

ТЕОРИЯ АВТОМАТИЧЕСКОГО УПРАВЛЕНИЯ ДЛЯ «ЧАЙНИКОВ»

К.Ю. Поляков

*Санкт-Петербург
2008*

«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 предназначена для первого знакомства с предметом. Ее задача – объяснить «на пальцах» основные понятия *теории автоматического регулирования* и сделать так, чтобы после ее прочтения вы смогли воспринимать профессиональную литературу на эту тему. Нужно рассматривать это пособие только как фундамент, стартовую площадку для серьезного изучения серьезного предмета, который может стать очень интересным и увлекательным.

Есть сотни учебников по автоматическому управлению. Но вся проблема в том, что мозг при восприятии новой информации ищет что-то знакомое, за что можно «зацепиться», и на этой основе «привязать» новое к уже известным понятиям. Практика показывает, что читать серьезные учебники современному студенту сложно. Не за что зацепиться. Да и за строгими научными доказательствами часто ускользает суть дела, которая обычно достаточно проста. Автор попытался «спуститься» на уровень ниже и выстроить цепочку от «житейских» понятий к понятиям теории управления.

Изложение на каждом шагу грешит нестрогостью, доказательства не приводятся, формулы используются только там, где без них нельзя. Математик найдет здесь много недоговоренностей и упущений, поскольку (в соответствии с целями пособия) между строгостью и понятностью выбор всегда делается в пользу понятности.

От читателя требуются небольшие предварительные знания. Нужно иметь представление о некоторых разделах курса высшей математики:

- 1) производных и интегралах;
- 2) дифференциальных уравнениях;
- 3) линейной алгебре, матрицах;
- 4) комплексных числах.

Благодарности

Автор выражает глубокую признательность д.ф.-м.н. А.Н. Чурилову, к.т.н. В.Н. Калиниченко и к.т.н. В.О. Рыбинскому, которые внимательно прочитали предварительную версию пособия и высказали много ценных замечаний, которые позволили улучшить изложение и сделать его более понятным.

Содержание

| | |
|--|----|
| 1. ОСНОВНЫЕ ПОНЯТИЯ | 4 |
| 1.1. Введение | 4 |
| 1.2. Системы управления | 4 |
| 1.3. Какие бывают системы управления? | 7 |
| 2. МАТЕМАТИЧЕСКИЕ МОДЕЛИ | 10 |
| 2.1. Что нужно знать для управления? | 10 |
| 2.2. Связь входа и выхода | 10 |
| 2.3. Как строятся модели? | 11 |
| 2.4. Линейность и нелинейность | 12 |
| 2.5. Линеаризация уравнений | 13 |
| 2.6. Управление | 17 |
| 3. МОДЕЛИ ЛИНЕЙНЫХ ОБЪЕКТОВ | 20 |
| 3.1. Дифференциальные уравнения | 20 |
| 3.2. Модели в пространстве состояний | 21 |
| 3.3. Переходная функция | 22 |
| 3.4. Импульсная характеристика (весовая функция) | 24 |
| 3.5. Передаточная функция | 25 |
| 3.6. Преобразование Лапласа | 26 |
| 3.7. Передаточная функция и пространство состояний | 29 |
| 3.8. Частотные характеристики | 31 |
| 3.9. Логарифмические частотные характеристики | 32 |
| 4. ТИПОВЫЕ ДИНАМИЧЕСКИЕ ЗВЕНЬЯ | 34 |
| 4.1. Усилитель | 34 |
| 4.2. Аperiodическое звено | 34 |
| 4.3. Колебательное звено | 36 |
| 4.4. Интегрирующее звено | 38 |
| 4.5. Дифференцирующие звенья | 39 |
| 4.6. Запоздывание | 40 |
| 4.7. «Обратные» звенья | 41 |
| 4.8. ЛАФЧХ сложных звеньев | 42 |
| 5. СТРУКТУРНЫЕ СХЕМЫ | 43 |
| 5.1. Условные обозначения | 43 |
| 5.2. Правила преобразования | 44 |
| 5.3. Типовая одноконтурная система | 45 |
| 6. АНАЛИЗ СИСТЕМ УПРАВЛЕНИЯ | 47 |
| 6.1. Требования к управлению | 47 |
| 6.2. Процесс на выходе | 47 |
| 6.3. Точность | 48 |
| 6.4. Устойчивость | 50 |
| 6.5. Критерии устойчивости | 57 |
| 6.6. Переходный процесс | 62 |
| 6.7. Частотные оценки качества | 63 |
| 6.8. Корневые оценки качества | 65 |
| 6.9. Робастность | 66 |
| 7. СИНТЕЗ РЕГУЛЯТОРОВ | 69 |
| 7.1. Классическая схема | 69 |
| 7.2. ПИД-регуляторы | 70 |
| 7.3. Метод размещения полюсов | 71 |
| 7.4. Коррекция ЛАФЧХ | 72 |
| 7.5. Комбинированное управление | 75 |
| 7.6. Инвариантность | 75 |
| 7.7. Множество стабилизирующих регуляторов | 76 |
| ЗАКЛЮЧЕНИЕ | 79 |
| ЛИТЕРАТУРА ДЛЯ ПОСЛЕДУЮЩЕГО ЧТЕНИЯ | 80 |

1. Основные понятия

1.1. Введение

С древних времен человек хотел использовать предметы и силы природы в своих целях, то есть *управлять* ими. Управлять можно неодушевленными предметами (например, перекачивая камень на другое место), животными (дрессировка), людьми (начальник – подчиненный). Множество задач управления в современном мире связано с техническими системами – автомобилями, кораблями, самолетами, станками. Например, нужно поддерживать заданный курс корабля, высоту самолета, частоту вращения двигателя, температуру в холодильнике или в печи. Если эти задачи решаются без участия человека, говорят об *автоматическом управлении*.

Теория управления пытается ответить на вопрос «как нужно управлять?». До XIX века науки об управлении не существовало, хотя первые системы автоматического управления уже были (например, ветряные мельницы «научили» разворачиваться навстречу ветру). Развитие *теории* управления началось в период промышленной революции. Сначала это направление в науке разрабатывалось механиками для решения задач *регулирования*, то есть поддержания заданного значения частоты вращения, температуры, давления в технических устройствах (например, в паровых машинах). Отсюда происходит название «теория автоматического регулирования».

Позднее выяснилось, что принципы управления можно успешно применять не только в технике, но и в биологии, экономике, общественных науках. Процессы управления и обработки информации в системах любой природы изучает наука *кибернетика*. Один из ее разделов, связанный главным образом с техническими системами, называется *теорией автоматического управления*. Кроме классических задач регулирования, она занимается также оптимизацией законов управления, вопросами приспособляемости (адаптации).

Иногда названия «теория автоматического управления» и «теория автоматического регулирования» используются как синонимы. Например, в современной зарубежной литературе вы встретите только один термин – *control theory*.

1.2. Системы управления

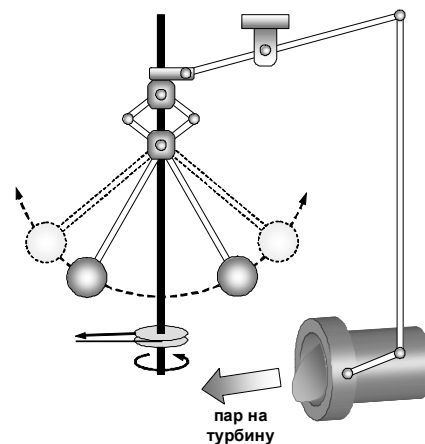
1.2.1. Из чего состоит система управления?

В задачах управления всегда есть два объекта – управляемый и управляющий. Управляемый объект обычно называют *объектом управления* или просто *объектом*, а управляющий объект – *регулятором*. Например, при управлении частотой вращения объект управления – это двигатель (электромотор, турбина); в задаче стабилизации курса корабля – корабль, погруженный в воду; в задаче поддержания уровня громкости – динамик.

Регуляторы могут быть построены на разных принципах. Самый знаменитый из первых механических регуляторов – центробежный регулятор Уатта для стабилизации частоты вращения паровой турбины (на рисунке справа). Когда частота вращения увеличивается, шарики расходятся из-за увеличения центробежной силы. При этом через систему рычагов немного закрывается заслонка, уменьшая поток пара на турбину.

Регулятор температуры в холодильнике или термостате – это электронная схема, которая включает режим охлаждения (или нагрева), если температура становится выше (или ниже) заданной.

Во многих современных системах регуляторы – это микропроцессорные устройства, компьютеры. Они успешно управляют самолетами и космическими кораблями без участия челове-



ка. Современный автомобиль буквально «напичкан» управляющей электроникой, вплоть до бортовых компьютеров.

Обычно регулятор действует на объект управления не прямо, а через исполнительные механизмы (*приводы*), которые могут усиливать и преобразовывать сигнал управления, например, электрический сигнал может «превращаться» в перемещение клапана, регулирующего расход топлива, или в поворот руля на некоторый угол.

Чтобы регулятор мог «видеть», что фактически происходит с объектом, нужны *датчики*. С помощью датчиков чаще всего измеряются те характеристики объекта, которыми нужно управлять. Кроме того, качество управления можно улучшить, если получать дополнительную информацию – измерять внутренние свойства объекта.

1.2.2. Структура системы

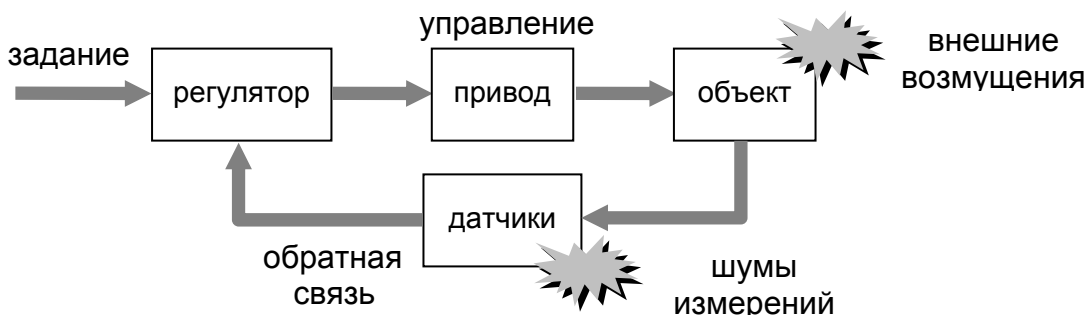
Итак, в типичную систему управления входят объект, регулятор, привод и датчики. Однако, набор этих элементов – еще не система. Для превращения в систему нужны *каналы связи*, через них идет обмен информацией между элементами. Для передачи информации могут использоваться электрический ток, воздух (пневматические системы), жидкость (гидравлические системы), компьютерные сети.

Взаимосвязанные элементы – это уже *система*, которая обладает (за счет связей) особыми свойствами, которых нет у отдельных элементов и любой их комбинации.

Основная интрига управления связана с тем, что на объект действует окружающая среда – *внешние возмущения*, которые «мешают» регулятору выполнять поставленную задачу. Большинство возмущений заранее непредсказуемы, то есть носят случайный характер.

Кроме того, датчики измеряют параметры не точно, а с некоторой ошибкой, пусть и малой. В этом случае говорят о «шумах измерений» по аналогии с шумами в радиотехнике, которые искажают сигналы.

Подводя итог, можно нарисовать структурную схему системы управления так:



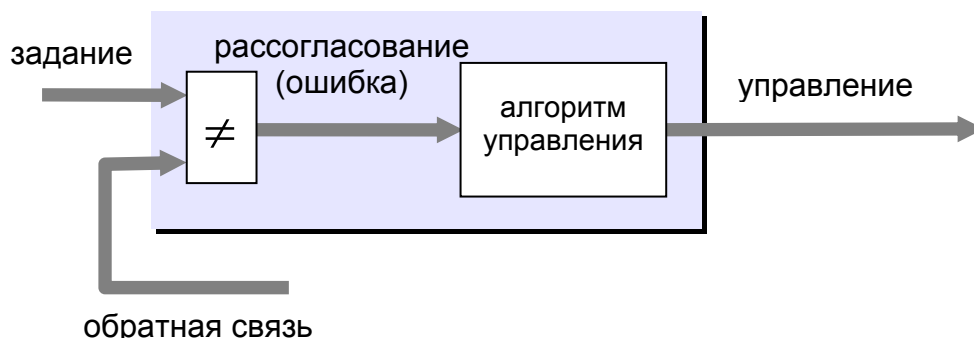
Например, в системе управления курсом корабля

- *объект управления* – это сам корабль, находящийся в воде; для управления его курсом используется руль, изменяющий направление потока воды;
- *регулятор* – цифровая вычислительная машина;
- *привод* – рулевое устройство, которое усиливает управляющий электрический сигнал и преобразует его в поворот руля;
- *датчики* – измерительная система, определяющая фактический курс;
- *внешние возмущения* – это морское волнение и ветер, отклоняющие корабль от заданного курса;
- *шумы измерений* – это ошибки датчиков.

Информация в системе управления как бы «ходит по кругу»: регулятор выдает сигнал управления на привод, который воздействует непосредственно на объект; затем информация об объекте через датчики возвращается обратно к регулятору и все начинается заново. Говорят, что в системе есть *обратная связь*, то есть регулятор использует информацию о состоянии объекта для выработки управления. Системы с обратной связью называют *замкнутыми*, поскольку информация передается по замкнутому контуру.

1.2.3. Как работает регулятор?

Регулятор сравнивает задающий сигнал («задание», «уставку», «желаемое значение») с сигналами обратной связи от датчиков и определяет *рассогласование (ошибку управления)* – разницу между заданным и фактическим состоянием. Если оно равно нулю, никакого управления не требуется. Если разница есть, регулятор выдает управляющий сигнал, который стремится свести рассогласование к нулю. Поэтому схему регулятора во многих случаях можно нарисовать так:



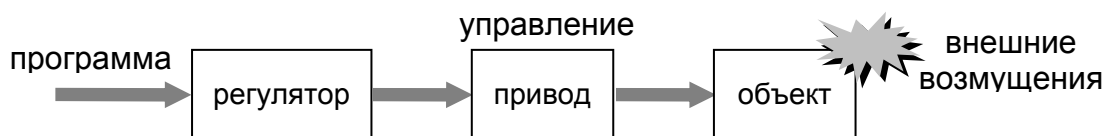
Такая схема показывает *управление по ошибке (или по отклонению)*. Это значит, что для того, чтобы регулятор начал действовать, нужно, чтобы управляемая величина *отклонилась* от заданного значения. Блок, обозначенный знаком \neq , находит рассогласование. В простейшем случае в нем из заданного значения вычитается сигнал обратной связи (измеренное значение).

Можно ли управлять объектом так, чтобы не было ошибки? В реальных системах – нет. Прежде всего, из-за внешних воздействий и шумов, которые заранее неизвестны. Кроме того, объекты управления обладают *инерционностью*, то есть, не могут мгновенно перейти из одного состояния в другое. Возможности регулятора и приводов (то есть мощность сигнала управления) всегда ограничены, поэтому быстродействие системы управления (скорость перехода на новый режим) также ограничена. Например, при управлении кораблем угол перекладки руля обычно не превышает $30-35^\circ$, это ограничивает скорость изменения курса.

Мы рассмотрели вариант, когда обратная связь используется для того, чтобы уменьшить разницу между заданным и фактическим состоянием объекта управления. Такая обратная связь называется *отрицательной*, потому что сигнал обратной связи *вычитается* из задающего сигнала. Может ли быть наоборот? Оказывается, да. В этом случае обратная связь называется *положительной*, она *увеличивает* рассогласование, то есть, стремится «раскачать» систему. На практике положительная обратная связь применяется, например, в генераторах для поддержания незатухающих электрических колебаний.

1.2.4. Разомкнутые системы

Можно ли управлять, не используя обратную связь? В принципе, можно. В этом случае регулятор не получает никакой информации о реальном состоянии объекта, поэтому должно быть точно известно, как этот объект себя ведет. Только тогда можно заранее рассчитать, как им нужно управлять (построить нужную *программу* управления). Однако при этом нельзя гарантировать, что задание будет выполнено. Такие системы называют *системами программного управления* или *разомкнутыми системами*, поскольку информация передается не по замкнутому контуру, а только в одном направлении.



Слепой и глухой водитель тоже может вести машину. Некоторое время. Пока он помнит дорогу и сможет правильно рассчитать свое место. Пока на пути не встретятся пешеходы или другие машины, о которых он заранее не может знать. Из этого простого примера ясно, что без

обратной связи (информации с датчиков) невозможно учесть влияние неизвестных факторов, неполноту наших знаний.

Несмотря на эти недостатки, разомкнутые системы применяются на практике. Например, информационное табло на вокзале. Или простейшая система управления двигателем, в которой не требуется очень точно поддерживать частоту вращения. Однако с точки зрения теории управления разомкнутые системы малоинтересны, и мы не будем больше про них вспоминать.

1.3. Какие бывают системы управления?

Автоматическая система – это система, работающая без участия человека. Есть еще *автоматизированные* системы, в которых рутинные процессы (сбор и анализ информации) выполняет компьютер, но управляет всей системой человек-оператор, который и принимает решения. Мы будем далее изучать только автоматические системы.

1.3.1. Задачи систем управления

Автоматические системы управления применяются для решения трех типов задач:

- *стабилизация*, то есть поддержание заданного режима работы, который не меняется длительное время (задающий сигнал – постоянная, часто нуль);
- *программное управление* – управление по заранее известной программе (задающий сигнал меняется, но заранее известен);
- *слежение* за неизвестным задающим сигналом.

К системам стабилизации относятся, например, авторулевые на кораблях (поддержание заданного курса), системы регулирования частоты вращения турбин. Системы программного управления широко используются в бытовой технике, например, в стиральных машинах. Следящие системы служат для усиления и преобразования сигналов, они применяются в приводах и при передаче команд через линии связи, например, через Интернет.

1.3.2. Одномерные и многомерные системы

По количеству входов и выходов бывают

- *одномерные* системы, у которых один вход и один выход (они рассматриваются в так называемой *классической* теории управления);
- *многомерные* системы, имеющие несколько входов и./или выходов (главный предмет изучения *современной* теории управления).

Мы будем изучать только одномерные системы, где и объект, и регулятор имеют один входной и один выходной сигнал. Например, при управлении кораблем по курсу можно считать, что есть одно управляющее воздействие (поворот руля) и одна регулируемая величина (курс).

Однако, в самом деле это не совсем верно. Дело в том, что при изменении курса меняется также крен и дифферент корабля. В одномерной модели мы пренебрегаем этими изменениями, хотя они могут быть очень существенными. Например, при резком повороте крен может достигнуть недопустимого значения. С другой стороны, для управления можно использовать не только руль, но и различные подруливающие устройства, стабилизаторы качки и т.п., то есть объект имеет несколько входов. Таким образом, реальная система управления курсом – многомерная.

Исследование многомерных систем – достаточно сложная задача и выходит за рамки этого пособия. Поэтому в инженерных расчетах стараются иногда упрощенно представить многомерную систему как несколько одномерных, и довольно часто такой метод приводит к успеху.

1.3.3. Непрерывные и дискретные системы

По характеру сигналов системы могут быть

- *непрерывными*, в которых все сигналы – функции непрерывного времени, определенные на некотором интервале;
- *дискретными*, в которых используются дискретные сигналы (последовательности чисел), определенные только в отдельные моменты времени;

- *непрерывно-дискретными*, в которых есть как непрерывные, так и дискретные сигналы.

Непрерывные (или *аналоговые*) системы обычно описываются дифференциальными уравнениями. Это все системы управления движением, в которых нет компьютеров и других элементов дискретного действия (микропроцессоров, логических интегральных схем).

Микропроцессоры и компьютеры – это дискретные системы, поскольку в них вся информация хранится и обрабатывается в дискретной форме. Компьютер не может обрабатывать непрерывные сигналы, поскольку работает только с *последовательностями* чисел. Примеры дискретных систем можно найти в экономике (период отсчета – квартал или год) и в биологии (модель «хищник-жертва»). Для их описания применяют *разностные* уравнения.

Существуют также и гибридные *непрерывно-дискретные* системы, например, компьютерные системы управления движущимися объектами (кораблями, самолетами, автомобилями и др.). В них часть элементов описывается дифференциальными уравнениями, а часть – разностными. С точки зрения математики это создает большие сложности для их исследования, поэтому во многих случаях непрерывно-дискретные системы сводят к упрощенным чисто непрерывным или чисто дискретным моделям.

1.3.4. Стационарные и нестационарные системы

Для управления очень важен вопрос о том, изменяются ли характеристики объекта со временем. Системы, в которых все параметры остаются постоянными, называются *стационарными*, что значит «не изменяющиеся во времени». В этом пособии рассматриваются только стационарные системы.

В практических задачах часто дело обстоит не так радужно. Например, летящая ракета расходует топливо и за счет этого ее масса изменяется. Таким образом, ракета – нестационарный объект. Системы, в которых параметры объекта или регулятора изменяются со временем, называются *нестационарными*. Хотя *теория* нестационарных систем существует (формулы написаны), применить ее на практике не так просто.

1.3.5. Определенность и случайность

Самый простой вариант – считать, что все параметры объекта определены (заданы) точно, так же, как и внешние воздействия. В этом случае мы говорим о *детерминированных* системах, которые рассматривались в классической теории управления.

Тем не менее, в реальных задачах точных данных у нас нет. Прежде всего, это относится к внешним воздействиям. Например, для исследования качки корабля на первом этапе можно считать, что волна имеет форму синуса известной амплитуды и частоты. Это детерминированная модель. Так ли это на практике? Естественно нет. С помощью такого подхода можно получить только приближенные, грубые результаты.

По современным представлениям форма волны приближенно описывается как сумма синусоид, которые имеют *случайные*, то есть неизвестные заранее, частоты, амплитуды и фазы. Помехи, шум измерений – это тоже случайные сигналы.

Системы, в которых действуют случайные возмущения или параметры объекта могут изменяться случайным образом, называются *стохастическими* (вероятностными). Теория стохастических систем позволяет получать только вероятностные результаты. Например, нельзя гарантировать, что отклонение корабля от курса *всегда* будет составлять не более 2° , но можно попытаться обеспечить такое отклонение с некоторой *вероятностью* (вероятность 99% означает, что требование будет выполнено в 99 случаях из 100).

1.3.6. Оптимальные системы

Часто требования к системе можно сформулировать в виде *задачи оптимизации*. В *оптимальных* системах регулятор строится так, чтобы обеспечить минимум или максимум какого-то критерия качества. Нужно помнить, что выражение «оптимальная система» не означает, что она действительно идеальная. Все определяется принятым критерием – если он выбран удачно, система получится хорошая, если нет – то наоборот.

1.3.7. Особые классы систем

Если параметры объекта или возмущений известны неточно или могут изменяться со временем (в нестационарных системах), применяют *адаптивные* или самонастраивающиеся регуляторы, в которых закон управления меняется при изменении условий. В простейшем случае (когда есть несколько заранее известных режимов работы) происходит простое переключение между несколькими законами управления. Часто в адаптивных системах регулятор оценивает параметры объекта в реальном времени и соответственно изменяет закон управления по заданному правилу.

Самонастраивающаяся система, которая пытается настроить регулятор так, чтобы «найти» максимум или минимум какого-то критерия качества, называется *экстремальной* (от слова *экстремум*, обозначающего максимум или минимум).

Во многих современных бытовых устройствах (например, в стиральных машинах) используются *нечеткие регуляторы*, построенные на принципах *нечеткой логики*. Этот подход позволяет формализовать человеческий способ принятия решения: «если корабль ушел сильно вправо, руль нужно сильно переложить влево».

Одно из популярных направлений в современной теории – применение достижений искусственного интеллекта для управления техническими системами. Регулятор строится (или только настраивается) на основе *нейронной сети*, которую предварительно обучает человек-эксперт.

2. Математические модели

2.1. Что нужно знать для управления?

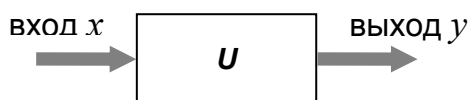
Цель любого управления – изменить состояние объекта нужным образом (в соответствии с заданием). Теория автоматического регулирования должна ответить на вопрос: «как построить регулятор, который может управлять данным объектом так, чтобы достичь цели?» Для этого разработчику необходимо знать, как система управления будет реагировать на разные воздействия, то есть нужна *модель* системы: объекта, привода, датчиков, каналов связи, возмущений, шумов.

Модель – это объект, который мы используем для изучения другого объекта (*оригинала*). Модель и оригинал должны быть в чем-то похожи, чтобы выводы, сделанные при изучении модели, можно было бы (с некоторой вероятностью) перенести на оригинал. Нас будут интересовать в первую очередь *математические модели*, выраженные в виде формул. Кроме того, в науке используются также описательные (словесные), графические, табличные и другие модели.

2.2. Связь входа и выхода

Любой объект взаимодействует с внешней средой с помощью входов и выходов. Входы – это возможные воздействия на объект, выходы – это те сигналы, которые можно измерить. Например, для электродвигателя входами могут быть напряжение питания и нагрузка, а выходами – частота вращения вала, температура.

Входы независимы, они «приходят» из внешней среды. При изменении информации на входе меняется внутреннее *состояние объекта* (так называют его изменяющиеся свойства) и, как следствие, выходы:



Это значит, что существует некоторое правило, по которому элемент преобразует вход x в выход y . Это правило называется *оператором*. Запись $y = U[x]$ означает, что выход y получен в результате применения оператора U ко входу x .

Построить модель – это значит найти оператор, связывающий входы и выходы. С его помощью можно предсказать реакцию объекта на любой входной сигнал.

Рассмотрим электродвигатель постоянного тока. Вход этого объекта – это напряжение питания (в вольтах), выход – частота вращения (в оборотах в секунду). Будем считать, что при напряжении 1 В частота вращения равна 1 об/сек, а при напряжении 2 В – 2 об/сек, то есть частота вращения равна по величине напряжению¹. Легко видеть, что действие такого оператора можно записать в виде

$$U[x] = x.$$

Теперь предположим, что этот же двигатель вращает колесо и в качестве выхода объекта мы выбрали число оборотов колеса относительно начального положения (в момент $t = 0$). В этом случае при равномерном вращении произведение $x \cdot \Delta t$ дает нам количество оборотов за время Δt , то есть $y(t) = x \cdot \Delta t$ (здесь запись $y(t)$ явно обозначает зависимость выхода от времени t). Можно ли считать, что этой формулой мы определили оператор U ? Очевидно, что нет, потому что полученная зависимость справедлива только для постоянного входного сигнала. Если напряжение на входе $x(t)$ меняется (все равно как!), угол поворота запишется в виде интеграла

¹ Конечно, это будет справедливо только в некотором диапазоне напряжений.

$$U[x] = \int_0^t x(t) dt.$$

Оператор, который действует по такому правилу, называется *оператором интегрирования*. С помощью этого оператора можно, например, описать наполнение пустого бака водой. Если сечение бака S (в m^2) постоянно по всей его высоте, то уровень воды h определяется как интеграл от потока воды q (в m^3/c), деленный на S :

$$h(t) = \frac{1}{S} \int_0^t q(t) dt,$$

Обратный оператор – **оператор дифференцирования** – вычисляет производную:

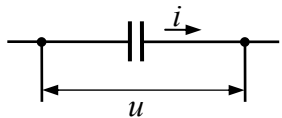
$$U[x(t)] = \dot{x}(t) = \frac{dx(t)}{dt}.$$

Как мы увидим, этот оператор играет очень важную роль в описании объектов управления.

Обычно оператор дифференцирования обозначается буквой p . Запись $y(t) = p x(t)$ внешне выглядит как «умножение» оператора p на сигнал $x(t)$, но на самом деле обозначает *действие* этого оператора, то есть дифференцирование:

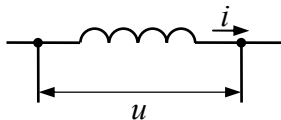
$$p x(t) = \frac{dx(t)}{dt}. \quad (1)$$

Где встречаются такие операторы? Приведем примеры из электротехники. Например, известно, что ток i (в *амперах*), проходящий по цепи с конденсатором, пропорционален производной от разности потенциалов u (в *вольтах*) на его пластинах:



$$i(t) = C \frac{du(t)}{dt} = C p u(t)$$

Здесь C – емкость конденсатора (измеряется в *фарадах*). Кроме того, падение напряжения u на катушке индуктивности пропорционально производной от проходящего тока i :



$$u(t) = L \frac{di(t)}{dt} = L p i(t)$$

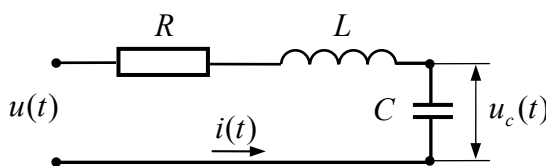
где L – индуктивность (измеряется в *генри*).

Оператор дифференцирования – это идеальный (*физически нереализуемый*) оператор, его невозможно реализовать на практике. Чтобы понять это вспомним, что при мгновенном изменении сигнала его производная (скорость возрастания) будет равна бесконечности, а никакое реальное устройство не может работать с бесконечными сигналами.

2.3. Как строятся модели?

Во-первых, математические модели могут быть получены теоретически **из законов физики** (законы сохранения массы, энергии, импульса). Эти модели описывают внутренние связи в объекте и, как правило, наиболее точны.

Рассмотрим RLC -цепочку, то есть последовательное соединение резистора с сопротивлением R (в *омах*), катушки индуктивности с индуктивностью L и конденсатора с емкостью C . Она может быть описана с помощью двух уравнений:



$$u(t) = u_c(t) + L \frac{di(t)}{dt} + R \cdot i(t)$$

$$i(t) = C \frac{du_c(t)}{dt}$$

Первое уравнение означает, что разность потенциалов на концах RLC -цепочки равна сумме разностей потенциалов на всех промежуточных участках. Разность потенциалов $R \cdot i(t)$ на рези-

сторе вычисляется по закону Ома, а на катушке – по формуле, приведенной в предыдущем параграфе. Второе уравнение описывает связь между напряжением и током для конденсатора. Вход этого объекта – напряжение $u(t)$ на концах цепочки, а выход – разность потенциалов $u_c(t)$ на пластинах конденсатора.

Второй способ – построение модели в результате **наблюдения за объектом** при различных входных сигналах (этим занимается *теория идентификации*). Объект рассматривается как «черный ящик», то есть, его внутреннее устройство неизвестно. Мы смотрим, как он реагирует на входные сигналы, и стараемся подстроить модель так, чтобы выходы модели и объекта совпадали как можно точнее при разнообразных входах.

На практике часто используется смешанный способ: структура модели (вид уравнения, связывающего вход и выход) определяется из теории, а коэффициенты находят опытным путем. Например, общий вид уравнений движения корабля хорошо известен, однако в этих уравнениях есть коэффициенты, которые зависят от многих факторов (формы корпуса, шероховатости поверхности и т.п.), так что их крайне сложно (или невозможно) найти теоретически. В этом случае для определения неизвестных коэффициентов строят масштабные модели и испытывают их в бассейнах по специальным методикам. В авиастроении для тех же целей используют аэродинамические трубы.

Для любого объекта управления можно построить множество различных моделей, которые будут учитывать (или не учитывать) те или иные факторы. Обычно на первом этапе стараются описать объект как можно более подробно, составить детальную модель. Однако при этом будет трудно теоретически рассчитать закон управления, который отвечает заданным требованиям к системе. Даже если мы сможем его рассчитать, он может оказаться слишком сложным для реализации или очень дорогим.

С другой стороны, можно упростить модель объекта, отбросив некоторые «детали», которые кажутся разработчику маловажными. Для упрощенной модели закон управления также получается проще, и с его помощью часто можно добиться желаемого результата. Однако в этом случае нет гарантии, что он будет так же хорошо управлять полной моделью (и реальным объектом).

Обычно используется компромиссный вариант. Начинают с простых моделей, стараясь спроектировать регулятор так, чтобы он «подходил» и для сложной модели. Это свойство называют *робастностью* (*грубостью*) регулятора (или системы), оно означает нечувствительность к ошибкам моделирования. Затем проверяют работу построенного закона управления на полной модели или на реальном объекте. Если получен отрицательный результат (простой регулятор «не работает»), усложняют модель, вводя в нее дополнительные подробности. И все начинается сначала.

2.4. Линейность и нелинейность

Из школьной математики известно, что проще всего решать *линейные* уравнения. С нелинейными уравнениями (квадратными, кубическими и др.) работать намного сложнее, многие типы уравнений математика пока не умеет решать аналитически (точно).

Среди операторов самые простые – также *линейные*. Они обладают двумя свойствами²:

- **умножение на константу**: $U[\alpha \cdot x] = \alpha \cdot U[x]$, где α – любая постоянная (то есть, при увеличении входа в несколько раз выход увеличивается во столько же раз);
- **принцип суперпозиции**: если на вход подать сумму двух сигналов, выход будет представлять собой сумму реакций того же оператора на отдельные сигналы:

$$U[x_1 + x_2] = U[x_1] + U[x_2].$$

Модели, которые описываются линейными операторами, называются *линейными*. С ними можно работать с помощью методов *теории линейных систем*, которая наиболее развита и позволяет точно решать большинство известных практических задач.

² В математике эти свойства называют *однородность* и *аддитивность*.

Однако, все модели реальных систем – *нелинейные*. Это легко понять хотя бы потому, что всегда есть предельно допустимое значение входного сигнала – при его превышении объект может просто выйти из строя или даже разрушиться (линейность нарушается). Методы исследования нелинейных операторов очень сложны математически, в *теории нелинейных систем* точные решения известны только для достаточно узкого круга задач. Здесь пока больше «белых пятен», чем полученных результатов, хотя это научное направление активно развивается в последние годы.

Что же делать? Чаще всего сначала проводят *линеаризацию* нелинейной модели объекта (привода), то есть строят приближенную линейную модель. Затем на основе этой модели проектируют закон управления, применяя точные методы теории линейных систем. Наконец, проверяют полученный регулятор с помощью компьютерного моделирования на полной нелинейной модели.

Нужно отметить, что если объект или привод имеют так называемую «существенную» нелинейность, этот подход может не сработать. Тогда приходится использовать методы нелинейной теории, а также компьютерное моделирование. Моделирование стало очень популярным в последнее время, поскольку появились мощные компьютерные программы для проведения вычислительных экспериментов, и можно проверить поведение системы при разнообразных допустимых входных сигналах.

Таким образом, в классификацию систем управления в разделе 1.3 нужно добавить еще одно деление, может быть, самое существенное – системы бывают *линейные* и *нелинейные*. В линейных системах все звенья описываются линейными операторами, и это значительно упрощает работу с ними.

2.5. Линеаризация уравнений

Вы уже знаете, что в теории управления лучше всего разработаны методы исследования линейных систем. Однако строго линейных систем в окружающем нас мире не существует. Поэтому для того, чтобы эти методы можно было применить на практике, нужно выполнить *линеаризацию* – построить приближенную линейную модель на основе более реалистичной нелинейной модели объекта.

2.5.1. Алгебраические уравнения

Представим себе бак с водой. В нижней части бака просверлено отверстие, через которое вытекает вода. Площадь сечения бака обозначим через S , а площадь сечения отверстия – через S_0 .

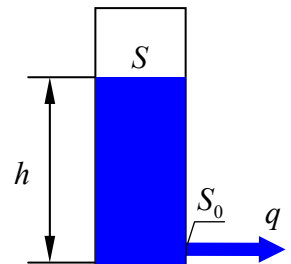
Построим модель, которая связывает уровень воды в баке h (в метрах) и расход вытекающей воды q (в $\text{м}^3/\text{с}$). Эту связь можно найти с помощью закона Бернулли, который в данном случае принимает вид

$$\rho g h = \frac{\rho v^2}{2}.$$

Здесь ρ – плотность жидкости (в $\text{кг}/\text{м}^3$), $g \approx 9,81 \text{ м}/\text{с}^2$ – ускорение свободного падения, v – скорость вытекания жидкости (в $\text{м}/\text{с}$). Отсюда получаем $v = \sqrt{2gh}$. Учитывая, что расход воды вычисляется как $q = S_0 \cdot v$, находим

$$q = \alpha \sqrt{h}, \quad (2)$$

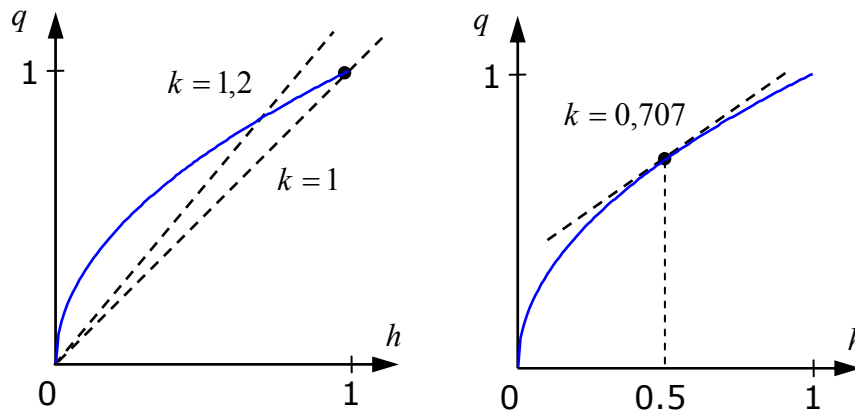
где $\alpha = S_0 \sqrt{2g}$ – постоянная величина. Это *статическая* модель, потому что она не содержит производных, характеризующих изменение сигналов во времени. Статическая модель описывает *установившееся* состояние (*статический режим*), когда в баке поддерживается постоянный уровень воды и поток вытекающей воды тоже постоянный.



Очевидно, что модель (2) – нелинейная, поскольку содержит \sqrt{h} . Линеаризовать ее – значит приближенно заменить уравнение (2) линейным уравнением $q = k \cdot h$, где k – некоторый коэффициент. Как его выбрать? На этот вопрос нет однозначного ответа.

Предположим, что уровень воды изменяется в интервале от 0 до 1 м. Тогда один из вариантов – вычислить коэффициент как угол наклона отрезка, соединяющего точки кривой $q = \alpha\sqrt{h}$ на концах этого интервала. Для определенности далее везде принимаем $\alpha = 1$, тогда получаем $k = 1$.

Конечно, эта модель очень грубая и дает большую ошибку, особенно для уровней в диапазоне от 0,1 до 0,6. Чтобы уменьшить ошибку, можно попробовать несколько изменить k (например, увеличив его до 1,2), однако точность приближения по-прежнему будет невысока, хотя и чуть-чуть лучше, чем в первом случае.



Теперь предположим, что обычно уровень мало изменяется вблизи среднего значения $h = 0,5$ м. В этом случае можно применить другой подход. Заметим, что в этой области кривая $q = \alpha\sqrt{h}$ почти совпадает с касательной в точке $(0,5; \frac{\sqrt{2}}{2})$, угол наклона которой равен производной

$$k = \left. \frac{dq}{dh} \right|_{h=0,5} = \left. \frac{1}{2\sqrt{h}} \right|_{h=0,5} = \frac{\sqrt{2}}{2} \approx 0,707.$$

Касательная – это прямая с наклоном k , проходящая через точку $(0,5; \frac{\sqrt{2}}{2})$, ее уравнение имеет вид $q = kh + b$. Свободный член b определим из равенства

$$\frac{\sqrt{2}}{2} = kh + b = \frac{\sqrt{2}}{2} \cdot 0,5 + b \quad \Rightarrow \quad b = \frac{\sqrt{2}}{4} \approx 0,354,$$

так что получаем модель

$$q = \frac{\sqrt{2}}{2}h + \frac{\sqrt{2}}{4}. \quad (3)$$

Это линейное уравнение, однако модель (3) – нелинейная, поскольку для нее не выполняется, например, свойство умножения на константу. Это легко проверить, сравнив $U[2 \cdot h]$ и $2 \cdot U[h]$:

$$U[2 \cdot h] = \sqrt{2}h + \frac{\sqrt{2}}{4}, \quad 2 \cdot U[h] = \sqrt{2}h + \frac{\sqrt{2}}{2} \neq U[2 \cdot h].$$

Принцип суперпозиции также не выполняется.

Для того, чтобы получить из (3) линейную модель, нужно записать **уравнения в отклонениях** от рабочей точки $(h_0; q_0)$, в которой мы определяли наклон касательной. Из (3) следует, что

$$q_0 + \Delta q = \frac{\sqrt{2}}{2} \cdot (h_0 + \Delta h) + \frac{\sqrt{2}}{4}. \quad (4)$$

Поскольку график зависимости (3) проходит через точку $(h_0; q_0)$, можно применить равенство

$$q_0 = \frac{\sqrt{2}}{2} h_0 + \frac{\sqrt{2}}{4}. \text{ Тогда из (4) находим}$$

$$\Delta q = \frac{\sqrt{2}}{2} \cdot \Delta h. \quad (5)$$

Полученное таким образом уравнение – это *линейная* модель объекта, записанная в отклонениях входа и выхода от *номинальной (рабочей) точки* $(h_0; q_0)$. Приближенная модель (5) точнее всего соответствует объекту вблизи этой точки, а при больших отклонениях от нее ошибка может значительно возрастать.

На этом простом примере мы познакомились с основными принципами линеаризации нелинейных алгебраических уравнений. В следующем параграфе те же самые идеи используются для более сложной модели, которая описывает *динамику* системы (изменение во времени).

2.5.2. Дифференциальные уравнения

Реальные объекты не могут мгновенно изменять свое состояние, поэтому вместо статических моделей типа (2) для их исследования используют *динамические* модели, которые описываются дифференциальными уравнениями, содержащими производные (скорости изменения сигналов). Как мы видели в разделе 2.3, такие модели могут быть получены из физических законов. Во многих случаях более или менее точные модели представляют собой *нелинейные* дифференциальные уравнения, поэтому для того, чтобы применить теорию линейных систем, требуется *линеаризация*. При этом применяется почти та же методика, что и для алгебраических уравнений.

Идея линеаризации заключается в том, что в системах регулирования (поддержания заданных значений величин) сигналы мало отклоняются от *рабочей точки* – некоторого положения равновесия, в котором все сигналы имеют «правильные» значения и их производные равны нулю. Поэтому для решения задач управления часто достаточно использовать линейную модель в отклонениях от этой рабочей точки.

Модель, только что построенная для бака с водой, не совсем правильная, потому что не учитывает, что уровень в баке изменяется – уменьшается по мере вытекания воды. Кроме того, предположим, что для поддержания уровня используется насос, который подкачивает воду в бак, его расход обозначим через Q . Для такого объекта входом является расход Q , а выходом – изменение уровня h .

Предположим, что в течение маленького интервала Δt расходы Q и q можно считать постоянными. За это время объем воды, добавленной в бак насосом, равен $Q \cdot \Delta t$, а объем «ушедшей» воды – $q \cdot \Delta t$. Учитывая, что площадь сечения бака равна S , получаем изменение уровня:

$$\Delta h = \frac{(Q - q)}{S} \cdot \Delta t. \text{ Переходя к пределу при } \Delta t \rightarrow 0, \text{ получаем дифференциальное уравнение}$$

$$\frac{dh(t)}{dt} = \frac{1}{S} [Q(t) - q(t)].$$

Эта модель учитывает, что уровень воды и расходы изменяются во времени. Вспомним, что расход вытекающей жидкости $q(t)$ зависит от уровня воды в баке $h(t)$ и связан с ним нелинейной зависимостью $q(t) = \alpha \cdot \sqrt{h(t)}$. Поэтому уравнение можно записать в виде

$$\frac{dh(t)}{dt} = \frac{1}{S} Q(t) - \frac{\alpha}{S} \sqrt{h(t)}. \quad (6)$$

Здесь остались только две изменяющиеся величины: расход насоса $Q(t)$ (вход объекта) и уровень воды $h(t)$ (выход). Далее для упрощения записи мы не будем явно указывать зависимость этих сигналов от времени.

В установившемся (*статическом*) режиме, когда сигналы не изменяются, все производные равны нулю. В нашем случае, приняв $\frac{dh(t)}{dt} = 0$ в (6), получаем

$$0 = Q - \alpha \cdot \sqrt{h} \Rightarrow h = \frac{Q^2}{\alpha^2}. \quad (7)$$

Эта зависимость между *установившимися* значениями входа Q и выхода h называется **статической характеристикой**. Она позволяет для любого заданного *постоянного* значения Q на входе получить значение выхода h .

Теперь предположим, что задана некоторая рабочая точка, то есть, значения входа $Q = Q_0$ и выхода $h = h_0$ удовлетворяют уравнению (7), и система все время работает около этого положения равновесия. Вблизи этой точки

$$Q = Q_0 + \Delta Q \quad \text{и} \quad h = h_0 + \Delta h,$$

где ΔQ и Δh – малые отклонения входа и выхода от рабочей точки.

Дальше для линеаризации используется разложение функций в *ряд Тейлора*. Для некоторой функции $f(x, y)$ в окрестности точки (x_0, y_0) этот ряд имеет вид:

$$f(x, y) = f(x_0, y_0) + \frac{\partial f(x_0, y_0)}{\partial x} \cdot \Delta x + \frac{\partial f(x_0, y_0)}{\partial y} \cdot \Delta y + F(x, y),$$

где $\frac{\partial f(x_0, y_0)}{\partial x}$ и $\frac{\partial f(x_0, y_0)}{\partial y}$ – частные производные функции $f(x, y)$ по x и по y в точке (x_0, y_0) , а $F(x, y)$ зависит от высших производных в той же точке (второй, третьей и т.д.). При малых значениях Δx и Δy можно считать, что «хвост» этого ряда $F(x, y)$ очень мал, примерно равен нулю, поэтому

$$f(x, y) \approx f(x_0, y_0) + \frac{\partial f(x_0, y_0)}{\partial x} \cdot \Delta x + \frac{\partial f(x_0, y_0)}{\partial y} \cdot \Delta y. \quad (8)$$

Применим формулу (8) для линеаризации правой части уравнения (6), где в роли x выступает расход Q , а в роли y – уровень h . Выполняя дифференцирование, находим

$$\frac{\partial}{\partial Q} \left[\frac{1}{S} Q - \frac{\alpha}{S} \sqrt{h} \right] = \frac{1}{S}, \quad \frac{\partial}{\partial h} \left[\frac{1}{S} Q - \frac{\alpha}{S} \sqrt{h} \right] = -\frac{\alpha}{2S\sqrt{h}}.$$

Тогда с помощью формулы (8) получаем

$$\frac{1}{S} Q - \frac{\alpha}{S} \sqrt{h} \approx \frac{1}{S} Q_0 - \frac{\alpha}{S} \sqrt{h_0} + \frac{1}{S} \cdot \Delta Q - \frac{\alpha}{2S\sqrt{h_0}} \cdot \Delta h.$$

Подставим $Q = Q_0 + \Delta Q$ и $h = h_0 + \Delta h$ в уравнение (6) и учтем, что $\frac{d(h_0 + \Delta h)}{dt} = \frac{d\Delta h}{dt}$. Тогда

$$\frac{d\Delta h}{dt} \approx \frac{1}{S} Q_0 - \frac{\alpha}{S} \sqrt{h_0} + \frac{1}{S} \cdot \Delta Q - \frac{\alpha}{2S\sqrt{h_0}} \cdot \Delta h.$$

Вспомогая, что Q_0 и h_0 соответствуют статическому режиму, то есть $\frac{1}{S} Q_0 - \frac{\alpha}{S} \sqrt{h_0} = 0$, получаем **линеаризованное уравнение** в отклонениях от рабочей точки:

$$\frac{d\Delta h}{dt} + k_h \cdot \Delta h \approx k_Q \cdot \Delta Q, \quad (9)$$

где $k_h = \frac{\alpha}{2S\sqrt{h_0}}$ и $k_Q = \frac{1}{S}$. Заметим, что коэффициент k_h зависит от h_0 , то есть от выбора рабочей точки. В этом проявляется нелинейность объекта.

Обычно при записи линеаризованного уравнения знак Δ (обозначающий отклонение) не пишут. Таким образом, окончательно получаем линеаризованную модель

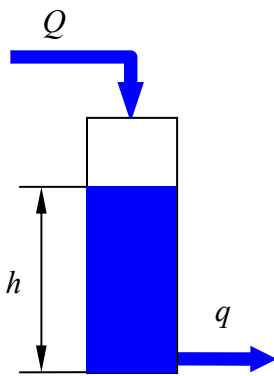
$$\frac{dh(t)}{dt} + k_h \cdot h(t) = k_Q \cdot Q(t). \quad (10)$$

Но нужно помнить, что это *уравнение в отклонениях*, и оно справедливо только при малых отклонениях от рабочей точки (Q_0, h_0) . При выборе другой рабочей точки коэффициент k_h получится другой.

2.6. Управление

Посмотрим на примере, как можно управлять объектом и что из этого получается. Немного изменим предыдущую задачу, разрешив потоку вытекающей жидкости q изменяться независимо (в теории управления это называется *нагрузкой* на объект).

Для того, чтобы обеспечить водой всех жителей деревни, построили водонапорную башню, в которую насосом закачивается вода из реки. Каждый житель может в любой момент включить воду на своем участке, например, для полива. Нужно построить систему, которая автоматически поддерживает заданный уровень h_0 воды в цистерне (в метрах).



Будем считать, что жителей довольно много, поэтому у кого-то всегда включена вода и насос постоянно работает на закачку воды в цистерну. Для управления уровнем воды h мы можем изменять его поток Q (в $\text{м}^3/\text{с}$). Таким образом, уровень h – это регулируемая величина, а поток Q – сигнал управления. Для обратной связи используем датчик, измеряющий уровень воды h в цистерне.

Построим математическую модель объекта, то есть цистерны. Поток на выходе q (в $\text{м}^3/\text{с}$) показывает, сколько воды вытекает из цистерны за 1 с – это *нагрузка*.

Изменение уровня Δh зависит от разности потоков $Q - q$ и площади сечения цистерны S .

Если разность потоков постоянна в течение интервала времени Δt , то $\Delta h(t) = \frac{Q(t) - q(t)}{S} \cdot \Delta t$. В общем случае нужно использовать интеграл:

$$\Delta h(t) = \frac{1}{S} \int_0^t (Q(t) - q(t)) dt.$$

Пусть в момент времени $t = 0$ уровень воды равен заданному значению, а входной и выходной потоки равны ($Q(0) = q(0) = q_0$), так что уровень не меняется. Этот режим мы примем за *номинальный* (рабочую точку). Для того, чтобы получить уравнение в отклонениях, представим потоки в виде

$$Q(t) = q_0 + \Delta Q(t), \quad q(t) = q_0 + \Delta q(t),$$

где $\Delta Q(t)$ и $\Delta q(t)$ – малые отклонения потоков от номинального режима. Тогда, опуская знак приращения Δ , можно записать *модель объекта управления* в форме

$$h(t) = \frac{1}{S} \int_0^t (Q(t) - q(t)) dt.$$

Здесь $h(t)$, $Q(t)$ и $q(t)$ обозначают отклонения этих величин от номинальных значений. Заметим, что эта модель может быть записана как дифференциальное уравнение (если найти производные обеих частей равенства):

$$\frac{dh(t)}{dt} = \frac{1}{S} [Q(t) - q(t)].$$

Для упрощения далее примем $S = 1 \text{ м}^2$.

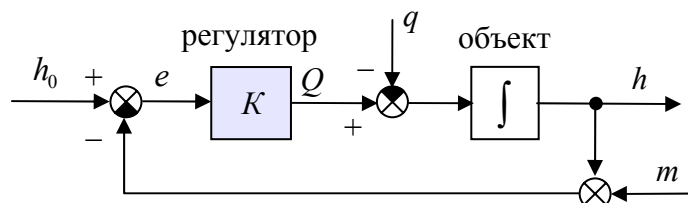
В качестве обратной связи мы будем использовать сигнал с датчика уровня. Ошибка управления вычисляется как разница между заданным и измеренным уровнями воды:

$$e(t) = h_0(t) - h(t).$$

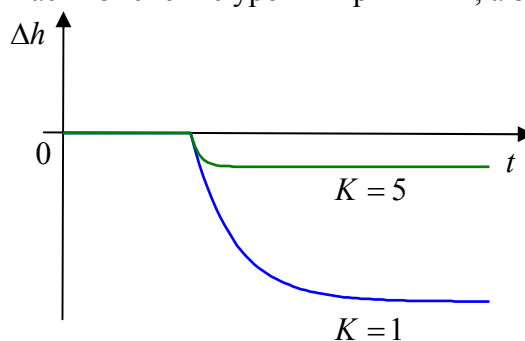
Применим самый простой регулятор – усилитель с коэффициентом K (или *пропорциональный регулятор, П-регулятор*), который управляет потоком по закону

$$q(t) = K \cdot e(t) = K \cdot [h_0(t) - h(t)].$$

Структурная схема системы управления показана на рисунке ниже. Знак интеграла обозначает звено, модель которого – оператор интегрирования. С помощью кружка с секторами обозначается сложение сигналов. Если какой-то сектор закрашен черным цветом, входящий в него сигнал вычитается (учитывается в сумме со знаком «минус»). Кроме сигналов, о которых уже шла речь, на рисунке показан также шум измерения $m(t)$, искажающий показания датчика.



Проверим работу этого регулятора при различных значениях коэффициента K . Сначала будем считать, что шума измерений нет, то есть уровень измеряется точно. Предположим, что расход воды на выходе q увеличивается скачком (все начали поливать огороды). Синяя линия на рисунке (см. ниже) показывает изменение уровня при $K = 1$, а зеленая – при $K = 5$.

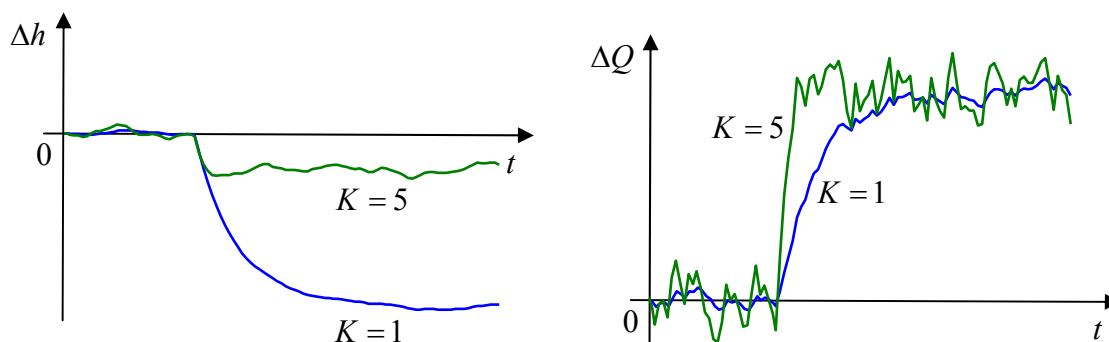


По этим данным можно сделать некоторые выводы:

- при изменении нагрузки (потребления воды, потока q) регулятор-усилитель не может поддерживать заданный уровень (графики не приходят к значению $\Delta h = 0$);
- чем больше K , тем меньше ошибка регулирования Δh в установившемся режиме; можно ожидать, что при $K \rightarrow \infty$ ошибка должна уменьшиться до нуля;
- чем больше K , тем быстрее заканчивается переход на новый режим.

Кажется, что для улучшения управления нужно увеличивать K , однако это только первое впечатление.

Теперь посмотрим, что будет, если есть шум измерений (случайная ошибка датчика).



По графикам видно, что при неточных измерениях уровень колеблется около некоторого среднего значения (того, что было получено без шума), причем при бóльшем K колебания увеличиваются. Этот эффект особенно хорошо виден на графике изменения расхода насоса Δq (рисунок справа).

При увеличении K повышение точности (уменьшение установившейся ошибки) достигается за счет повышенной активности насоса, который все время «дергается». При этом механические части изнашиваются, и существенно уменьшается его срок службы. Поэтому коэффициент K нельзя сильно увеличивать.

Один из главных выводов этого примера: управление чаще всего связано с *компромиссом*. Здесь, с одной стороны, нужно **увеличивать K , чтобы повысить точность**, а с другой – **нужно уменьшать K , чтобы уменьшить влияние шума измерения**.

При выборе управления мы шли самым простым путем, остановившись на регуляторе-усилителе (П-регуляторе). У вдумчивого читателя неизбежно должны были возникнуть вопросы следующего характера:

- любым ли объектом можно управлять с помощью регулятора-усилителя?
- как правильно выбрать коэффициент K (на каком значении остановиться)?
- можно ли добиться улучшения управления с помощью более сложного регулятора?
- какой регулятор нужно применить, чтобы улучшить управление?
- как обеспечить нулевую установившуюся ошибку (постоянный уровень при любом расходе q) и можно ли это сделать вообще?
- как подавить шумы измерений, чтобы они не приводили к «дерганию» насоса?

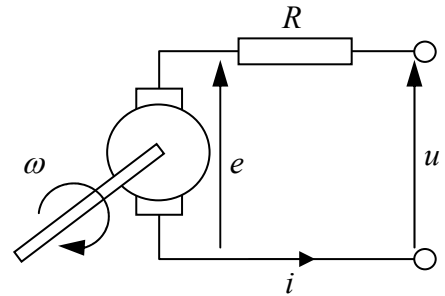
В следующих разделах представлены основы теории автоматического управления, которая отвечает на такие вопросы и предлагает надежные методы проектирования регуляторов, решающих задачу управления в соответствии с заданными требованиями.

3. Модели линейных объектов

3.1. Дифференциальные уравнения

Составляя модель объекта на основании физических законов, мы чаще всего получаем систему дифференциальных уравнений первого и второго порядка.

Для примера покажем, как построить модель *двигателя постоянного тока*, используя законы механики и электротехники. Вход этого объекта – напряжение якоря $u(t)$ (в *вольтах*), выход – угол поворота вала $\theta(t)$ (в *радианах*).



Сначала вспомним некоторые «житейские» знания об электродвигателях. Вал двигателя начинает вращаться, когда приложено напряжение питания. Если напряжение не меняется, угловая скорость вращения $\omega(t)$ (в *радианах в секунду*) остается постоянной, при этом угол $\theta(t)$ равномерно увеличивается.

Чем больше напряжение, тем быстрее вращается вал. Если зажать вал рукой (или подключить нагрузку, например, заставить двигатель вращать турбину), скорость вращения постепенно уменьшается до нового значения, при котором вращающий момент двигателя будет равен моменту сопротивления (нагрузки). Пока эти моменты равны, скорость вращения остается постоянной и ее производная равна нулю.

Теперь переведем эти рассуждения на строгий язык математики. Угловая скорость вращения $\omega(t)$ вычисляется как производная от угла поворота вала $\theta(t)$, то есть $\omega(t) = \frac{d\theta(t)}{dt}$. Соответственно, угол $\theta(t)$ – это интеграл от угловой скорости. В механике уравнение вращательного движения обычно записывают в виде

$$J \frac{d\omega(t)}{dt} = M(t) - M_H(t),$$

где $M(t)$ – вращающий момент (измеряется в $H \cdot m$), $M_H(t)$ – момент нагрузки (возмущение, также в $H \cdot m$). Буквой J обозначен суммарный *момент инерции* якоря и нагрузки (в $кг \cdot м^2$). Величина момента инерции говорит о том, насколько легко «разогнать» двигатель (чем больше момент инерции, тем сложнее «разогнать»).

Перейдем к электротехнике. В нашем случае момент $M(t)$ – это электромагнитный момент двигателя, который вычисляется по формуле

$$M(t) = C_M \cdot \Phi \cdot i(t),$$

где C_M – коэффициент, Φ – магнитный поток, создаваемый обмоткой возбуждения (измеряется в *веберах*); $i(t)$ – ток якоря (в *амперах*), который может быть найден из уравнения

$$u(t) = e(t) + R \cdot i(t),$$

где $e(t)$ – электродвижущая сила (ЭДС) якоря (в *вольтах*) и R – сопротивление якорной цепи (в *омах*). В свою очередь, ЭДС рассчитывается через магнитный поток и частоту вращения:

$$e(t) = C_\omega \cdot \Phi \cdot \omega(t),$$

где C_ω – коэффициент. Вводя новые постоянные $k_1 = C_M \cdot \Phi$ и $k_2 = C_\omega \cdot \Phi$, можно записать модель двигателя в виде системы уравнений

$$J \frac{d\omega(t)}{dt} = k_1 \cdot i(t) - M_H(t), \quad e(t) = k_2 \cdot \omega(t), \quad \omega(t) = \frac{d\theta(t)}{dt}, \quad u(t) = e(t) + R \cdot i(t). \quad (11)$$

Модель (11) описывает связи реальных сигналов в системе, ее внутреннее устройство.

Часто нам достаточно знать, как будет реагировать объект на заданный входной сигнал (управление). При этом его внутреннее устройство нас не очень интересует, то есть мы рас-

смаатриваем объект в качестве «черного ящика». Подставив второе уравнение из системы (11) в третье, найдем $i(t)$ и подставим в первое уравнение. Переходя к переменной $\theta(t)$, получаем:

$$J \frac{d^2 \theta(t)}{dt^2} = \frac{k_1}{R} \cdot \left[u(t) - k_2 \cdot \frac{d\theta(t)}{dt} \right] - M_H(t)$$

или, перенося все члены, зависящие от $\theta(t)$, в левую часть равенства

$$J \frac{d^2 \theta(t)}{dt^2} + \frac{k_1 k_2}{R} \cdot \frac{d\theta(t)}{dt} = k_2 \cdot u(t) - M_H(t). \quad (12)$$

Это дифференциальное уравнение второго порядка, связывающее вход $u(t)$ и нагрузку $M_H(t)$ с выходом $\theta(t)$. В сравнении с системой (11), все внутренние сигналы исходной модели ($e(t)$ и $i(t)$) были исключены из уравнений. Поэтому уравнение (12) называется **уравнением «вход-выход»**.

Порядком модели называют порядок соответствующего дифференциального уравнения. В данном случае мы получили модель второго порядка.

В этом разделе на простом примере мы посмотрели, как на основе физических законов строятся математические модели объектов управления. Как правило, они представляют собой дифференциальные уравнения. В дальнейшем мы будем использовать готовые модели объектов управления, предполагая, что они были кем-то получены ранее (например, предоставлены заказчиком).

3.2. Модели в пространстве состояний

Для того, чтобы было легче исследовать модель объекта, желательно привести ее к некоторому стандартному виду, для которого уже есть готовые общие решения. Таким «стандартом» в теории управления считается система дифференциальных уравнений первого порядка, которая называется *нормальной формой Коши*.

Рассмотрим снова модель электродвигателя, считая, что $M_H(t) = 0$ (нагрузки нет). Вспомнив, что $\omega(t) = \dot{\theta}(t)$, можно записать (12) в виде системы

$$\begin{aligned} \dot{\theta}(t) &= \omega(t) \\ \dot{\omega}(t) &= -\frac{k_1 k_2}{J R} \cdot \omega(t) + \frac{k_1}{J R} \cdot u(t) \end{aligned}$$

Эта система дифференциальных уравнений первого порядка может быть записана в матричной форме:

$$\begin{bmatrix} \dot{\theta}(t) \\ \dot{\omega}(t) \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 0 & -\frac{k_1 k_2}{J R} \end{bmatrix} \cdot \begin{bmatrix} \theta(t) \\ \omega(t) \end{bmatrix} + \begin{bmatrix} 0 \\ \frac{k_1}{J R} \end{bmatrix} \cdot u(t) \quad (13)$$

Значения $\theta(t)$ и $\omega(t)$ определяют *состояние* двигателя в момент времени t . Это значит, что зная их значения в некоторый момент времени t_0 и входной сигнал $u(t)$ при всех $t \geq t_0$ можно рассчитать поведение объекта для любого последующего момента. При этом предыдущие значения $\theta(t)$, $\omega(t)$ и $u(t)$ (при $t < t_0$) не играют никакой роли. Поэтому $\theta(t)$ и $\omega(t)$ называются *переменными состояния*, а вектор $\begin{bmatrix} \theta(t) \\ \omega(t) \end{bmatrix}$ – *вектором состояния*.

В теории управления принято обозначать вектор состояния через $x(t)$, вход объекта (сигнал управления) – через $u(t)$. Тогда модель (13) может быть записана в виде

$$\dot{x}(t) = A \cdot x(t) + B \cdot u(t) \quad (14)$$

где $x(t) = \begin{bmatrix} \theta(t) \\ \omega(t) \end{bmatrix}$, $A = \begin{bmatrix} 0 & 1 \\ 0 & -\frac{k_1 k_2}{J R} \end{bmatrix}$ и $B = \begin{bmatrix} 0 \\ \frac{k_1}{J R} \end{bmatrix}$. Модель (14) связывает вход $u(t)$ и вектор состояния $x(t)$, поэтому она называется *моделью вход-состояние*.

Полная модель объекта в пространстве состояний содержит еще одно уравнение – уравнение выхода, которое показывает, как формируется выход объекта $y(t)$:

$$\begin{aligned}\dot{x}(t) &= A \cdot x(t) + B \cdot u(t) \\ y(t) &= C \cdot x(t) + D \cdot u(t)\end{aligned}\tag{15}$$

Эта модель называется моделью *вход-состояние-выход*. Выходная координата для двигателя постоянного тока – это угол поворота вала:

$$y(t) = \theta(t) = \begin{bmatrix} 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} \theta(t) \\ \omega(t) \end{bmatrix} = \begin{bmatrix} 1 & 0 \end{bmatrix} \cdot x(t),$$

так что $C = \begin{bmatrix} 1 & 0 \end{bmatrix}$ и $D = 0$. Если же в качестве выхода принять угловую скорость, то $C = \begin{bmatrix} 0 & 1 \end{bmatrix}$.

С помощью модели (15), изменяя матрицы C и D , можно принять за выход любую линейную комбинацию переменных состояния и входа. Во многих практических задачах выход – это одна или несколько переменных состояния, которые мы можем измерить.

Поскольку момент инерции J , сопротивление якоря R и коэффициенты k_1 и k_2 не зависят от времени, матрицы A , B , C и D в модели (15) – постоянные. Такие объекты называются *стационарными*, в отличие от *нестационарных* объектов, параметры которых изменяются во времени.

Запись моделей в единой форме (15) позволяет отвлечься от смысла переменных состояния и исследовать системы разной природы стандартными методами, которые хорошо разработаны и реализованы в современных компьютерных программах.

Покажем, как уравнения вида (15) могут быть решены и чем удобна именно такая форма записи. Предположим, что мы знаем начальные условия, то есть вектор состояния $x(0)$ при $t = 0$. Вспомним, что знание $x(0)$ и входа $u(t)$ при всех $t > 0$ дает возможность однозначно определить дальнейшее поведение этого объекта.

Первое уравнение в (15) позволяет найти производную, то есть, скорость изменения вектора состояния $x(t)$ в любой момент времени. Будем считать, что при $0 \leq t \leq \Delta t$, где Δt – малый интервал времени, эта производная не меняется. Тогда значение вектора состояния при $t = \Delta t$ приближенно определяется формулой

$$x(\Delta t) \approx x(0) + \dot{x}(0) \cdot \Delta t = x(0) + [A \cdot x(0) + B \cdot u(0)] \cdot \Delta t,$$

то есть, его можно легко вычислить. Зная $x(\Delta t)$ и сигнал управления $u(\Delta t)$, находим выход системы в тот же момент

$$y(\Delta t) \approx C \cdot x(\Delta t) + D \cdot u(\Delta t).$$

Эту методику можно применять и дальше, в конце второго интервала получаем

$$\begin{aligned}x(2 \cdot \Delta t) &\approx x(\Delta t) + \dot{x}(\Delta t) \cdot \Delta t = x(\Delta t) + [A \cdot x(\Delta t) + B \cdot u(\Delta t)] \cdot \Delta t, \\ y(2 \cdot \Delta t) &\approx C \cdot x(2 \cdot \Delta t) + D \cdot u(2 \cdot \Delta t).\end{aligned}$$

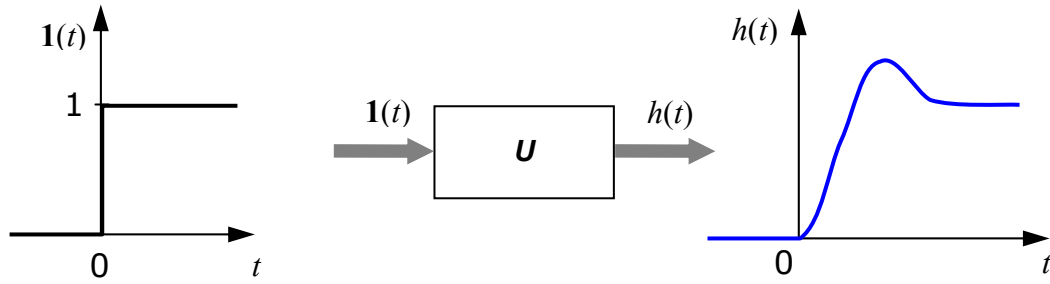
Таким образом, можно (приближенно) рассчитать выход системы при всех $t > 0$. Конечно, точность будет тем выше, чем меньше Δt , однако объем вычислений при этом также увеличится. Этот метод приближенного решения дифференциальных уравнения называется *методом Эйлера*. Так как мы не делали никаких предположений о постоянных матрицах A , B , C и D , его (как и другие, более совершенные методы) можно использовать без изменений для решения любых уравнений вида (15).

3.3. Переходная функция

Один из методов построения моделей «вход-выход» – определение реакции объекта на некоторый стандартный сигнал. Один из простейших сигналов – так называемый «единичный скачок» («единичный ступенчатый сигнал»), то есть мгновенное изменение входного сигнала с 0 до 1 в момент $t = 0$. Формально этот сигнал определяется так:

$$1(t) = \begin{cases} 0, & t < 0 \\ 1, & t \geq 0 \end{cases}$$

Реакция объекта на единичный скачок называется **переходной функцией** и обозначается $h(t)$:



При этом предполагается, что объект в начальный момент находится в состоянии покоя, то есть, имеет *нулевые начальные условия*. Это значит, что все его переменные состояния равны нулю и внутренняя энергия также нулевая.

Если начальные условия ненулевые, то для построения сигнала выхода при любом входе нужно использовать дифференциальные уравнения объекта или модель в пространстве состояний. Это значит, что переходная характеристика дает меньше информации, чем исходные уравнения.

Пусть модель объекта задана дифференциальным уравнением первого порядка:

$$T \frac{dy(t)}{dt} + y(t) = k \cdot x(t), \quad (16)$$

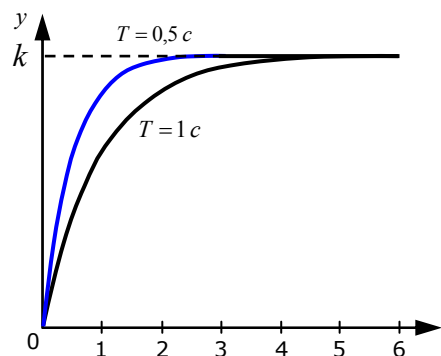
где k – безразмерный коэффициент, а T – некоторая постоянная, которая имеет размерность времени (измеряется в *секундах*). Найдем переходную характеристику этого звена. Решая уравнение (16) при $x(t) = 1$ ($t > 0$), получаем

$$y(t) = k + C_1 \cdot \exp\left(-\frac{t}{T}\right),$$

где постоянная C_1 должна определяться из начальных условий. Поскольку нас интересует переходная характеристика, начальные условия считаем нулевыми, то есть $y(0) = 0$, что дает $C_1 = -k$ и поэтому

$$h(t) = y(t) = k \left[1 - \exp\left(-\frac{t}{T}\right) \right]. \quad (17)$$

На рисунке показаны переходные характеристики (17) при различных значениях параметра T , который называется **постоянной времени** звена:

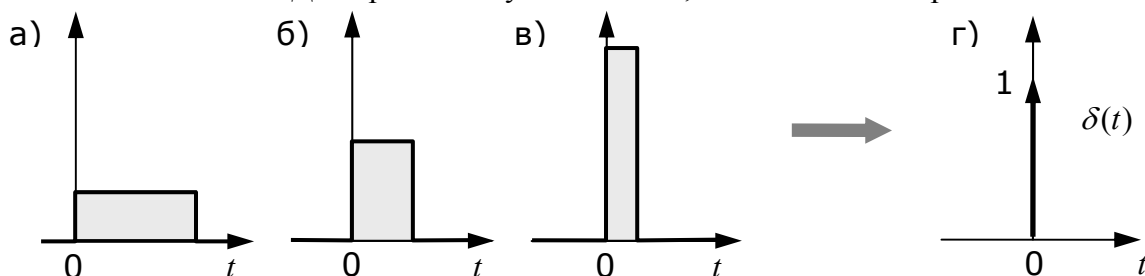


Видно, что при увеличении T выход y медленнее достигает установившегося значения, равного k , то есть постоянная времени характеризует *инерционность* звена (16). Чем больше постоянная времени, тем медленнее реагирует объект на управление и тем больше усилий нужно для того, чтобы перевести его в новое состояние.

Заметим, что ступенчатый сигнал легко получить на практике, поэтому переходную характеристику можно снять экспериментально.

3.4. Импульсная характеристика (весовая функция)

В качестве тестового сигнала можно, в принципе, использовать любой сигнал. Например, можно изучать реакцию системы на прямоугольный импульс. Вопрос в том, чтобы определить некоторый стандартный вид этого импульса. На рисунках а)-в) показаны три импульса, имеющих одинаковые площади. Для простоты будем считать, что эта площадь равна единице.



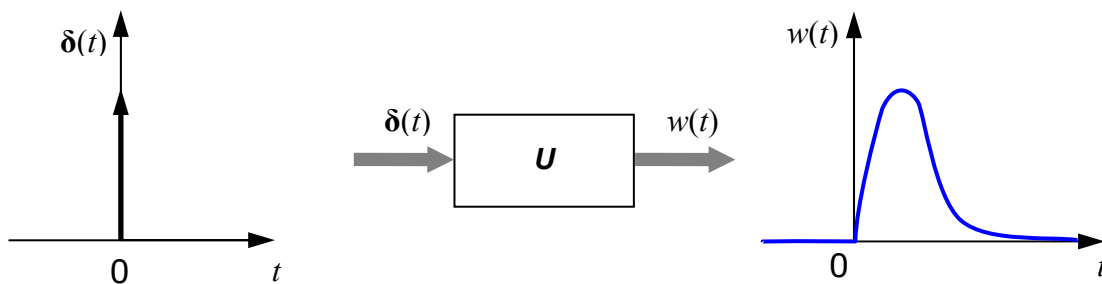
Что будет, если мы будем уменьшать ширину импульса, сохраняя его площадь? Очевидно, что высота импульса будет расти и в пределе (когда ширина стремится к нулю) станет бесконечной. Таким образом, мы получили еще один классический тестовый сигнал – *единичный импульс* или *дельта-функцию Дирака* $\delta(t)$. Это идеальный (невозможный в реальной жизни) сигнал, который равен нулю во всех точках, кроме $t = 0$, где он уходит к бесконечности, причем его площадь (интеграл по всей оси времени) равен единице:

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

Поскольку бесконечный импульс невозможно нарисовать, на графике он изображается стрелкой, высота которой равна единице (см. рисунок г).

Иногда определяют дельта-функцию как производную от единичного ступенчатого сигнала $1(t)$. Действительно, эта производная равна нулю при всех значениях t , кроме нуля, где она обращается в бесконечность.

Реакция системы на единичный импульс (дельта-функцию) называется **импульсной характеристикой** и обозначается $w(t)$:



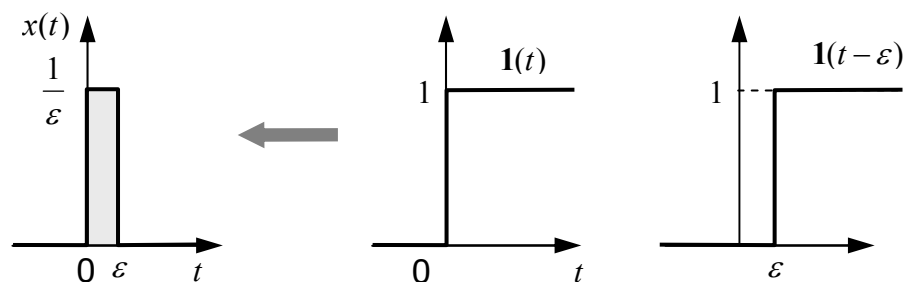
Импульсная характеристика, так же, как и переходная характеристика, определяется при нулевых начальных условиях, то есть, объект должен находиться в состоянии покоя.

Рассматривая дельта-функцию как предельный случай прямоугольного сигнала единичной площади, можно найти связь между переходной функцией и импульсной характеристикой.

Пусть ширина прямоугольного импульса равна ε , а высота – $1/\varepsilon$. Такой импульс можно представить в виде разности двух ступенчатых сигналов

$$x(t) = \frac{1}{\varepsilon} [1(t) - 1(t - \varepsilon)],$$

где $1(t - \varepsilon)$ – это единичный ступенчатый сигнал, который приходит в момент $t = \varepsilon$, то есть, смещен по времени на ε (см. рисунок далее).



Так как для линейных систем справедлив принцип суперпозиции, сигнал на выходе будет равен разности реакций системы на входы $1(t)$ и $1(t-\varepsilon)$, умноженной на коэффициент $1/\varepsilon$. Учитывая, что реакция на сигнал $1(t)$ – это переходная функция $h(t)$, получаем

$$y(t) = \frac{1}{\varepsilon} [h(t) - h(t-\varepsilon)].$$

Переходя к пределу при $\varepsilon \rightarrow 0$, наодим, что импульсная характеристика

$$w(t) = \lim_{\varepsilon \rightarrow 0} \frac{h(t) - h(t-\varepsilon)}{\varepsilon} = \frac{dh(t)}{dt},$$

как оказывается, равна производной от переходной функции. Наоборот, переходная функция – это интеграл от импульсной характеристики на интервале от 0 до t :

$$h(t) = \int_0^t w(\tau) d\tau.$$

Дифференцируя переходную характеристику (17) звена первого порядка, получаем соответствующую импульсную характеристику:

$$w(t) = \frac{d}{dt} \left(k \left[1 - \exp\left(-\frac{t}{T}\right) \right] \right) = \frac{k}{T} \exp\left(-\frac{t}{T}\right).$$

Другое название импульсной характеристики – *весовая функция*. Это название связано с тем, что для произвольного входного сигнала $x(t)$ выход системы $y(t)$ при нулевых начальных условиях вычисляется как интеграл

$$y(t) = \int_{-\infty}^t x(\tau) w(t-\tau) d\tau = \int_0^{\infty} x(t-\tau) w(\tau) d\tau.$$

Здесь функция $w(t)$ как бы «взвешивает» входной сигнал $x(t)$ в подынтегральном выражении. Заметим, что импульсная характеристика дает неполную информацию об объекте, поскольку не учитывает ненулевые начальные условия.

В отличие от ступенчатого сигнала, мгновенный импульс бесконечной величины невозможно получить на реальном устройстве, поэтому снять импульсную характеристику системы, строго говоря, экспериментально не удастся.

3.5. Передаточная функция

Вы уже знаете, выходной сигнал системы можно представить как результат действия некоторого оператора на ее вход. Для линейных моделей такой оператор можно записать следующим образом.

Пусть модель объекта задана линейным дифференциальным уравнением второго порядка, связывающим вход $x(t)$ и выход $y(t)$:

$$b_2 \frac{d^2 y(t)}{dt^2} + b_1 \frac{dy(t)}{dt} + b_0 y(t) = a_1 \frac{dx(t)}{dt} + a_0 x(t) \quad (18)$$

где a_i ($i = 0,1$) и b_i ($i = 0,1,2$) – постоянные.

Введем *оператор дифференцирования* $p = \frac{d}{dt}$, который действует на сигнал $x(t)$ по правилу $p x(t) = \frac{dx(t)}{dt}$. Обратите внимание, что запись $p x(t)$ обозначает *не умножение* оператора p на $x(t)$, а *действие* этого оператора, то есть дифференцирование $x(t)$.

Теперь запишем производные сигналов $x(t)$ и $y(t)$ по времени в операторной форме

$$\dot{y}(t) = \frac{dy(t)}{dt} = py(t), \quad \ddot{y}(t) = \frac{d^2 y(t)}{dt^2} = p^2 y(t), \quad \dot{x}(t) = \frac{dx(t)}{dt} = px(t).$$

Подставляя эти выражения в (18), получим

$$b_2 p^2 y(t) + b_1 p y(t) + b_0 y(t) = a_1 p x(t) + a_0 x(t). \quad (19)$$

Можно формально вынести за скобки $y(t)$ в левой части равенства (19) и $x(t)$ в правой части:

$$(b_2 p^2 + b_1 p + b_0) y(t) = (a_1 p + a_0) x(t). \quad (20)$$

Левая часть (20) означает, что оператор $b_2 p^2 + b_1 p + b_0$ действует на сигнал $y(t)$, а в правой части оператор $a_1 p + a_0$ действует на сигнал $x(t)$. «Разделив» (условно, конечно) обе части (20) на оператор $b_2 p^2 + b_1 p + b_0$, связь выхода и входа можно записать в виде

$$y(t) = \frac{a_1 p + a_0}{b_2 p^2 + b_1 p + b_0} x(t) = W(p) x(t), \quad (21)$$

где запись $W(p) x(t)$ означает не умножение, а действие сложного оператора

$$W(p) = \frac{a_1 p + a_0}{b_2 p^2 + b_1 p + b_0}. \quad (22)$$

на сигнал $x(t)$. Иначе говоря, формула $y(t) = W(p) x(t)$ – это не что иное, как символическая запись уравнения (18), которую удобно использовать.

Функция $W(p)$ называется **передаточной функцией** объекта, который описывается уравнением (18). Она полностью описывает связи между выходом и входом объекта при нулевых начальных условиях, но не учитывает его внутреннее устройство.

Часто передаточной функцией называют функцию $W(\lambda)$, которая получается из (22) в результате замены оператора p на некоторую независимую переменную λ . Эта функция представляет собой отношение двух полиномов (многочленов) от λ .

Передаточная функция $W(\lambda)$ называется **правильной**, если степень ее числителя *не больше*, чем степень знаменателя; **строго правильной**, если степень числителя *меньше* степени знаменателя; **неправильной**, если степень числителя *больше*, чем степень знаменателя. Например, функция $\frac{1}{\lambda+1}$ – строго правильная и одновременно правильная; $\frac{\lambda}{\lambda+1}$ – правильная, но не строго правильная (иногда такие функции называют *биправильными*), а $\frac{\lambda^2 + \lambda + 1}{\lambda + 1}$ – неправильная.

Нулями передаточной функции называются корни ее числителя, а **полюсами** – корни знаменателя. Например, функция $W(\lambda) = \frac{\lambda - 1}{\lambda^2 + 3\lambda + 2}$ имеет нуль в точке $\lambda = 1$ и два полюса в точках $\lambda = -1$ и $\lambda = -2$.

3.6. Преобразование Лапласа

3.6.1. Что такое преобразование Лапласа?

Одна из первых задач, которые были поставлены в теории управления – вычисление выхода системы при известном входе. Мы видели, что для ее решения нужно решать дифференци-

альные уравнения. Чтобы упростить процедуру, математики придумали преобразование, которое позволило заменить решение дифференциальных уравнений алгебраическими вычислениями, то есть, операциями с полиномами (многочленами) и рациональными функциями.

Для функции $f(t)$ вводится **преобразование Лапласа**, которое обозначается как $\mathcal{L}\{f(t)\}$:

$$F(s) = \mathcal{L}\{f(t)\} = \int_0^{\infty} f(t) e^{-st} dt. \quad (23)$$

Функция $F(s)$ называется *изображением* для функции $f(t)$ (оригинала). Здесь s – это комплексная переменная, которая выбирается так, чтобы интеграл (23) сходилась³.

Обратное преобразование Лапласа $\mathcal{L}^{-1}\{F(s)\}$ позволяет вычислить оригинал $f(t)$ по известному изображению $F(s)$:

$$f(t) = \mathcal{L}^{-1}\{F(s)\} = \frac{1}{2\pi j} \int_{\sigma-j\infty}^{\sigma+j\infty} F(s) e^{st} ds, \quad (24)$$

где $j = \sqrt{-1}$, а постоянная σ выбирается так, чтобы интеграл сходилась⁴.

На практике вместо интеграла (24) чаще всего используют готовые таблицы, по которым можно сразу определить изображение по оригиналу и наоборот. Например, изображения по Лапласу для дельта-функции, единичного скачка и функции e^{-at} равны, соответственно

$$\mathcal{L}\{\delta(t)\} = 1, \quad \mathcal{L}\{1(t)\} = \frac{1}{s}, \quad \mathcal{L}\{e^{-at}\} = \frac{1}{s+a}. \quad (25)$$

3.6.2. Свойства преобразования Лапласа

Преобразование Лапласа имеет несколько замечательных свойств. Во-первых, используя (23) и (24), легко доказать, что принцип суперпозиции выполняется как для прямого, так и для обратного преобразования Лапласа:

$$\mathcal{L}\{f_1(t) + f_2(t)\} = \mathcal{L}\{f_1(t)\} + \mathcal{L}\{f_2(t)\}, \quad (26)$$

$$\mathcal{L}^{-1}\{F_1(s) + F_2(s)\} = \mathcal{L}^{-1}\{F_1(s)\} + \mathcal{L}^{-1}\{F_2(s)\}. \quad (27)$$

Во-вторых, изображение для производной функции $f(t)$ равно

$$\mathcal{L}\left\{\frac{df(t)}{dt}\right\} = s \cdot F(s) - f(0),$$

где $F(s)$ – изображение функции $f(t)$, и $f(0)$ – ее значение⁵ при $t = 0$. Поэтому *при нулевых начальных условиях* изображение производной равно изображению самой функции, умноженному на s . Аналогично для построения изображения i -ой производной нужно умножить изображение функции на s^i (это также справедливо только при нулевых начальных условиях).

Кроме того, с помощью преобразование Лапласа можно сразу найти **начальное и конечное значения** функции-оригинала (при $t = 0$ и $t \rightarrow \infty$), не вычисляя самого оригинала:

$$f(0) = \lim_{s \rightarrow \infty} s \cdot F(s), \quad f(\infty) = \lim_{s \rightarrow 0} s \cdot F(s). \quad (28)$$

³ Преобразование Лапласа определяется для функций ограниченного роста, таких что $f(t) < Me^{\alpha t}$, где M и α – постоянные, и α называется *показателем роста* функции $f(t)$. Для всех s , вещественная часть которых больше α (в области $\operatorname{Re} s > \alpha$) функция $f(t)e^{-st}$ затухает при $t \rightarrow \infty$ и интеграл (23) сходится.

⁴ Постоянная σ должна быть больше, чем показатель роста α функции-оригинала $f(t)$. При этом можно показать, что значение интеграла (24) не зависит от выбора σ .

⁵ Если функция имеет разрыв при $t = 0$, нужно брать предел слева, то есть ее значение при бесконечно малом отрицательном t .

3.6.3. Снова передаточная функция

Рассмотрим снова уравнение (18):

$$b_2 \frac{d^2 y(t)}{dt^2} + b_1 \frac{dy(t)}{dt} + b_0 y(t) = a_1 \frac{dx(t)}{dt} + a_0 x(t) \quad (29)$$

Применим к левой и правой частям преобразование Лапласа, считая, что все начальные условия нулевые. Получается *уравнение в изображениях*, связывающее преобразования Лапласа входа $X(s)$ и выхода $Y(s)$:

$$b_2 \cdot s^2 Y(s) + b_1 \cdot s Y(s) + b_0 \cdot Y(s) = a_1 \cdot s X(s) + a_0 \cdot X(s)$$

Можно вынести за скобки $Y(s)$ в левой части и $X(s)$ в правой части:

$$(b_2 s^2 + b_1 s + b_0) \cdot Y(s) = (a_1 s + a_0) \cdot X(s).$$

Разделив обе части этого равенства на $b_2 s^2 + b_1 s + b_0$, получаем

$$Y(s) = \frac{a_1 s + a_0}{b_2 s^2 + b_1 s + b_0} \cdot X(s) = W(s) \cdot X(s), \quad \text{где } W(s) = \frac{a_1 s + a_0}{b_2 s^2 + b_1 s + b_0}. \quad (30)$$

Сравнение (22) и (30) показывает, что $W(s)$ – это *передаточная функция* объекта, записанная в виде функции от комплексной переменной s , а не от оператора дифференцирования p , как в (22).

Таким образом, при нулевых начальных условиях *изображение выхода линейного объекта вычисляется как произведение его передаточной функции на изображение входного сигнала*.

Из (30) следует и другой важный вывод: *передаточная функция равна отношению изображений по Лапласу выхода и входа при нулевых начальных условиях*.

3.6.4. Пример

Рассмотрим пример использования преобразования Лапласа для вычисления выхода системы при известном входном сигнале. Пусть объект управления описывается уравнением первого порядка (16):

$$T \frac{dy(t)}{dt} + y(t) = k \cdot x(t) \quad (31)$$

и на его вход поступает единичный ступенчатый сигнал $x(t) = 1(t)$. Требуется найти сигнал выхода $y(t)$, который в данном случае представляет собой переходную характеристику.

Решим эту задачу с помощью передаточных функций и изображений сигналов по Лапласу. Чтобы найти изображение выхода по формуле (30), нужно знать изображение входного сигнала $X(s)$ и передаточную функцию звена $W(s)$. Изображение входа находим по табличным данным (см. (25)), а передаточную функцию – из (31), повторяя приведенные выше рассуждения:

$$X(s) = \frac{1}{s}, \quad W(s) = \frac{k}{Ts + 1}.$$

Теперь находим изображение выхода

$$Y(s) = \frac{1}{s} \cdot \frac{k}{Ts + 1} = \frac{k}{s} - \frac{kT}{Ts + 1}.$$

и представляем его в виде суммы элементарных дробей:

$$Y(s) = \frac{k}{s} - \frac{k}{s + 1/T}.$$

Используя принцип суперпозиции для изображений (27), вычисляем оригинал – сигнал выхода:

$$y(t) = k \cdot \mathcal{L}^{-1} \left\{ \frac{1}{s} \right\} - k \cdot \mathcal{L}^{-1} \left\{ \frac{1}{s + 1/T} \right\}.$$

Обратные преобразования Лапласа находим по таблице (25):

$$y(t) = k - k \cdot \exp\left(-\frac{t}{T}\right) \quad \text{при } t > 0,$$

что совпадает с (17). Таким способом можно вычислять реакцию системы на известный входной сигнал без прямого решения дифференциального уравнения.

Применяя формулы (28) для вычисления начального и конечного значений сигнала выхода $y(t)$:

$$y(0) = \lim_{s \rightarrow \infty} s \cdot Y(s), \quad y(\infty) = \lim_{s \rightarrow 0} s \cdot Y(s).$$

При ступенчатом входном сигнале с изображением $X(s) = \frac{1}{s}$ получаем

$$y(0) = \lim_{s \rightarrow \infty} W(s), \quad y(\infty) = W(0).$$

Таким образом, для рассмотренного выше примера

$$y(0) = \lim_{s \rightarrow \infty} \frac{k}{Ts + 1} = 0, \quad y(\infty) = W(0) = k.$$

Значение $W(0)$ называют **статическим коэффициентом усиления** звена, поскольку он показывает, во сколько раз усиливается постоянный сигнал.

3.7. Передаточная функция и пространство состояний

Используя преобразование Лапласа, можно построить передаточную функцию для модели объекта в пространстве состояний

$$\dot{x}(t) = A \cdot x(t) + B \cdot u(t)$$

$$y(t) = C \cdot x(t) + D \cdot u(t)$$

Напомним, что здесь $u(t)$, $y(t)$ и $x(t)$ обозначают соответственно вход, выход и вектор состояния объекта. Преобразуя левые и правые части каждого уравнения по Лапласу (переходя к изображениям сигналов по Лапласу при нулевых начальных условиях), получаем

$$s \cdot X(s) = A \cdot X(s) + B \cdot U(s)$$

$$Y(s) = C \cdot X(s) + D \cdot U(s) \quad (32)$$

В первом уравнении перенесем все члены, зависящие от $X(s)$, в левую часть:

$$(s \cdot I - A) \cdot X(s) = B \cdot U(s),$$

где I обозначает *единичную матрицу*, у которой на главной диагонали стоят единицы, а все остальные элементы – нули. Умножая обе части последнего равенства на $(s \cdot I - A)^{-1}$, получим выражение для $X(s)$:

$$X(s) = (s \cdot I - A)^{-1} \cdot B \cdot U(s)$$

которое при подстановке во второе уравнение в (32) дает

$$Y(s) = C \cdot (s \cdot I - A)^{-1} \cdot B \cdot U(s) + D \cdot U(s) = [C \cdot (s \cdot I - A)^{-1} \cdot B + D] \cdot U(s).$$

Чтобы определить передаточную функцию, найдем отношение изображений выхода и входа:

$$W(s) = \frac{Y(s)}{U(s)} = C \cdot (s \cdot I - A)^{-1} \cdot B + D. \quad (33)$$

Обратный переход, от передаточной функции к модели в пространстве состояний, более сложен и неоднозначен. Дело в том, что каждой передаточной функции соответствует бесчисленное множество моделей в пространстве состояний. Одну из них можно найти следующим образом. Для передаточной функции

$$W(s) = d + \frac{a_2 s^2 + a_1 s + a_0}{s^3 + b_2 s^2 + b_1 s + b_0},$$

где d , a_i ($i = 0, 1, 2$) и b_i ($i = 0, 1, 2$) – постоянные коэффициенты, модель в пространстве состояний задается матрицами

$$A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -b_0 & -b_1 & -b_2 \end{bmatrix}, \quad B = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, \quad C = [a_0 \quad a_1 \quad a_2], \quad D = d. \quad (34)$$

При увеличении порядка передаточной функции (степени ее знаменателя), эти матрицы расширяются. В нижней строке матрицы A записываются коэффициенты знаменателя с обратным знаком, над главной диагональю – единицы, а остальные элементы – нули. В матрице B только самый последний элемент – единица, а остальные – нули. Наконец, матрица C строится из коэффициентов числителя передаточной функции.

Отметим, что модель, заданную *неправильной* передаточной функцией (у которой степень числителя *больше* степени знаменателя) нельзя представить в пространстве состояний.

Рассмотрим простой объект, модель которого задана в пространстве состояний матрицами

$$A = \begin{bmatrix} 3 & -0,5 \\ 4 & 0 \end{bmatrix}, \quad B = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad C = [1 \quad 0,25], \quad D = 0.$$

Используя формулу (33), получаем

$$W(s) = C(sI - A)B + D = [1 \quad 0,25] \left(\begin{bmatrix} s & 0 \\ 0 & s \end{bmatrix} - \begin{bmatrix} 3 & -0,5 \\ 4 & 0 \end{bmatrix} \right)^{-1} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \frac{s+1}{s^2+3s+2}.$$

Теперь выполним обратный переход. По формулам (34) сразу находим матрицы

$$\tilde{A} = \begin{bmatrix} 0 & 1 \\ -2 & -3 \end{bmatrix}, \quad \tilde{B} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \quad \tilde{C} = [1 \quad 1], \quad \tilde{D} = 0.$$

Заметим, что эти матрицы отличаются от исходных, однако если найти передаточную функцию по формулам (33), мы получим тот же самый результат. Это говорит о том, что одной и той же передаточной функции могут соответствовать разные модели в пространстве состояний. Если известна одна такая модель с матрицами A, B, C и D , то все остальные модели могут быть получены по формулам

$$\tilde{A} = PAP^{-1}, \quad \tilde{B} = PB, \quad \tilde{C} = CP^{-1}, \quad \tilde{D} = D,$$

где P – некоторая обратимая матрица (ее определитель должен быть ненулевым). При таком преобразовании передаточная функция не меняется (проверьте это!). Фактически мы переходим к другому вектору состояния $x'(t)$, который связан с исходным зависимостью $x'(t) = P \cdot x(t)$. Легко проверить, что в данном случае нужное преобразование выполняет матрица

$$P = \begin{bmatrix} 0 & 0,25 \\ 1 & 0 \end{bmatrix}.$$

Внимательно посмотрев на функцию $W(s)$, можно заметить, что ее числитель и знаменатель имеют одинаковый множитель $s+1$, который можно сократить. Таким образом,

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

Для этой передаточной функции модель в пространстве состояний выглядит так:

$$A = -2, \quad B = 1, \quad C = 1, \quad D = 0. \quad (35)$$

Вместо исходной модели второго порядка (два уравнения, две переменные состояния) мы получили модель первого порядка. Что же произошло? Оказалось, что *при нулевых начальных условиях* состояние объекта определяется одной переменной, а зависимость между входом и выходом системы – одним уравнением первого порядка. Поэтому произошло сокращение числителя и знаменателя передаточной функции.

Если нас интересуют только связь входа и выхода (а не внутренние сигналы в объекте) и начальные условия нулевые, можно использовать модель первого порядка. Однако при ненулевых начальных условиях нужно использовать исходную модель в пространстве состояний, потому что передаточная функция дает неполную информацию. Это особенно важно при анализе устойчивости системы (см. разд. 6.4).

3.8. Частотные характеристики

Еще один популярный эталонный сигнал – гармонический (синус, косинус), например:

$$x(t) = \sin \omega t, \quad (36)$$

где ω – угловая частота (в радианах в секунду). Можно показать, что при таком входе на выходе линейной системы в установившемся режиме (при больших t) будет синус той же частоты⁶, но с другой амплитудой A и сдвигом фазы ϕ :

$$y(t) = A(\omega) \cdot \sin(\omega t + \phi(\omega)).$$

Для каждой частоты входного сигнала будет своя амплитуда и свой сдвиг фазы. Чтобы определить по графику фазовый сдвиг ϕ , нужно найти расстояние Δt по оси времени между соответствующими точками синусоид (например, точками пересечения с осью t или вершинами). Если Δt умножить на частоту ω , получаем сдвиг фазы ϕ (в радианах).

На рисунке показан случай $\phi > 0$ (опережение по фазе), когда выход сдвинут «влево» по оси времени относительно входа, то есть, «идет раньше» входного.

Зная передаточную функцию системы $W(s)$, можно вычислить амплитуду и сдвиг фазы по формулам

$$A(\omega) = |W(j\omega)|, \quad \phi(\omega) = \arg W(j\omega) = \operatorname{arctg} \frac{\operatorname{Im} W(j\omega)}{\operatorname{Re} W(j\omega)}.$$

Запись $W(j\omega)$ означает, что в передаточную функцию $W(s)$ подставляется чисто мнимое число $s = j\omega$, где $j = \sqrt{-1}$. Для каждой частоты ω значение $W(j\omega) = P + jQ$ – это некоторое комплексное число, имеющее амплитуду $|W(j\omega)| = \sqrt{P^2 + Q^2}$ и фазу $\arg W(j\omega) = \operatorname{arctg} \frac{Q}{P}$.

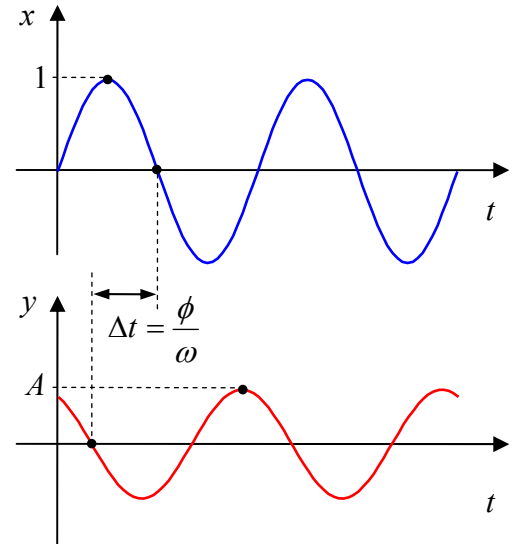
Функция $W(j\omega)$ называется **частотной характеристикой** звена, поскольку она характеризует выход системы при гармонических сигналах разной частоты. Зависимости $P(\omega)$ и $Q(\omega)$ (вещественная и мнимая части $W(j\omega)$) – это *вещественная и мнимая частотные характеристики*.

Функции $A(\omega)$ и $\phi(\omega)$ (они для каждой частоты принимают вещественные значения) называются соответственно **амплитудной** и **фазовой частотными характеристиками** (АЧХ и ФЧХ). Амплитудная частотная характеристика – это коэффициент усиления гармонического сигнала. Если на какой-то частоте ω значение $A(\omega) > 1$, входной сигнал усиливается, если $A(\omega) < 1$, то вход данной частоты ослабляется.

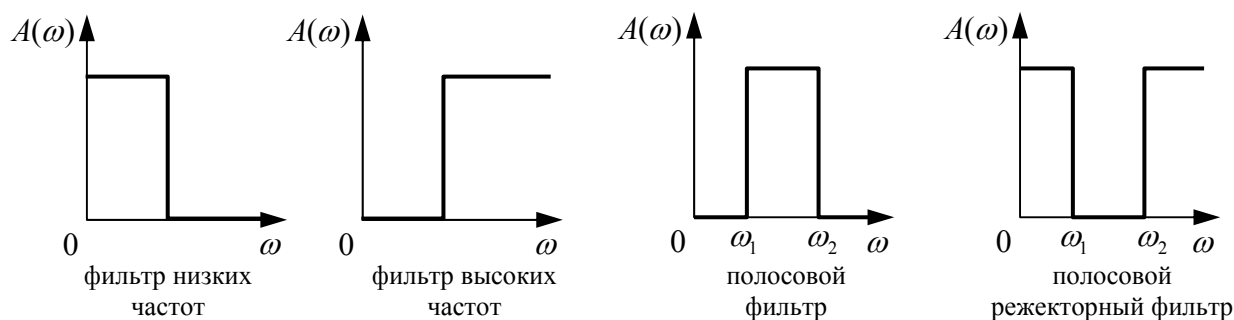
По форме АЧХ различают несколько основных типов звеньев:

- 1) *фильтр низких частот* – пропускает низкочастотные сигналы примерно с одинаковым коэффициентом усиления, блокирует высокочастотные шумы и помехи;
- 2) *фильтр высоких частот* – пропускает высокочастотные сигналы, блокирует сигналы низкой частоты;
- 3) *полосовой фильтр* – пропускает только сигналы с частотами в полосе от ω_1 до ω_2 ;
- 4) *полосовой режсекторный фильтр* – блокирует только сигналы с частотами в полосе от ω_1 до ω_2 , остальные пропускает.

На рисунке показаны амплитудные частотные характеристики идеальных фильтров этих четырех типов:

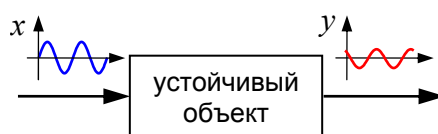


⁶ Здесь, конечно, предполагается, что при синусоидальном входном сигнале система не «идет вразнос», то есть, ее выходной сигнал не растет неограниченно (система является *устойчивой*).

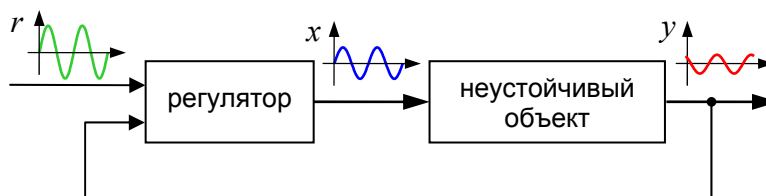


В радиотехнике используется понятие *полосы пропускания* – это ширина полосы частот, в которой значение АЧХ больше, чем $1/\sqrt{2}$ от ее максимального значения.

Частотные характеристики во многих случаях можно снять экспериментально. Если объект *устойчивый*, на его вход подается гармонический сигнал (36) и записывается сигнал $y(t)$ на выходе. Определив амплитуду и сдвиг фазы для разных частот, можно построить по точкам амплитудную и фазовую частотные характеристики.



Если объект *неустойчив*, то при подаче на вход синуса амплитуда колебаний на выходе будет неограниченно расти. Однако частотную характеристику все равно можно определить экспериментально. Для этого нужно сначала найти какой-нибудь регулятор, который сделает замкнутую систему устойчивой. Затем на вход $r(t)$ подают синусоидальный сигнал и сравнивают сигналы $x(t)$ и $y(t)$ на входе и выходе интересующего нас объекта, определяя для каждой частоты ω «коэффициент усиления» $A(\omega)$ (отношение амплитуд сигналов $x(t)$ и $y(t)$) и сдвиг фазы $\phi(\omega)$.



3.9. Логарифмические частотные характеристики

Частотные характеристики достаточно сложно строить вручную. В 60-е годы, когда развивалась классическая теория управления, не было мощных компьютеров, поэтому наибольшую популярность приобрели приближенные методы, с помощью которых можно было проектировать регуляторы с помощью ручных вычислений и построений. Один из таких подходов основан на использовании **логарифмических частотных характеристик**.

Вместо $A(\omega)$ было предложено использовать *логарифмическую амплитудную частотную характеристику* (ЛАЧХ): график, на котором по оси абсцисс откладывается десятичный логарифм частоты ($\lg \omega$), а по оси ординат – величина $L_m(\omega) = 20 \lg A(\omega)$, измеряемая в децибелах (дБ). При построении логарифмической фазовой частотной характеристики (ЛФЧХ) по оси абсцисс также откладывается логарифм частоты $\lg \omega$.

Единицей отсчета на логарифмической оси частот является *декада* – диапазон, на котором частота увеличивается в 10 раз (а значение ее логарифма увеличивается на единицу). Вместе ЛАЧХ и ЛФЧХ называются логарифмической амплитудно-фазовой частотной характеристикой (ЛАФЧХ) или *диаграммой Боде*.

Логарифмические характеристики обладают двумя ценными свойствами:

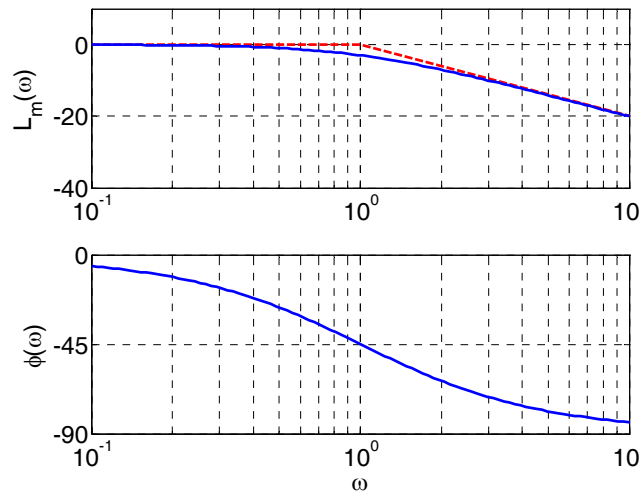
- 1) ЛАЧХ и ЛФЧХ для произведения $W_1(s)W_2(s)$ вычисляются как суммы ЛАЧХ и ЛФЧХ отдельных звеньев:

$$20 \lg A(\omega) = 20 \lg A_1(\omega) + 20 \lg A_2(\omega); \quad (37)$$

$$\phi(\omega) = \phi_1(\omega) + \phi_2(\omega); \quad (38)$$

- 2) в области высоких и низких частот ЛАЧХ асимптотически приближаются к прямым, наклон которых составляет ± 20 дБ/дек (децибел на декаду), ± 40 дБ/дек и т.д.

В классической теории управления хорошо разработаны методы анализа и синтеза систем на основе *асимптотических ЛАЧХ*, которые представляют собой ломанные линии и легко строятся вручную. С появлением компьютерных средств расчета практическая ценность ЛАФЧХ несколько снизилась, однако они по сей день остаются простейшим инструментом прикладных расчетов для инженера.



На рисунке показаны точная (сплошная синяя линия) и асимптотическая (штриховая красная линия) ЛАФЧХ для звена первого порядка с передаточной функцией

$$W(s) = \frac{1}{Ts + 1} \text{ при } T = 1 \text{ с.}$$

Первая асимптота, определяющая поведение ЛАЧХ **на низких частотах**, имеет нулевой наклон, потому что звено относится к классу *позиционных* звеньев, имеющих постоянный ненулевой статический коэффициент усиления, то есть

$$W(0) = 1 \neq 0.$$

Если $W(0) = 0$, передаточная функция содержит множитель s^k ($k > 0$), который соответствует производной порядка k . В этом случае наклон ЛАЧХ на низких частотах равен $k \cdot 20$ дБ/дек.

Если $W(0) = \infty$, звено содержит один или несколько *интеграторов*, то есть в знаменателе есть сомножитель s^k . Тогда наклон ЛАЧХ на низких частотах равен $-k \cdot 20$ дБ/дек.

Наклон ЛАЧХ **на высоких частотах** определяется разностью степеней числителя и знаменателя передаточной функции. Если числитель имеет степень m , а знаменатель – степень n , то наклон последней асимптоты равен $20 \cdot (m - n)$ дБ/дек. В нашем примере $m - n = 0 - 1 = -1$. Поэтому вторая асимптота, определяющая свойства звена на высоких частотах, имеет наклон -20 дБ/дек, то есть, за одну декаду значение уменьшается на 20 дБ (проверьте по графику!).

4. Типовые динамические звенья

Обычно система управления состоит из отдельных блоков, каждый из которых описывается уравнениями низкого порядка (чаще всего – первого или второго). Для понимания работы системы в целом желательно хорошо представлять, как ведут себя ее отдельные элементы. Кроме того, при построении ЛАФЧХ сложной системы передаточную функцию разбивают на простейшие сомножители

$$W(s) = W_1(s) \cdot W_2(s) \dots W_N(s)$$

и далее, воспользовавшись свойствами ЛАФЧХ, строят характеристики для всей системы как суммы ЛАЧХ и ЛФЧХ отдельных звеньев.

4.1. Усилитель

Звенья, имеющие конечный ненулевой коэффициент усиления постоянного сигнала, то есть $W(0) = k \neq 0$, называются *позиционными*. Это значит, что числитель и знаменатель передаточной функции имеют ненулевые свободные члены (постоянные слагаемые).

Простейшее позиционное звено – идеальный (безынерционный) **усилитель**. Его передаточная функция $W(s) = k$. Строго говоря, он не является динамическим звеном, поскольку изменение выхода происходит мгновенно, сразу вслед за изменением входа. При действии на вход единичного ступенчатого сигнала $1(t)$ (или дельта-функции $\delta(t)$) на выходе будет такой же сигнал, усиленный в k раз, поэтому переходная и импульсная характеристики звена равны

$$h(t) = k \quad (t > 0) \quad \text{и} \quad w(t) = k \cdot \delta(t).$$

Если на вход усилителя действует синусоидальный сигнал, на выходе он усиливается в k раз без изменения фазы, поэтому амплитудная и фазовая частотная характеристики не зависят от частоты входного сигнала:

$$A(\omega) = k, \quad \phi(\omega) = 0.$$

4.2. Аперриодическое звено

Одно из самых часто встречающихся звеньев – **аперриодическое**, которое описывается дифференциальным уравнением

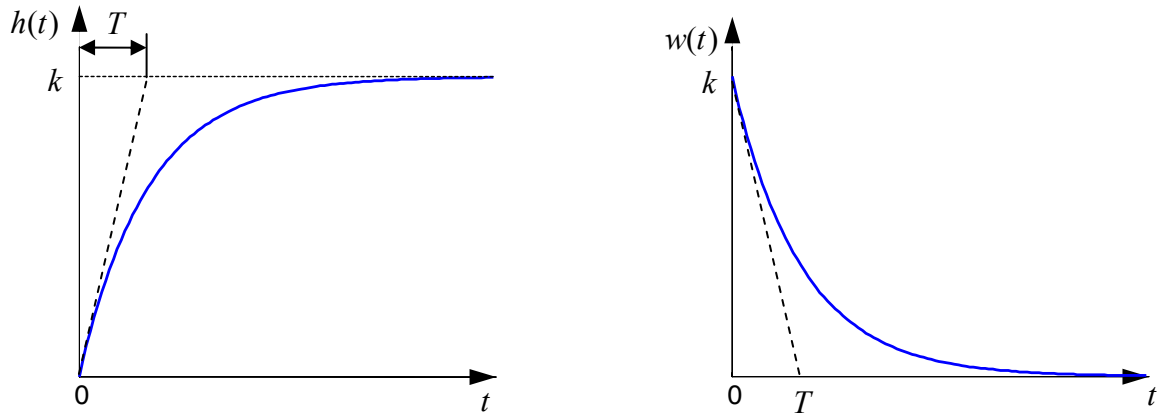
$$T \frac{dy(t)}{dt} + y(t) = k \cdot x(t) \quad (39)$$

и имеет передаточную функцию $W(s) = \frac{k}{Ts + 1}$. Здесь k – безразмерный коэффициент, а $T > 0$ – постоянная, которая называется *постоянной времени* звена. Постоянная времени – размерная величина, она измеряется в секундах и характеризует *инерционность* объекта, то есть скорость его реакции на изменение входного сигнала.

В разд. 3.3 и 3.4 мы уже нашли переходную и весовую функции аперриодического звена

$$h(t) = k \left[1 - \exp\left(-\frac{t}{T}\right) \right], \quad w(t) = \frac{k}{T} \exp\left(-\frac{t}{T}\right).$$

Они показаны на рисунке:

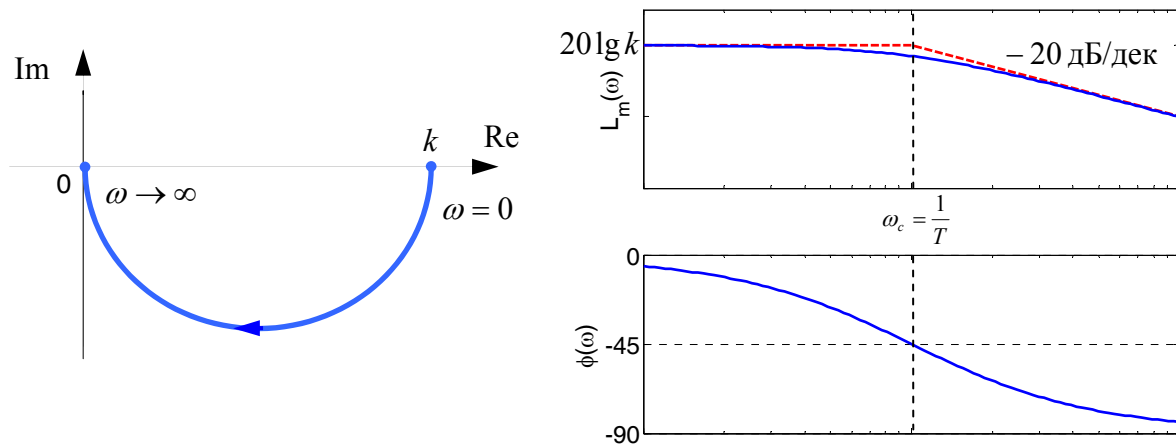


Обратите внимание, что предельное значение переходной характеристики равно k , а касательная к ней в точке $t = 0$ пересекается с линией установившегося значения при $t = T$. Переходная и импульсная характеристики выходят на установившееся значение (с ошибкой не более 5%) примерно за время $3T$. Эти факты позволяют определять постоянную времени экспериментально, по переходной характеристике звена.

Частотная характеристика определяется выражением

$$W(j\omega) = \frac{k}{Tj\omega + 1} = \frac{k(1 - Tj\omega)}{T^2\omega^2 + 1} = \frac{k}{T^2\omega^2 + 1} - \frac{jkT\omega}{T^2\omega^2 + 1}.$$

Для каждой частоты ω значение $W(j\omega)$ – это точка на комплексной плоскости. При изменении ω от 0 до ∞ получается кривая, которая называется *годографом Найквиста* (диаграммой Найквиста). В данном случае можно показать, что частотная характеристика – это полуокружность с центром в точке $(0,5k; 0)$ радиуса $0,5k$. Годограф начинается (на нулевой частоте) в точке $(k; 0)$ и заканчивается в начале координат (при $\omega \rightarrow \infty$).



Асимптотическая ЛАЧХ этого звена образована двумя прямыми, которые пересекаются на *сопрягающей частоте* $\omega_c = \frac{1}{T}$. На низких частотах она имеет нулевой наклон (так как звено позиционное), причем в этой области $L_m \approx 20 \lg k$.

На высоких частотах наклон ЛАЧХ равен -20 дБ/дек, так как степень знаменателя передаточной функции на единицу больше степени ее числителя. Фазовая характеристика меняется от 0 до -90° , причем на сопрягающей частоте ω_c она равна -45° .

Поскольку ЛАЧХ уменьшается на высоких частотах, апериодическое звено подавляет высокочастотные шумы, то есть обладает свойством *фильтра низких частот*.

Для сравнения рассмотрим также *неустойчивое апериодическое звено*, которое задается уравнением

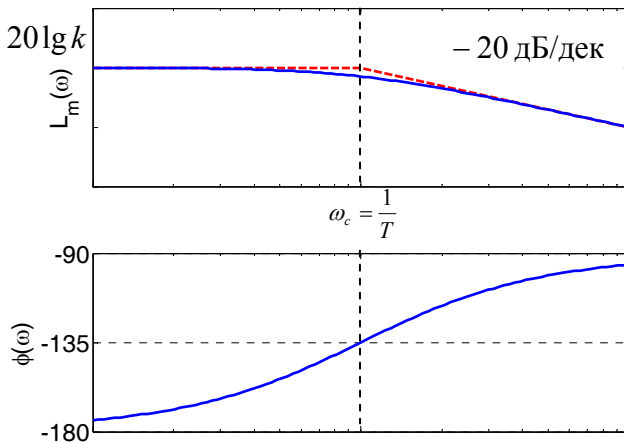
$$T \frac{dy(t)}{dt} - y(t) = k \cdot x(t) \quad (40)$$

Как видим, все отличие от (39) – только в знаке в левой части уравнения (плюс сменился на минус). Однако при этом кардинально меняются переходная и импульсная характеристики:

$$h(t) = k \left[\exp\left(\frac{t}{T}\right) - 1 \right], \quad w(t) = \frac{k}{T} \exp\left(\frac{t}{T}\right).$$

Обычно предполагается, что постоянная времени $T > 0$, тогда экспоненты в этих выражениях бесконечно возрастают с ростом t . Поэтому звено названо «неустойчивым»: в покое оно находится в неустойчивом равновесии, а при малейшем возмущении «идет вразнос».

Интересно сравнить частотные характеристики устойчивого и неустойчивого апериодических звеньев с теми же коэффициентами усиления и постоянными времени.



Из этого графика видно, что ЛАЧХ неустойчивого звена точно совпадает с ЛАЧХ аналогичного устойчивого, но отрицательный фазовый сдвиг значительно больше. Устойчивое апериодическое звено относится к *минимально-фазовым* звеньям, то есть его фаза по модулю меньше, чем фаза любого звена с такой же амплитудной характеристикой. Соответственно, неустойчивое звено – *неминимально-фазовое*.

К **неминимально-фазовым** звеньям относятся все звенья, передаточные функции которых имеют *нули или полюса* в правой полуплоскости, то есть, с положительной вещественной частью.

Для *минимально-фазовых* звеньев все нули и полюса передаточной функции находятся в левой полуплоскости (имеют отрицательные вещественные части). Например, при положительных постоянных времени T_1 , T_2 и T_3 звено с передаточной функцией

$$W(s) = \frac{T_1 s + 1}{(T_2 s + 1)(T_3 s + 1)}$$

– минимально-фазовое, а звенья с передаточными функциями

$$W_1(s) = \frac{T_1 s + 1}{(T_2 s + 1)(T_3 s - 1)}, \quad W_2(s) = \frac{T_1 s - 1}{(T_2 s + 1)(T_3 s + 1)}, \quad W_3(s) = \frac{T_1 s + 1}{(T_2 s - 1)(T_3 s - 1)}$$

– неминимально-фазовые.

4.3. Колебательное звено

Колебательное звено – это звено второго порядка с передаточной функцией вида

$$W(s) = \frac{k}{b_2 s^2 + b_1 s + 1},$$

знаменатель которой имеет комплексно-сопряженные корни (то есть, $b_1^2 - 4b_2 < 0$). Как известно из теории дифференциальных уравнений, свободное движение такой системы содержит гармонические составляющие (синус, косинус), что дает колебания выхода при изменении входного сигнала.

Несложно представить передаточную функцию колебательного звена в форме

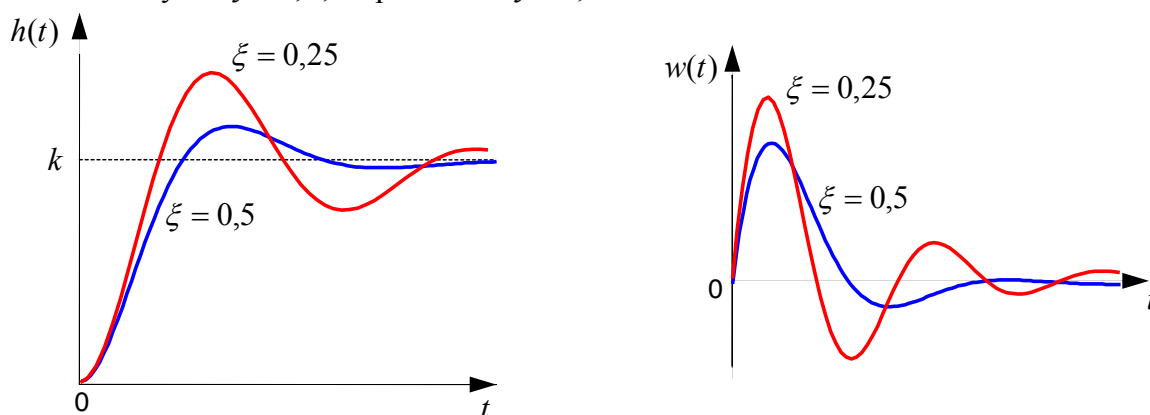
$$W(s) = \frac{k}{T^2 s^2 + 2T\xi s + 1} \quad (41)$$

где k – коэффициент, T – постоянная времени (в секундах), ξ – параметр затухания ($0 < \xi < 1$). Постоянная времени определяет инерционность объекта, чем она больше, тем медленнее изменяется выход при изменении входа. Чем больше ξ , тем быстрее затухают колебания.

При $\xi = 0$ в (41) получается *консервативное* звено, которое дает незатухающие колебания на выходе. Если $\xi \geq 1$, модель (41) представляет *апериодическое звено второго порядка*, то есть последовательное соединение двух апериодических звеньев.

Колебательное звено относится к позиционным звеньям, его статический коэффициент усиления равен $W(0) = k$.

Переходная и импульсная характеристики отличаются выраженной колебательностью, особенно при малых значениях параметра затухания ξ . На следующих двух графиках синие линии соответствуют $\xi = 0,5$, а красные – $\xi = 0,25$.

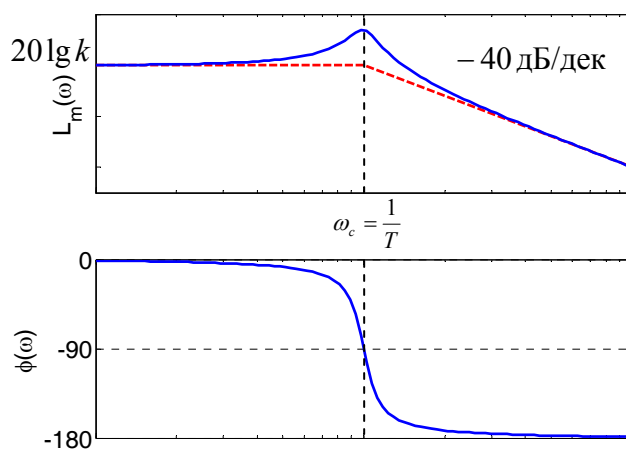


Асимптотическая ЛАЧХ этого звена образована двумя прямыми, которые пересекаются на *сопрягающей частоте* $\omega_c = \frac{1}{T}$. На низких частотах она имеет нулевой наклон (так как звено позиционное), причем в этой области $L_m \approx 20 \lg k$.

На высоких частотах наклон ЛАЧХ равен -40 дБ/дек, так как степень знаменателя передаточной функции на два больше степени ее числителя. Фазовая характеристика меняется от 0 до -180° , причем на сопрягающей частоте ω_c она равна -90° .

При значениях $\xi < 0,5$ ЛАЧХ имеет так называемый «горб» в районе сопрягающей частоты, причем его высота увеличивается с уменьшением ξ . Это означает, что при частоте входного сигнала, равной ω_c , наблюдается *резонанс*, то есть частота возмущения совпадает с частотой собственных колебаний системы.

В предельном случае при $\xi = 0$ (*консервативное звено*) ЛАЧХ терпит разрыв (обращается в бесконечность) на частоте ω_c , при таком входе амплитуда колебаний неограниченно растет и на практике объект разрушается.



4.4. Интегрирующее звено

Простейший пример интегрирующего звена – ванна, в которую набирается вода. Входной сигнал – это поток воды через кран, выход системы – уровень воды в ванне. При поступлении воды уровень растет, система «накапливает» (интегрирует) входной сигнал.

Интегрирующее звено описывается уравнением

$$\frac{dy(t)}{dt} = k \cdot x(t), \quad (42)$$

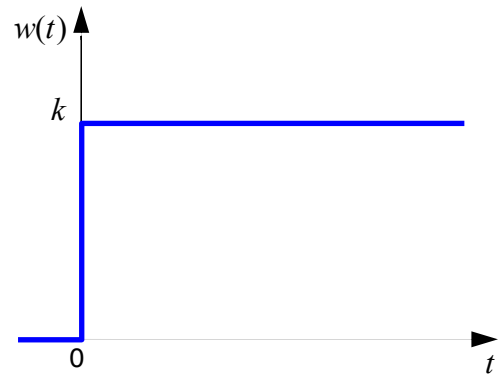
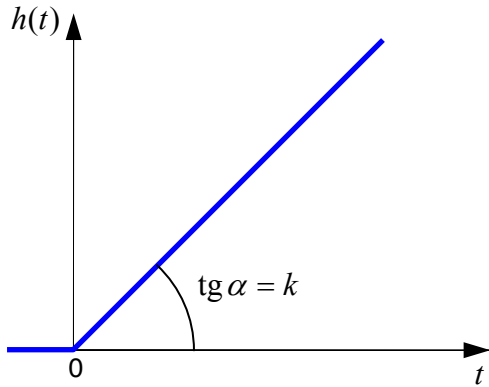
которому соответствует передаточная функция $W(s) = \frac{k}{s}$. Решение уравнения (42) дает

$$y(t) = y(0) + k \int_0^t x(\tau) d\tau.$$

Используя это решение для единичного скачка ($x(t) = 1$ при $t \geq 0$) при нулевых начальных условиях ($y(0) = 0$), получаем линейно возрастающую *переходную характеристику*:

$$h(t) = k \cdot t.$$

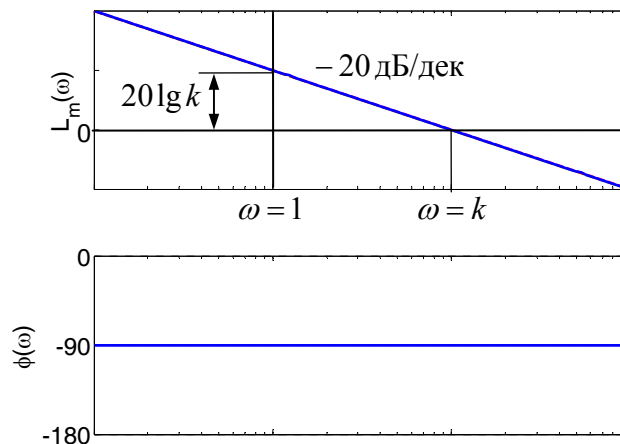
Для того, чтобы найти *импульсную характеристику*, вспомним, что интеграл от дельта-функции на любом интервале, включающем $t = 0$, равен 1. Поэтому $w(t) = k$ (при $t \geq 0$).



Частотная характеристика интегрирующего звена определяется формулой

$$W(j\omega) = \frac{k}{j\omega} = -j \frac{k}{\omega}.$$

Можно показать, что его логарифмическая амплитудная частотная характеристика – это прямая с наклоном -20 дБ/дек. На низких частотах усиление максимально, теоретически на частоте $\omega = 0$ оно равно бесконечности. Высокие частоты, наоборот, подавляются интегратором.



На частоте $\omega = 1$ значение ЛАЧХ равно $20 \lg k$, а при $\omega = k$ ЛАЧХ обращается в нуль, поскольку $|W(j\omega)| = 1$. Фазовая характеристика $\phi(\omega) = -90^\circ$ – говорит о постоянном сдвиге фазы на всех частотах.

4.5. Дифференцирующие звенья

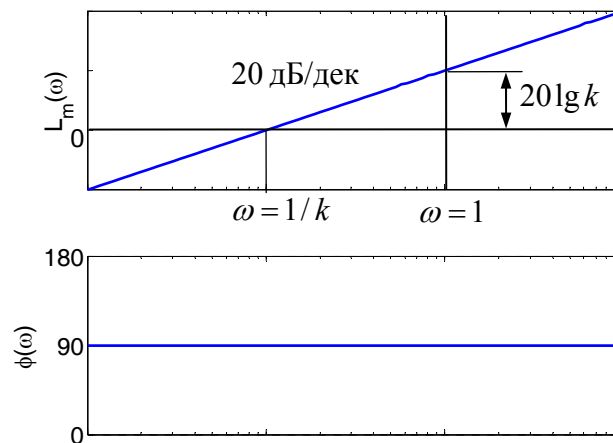
Дифференцирующее звено дает на выходе производную входного сигнала. Уравнение идеального дифференцирующего звена $y(t) = k \frac{dx(t)}{dt}$, его операторная запись $y(t) = k \cdot p x(t)$, а передаточная функция $W(s) = k \cdot s$.

Известно, что производная единичного ступенчатого сигнала $1(t)$ в точке $t = 0$ – это дельта-функция $\delta(t)$. Поэтому переходная и весовая функции дифференцирующего звена

$$h(t) = k\delta(t), \quad w(t) = k \frac{d\delta(t)}{dt}.$$

Это физически нереализуемые функции, так как дельта-функцию и ее производную, имеющие бесконечные значения, невозможно получить на реальном устройстве. Поэтому идеальное дифференцирующее относится к *физически нереализуемым* звеньям.

Логарифмическая амплитудная частотная характеристика дифференцирующего звена – прямая с наклоном 20 дБ/дек, пересекающая ось абсцисс $L_m(\omega) = 0$ на частоте $\omega = \frac{1}{k}$. При $\omega = 1$ ЛАЧХ равна $L_m(1) = 20 \lg k$. Дифференцирующее звено подавляет низкие частоты (производная от постоянного сигнала равна нулю) и бесконечно усиливает высокочастотные сигналы, что требует бесконечной энергии и невозможно в физически реализуемых системах.



Фазовая характеристика не зависит от частоты, звено дает положительный сдвиг фазы на 90° . Действительно, при дифференцировании сигнала $x(t) = \sin \omega t$ получаем

$$y(t) = \cos \omega t = \sin(\omega t + 90^\circ).$$

Дифференцирующее звено реагирует не на изменение самой входной величины, а на изменение ее производной, то есть на *тенденцию* развития событий. Поэтому говорят, что дифференцирующее звено обладает *упреждающим, прогнозирующим* действием. С его помощью можно ускорить реакцию системы.

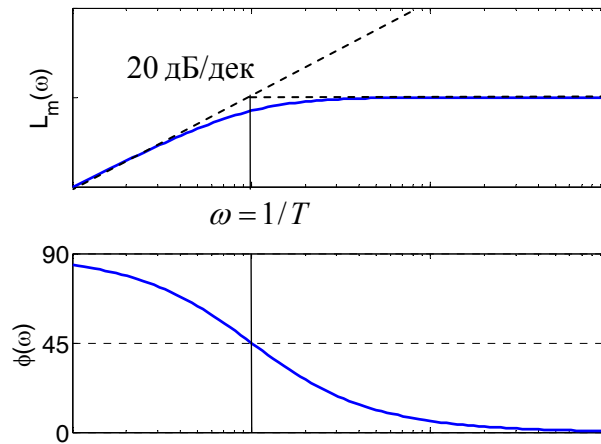
В технике не могут использоваться физически нереализуемые звенья. Поэтому важно рассмотреть аналогичное звено, которое выполняет дифференцирование низкочастотных сигналов и одновременно имеет ограниченное усиление на высоких частотах. *Инерционное дифференцирующее звено* описывается уравнением

$$T \frac{dy(t)}{dt} + y(t) = k \frac{dx(t)}{dt}$$

и имеет передаточную функцию $W(s) = \frac{ks}{Ts + 1}$. Фактически это последовательное соединение идеального дифференцирующего и апериодического звеньев.

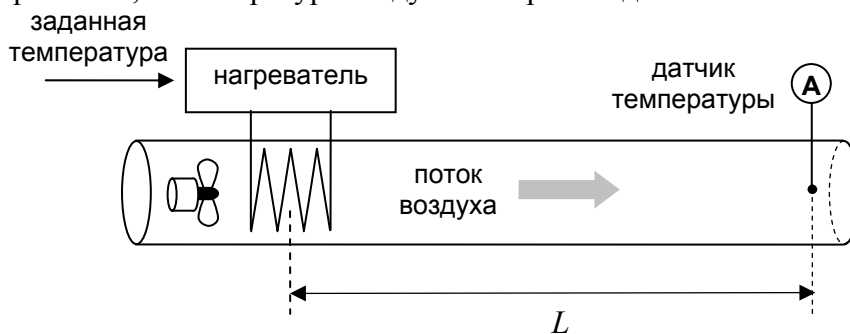
Апериодическое звено добавляет инерционность: обладая свойствами фильтра низких частот, оно ограничивает усиление на высоких частотах. Поскольку передаточная функция

имеет равные степени числителя и знаменателя, на высоких частотах (выше *сопрягающей частоты* $\omega_c = 1/T$) ЛАЧХ имеет нулевой наклон, поэтому неограниченного роста коэффициента усиления не происходит. Одновременно теряется точность дифференцирования, так как фазовая характеристика изменяется от 90° до нуля.



4.6. Запаздывание

Представим себе трубу, через которую вентилятор прокачивает воздух. В начале трубы установлен нагреватель, а температура воздуха измеряется датчиком в точке А.



Очевидно, что при изменении температуры воздуха датчик обнаружит это не сразу, а через время $\tau = L/v$, где L – длина трубы (в метрах), а v – скорость потока воздуха (в м/с). В этом случае говорят, что в системе есть транспортное **запаздывание** на величину τ (в секундах).

Другой распространенный пример – вычислительное запаздывание в компьютере. Так называется время, которое необходимо для расчета нового управляющего сигнала после получения всех исходных данных.

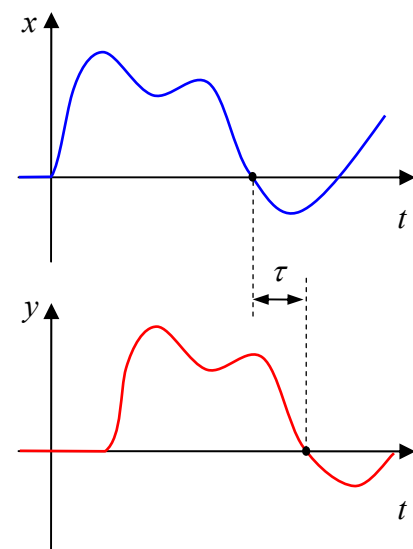
Запаздывание в системе просто сдвигает сигнал вправо на временной оси, не меняя его формы. Математически это можно записать в виде

$$y(t) = x(t - \tau).$$

Изображение сигнала на выходе звена запаздывания вычисляется по *теореме о смещении аргумента* для преобразования Лапласа:

$$Y(s) = \mathcal{L}\{y(t)\} = \int_0^\infty x(t - \tau) e^{-st} dt = e^{-s\tau} \int_0^\infty x(t) e^{-st} dt = e^{-s\tau} X(s),$$

поэтому передаточная функция звена чистого запаздывания равна $W_\tau(s) = e^{-s\tau}$.



Очевидно, что при гармоническом входном сигнале запаздывание не изменяет амплитуду, но вносит дополнительный отрицательный сдвиг фазы. Частотная характеристика этого звена имеет вид $W_\tau(j\omega) = e^{-j\omega\tau}$. По общим формулам находим:

$$A(j\omega) = |W_\tau(j\omega)| = 1, \quad \phi(j\omega) = \arg W_\tau(j\omega) = -\omega\tau.$$

Таким образом, фазовая частотная характеристика звена запаздывания – линейная функция частоты ω , чем больше частота, тем больше фазовый сдвиг.

4.7. «Обратные» звенья

Звено с передаточной функцией $\tilde{W}(s) = \frac{1}{W(s)}$ назовем «обратным» звеном для звена с передаточной функцией $W(s)$ (или *инверсией* для этого звена). Предположим, что мы знаем ЛАФЧХ для исходного звена и хотим найти ЛАФЧХ «обратного» звена без вычислений. Эта задача имеет простое решение.

Для исходного звена $W(j\omega) = P(\omega) + jQ(\omega)$, где $P(\omega)$ и $Q(\omega)$ – соответственно вещественная и мнимая частотные характеристики. Амплитудная и фазовая характеристики имеют вид

$$A(\omega) = \sqrt{P^2(\omega) + Q^2(\omega)}, \quad \phi(\omega) = \arctg \frac{Q(\omega)}{P(\omega)}.$$

Для «обратного» звена получим

$$\tilde{W}(j\omega) = \frac{1}{W(j\omega)} = \frac{1}{P(\omega) + jQ(\omega)} = \frac{P(\omega) - jQ(\omega)}{P^2(\omega) + Q^2(\omega)},$$

что после простых преобразований дает

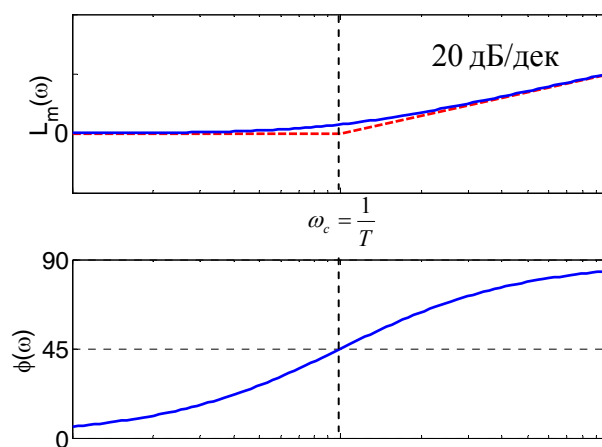
$$\tilde{A}(\omega) = \frac{1}{\sqrt{P^2(\omega) + Q^2(\omega)}} = \frac{1}{A(\omega)}, \quad \tilde{\phi}(\omega) = -\arctg \frac{Q(\omega)}{P(\omega)} = -\phi(\omega).$$

Таким образом, для логарифмических характеристик получаем

$$20 \lg \tilde{A}(\omega) = 20 \lg \frac{1}{A(\omega)} = -20 \lg A(\omega), \quad \tilde{\phi}(\omega) = -\phi(\omega).$$

Это значит, что при переходе к «обратной» передаточной функции ЛАЧХ и ЛФЧХ просто меняют знак.

Рассмотрим, например, звено с передаточной функцией $W(s) = Ts + 1$. Оно является «обратным» для апериодического звена, поэтому можно сразу нарисовать его ЛАФЧХ так, как на рисунке.



Для звена чистого запаздывания «обратным» будет звено с передаточной функцией $\tilde{W}_\tau(s) = e^{s\tau}$, его амплитудная частотная характеристика равна 1 на всех частотах, а фазовая вычисляется как $\phi(\omega) = \omega\tau$. Положительный сдвиг фазы говорит о том, что сигнал на выходе по-

является *раньше*, чем на входе. Такое звено называется *звеном упреждения* или *предсказания*. Понятно, что в реальных системах нельзя «заглянуть в будущее», поэтому звено упреждения физически нереализуемо. Тем не менее, модели некоторых практических задач могут включать звенья упреждения. Например, известны «автопилоты» для автомобилей, которые используют данные о рельефе дороги на некотором расстоянии *вперед* машины (будущие значения!), полученные с помощью лазерного измерителя.

4.8. ЛАФЧХ сложных звеньев

Для построения ЛАФЧХ звеньев со сложными передаточными функциями их числитель и знаменатель разбивают на сомножители первого и второго порядков. Фактически сложное звено при этом представляется как последовательное соединение простых звеньев, для которых известны все характеристики. При этом асимптотическую ЛАЧХ можно легко построить даже вручную.

Рассмотрим звено второго порядка с передаточной функцией

$$W(s) = \frac{a_1 s + a_0}{b_2 s^2 + b_1 s + b_0} = \frac{k(T_2 s - 1)}{(T_1 s + 1)(T_3 s + 1)}.$$

Здесь $T_i (i=1, \dots, 3)$ – положительные постоянные времени. Для определенности примем $T_1 > T_2 > T_3$.

Представим передаточную функцию в виде произведения

$$W(s) = k \cdot \frac{1}{T_1 s + 1} \cdot (T_2 s - 1) \cdot \frac{1}{T_3 s + 1}. \quad (43)$$

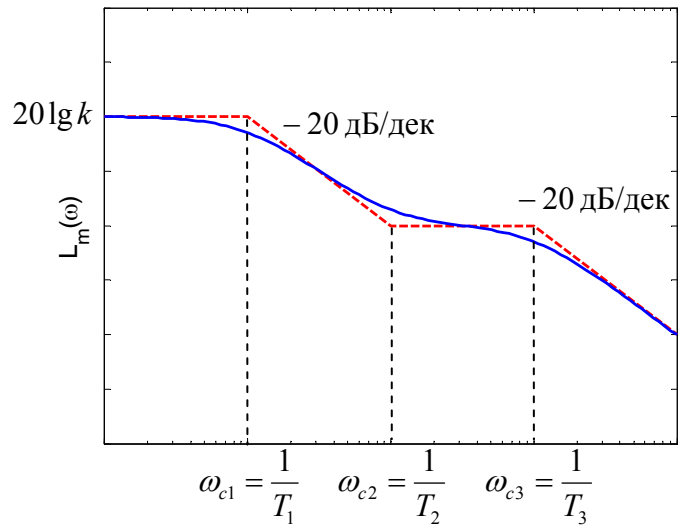
Таким образом, это звено представляет собой последовательное соединение усилителя, двух апериодических звеньев и усилителя с дифференцированием (его передаточная функция $T_2 s - 1$).

Как следует из свойств ЛАФЧХ (37)–(38), для построения ЛАЧХ системы с передаточной функцией $W(s)$ (43) достаточно сложить ЛАЧХ всех ее сомножителей.

На низких частотах, до первой сопрягающей частоты $\omega_{c1} = \frac{1}{T_1}$, «работает» только усилитель, и асимптотическая ЛАЧХ идет на постоянном уровне $20 \lg k$. Начиная с частоты $\omega_{c1} = \frac{1}{T_1}$ первое апериодическое звено дает наклон ЛАЧХ -20 дБ/дек, а с

частоты $\omega_{c2} = \frac{1}{T_2}$ звено $T_2 s - 1$ восстанавливает нулевой наклон. На частотах выше $\omega_{c3} = \frac{1}{T_3}$ включается второе (быстродействующее) апериодическое звено, которое определяет наклон -20 дБ/дек оставшейся высокочастотной части ЛАЧХ.

Для построения фазовой характеристики желательно использовать компьютерные программы. Однако принцип остается тот же, что и для ЛАЧХ: полная фазовая характеристика равна сумме фазовых характеристик отдельных звеньев, входящих в произведение.



5. Структурные схемы

5.1. Условные обозначения

Систему управления можно разбить на блоки, имеющие вход и выход (объект, регулятор, привод, измерительная система). Для того, чтобы показать взаимосвязи этих блоков, используют структурные схемы. На них каждый элемент изображается в виде прямоугольника, внутри которого записывается его передаточная функция. Вход и выход блока показывают соответственно «входящей» и «выходящей» стрелками.

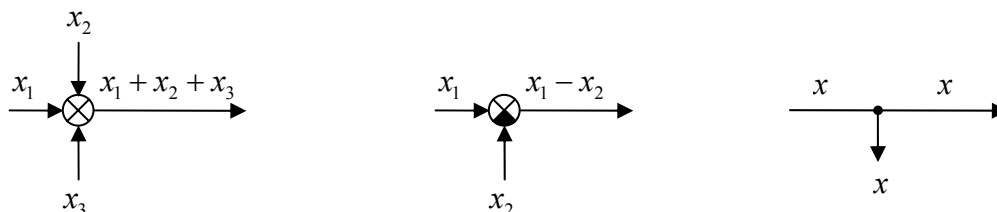


Строго говоря, есть две формы записи:

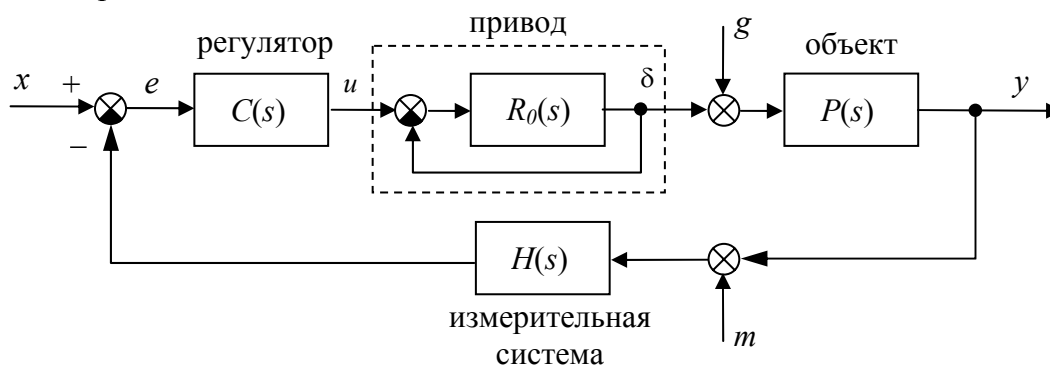
- *операторная запись*, когда передаточная функция записывается как функция оператора дифференцирования p , входы и выходы блоков – функции времени;
- *запись в изображениях*, когда передаточная функция записывается как функция комплексной переменной s , а для обозначения входов и выходов используют их изображения по Лапласу.

Однако суть дела от этого не меняется. Поэтому дальше при обозначении сигналов мы, несколько жертвуя строгостью ради простоты записи, будем обозначать сигналы строчными буквами, не указывая независимую переменную (t или s), а в записи передаточных функций будем использовать переменную s , как принято в литературе.

Для суммирующих элементов используют специальное обозначение – круг, разбитый на сектора. Если сектор залит черным цветом, поступающий в него сигнал вычитается, а не складывается с другими. Разветвление сигнала обозначается точкой, как и радиотехнике.



На следующем рисунке показана типичная схема системы управления кораблем по курсу. Здесь вход x – заданный курс, выход y – фактический курс. Сигналы e , u и δ обозначают соответственно ошибку регулирования, сигнал управления и управляющее воздействие привода на объект (угол поворота руля). Сигнал g – это возмущение (влияние ветра и морского волнения), а m – шум измерений.

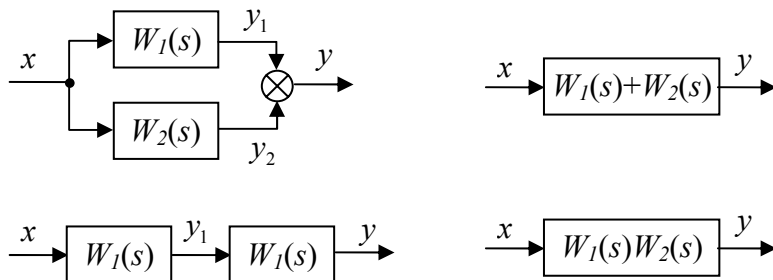


В этой системе кроме «большого» контура управления (регулятор – привод – объект) есть еще внутренний контур привода (звено с передаточной функцией $R_0(s)$ охвачено отрицательной обратной связью).

5.2. Правила преобразования

Многие инженерные (классические) методы исследования систем управления основаны на использовании передаточных функций. Для построения передаточной функции системы между заданными входом и выходом нужно преобразовать структурную схему так, чтобы в конечном счете остался один блок с известной передаточной функцией. Для этого используют *структурные преобразования*.

Легко показать, что передаточные функции параллельного и последовательного соединений равны соответственно сумме и произведению исходных передаточных функций:



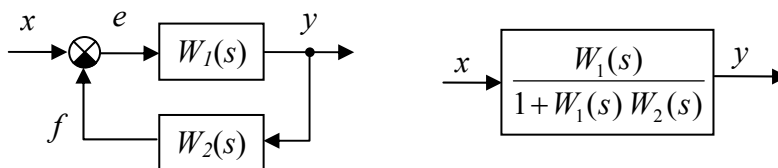
Действительно, в изображениях по Лапласу для параллельного соединения получаем

$$Y(s) = Y_1(s) + Y_2(s) = W_1(s)X(s) + W_2(s)X(s) = [W_1(s) + W_2(s)]X(s),$$

а для последовательного

$$Y(s) = W_2(s)Y_1(s) = W_1(s)W_2(s)X(s).$$

Для контура с отрицательной обратной связью имеем



Для доказательства заметим, что $Y(s) = W_1(s)E(s)$, а изображение ошибки равно

$$E(s) = X(s) - F(s) = X(s) - W_2(s)Y(s).$$

Поэтому

$$Y(s) = W_1(s)[X(s) - W_2(s)Y(s)].$$

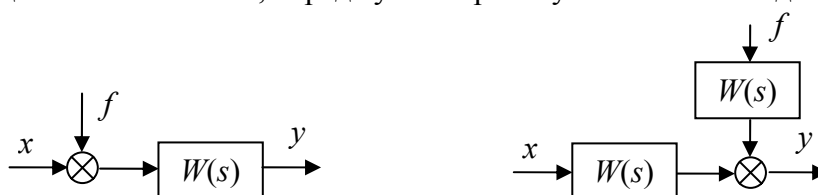
Переносим $X(s)$ в левую часть, получаем

$$Y(s)[1 + W_1(s)W_2(s)] = W_1(s)X(s) \quad \Rightarrow \quad Y(s) = \frac{W_1(s)}{1 + W_1(s)W_2(s)}X(s).$$

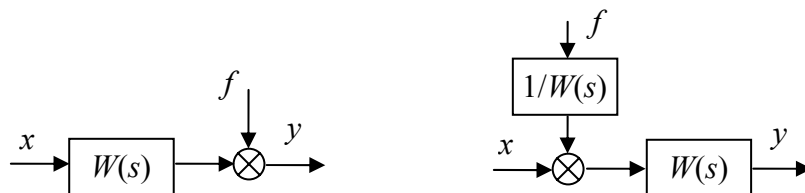
Если обратная связь – положительная (сигналы x и f складываются), в знаменателе будет стоять знак «минус»:

$$W(s) = \frac{W_1(s)}{1 - W_1(s)W_2(s)}.$$

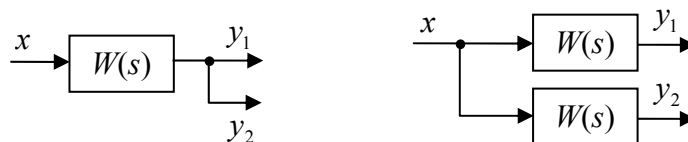
Звено можно переносить через сумматор как вперед, так и назад. Чтобы при этом передаточные функции не изменились, перед сумматором нужно поставить дополнительное звено:



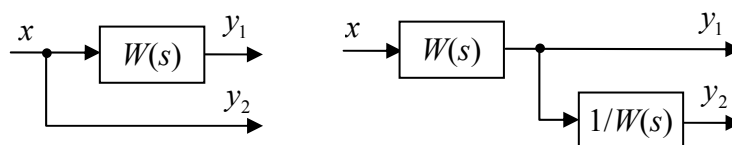
Обратите внимание, что передаточные функции от обоих входов к выходу на двух схемах одинаковые. Для следующей пары это условие тоже выполняется:



Звено можно переносить также через точку разветвления, сохраняя все передаточные функции:

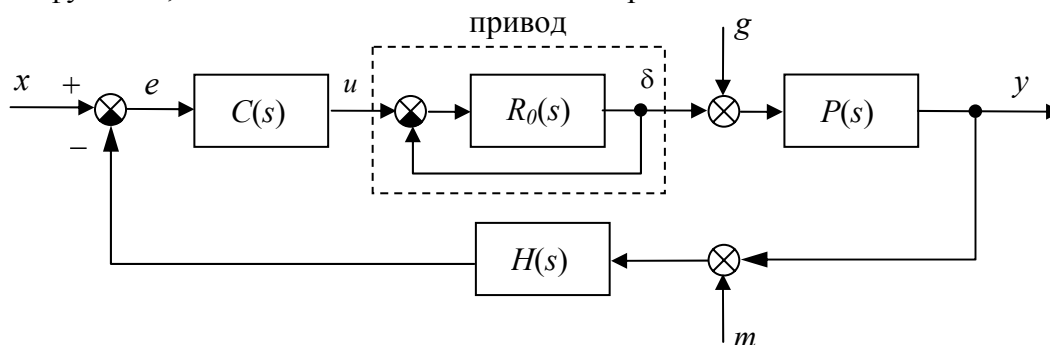


Эти две схемы тоже равносильны:



5.3. Типовая одноконтурная система

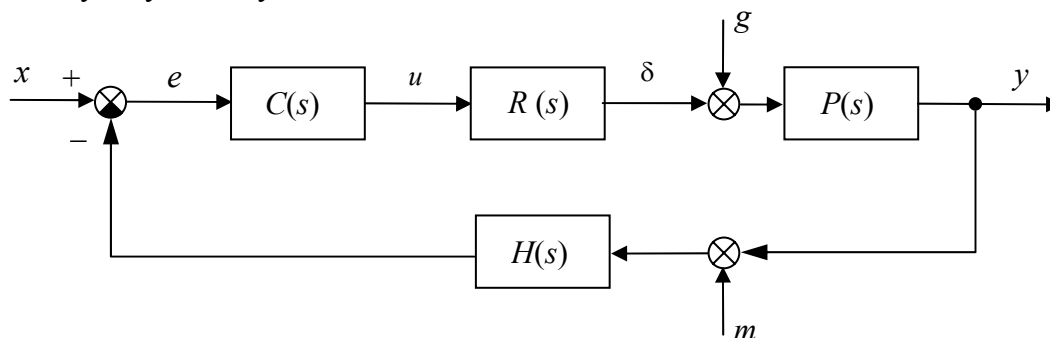
Применим показанные выше приемы для вычисления передаточных функций рассмотренной выше системы. Здесь три входа (x , g и m), а в качестве выходов обычно рассматривают выход системы y , сигнал управления u и ошибку e . Таким образом, всего можно записать 9 передаточных функций, соединяющих все возможные пары вход-выход.



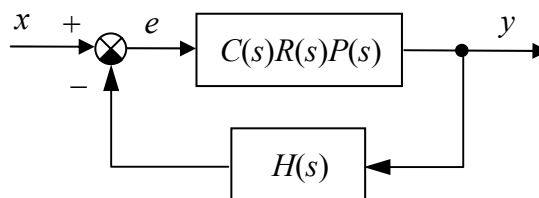
Сначала найдем полную передаточную функцию привода (обведенного штриховой рамкой), используя формулу для контура с отрицательной обратной связью:

$$R(s) = \frac{R_0(s)}{1 + R_0(s)}.$$

Получаем следующую схему:



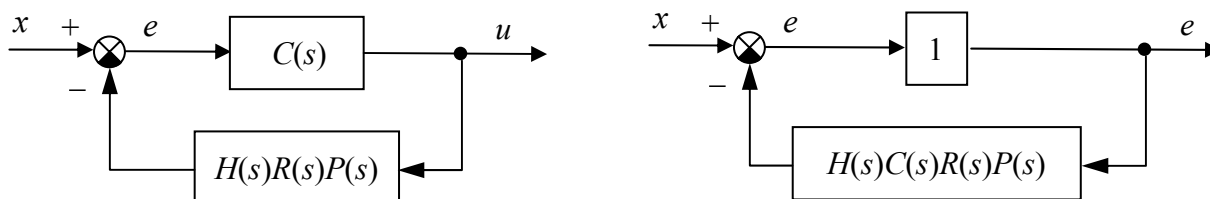
Теперь найдем передаточные функции от входа x ко всем выходам. Для этого все остальные входы будем считать нулевыми и удалим со схемы. Кроме того, заменим последовательное соединение звеньев с передаточными функциями $C(s)$, $R(s)$ и $P(s)$ на одно звено:



Для получения окончательного результата снова используем формулу для контура с отрицательной обратной связью:

$$W(s) = \frac{C(s)R(s)P(s)}{1 + H(s)C(s)R(s)P(s)}.$$

Принимая в качестве выходов управления u и ошибку e , получим похожие схемы:



Первая из этих схем дает передаточную функцию по управлению $W_u(s)$, а вторая – передаточную функцию по ошибке $W_e(s)$ (здесь блок с передаточной функцией, равной единице, можно было вообще не рисовать). Снова применяя формулу для контура с отрицательной обратной связью, получаем:

$$W_u(s) = \frac{C(s)}{1 + H(s)C(s)R(s)P(s)}, \quad W_e(s) = \frac{1}{1 + H(s)C(s)R(s)P(s)}.$$

Используя этот подход, легко найти передаточные функции для других входов. Теперь вы вполне можете сделать это самостоятельно.

6. Анализ систем управления

6.1. Требования к управлению

Что мы хотим от управления? Это зависит, прежде всего, от решаемой задачи. В задаче стабилизации наиболее важны свойства установившегося режима. Для следящих систем в первую очередь нужно обеспечить высокое качество переходных процессов при изменении задающего сигнала (*уставки*).

В целом можно выделить четыре основных требования:

- **точность** – в установившемся режиме система должна поддерживать заданное значение выхода системы, причем ошибка (разница между заданным и фактическим значением) не должна превышать допустимую;
- **устойчивость** – система должна оставаться устойчивой на всех режимах, не должна идти «вразнос» (корабль не должен идти по кругу при смене курса);
- **качество переходных процессов** – при смене заданного значения система должна переходить в нужное состояние по возможности быстро и плавно;
- **робастность** – система должна сохранять устойчивость и приемлемое качество даже в том случае, если динамика объекта и свойства внешних возмущений немного отличаются от тех, что использовались при проектировании.

6.2. Процесс на выходе

Начнем с простого – покажем, как вычислить процесс на выходе системы с передаточной функцией $W(s)$ при входном сигнале, для которого известно изображение по Лапласу $X(s)$.

При нулевых начальных условиях изображение выхода равно $Y(s) = W(s) X(s)$. Предположим, что $W(s)$ и $X(s)$ – рациональные функции, то есть их можно представить в виде отношения полиномов

$$W(s) = \frac{n_W(s)}{\Delta(s)}, \quad X(s) = \frac{n_X(s)}{d_X(s)}.$$

Для простоты будем считать, что полиномы $\Delta(s)$ и $d_X(s)$ имеют только простые вещественные корни, так что

$$\Delta(s) = (s - \alpha_1)(s - \alpha_2) \dots (s - \alpha_N), \quad d_X(s) = (s - \beta_1)(s - \beta_2) \dots (s - \beta_M),$$

причем общих корней у них нет⁷. Числа α_i ($i = 1, \dots, N$) и β_j ($j = 1, \dots, M$) называются *полюсами* функций $W(s)$ и $X(s)$ соответственно.

При этих условиях произведение $Y(s) = W(s) X(s)$ можно разложить на простые дроби

$$Y(s) = \frac{a_1}{s - \alpha_1} + \frac{a_2}{s - \alpha_2} + \dots + \frac{a_N}{s - \alpha_N} + \frac{b_1}{s - \beta_1} + \frac{b_2}{s - \beta_2} + \dots + \frac{b_M}{s - \beta_M}.$$

Здесь a_i ($i = 1, \dots, N$) и b_j ($j = 1, \dots, M$) – постоянные, которые в данном случае определяются по формулам:

$$a_i = W(s)X(s)(s - \alpha_i) \Big|_{s=\alpha_i}, \quad b_j = W(s)X(s)(s - \beta_j) \Big|_{s=\beta_j}.$$

Далее мы предположим, что произведение $W(s) X(s)$ несократимо. В этом случае все числа a_i и b_j не равны нулю.

Чтобы найти выход $y(t)$, нужно вычислить обратное преобразование Лапласа для $Y(s)$. По таблицам (см., например, формулы (25)) находим

⁷ Более сложные случаи (комплексно-сопряженные и кратные полюса) рассматриваются аналогично.

$$y(t) = a_1 e^{\alpha_1 t} + a_2 e^{\alpha_2 t} + \dots + a_N e^{\alpha_N t} + b_1 e^{\beta_1 t} + b_2 e^{\beta_2 t} + \dots + b_M e^{\beta_M t}. \quad (44)$$

Вспомним, что функция $e^{\lambda t}$ при $t \rightarrow \infty$ стремится к нулю, если $\lambda < 0$; остается постоянной (равной 1) при $\lambda = 0$ и уходит в бесконечность при $\lambda > 0$. Поэтому выражение (44) позволяет сделать следующие выводы:

- сигнал на выходе системы зависит как от свойств передаточной функции системы, так и от входного сигнала;
- для того, чтобы переходный процесс затухал (функция $y(t)$ стремилась к нулю), все числа α_i ($i = 1, \dots, N$) и β_i ($i = 1, \dots, M$) должны быть отрицательными (иметь отрицательные вещественные части);
- если один из полюсов $W(s)$ или $X(s)$ равен нулю, $y(t)$ может иметь постоянную (незатухающую) составляющую;
- если хотя бы один из полюсов $W(s)$ или $X(s)$ больше нуля (имеет положительную вещественную часть), выход системы неограниченно растет.

Еще раз отметим, что мы предполагали несократимость произведения $W(s)X(s)$, иначе некоторые коэффициенты a_i и/или b_j могут оказаться нулевыми и соответствующие экспоненты «исчезают» из формулы (44). Тогда, например, может оказаться, что выход не «уходит в бесконечность» даже если $W(s)$ или $X(s)$ имеет полюс с положительной вещественной частью (и он сократился в произведении $W(s)X(s)$).

Как следует из (44), часть показателей экспонент (числа α_i ($i = 1, \dots, N$)) полностью определяются свойствами системы – это корни полинома $\Delta(s)$. Если среди них есть числа с положительной вещественной частью, сигнал выхода будет неограниченно возрастать при *любом* входе, для которого произведение $W(s)X(s)$ несократимо. В этом случае говорят, что *система неустойчива*, а соответствующие полюса также называют неустойчивыми. Полином $\Delta(s)$ называется **характеристическим полиномом**, так как расположение его корней определяет устойчивость (или неустойчивость) системы (подробнее см. разд. 6.4).

6.3. Точность

Точность системы обычно оценивается для одного из эталонных входных сигналов. Это может быть, например, единичный скачок

$$x(t) = \mathbf{1}(t) = \begin{cases} 0, & t < 0 \\ 1, & t \geq 0 \end{cases}, \quad X(s) = \frac{1}{s}$$

или линейно возрастающий сигнал

$$x(t) = \begin{cases} 0, & t < 0 \\ t, & t \geq 0 \end{cases}, \quad X(s) = \frac{1}{s^2}$$

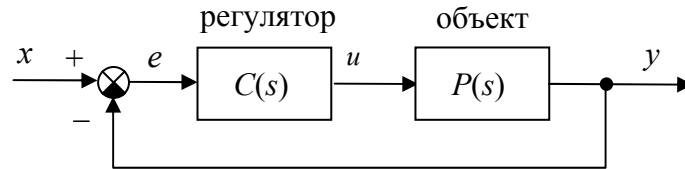
или гармонический сигнал с частотой ω :

$$x(t) = \sin \omega t, \quad X(s) = \frac{\omega}{s^2 + \omega^2}.$$

Точность системы в установившемся режиме определяется ошибкой $e(t)$ или ее изображением $E(s)$. Для ее исследования используют *передаточную функцию по ошибке* $W_e(s)$, которая связывает изображения ошибки и входного сигнала:

$$E(s) = W_e(s) X(s).$$

Рассмотрим контур управления, состоящий из регулятора и объекта:



Представим передаточные функции $C(s)$ и $P(s)$, а также изображение входа $X(s)$ в виде отношения полиномов

$$C(s) = \frac{n_C(s)}{d_C(s)}, \quad P(s) = \frac{n(s)}{d(s)}, \quad X(s) = \frac{n_X(s)}{d_X(s)}.$$

В данном случае передаточная функция по ошибке равна

$$W_e(s) = \frac{1}{1 + C(s)P(s)} = \frac{d_C(s)d(s)}{\Delta(s)},$$

где $\Delta(s) = d_C(s)d(s) + n_C(s)n(s)$ – *характеристический полином* замкнутой системы.

Рассмотрим реакцию системы на единичный ступенчатый входной сигнал, изображение которого равно $X(s) = 1/s$. Как следует из разд. 6.2, сигнал ошибки определяется полюсами передаточной функции $W_e(s)$ (то есть корнями характеристического полинома $\Delta(s)$) и полюсами изображения $X(s)$. На практике все полюса $W_e(s)$ должны иметь отрицательные вещественные части, иначе система будет неустойчивой (подробнее см. в разд. 6.4). Поэтому нулевых полюсов у функции $W_e(s)$ быть не может. Тогда

$$W_e(s)X(s) = \frac{1}{1 + C(s)P(s)} \cdot \frac{1}{s} = Y_0(s) + \frac{b}{s}.$$

Здесь изображение $Y_0(s)$ имеет полюса только с отрицательной вещественной частью, а постоянная b рассчитывается по формуле разложения на простые дроби:

$$b = \frac{1}{1 + C(0)P(0)} = \frac{d_C(0)d(0)}{\Delta(0)}.$$

Как следует из разд. 6.2, после затухания всех экспонент с отрицательными показателями получим $\lim_{t \rightarrow \infty} e(t) = b$.

Заметим, что для того, чтобы сделать нулевой статическую ошибку, достаточно обеспечить $d_C(0) = 0$ (то есть регулятор должен содержать интегратор) или $d(0) = 0$ (объект содержит интегратор).

Этот результат можно обобщить для любых незатухающих входных сигналов, изображения которых имеют полюса на мнимой оси (в точке $s = 0$ или в точках $s = \pm j\omega$). Для того, чтобы ошибка стремилась к нулю при $t \rightarrow \infty$ необходимо, чтобы эти полюса сократились в произведении

$$W_e(s)X(s) = \frac{d_C(s)d(s)}{\Delta(s)} \cdot \frac{n_X(s)}{d_X(s)}.$$

А это, в свою очередь, возможно только тогда, когда они являются корнями полинома $d_C(s)d(s)$, то есть, *внутри* системы есть модель входного сигнала. Этот принцип называется **принципом внутренней модели**.

Например, для точного отслеживания ступенчатого сигнала нужно, чтобы объект или регулятор содержали интегрирующее звено (с передаточной функцией $1/s$). Тогда произведение $d_C(s)d(s)$ имеет множитель s , и полюс $X(s)$ в точке $s = 0$ сократится в произведении $W_e(s)X(s)$. Таким образом, если передаточная функция разомкнутого контура $C(s)P(s)$ содержит множитель s в знаменателе, обеспечивается нулевая ошибка слежения за постоянным сигналом (нулевая статическая ошибка). Поэтому такую систему называют **астатической**.

Для отслеживания *линейно возрастающего сигнала* в контуре должно быть уже два интегратора (нужно сократить двойной полюс $X(s)$ в точке $s = 0$). Такая система обладает *астатизмом второго порядка*. В общем случае система, в которой

$$C(s)P(s) = \frac{1}{s^\nu} G(s),$$

где $\nu > 0$ – натуральное число и функция $G(s)$ не имеет нулей и полюсов в точке $s = 0$, называется *астатической системой ν -ого порядка*. Такая система в установившемся режиме без ошибки отслеживает сигнал вида

$$x(t) = x_0 + x_1 t + x_2 t^2 + \dots + x_{\nu-1} t^{\nu-1}$$

при любых значениях коэффициентов $x_i (i = 0, \dots, \nu - 1)$.

Казалось бы, для повышения точности можно поставить много интеграторов, и все проблемы будут решены. Но при этом нужно учесть, что мы говорили только о точности *в установившемся режиме*, не затрагивая переходные процессы (переход с режима на режим) и вопросы устойчивости. Добавление каждого нового интегратора ухудшает переходные процессы, осложняет стабилизацию системы, снижает быстродействие. Например, двойным интегратором в принципе невозможно управлять с помощью простого регулятора-усилителя (так называемого *пропорционального регулятора* или П-регулятора). Кроме того, если разомкнутая система включает два интегратора и более, для сигнала ошибки $e(t)$ справедливо ограничение

$$\int_0^\infty e(t) dt = 0.$$

На вопрос «ну и что?» можно ответить так: поскольку интеграл от ошибки равен нулю, часть времени ошибка должна быть положительной, а часть – отрицательной. Поэтому при *любом* управлении не удастся получить монотонный переходный процесс (когда сигнал выхода подходит к заданному значению «с одной стороны», как у апериодического звена).

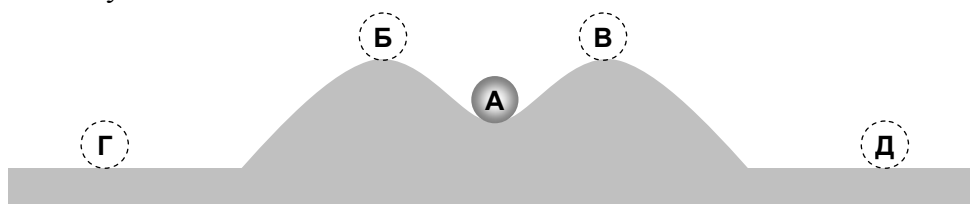
Для стохастической системы, в которой все процессы имеют случайный характер, точность оценивается с помощью математического ожидания и дисперсии ошибки. Но эти вопросы выходят за рамки пособия.

6.4. Устойчивость

6.4.1. Что такое устойчивость?

«Бытовое» понятие устойчивости известно нам с детства. Например, табуретка с двумя ножками неустойчива, она упадет при малейшем дуновении ветра, а с тремя – устойчива. Всем знакомый пример неустойчивой системы – близко расположенные микрофон и колонки, которые начинают «свистеть». Неустойчивость может привести к трагическим последствиям. Достаточно вспомнить аварии самолетов, попавших в грозовой фронт или в штопор, взрыв ядерного реактора на Чернобыльской атомной станции в 1986 г.

Термин «устойчивость» используется в численных методах, механике, экономике, социологии, психологии. Во всех этих науках имеют в виду, что устойчивая система возвращается в состояние равновесия, если какая-то сила выведет ее из этого состояния. Шарик на рисунке находится в *устойчивом равновесии* в положении А – если немного сдвинуть его с места, он скатится обратно в ямку.



Однако мы можем заметить, что если шарик *сильно* отклонить от равновесия, он может свалиться через горку вбок, то есть устойчивость нарушится.

В положениях Б и В шарик также находится в положении равновесия, но оно *неустойчиво*, так как при малейшем сдвиге в сторону шарик скатывается с вершины.

В положениях Г и Д равновесие шарика *нейтральное* – при небольшом смещении он остается в новом положении. При этом говорят, что система *нейтрально устойчива*, то есть находится на границе устойчивости.

Можно показать, что система «шарик-горка» – **нелинейная**. Как мы увидели, для нее

- устойчивость – не свойство системы, а свойство некоторого положения равновесия;
- может быть несколько положений равновесия, из них некоторые – устойчивые, а некоторые – нет;
- положение равновесия может быть устойчиво при малых отклонениях (система устойчива «в малом») и неустойчиво при больших («в большом»).

6.4.2. Устойчивость бывает разная

Известно несколько определений устойчивости, которые отличаются некоторыми деталями. Если рассматривается только выход системы при различных ограниченных входах, говорят об устойчивости «выход-выход».

Кроме того, часто изучают устойчивость *автономной* системы, на которую не действуют внешние сигналы (все входы нулевые). Предполагается, что систему вывели из положения равновесия (задали *ненулевые начальные условия*) и «отпустили». Система, которая сама возвращается в исходное положение равновесия, называется устойчивой. Если при этом рассматривается только выход системы (а не ее внутренние сигналы), говорят о «*технической устойчивости*» (или *устойчивости по выходу*). Напротив, *внутренняя* или *математическая устойчивость* означает, что не только выход, но и все внутренние переменные (переменные состояния) приближаются к своим значениям в положении равновесия.

В некоторых задачах основной рабочий режим – это периодические колебания, поэтому можно рассматривать *устойчивость процессов*, а не только положения равновесия. Однако почти все такие системы – нелинейные, и эти вопросы выходят за рамки нашего пособия.

6.4.3. Устойчивость «вход-выход»

Обычно для инженеров практиков в первую очередь важно, чтобы система не «пошла вразнос», то есть, чтобы управляемая величина не росла неограниченно при всех допустимых входных сигналах. Если это так, говорят, что система обладает *устойчивостью «вход-выход»* (при ограниченном входе выход также ограничен). Заметим, что при этом нас не интересует, как меняются внутренние переменные объекта, важен только вход и выход.

Рассмотрим ванну, которая наполняется водой из крана. Модель этой системы – интегрирующее звено. При постоянном (ограниченном по величине!) входном потоке уровень воды в ванне будет неограниченно увеличиваться (пока вода не польется через край), поэтому такая система не обладает устойчивостью «вход-выход».

6.4.4. «Техническая» устойчивость

В отличие от устойчивости «вход-выход», понятие «техническая устойчивость» относится к автономной системе, у которой все входные сигналы равны нулю.

Положением равновесия называют состояние системы, которая находится в покое, то есть, сигнал выхода $y(t)$ – постоянная величина, и все его производные равны нулю.

Систему выводят из положения равновесия и убирают все возмущения. Если при этом с течением времени (при $t \rightarrow \infty$) система возвращается в положение равновесия, она называется *устойчивой*. Если выходная координата остается ограниченной (не уходит в бесконечность), система называется *нейтрально устойчивой*, а если выход становится бесконечным – *неустойчивой*.

Если вернуться к примеру с ванной, становится понятно, что эта система – нейтрально устойчива, потому что уровень воды остается постоянным, когда мы перекроем кран. С одной стороны, уровень воды не возвращается к предыдущему значению, а с другой – не растет бесконечно (система не является неустойчивой).

6.4.5. Внутренняя устойчивость

Говоря о внутренней устойчивости, рассматривают не только выход, но и все переменные, описывающие состояние системы. В математической теории систем вектор состояния обозначают через $x(t)$, а уравнение движения системы записывают в виде

$$\frac{dx(t)}{dt} = f(x, t) \quad (45)$$

Фактически это система дифференциальных уравнений первого порядка, в нем правая часть зависит только от значений t и $x(t)$, но не от производных. Если вектор состояния $x(t)$ состоит из двух компонентов, $x_1(t)$ и $x_2(t)$, уравнение (45) можно записать в развернутой форме

$$\begin{cases} \frac{dx_1(t)}{dt} = f_1(x, t) \\ \frac{dx_2(t)}{dt} = f_2(x, t) \end{cases}$$

где функции $f_1(x, t)$ и $f_2(x, t)$ зависят от вектора состояния и времени.

Устойчивость определяется для некоторого положения равновесия. Для нелинейной системы может быть несколько положений равновесия, причем некоторые из них могут быть устойчивы, а некоторые – нет. В положении равновесия все производные равны нулю, то есть $f(x^*, t) = 0$, где x^* – соответствующий вектор состояния⁸.

Предположим, что систему вывели в некоторое начальное состояние $x_0 = x(0)$ (задали начальные условия), а потом внешнее воздействие прекратили. Дальнейшее изменение координат («движение» системы $x(t)$) можно найти как решение уравнения (45) при заданных начальных условиях.

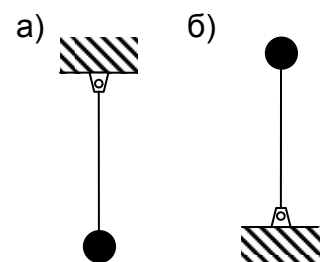
Нестрого говоря, *устойчивость* означает, что все движения $x(t)$, которые начинаются близко от положения равновесия x^* , при всех t остаются в некоторой окрестности x^* .

Лучше, конечно, если система не просто устойчива, а еще и возвращается в положение равновесия, то есть, $x(t)$ стремится к x^* при $t \rightarrow \infty$. В этом случае говорят об *асимптотической устойчивости*.

Рассмотрим маятник на рисунке а) справа, состоящий из подвешенного металлического стержня и шарика. Здесь положение равновесия – шарик в нижней точке. Если не учитывать трение, маятник, выведенный из положения равновесия, будет качаться бесконечно долго, причем амплитуда колебаний не будет увеличиваться, то есть, система устойчива.

В реальности трение, конечно, есть, поэтому колебания маятника будут постепенно затухать (амплитуда уменьшается), и система в конце концов возвращается в положение равновесия. Это значит, что маятник с трением – асимптотически устойчивая система.

Маятник на рисунке б) тоже находится в положении равновесия, но оно неустойчиво: при малейшем отклонении маятник упадет вниз.



⁸ Часто для удобства считают, что $x^* = 0$. Если это не так, можно заменить $x(t)$ на $\tilde{x}(t) = x(t) - x^*$, так чтобы в положении равновесия все координаты нового вектора $\tilde{x}(t)$ были нулевые.

Формальное определение внутренней устойчивости было введено в работах А.М. Ляпунова⁹, поэтому такое понятие устойчивости принято называть *устойчивостью по Ляпунову*.

Для простоты рассмотрим систему первого порядка, с одной переменной состояния $x(t)$. Система называется устойчивой по Ляпунову в положении равновесия x^* , если при начальном отклонении от положения равновесия x^* не более, чем на δ , траектория движения отклоняется от x^* не более, чем на ε , причем для каждого ε можно найти соответствующее ему $\delta(\varepsilon)$:

$$|x_0 - x^*| < \delta \Rightarrow |x(t) - x^*| < \varepsilon \text{ при всех } t > 0.$$

Фактически это означает, что чем меньше начальное отклонение, тем меньше траектория движения отклоняется от положения равновесия.

Если кроме того вектор состояния стремится к положению равновесия, то есть,

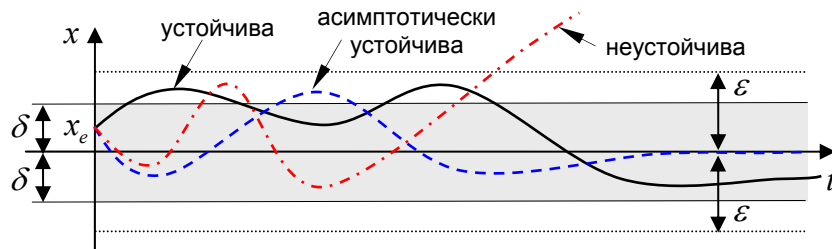
$$|x(t) - x^*| \rightarrow 0 \text{ при } t \rightarrow \infty, \quad (46)$$

система называется *асимптотически устойчивой* в положении равновесия x^* . Заметим, что выполнение условия сходимости (46) не гарантирует устойчивость по Ляпунову. Существуют примеры достаточно сложных нелинейных систем, в которых даже при очень малых отклонениях от положения равновесия сначала наблюдается большой «выброс», а затем траектория сходится к точке равновесия.

Очевидно, что асимптотическая устойчивость – более сильное требование. Положения равновесия, которые устойчивы по Ляпунову, но не асимптотически устойчивы, иногда называются *нейтрально устойчивыми* (маятник без трения, ванна с водой).

Положение равновесия *неустойчиво*, если для него не выполняется условие устойчивости Ляпунова. Это значит, что существует такое $\varepsilon > 0$, что траектория $x(t)$ выходит за границы области $|x(t) - x^*| < \varepsilon$ при сколь угодно малом отклонении начального состояния x_0 от положения равновесия x^* . Например, система переходит в *другое* положение равновесия, или $x(t)$ неограниченно возрастает.

На следующем рисунке показаны движения устойчивой, асимптотически устойчивой и неустойчивой систем первого порядка (с одной координатой $x(t)$).



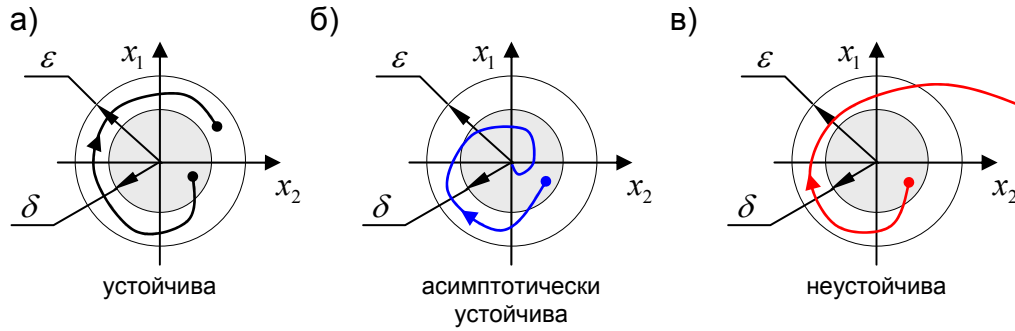
Если вектор состояния содержит несколько переменных, для оценки разности векторов $x_0 - x^*$ и $x(t) - x^*$ вместо модуля используют евклидову норму (корень из суммы квадратов отклонений по каждой координате). Например, для системы второго порядка

$$\|x(t) - x_e\| = \sqrt{(x_1(t) - x_1^*)^2 + (x_2(t) - x_2^*)^2},$$

где x_1^* и x_2^* – компоненты вектора x^* .

Траекторию движения систем второго порядка обычно изображают на *фазовой плоскости*, где по одной оси откладывается $x_1(t)$, а по другой – $x_2(t)$. На следующем рисунке показаны движения устойчивой, асимптотически устойчивой и неустойчивой систем. Для простоты предполагается, что положение равновесия – это начало координат, где $x_1 = x_2 = 0$.

⁹ А.М. Ляпунов (1857-1918) – русский математик и механик.



6.4.6. Устойчивость линейных систем

Линейные системы обладают рядом особенностей, которые во многих случаях упрощают анализ устойчивости:

- автономная линейная система (на которую не действуют внешние силы) может иметь единственное положение равновесия (в котором все сигналы равны нулю) или бесконечно много положений равновесия (шарик на плоской поверхности);
- устойчивость – это свойство линейной системы, а не отдельного положения равновесия: или все ее движения устойчивы (асимптотически устойчивы), или все неустойчивы;
- асимптотическая устойчивость линейной системы «в малом» сразу означает ее устойчивость «в целом», то есть, при любых отклонениях от положения равновесия;
- асимптотически устойчивая система также обладает устойчивостью «вход-выход», а просто устойчивая (нейтрально устойчивая, не асимптотически устойчивая) – нет.

Для того, чтобы получить условия устойчивости, рассмотрим уравнение движения линейной системы, на которую не действуют возмущения. Пусть $W(s)$ – ее передаточная функция. Будем считать, что она имеет только простые (не кратные) полюса α_i ($i = 1, \dots, N$) (корни знаменателя):

$$W(s) = \frac{n_w(s)}{\Delta(s)} = \frac{n_w(s)}{(s - \alpha_1)(s - \alpha_2) \dots (s - \alpha_N)},$$

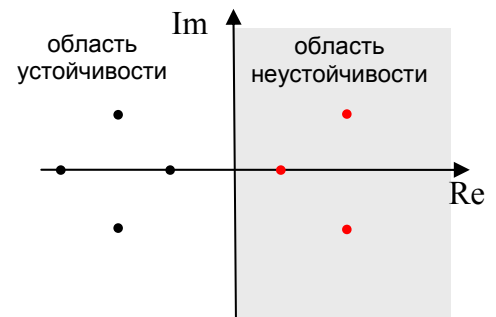
где $n_w(s)$ и $\Delta(s)$ – полиномы. Из теории линейных дифференциальных уравнений известно, что при отсутствии возмущений выход такой системы можно представить в виде:

$$y(t) = a_1 e^{\alpha_1 t} + a_2 e^{\alpha_2 t} + \dots + a_N e^{\alpha_N t}, \quad (47)$$

где a_i ($i = 1, \dots, N$) – постоянные, которые определяются начальными условиями. Таким образом, процесс $y(t)$ затухает при любых начальных условиях тогда и только тогда, когда все корни α_i ($i = 1, \dots, N$) имеют *отрицательные вещественные части*. В этом случае система *асимптотически устойчива*.

Поскольку устойчивость линейной системы определяют корни полинома $\Delta(s)$ – знаменателя передаточной функции $W(s)$, этот полином называется *характеристическим полиномом* системы.

Если показать корни характеристического полинома (в общем случае – комплексные числа¹⁰) на комплексной плоскости, то слева от мнимой оси будут *устойчивые корни* (с отрицательной вещественной частью), а справа – неустойчивые. Таким образом, область устойчивости –



¹⁰ Так как все коэффициенты полинома $\Delta(s)$ – вещественные, комплексные корни всегда будут парными, то есть, вместе с корнем $\alpha + j\beta$ всегда будет присутствовать $\alpha - j\beta$. На комплексной плоскости эти точки расположены симметрично относительно оси абсцисс. Здесь и далее $j = \sqrt{-1}$ – мнимая единица.

это левая полуплоскость.

Предположим, что один из корней полинома $\Delta(s)$ равен нулю (скажем, $\alpha_1 = 0$), а остальные *устойчивы*, то есть, их вещественные части отрицательные. Это значит, что система содержит интегрирующее звено. Учитывая, что $e^{\alpha_1 t} = e^0 = 1$ при всех t , получаем

$$y(t) = a_1 + a_2 e^{\alpha_2 t} + \dots + a_N e^{\alpha_N t}.$$

Здесь все слагаемые в правой части, кроме первого, затухают с течением времени, а постоянная составляющая a_1 остается. С другой стороны, выход не возрастает неограниченно, поэтому система *нейтрально устойчива*.

Теперь допустим, что характеристический полином имеет пару чисто мнимых корней: $\alpha_1 = j\omega$ и $\alpha_2 = -j\omega$. Это значит, что система содержит *консервативное звено* – генератор колебаний. При этом процесс (47) на выходе системы содержит слагаемые $a_1 e^{j\omega t}$ и $a_2 e^{-j\omega t}$, которые могут быть (с помощью формулы Эйлера) представлены в виде

$$a_1 e^{j\omega t} = a_1 (\cos \omega t + j \sin \omega t), \quad a_2 e^{-j\omega t} = a_2 (\cos \omega t - j \sin \omega t).$$

Эти составляющие дают незатухающие колебания (по крайней мере, для некоторых начальных условий), поэтому система находится на границе устойчивости (*нейтрально устойчива*). Заметим, что постоянные a_1 и a_2 – комплексно-сопряженные, то есть, если $a_1 = b + jc$, то $a_2 = b - jc$. При этом сумма

$$a_1 e^{j\omega t} + a_2 e^{-j\omega t} = 2b \cos \omega t - 2c \sin \omega t$$

не содержит мнимой части.

6.4.7. Внутренняя устойчивость линейных систем

В предыдущем параграфе мы фактически рассмотрели *техническую устойчивость*, то есть, устойчивость по выходу при ненулевых начальных условиях.

Теперь посмотрим, как определить *внутреннюю устойчивость* линейной системы, то есть, устойчивость внутренних процессов. Поскольку выход системы нас пока не интересует, используем модель «вход-состояние»:

$$\dot{x}(t) = Ax(t) + Bu(t),$$

где $x(t)$ – вектор состояния, $u(t)$ – входной сигнал, A и B – постоянные матрицы. Если вход равен нулю (нет возмущений), уравнение упрощается

$$\dot{x}(t) = Ax(t). \quad (48)$$

Таким образом, свободное движение определяется только свойствами матрицы A .

Сначала для простоты будем считать, что матрица A имеет вид $A = \begin{bmatrix} \alpha_1 & 0 \\ 0 & \alpha_2 \end{bmatrix}$. Тогда уравнение (48) распадается на два независимых уравнения (две подсистемы):

$$\dot{x}_1(t) = \alpha_1 x_1(t)$$

$$\dot{x}_2(t) = \alpha_2 x_2(t)$$

Здесь устойчивость определяется значениями α_1 и α_2 . Если они оба отрицательны, то система асимптотически устойчива. Если одно из них – нуль, а второе отрицательно (или оба нулевых), то система нейтрально устойчива.

В общем случае внутренняя устойчивость зависит от *собственных чисел* матрицы A , то есть, от корней *характеристического уравнения* $\det(\lambda I - A) = 0$, где I – единичная матрица, а «det» обозначает определитель квадратной матрицы. Полином $\det(\lambda I - A)$ от переменной λ называют *характеристическим полиномом*. Например, для рассмотренной выше диагональной матрицы A

$$\det(\lambda I - A) = \det \left(\lambda \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \begin{bmatrix} \alpha_1 & 0 \\ 0 & \alpha_2 \end{bmatrix} \right) = \det \begin{bmatrix} \lambda - \alpha_1 & 0 \\ 0 & \lambda - \alpha_2 \end{bmatrix} = (\lambda - \alpha_1)(\lambda - \alpha_2)$$

Очевидно, что корни этого полинома – это α_1 и α_2 .

Если все корни характеристического полинома устойчивы (имеют отрицательные вещественные части, расположены в левой полуплоскости), то система *асимптотически устойчива*. Если есть неустойчивые корни (с положительной вещественной частью), то система *неустойчива*. Если характеристический полином имеет один нулевой корень или пару комплексно-сопряженных корней на мнимой оси, система *нейтрально устойчива*¹¹.

Внутренняя устойчивость – более сильное требование, чем техническая устойчивость, потому что определяет ограниченность не только выхода, но и всех внутренних переменных при любых начальных условиях. Рассмотрим, например, такую модель в пространстве состояний

$$\begin{aligned} \dot{x}(t) &= \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} x(t) + \begin{bmatrix} 1 \\ 0 \end{bmatrix} u(t) \\ y(t) &= \begin{bmatrix} 0 & 1 \end{bmatrix} x(t) \end{aligned}$$

Здесь матрица $A = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$ имеет собственные числа 1 и -1 , причем первое из них – неустойчиво, поэтому система внутренне неустойчива.

Теперь найдем передаточную функцию (см. раздел 3.7):

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

Ее знаменатель (характеристический полином) $\Delta(s) = s+1$ устойчив, так как имеет единственный устойчивый корень -1 , хотя система внутренне неустойчива! Обратите внимание, что система имеет порядок 2, а знаменатель передаточной функции – порядок 1. В данном случае это означает, что некоторые внутренние движения системы *не наблюдаемы* на выходе, не влияют на него.

Вспомним, что передаточная функция описывает свойства системы только *при нулевых начальных условиях*. Поэтому выводы об устойчивости внутренних процессов в системе, сделанные по передаточной функции, могут оказаться неверными, если степень ее знаменателя меньше порядка исходного дифференциального уравнения.

6.4.8. Устойчивость линеаризованных систем

Устойчивость нелинейной системы можно во многих случаях оценивать с помощью линеаризованной системы. Для этого применяют **теоремы Ляпунова**, которые связывают корни характеристического полинома $\Delta(s)$ линейной модели и устойчивость нелинейной системы в окрестности точки линеаризации:

- 1) если все корни имеют отрицательные вещественные части, то нелинейная система также устойчива;
- 2) если есть хотя бы один корень с положительной вещественной частью, то нелинейная система неустойчива;
- 3) если нет корней с положительной вещественной частью, но есть хотя бы один корень с нулевой вещественной частью, то об устойчивости нелинейной системы ничего нельзя сказать без дополнительного исследования.

Таким образом, для исследования устойчивости положения равновесия нелинейной системы нужно линеаризовать модель в окрестности этой точки и найти корни характеристического полинома.

¹¹ Если есть несколько нулевых корней (или несколько одинаковых пар мнимых корней) система может быть как нейтрально устойчива, так и неустойчива.

6.5. Критерии устойчивости

Итак, для исследования устойчивости линейной системы достаточно найти корни ее характеристического полинома. Если все корни имеют отрицательные вещественные части (находятся в левой полуплоскости, слева от мнимой оси), такой полином называется *устойчивым*, потому что соответствующая линейная система устойчива. Полиномы, имеющие хотя бы один корень с положительной вещественной частью (в правой полуплоскости) называются *неустойчивыми*.

На ранней стадии развития теории управления актуальной была задача определения устойчивости полинома без вычисления его корней. Конечно, сейчас легко найти корни характеристического полинома с помощью компьютерных программ, однако такой подход дает нам только количественные (а не качественные) результаты и не позволяет исследовать устойчивость теоретически, например, определять границы областей устойчивости.

6.5.1. Критерий Гурвица

Существует несколько алгоритмов, позволяющих проверить устойчивость полинома

$$\Delta(s) = a_0 s^n + a_1 s^{n-1} + \dots + a_{n-1} s + a_n,$$

не вычисляя его корни. Прежде всего, для устойчивости все коэффициенты a_i ($i = 0, \dots, n$) должны быть одного знака, обычно считают, что они положительные. Это *необходимое условие устойчивости полинома*. Однако при $n > 2$ это условие недостаточно, если полином имеет комплексно-сопряженные корни. Поэтому были разработаны *необходимые и достаточные* условия (критерии) устойчивости полиномов.

Один из самых известных критериев – **критерий Гурвица** – использует матрицу H_n размером $n \times n$, составленную из коэффициентов полинома $\Delta(s)$ следующим образом:

- первая строка содержит коэффициенты a_1, a_3, a_5, \dots (все с нечетными номерами), оставшиеся элементы заполняются нулями;
- вторая строка содержит коэффициенты a_0, a_2, a_4, \dots (все с четными номерами);
- третья и четвертая строка получаются сдвигом первой и второй строк на 1 позицию вправо, и т.д.

Например, для полинома пятого порядка ($n = 5$) эта матрица имеет вид

$$H_5 = \begin{bmatrix} a_1 & a_3 & a_5 & 0 & 0 \\ a_0 & a_2 & a_4 & 0 & 0 \\ 0 & a_1 & a_3 & a_5 & 0 \\ 0 & a_0 & a_2 & a_4 & 0 \\ 0 & 0 & a_1 & a_3 & a_5 \end{bmatrix} \quad (a_0 > 0)$$

Критерий Гурвица. Все корни полинома $\Delta(s)$ имеют отрицательные вещественные части тогда и только тогда, когда все n главных миноров матрицы H_n (*определителей Гурвица*) положительны.

Вспомним, что для устойчивости полинома *необходимо*, чтобы все его коэффициенты были положительными. Поэтому достаточно проверить только $n-1$ первых определителей Гурвица. Например, для $n = 5$ речь идет об определителях

$$D_1 = a_1 > 0, \quad D_2 = \begin{vmatrix} a_1 & a_3 \\ a_0 & a_2 \end{vmatrix} > 0, \quad D_3 = \begin{vmatrix} a_1 & a_3 & a_5 \\ a_0 & a_2 & a_4 \\ 0 & a_1 & a_3 \end{vmatrix} > 0, \quad D_4 = \begin{vmatrix} a_1 & a_3 & a_5 & 0 \\ a_0 & a_2 & a_4 & 0 \\ 0 & a_1 & a_3 & a_5 \\ 0 & a_0 & a_2 & a_4 \end{vmatrix} > 0.$$

Раскрывая определитель матрицы H_5 по последнему столбцу, получаем $D_5 = \det H_5 = a_5 D_4$. Так как $a_5 > 0$, из условия $D_4 > 0$ сразу следует $D_5 > 0$.

Таким образом, условия устойчивости сводятся к нескольким неравенствам. Это очень удобно для систем низкого порядка. Например, для $n = 2$ необходимое и достаточное условие устойчивости – положительность всех коэффициентов полинома. Для $n = 3$ характеристический полином имеет вид $\Delta(s) = a_0 s^3 + a_1 s^2 + a_2 s + a_3$, поэтому условия Гурвица определяются матрицей

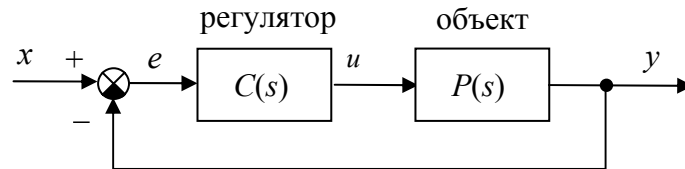
$$H_3 = \begin{bmatrix} a_1 & a_3 & 0 \\ a_0 & a_2 & 0 \\ 0 & a_1 & a_3 \end{bmatrix} \quad (a_0 > 0).$$

Полином устойчив, если все коэффициенты положительны и

$$D_2 = \begin{vmatrix} a_1 & a_3 \\ a_0 & a_2 \end{vmatrix} = a_1 a_2 - a_0 a_3 > 0. \quad (49)$$

Рассмотрим систему, в которой объект и регулятор задаются передаточными функциями:

$$P(s) = \frac{1}{(T_1 s + 1)(T_2 s + 1)}, \quad C(s) = \frac{K}{s}.$$



С помощью критерия Гурвица можно определить, при каких значениях K замкнутая система (с отрицательной обратной связью) устойчива. Передаточная функция замкнутой системы равна

$$W(s) = \frac{C(s)P(s)}{1 + C(s)P(s)} = \frac{K}{\Delta(s)},$$

где характеристический полином имеет вид

$$\Delta(s) = (T_1 s + 1)(T_2 s + 1)s + K = T_1 T_2 s^3 + (T_1 + T_2)s^2 + s + K.$$

Необходимое условие устойчивости дает $K > 0$. Применяя критерий Гурвица для системы третьего порядка (49), получаем

$$T_1 + T_2 > K T_1 T_2 \Rightarrow K < \frac{1}{T_1} + \frac{1}{T_2}.$$

Таким образом, система устойчива при $0 < K < \frac{1}{T_1} + \frac{1}{T_2}$.

Теперь предположим, что модель системы задана в пространстве состояний:

$$\dot{x}(t) = A \cdot x(t) + B \cdot u(t)$$

$$y(t) = C \cdot x(t) + D \cdot u(t)$$

Как проверить ее устойчивость? Используя результаты раздела 3.7, построим передаточную функцию

$$W(s) = C(sI - A)^{-1}B + D.$$

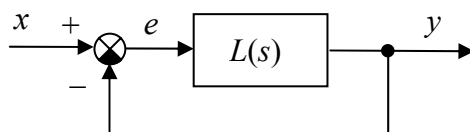
Характеристический полином этой системы (знаменатель $W(s)$), определяется формулой

$$\Delta(s) = \det(sI - A),$$

где \det обозначает определитель квадратной матрицы. Чтобы определить, устойчива ли система, нужно применить к этому полиному критерий Гурвица.

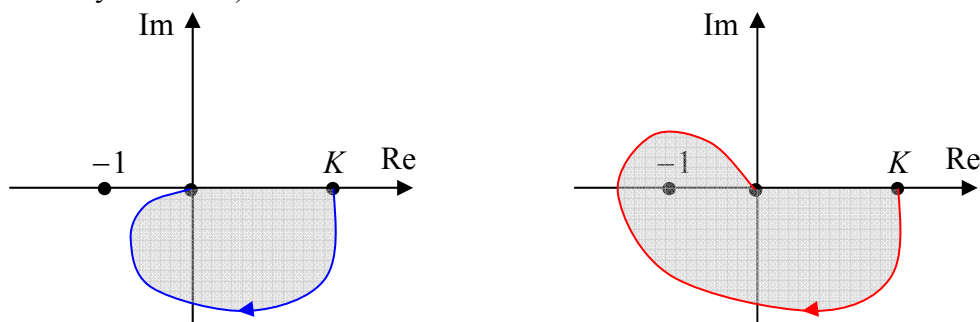
6.5.2. Критерий Найквиста

Критерий Найквиста позволяет определить устойчивость замкнутой системы, построив частотную характеристику *разомкнутой* системы. Пусть $L(s)$ – передаточная функция разомкнутой системы, а $L(j\omega)$ – ее частотная характеристика.

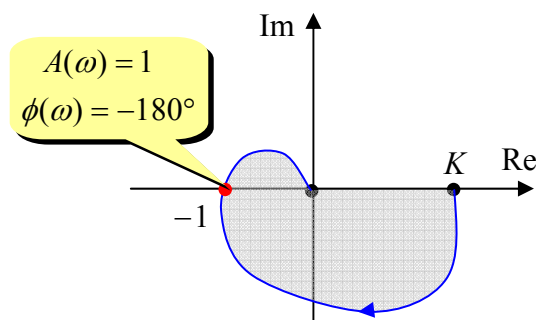


Для простоты сначала будем считать, что разомкнутая система устойчива и не содержит интегрирующих звеньев, то есть $L(0) = K \neq \infty$, где K – некоторое число.

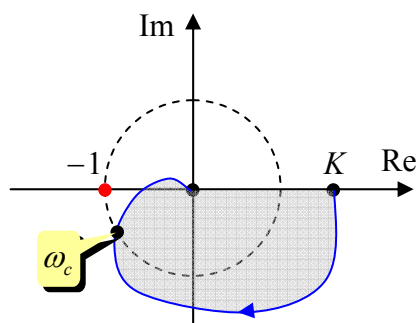
Для каждой частоты ω значение $L(j\omega)$ – это комплексное число, которое можно изобразить точкой на комплексной плоскости. При изменении частоты от 0 до ∞ из этих точек складывается *годограф Найквиста* – некоторая кривая, которая начинается в точке $(K; 0)$ на вещественной оси и заканчивается в начале координат (если $L(s)$ – строго правильная функция, то есть степень ее числителя меньше степени знаменателя). Можно доказать, что система устойчива тогда и только тогда, когда годограф $L(j\omega)$ не охватывает точку $(-1; 0)$. На рисунке слева годограф не охватывает эту точку (и замкнутая система устойчива), а на рисунке справа – охватывает (система неустойчива).



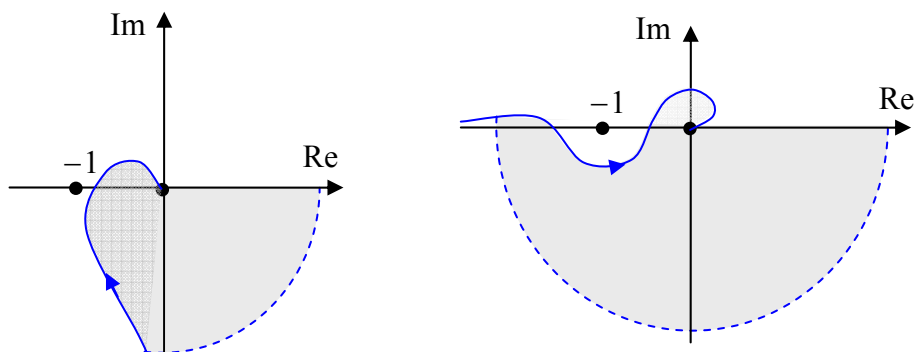
Выражение «система находится на границе устойчивости» означает, что частотная характеристика проходит через точку $(-1; 0)$. В этом случае для некоторой частоты ω мы имеем $A(\omega) = 1$ и $\phi(\omega) = -180^\circ$. Это говорит о том, что после прохождения контура величина сигнала меняет знак, сохраняя абсолютную величину (энергию), то есть устанавливаются незатухающие колебания.



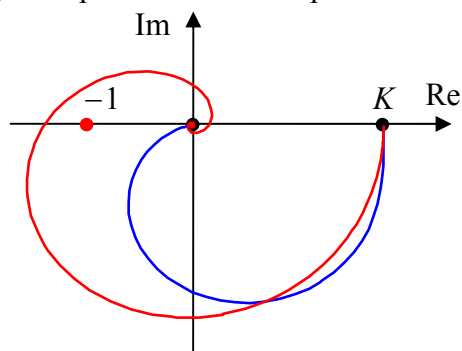
Частота ω_c , для которой $A(\omega_c) = 1$, называется **частотой среза**. Для устойчивой системы значение фазы на частоте среза должно быть больше, чем -180° ; в этом случае годограф не охватит точку $(-1; 0)$.



Если передаточная функция $L(s)$ имеет полюса в точке $s = 0$ (то есть обращается в бесконечность в этой точке), ситуация усложняется. Теперь годограф начинается не на вещественной оси, а приходит из бесконечности. Тогда в контур необходимо включить не только полученную кривую, но и часть окружности бесконечного радиуса от вещественной оси до годографа в порядке обхода по часовой стрелке. Если функция $L(s)$ имеет k полюсов в точке $s = 0$, нужно добавить k секторов по 90° . На рисунках показаны годографы Найквиста устойчивых систем, в которых функция $L(s)$ имеет соответственно 1 и 2 полюса в точке $s = 0$. Эти годографы не охватывают точку $(-1; 0)$.



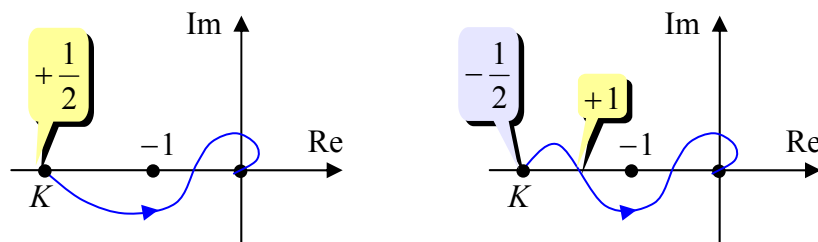
Если в системе есть запаздывание на время τ , на любой частоте появляется дополнительный сдвиг фазы на $-\tau\omega$ (без изменения амплитуды). Это значит, что каждая точка годографа поворачивается на некоторый угол против часовой стрелки.



На рисунке синяя линия – частотная характеристика системы без запаздывания, а красная – аналогичная характеристика для системы с запаздыванием. Видно, что запаздывание привело к неустойчивости системы (годограф охватил критическую точку $(-1; 0)$). Таким образом, система может потерять устойчивость из-за «медленного» датчика. Можно говорить о том, что *запаздывание всегда ухудшает устойчивость системы*, и этот факт важно учитывать при проектировании.

Если $L(s)$ имеет полюса с положительной вещественной частью (разомкнутая система неустойчива), нужно считать, сколько раз годограф пересекает ось абсцисс левее точки $(-1; 0)$. Причем переходы «сверху вниз» считаются положительными, а переходы «снизу вверх» – отрицательными.

Для того, чтобы замкнутая система была устойчива, необходимо и достаточно, чтобы разница между числом положительных и отрицательных переходов была равна $\ell/2$, где ℓ – число неустойчивых полюсов функции $L(s)$. Начальная точка на оси абсцисс левее точки $(-1; 0)$ считается за половину перехода. На рисунке показаны годографы устойчивых систем для случая $\ell = 1$.

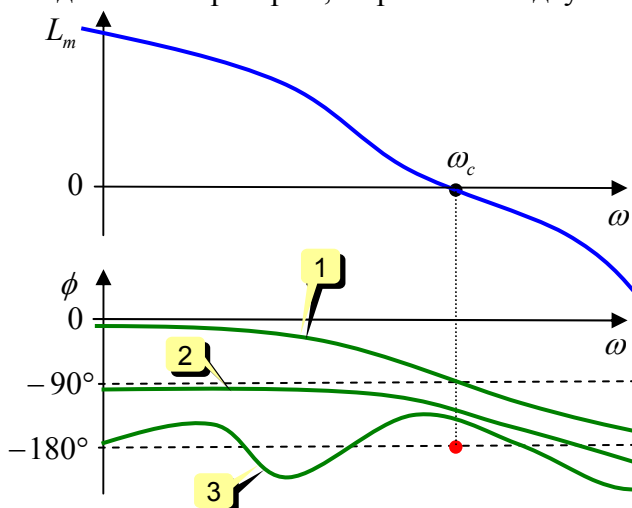


Частотная характеристика начинается на вещественной оси левее точки $(-1; 0)$. На рисунке слева годограф сначала идет вниз (половина положительного перехода) и больше нигде не пересекает ось абсцисс левее точки $(-1; 0)$, поэтому разница переходов равна $1/2 = \ell/2$ и замкнутая система устойчива.

На правом рисунке частотная характеристика сначала идет вверх (считаем это за половину отрицательного перехода), а затем переходит в нижнюю полуплоскость (положительный переход). Разница снова равна $1/2 = \ell/2$ и система устойчива.

6.5.3. Критерий Найквиста для ЛАФЧХ

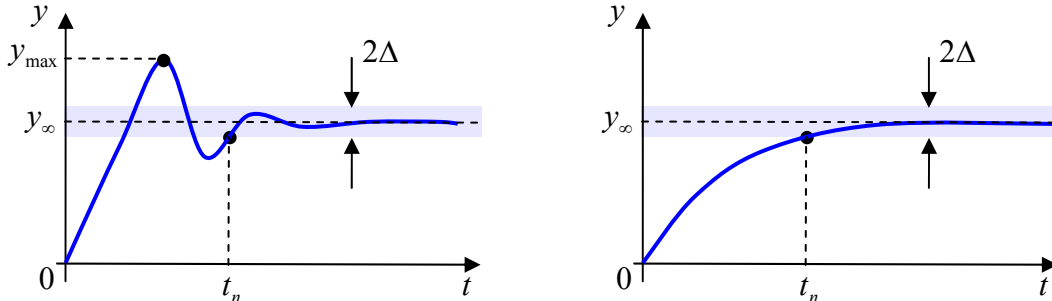
Критерий Найквиста часто используется для логарифмических частотных характеристик. Сначала предположим, что передаточная функция разомкнутой системы не имеет неустойчивых полюсов. Как мы уже знаем, для анализа устойчивости наиболее важно поведение частотной характеристики в районе частоты среза ω_c , где $A(\omega_c) = 1$ и $L_m(\omega_c) = 20 \lg A(\omega_c) = 0$. Для устойчивой системы значение фазы на частоте среза должно быть больше, чем -180° . На графике представлены три фазовых характеристики устойчивых систем. Кривая 1 соответствует случаю, когда в разомкнутой системе нет интеграторов (и фазовая характеристика начинается с нуля), кривая 2 – системе с одним интегратором, а кривая 3 – с двумя.



Если разомкнутая система имеет неустойчивые звенья, нужно считать переходы фазовой характеристики через линию $\phi(\omega) = -180^\circ$ левее частоты среза. Здесь положительным считается переход снизу вверх, а отрицательным – сверху вниз. Если фазовая характеристика начинается на линии $\phi(\omega) = -180^\circ$ (на нулевой частоте), это считается за половину перехода. Для устойчивой системы разность между числом положительных и отрицательных переходов должна быть равна $\ell/2$, где ℓ – число неустойчивых полюсов передаточной функции $L(s)$.

6.6. Переходный процесс

Хорошо спроектированная система должна не только быть устойчивой и поддерживать заданную точность в установившемся режиме, но и плавно переходить на новый режим при изменении заданного значения выхода (*уставки*). Качество переходных процессов обычно оценивается по переходной характеристике (реакции системы на единичный ступенчатый входной сигнал).



В первую очередь нас интересует, насколько быстро заканчивается переход на другой режим (время переходного процесса t_n). Оно определяется как время, через которое регулируемая величина «входит в коридор» шириной 2Δ вокруг установившегося значения y_∞ . Это значит, что при $t > t_n$ значение выхода отличается от установившегося не более, чем на Δ . Обычно величина Δ задается в процентах от установившегося значения, чаще всего 2% или 5%. Заметим, что для апериодического звена с постоянной времени T время переходного процесса равно $t_n = 3T$ (с точностью 5%).

Другая важная характеристика – **перерегулирование** σ – показывает, на сколько процентов максимальное значение выхода y_{\max} превышает установившееся значение¹² y_∞ :

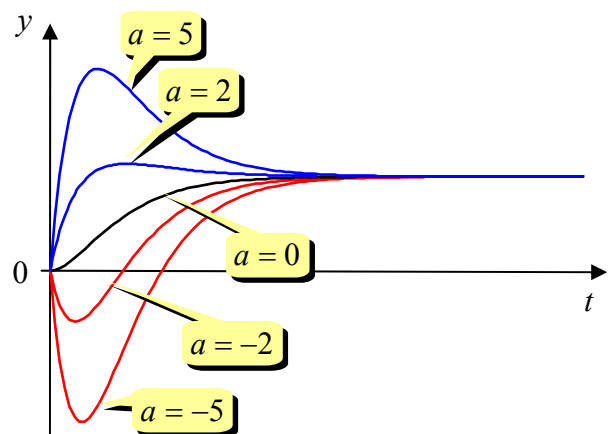
$$\sigma = \frac{y_{\max} - y_\infty}{y_\infty} \cdot 100\%.$$

Иногда удастся обеспечить нулевое перерегулирование (*апериодический* переходный процесс, как у апериодического звена). Нужно помнить, что увеличение быстродействия обычно приводит к увеличению перерегулирования.

Вы уже знаете, что устойчивость линейной системы определяется полюсами ее передаточной функции $W(s)$, однако на переходные процесс влияют и *нули*, причем в некоторых случаях очень существенно. Для примера рассмотрим передаточную функцию

$$W(s) = \frac{as + 1}{(s + 1)^2} = \frac{a(s + 1/a)}{(s + 1)^2},$$

где a может принимать как положительные, так и отрицательные значения. Такая передаточная функция имеет нуль в точке $s = -1/a$. Нули, находящиеся в левой полуплоскости (при $a > 0$) часто называют *устойчивыми* (по аналогии с полюсами), а нули в правой полуплоскости (при $a < 0$) – *неустойчивыми*. Очевидно, что при $a = 0$ мы получаем апериодическое звено второго порядка. Теперь построим переходные характеристики этого звена при разных значениях a . Заметим, что при любом a установившееся значение выхода равно $W(0) = 1$.



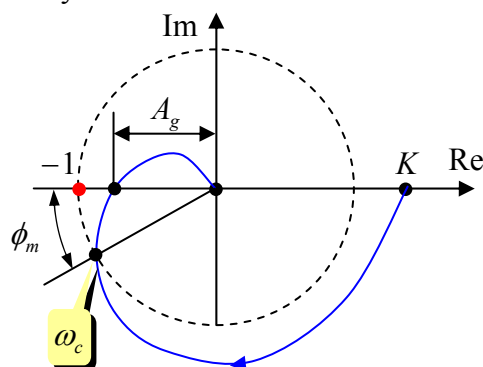
¹² Понятие «перерегулирование» обычно вводится для случая, когда установившееся значение выхода больше нуля, хотя, в принципе, оно может быть и отрицательным – тогда перерегулирование показывает, насколько «ниже» установившегося значения ушла переходная функция в точке минимума.

По графикам видно, что при нулевом значении a переходный процесс – аperiodический. При $a > 0$ (устойчивый нуль) наблюдается перерегулирование, причем оно тем больше, чем больше модуль a . При отрицательных значениях a в переходном процессе есть *недорегулирование*. Это значит, что в первый момент времени регулируемая переменная начинает изменяться в сторону, противоположную заданному значению.

6.7. Частотные оценки качества

Качество системы можно оценивать не только во временной области (переходный процесс во времени), но и в частотной (по частотной характеристике). Из частотных оценок наиболее важны **запасы устойчивости**. Дело в том, что поведение реального объекта всегда несколько отличается от принятой модели, более того, динамика может меняться во времени, например, когда корабль расходует топливо в ходе рейса. Поэтому недостаточно спроектировать просто устойчивую систему, нужно, чтобы система сохранила устойчивость при некоторых изменениях параметров объекта и регулятора в сравнении с расчетными, то есть, обладала запасами устойчивости.

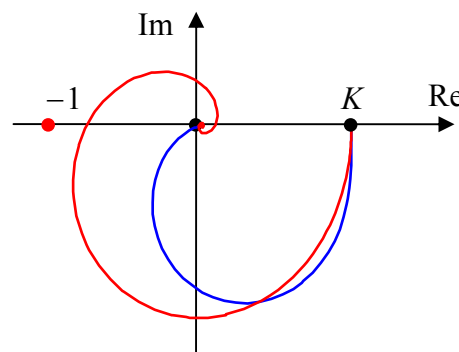
Обычно рассматривают запасы устойчивости по амплитуде и по фазе. **Запас устойчивости по амплитуде** g_m – это дополнительное усиление контура, которое необходимо, чтобы вывести систему на границу области устойчивости. Эта величина измеряется в децибелах.



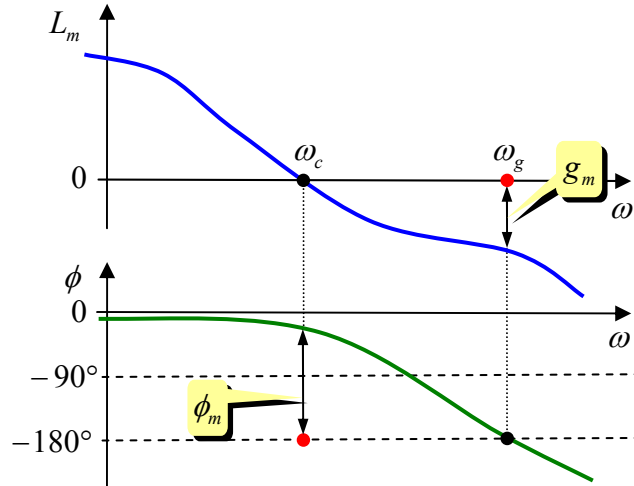
Запас по амплитуде вычисляется по формуле $g_m = 20 \lg \frac{1}{A_g}$, где $A_g < 1$ – значение амплитудной характеристики на частоте ω_g , где фазовая характеристика равна -180° . В практических задачах нужно обеспечивать запас по амплитуде не менее 6 дБ.

Запас устойчивости по фазе ϕ_m – это дополнительный сдвиг фазы («поворот» частотной характеристики против часовой стрелки), который необходим для того, чтобы вывести систему на границу устойчивости. Он определяется на *частоте среза* ω_c , где $A(\omega_c) = 1$. Запас по фазе должен быть не менее 30° .

Если в системе есть запаздывание на время τ , каждая точка годографа частотной характеристики дополнительно поворачивается против часовой стрелки на угол, равный $\tau\omega$ для частоты ω . Поэтому запасы устойчивости (как по амплитуде, так и по фазе) уменьшаются. На рисунке синяя линия соответствует системе без запаздывания, а красная – той же системе с запаздыванием. Видно, что во втором случае запасы устойчивости существенно меньше.

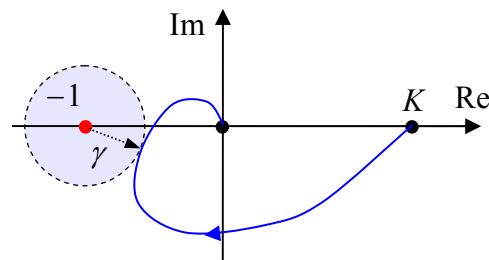


Запасы устойчивости легко определяются по логарифмическим частотным характеристикам:

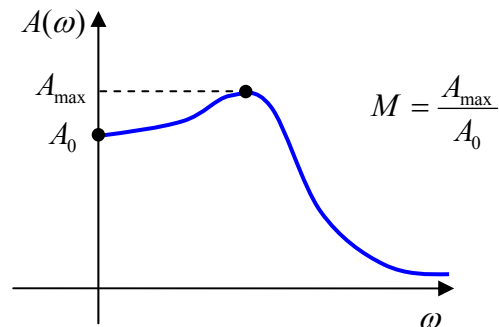


Заметим, что запас по амплитуде может быть равен бесконечности, если фазовая характеристика не пересекает линию -180° .

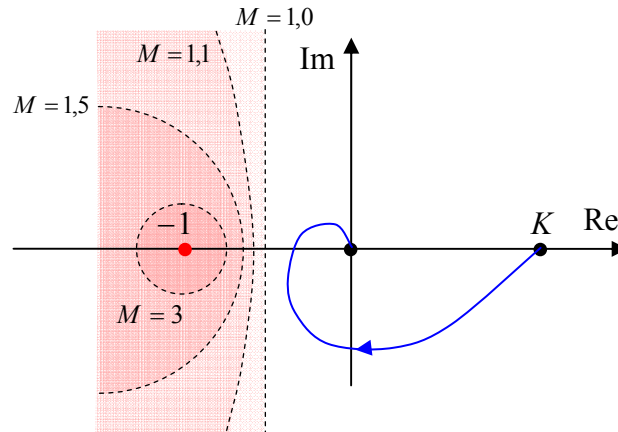
К сожалению, в некоторых случаях классические запасы устойчивости (по амплитуде и фазе) дают не совсем верное представление о том, насколько система действительно близка к границе устойчивости. Поэтому в качестве единой характеристики иногда используют кратчайшее расстояние γ от годографа до точки $(-1; 0)$.



Еще одна аналогичная характеристика называется **показателем колебательности** M . Она определяется по амплитудной частотной характеристике *замкнутой системы* как отношение ее максимума к значению на нулевой частоте:



Для каждого значения M можно нарисовать «запретную области», в которую не должна заходить частотная характеристика *разомкнутой системы*, если ее показатель колебательности должен быть меньше M . Эта область имеет форму круга радиуса $R = \frac{M}{M^2 - 1}$, центр которого находится в точке $\left(-\frac{M^2}{M^2 - 1}; 0\right)$. На рисунке показаны границы запретных областей для различных значений M .

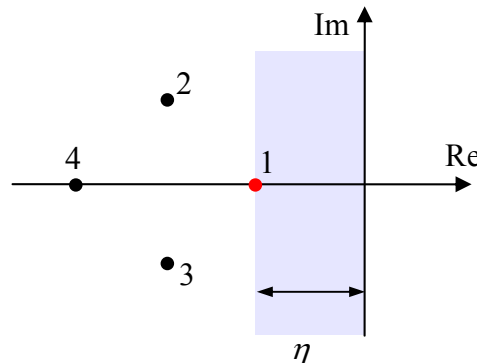


При $M = 1$ окружность имеет бесконечный радиус (превращается в вертикальную линию) и проходит через точку $(-0,5; 0)$. При увеличении M радиус окружности уменьшается.

6.8. Корневые оценки качества

Многие свойства системы можно предсказать, посмотрев на расположение корней характеристического полинома $\Delta(s)$ на комплексной плоскости. Прежде всего, все корни $\Delta(s)$ для устойчивой системы должны находиться в левой полуплоскости, то есть слева от мнимой оси. Быстродействие системы определяется **степенью устойчивости** η – так называется расстояние мнимой оси до ближайшего корня (или пары комплексно-сопряженных корней).

На рисунке точками отмечены положения корней характеристического полинома. Он имеет два вещественных корня (обозначенных номерами 1 и 4) и пару комплексно сопряженных корней (2 и 3). Степенью устойчивости определяется вещественным корнем 1, потому что он находится ближе всех к мнимой оси.



Этот корень называется *доминирующим*, он определяет самые медленные движения в системе и время переходного процесса, которое может быть примерно рассчитано по формуле $t_n = \frac{3}{\eta}$.

Корни 2, 3 и 4 соответствуют более быстрым движениям.

Обратите внимание, что степень устойчивости, несмотря на название, ничего не говорит о близости системы к границе устойчивости, она только характеризует быстродействие.

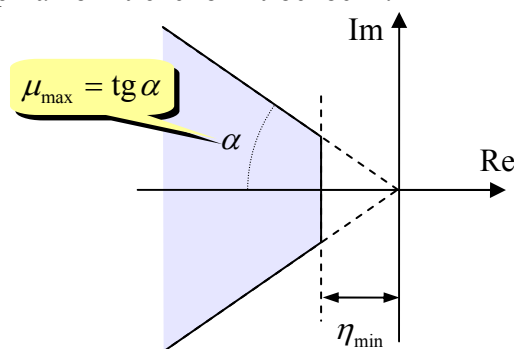
Параметр, определяющий скорость затухания колебаний в системе, называется *колебательностью*. Колебательность μ для пары комплексно-сопряженных корней $\alpha \pm j\beta$ вычисляется как отношение мнимой и вещественной частей корня (по модулю):

$$\mu = \left| \frac{\beta}{\alpha} \right|.$$

Чем больше эта величина, тем слабее затухают колебания, вызванные этими корнями, за 1 период колебаний.

Линии постоянной колебательности – это лучи, выходящие из начала координат. При проектировании систем обычно требуется обеспечить быстродействие не ниже заданного (степень

устойчивости не меньше заданной η_{\min}) и колебательность не выше заданной μ_{\max} . Эти условия определяют усеченный сектор на комплексной плоскости.



6.9. Робастность

6.9.1. Что такое робастность?

Обычно регулятор строится на основе некоторых приближенных (*номинальных*) моделей объекта управления (а также приводов и датчиков) и внешних возмущений. При этом поведение реального объекта и характеристики возмущений могут быть несколько иными. Поэтому требуется, чтобы разработанный регулятор обеспечивал устойчивость и приемлемое качество системы при малых отклонениях свойств объекта и внешних возмущений от номинальных моделей. В современной теории управления это свойство называют *робастностью* (грубостью). Иначе его можно назвать нечувствительностью к малым ошибкам моделирования объекта и возмущений.

Различают несколько задач, связанных с робастностью:

- **робастная устойчивость** – обеспечить *устойчивость* системы при всех допустимых отклонениях модели *объекта* от номинальной;
- **робастное качество** – обеспечить *устойчивость* и заданные *показатели качества* системы при всех допустимых отклонениях модели *объекта* от номинальной;
- **гарантирующее управление** – обеспечить заданные показатели качества системы при всех допустимых отклонениях модели *возмущения* от номинальной (считая, что модель объекта известна точно).

Для того, чтобы исследовать робастность системы, нужно как-то определить возможную ошибку моделирования (неопределенность). Ее можно задать различными способами.

6.9.2. Параметрическая неопределенность

Параметрическая неопределенность означает, что структура модели известна, а параметры могут отличаться от номинальных, например:

$$P(s) = \frac{k_0 + \varepsilon_1}{(T_0 + \varepsilon_2)s + 1},$$

где k_0 и T_0 – номинальные значения коэффициента усиления и постоянной времени, а ε_1 и ε_2 – малые ошибки моделирования.

Предположим, что такой объект управляется регулятором-усилителем с передаточной функцией $C(s) = K$. Тогда характеристический полином замкнутой системы принимает вид

$$\Delta(s) = (T_0 + \varepsilon_2)s + 1 + K(k_0 + \varepsilon_1).$$

Робастный регулятор должен обеспечивать устойчивость этого полинома при всех допустимых ε_1 и ε_2 . В данном случае условия устойчивости сводятся к тому, что коэффициенты полинома, $T_0 + \varepsilon_2$ и $1 + K(k_0 + \varepsilon_1)$, имеют одинаковый знак (оба положительные или оба отрицательные). Будем считать, что $k_0 > 0$ и $T_0 > 0$, а отклонения ε_1 и ε_2 малы в сравнении с k_0 и T_0 соответ-

венно. Таким образом, $T_0 + \varepsilon_2 > 0$ при всех возможных ε_2 . Следовательно, замкнутая система устойчива при

$$1 + K(k_0 + \varepsilon_1) > 0 \Rightarrow K > \frac{-1}{k_0 + \varepsilon_1}.$$

Наибольшее значение в правой части последнего неравенства будет при максимальном значении ε_1 , поэтому условие робастной устойчивости принимает вид

$$K > K_{\min} = \frac{-1}{k_0 + \varepsilon_{1\max}}.$$

Таким образом, любой регулятор-усилитель, имеющий коэффициент усиления $K > K_{\min}$, обеспечивает робастную устойчивость системы в том смысле, что устойчивость сохраняется при всех допустимых ошибках ε_1 и ε_2 .

В более сложных случаях часто используют *теорему Харитонова*, которая позволяет проверить робастную устойчивость характеристического полинома

$$\Delta(s) = a_0 + a_1 s + \dots + a_n s^{n-1} + a_n s^n,$$

где коэффициенты a_0, a_1, \dots, a_n точно неизвестны, но принадлежат интервалам

$$\ell_i < a_i < u_i \quad (i = 1, \dots, n).$$

Оказывается, полином $\Delta(s)$ устойчив при всех возможных значениях коэффициентов тогда и только тогда, когда устойчивы четыре полинома Харитонова:

$$\Delta_1(s) = \ell_0 + \ell_1 s + u_2 s^2 + u_3 s^3 + \ell_4 s^4 + \ell_5 s^5 + \dots$$

$$\Delta_2(s) = u_0 + u_1 s + \ell_2 s^2 + \ell_3 s^3 + u_4 s^4 + u_5 s^5 + \dots$$

$$\Delta_3(s) = \ell_0 + u_1 s + u_2 s^2 + \ell_3 s^3 + \ell_4 s^4 + u_5 s^5 + \dots$$

$$\Delta_4(s) = u_0 + \ell_1 s + \ell_2 s^2 + u_3 s^3 + u_4 s^4 + \ell_5 s^5 + \dots$$

Таким образом, для проверки устойчивости *бесконечного* числа возможных характеристических полиномов достаточно проверить устойчивость *четырёх* полиномов Харитонова.

6.9.3. Непараметрическая неопределенность

Непараметрическая неопределенность задает допустимую ошибку в частотной области, то есть ошибку в частотных характеристиках. Для номинальной модели $P_0(j\omega)$ различают *аддитивную* неопределенность (абсолютную ошибку) $\Delta_a(j\omega)$:

$$P(j\omega) = P_0(j\omega) + \Delta_a(j\omega)$$

и *мультипликативную* неопределенность (относительную ошибку) $\Delta_m(j\omega)$:

$$P(j\omega) = [1 + \Delta_m(j\omega)] P_0(j\omega).$$

Для мультипликативной неопределенности известен очень простой **критерий робастной устойчивости**: система с регулятором $C(s)$ и номинальный объектом $P_0(s)$ робастно устойчива, если для любой частоты ω выполняется неравенство

$$|W_0(j\omega) \Delta_m(j\omega)| < 1, \quad (50)$$

где $W_0(s)$ – передаточная функция номинальной замкнутой системы:

$$W_0(s) = \frac{C(s) P_0(s)}{1 + C(s) P_0(s)}.$$

Этот результат называется *теоремой о малом коэффициенте усиления*. При этом также требуется, чтобы реальная и номинальная модели объекта, $P(s)$ и $P_0(s)$, имели одинаковые неустойчивые полюса, то есть неопределенность не должна вносить новые источники неустойчивости.

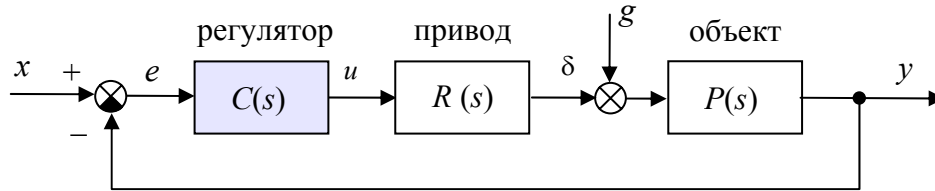
Условие (50) – это *достаточное* условие робастной устойчивости, то есть, его выполнение гарантирует устойчивость, но для некоторых робастно устойчивых систем оно может не выполняться.

Обычно модель строится так, чтобы хорошо описывать свойства реального объекта на низких частотах, а для высоких частот ошибка моделирования $\Delta_m(j\omega)$ может быть значительной. Тогда, учитывая (50), можно сделать вывод, что с точки зрения робастной устойчивости значение $|W_0(j\omega)|$ должно быть мало на высоких частотах, где велика неопределенность модели.

7. Синтез регуляторов

7.1. Классическая схема

Чаще всего регулятор включается перед объектом, как показано на схеме:



Задача системы управления состоит в том, чтобы подавить действие внешнего возмущения g и обеспечить быстрые и качественные переходные процессы. К сожалению, эти задачи часто противоречивы. Фактически нам нужно скорректировать систему так, чтобы она имела нужные передаточные функции по возмущению ($W_g(s)$, от входа g к выходу y) и по задающему воздействию ($W(s)$, от входа x к выходу y)

$$W_g(s) = \frac{P(s)}{1 + C(s)R(s)P(s)}, \quad W(s) = \frac{C(s)R(s)P(s)}{1 + C(s)R(s)P(s)}.$$

Для этого мы можем использовать только *один* регулятор $C(s)$, поэтому такую систему называют системой *с одной степенью свободы*.

Легко проверить, что эти две передаточные функции связаны равенством

$$W_g(s) = [1 - W(s)]P(s).$$

Поэтому, изменяя одну из передаточных функций, мы автоматически меняем и вторую. Таким образом, их невозможно сформировать независимо и решение всегда будет некоторым *компромиссом*.

Посмотрим, можно ли в такой системе обеспечить нулевую ошибку, то есть, абсолютно точное отслеживание входного сигнала. Передаточная функция по ошибке (от входа $x(t)$ к ошибке $e(t)$) равна

$$W_e(s) = \frac{1}{1 + C(s)R(s)P(s)}.$$

Для того, чтобы ошибка *всегда* была нулевой, требуется, чтобы эта передаточная функция была равна нулю. Поскольку ее числитель – не нуль, сразу получаем, что знаменатель должен обращаться в бесконечность. Мы можем влиять только на регулятор $C(s)$ (остальные элементы заданы заранее), поэтому получаем $C(s) \rightarrow \infty$. Таким образом, для *уменьшения ошибки нужно увеличивать коэффициент усиления регулятора*. Это так называемый принцип *глубокой обратной связи*.

Однако нельзя увеличивать усиление до бесконечности. Во-первых, все реальные устройства имеют предельно допустимые значения входных и выходных сигналов. Во-вторых, при большом усилении контура ухудшается качество переходных процессов, усиливается влияние возмущений и шумов, система может потерять устойчивость. Поэтому в схеме с одной степенью свободы обеспечить нулевую ошибку слежения невозможно.

Посмотрим на задачу с точки зрения частотных характеристик. С одной стороны, для качественного отслеживания задающего сигнала $x(t)$ желательно, чтобы частотная характеристика $W(j\omega)$ была примерно равна 1 (в этом случае $y(t) \approx x(t)$). С другой стороны, с точки зрения робастной устойчивости нужно обеспечить $W(j\omega) \approx 0$ на высоких частотах, где ошибка моделирования велика. Кроме того, передаточная функция по возмущению должна быть такой, чтобы эти возмущения подавлять, в идеале мы должны обеспечить $W_g(j\omega) \approx 0$.

Выбирая компромиссное решение, обычно поступают следующим образом:

- 1) на *низких* частотах добиваются выполнения условия $W(j\omega) \approx 1$, что обеспечивает хорошее слежение за низкочастотными сигналами; при этом $W_g(j\omega) \approx 0$, то есть, низкочастотные возмущения подавляются;
- 2) на *высоких* частотах стремятся сделать $W(j\omega) \approx 0$, чтобы обеспечить робастную устойчивость и подавление шума измерений; при этом $W_g(j\omega) \approx P(j\omega)$, то есть система фактически работает как разомкнутая, регулятор не реагирует на высокочастотные помехи.

7.2. ПИД-регуляторы

Несмотря на развитые современные методы проектирования сложных регуляторов, подавляющее большинство промышленных систем управления основаны на регуляторах первого и второго порядка. Эти регуляторы во многих случаях могут обеспечить приемлемое управление, легко настраиваются и дешевы при массовом изготовлении.

Простейший регулятор – пропорциональный или **П-регулятор** – это простой усилитель с передаточной функцией $C(s) = K$. Его выход – это *ошибка управления* $e(t)$, умноженная на коэффициент K . С помощью П-регулятора можно управлять любым устойчивым объектом, однако он дает относительно медленные переходные процессы и ненулевую статическую ошибку.

Чтобы убрать статическую ошибку в установившемся режиме, в регулятор вводят *интегральный канал* с коэффициентом усиления K_I , так что

$$C(s) = K + \frac{K_I}{s}, \quad u(t) = Ke(t) + K_I \int_0^t e(t) dt.$$

Такой регулятор называется пропорционально-интегральным или **ПИ-регулятором**. Интегратор выдает сигнал, пропорциональный *накопленной* ошибке, поэтому переходный процесс несколько замедляется. Однако за счет интегрального канала обеспечивается нулевая ошибка в установившемся состоянии при ступенчатом возмущении и ступенчатом изменении задающего сигнала-уставки.

Для ускорения переходных процессов добавляют *дифференциальный канал* с коэффициентом усиления K_D :

$$C(s) = K + \frac{K_I}{s} + K_D s, \quad u(t) = Ke(t) + K_I \int_0^t e(t) dt + K_D \frac{de(t)}{dt}.$$

Такой регулятор называется **ПИД-регулятором** (пропорционально-интегрально-дифференциальный). Регуляторы этого типа очень хорошо зарекомендовали себя в практических задачах. Кроме того, иногда используются **ПД-регуляторы** (пропорционально-дифференциальные), у которых нет интегрального канала.

Управление по производной – это быстрый способ управления. Сигнал дифференциального канала наиболее важен при изменениях входов и исчезает в установившемся режиме. Он позволяет реагировать не на само увеличение ошибки, а на *тенденцию* ее изменения, и принять «превентивные меры». Главный недостаток дифференциального канала – большое влияние высокочастотных помех, например, шумов измерений. Для того, чтобы сделать регулятор физически реализуемым, вместо чистого дифференцирования используют инерционное дифференцирующее звено:

$$C(s) = K + \frac{K_I}{s} + \frac{K_D s}{T_D s + 1},$$

где T_D – малая постоянная времени. Чем меньше T_D , тем в большем частотном диапазоне выполняется точное дифференцирование, но сильнее влияют высокочастотные помехи.

Для устойчивого объекта можно выбрать коэффициенты регулятора опытным путем, выполняя эксперименты с реальным объектом. Предложено несколько методов решения этой задачи, например, правила Зиглера-Никольса или Коэна-Куна.

Можно показать (сделайте это самостоятельно), что любой регулятор второго порядка с интегратором может быть представлен в форме ПИД-регулятора:

$$C(s) = \frac{a_2 s^2 + a_1 s + a_0}{s(b_1 s + b_0)} \Leftrightarrow C(s) = K + \frac{K_I}{s} + \frac{K_D s}{T_D s + 1}.$$

7.3. Метод размещения полюсов

Один из простых методов синтеза регулятора – *размещение полюсов* передаточной функции замкнутой системы, которые во многом определяют ее динамику, например, быстродействие и степень затухания колебаний (см. разд. 6.8). Смысл в том, чтобы разместить эти полюса в заданных точках комплексной плоскости с помощью специально выбранного регулятора. Эта задача сводится к решению системы линейных уравнений.

Пусть передаточная функция объекта задана в виде отношения полиномов

$$P(s) = \frac{n(s)}{d(s)} = \frac{n_1 s + n_0}{s^2 + d_1 s + n_0}.$$

Выберем регулятор вида

$$C(s) = \frac{n_c(s)}{d_c(s)} = \frac{a_1 s + a_0}{b_1 s + b_0},$$

где a_0, a_1, b_0 и b_1 – неизвестные коэффициенты, которые нужно определить. Характеристический полином замкнутой системы равен

$$\begin{aligned} \Delta(s) &= n(s) n_c(s) + d(s) d_c(s) = (n_1 s + n_0)(a_1 s + a_0) + (s^2 + d_1 s + n_0)(b_1 s + b_0) \\ &= b_1 s^3 + (n_1 a_1 + d_1 b_1 + b_0) s^2 + (n_0 a_1 + n_1 a_0 + d_0 b_1 + d_1 b_0) s + n_0 a_0 + d_0 b_0. \end{aligned}$$

Предположим, что мы хотим выбрать регулятор так, чтобы разместить корни полинома $\Delta(s)$ в заданных точках, то есть добиться выполнения равенства

$$\Delta(s) = s^3 + \delta_2 s^2 + \delta_1 s + \delta_0,$$

где δ_i ($i = 0, \dots, 2$) – заданные числа. Приравнявая коэффициенты при одинаковых степенях s в последних двух равенствах, получаем

$$\begin{aligned} s^3 : b_1 &= 1 \\ s^2 : n_1 a_1 + d_1 b_1 + b_0 &= \delta_2 \\ s^1 : n_0 a_1 + n_1 a_0 + d_0 b_1 + d_1 b_0 &= \delta_1 \\ s^0 : n_0 a_0 + d_0 b_0 &= \delta_0 \end{aligned}$$

или в матричном виде

$$\begin{bmatrix} 0 & 0 & 1 & 0 \\ n_1 & 0 & d_1 & 1 \\ n_0 & n_1 & d_0 & d_1 \\ 0 & n_0 & 0 & d_0 \end{bmatrix} \cdot \begin{bmatrix} a_1 \\ a_0 \\ b_1 \\ b_0 \end{bmatrix} = \begin{bmatrix} 1 \\ \delta_2 \\ \delta_1 \\ \delta_0 \end{bmatrix}.$$

Решение уравнения имеет вид

$$\begin{bmatrix} a_1 \\ a_0 \\ b_1 \\ b_0 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 & 0 \\ n_1 & 0 & d_1 & 1 \\ n_0 & n_1 & d_0 & d_1 \\ 0 & n_0 & 0 & d_0 \end{bmatrix}^{-1} \cdot \begin{bmatrix} 1 \\ \delta_2 \\ \delta_1 \\ \delta_0 \end{bmatrix}.$$

Конечно, квадратная матрица в этом выражении (она называется *матрицей Сильвестра*) должна быть обратима. Можно доказать, что она действительно обратима тогда и только тогда, когда полиномы $n(s)$ и $d(s)$ не имеют общих корней, то есть передаточная функция объекта

$P(s)$ несократима. В противном случае общий корень этих полиномов неизбежно будет корнем характеристического полинома $\Delta(s)$.

Кроме того, для того, чтобы количество неизвестных коэффициентов было равно числу уравнений, порядок регулятора нужно выбирать не меньше, чем $N-1$, где N – порядок модели объекта управления:

$$N = \max\{\deg n(s), \deg d(s)\},$$

где \deg обозначает степень полинома. Иначе полученное уравнение будет разрешимо только при специально выбранном полиноме $\Delta(s)$.

Заметим, что при размещении полюсов мы никак не учитываем *нули* передаточной функции, которые также влияют на динамику системы.

7.4. Коррекция ЛАФЧХ

На протяжении многих лет самым популярным инженерным методом синтеза регуляторов был метод, основанный на использовании логарифмических частотных характеристик (ЛАФЧХ). Он основан на двух свойствах ЛАФЧХ:

- 1) логарифмические амплитудные и фазовые частотные характеристики для последовательного соединения двух блоков (например, регулятора и объекта управления) равны сумме ЛАЧХ и ЛФЧХ этих блоков;
- 2) если передаточная функция объекта не имеет неустойчивых нулей и полюсов (с положительной вещественной частью), то амплитудная частотная характеристика однозначно определяет фазовую; отсюда следует что можно свести выбор регулятора к изменению *только* амплитудной характеристики нужным образом.

Пусть $G(s) = P(s) R(s)$ – передаточная функция объекта вместе с приводом, причем будет предполагать, что она не имеет неустойчивых нулей и полюсов (то есть, является *минимально-фазовой*). ЛАЧХ такого расширенного объекта обозначим как $L_0(\omega) = 20 \lg |G(j\omega)|$. Если мы сможем каким-то образом найти *желаемую* ЛАЧХ $L_{\text{ж}}(\omega)$, то разница между этими двумя характеристиками – это и есть ЛАЧХ необходимого последовательного регулятора:

$$L_c(\omega) = L_{\text{ж}}(\omega) - L_0(\omega). \quad (51)$$

Таким образом, для решения задачи требуется ответить на два вопроса:

- 1) как выбрать желаемую ЛАЧХ $L_{\text{ж}}(j\omega)$ так, чтобы обеспечить устойчивость и требуемое качество замкнутой системы?
- 2) как получить передаточную функцию регулятора $C(s)$ по его ЛАЧХ (51)?

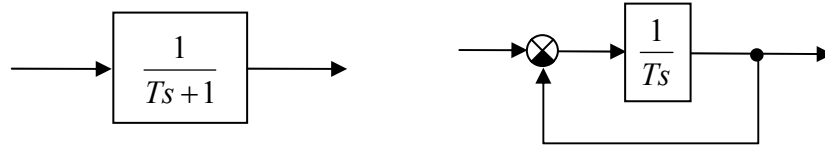
Чтобы ответить на первый вопрос, вспомним типичные требования к системе управления:

- устойчивость;
- нулевая ошибка в установившемся режиме;
- быстрый и плавный (в идеале – монотонный) переходный процесс;
- подавление шумов;
- робастность (нечувствительность к ошибкам модели).

Эти требования нужно связать с формой ЛАЧХ.

Постоянный сигнал можно рассматривать как предельный случай гармонического (синуса), только с нулевой частотой. Поэтому для обеспечения нулевой установившейся ошибки цепочка «регулятор-объект» должна иметь бесконечное усиление на нулевой частоте, то есть передаточная функция $G(s) C(s)$ должна содержать *интегратор* (вспомните *принцип внутренней модели*).

Обычно хочется, чтобы переходный процесс был монотонным, без перерегулирования. Такой процесс дает апериодическое звено. Легко проверить, что передаточная функция апериодического звена (слева) равна передаточной функции интегратора, охваченного единичной обратной связью (справа):



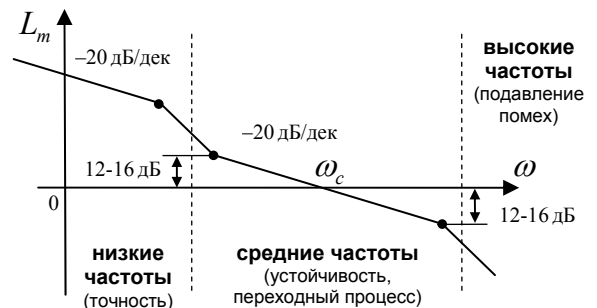
Таким образом, для получения монотонного переходного процесса ЛАЧХ разомкнутой системы должна быть похожа на ЛАЧХ интегратора – это прямая линия с наклоном -20 дБ/дек, которая пересекает ось абсцисс на частоте $\omega_c = 1/T$. Эта частота называется *частотой среза*. Заметим, что для апериодического звена легко определить время переходного процесса: оно примерно равно $3T$. Таким образом, *частота среза определяет время переходного процесса*. Вспомним, что устойчивость системы также определяется поведением ЛАЧХ в районе частоты среза. В результате имеем:

- устойчивость и качество переходного процесса (время, перерегулирование) определяются формой ЛАЧХ в районе частоты среза, где она пересекает ось $L_m = 0$; эта область называется *областью средних частот*;
- для получения качественного переходного процесса желательно, чтобы наклон ЛАЧХ около частоты среза был равен -20 дБ/дек;
- если задано время переходного процесса t_n , нужно выбирать $\omega_c = \frac{3}{t_n}$.

Теперь разберемся с шумами и робастностью. Как мы знаем, шумы – это высокочастотные сигналы. Кроме того, обычно именно в области высоких частот характеристики объекта и модели могут сильно расходиться. Поэтому для подавления помех и уменьшения влияния ошибок модели нужно по возможности уменьшать усиление системы в области высоких частот, то есть ЛАЧХ должна резко идти вниз.

На рисунке показана типовая желаемая ЛАЧХ. Это *асимптотическая* ЛАЧХ, состоящая из отрезков. В выделенных точках стыкуются два отрезка разного наклона. На низких частотах она имеет наклон -20 дБ/дек, то есть система содержит интегратор, который обеспечивает нулевую ошибку в установившемся режиме.

ЛАЧХ пересекает ось абсцисс под наклоном -20 дБ/дек. Для обеспечения устойчивости и приемлемого показателя колебательности ($M < 1,2$) точки излома ЛАЧХ должны находиться на расстоянии 12-16 дБ от оси абсцисс (см. рисунок).



Продemonстрируем метод коррекции ЛАЧХ на простом примере. Пусть объект управления – апериодическое звено с передаточной функцией $G(s) = \frac{1}{T_0 s + 1}$, где $T_0 = 5$ с. Передаточная функция замкнутой системы без коррекции (то есть, с регулятором $C_0(s) = 1$) равна:

$$W(s) = \frac{1}{T_0 s + 2} = \frac{0,5}{0,5 T_0 s + 1}.$$

Видим, что статический коэффициент усиления $W(0) = 0,5$ (а не 1), так что точного отслеживания входного сигнала не получается. Время переходного процесса можно приближенно подсчитать как $t_n = 3 \cdot 0,5 \cdot T_0 = 7,5$ с.

Поставим задачу следующим образом: выбрать регулятор $C(s)$, который обеспечивает

- нулевую ошибку в установившемся режиме;
- время переходного процесса около 1,5 с;

- наклон ЛАЧХ -40 дБ/дек на высоких частотах для подавления помех.

Для решения используем метод коррекции ЛАЧХ. Синяя линия на рисунке обозначает нескорректированную ЛАЧХ, совпадающую с ЛАЧХ апериодического звена $G(s)$.

Желаемая ЛАЧХ (зеленая линия) должна иметь наклон -20 дБ/дек на низких частотах, чтобы обеспечить нулевую статическую ошибку. Частота среза ω_c определяется требуемым быстродействием: $\omega_c = 3/t_n = 2$ рад/с. Таким образом, начальный участок желаемой ЛАЧХ совпадает с ЛАЧХ интегрирующего звена с передаточной функцией $\frac{\omega_c}{s}$, то есть

$$L_{жс}(\omega) = 20 \lg \frac{\omega_c}{\omega} \quad (\text{на низких частотах}).$$

На высоких частотах нужно изменить наклон ЛАЧХ с -20 до -40 дБ/дек на частоте ω_1 , где $L_m(\omega) = -15$ дБ. Из этого условия находим

$$20 \lg \frac{\omega_c}{\omega_1} = -15 \Rightarrow \frac{\omega_c}{\omega_1} = 10^{-15/20} \Rightarrow \omega_1 = 10^{3/4} \cdot \omega_c = 5,62 \cdot 2 = 11,24 \text{ рад/с}.$$

Таким образом, мы полностью построили желаемую ЛАЧХ, удовлетворяющую требованиям к системе. Вычитая из нее исходную ЛАЧХ (без коррекции, синяя линия), получим ЛАЧХ регулятора, которая показана красной линией на нижнем графике.

Остается перейти от ЛАЧХ регулятора к его передаточной функции. На низких частотах ($\omega < 1/T_0$) ЛАЧХ регулятора имеет наклон -20 дБ/дек и проходит через точку $(\omega_c; 0)$, то есть

$$C(s) = \frac{\omega_c}{s} \cdot C_1(s),$$

где $C_1(s)$ не изменяет асимптотическую ЛАЧХ на частотах, меньших $1/T_0$. На частоте $\omega_0 = 1/T_0$ ЛАЧХ регулятора меняет наклон с -20 дБ/дек до нуля, то есть в числитель добавляется множитель $T_0s + 1$:

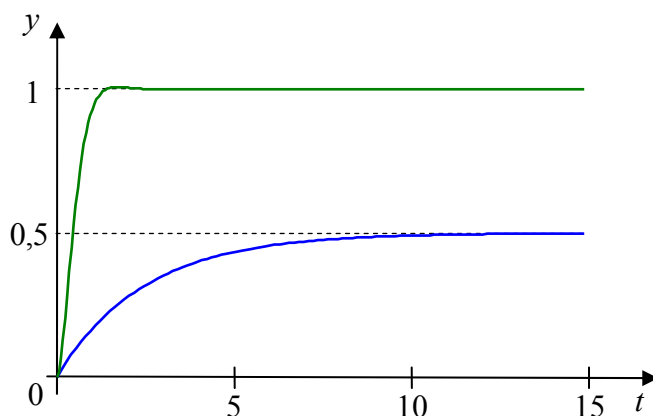
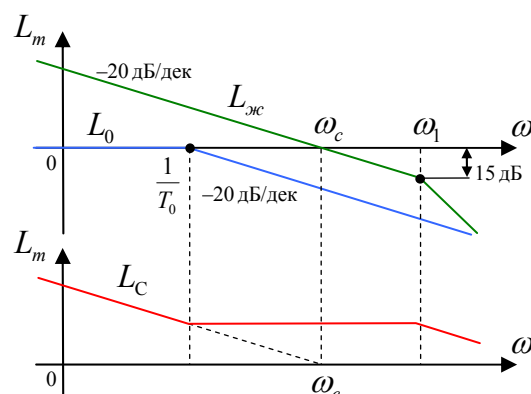
$$C(s) = \frac{\omega_c(T_0s + 1)}{s} \cdot C_2(s).$$

Здесь $C_2(s)$ – регулятор, не влияющий на ЛАЧХ для частот, меньших ω_1 . Наконец, на частоте ω_1 наклон увеличивается с нуля до -20 дБ/дек. Для того, чтобы на этой частоте «загнуть» вниз ЛАЧХ, нужно добавить в регулятор апериодическое звено с постоянной времени T_1 , где

$$T_1 = \frac{1}{\omega_1} = 0,09 \text{ с}. \quad \text{Таким образом, окончательно}$$

$$C(s) = \frac{\omega_c(T_0s + 1)}{s(T_1s + 1)}.$$

На рисунке показаны переходные процессы при единичном ступенчатом входном сигнале в нескорректированной системе (синяя линия) и в системе с полученным регулятором $C(s)$ (зеленая линия). Графики показывают, что найденный регулятор значительно ускорил переходный процесс и обеспечил нулевую статическую ошибку (установившееся значение

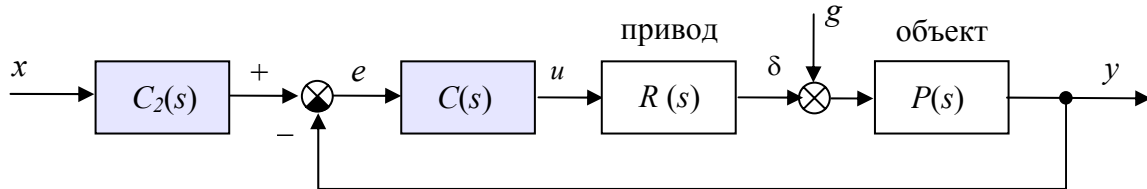


ние выхода равно 1).

Нужно отметить, что алгоритм коррекции ЛАЧХ существенно усложняется, если объект содержит неустойчивые или неминимально-фазовые звенья.

7.5. Комбинированное управление

Один из способов улучшить качество управления – изменить *структуру* системы, добавив в нее второй регулятор $C_2(s)$ на входе:



Теперь

$$W(s) = \frac{C_2(s)C(s)R(s)P(s)}{1 + C(s)R(s)P(s)}.$$

Регулятор $C_2(s)$ не влияет на свойства *контура* управления (запасы устойчивости, подавление возмущений, робастность), а влияет только на переходные процессы при изменении задающего воздействия. Поэтому сначала можно, не обращая внимание на переходные процессы, построить регулятор в контуре $C(s)$ так, чтобы обеспечить нужный уровень подавления возмущений и робастность, а затем сформировать нужные качества передаточной функции $W(s)$ с помощью регулятора $C_2(s)$. Поскольку две передаточные функции можно изменять независимо друг от друга, такая схема называется *комбинированным управлением* (или *управлением с двумя степенями свободы*).

В идеале мы хотим, чтобы система точно воспроизводила сигнал $x(t)$ на выходе $y(t)$, то есть, нужно обеспечить $W(s) \equiv 1$. Для этого требуется, чтобы

$$C_2(s) = \frac{1 + C(s)R(s)P(s)}{C(s)R(s)P(s)} = \frac{1}{W_1(s)}, \quad (52)$$

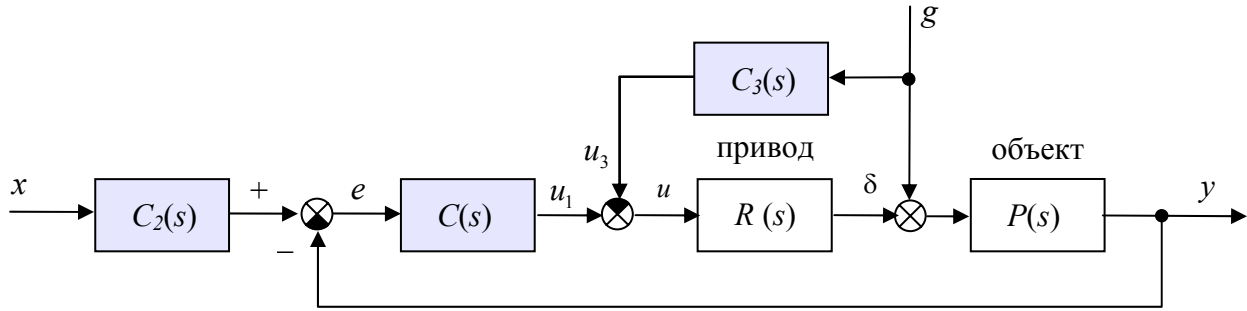
где $W_1(s) = \frac{C(s)R(s)P(s)}{1 + C(s)R(s)P(s)}$ – передаточная функция замкнутой системы с одной степенью свободы.

Из (52) следует, что регулятор $C_2(s)$ должен быть *обратной* системой (*инверсией*) для $W_1(s)$. Частотная характеристика $W_1(j\omega)$ в реальных системах близка к нулю на высоких частотах, следовательно, регулятор $C_2(s)$ должен иметь в этом частотном диапазоне огромное усиление. Например, для $W_1(s) = \frac{1}{Ts + 1}$ получим $C_2(s) = Ts + 1$, то есть регулятор содержит физически нереализуемое дифференцирующее звено. Таким образом, точная инверсия (52) не может применяться в практических задачах. Обычно стараются приближенно обеспечить равенство (52) для тех частот, где важно точно отследить задающий сигнал.

Отметим, что существуют и другие схемы с двумя степенями свободы, но можно доказать, что все они эквивалентны, разница только в реализации.

7.6. Инвариантность

Если возмущение g можно как-то измерить, для улучшения качества системы иногда вводится третий регулятор (*третья степень свободы*):



Теперь передаточная функция по возмущению равна

$$W_g(s) = \frac{[1 - C_3(s)R(s)]P(s)}{1 + C(s)R(s)P(s)}.$$

В этом случае теоретически есть возможность обеспечить полную компенсацию возмущения g , выбрав

$$C_3(s) = \frac{1}{R(s)}, \quad (53)$$

так что $W_g(s) = 0$. Это условие называется *условием инвариантности* (неизменности), поскольку в этом случае система абсолютно подавляет *любые* возмущения по входу g . Заметим, что мы снова пришли к идее *инверсии* (построения *обратной* системы), как и в (52).

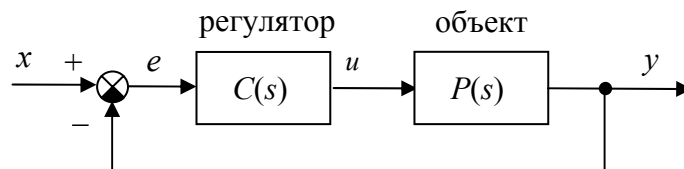
К сожалению, на практике условие инвариантности чаще всего невыполнимо, потому что регулятор $C_3(s)$ должен быть *предсказывающим*, так как нужно подать компенсирующий сигнал на привод *раньше*, чем внешнее возмущение успеет повлиять на объект.

Чаще всего получается, что числитель передаточной функции $C_3(s)$ (53) должен иметь более высокую степень, чем знаменатель. Это значит, что такой регулятор включает звенья чистого дифференцирования, которые не являются физически реализуемыми. Обычно подбирают регулятор $C_3(s)$ так, чтобы он был физически реализуемым, но условие (53) приближенно выполнялось в наиболее важном диапазоне частот.

7.7. Множество стабилизирующих регуляторов

Как известно, не каждый регулятор стабилизирует систему. Поэтому важно выделить множество регуляторов, которые обеспечивают устойчивость замкнутого контура. Такие регуляторы называются *стабилизирующими*. Желательно также получить *параметризацию*, то есть, представить все множество стабилизирующих регуляторов в виде формулы, зависящей от параметра, который может выбираться произвольно в некоторой допустимой области.

Рассмотрим простейшую замкнутую систему:



Ее передаточная функция равна

$$W(s) = \frac{C(s)P(s)}{1 + C(s)P(s)}. \quad (54)$$

Регулятор входит в нее *нелинейно*, что значительно осложняет анализ и синтез системы. Заметим, что эту функцию можно представить в виде

$$W(s) = Q(s)P(s), \quad \text{где} \quad Q(s) = \frac{C(s)}{1 + C(s)P(s)} \quad (55)$$

Выражение (55) внешне выглядит как передаточная функция последовательного соединения объекта $P(s)$ и «регулятора» $Q(s)$, причем оно *линейно* зависит от $Q(s)$. Поэтому естественно возникает вопрос: нельзя ли сначала выбрать нужным образом $Q(s)$, а затем найти соответствующий ей регулятор, выразив его передаточную функцию из (55):

$$C(s) = \frac{Q(s)}{1 - Q(s)P(s)}. \quad (56)$$

Очевидно, что функция $Q(s)$ должна быть устойчивой, иначе передаточная функция замкнутой системы $W(s)$ (55) также окажется неустойчивой. Оказывается, если объект $P(s)$ устойчив, то регулятор, полученный из (56), всегда будет стабилизирующим. Более того, форма (56) охватывает *все возможные* стабилизирующие регуляторы. Поэтому (56) – это параметризация множества стабилизирующих регуляторов для устойчивого объекта, она называется **параметризацией Юла** (*D.C. Youla*).

Параметром в (56) является устойчивая функция $Q(s)$, которая может выбираться произвольно. На практике регулятор (56) должен быть *физически реализуемым*. Это значит, что передаточная функция $C(s)$ должна быть *правильной* (степень ее числителя не больше степени знаменателя). Для этого функция $Q(s)$ также должна быть *правильной*.

Теоретически для оптимального слежения нужно выбрать $Q(s) = 1/P(s)$, что дает $W(s) = 1$, однако чаще всего это невозможно. Дело в том, что передаточная функция объекта в практических задачах – строго правильная (степень числителя *меньше* степени знаменателя), и $Q(s)$ получается *неправильной*. Поэтому используют компромиссные решения, обеспечивая приближенную инверсию только для наиболее важной полосы частот.

Существует множество методов синтеза, в которых устойчивая и правильная функция $Q(s)$ выбирается в результате *численной оптимизации* по какому-либо критерию. Затем передаточная функция регулятора рассчитывается по формуле (56).

Посмотрим, что получится, если попробовать применить такой подход для *неустойчивого* объекта с передаточной функцией $P(s) = \frac{1}{s-1}$. Выбрав $Q(s) = 1$, из (56) получаем

$$C(s) = \frac{s-1}{s-2}.$$

При этом в произведении $C(s)P(s) = \frac{s-1}{s-2} \cdot \frac{1}{s-1}$ неустойчивый полюс модели объекта сокращается (*компенсируется*) неустойчивым нулем регулятора. Характеристический полином

$$\Delta(s) = s - 1 + (s - 2)(s - 1) = (s - 1)^2$$

будет неустойчивым, как и вся замкнутая система. Следовательно, параметризацию (56) в этом случае использовать нельзя.

Для неустойчивых объектов используют другую, более сложную параметризацию. Пусть $P(s) = \frac{n(s)}{d(s)}$, где $n(s)$ и $d(s)$ – полиномы. Выберем произвольный *устойчивый* полином $f(s)$, степень которого равна наибольшей из степеней $n(s)$ и $d(s)$. Представим функцию $P(s)$ в виде отношения рациональных функций

$$P(s) = \frac{U(s)}{V(s)}, \quad \text{где } U(s) = \frac{n(s)}{f(s)} \text{ и } V(s) = \frac{d(s)}{f(s)}.$$

Можно показать, что существуют такие *правильные* устойчивые функции $X(s)$ и $Y(s)$, для которых выполняется равенство

$$U(s)X(s) + V(s)Y(s) = 1. \quad (57)$$

Тогда множество всех стабилизирующих регуляторов описывается формулой

$$C(s) = \frac{X(s) + V(s)Q(s)}{Y(s) - U(s)Q(s)}, \quad (58)$$

где $Q(s)$ – произвольная *правильная устойчивая* функция. Выражение (58) определяет *параметризацию множества стабилизирующих регуляторов* (параметризацию Юла) в общем случае, даже для неустойчивых объектов. Подставив (58) в формулу (54), получаем, учитывая (57),

$$W(s) = [X(s) + V(s)Q(s)]U(s).$$

При синтезе можно выбирать *устойчивую правильную* функцию $Q(s)$, при которой передаточные функции замкнутой системы (по входу, по возмущению, по ошибке) имеют нужные свойства, а затем вычислять передаточную функцию регулятора, используя (58).

Для примера рассмотрим снова неустойчивый объект с передаточной функцией $P(s) = \frac{1}{s-1}$, которую можно записать в виде

$$P(s) = \frac{U(s)}{V(s)}, \quad \text{где } U(s) = \frac{1}{s+1}, \quad V(s) = \frac{s-1}{s+1}.$$

Решением уравнения (57) может быть, например, такая пара устойчивых функций

$$X(s) = \frac{4}{s+1}, \quad Y(s) = \frac{s+3}{s+1}.$$

При выборе $Q(s) = 1$ по формуле (58) получаем

$$C(s) = \frac{s+3}{s+2}.$$

Теперь в произведении $C(s)P(s)$ нет никаких сокращений, система устойчива.

Заключение

Шаг за шагом, мы рассмотрели основные понятия классической теории автоматического управления. Нужно понимать, что вы прочитали не учебник, а небольшое введение, призванное *познакомить* с основными понятиями и дать общее представление о предмете. Тот, кто серьезно собирается изучать методы теории управления и использовать их в своей работе, должен продолжить изучение, взяв «нормальные» учебники (см. список литературы), в которых эти и другие вопросы изложены значительно более строго и научно.

За рамками пособия остались многие 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.

При проектировании обязательно должны учитываться случайные воздействия, которые не обсуждались в пособии. Внедрение цифровых компьютеров позволило использовать *адаптивные* системы со сложными алгоритмами управления, требующими объемных вычислений. Развиваются и новые классы систем, в которых для управления используется методы искусственного интеллекта и теории нечетких множеств.

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. *Попов Е.П.* Теория линейных систем автоматического регулирования и управления. М.: Наука, 1989.
2. *Мирошник И.В.* Теория автоматического управления. Линейные системы. СПб.: Питер, 2005.
3. *Первозванский А.А.* Курс теории автоматического управления – М.: Наука, 1986.
4. *Бесекерский В.А., Попов Е.П.* Теория систем автоматического управления – 4-е изд. СПб.: Профессия, 2003.
5. *Дорф Р., Бишоп Р.* Современные системы управления – М.: Бином, Лаборатория базовых знаний, 2004.
6. *Гудвин Г.К., Гребе С.Ф., Сальгадо М.Э.* Проектирование систем управления. М.: Бином, Лаборатория базовых знаний, 2004.