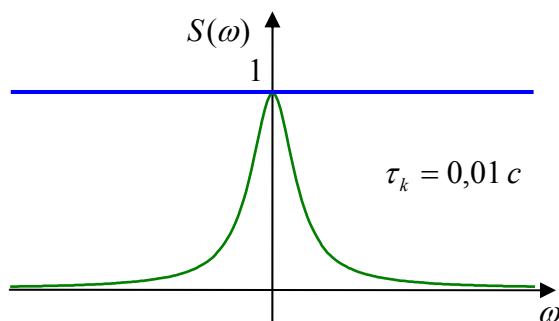


Конечно, это далеко не белый шум, у которого спектральная плотность должна быть постоянной на всех частотах. Тем не менее, при уменьшении интервала корреляции τ_k «колокол» расширяется, и для низких частот можно считать, что $S_X(\omega) \approx 1$. В пределе при $\tau_k \rightarrow 0$ спектр стремится к равномерному спектру единичного белого шума. Далее будет показано, что при грамотном выборе τ_k такой сигнал можно использовать в качестве источника вместо белого шума.

Для примера предположим, что нужно получить сигнал со спектральной плотностью $S_Y(\omega) = \frac{1}{\omega^2 + 1}$, то есть формирующий фильтр имеет передаточную функцию $F(s) = \frac{1}{s + 1}$. В качестве входного сигнала для этого звена будем использовать описанный выше ступенчатый сигнал при $\tau_k = 0,5$ с. На рисунке приведены графики спектральной плотности ступенчатого сигнала (синяя линия), желаемой спектральной плотности (сплошная зеленая линия) и фактической спектральной плотности выхода (штриховая линия).



По графику видно, что в существенной полосе частот (где частотная характеристика звена ненулевая) спектр входного сигнала существенно неравномерный, поэтому желаемый и фактический спектры на выходе системы *немного* различаются в области высоких частот. Приняв $\tau_k = 0,01$ с, имеем совершенно другую картину:



Спектр входного сигнала в интересующей нас области практически равномерный, в спектр реального выхода практически точно совпадает с заданным.

Очевидно, что при выборе τ_k нужно учитывать частотные свойства формирующего фильтра, точнее, полосу частот, где его частотная характеристика достаточно отличается от нуля. Для этого используют понятие *полосы пропускания* системы ω_b – так называется частота, для которой амплитудная частотная характеристика уменьшается на 3 дБ (децибела) в сравнении с максимальным значением (составляет примерно 0,708 от максимума). Разработчики MATLAB рекомендуют при моделировании использовать значение

$$\tau_k = \frac{1}{100} \cdot \frac{2\pi}{\omega_b}.$$

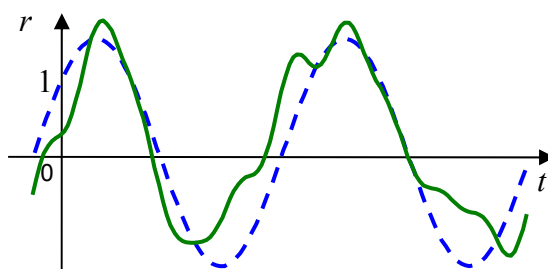
В нашем случае амплитудной частотная характеристика (апериодического звена) имеет вид $|F(j\omega)| = \frac{1}{\sqrt{\omega^2 + 1}}$, ее максимум равен 1 (при $\omega = 0$), поэтому полоса пропускания определяется равенством

$\frac{1}{\sqrt{\omega_b^2 + 1}} = 0,708$. Отсюда следует $\omega_b = \sqrt{\frac{1}{0,708^2} - 1} \approx 0,998$, так что $\tau_k \approx 0,063$.

4. Морское волнение

4.1. Что такое морское волнение?

Морское волнение – это колебания поверхности воды, вызванные ветром (а также приливами, отливами и другими причинами). Простейшая модель морского волнения – гармонические колебания поверхности, когда *волновая ордината* (вертикальная координата точек поверхности) изменяется по закону синуса. Такое волнение называют *регулярным* (синяя штриховая линия на рисунке).



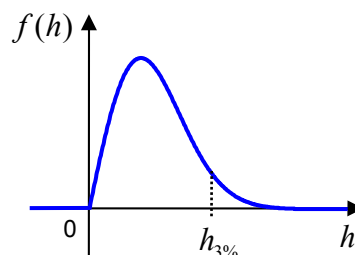
Однако, в самом деле волновая ордината меняется по более сложному закону, его можно (опять-таки приближенно) представить как сумму большого количества гармоник (синусоид) с разными амплитудами и фазами (теоретически – это сумма бесконечного числа гармоник). Это так называемое *нерегулярное* волнение, которое чаще всего описывается как случайный процесс. Можно предположить, случайные волновые ординаты распределены по нормальному закону, и это действительно подтверждается экспериментами. При этом плотность распределения высот волн подчиняется закону Рэлея.

В нашей стране для характеристики волнения используется *высота волны 3%-ной обеспеченности* $h_{3\%}$ – это высота волны, вероятность превышения которой составляет 3%. Значения $h_{3\%}$ определяются по таблице в зависимости от интенсивности волнения (в баллах):

Волнение, баллы	$h_{3\%}$, м	Волнение, баллы	$h_{3\%}$, м	Волнение, баллы	$h_{3\%}$, м
1	0 – 0,25	4	1,25 – 2,0	7	6,0 – 8,5
2	0,25 – 0,75	5	2,0 – 3,5	8	8,5 – 11
3	0,75 – 1,25	6	3,5 – 6,0	9	более 11

По значению $h_{3\%}$ можно определить *дисперсию волновой ординаты* D_r . Известно, что высота волны h – это случайная величина, распределенная по закону Рэлея:

$$f(h) = \frac{h}{4D_r} \exp\left(-\frac{h^2}{8D_r}\right), \quad h \geq 0.$$



Вероятность превышения некоторого уровня $h_{3\%}$ можно рассчитать как интеграл от плотности вероятности на интервале $[h_{3\%}; \infty)$:

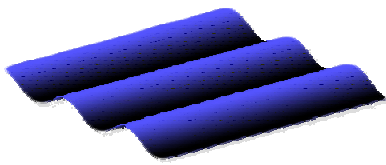
$$p(h > h_{3\%}) = \int_{h_{3\%}}^{\infty} \frac{h}{4D_r} \exp\left(-\frac{h^2}{8D_r}\right) dh = \exp\left(-\frac{h_{3\%}^2}{8D_r}\right) = \exp\left(-\frac{h_{3\%}^2}{8D_r}\right).$$

Эта вероятность должна быть равна 3% или 0,03; поэтому

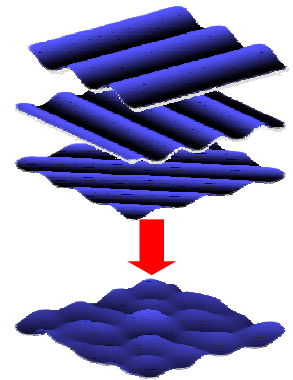
$$\exp\left(-\frac{h_{3\%}^2}{8D_r}\right) = 0,03 \quad \Rightarrow \quad D_r = -\frac{h_{3\%}^2}{8 \ln 0,03} \approx 0,0356 h_{3\%}^2.$$

Значение $h_{3\%}$ определяет также *средний период волнения* $\bar{T}$, *среднюю частоту волнения* $\bar{\omega}$ и *частоту максимума спектра* ω_m :

$$\bar{T} \approx 3,1 \sqrt{h_{3\%}}, \quad \bar{\omega} = \frac{2\pi}{\bar{T}}, \quad \omega_m \approx 0,71 \bar{\omega}.$$



Существуют два типа моделей нерегулярного волнения, двухмерная и трехмерная. В *двухмерной* модели предполагается, что гребни волн имеют бесконечную длину и перемещаются параллельно друг другу в одном направлении (рисунок слева). Каждый, кто видел реальное волнение, знает, что на практике это не так. Более точная (но и более сложная) *трехмерная* модель учитывает сложение множества двухмерных волн, идущих в разных направлениях (рисунок справа).



Строго говоря, волнение – это нестационарный процесс. Оно начинается с ряби, затем ветер за счет завихрений потоков воздуха разгоняет волны. Но если ветер с постоянными свойствами действует достаточно длительное время (несколько часов) на большой акватории (десятки километров), можно говорить о *развитом* (установившемся) волнении, которое считают стационарным эргодическим процессом. Такой подход позволяет использовать методы анализа случайных процессов на основе корреляционных функций и спектральных плотностей.

Под термином «спектр морского волнения» обычно понимают спектральную плотность волновой ординаты. С экспериментальными данными лучше всего согласуются экспоненциальные спектры вида

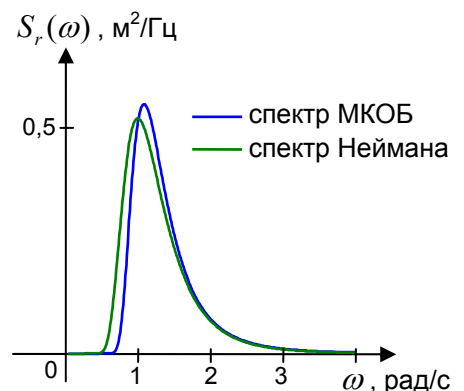
$$S_r(\omega) = A \omega^{-k} \exp(-B \omega^{-n}),$$

где параметры A , B характеризуют интенсивность волнения, а k и n зависят от его особенностей.

Вообще говоря, спектры волнения в разных районах отличаются. Если информации недостаточно, рекомендуют использовать типовой спектр Пирсона и Мошковица ($k = 5$ и $n = 4$), рекомендованный 12-й Международной конференцией опытовых бассейнов (МКОБ):

$$S_r(\omega) = 7,06\pi \frac{D_r}{\bar{\omega}} \left(\frac{\omega_m}{\omega}\right)^5 \exp\left[-1,25\left(\frac{\omega_m}{\omega}\right)^4\right].$$

Известно множество других выражений для спектров. Часто используют, например, спектр Неймана:



$$S_r(\omega) = 2\pi \frac{1,2}{\omega^6} \exp\left[-\frac{4,575}{\omega^2 h_{3\%}^{0,8}}\right].$$

На рисунке справа показаны графики спектра МКОБ (синяя линия) и спектра Неймана (зеленая линия) для волнения 4 балла ($h_{3\%} = 2$ м).

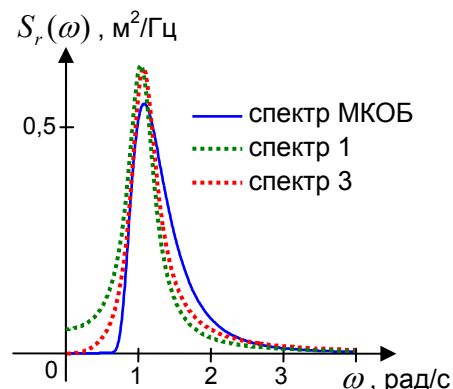
Экспоненциальные спектры хорошо согласуются с экспериментальными данными, однако неудобны для моделирования и расчета систем управления. Вместо них используются дробно-рациональные спектры

$$1: \quad S_r(\omega) = \frac{2D_r\alpha(\alpha^2 + \beta^2 + \omega^2)}{\omega^4 + 2(\alpha^2 - \beta^2)\omega^2 + (\alpha^2 + \beta^2)^2},$$

$$2: \quad S_r(\omega) = \frac{4D_r\alpha(\alpha^2 + \beta^2)}{\omega^4 + 2(\alpha^2 - \beta^2)\omega^2 + (\alpha^2 + \beta^2)^2},$$

$$3: \quad S_r(\omega) = \frac{4D_r\alpha\omega^2}{\omega^4 + 2(\alpha^2 - \beta^2)\omega^2 + (\alpha^2 + \beta^2)^2}.$$

Здесь D_r – дисперсия волновой ординаты, α – коэффициент затухания и β – угловая частота корреляционных функций. Для развитого морского волнения $\beta = 1,02\omega_m$ и $\alpha = 0,21\beta$.



На рисунке справа показаны спектр МКОБ и дробно-рациональные спектры 1 и 3 (кривая для спектра 2 проходит между ними). Графики дробно-рациональных спектров проходят выше экспоненциальных на низких частотах, которые больше всего влияют на поведение инерционных морских объектов (судов). Поэтому считается, что использование дробно-рациональных спектров приводит к несколько завышенным оценкам параметров движения (качки, рыскания).

Заметим, что спектр 3 равен нулю на нулевой частоте, так же, как и экспоненциальные спектры. Однако, если учитывать трехмерность волнения, по современным представлениям спектр на нулевой частоте в большинстве случаев отличен от нуля. Это оправдывает использование дробно-рациональных спектров 1 и 2.

4.2. Кажущиеся спектры

До этого мы рассматривали спектры волнения без учета движения судна. Очевидно, что двигаться против волнения труднее, чем «по волне», а движение судна «лагом» к волне (при бортовой волне) вызывает сильную качку. Это означает, что эффект действия волнения на судно зависит от скорости и направления его движения по отношению к основному направлению распространения волн. Математически это выражается в изменении спектральной плотности возмущения.

В физике хорошо известен эффект Доплера, который состоит в том, что при движении датчика измеренная им частота волны изменяется в зависимости от его собственной скорости по закону $\omega_k = \omega \cdot \left(1 + \frac{v}{u}\right)$, где ω – истинная частота волны, ω_k – кажущаяся (измеренная) частота, v – скорость движения датчика в направлении источника волн, а u – скорость распространения самих волн. При движении судна против волнения ($v > 0$) кажущаяся частота волн будет больше, чем истинная, а при движении «по волнению» ($v < 0$) – меньше истинной.

Из гидродинамики известно, что скорость распространения волны с частотой ω (на глубокой воде) равна $u = \frac{g}{\omega}$. Кроме того, если судно движется со скоростью V (в м/с) под углом ξ к направлению распространения волн (считается, что $\xi = 0$ соответствует движению против волны), оно приближается к источнику со скоростью $v = V \cos \xi$. Таким образом, формула преобразования истинной частоты в кажущуюся приобретает вид $\omega_k = \omega \cdot \left(1 + \frac{V \cos \xi}{g} \omega \right)$.

Если судно «убегает» от волн, то при увеличении скорости кажущаяся частота, вычисленная по этой формуле, оказывается отрицательной – это означает, что судно обгоняет волны. Поэтому для того, чтобы работать только с положительными частотами, нужно взять модуль выражения в правой части. Окончательно получаем

$$\omega_k = \left| \omega \cdot (1 + a\omega) \right|, \quad (8)$$

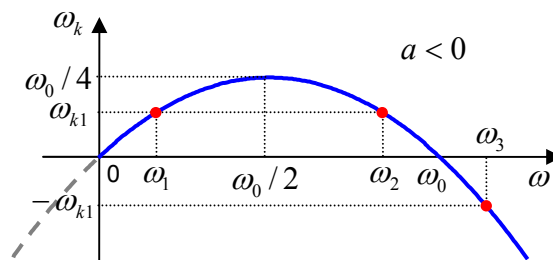
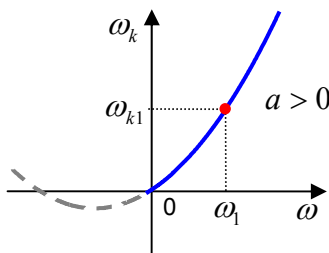
где $a = \frac{V \cos \xi}{g}$ – фактор относительного движения. Итак, составляющую волнения с частотой ω судно воспринимает с кажущейся частотой ω_k .

Чтобы построить кажущийся спектр, нужно научиться решать обратную задачу – по заданной кажущейся частоте ω_k определить частоту (или частоты!) исходного спектра, которые судно воспринимает как ω_k . Используя (8), формально получаем два квадратных уравнения

$$a\omega^2 + \omega + \omega_k = 0 \quad \text{и} \quad a\omega^2 + \omega - \omega_k = 0, \quad (9)$$

Нас интересуют все вещественные и положительные решения этих уравнений (фактически таких «подходящих» частот может быть от одной до трех).

Предположим, что судно идет под острым углом к волне, так что $\cos \xi > 0$ и $a > 0$. В этом случае первое уравнение в (9) заведомо не имеет подходящих решений (поскольку $a\omega^2 + \omega \geq 0$), а второе имеет только одно положительное решение. На рисунке слева изображен график функции $\omega_k = a\omega^2 + \omega$, который ясно показывает, что одной кажущейся частоте $\omega_{k1} > 0$ соответствует только одна частота ω_1 исходного спектра и наоборот.



Теперь предположим, что судно идет под тупым углом к волне («убегает» от волн), так что $\cos \xi < 0$ и $a < 0$. На рисунке справа показан график функции $\omega_k = a\omega^2 + \omega$ для этого случая. Используя знания школьной математики, легко увидеть, что парабола пересекает ось $\omega_k = 0$ при $\omega = \omega_0 = -\frac{1}{a}$, а ее вершина находится в точке $\left(\frac{\omega_0}{2}, \frac{\omega_0}{4} \right)$. Все волны с частотами больше ω_0 судно опережает, поэтому кажущаяся частота становится отрицательной.

Если интересующая нас кажущаяся частота ω_{k1} меньше максимальной (равной $\omega_0/4$), то энергия волн на трех частотах (ω_1 , ω_2 и ω_3) складывается на частоте ω_{k1} кажущегося спектра. Если $\omega_{k1} > \omega_0/4$, остается только одно решение – ω_3 .

Нужно понимать, что мощность волнения (средний квадрат волновой ординаты) не зависит от скорости движения судна. Для движущегося объекта эта мощность просто перераспределяется по частотам. Поэтому при расчете кажущегося спектра важно сохранить мощность. Это значит, что средние квадраты волновой ординаты для исходного и кажущегося спектров должны быть равны:

$$\frac{1}{\pi} \int_0^{\infty} S_r(\omega) d\omega = \frac{1}{\pi} \int_0^{\infty} S_{rk}(\omega_k) d\omega_k.$$

Чтобы добиться этого, приравнивают «элементарные» мощности:

$$S_r(\omega) d\omega = S_{rk}(\omega_k) d\omega_k, \quad \text{что дает} \quad S_{rk}(\omega_k) = S_r(\omega) \frac{d\omega}{d\omega_k}.$$

Выражение $d\omega/d\omega_k$ в правой части последнего равенства можно найти как обратное значение для производной (при $a > 0$)

$$\frac{d\omega_k}{d\omega} = \frac{d}{d\omega} (a\omega^2 + \omega) = 2a\omega + 1, \quad \text{так что} \quad \frac{d\omega}{d\omega_k} = \frac{1}{2a\omega + 1}.$$

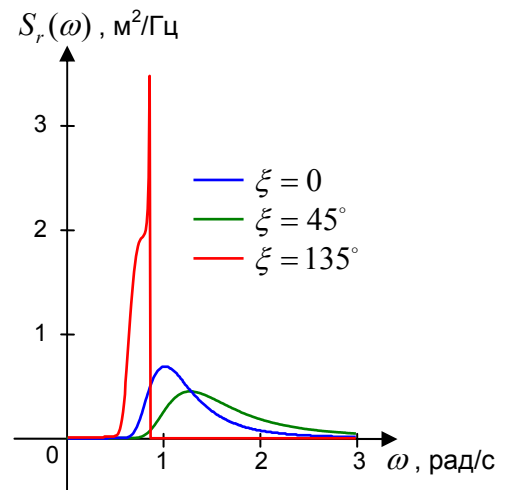
При $a < 0$ нужно взять эту производную по модулю (чтобы все составляющие спектра складывались, а не вычитались). В самом сложном случае, когда на частоте ω_{k1} кажущегося спектра складываются *три* составляющие исходного спектра (соответствующие частотам ω_1 , ω_2 и ω_3), при численном пересчете спектра имеем:

$$S_{rk}(\omega_{k1}) = \frac{S_r(\omega_1)}{|1 + 2a\omega_1|} + \frac{S_r(\omega_2)}{|1 + 2a\omega_2|} + \frac{S_r(\omega_3)}{|1 + 2a\omega_3|}.$$

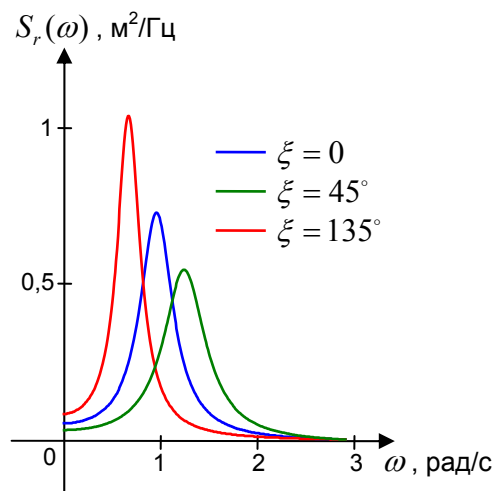
На рисунке показаны спектры, полученные преобразованием спектра МКОБ (Пирсона и Мошковица) при разных углах встречи с волной (волнение 4 балла, $h_{3\%} = 2$ м, скорость судна 4 м/с).

При встречном волнении (зеленая линия) максимум спектра смещается в область высоких частот, спектр «размазывается» по оси ω . Учитывая, что инерционный объект подавляет высокочастотные возмущения, можно считать, что влияние волнения на судно уменьшается.

При попутном волнении (красная линия) спектр имеет разрыв (скачок), такие разрывы подтверждаются экспериментально. Вся энергия волн сосредоточена в узкой полосе на низких частотах. При этом возмущение оказывает очень сильное влияние на судно, и это влияние сложно (а иногда и невозможно) скомпенсировать с помощью управления. Поэтому движение на попутном волнении – это один из самых сложных режимов движения в точки зрения стабилизации курса.



Пересчет дробно-рациональных спектров выполняется достаточно просто. Вычисляют частоту максимума кажущегося спектра $\omega_{mk} = |\omega_m + a\omega_m^2|$, и в приведенных выражениях для спектров используют $\beta_k = 1,02\omega_{mk}$ и $\alpha_k = 0,21\beta_k$ вместо α и β . Дисперсия D_r сохраняется та же самая. На рисунке слева показаны дробно-рациональные спектры (типа 1) для разных углов встречи с волной (волнение 4 балла, $h_{3\%} = 2$ м, скорость судна 4 м/с). Видно, что общий характер смещения спектров такой же, как и для пересчитанного спектра МКОБ.



4.3. Моделирование действия морского волнения на судно

Спектры, рассмотренные выше, относились к изменению волновой ординаты, тогда как для разработчика систем управления важно выяснить влияние волнения на динамику судна (угол рыскания, качку и т.п.). Как связано изменение волновой ординаты и рыскание (качка) судна? К сожалению, ответ на этот вопрос не прост и требует довольно полной информации о судне, например, об обводах корпуса. В конечном счете, (для линейной модели) требуется построить частотную характеристику $H(j\omega)$, которая связывает волновую ординату и требуемую характеристику движения судна, например, угол рыскания или угол крена.

Рассмотрим более подробно задачу стабилизации судна по курсу. Простейшая линейная модель судна – это модель первого порядка (модель Номото), которая связывает угловую скорость вращения вокруг вертикальной оси (оси Z) и угол перекладки руля:

$$\dot{\omega}_z = -\frac{1}{T_S} \omega_z + \frac{K}{T_S} \delta,$$

где ω_z – угловая скорость (в рад/с), δ – угол поворота руля (в радианах), T_S – постоянная времени (в секундах) и K – безразмерный коэффициент. Угловая скорость (в такой простейшей модели) равна производной от угла рыскания φ (так называется угол отклонения от заданного курса). Поэтому к модели нужно добавить еще одно уравнение:

$$\dot{\varphi} = \omega_z.$$

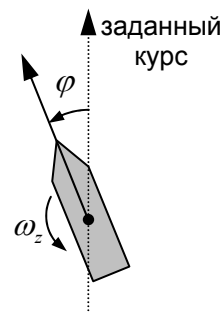
Обычно угол рыскания считается положительным при вращении против часовой стрелки.

Волновое возмущение приводит к тому, что появляется дополнительное вращение, отклоняющее судно от заданного курса, то есть, уравнение Номото с учетом возмущения принимает вид

$$\dot{\omega}_z = -\frac{1}{T_S} \omega_z + \frac{K}{T_S} \delta + w, \quad (10)$$

где w – возмущение, вызванное морским волнением. Будем считать, что все случайные процессы в системе – центрированные, то есть, их математические ожидания равны нулю. Спектр $S_w(\omega)$ возмущения w рассчитывается по общей формуле

$$S_w(\omega) = H(j\omega)H(-j\omega)S_r(\omega),$$



где $S_r(\omega)$ – спектральная плотность волновой ординаты, а $H(j\omega)$ – частотная характеристика, описывающая реакцию судна на волнение. Для построения $H(j\omega)$ нужно использовать уравнения движения конкретного судна.

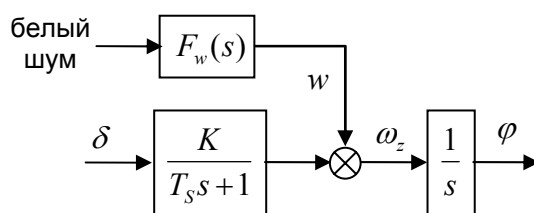
Спектральную плотность $S_w(\omega)$ часто удается достаточно успешно аппроксимировать дробно-рациональным спектром вида

$$\hat{S}_w(\omega) = \frac{K_w^2 \omega^2}{\omega^4 + 2\omega_0^2(\lambda^2 - 1)\omega^2 + \omega_0^4},$$

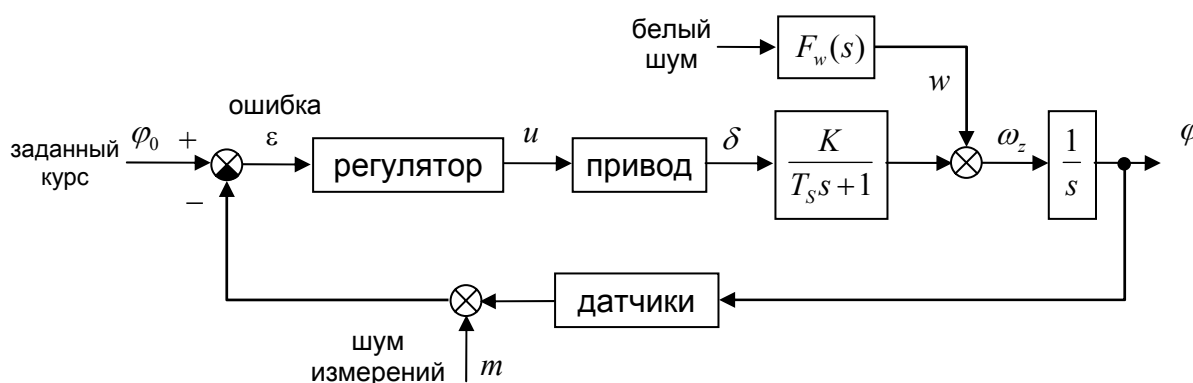
где ω_0 – доминирующая частота волн, λ – коэффициент демпфирования ($0 < \lambda < 1$), $K_w = 2\lambda\omega_0\sigma_w$ и σ_w – коэффициент, определяющий интенсивность волнения. Формирующий фильтр для такой спектральной плотности имеет вид

$$F(s) = \frac{K_w s}{s^2 + 2\lambda\omega_0 s + \omega_0^2}.$$

Легко проверить, что $S_w(\omega) = F(j\omega)F(-j\omega)$. Таким образом, для моделирования волнения нужно пропустить белый шум через формирующий фильтр с передаточной функцией $F(s)$ и полученный сигнал $w(t)$ подключить к модели Номото (10). Соответствующая блок-схема системы выглядит так:



Эта схема описывает разомкнутую систему (без обратной связи). Система управления с обратной связью включает также регулятор, привод руля, измерительную систему (датчики):



Точность стабилизации на курсе определяется среднеквадратическим отклонением (СКВО) σ_φ угла рыскания (или его дисперсией). Кроме того, требуется, чтобы мощность управления (СКВО угла перекладки руля σ_δ) не была слишком большой. Как правило, эти задачи противоречивы. Это значит, что для уменьшения рыскания нужно увеличивать мощность управления, а при уменьшении активности руля увеличивается рыскание.

Кроме внешних возмущений, нужно обязательно учитывать шум (помехи) измерений (сигнал m на схеме). Дело в том, что неправильно спроектированный регулятор может хорошо работать в идеальных условиях, но приводить к «раскачиванию» системы при малейших ошибках измерений.

Шум измерений обычно моделируется как белый шум, имеющий равномерный спектр на всех частотах. Поскольку объект (апериодическое звено) является фильтром низких частот, сигнал φ (изменение курса) не может быть высокочастотным. Поэтому регулятор должен подавлять высокочастотные сигналы, которые явно вызваны ошибками измерений.

5. Оптимизация систем при случайных возмущениях

5.1. Что такое оптимальная система?

Слово «оптимальный» означает «наилучший в некотором отношении». Для того, чтобы этот термин имел смысл, нужно определить, как именно (по какому *критерию*) мы будем оценивать систему управления. Критерий $I(x)$ – это оценка качества системы в виде числа, зависящая от выбора некоторых изменяемых параметров x . Эти параметры нужно выбрать так, чтобы обеспечить минимум или максимум критерия:

$$I(x) \rightarrow \min \quad \text{или} \quad I(x) \rightarrow \max .$$

В первом случае критерий выражает потери (расходы, убыток), в этом случае функция $I(x)$ часто называется *функцией потерь*. Во втором случае $I(x)$ – это доходы (прибыль и т.п.). Если критерий $I(x)$ зависит от выбора некоторой *функции* (например, от неизвестной передаточной функции регулятора), величину $I(x)$ называют *функционалом*.

В теории управления чаще всего рассматривается задача на минимум какого-то критерия (определяющего ошибку). Решение называется *оптимальным*, если выбранное x обеспечивает минимальное значение $I(x)$ среди всех допустимых решений. При этом нельзя сказать, что такая система «самая лучшая». Она является лучшей только по выбранному критерию, а по другим показателям она может вести себя неудовлетворительно.

На практике к системе обычно предъявляется много разных требований, то есть в идеале хочется сделать минимальными несколько величин (например, уменьшить рыскание судна и одновременно снизить активность руля). Такие задачи называются задачами *многокритериальной* или *многоцелевой* оптимизации. Как правило, отдельные требования противоречивы (чтобы уменьшить рыскание, нужно увеличивать мощность управления). В этом случае задача не имеет единственного решения; оптимальным (или *Парето-оптимальным*¹²) называют любое решение, для которого улучшение по любому показателю невозможно без ухудшения какого-либо другого.

Обычно стараются составить один общий критерий, который отражает все наиболее важные требования к системе. Однако это тоже «палка о двух концах» – любое усложнение критерия затрудняет поиск оптимального решения, особенно аналитическими методами.

Часто такой единый критерий представляет собой *взвешенную сумму* отдельных показателей качества $I_1(x), I_2(x), \dots, I_N(x)$, которые складываются после умножения на весовые коэффициенты $k_1, k_2, \dots, k_N$ (в большинстве случаев все они неотрицательны):

$$I(x) = k_1 I_1(x) + k_2 I_2(x) + \dots + k_N I_N(x) \rightarrow \min .$$

Любое решение этой задачи при некоторых коэффициентах является Парето-оптимальным, поскольку для него невозможно *одновременно* улучшить все показатели. Коэффициенты выбираются методом проб и ошибок так, чтобы полученное компромиссное решение удовлетворяло техническому заданию.

5.2. Оптимальная фильтрация

5.2.1. Постановка задачи

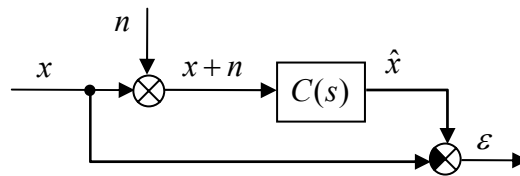
Задача любой системы управления – поддерживать заданный режим. В идеале ошибка – отклонение от этого режима – должна быть нулевой. Если на систему действуют случайные возмущения, даже в установившемся режиме нельзя обеспечить нулевую ошибку, поскольку

¹² В честь итальянского экономиста В. Парето, предложившего такой подход.

эти возмущения все время будут выводить систему из состояния равновесия. Таким образом, ошибка тоже будет случайным процессом.

В этом случае часто требуется выбрать регулятор так, чтобы сделать ошибку минимальной в статистическом смысле: уменьшить, насколько возможно, ее дисперсию (или, эквивалентно, среднеквадратическое отклонение). Заметим, что в этой задаче могут быть выбраны и другие критерии качества. Например, можно потребовать, чтобы дисперсия ошибки не превышала заданное значение при всех возможных возмущениях (задача *гарантирующего управления*). В *минимаксной задаче* нужно обеспечить минимум ошибки в самом худшем случае.

Рассмотрим простейшую задачу оптимизации линейной системы при случайных возмущениях – задачу *оптимальной фильтрации*. При измерениях полезный сигнал $x(t)$ искажается помехой $n(t)$, причем и полезный сигнал, и помеха – стационарные центрированные случайные процессы. Предполагается, что помеха и полезный сигнал статистически независимы, то есть, никак не связаны, порождаются разными источниками.



Задача состоит в том, чтобы с помощью некоторого фильтра выделить сигнал из его смеси с помехой наилучшим образом – построить его наилучшую оценку $\hat{x}(t)$. Наилучшей оценкой будем считать такую, при которой дисперсия D_ε ошибки $\varepsilon = \hat{x} - x$ минимальна.

Предполагается, что известны только спектральные плотности полезного сигнала $S_X(\omega)$ и помехи $S_N(\omega)$. Если фильтр $C(s)$ – *линейный*, этих данных достаточно для того, чтобы

- 1) найти дисперсию ошибки при известном фильтре;
- 2) построить оптимальный линейный фильтр.

При этом результат не зависит от плотностей распределения вероятности входных сигналов. Если фильтр нелинейный, для решения обеих задач требуется знать плотности вероятности. Тем не менее, оптимальный фильтр все равно будет линейным, если сигнал и помеха имеют нормальное (гауссово) распределение.

Таким образом, при заданных спектральных плотностях $S_X(\omega)$ и $S_N(\omega)$ требуется найти передаточную функцию линейного фильтра $C(s)$, который обеспечивает минимальную дисперсию D_ε ошибки ε . Эту задачу впервые решил американский математик Норберт Винер в 1940-х годах, поэтому она называется задачей Винера, а соответствующий оптимальный фильтр – *фильтром Винера*.

5.2.2. Фильтр Винера

Вспомним, что знание спектральной плотности $S_\varepsilon(\omega)$ ошибки позволяет найти ее средний квадрат, равный дисперсии для центрированных процессов:

$$D_\varepsilon = \frac{1}{2\pi} \int_{-\infty}^{\infty} S_\varepsilon(\omega) d\omega = \frac{1}{2\pi j} \int_{-j\infty}^{j\infty} S_\varepsilon(s) ds. \quad (11)$$

В последнем равенстве от угловой частоты ω мы перешли к комплексной переменной $s = j\omega$.

Если помехи нет, спектральная плотность ошибки вычисляется по передаточной функции от входа x к выходу ε , которая равна $C(s) - 1$:

$$S_{\varepsilon X}(s) = (C-1)(C^*-1) S_X.$$

Здесь и далее для сокращения записи у функций переменной s опущен аргумент, а звездочка (верхний индекс) обозначает замену s на $-s$. Если же нет полезного сигнала, спектральная плотность ошибки вычисляется по формуле

$$S_{\varepsilon N}(s) = C C^* S_N.$$

Когда действуют оба сигнала, но они статистически независимы, спектральная плотность ошибки равна сумме приведенных выше «отдельных» спектров:

$$S_{\varepsilon}(s) = S_{\varepsilon X} + S_{\varepsilon N} = (C-1)(C^*-1)S_X + CC^* S_N. \quad (12)$$

Можно сгруппировать слагаемые немного по-другому

$$S_{\varepsilon}(s) = CC^*(S_X + S_N) - CS_X - C^*S_X + S_X.$$

Выражение $S_{XN}(s) = S_X + S_N$ – это спектральная плотность смеси сигнала и шума. Поэтому можно построить соответствующий формирующий фильтр $\Psi(s)$, все нули и полюса которого находятся в левой полуплоскости:

$$S_X + S_N = \Psi(s)\Psi(-s). \quad (13)$$

Эта операция называется *спектральной факторизацией*. Тогда выражение для спектра ошибки можно записать в виде

$$S_{\varepsilon}(s) = CC^*\Psi\Psi^* - CS_X - C^*S_X + S_X = \left(C\Psi - \frac{S_X}{\Psi^*}\right)\left(C^*\Psi^* - \frac{S_X}{\Psi}\right) + S_0, \quad (14)$$

где $S_0(s) = S_X - \frac{S_X S_X}{\Psi\Psi^*} = S_X - \frac{S_X S_X}{S_X + S_N} = \frac{S_X S_N}{S_X + S_N}$. Заметим, что

- 1) $S_0(s)$ не зависит от выбора фильтра;
- 2) $S_0(s)$ можно рассматривать как некоторую спектральную плотность, поэтому она всегда неотрицательна на мнимой оси, где проводится интегрирование (при $s = j\omega$);
- 3) выражение $Z(s)Z(-s)$ тоже всегда неотрицательно на мнимой оси, поскольку при подстановке $s = j\omega$ представляет собой квадрат амплитудной частотной характеристики; в нашем случае $Z(s) = C\Psi - \frac{S_X}{\Psi^*}$.

Таким образом, спектральная плотность ошибки (14) – это сумма двух неотрицательных величин, одна из которых не зависит от выбора фильтра. Поэтому лучшее возможное решение – выбрать фильтр из условия $Z(s) = 0$, то есть

$$C(s) = \frac{S_X}{\Psi\Psi^*} = \frac{S_X}{S_X + S_N}. \quad (15)$$

Обычно полезный сигнал – низкочастотный, а помеха – высокочастотная. Их этой формулы следует, что на тех частотах, где $S_X(\omega) \gg S_N(\omega)$ (сигнал значительно мощнее помехи), усиление фильтра близко к единице (помеху можно не учитывать). В то же время частоты, где $S_N(\omega) \gg S_X(\omega)$ (помеха мощнее сигнала) подавляются, так как коэффициент усиления фильтра (15) стремится к нулю.

Заметим, что передаточная функция фильтра (15) содержит полюса как слева, так и справа от мнимой оси. Импульсная характеристика такого фильтра не будет равна нулю при $t < 0$, то есть, фильтр должен использовать *будущие* значения входного сигнала. Поэтому при обработке

в реальном времени его невозможно реализовать практически. Такой фильтр называют *оптимальным физически нереализуемым фильтром*. Он имеет «право на жизнь» только тогда, когда обрабатывается предварительно записанный сигнал, и для каждого момента времени известны все его прошлые и будущие значения.

В прикладных задачах чаще всего требуется построить оптимальный *физически реализуемый фильтр*, передаточная функция которого содержит полюса только в левой полуплоскости. При этом мы сужаем множество допустимых решений, поэтому такой фильтр заведомо будет работать не лучше, чем фильтр (15).

Доказано¹³, что оптимальным будет фильтр, при котором произведение $C\Psi$ равно устойчивой части $\frac{S_X}{\Psi^*}$, обозначаемой как $\left[\frac{S_X}{\Psi^*}\right]_+$. Это значит, что нужно выполнить *сепарацию* (расщепление) этой функции на два слагаемых, устойчивое (с полюсами только в левой полуплоскости) и неустойчивое (все полюса – в правой полуплоскости):

$$\frac{S_X}{\Psi^*} = \left[\frac{S_X}{\Psi^*}\right]_+ + \left[\frac{S_X}{\Psi^*}\right]_- . \quad (16)$$

Неустойчивая часть $\left[\frac{S_X}{\Psi^*}\right]_-$ должна быть *строго правильной* функцией (степень ее числителя меньше степени знаменателя). Условие $C\Psi = \left[\frac{S_X}{\Psi^*}\right]_+$ дает

$$C(s) = \frac{1}{\Psi} \cdot \left[\frac{S_X}{\Psi^*}\right]_+ . \quad (17)$$

Эта формула определяет оптимальный *физически реализуемый* фильтр Винера.

Для вычисления дисперсии ошибки нужно для полученной функции $C(s)$ найти спектральную плотность (12) и вычислить интеграл (11). В среде МАТЛАВ существует и другой способ: факторизовать полученную спектральную плотность ошибки

$$S_\varepsilon(s) = F_\varepsilon(s) F_\varepsilon(-s) ,$$

а затем вычислить (с помощью функции **norm**) *норму* устойчивой передаточной функции $F_\varepsilon(s)$. Эта норма представляет собой СКВО выхода системы при действии на ее вход единичного белого шума, то есть, $\sqrt{D_\varepsilon}$.

Пусть, например, спектральная плотность полезного сигнала имеет вид $S_X(\omega) = \frac{1}{\omega^2 + 1}$, а помеха – белый шум с интенсивностью 1, то есть, $S_N(\omega) = 1$. Переходя к переменной $s = j\omega$, находим оптимальный физически нереализуемый фильтр (15):

$$C(s) = \frac{S_X}{S_X + S_N} = \frac{1}{-s^2 + 1} \cdot \frac{-s^2 + 1}{-s^2 + 2} = \frac{1}{-s^2 + 2} .$$

Соответствующая ему дисперсия ошибки равна

$$D_0 = \frac{1}{\pi} \int_0^\infty \frac{S_X S_N}{S_X + S_N} d\omega = \frac{1}{\pi} \int_0^\infty \frac{1}{\omega^2 + 2} d\omega = 0,354 .$$

Теперь построим физически реализуемый фильтр. Выполняем факторизацию (13)

¹³ Это решение предложили Х. Бode и К. Шеннон.

$$S_X + S_N = \frac{-s^2 + 2}{-s^2 + 1} = \Psi\Psi^* \Rightarrow \Psi(s) = \frac{s + \sqrt{2}}{s + 1},$$

а затем сепарацию (16)

$$\frac{S_X}{\Psi^*} = \frac{1}{-s^2 + 1} \cdot \frac{-s + 1}{-s + \sqrt{2}} = \frac{1}{(s + 1)(-s + \sqrt{2})} = \frac{k}{s + 1} + \frac{k}{-s + \sqrt{2}},$$

где $k = \sqrt{2} - 1$. По формуле (17) находим оптимальный фильтр:

$$C(s) = \frac{1}{\Psi} \left[\frac{S_X}{\Psi^*} \right]_+ = \frac{s + 1}{s + \sqrt{2}} \cdot \frac{k}{s + 1} = \frac{k}{s + \sqrt{2}}.$$

Подставляя эту функцию в формулу для спектра ошибки (12), после сокращений получаем

$$S_\varepsilon(s) = \frac{2(2 - \sqrt{2})}{(s + \sqrt{2})(-s + \sqrt{2})}.$$

Вычисление дисперсии ошибки дает $D_\varepsilon = 0,414 > D_0$. Видим, что в данном случае проигрыш по сравнению с оптимальным физически нереализуемым фильтром небольшой. Это связано с тем, что мощность помехи значительна в сравнении с мощностью сигнала, поэтому эффективность того и другого фильтров невысока.

При очень больших помехах усиление оптимального фильтра будет уменьшаться (вплоть до нуля). Это значит, что сигнал выделить практически невозможно и фильтр просто «давит» помеху.

5.2.3. Функционал общего вида

В этом пункте мы рассмотрим задачу оптимизации в более общей форме для того, чтобы далее использовать полученные результаты для других аналогичных по сути задач.

В выражении (12) можно выделить неизвестную передаточную функцию $C(s)$ и «все остальное», записав его в общем виде:

$$X(s) = ACC^* - BC - B^*C^* + E, \quad (18)$$

где, $B(s)$ и $E(s)$ – некоторые функции, причем $A(s) = A(-s)$ и $E(s) = E(-s)$. Требуется найти устойчивую передаточную функцию $C(s)$, которая обеспечивает минимум функционала

$$I = \frac{1}{2\pi j} \int_{-j\infty}^{j\infty} X(s) ds. \quad (19)$$

Функция $A(s)$ представляет собой некоторый спектр, для которого можно выполнить спектральную факторизацию, аналогичную (13):

$$A(s) = \Psi\Psi^*, \quad (20)$$

где у функции $\Psi(s)$ все нули (корни числителя) и полюса (корни знаменателя) имеют отрицательные вещественные части. Тогда выражение для спектра ошибки можно представить в виде

$$X(s) = \Psi\Psi^*CC^* - BC - B^*C^* + E = ZZ^* + E_0,$$

где $Z(s) = C\Psi - \frac{B^*}{\Psi^*}$ и $E_0(s) = E - \frac{BB^*}{A}$. Заметим, что E_0 не зависит от выбора фильтра, а выражение ZZ^* неотрицательно на мнимой оси (квадрат амплитудной частотной характеристики). Поэтому лучшее, что можно сделать – обеспечить $Z(s) = 0$, что дает оптимальный физически нереализуемый фильтр

$$C(s) = \frac{B^*}{\Psi \Psi^*} = \frac{B^*}{A}.$$

Для построения оптимального *физически реализуемого* фильтра выполним *сепарацию*

$$\frac{B^*}{\Psi^*} = \left[\frac{B^*}{\Psi^*} \right]_+ + \left[\frac{B^*}{\Psi^*} \right]_- ,$$

где первое слагаемое соответствует всем устойчивым полюсам, а второе – строго правильная функция – всем неустойчивым. Оптимальный фильтр вычисляется по формуле

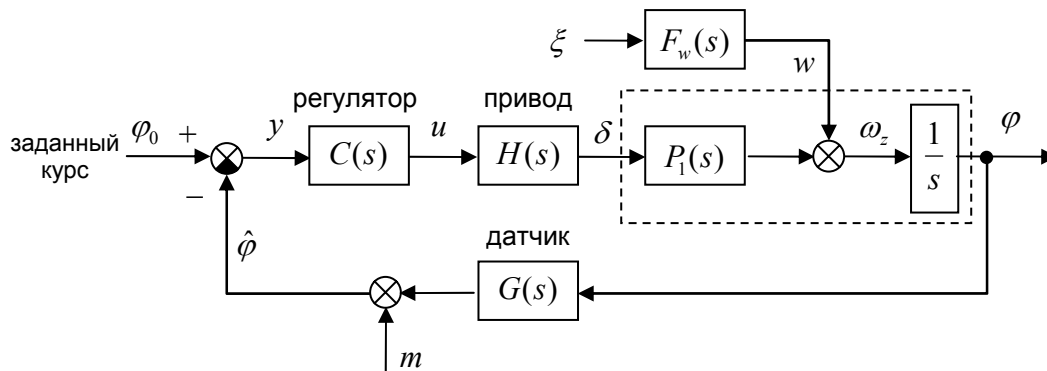
$$C(s) = \frac{1}{\Psi} \left[\frac{B^*}{\Psi^*} \right]_+ .$$

5.3. Оптимальное управление в замкнутых системах

Расчет оптимального регулятора для замкнутых систем (систем управления с обратной связью) выполняется почти так же, как и синтез оптимальных фильтров. Однако нужно учитывать две особенности:

- 1) регулятор должен обеспечивать устойчивость *замкнутой системы* (в задаче фильтрации обычно требуется устойчивость самого фильтра);
- 2) уменьшение дисперсии ошибки, как правило, достигается за счет увеличения мощности управления, которая ограничена в реальных задачах; например, угол перекладки руля судна не может быть больше 30-35°.

Рассмотрим систему стабилизации судна на курсе в условиях морского волнения.



Модель судна (обведена штриховой линией) состоит из двух блоков с передаточными функциями $P_1(s)$ и $P_2(s) = 1/s$ (интегрирующее звено). Истинный курс φ измеряется датчиком с передаточной функцией $G(s)$. Для компенсации *измеренной* ошибки $y = \varphi_0 - \hat{\varphi}$ регулятор $C(s)$ выдает на привод (с передаточной функцией $H(s)$) управляющий сигнал u , который преобразуется в угол поворота руля δ . Заметим, что сама ошибка не измеряется, вместо нее регулятор получает сигнал y , который можно назвать оценкой ошибки.

На систему действуют два случайных сигнала: возмущение w , вызванное морским волнением и имеющее спектральную плотность $S_w(\omega)$, и шум измерений m . Для моделирования возмущения используется формирующий фильтр $F_w(s)$ (такой, что $S_w(\omega) = F_w(s) F_w(-s)$), на вход которого поступает единичный белый шум ξ . Далее мы будем считать, что $\varphi_0 = 0$ и все случайные сигналы – центрированные.

Какова же цель управления? Очевидно, что нужно сделать минимальной дисперсию D_ε ошибки $\varepsilon = \varphi - \varphi_0 = \varphi$ (при $\varphi_0 = 0$). Это может быть обеспечено только с помощью управления, то есть, за счет увеличения активности руля. В то же время, постоянные «дергания» руля крайне нежелательны, поэтому нужно ограничить дисперсию D_u (мощность) сигнала управления u . Обычно в таких случаях для решения задачи оптимизации составляют единый критерий

$$I = D_\varepsilon + k_u^2 D_u \rightarrow \min. \quad (21)$$

где k_u^2 – неотрицательный весовой коэффициент. Задача состоит в том, чтобы выбрать передаточную функцию регулятора $C(s)$, обеспечивающую минимум критерия (21).

Величина k_u^2 позволяет ограничить сигнал управления. Если этого не делать, то есть принять $k_u^2 = 0$, мощность управления в оптимальной системе практически всегда будет неограниченно расти, что неприемлемо.

Вычисление дисперсий ошибки и управления выполним через их спектральные плотности. Чтобы упростить выводы, пока не будем учитывать шум измерений. Тогда

$$I = \frac{1}{2\pi j} \int_{-j\infty}^{j\infty} (W_\varphi W_\varphi^* + k_u^2 W_u W_u^*) ds,$$

где $W_\varphi(s)$ и $W_u(s)$ – передаточные функции от входа ξ (белого шума) к углу рыскания φ и сигналу управления u . Используя правила преобразования структурных схем, находим

$$W_\varphi(s) = P_2 F_w - \frac{C}{1 + CHPG} PHGP_2 F_w, \quad W_u(s) = -\frac{C}{1 + CHPG} GP_2 F_w.$$

где $P(s) = P_1(s) P_2(s)$. Здесь от выбора регулятора $C(s)$ зависит только функция

$$\tilde{C}(s) = \frac{C}{1 + CHPG}, \quad (22)$$

поэтому спектральные плотности угла рыскания и управления можно вычислить как

$$S_\varphi(s) = (1 - PHG\tilde{C})(1 - P^* H^* G^* \tilde{C}^*) P_2^* P_2^* F_w^* F_w^*, \quad S_u(s) = \tilde{C}\tilde{C}^* G G^* P_2^* P_2^* F_w^* F_w^*.$$

После простых преобразований, критерий качества (21) можно записать в виде (19), где

$$X(s) = A\tilde{C}\tilde{C}^* - B\tilde{C} - B^* \tilde{C}^* + E. \quad (23)$$

Это значит, что подынтегральное выражение имеет вид (18), где вместо $C(s)$ фигурирует $\tilde{C}(s)$, а функции $A(s)$, $B(s)$ и $E(s)$ известны (конечно, они отличаются от тех, что были в задаче фильтрации). При этом сразу возникает вопрос: нельзя ли использовать здесь тот же способ оптимизации, который применяется для построения фильтра Винера?

Несложно показать, что $\tilde{C}(s)$ – это в самом деле передаточная функция замкнутой системы от u к u , то есть, для устойчивости системы она должна быть устойчива, как и фильтр Винера. Тогда можно было бы сначала найти оптимальную устойчивую функцию $\tilde{C}(s)$ как оптимальный фильтр Винера, а затем выразить передаточную функцию регулятора $C(s)$ из (22). Однако такой метод «проходит» только тогда, когда сам объект устойчив. В общем случае (если объект может быть неустойчивым) приходится использовать различные «хитрости», чтобы свести задачу оптимизации замкнутой системы к задаче Винера и при этом гарантировать устойчивость системы. По существу, сначала строится вспомогательный регулятор, который стабилизирует объект, а затем используется методика синтеза оптимального фильтра Винера.

Современный подход к решению этой задачи основан на использовании *параметризации всех стабилизирующих регуляторов*. Так называется выражение для передаточной функции регулятора, которое зависит от неизвестной устойчивой функции. Изменяя эту функцию произвольным образом (но сохраняя ее устойчивость), мы можем получить любой регулятор, который стабилизирует систему. С другой стороны, в такой форме можно представить любой регулятор, который стабилизирует систему.

Для построения параметризации передаточную функцию *разомкнутого контура* (без регулятора) представляют в виде отношения полиномов, $n(s)$ и $d(s)$:

$$P(s) G(s) H(s) = \frac{n(s)}{d(s)}. \quad (24)$$

Если полиномы $n(s)$ и $d(s)$ не имеют общих корней, то существуют полиномы $a_0(s)$ и $b_0(s)$, удовлетворяющие уравнению

$$a_0(s) n(s) + b_0(s) d(s) = 1. \quad (25)$$

Это *полиномиальное* уравнение, то есть уравнение, в котором неизвестными являются полиномы (многочлены). Степени полиномов $a_0(s)$ и $b_0(s)$ на единицу меньше, чем степени $d(s)$ и $n(s)$ соответственно. Приравнивая коэффициенты при одинаковых степенях s в левой и правой частях уравнения, получаем линейную систему уравнений, которая легко решается современными программными средствами.

Параметризация всех стабилизирующих регуляторов имеет вид

$$C(s) = \frac{a_0 + d\Phi}{b_0 - n\Phi}, \quad (26)$$

где $\Phi(s)$ произвольная устойчивая дробно-рациональная функция (то есть, все корни ее знаменателя имеют отрицательные вещественные части). Любой регулятор, построенный по этой формуле – стабилизирующий, в то же время любой стабилизирующий регулятор может быть представлен в таком виде для некоторой функции $\Phi(s)$.

Подставляя выражение (26) для $C(s)$ в (22), находим, учитывая (24) и (25):

$$\tilde{C}(s) = d(a_0 + d\Phi).$$

В свою очередь, подставляя это выражение в (23), получаем

$$X(s) = \tilde{A}\Phi\Phi^* - \tilde{B}\Phi - \tilde{B}^*\Phi^* + \tilde{E},$$

где $\tilde{A}(s)$, $\tilde{B}(s)$ и $\tilde{E}(s)$ – известные функции. Поскольку любая устойчивая функция $\Phi(s)$ дает стабилизирующий регулятор (26), мы свели исходную задачу к задаче Винера, которую можно решать известными средствами (применяя факторизацию и сепарацию).

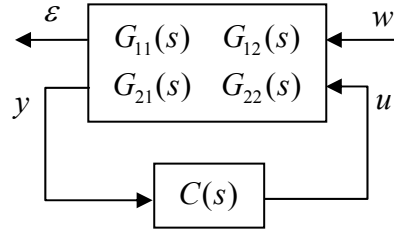
5.4. Стандартная система

Чтобы не «привязываться» к конкретной структурной схеме, программы для проектирования оптимальных регуляторов обычно используют так называемую стандартную систему, где выделяют 4 типа сигналов и связанных с ними передаточных функций:

- ε – сигнал ошибки, его нужно сделать «минимальным» в некотором смысле;
- y – измеренный сигнал обратной связи, поступающий на вход регулятора;
- w – внешнее возмущение;
- u – управляющий сигнал на выходе регулятора.

Уравнения системы (в изображениях по Лапласу) можно записать так:

$$\begin{aligned}\varepsilon &= G_{11}(s) w + G_{12}(s) u \\ y &= G_{21}(s) w + G_{22}(s) u\end{aligned}$$



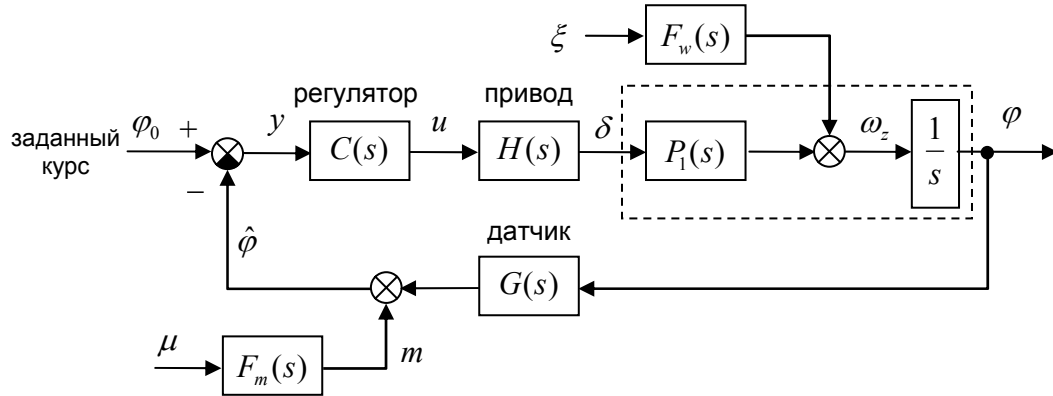
Справа показано общепринятое обозначение стандартной системы. Обратите внимание, что в ней (формально) используется положительная обратная связь. При построении передаточных функций $G_{ij}(s)$ ($i, j = 1, 2$) предполагается, что регулятора в контуре нет (цепь разорвана).

Каждый из четырех сигналов может быть векторным, то есть, содержать несколько компонент. Например, в рассмотренной выше задаче стабилизации судна на курсе требовалось ограничить сумму дисперсий $D_\varphi + k_u^2 D_u$. Ее можно представить как дисперсию векторного сигнала $\varepsilon = \begin{bmatrix} \varphi \\ k_u u \end{bmatrix}$, которая по определению равна:

$$D_\varepsilon = E\{\varepsilon^T \varepsilon\} = E\left\{\begin{bmatrix} \varphi & k_u u \end{bmatrix} \begin{bmatrix} \varphi \\ k_u u \end{bmatrix}\right\} = E\{\varphi^2 + k_u^2 u^2\} = D_\varphi + k_u^2 D_u.$$

Здесь $E\{\}$ обозначает математическое ожидание выражения в фигурных скобках, верхний индекс T – операцию транспонирования. Также учитывается, что сигналы центрированные, поэтому математическое ожидание квадрата сигнала равно его дисперсии.

Построим стандартную систему для задачи оптимальной стабилизации судна на курсе, которая рассматривалась выше. Здесь добавлен фильтр $F_m(s)$, который формирует спектр помехи измерений. Его входной сигнал μ описывается как единичный белый шум, независимый от ξ .



Вектор внешних воздействий в задаче оптимизации при случайных возмущениях должен описываться моделью единичного белого шума. В данном случае он состоит из двух компонент, независимых сигналов ξ и μ . Уравнения системы выглядят так:

$$\begin{aligned}\varepsilon_1 = \varphi &= P_2 F_w \xi + P H u \\ \varepsilon_2 &= k_u u \\ y &= -G P_2 F_w \xi - F_m \mu - G P H u\end{aligned}$$

Поэтому

$$G_{11}(s) = \begin{bmatrix} P_2 F_w & 0 \\ 0 & 0 \end{bmatrix}, \quad G_{12}(s) = \begin{bmatrix} P H \\ k_u \end{bmatrix}, \quad G_{21}(s) = \begin{bmatrix} -G P_2 F_w & -F_m \end{bmatrix}, \quad G_{22}(s) = -G P H.$$

Обратите внимание, что отрицательная обратная связь здесь выражается в том, что в записи $G_{22}(s)$ появляется знак «минус». Можно показать, что $G_{22}(s)$ – это передаточная функция разомкнутого контура (без регулятора).

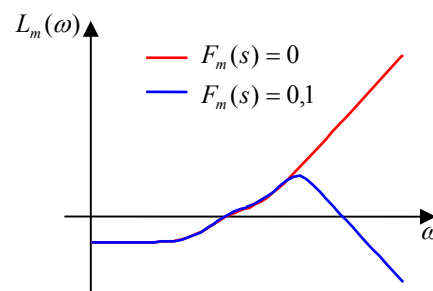
5.5. Особенности задачи оптимизации

5.5.1. Идеальные датчики

Что будет, если в задаче оптимизации предположить, что датчик идеально измеряет сигнал (то есть, принять $F_m(s) = 0$)? Если выполнить синтез, передаточная функция оптимального регулятора может получиться, например, такая:

$$C(s) = \frac{0,5s^2 + s + 0,1}{s + 0,4}.$$

В ней степень числителя выше, чем степень знаменателя. Разделив числитель на знаменатель, можно выделить дифференцирующее звено с передаточной функцией $0,5s$. Вроде бы здесь нет ничего плохого. Однако при построении частотной характеристики такого регулятора (см. ЛАФЧХ на рисунке справа, красная линия) становится видно, что он усиливает высокочастотные сигналы. Во-первых, это приводит к тому, что сигнал управления будет очень большим и в реальной системе будет превышено его максимально допустимое значение. Во-вторых, любой высокочастотный шум измерения (который всегда присутствует) будет усиливаться регулятором, который начнет «раскачивать» систему. Таким образом, с точки зрения практического использования получилось плохое решение, хотя и «оптимальное».



Чтобы избежать подобных проблем и сделать регулятор нечувствительным к высокочастотным помехам измерительной системы, в задаче оптимизации учитывают сигнал помехи типа «белого шума». Если принять $F_m(s) = 0,1$ при тех же условиях мы получаем регулятор

$$C(s) = \frac{50s^2 + 100s + 10}{s^3 + 14s^2 + 100s + 40}.$$

Поскольку эта функция *строго правильная* (степень ее числителя меньше степени знаменателя), амплитудная частотная характеристика (синяя линия на рисунке) «скатывается» вниз на высоких частотах. Поэтому регулятор будет нечувствителен к помехам измерений. Для обеспечения «ската» частотной характеристики регулятора на высоких частотах (англ. *roll-off*), передаточная функция фильтра $F_m(s)$ должна иметь *равные* степени числителя и знаменателя, то есть, модель помехи должна содержать составляющую типа «белого шума». Если в системе несколько измеряемых сигналов, это относится к каждому из них.

5.5.2. Фиксированные полюса

Анализ оптимальных систем показывает, что все устойчивые полюса $H(s)$, $G(s)$ и $P_1(s)$ автоматически становятся корнями характеристического уравнения оптимальной замкнутой системы. Это – так называемые *фиксированные полюса*, они определяются особенностями структуры системы. Если эти передаточные функции имеют неустойчивые полюса (с положительной вещественной частью), корнями характеристического уравнения становятся их (устой-

чивые) *отражения* от мнимой оси. Если же в $H(s)$, $G(s)$ или $P_1(s)$ есть интегрирующее звено, задача не имеет практически ценного решения – характеристическое уравнение «оптимальной» системы будет иметь нулевой корень, система находится на границе области устойчивости.

Полюса функций $P_2(s)$, $F_w(s)$ и $F_m(s)$ в явном виде не входят в характеристическое уравнение. Это объясняется (на интуитивном уровне) тем, что блоки с передаточными функциями находятся $F_w(s)$ и $F_m(s)$ вне контура управления. Особенность блока $P_2(s)$ состоит в том, что входной сигнал ξ проходит *через него* на выход φ . Путем достаточно сложных выкладок можно доказать, что при этом его полюса не будут корнями характеристического уравнения. Из этого правила есть одно исключение: формирующий фильтр возмущения не должен содержать дифференцирующее звено, то есть передаточная функция $F_w(s)$ не должна иметь сомножителя s в числителе.

Вспомним, что для моделирования воздействия морского волнения чаще всего используют формирующий фильтр с передаточной функцией

$$F_w(s) = \frac{K_w s}{s^2 + 2\lambda\omega_0 s + \omega_0^2},$$

где ω_0 – доминирующая частота волн, λ – коэффициент демпфирования ($0 < \lambda < 1$), $K_w = 2\lambda\omega_0\sigma_w$ и σ_w – коэффициент, определяющий интенсивность волнения. При этом

$$F_w(s) P_2(s) = \frac{K_w s}{s^2 + 2\lambda\omega_0 s + \omega_0^2} \cdot \frac{1}{s} = \frac{K_w}{s^2 + 2\lambda\omega_0 s + \omega_0^2},$$

то есть, полюс $P_2(s)$ в точке $s = 0$ *сократился* нулем передаточной функции $F_w(s)$. Это значит, что входной сигнал ξ уже *не действует* на интегратор, и полюс в точке $s = 0$ в модели судна должен быть (с формально-математической точки зрения) отнесен к блоку $P_1(s)$. Так как все полюса $P_1(s)$ (в исходном или «отраженном» виде) становятся корнями характеристического уравнения, в этом случае оптимального стабилизирующего регулятора не существует.

5.5.3. Регуляризация

Что же делать? Нужно построить «квазиоптимальный» (лат. *как бы оптимальный, похожий на оптимальный*) регулятор, который будет стабилизировать систему, хотя и будет несколько хуже, чем «оптимальный» (но нестабилизирующий!) по выбранному критерию качества. Для этого обычно выполняют *регуляризацию* задачи, то есть, немного меняют условие так, чтобы она стала решаемой. В данном случае возможно два приема:

- 1) вместо $F_w(s)$ использовать измененную передаточную функцию фильтра

$$\tilde{F}_w(s) = \frac{K_w s + \theta}{s^2 + 2\lambda\omega_0 s + \omega_0^2},$$

которая не имеет нуля в точке $s = 0$ при любом малом θ ;

- 2) воспользоваться другой моделью спектра морского волнения, которая дает ненулевое значение на частоте $\omega = 0$, например, дробно-рациональными спектрами типов 1 и 2, которым соответствуют формирующие фильтры

$$\tilde{F}_w(s) = \frac{\gamma_0}{s^2 + \delta_1 s + \delta_0} \quad \text{и} \quad \tilde{F}_w(s) = \frac{\gamma_1 s + \gamma_0}{s^2 + \delta_1 s + \delta_0}.$$

5.5.4. Все не может быть оптимальным

Отметим еще одну особенность рассматриваемой задачи. Мы стремились к тому, чтобы система была устойчива и обеспечивался минимум функции потерь $I = D_\varepsilon + k_u^2 D_u$. Одновременно все остальные требования к системе, например, скорость и качество переходных процессов, не учитывались. Поэтому «оптимальный» регулятор может приводить к затянутому или колебательному переходному процессу.

5.6. Кривая качества

Важно понимать, что в любой задаче есть принципиальные (*фундаментальные*) ограничения, которые определяются особенностями структуры системы и не могут быть преодолены никаким регулятором. Например, ясно, что задачу «обеспечить рыскание судна не более 1° при волнении 8 баллов» в реальной ситуации решить нельзя.

Пусть известны все характеристики системы и возмущений. За счет чего можно уменьшить ошибку стабилизации? Как правило, только за счет увеличения мощности управления, которая ограничена¹⁴. Поэтому для уменьшения ошибки нужно увеличивать управляющий сигнал.

С другой стороны, активность руля тоже нужно всячески уменьшать, потому что механические части быстро изнашиваются и приходят в негодность. Кроме того, на управление затрачивается дополнительная энергия.

Таким образом, мы пришли к задаче *многоцелевой оптимизации* – хочется одновременно обеспечить минимум ошибки и минимум мощности управления. Как мы видели, эти две цели противоречивы. В таком случае чаще всего составляется единый критерий качества, который включает все величины, которые нужно минимизировать, с различными весовыми коэффициентами. Например, в задаче стабилизации судна при случайных возмущениях он выглядит так:

$$I = D_\varepsilon + k_u^2 D_u. \quad (27)$$

Здесь D_ε и D_u – дисперсии ошибки и сигнала управления, а k_u^2 – неотрицательный весовой коэффициент.

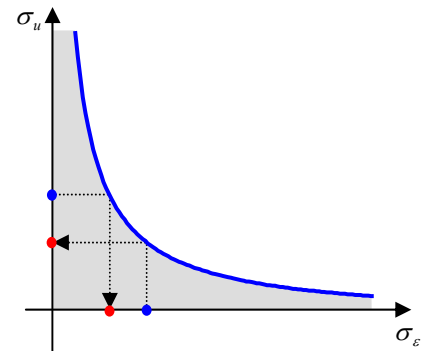
Предположим, что мы нашли оптимальный регулятор $C_{opt}(s)$, который минимизирует этот критерий при некотором фиксированном коэффициенте k_u^2 . Можно ли выбором какого-то *другого* регулятора $\hat{C}(s)$ *одновременно* уменьшить и дисперсию ошибки D_ε , и дисперсию управления D_u ? Если предположить, что можно, получается, что регулятор $\hat{C}(s)$ дает меньшее значение критерия качества (при том же k_u^2), чем $C_{opt}(s)$, то есть, $C_{opt}(s)$ – это *не* оптимальный регулятор и мы пришли к противоречию. Таким образом, для полученной оптимальной системы *нельзя* одновременно уменьшить и дисперсию ошибки D_ε , и дисперсию управления D_u . Такие регуляторы называются *Парето-оптимальными*.

Для каждого коэффициента k_u^2 будут свои значения D_ε и D_u , так что можно построить график зависимости D_u от D_ε в оптимальных системах. Отметим, что чаще всего удобнее вме-

¹⁴ Например, угол перекаладки руля судна чаще всего не может быть более $30\text{--}35^\circ$ (при больших углах руль становится неэффективен).

сто дисперсий использовать соответствующие среднеквадратические отклонения – $\sigma_\varepsilon = \sqrt{D_\varepsilon}$ и $\sigma_u = \sqrt{D_u}$, которые измеряются в тех же единицах, что и исходные величины. Этот график показывает, какая мощность управления требуется, чтобы обеспечить заданную точность стабилизации. Наоборот, по графику можно определить, какую точность можно обеспечить, имея заданную мощность управления. В этом смысле можно называть эту кривую *кривой качества* системы.

Каждая точка этой кривой соответствует какому-то Парето-оптимальному регулятору. Поскольку для таких регуляторов нельзя одновременно уменьшить оба показателя качества, выпуклость кривой всегда направлена в сторону начала координат. Серая область недостижима, то есть, ни один регулятор в такой системе не может обеспечить соответствующее качество.

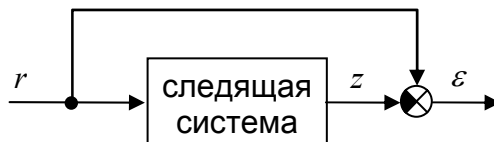


6. Оптимальные следящие системы

6.1. Постановка задачи

Задача следящей системы – отслеживать на выходе z сигнал r , подаваемый на вход. Например, систему автоматического управления курсом корабля (автопилот) можно рассматривать как следящую систему (r – заданный курс, z – фактический курс).

Точность следящей системы определяется свойствами сигнала ошибки $\varepsilon = r - z$:

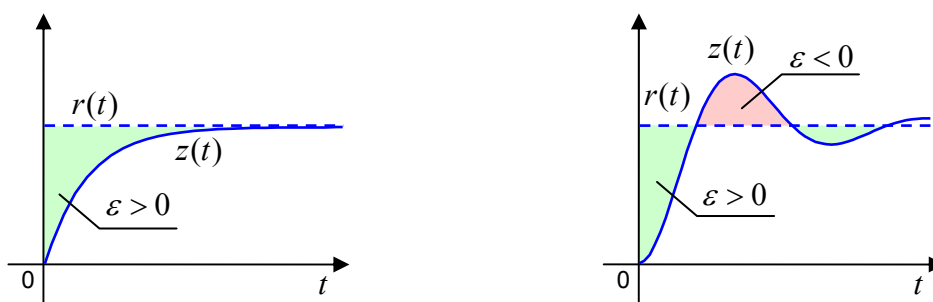


Если эталонный (задающий) сигнал r – случайный процесс с известной спектральной плотностью, мы получаем задачу оптимизации при случайных возмущениях, варианты которой были рассмотрены ранее.

Здесь мы остановимся на задаче оптимизации при *детерминированных* (известных, определенных, неслучайных) возмущениях. Это означает, что мы знаем входной сигнал r (например, его изображение по Лапласу $R(s)$). При этом требуется обеспечить «малость» ошибки ε в некотором смысле. В идеальном случае ошибка равна нулю для любого момента времени. В реальных системах этот результат чаще всего недостижим, поскольку требует бесконечно большого управления.

Предположим, что входной сигнал имеет ступенчатый вид, причем можно считать, что его изменение происходит достаточно редко, так что при очередном скачке переходный процесс, вызванный предыдущим изменением, уже закончился. В этом случае имеет смысл строить систему, оптимальную для единичного скачка на входе. Так как система линейная, при любом изменении величины скачка она останется оптимальной (изменится только величина сигналов).

В идеале мы хотим, чтобы изменение входного сигнала *мгновенно* привело к такому же изменению на выходе. Можно догадаться, что для мгновенного перевода *инерционной* системы (а не просто усилителя) в новое состояние требуется бесконечное управление. Этот вариант неприемлем с практической точки зрения и не реализуем, поскольку управляющий сигнал всегда ограничен. Таким образом, реальный переходный процесс будет отличаться от идеального. Как же измерить эту ошибку, оценив ее одним числом?



Казалось бы, можно взять интеграл от сигнала ошибки $\varepsilon = r - z$ на интервале от 0 до бесконечности¹⁵

$$I_1 = \int_0^{\infty} \varepsilon(t) dt ,$$

¹⁵ Чтобы такой интеграл сходил, необходимо, чтобы ошибка $\varepsilon(t)$ стремилась к нулю при $t \rightarrow \infty$.

однако он может служить для оценки ошибки только при монотонном переходном процессе, когда ошибка всегда остается положительной (см. рисунок слева). Если процесс колебательный, на разных интервалах ошибка может принимать как положительные, так и отрицательные значения, поэтому использовать этот интеграл для оценки ошибки нельзя (см. рисунок справа).

Мы можем справиться с этой проблемой, если интегрировать *модуль* ошибки:

$$I_2 = \int_0^{\infty} |\varepsilon(t)| dt.$$

Такую оценку иногда используют при численной оптимизации. К сожалению, получить оптимальную передаточную функцию регулятора аналитически (по формулам) в этом случае не удастся.

Удобнее всего минимизировать интеграл от *квадрата* ошибки (его также называют *интегральной квадратической ошибкой*):

$$I = \int_0^{\infty} \varepsilon^2(t) dt. \quad (28)$$

Далее мы увидим, что задачу оптимизации по такому критерию удастся свести к задаче фильтрации Винера.

6.2. Теорема Парсеваля

Для большинства задач в теории управления существует два типа решений – *временное* (когда рассматривается изменение сигналов во времени) и *частотный* (работа с передаточными функциями и частотными характеристиками).

Эти подходы не исключают, а взаимно дополняют друг друга, позволяя увидеть разные стороны одной задачи. Для построения оптимального фильтра Винера мы использовали операции с передаточными функциями (частотный метод Боде и Шеннона), хотя сам Винер впервые предложил решение этой задачи с помощью временного метода (на основе корреляционных функций).

Для того, чтобы использовать уже рассмотренные алгоритмы, нужно «перевести» задачу с критерием (28) в частотную область, то есть, выразить критерий через изображения сигналов по Лапласу и передаточные функции. Это позволяет сделать *теорема Парсеваля*, которая утверждает, что интеграл от квадрата функции $\varepsilon(t)$, которая равна нулю при $t < 0$ и стремится к нулю при $t \rightarrow \infty$, равен интегралу от «квадрата» ее преобразования Лапласа $\mathcal{E}(s)$:

$$I = \int_0^{\infty} \varepsilon^2(t) dt = \frac{1}{2\pi j} \int_{-j\infty}^{j\infty} \mathcal{E}(s) \mathcal{E}(-s) ds. \quad (29)$$

Выражение $\mathcal{E}(s) \mathcal{E}(-s)$ можно назвать «квадратом» потому что оно на мнимой оси (где берется интеграл), при подстановке $s = j\omega$, действительно является квадратом частотной характеристики $\mathcal{E}(j\omega)$:

$$\mathcal{E}(j\omega) \mathcal{E}(-j\omega) = |\mathcal{E}(j\omega)|^2.$$

Заметим, что функционал (29) в нашей задаче (при известном входном сигнале) совпадает по форме (при $X(s) = \mathcal{E}(s) \mathcal{E}(-s)$) с функционалом (19), который получен в задаче фильтрации при случайных возмущениях. Поэтому для решения задачи можно использовать алгоритм, применявшийся при расчете фильтра Винера.

6.3. Эквивалентность двух задач

Теорема Парсеваля позволяет обнаружить связь между задачами анализа и синтеза при случайных (или *стохастических*) и детерминированных возмущениях.

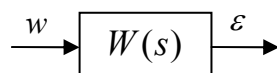
Поскольку ошибка $\varepsilon(t)$ стремится к нулю при $t \rightarrow \infty$, ее преобразование Лапласа $\mathcal{E}(s)$ – устойчивая функция (все ее полюса имеют отрицательные вещественные части). Если рассмат-

ривать $\mathcal{E}(s)$ как передаточную функцию формирующего фильтра, то можно построить соответствующую спектральную плотность $S_{\varepsilon}(\omega) = \mathcal{E}(j\omega) \mathcal{E}(-j\omega)$, так что

$$\int_0^{\infty} \varepsilon^2(t) dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} S_{\varepsilon}(\omega) d\omega = \frac{1}{\pi} \int_0^{\infty} S_{\varepsilon}(\omega) d\omega.$$

Таким образом, интеграл от квадрата функции $\varepsilon(t)$, стремящейся к нулю при $t \rightarrow \infty$, равен среднему квадрату случайного процесса, имеющего спектральную плотность $S_{\varepsilon}(\omega)$.

Пусть существует некоторая система с передаточной функцией $W(s)$. Обозначим ее входной сигнал через $w(t)$, а выходной – через $\varepsilon(t)$.



Если w – единичный центрированный белый шум, то спектральная плотность выхода ε равна $S_{\varepsilon}(s) = W(s) W(-s)$, а дисперсия выхода (средний квадрат) – интегралу от спектральной плотности по мнимой оси.

В то же время, если w – единичный импульс (дельта-функция), изображение выхода по Лапласу равно $\mathcal{E}(s) = W(s)$, а интегральная квадратическая ошибка равна тому же самому интегралу от $W(s) W(-s)$. Таким образом, вместо вычисления дисперсии выхода при белом шуме на входе можно вычислить интеграл от квадрата выходного сигнала при импульсном входе, и наоборот. Квадратный корень из этой величины называется H_2 -нормой передаточной функции $W(s)$:

$$\|W(s)\|_2 = \sqrt{\frac{1}{2\pi j} \int_{-j\infty}^{j\infty} W(s) W(-s) ds}$$

и вычисляется с помощью функции **norm** среды MATLAB.

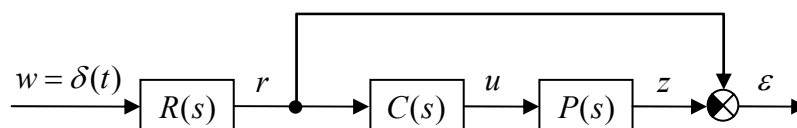
С другой стороны, пусть передаточная функция $W(s)$ зависит определенным образом от выбора регулятора $C(s)$. Тогда, как следует из сказанного, две следующие задачи оптимизации эквивалентны:

- 1) найти регулятор $C(s)$, минимизирующий дисперсию ошибки при единичном белом шуме на входе;
- 2) найти регулятор $C(s)$, минимизирующий интегральную квадратическую ошибку при поступлении на вход единичного импульса (дельта-функции).

Вторая задача в теории управления называется *задачей H_2 -оптимизации* или просто H_2 -задачей (поскольку требуется обеспечить минимум H_2 -нормы передаточной функции замкнутой системы), а о первой говорят как о стохастическом варианте H_2 -задачи.

6.4. Разомкнутые системы

Чтобы понять особенности задачи, сначала мы рассмотрим простейшую разомкнутую систему, состоящую только из регулятора $C(s)$ и объекта $F(s)$:



Входной сигнал r задан в виде преобразования Лапласа $R(s)$, так что его можно представить как результат прохождения единичного импульса (дельта-функции $\delta(t)$) через звено с переда-

точной функцией $R(s)$. Такой способ общепринят при моделировании входных сигналов и позволяет представить систему в стандартном виде: на входе – дельта-функция, на выходе – ошибка.

Передаточная функция системы от входа w к выходу ε равна

$$W(s) = (1 - CP)R.$$

Предполагается, что все звенья устойчивы. Это значит, что все полюса функций $P(s)$ и $R(s)$ имеют отрицательные вещественные части и регулятор $C(s)$, который требуется найти, также должен обладать этим свойством. Кроме того, регулятор должен обеспечить минимум интегральной квадратической ошибки

$$I = \int_0^{\infty} \varepsilon^2(t) dt = \frac{1}{2\pi j} \int_{-j\infty}^{j\infty} X(s) ds,$$

где $X(s) = W W^* = PP^*RR^*CC^* - PRR^*C - P^*RR^*C^* + RR^*$. Подынтегральное выражение имеет форму (18), как для задачи Винера, поэтому можно применить уже известный алгоритм поиска оптимальной устойчивой передаточной функции $C(s)$.

Попытаемся понять, какие результаты мы можем получить в результате оптимизации. Прежде всего, выбор регулятора из условия $CP = 1$ сразу дает $W(s) = 0$, то есть, ошибка будет нулевой не только при ступенчатом, но и при *любом* другом входе. Это так называемое *условие инвариантности*, при котором обеспечивается идеальное слежение за эталонным сигналом. Именно такой регулятор будет получен в результате оптимизации «в лоб», если все нули передаточной функции $P(s)$ находятся в левой полуплоскости.

К сожалению, все не так просто. Как правило, передаточные функции реальных объектов – строго правильные, то есть, степень их числителя меньше, чем степень знаменателя. Тогда передаточная функция регулятора $C(s) = 1/P(s)$ будет *неправильной* (степень числителя *больше* степени знаменателя). Как было показано при анализе задачи оптимизации при случайных возмущениях, такой регулятор неприменим в практических задачах по двум причинам:

- 1) регулятор содержит дифференцирующее звено, поэтому при скачкообразном изменении входного сигнала сигнал управления должен теоретически стать бесконечным;
- 2) регулятор усиливает высокочастотные помехи измерений вместо того, чтобы подавлять их; это делает систему неработоспособной.

В задаче оптимизации при случайных возмущениях мы добивались «ската» частотной характеристики регулятора на высоких частотах с помощью ограничения на сигнал управления: в критерий качества вводилась дисперсия управления с некоторым весовым коэффициентом, который подстраивался методом проб и ошибок.

Если формально составить критерий $I = \int_0^{\infty} [\varepsilon^2(t) + k_u^2 u^2(t)] dt$, добавив интеграл от квадрата

сигнала управления $u(t)$ с весовым коэффициентом k_u^2 , то ничего хорошего не получится, поскольку установившееся значение $u(t)$ при $t \rightarrow \infty$ не равно нулю. Следовательно, интеграл расходится и теорема Парсеваля неприменима. Оптимизация в частотной области не имеет смысла: если формально применить алгоритм синтеза, получится регулятор, для которого ошибка в установившемся режиме (при $t \rightarrow \infty$) не равна нулю.

Чтобы грамотно ограничить управляющий сигнал, нужно ввести *ошибку управления* $\varepsilon_u(t)$, то есть отклонение фактического сигнала управления $u(t)$ от некоторого эталонного сигнала $u_0(t)$, который должен быть выбран так, чтобы ошибка управления $\varepsilon_u(t) = u_0(t) - u(t)$ стремилась к нулю при $t \rightarrow \infty$. Тогда можно искать минимум критерия

$$I = \int_0^{\infty} [\varepsilon^2(t) + k_u^2 \varepsilon_u^2(t)] dt,$$

это позволит ограничить мощность управления и при этом минимизировать ошибку слежения. Изменяя коэффициент k_u , мы получим семейство Парето-оптимальных регуляторов, как и в задаче оптимизации при случайных возмущениях.

Как же выбрать сигнал $u_0(t)$? Для простоты рассмотрим только случай ступенчатого входного сигнала. Поскольку передаточная функция $C(s)$ устойчива, установившееся значение u_∞ сигнала $u(t)$ при $t \rightarrow \infty$ – постоянная величина, ее можно рассчитать следующим образом.

Если вход w – это единичный импульс, то установившееся значение эталонного сигнала $r(t)$ можно определить по теореме о конечном значении для преобразования Лапласа

$$r_\infty = \lim_{t \rightarrow \infty} r(t) = \lim_{s \rightarrow 0} s R(s).$$

Например, если $r(t)$ – единичный ступенчатый сигнал, для которого $R(s) = 1/s$, то $r_\infty = 1$. С другой стороны, для того, чтобы установившаяся ошибка была равна нулю, необходимо, чтобы предельное значение сигнала $z(t)$ тоже было равно r_∞ . Учитывая, что $P(s)$ – устойчивая передаточная функция, имеем

$$r_\infty = k_p \cdot u_\infty,$$

где u_∞ – нужное нам установившееся значение сигнала управления, а k_p – статический коэффициент усиления объекта, который вычисляется по формуле $k_p = \lim_{s \rightarrow 0} P(s)$ (в среде MATLAB это делает функция **dcgain**). Таким образом, в качестве эталонного сигнала управления можно выбрать любой сигнал $u_0(t)$, у которого предельное значение при $t \rightarrow \infty$ равно $u_\infty = \frac{r_\infty}{k_p}$, на-

пример, ступенчатый сигнал $u_0(t) = u_\infty (t > 0)$, изображение которого равно $U_0(s) = \frac{u_\infty}{s}$.

Теперь построим стандартную систему, соответствующую задаче оптимизации. Уравнения в изображениях имеют вид:

$$\varepsilon = R w - P u$$

$$k_u \varepsilon_u = k_u U_0 w - k_u u$$

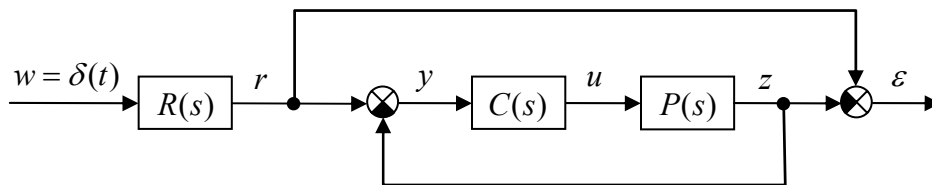
$$y = r = R w$$

Учитывая, что первые два уравнения определяют ошибки по выходу и по управлению, имеем

$$G_{11}(s) = \begin{bmatrix} R \\ k_u U_0 \end{bmatrix}, \quad G_{12}(s) = \begin{bmatrix} -P \\ -k_u \end{bmatrix}, \quad G_{21}(s) = R \quad \text{и} \quad G_{22}(s) = 0.$$

6.5. Замкнутые системы

Теперь рассмотрим аналогичную задачу для замкнутой системы.



Что изменилось с появлением замкнутого контура? Во-первых, уже не требуется, чтобы передаточные функции регулятора $C(s)$ и объекта $P(s)$ были устойчивыми. Во-вторых, требуется обеспечить устойчивость *замкнутой системы*. В-третьих, поскольку $P(s)$ может содержать множитель s в знаменателе (соответствующий интегрирующему звену), статический коэффициент усиления $k_p = \lim_{s \rightarrow 0} P(s)$ может оказаться равным бесконечности. Поэтому нужно скорректировать процедуру выбора эталонного управляющего сигнала.

Передаточная функция от входа w к ошибке ε равна

$$W(s) = \left[1 - \frac{C}{1 + CP} P \right] R = (1 - \tilde{C}P)R, \text{ где } \tilde{C}(s) = \frac{C}{1 + CP},$$

поэтому подынтегральное выражение в критерии качества будет зависеть от $\tilde{C}(s)$, а не от $C(s)$. Чтобы обеспечить устойчивость получаемой оптимальной системы в общем случае нужно использовать параметризацию стабилизирующих регуляторов (см. аналогичную задачу при случайных возмущениях).

При определении эталонного сигнала управления вроде бы все осталось по-прежнему: по теореме о предельном значении

$$r_\infty = z_\infty = \lim_{s \rightarrow 0} s P(s) U_0(s) \Rightarrow r_\infty = k_p \cdot u_\infty \Rightarrow u_\infty = \frac{r_\infty}{k_p}.$$

Однако, проблема в том, что объект может содержать интегрирующие звенья, поэтому статический коэффициент усиления $k_p = \lim_{s \rightarrow 0} P(s)$ обращается в бесконечность. В этом случае эталонный сигнал управления должен стремиться к нулю при $t \rightarrow \infty$. Например, можно принять $u_0(t) = 0$ при всех t .

Построим стандартную систему в задаче оптимизации по критерию

$$I = \int_0^\infty [\varepsilon^2(t) + k_u^2 \varepsilon_u^2(t)] dt \rightarrow \min.$$

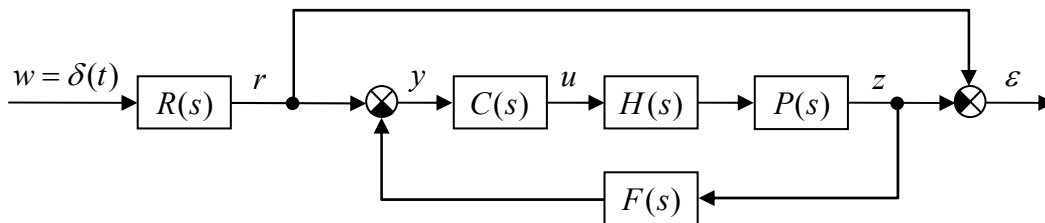
Уравнения системы имеют вид

$$\begin{aligned} \varepsilon &= R w - P u \\ k_u \varepsilon_u &= k_u U_0 w - k_u u \\ y &= R w - P u \end{aligned}$$

Учитывая, что первые два уравнения определяют ошибки по выходу и по управлению, имеем $G_{11}(s) = \begin{bmatrix} R \\ k_u U_0 \end{bmatrix}$, $G_{12}(s) = \begin{bmatrix} -P \\ -k_u \end{bmatrix}$, $G_{21}(s) = R$ и $G_{22}(s) = -P$. Как видим, все отличие от разомкнутой системы состоит в функции $G_{22}(s)$, которая в данном случае не равна нулю и представляет собой передающую функцию контура (без регулятора).

Можно показать, что устойчивые полюса передаточной функции объекта $P(s)$ (и неустойчивые полюса, «отраженные» от мнимой оси) становятся корнями характеристического уравнения оптимальной замкнутой системы. Вроде бы получается, что для объекта, включающего интегрирующее звено, задача не имеет решения. Однако из этого правила есть исключение: если модель входного сигнала $R(s)$ также содержит интегрирующее звено (например, для единичного ступенчатого сигнала $R(s) = 1/s$), в ходе синтеза происходит сокращение двух множителей и оптимальная система оказывается устойчивой.

Если учитывать динамику привода и датчиков, схема немного усложняется:



В этом случае стандартная система описывается матрицами

$$G_{11}(s) = \begin{bmatrix} R \\ k_u U_0 \end{bmatrix}, \quad G_{12}(s) = \begin{bmatrix} -PH \\ -k_u \end{bmatrix}, \quad G_{21}(s) = R \text{ и } G_{22}(s) = -FPH.$$

Заключение

Шаг за шагом, мы кратко рассмотрели основные понятия теории случайных процессов, а также принципы проектирования оптимальных линейных систем при случайных и детерминированных возмущениях.

Нужно понимать, что вы прочитали не учебник, а небольшое введение, призванное *познакомить* с основными идеями и дать общее представление о рассматриваемых вопросах. Тот, кто серьезно собирается изучать современные методы теории управления и использовать их в своей работе, должен продолжить изучение, взяв «нормальные» учебники (см. список литературы), в которых материал изложен значительно более строго и научно.

За рамками пособия остались многие родственные темы, с которыми должен быть знаком современный специалист по автоматическому управлению. Достаточно сказать, что мы рассмотрели только линейные непрерывные системы, тогда как практически все реальные системы содержат нелинейности и управляются цифровыми регуляторами, то есть являются непрерывно-дискретными. Ничего не было сказано о современных методах исследования многомерных систем (со многими входами и выходами) на основе моделей в пространстве состояний.

Автор будет считать свою задачу выполненной, если читатель почувствует в себе силы не остановиться на достигнутом и продолжить самообразование.

Литература для последующего чтения

(в порядке увеличения количества страниц)

1. *Купер Дж., Макгиллем К.* Вероятностные методы анализа сигналов и систем. М.: Мир, 1989.
2. *Вентцель Е.С., Овчаров Л.А.* Теория случайных процессов и ее инженерные приложения. М.: Наука, 1991.
3. *Мирошник И.В.* Теория автоматического управления. Нелинейные и оптимальные системы – СПб.: Питер, 2003.
4. *Бендат Дж., Пирсол А.* Прикладной анализ случайных данных – М.: Мир, 1989.
5. *Первозванский А.А.* Курс теории автоматического управления – М.: Наука, 1986.
6. *Квакернак Х., Сиван Р.* Линейные оптимальные системы управления / Пер с англ. – М.: Мир, 1977.
7. *Гудвин Г.К., Гребе С.Ф., Сальгадо М.Э.* Проектирование систем управления. М.: Бином, Лаборатория базовых знаний, 2004.