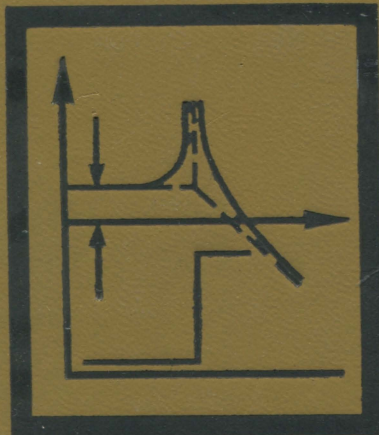
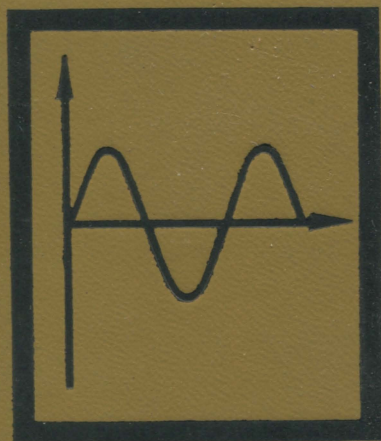


ОСНОВЫ ТЕХНИЧЕСКОЙ КИБЕРНЕТИКИ

Р. А. САПОЖНИКОВ



ОСНОВЫ ТЕХНИЧЕСКОЙ КИБЕРНЕТИКИ



**Р. А. САПОЖНИКОВ, П. Н. МАТВЕЕВ,
Б. П. РОДИН, Н. А. ФИЛАДЕЛЬФИНА**

ОСНОВЫ ТЕХНИЧЕСКОЙ КИБЕРНЕТИКИ

**Допущено
Министерством высшего и среднего
специального образования СССР
в качестве учебного пособия
для студентов
высших учебных заведений**



**ИЗДАТЕЛЬСТВО «ВЫСШАЯ ШКОЛА»
Москва — 1970**

6Ф0.1

0-75

УДК 62—50

0-75 **Основы технической кибернетики.** Учеб. пособие для студентов вузов. М., «Высшая школа», 1970.
464 с. с илл.

Перед загл. авт.: Р. А. Сапожников, П. Н. Матвеев,
Б. П. Родин, Н. А. Филадельфина

В книге в соответствии с современным уровнем знаний в области кибернетики изложены основные положения теории информации, общие принципы построения систем автоматического управления, математический аппарат для описания непрерывных и дискретных сигналов, проблема устойчивости, надежность систем автоматического регулирования и другие вопросы технической кибернетики. По теоретическим темам приведены расчетные примеры.

В книге приводятся все необходимые исходные сведения и понятия для проведения анализа и синтеза.

3-3-14

БЗ-35/25-70

6Ф0.1

Рецензенты:

Проф. Сергей Михайлович Смирнов;
кафедра автоматики и телемеханики
Московского энергетического института.

ПРЕДИСЛОВИЕ

В содержании известных книг по технической кибернетике [0.2, 0.6, 0.7, 0.9, 0.11] имеются значительные различия. Это естественно, поскольку современная кибернетика возникла недавно и ее разделы не получили общепризнанных границ. Однако бесспорно то, что кибернетика — наука об управлении динамическими системами. Поэтому в техническую кибернетику должна входить *теория автоматического управления*. Принимая это положение, авторы рассматривают регулирование, теория которого развивалась в течение многих десятилетий, как частный случай управления, хотя существует и другое понимание различия между управлением и регулированием [0.3]. Вместе с тем в технической кибернетике должна быть изложена применительно к автоматическому управлению теория информации, так как без информации не может быть управления.

Таким образом, содержание предлагаемой книги в основном определяется соединением теории информации с теорией автоматического управления. Это оставляет, однако, широкие возможности для того или иного выбора излагаемого материала. Составляя учебное пособие для студентов, изучающих автоматику, имелось в виду наличие дисциплин, в которых рассматривается синтез автоматических систем. Поэтому здесь ограничили изложением основных сведений и понятий, необходимых для анализа и синтеза автоматических систем, после чего уделили внимание лишь анализу. При этом отдавалось предпочтение новейшим и наиболее эффективным методам решения рассматриваемых задач.

При всех указанных ограничениях круг вопросов, затронутых в книге, все же довольно обширен. По каждому такому вопросу в книге дан лишь самый необходимый для учебного пособия материал. В выборе этого материала, конечно, сказались взгляды и опыт авторов.

Для расширения и углубления своих знаний читатели книги могут обратиться к литературе, указанной по главам, а также к особо рекомендуемым руководствам [0.1, 0.4, 0.5, 0.8, 0.10, 0.12, 0.13]. Изучая указанную литературу, можно, в частности, критически оценить некоторые новые методы, изложенные в этой книге.

Авторы работали над всей книгой совместно. В основном же гл. I, § 2.1, 2.4, 2.6, 3.2, 4.4—4.6, 5.1, 5.3, 6.1—6.4, 6.6, 7.1—7.5, 7.7, 7.8, 8.1 и 8.2 написаны Р. А. Сапожниковым; § 2.2, 2.3, 4.1 и 5.2 —

Р. А. Сапожниковым и П. Н. Матвеевым; § 2.5 и 8.3 — Б. П. Родиным и Р. А. Сапожниковым; § 3.1, 3.3, 4.2, 4.3, 5.5 и 6.5 — Р. А. Сапожниковым и Б. П. Родиным; § 3.4 и 3.5 — Р. А. Сапожниковым, Б. П. Родиным и Н. А. Филадельфиной; § 5.4 — Р. А. Сапожниковым и Н. А. Филадельфиной; § 7.6 и гл. IX — П. Н. Матвеевым; § 8.4—8.6 — Б. П. Родиным; гл. X — Н. А. Филадельфиной. Общее редактирование книги выполнил Р. А. Сапожников.

Авторы признательны проф. К. В. Егорову, проф. А. В. Нестушилу и проф. С. М. Смирнову за ряд ценных замечаний, сделанных при рецензировании рукописи книги и учтенных в ее окончательной редакции.

Отзывы по книге просим направлять по адресу: Москва, К-51, Неглинная, 29/14. Изд-во «Высшая школа».

Авторы

ВВЕДЕНИЕ В КИБЕРНЕТИКУ

§ 1.1. ПРЕДМЕТ И ИСТОРИЯ КИБЕРНЕТИКИ

В природе, в обществе и в технике — повсюду наблюдаются *системы*, состоящие из так или иначе связанных между собой элементов или составных частей, обуславливающих существование и действие каждой системы. Таковы Солнечная система, живой организм, промышленное предприятие, любое транспортное средство и т. д. Каждая такая система существует, пока сохраняет свое единство, но ее состояние всегда находится в процессе изменения. Состояние Солнечной системы, например, изменяется вследствие движения составляющих ее тел. Катер, идущий по волнам, также представляет собой систему, в части которой: корпус, двигатель, винт, руль и т. д., — обеспечивают ее сохранение в плавании и выполнение ее назначения — движения по требуемому курсу. Состояние катера при этом все время изменяется. Состояние всякой системы определяется или описывается с помощью характеризующих его величин, рассматриваемых в функции времени. Для катера, например, такими величинами будут скорость его движения по курсу, а также угловые и линейные перемещения, обусловленные качкой, и производные этих перемещений по времени. На примере катера видно, что состояние системы может изменяться под влиянием испытываемых ею действий. Здесь, с одной стороны, имеются *возмущающие воздействия* в виде ветра и волн, а с другой — *управляющие воздействия*, которые производит управляющий катером человек, пользуясь рулевым устройством и регулируя работу двигателя. Изменение состояния всякой системы изучают, рассматривая характеризующие ее величины и их производные в функции времени и воздействующих на нее факторов. Какова бы ни была природа таких величин и факторов, установление их связи во времени представляет собой задачу, вполне аналогичную тем, которые решаются в динамике.

Поэтому рассматриваемые в таком аспекте системы называют *динамическими*. *Состояние динамической системы* определяется взаимосвязью характеризующих ее и воздействующих на нее величин и их производных по времени.

Воздействия, влияющие на динамическую систему, могут иметь установившийся на длительное время характер, при котором возникающие в системе изменения периодически повторяются, как это происходит в Солнечной системе. В других случаях эти воздействия появляются эпизодически или же имеют случайный характер и вызывают неожиданные изменения. В первом случае имеет место *установившийся режим* или *установившийся процесс* изменения состояния системы, во втором — *переходный режим* или *переходный процесс*. Катер, идущий с постоянной скоростью по неизменяемому курсу при штиле, находится в установившемся режиме, проходя одинаковые расстояния за одинаковые отрезки времени; при наличии же ветрового волнения он не выходит из переходного режима.

Возмущающие воздействия могут нарушать действие системы и угрожать ее существованию. Так, волны и ветер могут не только замедлять движение катера и сбивать его с курса, но и опрокинуть его. В таких условиях система может сохранить свое существование или свои функции, лишь приспосабливаясь к возмущающим воздействиям. А это достигается благодаря управляющим воздействиям. Человек, управляющий катером, действует так, чтобы плавание продолжалось по заданному курсу, несмотря на ветер и волны.

Таким образом, *управление* обеспечивает целенаправленное *приспособление* системы к возмущающим воздействиям. Это осуществляет каждый живой организм, являющийся системой, которая обладает регулирующими органами, позволяющими сохранять жизнеспособность в различных условиях внешней среды. Достаточно указать на то, что температура тела теплокровных организмов колеблется в одних и тех же узких пределах при значительных изменениях температуры окружающей среды, так как это необходимо, чтобы организм не разрушался. Живой организм управляет своими движениями и многими другими функциями. Для этого служат управляющие системы, которыми он обладает. Ни один живой организм не мог бы существовать без таких управляющих систем.

Их расстройство приводит к различным заболеваниям и к гибели живых существ.

Чем сложнее задача приспособления к окружающим условиям и воздействиям, тем сложнее должно быть устройство динамической системы. Но каковы бы ни были эти системы, процессы управления ими подчиняются некоторым общим закономерностям и характеризуются сходными явлениями. Эти закономерности и явления изучает *кибернетика* — наука об управлении динамическими системами.

Кибернетика изучает процессы управления и системы, в которых эти процессы осуществляются. Первоначально греческое слово *κυβερνήτης*, означало кораблевождение, т. е. науку об управлении кораблем, но еще в древности этим словом стали называть искусство управления вообще [1.1]. Мысль о значении и необходимости изучения проблемы управления находила признание неоднократно. А. М. Ампер предложил назвать кибернетикой науку об управлении обществом, которая, как он полагал, должна появиться и которой он дал место в своей классификации наук [1.1].

Современная кибернетика находит и исследует закономерности, общие для всех процессов управления, в какой бы области они ни протекали. Она получила развитие с того времени, когда было обращено внимание на общность проблем, относящихся к управлению машинами и живыми организмами. Такие проблемы стала изучать с 1942 г. небольшая группа ученых, образовавшаяся вокруг американского математика Норберта Винера и мексиканского физиолога Артуро Розенблюта [1.9]. Впоследствии с той же точки зрения начали рассматривать социально-экономические проблемы управления и было обнаружено, как заметил известный английский кибернетик Уильям Росс Эшби, много «интересных и многообещающих параллелей между машиной, мозгом и обществом» [1.11]. «Кибернетика стала наукой об общих принципах управления и о применении их в технике, в человеческом обществе и в живых организмах», — так определил ее содержание академик А. И. Берг, возглавивший кибернетическое направление научных исследований в СССР [1.16]. Известны примеры успешного применения кибернетической теории в указанных областях.

Первое систематическое изложение идей этой науки дал Винер в изданной в 1948 г. книге, название которой

указывает, что кибернетика изучает управление и связь в живых организмах и в машинах [1.4]. В этой книге Винер впервые определил содержание и задачи новой науки и предложил ее назвать кибернетикой. Поэтому общепризнано, что эта наука начала свое существование в 1948 г., а Винер является ее основателем. Действительно, быстрое развитие кибернетики началось после появления книги Винера. Его заслуга состоит в том, что он подметил возникновение новой науки, дал ей имя, сформулировал ее задачи и сделал значительный вклад в разработку некоторых ее разделов.

§ 1.2 ЭНТРОПИЯ

Всякий процесс управления должен быть целенаправленным. У. Р. Эшби сказал, что кибернетика — «наука о том, как надо управлять очень сложной системой, чтобы в итоге она вела себя желательным для нас образом» [1.10]. В процессах управления, осуществляемых человеком, требуемое поведение управляемых систем определяется сознательно намеченными целями. В процессах управления живыми организмами, существующими в природе, обеспечивается поведение управляемых систем, определяемое условиями сохранения и развития организма или вида.

Конкретное выражение «желательного» поведения может быть весьма различным для различных объектов. Есть, однако, нечто общее во всех случаях управления, где бы и как бы оно ни происходило. Это общее состоит в том, что всякий процесс управления вносит в управляемый объект порядок или организованность, упорядоченность и придает смысл его действиям. В связи с этим возникает задача количественного выражения *упорядоченности* или *неупорядоченности* управляемых объектов. Чтобы решать такую задачу, необходимо прежде всего выделить совокупность признаков, по которым определяется состояние упорядоченности рассматриваемого объекта. Это состояние может быть различным по разным признакам. Например, упорядоченность библиотеки по топографическому признаку будет совершенной, когда все книги находятся на своих местах, но в то же время эта библиотека может быть очень неупорядоченной по признаку полноценности, если книги попорчены и в них не хватает страниц.

Предположим, что интересующий нас признак, определяющий состояние некоторого объекта, выбран. Пусть n — число равновероятных состояний, в которых по этому признаку может находиться объект. Например, полка книг, на которой поддерживается порядок, по топографическому признаку может находиться лишь в одном состоянии; если же порядок не поддерживается, то в пределе число ее состояний равно числу перестановок всех книг. В этом примере n может изменяться от 1 до $b!$, если b — количество книг на полке. Это может быть очень большое число, которое поэтому неудобно для количественной оценки состояния объекта. Составим меру неупорядоченности, т. е. количественное выражение неупорядоченности объекта, как некоторую функцию $F(n)$ возможного числа его равновероятных состояний по выбранному признаку. Потребуем, чтобы функция $F(n)$ удовлетворяла следующим условиям:

1) $F(1) = 0$ — неупорядоченность вполне упорядоченного объекта, находящегося в единственно возможном состоянии, равна нулю;

2) $F(n_2) > F(n_1)$, если $n_2 > n_1$ — неупорядоченность монотонно возрастает с увеличением числа возможных состояний объекта;

3) $F(n_1 n_2) = F(n_1) + F(n_2)$ — неупорядоченность объекта равна сумме неупорядоченностей его частей, так как при n_1 возможных состояниях одной из двух частей объекта и n_2 — другой число его возможных состояний равно $n_1 n_2$.

Этим условиям, очевидно, удовлетворяет логарифм числа n , и можно доказать, что им не удовлетворяет никакая другая функция от n [1.14]. Логарифмическая мера неупорядоченности применима ко всем объектам, неупорядоченность которых по выбранному признаку не зависит от способа соединения и взаимодействия их частей [1.24]. В противном случае третье из перечисленных условий не имеет места [1.18.].

Не рассматривая здесь подобные случаи и применяя натуральный логарифм, напомним выражение меры неупорядоченности в виде

$$S = k \ln n, \quad (1.1)$$

где k — постоянный коэффициент, определяемый выбором единиц измерения неупорядоченности.

Величину S , численно выражающую неупорядоченность рассматриваемого объекта по некоторому признаку, назовем его *энтропией* по этому признаку. Энтропия характеризует общее состояние объекта, с которым может совпадать n конкретных его состояний. Общее состояние определяется именно тем, что оно допускает n конкретных состояний. Может быть известно число n , можем знать энтропию как функцию этого числа, но нет возможности узнать, какое из n состояний имеет место, т. е. отличить эти состояния друг от друга [1.12]. Поэтому энтропия является мерой нашего незнания. Ее называют также мерой неупорядоченности, употребляя слово мера в смысле количественного выражения.

Энтропия растет с увеличением числа n возможных равновероятных состояний объекта, когда упорядоченность объекта уменьшается. Поэтому упорядоченность будет возрастать с увеличением $1/n$. Из предыдущего выражения имеем:

$$-S = k \ln \frac{1}{n}. \quad (1.2.)$$

Отсюда следует, что с возрастанием упорядоченности объекта увеличивается отрицательная энтропия или *негэнтропия* (негативная энтропия):

$$N = S, \quad (1.3.)$$

которая является, следовательно, мерой упорядоченности [1.3].

Поскольку все рассматриваемые состояния системы, число которых равно n , равновероятны, вероятность каждого из них

$$P = \frac{1}{n}. \quad (1.4)$$

Это позволяет выразить энтропию и негэнтропию через вероятность P :

$$S = -k \ln P, \quad (1.5)$$

$$N = k \ln P. \quad (1.6)$$

В более общем случае возможные состояния системы не являются равновероятными, а характеризуются вероят-

ностями P_i ($i = 1, 2, \dots, n$), причем

$$\sum_{i=1}^n P_i = 1. \quad (1.7)$$

В большом числе m реализаций этих состояний каждое из них будет иметь место mP_i раз (здесь предполагается, что mP_i —целое или округленное до целого число). Рассматривая совокупность mP_i равновероятных реализаций i -го состояния, находим, согласно (1.1), что энтропия этой совокупности равна $-kmP_i \ln P_i$.

Энтропия же всех m реализаций должна быть равна сумме энтропий, получаемых из последнего выражения при $i = 1, 2, \dots, n$.

Деля такую сумму на m , найдем, что энтропия рассматриваемой системы

$$S = -k \sum_{i=1}^n P_i \ln P_i, \quad (1.8)$$

а негэнтропия

$$N = k \sum_{i=1}^n P_i \ln P_i. \quad (1.9)$$

Формулы (1.5.) и (1.6) получаются, если $P_i = P$, и поэтому

$$\sum_{i=1}^n P_i \ln P_i = nP \ln P,$$

где согласно (1.4) $nP = 1$. Формула (1.1) выражает максимальное значение энтропии (1.8); когда все возможные состояния системы равновероятны, она наиболее неупорядоченна, а следовательно, ее энтропия должна иметь наибольшее значение.

Полагая

$$k = \log_2 e \approx 1,443, \quad (1.10)$$

имеем из (1.1)

$$S = \log_2 n, \quad (1.11)$$

а из (1.8)

$$S = - \sum_{i=1}^n P_i \log_2 P_i. \quad (1.12)$$

Энтропия, вычисляемая по (1.11)¹, равна единице при $n=2$, т. е. в том случае, когда объект может находиться в одном из двух равновероятных состояний. Такая единица называется *двоичной единицей* и обозначается сокращенно дв. ед. Другие названия двоичной единицы «*бит*» и «*бинит*» (от английского *binary digit*—двоичный разряд²). Энтропия, выраженная в двоичных единицах, безразмерна.

Из выражений энтропии (1.1), (1.8) и (1.12) следует, что эта величина всегда положительна. Поэтому негэнтропия, в силу (1.3), всегда отрицательна, имея наибольшее значение, равное нулю, как получается из (1.6) при $P=1$. Таким образом, всегда имеют место условия: $S \geq 0$, $N \leq 0$.

Любое состояние каждого из рассматриваемых объектов характеризуется по выбранному признаку некоторой неупорядоченностью и некоторой упорядоченностью. Энтропия S является мерой неупорядоченности, а мерой упорядоченности служит негэнтропия (1.3). Если энтропия уменьшается, т. е. получает отрицательное приращение, то негэнтропия получает такое же по абсолютному значению положительное приращение, т. е. увеличивается. Это означает, что упорядоченность объекта увеличивается, а неупорядоченность уменьшается. Можно, следовательно, улучшать упорядоченность объекта, воздействуя на него так, чтобы его негэнтропия увеличивалась, или, как еще говорят, питая его негэнтропией [1.3]. При этом энтропия объекта будет уменьшаться или, по крайней мере, не будет возрастать.

Можно сказать, следовательно, что задачей управления является борьба с неупорядоченностью управляемой системы посредством введения в эту систему негэнтропии. Тем самым управление борется с существующей в неуправляемой природе тенденцией к разрушению организованного, к возрастанию дезорганизованности или неупорядоченности [1.8]. В этом состоит основная задача кибернетики [1.13]. На несколько десятилетий раньше ее предвосхитил известный русский физик Н. А. Умов. Еще в 1900 г. он писал: «Вероятный исход всякого естественного процесса в природе определяется уничтожением

¹ Эшби называет ее «разнообразием» [1.11].

² По-немецки—Ja — Nein Einheit — «да—нет» единица.

стройных движений. Возникновение в природе стройного движения будет поэтому событием невероятным, но там, где происходят миллиарды событий, невероятное не есть невозможное... Организованная материя призвана создавать в природе невероятные состояния, хотя на жизни отдельного индивидуума отражается властвующий в природе закон разрушения стройных движений» [1.2]. В борьбе с неорганизованной природой, указывал Умов, человек «вводит в круг своих стройностей растительное и животное царство, в своих орудиях и машинах распространяет эти стройности на неорганизованную материю и борется во имя этих стройностей со случайным порядком событий в природе» [1.2].

К тем же мыслям вновь пришел Л. Бриллюэн, указавший, что «научные законы дают нам способ предсказания, с помощью которого мы имеем возможность создавать системы с высокой негэнтропией... Каждое из этих устройств представляет собой чрезвычайно маловероятную структуру, которая не может осуществиться в природе» [1.18]. В природе, как правило, происходит лишь то, что допускает упомянутая тенденция, выражаемая *вторым законом термодинамики (принципом Карно)*, согласно которому имеет место возрастание энтропии и убывание негэнтропии, происходящее параллельно с разрушением организованных структур, стремящихся к наиболее вероятному, хаотическому состоянию.

Процесс возрастания энтропии, рассматриваемый в термодинамике, в обычных условиях, необратим. Но это статистический процесс, он характеризует изменение макросостояния сложных систем, зависящего от множества микросостояний, существующих на молекулярном уровне, где все процессы обратимы [1.26]. Если же система является *открытой*, т. е. находится в состоянии обмена с окружающей средой, то ее энтропия может не возрастать, а даже уменьшаться.

Одним из примеров местного нарушения второго закона термодинамики служит образование пары электрон — позитрон из фотона тепловой радиации, т. е. превращение рассеянной тепловой энергии в энергию более ценного вида. При некоторых условиях могут существовать и макроскопические области убывания физической энтропии.

Открытыми являются все системы, воспринимающие любые внешние воздействия. Таковы, в частности, само-

настраивающиеся и самоорганизующиеся системы, которые будут рассматриваться в гл. IX.

Явления противодействия возрастанию энтропии или даже уменьшения ее имеют место в органическом мире, на что тоже указывал Умов [1.2]. Живой организм все время борется с возрастанием своей энтропии. Прекращение этой борьбы означает смерть, когда энтропия организма достигает наибольшего значения. Кибернетика изучает явления и законы убывания энтропии в природе [1.8] для их практического применения.

Энтропия, изучаемая в термодинамике и называемая далее *физической*, — частный случай рассмотренной выше энтропии в широком смысле, являющейся мерой любой неупорядоченности [1.5]. Когда какая-либо система получает некоторое количество тепла ΔQ , ее физическая энтропия увеличивается на

$$\Delta S = \frac{\Delta Q}{T}. \quad (1.13)$$

Таким образом, физическая энтропия имеет размерность энергии, деленной на температуру. При этом температура системы также увеличивается, а это всегда сопровождается, во-первых, возрастанием интенсивности теплового движения, во-вторых, увеличением мощности теплового излучения. Первое увеличивает неупорядоченность системы, приближает ее к хаотическому состоянию, второе означает рассеяние или обесценение (деградацию) энергии системы. Поэтому физическая энтропия служит не только мерой неупорядоченности, но и мерой обесценения энергии.

В применении к физической энтропии выражение (1.1) представляет собой формулу Больцмана—Планка, в которой n —статистический вес; k —*постоянная Больцмана*:

$$k = 1,38 \cdot 10^{-23} \text{ дж/град.}$$

Статистический вес определяется числом возможных микросостояний системы, зависящих от положений и скоростей всех ее атомов, квантовых состояний этих атомов и молекулярных структур и т. п.

Пример. Какова энтропия состояния некоторого пункта по признаку осадков, если выпадение их ожидается с вероятностью 0,5?

Вероятность того, что осадков не будет, тоже равна 0,5, т. е. возможны два равновероятных состояния. Поэтому искомая энтро-

пия

$$S = -2 \frac{1}{2} \log_2 \frac{1}{2} = 1 \text{ дв. ед.}$$

Этот пример наглядно показывает, что энтропия является мерой нашего незнания. В действительности в рассматриваемом пункте осадки либо выпадут, либо нет, так что состояние его будет вполне определенным. Но это состояние представляется неопределенным, так как известные нам вероятности обоих возможных событий меньше единицы. Найденная энтропия и характеризует эту неопределенность, обусловленную неточностью наших знаний.

§ 1.3. ИНФОРМАЦИЯ

Успешное управление осуществимо лишь при наличии достаточных знаний о свойствах управляемого объекта. Источники этих знаний — прежде всего наблюдения и опыт. Пусть происходит исследование какого-либо объекта. Ограниченность знаний заставляет допускать (см. § 1.2), что этот объект может находиться в некотором числе n равновероятных состояний. Наблюдения и опыты дают сведения, на основании которых определяем новое число возможных равновероятных состояний объекта, равное n . В действительности в каждый момент времени объект находится в каком-то одном состоянии. Если бы наблюдатель определил, какое это состояние, то в результате опыта он положил бы $n = 1$. Но исход опыта обычно характеризуется некоторой *неопределенностью*, из-за которой $n > 1$, т. е. наблюдатель может установить лишь то, что объект находится в одном из n состояний. Это связано с воздействием средств испытания на объект, и там, где этим воздействием нельзя пренебречь, необходимо учитывать обусловленную им неопределенность результатов испытаний. В частности, в физических экспериментах это определяется принципом неопределенности Гейзенберга. Кроме того, испытания и измерения всегда сопровождаются систематическими и случайными ошибками, также приводящими к некоторой неопределенности результата. Эта неопределенность может быть еще следствием быстро происходящих изменений объекта наблюдения и заметно сказывающихся в течение времени испытаний. Во многих случаях она бывает обусловлена помехами. Неопределенность возникает также вследствие выборочного способа испытаний, когда без этого способа нельзя обойтись. Наконец, неопределенность наблюдае-

мого состояния объекта зависит от его неупорядоченности: чем больше неупорядоченность, тем больше число возможных состояний и тем труднее определить, какое из них фактически имеется [1.24, 1.27].

Неопределенность результата испытаний означает неупорядоченность нашего представления о их объекте. В силу перечисленных причин эта неупорядоченность больше неупорядоченности объекта.

Если сведения, полученные об объекте исследования, содержат какие-либо новые данные, то представление о нем уточняется и

$$n < n_0. \quad (1.14)$$

Если в этих сведениях нет ничего нового, то

$$n = n_0. \quad (1.15)$$

Если же полученные сведения противоречивы и делают наше представление об объекте менее ясным, то

$$n > n_0, \quad (1.16)$$

Когда имеет место соотношение (1.14), знания об объекте исследования увеличиваются. Это увеличение может быть различным, в одних случаях оно больше, в других меньше.

Слова «больше» и «меньше» недостаточны, чтобы оценить прирост знаний наблюдателя. Для этого необходимо иметь количественное выражение прироста знаний. Такое выражение может быть получено из соображений, аналогичных тем, которые привели нас к выражению энтропии [1.21]. Но имея последнее, можно воспользоваться им для решения стоящей задачи. Действительно, рассматривая энтропию как меру незнания (см. § 1.2), следует оценить прирост знаний, получаемый при условии (1.14), разностью энтропий, вычисленных по числам состояний n_0 и n , или, что то же, разностью негэнтропий, вычисленных по числам n и n_0 . По (1.1) энтропия имеет при этом, соответственно, значения

$$\left. \begin{aligned} S_0 &= k \ln n_0; \\ S &= k \ln n. \end{aligned} \right\} \quad (1.17)$$

Следовательно, прирост знаний определяется положительной величиной

$$I = S_0 - S, \quad S_0 > S, \quad (1.18)$$

откуда

$$I = k \ln \frac{n_0}{n}, \quad n \leq n_0. \quad (1.19)$$

При условии (1.16) прирост знаний определяется отрицательной величиной

$$D = S_0 - S, \quad S_0 < S, \quad (1.20)$$

показывающей, что знания уменьшились. Отсюда

$$D = k \ln \frac{n_0}{n}, \quad n \leq n_0. \quad (1.21)$$

Величины I и D называют, соответственно, *информацией и дезинформацией* [1.24]. Они подчиняются условиям:

$$I \geq 0,$$

$$D \leq 0.$$

По (1.19) при каждом значении n_0 наблюдатель получает наибольшую информацию, если

$$n = 1. \quad (1.22)$$

Смысл этого ясен: представление наблюдателя об объекте становится в таком случае вполне определенным.

Принимая в (1.19) и (1.21) значение k по (1.10), получаем:

$$I = \log_2 \frac{n_0}{n}, \quad n \leq n_0, \quad (1.23)$$

$$D = \log_2 \frac{n_0}{n}, \quad n \geq n_0. \quad (1.24)$$

Информация, вычисленная по (1.23), равна единице, если

$$\frac{n_0}{n} = 2. \quad (1.25)$$

Это значит, что информация, получаемая об объекте, вдвое уменьшает предполагавшееся возможным число его состояний.

Формулы (1.23) и (1.24) становятся удобнее для применения, если вместо чисел возможных состояний n_0 и n ввести *априорные* и *апостериорные вероятности* P_0 и P этих состояний. Положим, что все n состояний равновероятны, так же как равновероятны все n_0 состояний.

Тогда, принимая во внимание (1.4) и полагая аналогично

$$P_0 = \frac{1}{n_0}, \quad (1.26)$$

из (1.23) и (1.24) получаем:

$$I = \log_2 \frac{P}{P_0}, \quad P \geq P_0, \quad (1.27)$$

$$D = \log_2 \frac{P}{P_0}, \quad P \leq P_0. \quad (1.28)$$

По (1.27) при $P_0 = \text{const}$ информация получает наибольшее значение

$$I = -\log_2 P_0, \quad (1.29)$$

если

$$P = 1. \quad (1.30)$$

Согласно (1.4) это означает выполнение условия (1.22). При этом информация равна единице, если $P_0 = 1/2$, т. е. выясняется наличие одного из двух возможных равновероятных состояний.

Важно заметить, что вероятность P может быть меньше единицы по уже упомянутым причинам, вследствие которых больше единицы число n , получаемое по данным опыта, в частности, из-за помех при передаче и приеме информации. Помехи, называемые также шумами, согласно (1.19), (1.23) и (1.27) приводят к уменьшению получаемой информации или к ее потере. А потеря информации ведет, по (1.18), к увеличению энтропии, определяющей незнание, чем бы эта потеря ни была вызвана.

Если рассматриваемые n состояний системы характеризуются различными вероятностями P_{0i} и P_i ($i = 1, 2, \dots, n_0, \dots, n$), то, вычисляя информацию по (1.18), следует выражать энтропию S по (1.8) или (1.12) и получать аналогично

$$S_0 = - \sum_{i=1}^n P_{0i} \log_2 P_{0i}. \quad (1.31)$$

В этом общем случае из (1.18) получим [1.24]:

$$I = \sum_{i=1}^n P_i \log_2 P_i - \sum_{i=1}^{n_0} P_{0i} \log_2 P_{0i}. \quad (1.32)$$

Если $S_0 < S$, то правая часть последнего выражения отрицательна, т. е. получаем не информацию, а дезинформацию D .

Если $S = 0$, из (1.18) следует, что

$$I = S_0. \quad (1.33)$$

Это согласуется с определением Больцмана, по которому энтропия есть мера недостатка информации. Такой недостаток означает, что возможно некоторое разнообразие состояний объекта, отличающихся по выбранному признаку. Эти состояния, как было выяснено в § 1.2, нет возможности отличить одно от другого.

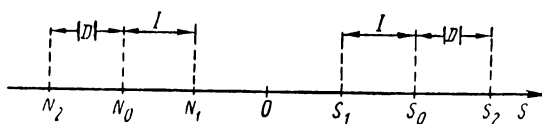


Рис. 1.1. Связь энтропии, негэнтропии, информации и дезинформации

На рис. 1.1. показана энтропийная ось, на положительной полуоси которой отложены значения энтропии, а на отрицательной — негэнтропии. Каждой точке положительной полуоси соответствует, согласно (1.3), находящаяся на таком же расстоянии от нуля точка отрицательной полуоси; первая характеризует неупорядоченность, вторая — упорядоченность рассматриваемой системы. Если

$$S = S_1 < S_0,$$

то полученная информация

$$I = S_0 - S_1 = N_1 - N_0. \quad (1.34)$$

Если же

$$S = S_2 > S_0,$$

то полученная дезинформация

$$D = S_0 - S_2 = N_2 - N_0.$$

Если энтропия получает отрицательное приращение, т. е. уменьшается, то негэнтропия получает такое же положительное приращение т. е. увеличивается, и наоборот.

Можно воздействовать на объект, изменяя его упорядоченность или неупорядоченность, но неупорядоченность в первом случае и упорядоченность во втором изменяются в соответствии с этим соотношением.

На место человека, наблюдающего некоторый объект и воздействующего на него, может быть поставлен автоматический прибор, исполняющий те же действия и определяющий энтропию объекта по тому же признаку, какой рассматривает человек. Поэтому далее будем говорить о связанном с объектом *операторе*, которым может быть либо человек, либо прибор.

Неупорядоченность представлений оператора зависит, с одной стороны, от него самого, а с другой, — от объекта. Следовательно, это неупорядоченность системы, состоящей из оператора и объекта его наблюдения.

После получения информации I энтропия системы оператор — объект уменьшается по (1.18) до

$$S = S_0 - I. \quad (1.35)$$

Это означает, с одной стороны, что представление оператора об объекте становится более упорядоченным. С другой стороны, это показывает, что информация действует так же, как негэнтропия, уменьшая энтропию системы оператор — объект. Негэнтропия N и информация I одинаково уменьшают энтропию, если

$$I = |N|, \quad (1.36)$$

но информация для этого вводится со знаком минус, а негэнтропия — со знаком плюс.

До сих пор рассматривался процесс получения информации. Предположим теперь, что имеется объект, состояние которого по некоторому признаку характеризуется энтропией S_0 . Если обладать знаниями свойств объекта, позволяющими упорядочить его состояние по тому же признаку, и эти знания выражать количественно информацией I , то возможно уменьшить энтропию рассматриваемого объекта по тому же признаку до S по формуле (1.35). Пусть о некотором объекте известно, что его энтропия равна S_1 . Знание свойств объекта и законов, которым подчинены происходящие в нем явления, позволяет предвидеть, что по прошествии времени t энтропия объекта достигнет значения $S_2 > S_1$. Количество информации

в этом предвидении, по (1.18),

$$I_{2,1} = S_2 - S_1. \quad (1.37)$$

Далее положим, что совокупность научных знаний, находящихся в нашем распоряжении, позволяет найти и осуществить способы воздействия на объект, в результате применения которых в течение времени t энтропия объекта достигнет значения $S_3 < S_2$. Количество информации в этих знаниях

$$I_{2,3} = S_2 - S_3. \quad (1.38)$$

Из (1.37)

$$S_2 = S_1 + I_{2,1}. \quad (1.39)$$

Подставляя выражение S_2 в (1.38), находим, что

$$S_3 = S_1 + I_{2,1} - I_{2,3}. \quad (1.40)$$

Если способы, позволяющие сделать энтропию объекта меньшей S_2 через время t , неизвестны, то информация $I_{2,3}$ будет равна нулю и из (1.39) и (1.40) следует, что $S_3 = S_2$ — естественный процесс возрастания энтропии. Чтобы повлиять на этот процесс, уменьшая возрастание энтропии объекта, необходимо дать последнему некоторую негэнтропию N . Уравнение (1.40) показывает, что

$$N = -I_{2,3}. \quad (1.41)$$

Если $I_{2,3} = I_{2,1}$, то негэнтропия N поддерживает энтропию объекта неизменной в течение времени t , при этом $S_3 = S_1$. Если же $I_{2,3} > I_{2,1}$, то в течение времени t энтропия объекта уменьшается, так что $S_3 < S_1$. Эти соотношения показаны на рис. 1.2, где изображен примерный ход изменения энтропии в четырех рассмотренных случаях.

В принципе информация $I_{2,3}$ всегда может быть достаточно большой, чтобы негэнтропия, определяемая по (1.41), дала возможность уменьшить энтропию объекта, характеризующую его состояние по выбранному признаку. Заметим, что как получение этой информации, так и основанные на ней действия, приводящие к уменьшению энтропии объекта, всегда требует совершения некоторой работы, т. е. некоторого, хотя бы и небольшого, расхода энергии. А при этом обязательно происходит обесценение энергии и возрастание физической энтропии.

Возрастание физической энтропии означает соответственное уменьшение физической негэнтропии объекта, наблюдателя и всех предметов, участвующих в рассматриваемых процессах. Согласно *негэнтропийному принципу информации*, указанному Л. Бриллюэном [1.12], информация может быть получена только за счет негэнтропии некоторой физической системы, но может быть превращена в физическую негэнтропию. Выведенные выше соотношения

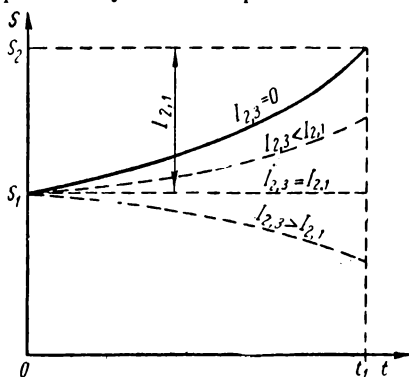


Рис. 1.2. Изменение энтропии управляемого объекта в зависимости от информации о нем

(1.18), (1.36) и связанные с ними являются обобщением этих представлений, получаемых, если рассматривать энтропию как меру неупорядоченности некоторой системы по любому признаку.

Итак, управление неизбежно сопровождается общим возрастанием физической энтропии. Но при этом достигается уменьшение энтропии, являющейся мерой неупорядоченности уп-

равляемого объекта по признаку, определяемому целью управления. И это достигается обязательно, хотя бы в течение ограниченного времени: без этого процесс управления был бы бессмысленным и бесполезным.

Заметим, что полученная информация может быть потеряна. Например, наблюдатель может забыть какие-либо из полученных данных или не может произойти разрушение материала, на котором эти данные зафиксированы. Очевидно, что *потеря информации* происходит вследствие явления роста неупорядоченности, обусловленных вторым законом термодинамики и сопровождаемых возрастанием физической энтропии. Согласно (1.36) всякая потеря информации означает равную по абсолютной величине потерю негэнтропии и эквивалентное возрастание энтропии. Поэтому факт потери информации является *кибернетическим выражением второго закона термодинамики* [1.8]: информация о замкнутой системе стихийно стремится к уменьшению, как энтропия такой системы

стремится к увеличению; информация сохраняется в обратимых процессах и частично теряется в необратимых [1.5].

Следует обратить внимание на различный смысл явлений потери информации и потери энергии. Потеря энергии означает ее преобразование в энергию другого вида по закону сохранения. Потеря информации означает ее уничтожение. Это показывает, что для информации нет закона ее сохранения: ее потеря приводит лишь к возрастанию энтропии системы оператор — объект.

Отличие от энергетики состоит здесь еще в том, что за счет некоторого количества энергии можно совершить работу только один раз, тогда как одной и той же информацией можно пользоваться в неограниченном числе случаев. Это наглядно поясняет следующий пример. Профессор, читающий лекцию, сообщает некоторое количество информации студентам. При этом обладателем этого количества информации остается сам профессор и становится каждый студент, слушающий лекцию, причем каждый из них может, если только информация не будет потеряна, пользоваться ею для эквивалентного уменьшения энтропии объектов, к которым информация относится, в любом числе случаев. Существует, таким образом, принцип *многократного применения информации*. Врач, например, пользуясь одними теми же данными медицины, лечит многих больных, достигая в каждом случае излечения, снижения энтропии организма больного. Уменьшение энтропии организма больного эквивалентно информации, которую содержат применяемые научные данные. Это вполне понятно, если исходить из определения информации как разности энтропий, простым следствием которого является принцип ее многократного применения.

Отсюда вытекают неограниченные возможности науки в борьбе со «случайным распорядком событий в природе», которые предвидел Умов (см. § 1.2), — возможности управления природой, реализуемые в кибернетике созданием технических систем управления, преодолевающих этот стихийный распорядок с помощью информации. В процессах получения и применения информации осуществляется наше приспособление к случайностям внешней среды и наша жизнедеятельность. «Действенно жить, — замечает в связи с этим Винер, — это значит жить, располагая правильной информацией» [1.8].

Пример. Топографический каталог библиотеки содержит сведения о месте, на котором должна находиться каждая книга. Найдем, чему равна заключенная в нем информация, если библиотека состоит из b томов. Для этого предположим, что все b книг побывали в чтении, после чего их ставили куда попало, не сообразуясь с топографическим каталогом. В результате этого библиотека пришла в состояние беспорядка, при котором возможное число ее равновероятных состояний $n = b!$ (см. § 1.2). При этом, по формуле (1.11), ее энтропия по топографическому признаку

$$S_0 = \log_2 (b!).$$

Расставив книги в соответствие с топографическим каталогом, приведем этим библиотеку в единственно возможное состояние, когда ее энтропия

$$S = 0,$$

как следует из той же формулы при $n = 1$. По формуле (1.18) информация, с помощью которой энтропия библиотеки уменьшилась с S_0 до S ,

$$I = \log_2 (b!).$$

Если, например, $b = 100\,000$, то, пользуясь формулой Стирлинга и принимая во внимание большое значение b , имеем:

$$\ln (b!) \approx 10^5 (\ln 10^5 - 1) \approx 1,05 \cdot 10^6.$$

Так как

$$\log_2 x = \frac{\ln x}{\ln 2},$$

а

$$\ln 2 = 0,6931,$$

из предыдущего результата получаем:

$$I = \frac{1,05 \cdot 10^6}{0,693},$$

т. е. приблизительно полтора миллиона двоичных единиц.

§ 1.4. ИНФОРМАЦИОННЫЕ МАШИНЫ

Всякое техническое устройство, совершающее целенаправленные действия, будем называть *машиной*. По сущности своих действий все машины разделяются на *энергетические* и *информационные*. Назначение первых — преобразование энергии, вторых — преобразование информации. Примерами того и другого рода машин могут служить паровоз и телевизор. В информационных машинах также происходят преобразования энергии, но не они определяют эффективность действия таких машин, всецело зависящую от преобразований информации. Совершенство

телефона, например, определяется не энергетикой преобразования акустических колебаний в электрические и обратно, а отсутствием искажений в передаче речи. Улучшение энергетических характеристик таких устройств не означает улучшения передачи информации, при ухудшении их информация может передаваться не хуже. В некоторых информационных машинах, например вычислительных, расходуемая энергия в конечном итоге полностью превращается в тепло.

Мощность энергетической машины определяется работой, которую эта машина может произвести в единицу времени. *Информационная мощность* есть количество информации в единицу времени и характеризует способность информационной машины выдавать информацию. Единицей информационной мощности может быть, например, бит в секунду.

Информационными машинами являются все управляющие системы или устройства. Они всегда имеют в своём составе энергетические элементы, производящие работу. Но действие этих систем состоит в получении, обработке и выдаче информации о том, какая работа, как и когда должна быть произведена. Как бы ни были совершенны энергетические элементы системы, она не сможет выполнять своих задач, если информация не будет доходить до этих элементов достаточно своевременно и точно. Следовательно, передача и преобразование информации — основа действия управляющей системы.

Функции управляющих систем в технике первоначально выполняли люди, управлявшие, например, поездами, самолетами, электростанциями, технологическими процессами и артиллерийскими установками. Затем началась автоматизация процессов контроля и управления. Автопилот и автомашинист заменяют человека в управлении самолетом и поездом; автоматизация электростанций и производственных процессов сокращает и здесь участие человека в управлении; артиллерийские следящие приводы заменяют наводчиков. Автоматизация процессов управления означает не только сокращение и облегчение человеческого труда, но и возможность решать задачи управления быстрее и точнее, чем это может сделать человек. Управление такими объектами, как самолет, энергосистема или артиллерийская установка, требует быстрого принятия решений на основе учета различных

и иногда многочисленных данных, т. е. на основе быстрой обработки большой информации. Люди вследствие физиологических ограничений не могут этого сделать так быстро, как требуется. То же имеет место в некоторых производственных процессах. Много больше позволяют достичь *автоматические управляющие системы (АУС)*, в особенности с применением *электронных управляющих машин (ЭУМ)*.

Электронные управляющие машины — это вычислительные машины, снабженные специальными устройствами, преобразующими информацию, поступающую в машину и выдаваемую ею. *Универсальные цифровые вычислительные машины* могут перерабатывать информацию по любой программе, выполняя вычисления и логические операции. Они являются поэтому универсальными устройствами, пригодными для решения разнообразных задач и для применения в АУС различного назначения. *Цифровые вычислительные машины (ЦВМ)* могут решать математические, в частности геометрические, задачи, участвовать в управлении технологическими процессами, например металлургическими и т. п.

ЦВМ моделируют, таким образом, мыслительные процессы, выполняя некоторые операции умственного труда (иногда это называют механизацией умственного труда). Программа, заложенная в машину, содержит совокупность правил, по которым машина перерабатывает поступающую в нее информацию. Пользуясь этими правилами, она может, так же как и человек, получить новые выводы. Кеплер открыл законы движения планет, обрабатывая по некоторым правилам результаты астрономических наблюдений Тихо де Браге. То же самое теперь могла бы сделать машина. Но машина сделала бы это значительно быстрее Кеплера, как несравненно быстрее людей машины предвычисляют элементы траекторий космических кораблей. Подобный же пример дает история открытия совершенных чисел (совершенное число равно сумме всех его делителей, кроме него самого; это, например, число $6 = 1 + 2 + 3$). За много веков было найдено лишь 12 таких чисел. Но примененная для этой цели электронная машина SWAC Калифорнийского университета нашла два новых совершенных числа за один день, а затем обнаружила еще три таких числа [1.17]. Тем самым машина, во-первых, получила новые результаты и, таким образом, сделала откры-

тие; во-вторых, она выполнила необходимую для этого умственную работу с быстротой, недоступной человеку.

В приведенных примерах информация, получаемая в результате открытий, содержалась в информации, обработка которой привела к этим открытиям. Кто бы ни обрабатывал информацию, полученную Тихо де Браге, — человек или машина, — они, действуя по определенной программе, лишь выявили бы то, что в скрытом виде содержалось в этой информации. Так и уравнение в скрытом виде содержит информацию, получаемую явно в его решении. Обработка информации по заданной программе — *детерминированный процесс*, цель которого — обнаружение *скрытой информации*. От этого отличается процесс *генерирования информации*. Создаваемая в таком процессе *новая информация* не может быть получена никакой программной обработкой какой-либо другой информации, которая могла бы быть предоставлена человеку или машине. Генерирование информации — *самоорганизующийся процесс* образования порядка из хаоса [1.19], осуществляемый *случайным поиском* с учетом вероятности успеха и с оценкой результата. Такие процессы имеют место в научном и художественном творчестве, а также в машинах, опознающих образы (*перцептроны*). «Научный закон, — замечает Бриллюэн, — является не только выражением определенного количества опытных фактов; в этих законах отражается мышление ученого: отбор фактов, сравнение, фантазия, проблеск гения» [1.18]. Очевидно, что мыслительная деятельность, как и все процессы, в которых возникает новая информация, не подчиняется второму закону термодинамики. Кибернетическая форма второго закона показывает это с особенной ясностью.

То обстоятельство, что в детерминированных процессах действие машин происходит по программе, данной им человеком, не означает, что они не могут открывать новое. Ведь человек, составляющий программу, не знает, какого вида результат получится на машине. Однако, давая машине программу, человек ставит перед ней определенную цель. Такова, например, цель — найти следующее совершенное число, но какое это число — составитель программы не знает. Отсюда ясно, что в таких случаях машина открывает новое по заданию человека. И отсюда следует, что она не может полностью заменить человека хотя бы в науке. Как заметил профессор В. В. Солодов-

ников, «человек может создать машину, а машина не может создать человека» [1.15]. Это относится и к *самопрограммирующим* машинам, действующим по автоматически ими же составляемым программам, целенаправленность которых в общем определяется конструктором машины.

Однако машина может превзойти своего создателя по скорости действия и точности вычислений. Кроме того, машина может обладать информацией об опыте и знаниях всех специалистов в какой-либо области, а не одного лишь из них. И наконец, машина, способная к самообучению, в связи с быстротой своих действий обучается значительно быстрее человека. Но в остальных отношениях никакого превосходства перед человеком у машин нет. Более того, машина весьма примитивна по сравнению с мозгом. Ее устройство и действие поддаются более или менее простому объяснению, тогда как человеческий мозг настолько сложен, что полностью еще нельзя постигнуть, как он работает. Работа мозга показывает, что возможно создание машины, действие которой основано на совершенно новых и высокоэффективных принципах [1.25]. Различие между человеком и машиной не только количественное, зависящее от сложности устройства, но и качественное. Если человек не может работать так, как работает машина, то и машина не может работать так, как работает человек. И наилучшее решение — работа человека с помощью машины.

Способность машин воздерживаться от нецелесообразных попыток, учитывать опыт, планировать исследования и т. п. позволяет говорить об *искусственном разуме*, которым они обладают. Однако привлекающий большое внимание вопрос о том, может ли машина мыслить, лишен практического значения. Винер дал на него следующий ответ: «Современные вычислительные машины могут обучаться, улучшать свою работу, оценивая ее результаты. Назвать ли это мышлением или нет — вопрос терминологии» [1.23].

Несомненно, что в соединении с машиной человеческий мозг может познавать больше того, что могут познать люди без машин. Информационная машина действует, следовательно, как *усилитель мыслительных способностей* [1.7]. Этот новый вид усилителя полезно сопоставить с давно известными усилителями мощности. Рассмотрим две

задачи: подъем тяжелого груза и длинное вычисление. Если груз достаточно тяжел, то один человек не сможет его поднять, сколько бы времени он ни старался это сделать. Но совместными усилиями многих людей груз может быть поднят, что и делалось, например, на строительстве египетских пирамид. С развитием техники ту же работу стал делать подъемный кран, во много раз сократив количество участвующих в ней людей. Вычисление производится, как правило, последовательными этапами: результат, полученный на первом этапе, передается на второй, полученный на втором — на третий и т. д. Поэтому время, потребное для такой работы отдельному человеку, не может быть существенно сокращено привлечением к ней сколь угодно большого числа других людей. Но вычислительная машина сокращает это время во много раз. В случае подъема груза машина выполняет мышечную работу многих людей. В случае вычисления машина исполняет в короткий срок умственную работу, на которую человеку не хватило бы, возможно, всей его жизни. За достаточно короткий срок информационная машина может сделать много таких вычислений и из сопоставления их результатов получить выводы, к которым человек никогда не мог бы прийти. Допуская, что в обоих рассмотренных случаях работой машины управляет один человек, видно, что подъемная машина усиливает механическую мощность его мышц, а информационная — информационную мощность его мозга. Подъемный кран усиливает механическую мощность во столько раз, сколько людей, могущих поднять тот же груз, он заменяет, так как он производит одинаковую с ними работу. Информационная машина выдает столько же информации за короткое время T_m , сколько ее выдает один человек за долгое время T_q . Поэтому усиление информационной мощности этой машиной выражается отношением T_q/T_m , т. е. определяется быстродействием машины.

Пример. Оценим информацию, получаемую в результате открытия совершенного числа N_2 , следующего за ранее известным совершенным числом N_1 . Если до нахождения числа N_2 можно было думать, что следующим совершенным числом является любое из целых чисел, следующих за N_1 , то после определения N_2 была бы получена информация, которую найдем по (1.23), полагая, что $n_0 = N_2 - N_1$ и $n = 1$:

$$I = \log_2 (N_2 - N_1).$$

Эту информацию получили бы и от вычислительной машины, которая нашла бы число N_2 . В действительности информация, получаемая при открытии числа N_2 , меньше, так как маловероятно, что совершенным окажется одно из чисел, близких к N_1 .

§ 1.5. КОДИРОВАНИЕ ИНФОРМАЦИИ

Как для передачи, так и для хранения информации всегда необходим тот или иной *код*, т.е. набор *символов*, с помощью которых может быть выражена и зафиксирована информация. Таким кодом является, например, любой алфавит, буквы которого служат символами для записи информации. Представление информации в символах какого-либо кода называют *кодированием*, переход к другому коду — *перекодированием*, возвращение к исходному коду — *декодированием*.

Количество символов в коде может быть различным. В простейших случаях достаточен код, состоящий лишь из двух символов. Это имело место, например, в железнодорожной сигнализации с помощью красного и зеленого сигналов семафоров. В общем случае может требоваться код, содержащий n символов. Рассмотрим сообщение, для передачи которого таким кодом нужна последовательность некоторого большого числа m символов. Положим, что символы кода встречаются в этой последовательности с априорными вероятностями P_{0i} ($i = 1, 2, \dots, n$), значения которых образуют полную систему вероятностей, т.е. должны удовлетворять условию

$$\sum_{i=1}^n P_{0i} = 1.$$

Тогда каждый из n символов встречается в рассматриваемой последовательности mP_{0i} раз, так что

$$\sum_{i=1}^n mP_{0i} = m.$$

Предположим, что прием сообщения происходит без помех. Тогда информация, получаемая от каждого символа, по (1.29), равна

$$-\log_2 P_{0i}.$$

Совокупность этих символов, входящая в рассматриваемую последовательность, содержит информацию

$$I_i = m P_{0i} \log_2 P_{0i},$$

а все сообщение — информацию

$$I = -m \sum_{i=1}^n P_{0i} \log_2 P_{0i}. \quad (1.42)$$

Отсюда следует *формула Шеннона*, выражающая *среднее количество информации на символ*

$$I_{\text{ср}} = - \sum_{i=1}^n P_{0i} \log_2 P_{0i}. \quad (1.43)$$

Эта формула может быть истолкована следующим образом [1.24]. Ожидая появления очередного символа в сообщении, предполагаем, что это может быть любой из n символов, характеризуемый вероятностью появления P_{0i} . Таким образом, предполагаем, что место ожидаемого символа в сообщении может иметь n различных состояний, каждое из которых ожидаем с вероятностью P_{0i} . Это значит, что рассматриваемое место сообщения характеризуется неупорядоченностью, мерой которой служит энтропия S_0 по (1.31). В результате приема, происходящего без помех, появляется некоторый вполне определенный k -й символ, для которого, следовательно, $P_k = 1$, так что для остальных $n - 1$ символов $P_i = 0$. Поэтому после приема очередного символа энтропия его места в сообщении, по (1.12), $S = 0$. Следовательно, информация, получаемая при появлении очередного символа, вычисляемая по (1.18),

$$I = S_0. \quad (1.44)$$

Подставляя сюда (1.31), находим, что каждый символ сообщения, принимаемого без помех, дает информацию, выражаемую формулой Шеннона¹.

Из формулы (1.43) следует, что в случае когда вероятность появления всех символов кода одинакова, т.е.

$$P_{0i} = \frac{1}{n},$$

¹ Информацию, получаемую по (1.44), Шеннон называет энтропией, подразумевая энтропию S_0 , которой эта информация численно равна.

$I_{\text{ср}}$ имеет наибольшее значение, как и энтропия S_0 по (1.31):

$$I_{\text{ср. макс}} = \log_2 n. \quad (1.45)$$

Это наибольшее значение среднего количества информации на символ, возможное при применении кода, состоящего из n символов, называемое *информационной емкостью кода*. *Экономичность* кода в конкретном случае его применения для передачи какого-либо сообщения оценивается отношением

$$\mathcal{E} = \frac{I_{\text{ср}}}{I_{\text{ср. макс}}}.$$

Для оптимального кода $\mathcal{E} = 1$. *Избыточность кода*

$$I = 1 - \mathcal{E}$$

служит мерой числа лишних символов в сообщении. У оптимального кода $I = 0$.

Из (1.45) следует, что наибольшая информация, которую может содержать сообщение, состоящее из m символов n — символьного кода,

$$I_{\text{макс}} = m \log_2 n.$$

Наименьшее же количество символов, необходимое для передачи данным кодом сообщения, содержащего информацию I ,

$$m_{\text{мин}} = \frac{I}{\log_2 n}.$$

Последние формулы показывают, что количество информации в сообщении зависит от кода. Это следует и из равенства (1.44), показывающего, что информация I в ее количественном выражении является лишь информацией о появлении в сообщении символов кода, с помощью которого оно передается.

Поэтому при передаче информации с заданной частотой появления символов, информация, передаваемая в единицу времени (*поток информации*), зависит от кода. При наличии помех канал связи (или любой элемент системы) может пропускать в единицу времени без искажения информацию, не превышающую некоторого значения, называемого *пропускной способностью канала* [1.22]. Решение, при котором поток информации менее пропускной

способности, неэкономично. Если же поток больше пропускной способности, передача становится несовершенной. Выбором подходящего кода можно приблизиться к наилучшим условиям передачи, при которых поток информации равен пропускной способности канала.

Выбор кода определяется конкретными условиями устройства и работы элементов системы, между которыми происходит передача информации. В каждом случае возможно применение различных кодов, один из которых может оказаться наиболее подходящим. Для дискретных сигналов, например, может быть применен цифровой код, основанный на той или иной позиционной *системе счисления*, причем каждый сигнал должен соответствовать одной из цифр системы и *разряду*, в котором эта цифра стоит. Применяются шестнадцатиричная, десятичная, восьмеричная, троичная, двоичная и некоторые другие системы счисления. Название каждой из этих систем определяется числом, принятым за ее *основание*. Основание системы счисления равно отношению между единицами соседних разрядов и количеству различающихся цифр, необходимых для представления всевозможных натуральных чисел в этой системе. Так, в десятичной системе (счет десятками) соседние разряды различаются в 10 раз и нужны 10 цифр (от 0 до 9), в восьмеричной (счет восьмерками) — в 8 раз и нужны 8 цифр (от 0 до 7), в двоичной (счет парами) — в два раза и нужны 2 цифры (0 и 1). Любое число представляется в каждой системе счисления суммой всех степеней основания от нулевой до высшей, нужной для получения данного числа, взятых с некоторыми коэффициентами, не превышающими основания. Например, число 16 в десятичной системе состоит из одного десятка в первой степени и шести десятков в нулевой степени

$$1 \cdot 10^1 + 6 \cdot 10^0 = 16.$$

То же число в восьмеричной системе надо представить как две восьмерки в первой степени и нуль восьмерок в нулевой степени, т. е. 20. В двоичной же системе это число изображается как одна двойка в четвертой степени и по нулю двоек в третьей, второй, первой и нулевой степенях, т. е. 10 000.

Преимуществом двоичных кодов, основанных на двоичной системе счисления, является наименьшее возможное

количество символов; для них $n = 2$. Поэтому устройства, в которых применяются такие коды, должны иметь лишь два различных устойчивых состояния, действуя по принципам «да» или «нет», «все» или «ничего» и т. п. Такие устройства существуют в технике и в органической природе.

Цепочки дизооксирибонуклеиновой кислоты, входящие в состав хромосом, имеют столько структурных элементов, что с помощью двоичного кода в единичной клетке человека можно зашифровать около тысячи больших учебников [1. 6]. «Ярким свидетельством единства процессов управления, — отметил академик М. В. Келдыш, — является то, что генетическими процессами управляют известные уже в технике коды. Это величайшее открытие означает, что и на эти процессы может быть распространена теория информации и теория управления» [1. 20].

Для технических устройств (в частности, для ЦВМ) важно, чтобы запись чисел достигалась не только наименьшим количеством символов, но и наименьшим количеством разрядов. А так как количество разрядов, необходимых для записи больших чисел, увеличивается с уменьшением основания системы счисления, оба условия осуществить нельзя и следует выбирать *оптимальную систему счисления*. Оптимальной целесообразно признать такую систему, которая позволяет обойтись наименьшим количеством аппаратуры. Это количество можно считать пропорциональным произведению числа символов n на число разрядов N . Наибольшее число, которое можно записать в N разрядах системы, имеющей n символов (т. е. основание n),

$$M = n^N - 1. \quad (1.46)$$

Так в десятичной системе, ограничиваясь тремя разрядами, можно записать число, не превышающее $M = 10^3 - 1 = 999$.

Из (1.46) находим, что

$$N = \frac{\ln(M + 1)}{\ln n}.$$

Следовательно, количество аппаратуры, необходимое для воспроизведения числа M в N разрядах системы

счисления с основанием n , пропорционально величине

$$nN = \frac{n}{\ln n} \ln(M + 1). \quad (1.47)$$

Приравнивая нулю производную правой части (1.47) по n , находим условие минимума количества аппаратуры при данных N и M :

$$\ln n - 1 = 0,$$

отсюда: $n = e$, т. е. оптимальной является система счисления с основанием, равным неперову числу. Но так как основание должно быть целым числом, наилучшее приближение к оптимальной системе осуществляется троичной системой счисления, а несколько худшее — двоичной.

Предположим, что требуется перекодировать сообщение, переданное некоторым кодом, в двоичный код. На первый взгляд наиболее простым решением этой задачи будет применение *равномерного кода*, в котором каждый из символов исходного кода выражается одинаковым и наименьшим возможным числом двоичных символов. Легко, однако, убедиться, что экономичнее будет *код Шеннона—Фано*, составляемый по следующему правилу. Выписываем символы исходного кода в порядке убывания их частот и разделяем их на две группы, суммарная частота символов каждой из которых точно или приблизительно равна $1/2$. В обозначении символов первой группы двоичным кодом на первое место ставим 0, а в обозначении символов второй группы на первое место ставим 1. Затем обе группы разделяем еще на две, чтобы суммарная частота символов каждой подгруппы была приблизительно равна $1/4$. На второе место в обозначении символов первой и второй подгруппы каждой группы ставим соответственно 0 и 1. Процесс образования подгрупп продолжается до тех пор, пока не получаются подгруппы, состоящие из одного символа. Легко видеть, что в коде Шеннона—Фано наименьшее число двоичных знаков расходуется на более вероятные символы и наоборот. Например, если в сообщении следующие четыре символа встречаются с подписанными под ними частотами [1.22]

$$\begin{array}{cccc} A & B & C & D \\ 1/2 & 1/4 & 1/8 & 1/8, \end{array} \quad (1.48)$$

то код Шеннона—Фано строится так:

$$\begin{array}{l} A \frac{1}{2} \} 0 \\ B \frac{1}{4} \} \} 10 \\ C \frac{1}{8} \} 1 \} \} 110 \\ D \frac{1}{8} \} \} 11 \} 111. \end{array} \quad (1.49)$$

Понятие информации позволяет количественно оценивать сообщения независимо от того, каким способом эти сообщения переданы и приняты, а также от их содержания и смысла. Для техники связи этого вполне достаточно, так как ее средства должны обеспечивать передачу сообщений любого содержания без потери информации. Для кибернетики этого мало. Сообщения, необходимые для управления, должны иметь надлежащий *смысл*. Следовательно, количественное выражение информации, которую содержит сообщение, должно быть связано с его смыслом. Чтобы выявить эту связь, заметим, что изменение количества информации при кодировании никак не влияет на смысл передаваемого сообщения, если не происходит искажений вследствие помех. Смысл сообщения вскрывается при декодировании, т. е. при переходе к исходным символам сообщения, содержащего информацию, связанную с энтропией системы, к которой оно относится [1.24]. Действительно, как показывает равенство (1.34), существует количество информации, непосредственно определяющее изменение состояния объекта, характеризуемого энтропией S или нэгэнтропией N . Именно это количество информации и связано со смыслом сообщения, определяющего указанное изменение состояния объекта. Такая связь утрачивается при перекодировании, когда количество информации изменяется, но восстанавливается при декодировании в тот код, в котором восстанавливается связь информации с энтропией или нэгэнтропией объекта управления. Количество информации, получаемое в таком коде, и является полезным, а *ценность сообщения* определяется возможностью его осмысленного декодирования в этот код и связана таким образом с полезной информацией. Такое декодирование даст осмысленный результат, если при перекодировании не произошло искажений. Количество информации в сообщении, переданном двоичным кодом, не изменяется, если нули и единицы почему-либо

поменялись местами. Но осмысленное декодирование при этом уже невозможно.

Вернемся к уже рассматривавшемуся примеру с библиотекой из b книг (см. § 1.2), упорядоченность которых определяется по топографическому признаку. Информацию, необходимую для поддержания порядка, содержит топографический каталог. Если им не пользуются, то по истечении времени, в течение которого все b книг побывали в чтении, энтропия библиотеки может достичь

$$S_0 = \log_2 (b!).$$

Поддерживая же порядок с помощью топографического каталога, можно иметь к тому же времени энтропию

$$S = 0.$$

Следовательно, по (1.18) топографический каталог содержит информацию, связанную с энтропией библиотеки по топографическому признаку:

$$I = \log_2 (b!). \quad (1.50)$$

Эта информация закодирована в каталоге, каждая карточка которого служит символом одной из книг. Код-каталог состоит, таким образом, из b символов, на каждый из которых приходится информация

$$I_{\text{ср}} = \frac{1}{b} \log_2 (b!),$$

так что информация, содержащаяся в каталоге, определяет возможное уменьшение энтропии библиотеки. Вместе с тем на каждой карточке топографического каталога имеется шифр книги и обозначение ее места в библиотеке. Это другой код — алфавитно-цифровой. В нем каждая карточка содержит информацию $I'_{\text{ср}}$, определяемую частотой появления отдельных букв и цифр, а каталог в целом — информацию

$$I' = bI'_{\text{ср}},$$

отличающуюся от I по (1.50) и не определяющую возможного уменьшения энтропии библиотеки. Но от алфавитно-цифрового кода можно перейти к коду-каталогу с его b символами-карточками. И такое декодирование даст информацию, определяющую возможности упорядочения библиотеки. Очевидно, что ценность или смысл для

упорядочения библиотеки имеет лишь информация I , которую содержит код-каталог. Информация же I' или получаемая в другом любом коде ценна для той же цели лишь при том условии, что ее содержание может быть декодировано в код-каталог.

При наличии помех выбор кода уже не может быть подчинен только стремлению к передаче сообщения наименьшим количеством символов. В этом случае необходимо обеспечить безошибочный прием сообщения. Безошибочный прием достигается с помощью добавочных символов, т. е. в ущерб скорости передачи сообщения. В двоичном коде, например, ошибка возникает всякий раз, когда вместо 0 применяют 1 или наоборот. Существуют коды, позволяющие обнаруживать и исправлять ошибки. Рассмотрим простейший пример такого кода.

Для обнаружения и исправления ошибок в переданных сообщениях добавляют к символам, несущим информацию (*информационным*), символы, предназначенные для проверки (*поверочные*). Таким путем получаются *самокорректирующиеся коды*, позволяющие обнаружить и исправить ошибку. В двоичном коде ошибка может состоять лишь в том, что 0 оказался на месте 1 или наоборот. Следовательно, если установлено место ошибки в сообщении, то она сейчас же может быть исправлена. Общий принцип построения двоичного кода, обнаруживающего ошибки, состоит в том, что из сообщения делаются выборки символов по определенному правилу, причем такие выборки должны удовлетворять некоторому признаку. Если этот признак оказывается нарушенным, то имеется ошибка, место которой обнаруживается по выборкам с нарушением. Как обнаружение, так и исправление ошибок может быть автоматизировано.

Предполагая, например, что в последовательности семи символов может быть не более одной ошибки, передают сообщение *блоками*, содержащими сначала четыре информационных символа, а затем три поверочных. Поверочные символы выбираются так, чтобы в выборке, состоящей из трех информационных символов и одного поверочного, число нулей, как и число единиц, было четным. Выборки же делаются так, чтобы в них всегда входил информационный символ, стоящий в блоке на первом месте, а каждый из прочих информационных символов входил не более, чем в две выборки; каждый поверочный

символ входит лишь в одну выборку. Таким образом, может быть три выборки. Это код «четыре из семи».

Пусть один из блоков сообщения, переданного таким кодом, записан в первой строке таблицы. Выборки, сделанные по описанному правилу, содержат следующие строки таблицы:

```

1 0 0 0 1 1 1
1 0 0   1
1 0   0  1
1   0 0   1.

```

Во всех выборках число нулей и единиц четное, следовательно, ошибки нет. Легко убедиться, что в случае нечетности в одной выборке должен быть ошибочным входящий в нее поверочный символ. Если же нечетность имеется в двух выборках, то ошибочен символ, входящий в обе. Если, наконец, нечетность есть во всех выборках, то ошибочен входящий в них первый информационный символ.

Заметим, наконец, что всякий код должен быть составлен так, чтобы не могло быть неопределенности при расшифровке символов. Нельзя, например, применять перекодирование

$$\begin{matrix} A & B & C \\ 0 & 1 & 01, \end{matrix}$$

так как тогда не будет известно, что означает последовательность 01: AB или C .

Пример. Перекодируем сообщение, переданное четырехбуквенным кодом, в три варианта двоичного кода, указанные в табл. 1.1 [1.24].

Таблица 1.1

Три варианта перекодирования четырехбуквенного кода в двоичный

Исходный код	A	B	C	D
Код I	00	01	10	11
» II	01	10	00	11
» III	0	10	110	111

Таблица содержит два варианта равномерного кода и код Шеннона — Фано, составленный, как сделано выше, по заданным ап-

приорным вероятностям появления символов исходного кода

$$P_0(A) = \frac{1}{2}; \quad P_0(B) = \frac{1}{4}; \quad P_0(C) = P_0(D) = \frac{1}{8}.$$

Полагая, что в соответствии с заданными вероятностями символов исходного кода на четыре появления буквы *A* должно быть два появления буквы *B* и по одному появлению *C* и *D*, а всего восемь символов, нетрудно найти, что в коде I при этом будет 11 нулей и 5 единиц, в коде II — по восемь нулей и единиц, а в коде III — по 7 нулей и единиц. Таким образом, на 8 символов исходного кода равномерные коды будут давать по 16 символов, а код Шеннона — Фано — 14 символов. Теперь легко подсчитать, что вероятности появления символов двоичного кода имеют значения, приведенные в табл. 1.2

Т а б л и ц а 1.2

Априорные вероятности появления символов
двоичных кодов в таблице 1.1

Код	I	II	III
$P_0(0)$	11/16	1/2	1/2
$P_0(1)$	5/16	1/2	1/2

Вычислим количество информации, получаемое от указанного числа символов рассматриваемых кодов, пользуясь формулой (1.43). Для исходного кода среднее количество информации, приходящееся на символ,

$$I_{\text{ср}} = - \left(\frac{1}{2} \log_2 \frac{1}{2} + \frac{1}{4} \log_2 \frac{1}{4} + \frac{2}{8} \log_2 \frac{1}{8} \right) = \frac{7}{4} \frac{\text{дв. ед.}}{\text{символ}}$$

и информация

$$I = \frac{7}{4} 8 = 14 \text{ дв. ед.}$$

Для кода I

$$I_{\text{ср}} = - \left(\frac{11}{16} \log_2 \frac{11}{16} + \frac{5}{16} \log_2 \frac{5}{16} \right) \approx 0,896 \frac{\text{дв. ед.}}{\text{символ}},$$

$$I = 0,896 \cdot 16 \approx 14,34 \text{ дв. ед.}$$

Для кодов II и III

$$I_{\text{ср}} = - 2 \frac{1}{2} \log_2 \frac{1}{2} = 1 \frac{\text{дв. ед.}}{\text{символ}};$$

в первом из них

$$I = 1 \cdot 16 = 16 \text{ дв. ед.},$$

во втором

$$I = 1 \cdot 14 = 14 \text{ дв. ед.}$$

ПРИНЦИПЫ ДЕЙСТВИЯ И СПОСОБЫ ОПИСАНИЯ
АВТОМАТИЧЕСКИХ СИСТЕМ

§ 2.1. ЗАДАЧИ ТЕХНИЧЕСКОЙ КИБЕРНЕТИКИ

Управляющие системы, существующие в органической природе, весьма совершенны. Действие их наблюдается в управлении движением, когда, например, птица налету ловит насекомое, собака гонится за добычей, человек играет в мяч. Не менее совершенно построено управление внутренними процессами в живом организме, в котором регулируется температура, состав и давление крови, деятельность желез внутренней секреции и т. п. Температура здорового человеческого тела, как и тела всякого теплокровного животного, остается неизменной, когда температура окружающей среды изменяется в широких пределах. В этом состоит свойство гомойотермии, делающее такие организмы независимыми от возможных в естественных условиях температурных воздействий. Та же задача осуществляется автоматическими регуляторами температуры в различных производственных процессах, в термостатах и холодильниках, в установках кондиционирования воздуха и других технических устройствах. Подобным же образом встречается необходимость технического осуществления и других функций *органических управляющих систем* (ОУС). Высокое совершенство этих систем указывает на полезность их изучения для применения в технике принципов их действия и устройства. Это и делает кибернетика, производя заимствования из области управления важнейшими функциями живых организмов. Так, тенденция развития живых организмов к ослаблению зависимости от воздействий внешней среды находит отражение в создании *помехоустойчивых автоматических систем*. Такое заимствование требует математического описания высшей нервной деятельности, которое Винер ввел в кибернетику, где оно получило дальнейшее развитие. Но первый шаг на этом пути сделал Н. А. Романов еще

до появления современной кибернетики [2.2], исходя из предвидения И. П. Павлова, утверждавшего возможность математического анализа явлений жизни [2.3].

Прежде были лишь отдельные попытки применения физико-математических методов в биологии и имели место лишь отдельные случаи переноса явлений живой природы в технику. В кибернетике началось планомерное объединение достижений биологических, физико-математических и технических наук. Это позволяет не только применять в АУС решения, находимые в ОУС, но и изучать последнее, моделируя их техническими средствами. Автомат, функционирующий подобно организму и похожий на него по структуре, служит моделью этого организма, позволяющей изучать хотя бы некоторые явления, недоступные исследованию в живой природе. Это относится прежде всего к явлениям высшей нервной деятельности и к работе мозга. Распространение тех же методов на другие явления органической природы привело затем к образованию новой науки—*бионики*. Примером решаемых ею задач может служить изучение свойств и особенностей дельфинов для применения этих данных в строительстве подводных лодок. Кибернетика заимствует у органической природы то, что относится к процессам управления. Развившаяся из нее бионика заимствует все остальное, что может быть полезным в технике. Полезным же тут может быть многое, поскольку в некоторых областях, как отметил академик А. И. Берг, «достижения живой природы намного выше наших результатов» [2.19]. К таким областям относится, например, локация.

В природе почерпнуто основное положение кибернетики о значении и возможности борьбы с возрастанием физической энтропии посредством управления (см. § 1.2). Это осуществляется с помощью устройств, сущность действия которых состоит в передаче и переработки информации. Такие устройства должны обладать свойством приспособления к условиям, в которых они действуют, а для этого — гибкостью структуры и целенаправленностью ее изменений, как это имеет место в ОУС [2.12]. Изучением и созданием таких устройств занимается *техническая кибернетика*. Иначе говоря, техническая кибернетика изучает теорию и методы исследования и построения информационных машин в виде автоматических управляющих

систем. При этом заимствование у органических систем не означает и не должно означать их простого копирования, которое никогда не приводит к успеху, как не привели к успеху попытки людей летать с помощью крыльев. Рационально построенные информационные машины, в которых применены указанные выше и другие принципы действия, наблюдаемые в природе, могут быть сделаны более экономичными и эффективными, чем системы, перерабатывающие информацию в ОУС.

Техническая кибернетика является основой *автоматизации*, общее значение которой было рассмотрено § 1.4. Автоматизация необходима в условиях, исключающих присутствие людей, функции которых в таких случаях передаются автоматическим устройствам. Управление процессами, развивающимися с большими скоростями, также требует обязательной автоматизации, так как человек в таких условиях не успевает ни управлять процессом, ни принимать меры, необходимые для предотвращения аварии, даже замечая ее угрозу. Это относится, например, к управлению ядерными реакторами. В других случаях введение автоматизации должно быть обосновано экономически. Технология автоматизируемого производства обычно должна быть перестроена применительно к условиям автоматического управления. Здесь требуется совместная работа технологов и кибернетиков, так как без участия последних достаточно совершенная автоматизация не может быть создана. Это относится в особенности к *комплексной автоматизации*, при которой должны быть автоматизированы не только технологические процессы, но и процессы управления производством.

Пример. Наименьшее время, необходимое человеку для осуществления реакции на принятую информацию, можно принять равным 0,1 сек. Найдем наименьшую скорость v самолета, при которой пилот должен быть заменен автопилотом, чтобы не произошло столкновение двух самолетов, движущихся навстречу друг другу, когда один из них показывается из облаков на расстоянии 150 м от другого [2.10].

Принимая во внимание, что относительная скорость машины при встречном движении равна $2v$, находим, что для предупреждения их столкновения должно выполняться неравенство

$$\frac{150}{2v} \leq 0,1,$$

откуда

$$v = 750 \text{ м/сек.}$$

§ 2.2. ПРИНЦИПЫ ДЕЙСТВИЯ АВТОМАТИЧЕСКИХ СИСТЕМ

Автоматической системой называют всякое устройство действующее без участия человека. Человек строит такую систему, придавая ей свойства, нужные для решения той или иной задачи, пускает ее в ход (включает) и затем вмешивается лишь при необходимости перестроить систему или устранить возникшие в ней неполадки.

По принципу действия все автоматические системы разделяются на две основные группы: *разомкнутые системы*, не корректирующие процесс управления по результатам своей работы, и *замкнутые системы*, осуществляющие такую корректировку с помощью контроля результатов управления.

В реальных условиях автоматическая система обычно находится под воздействием большого числа возмущающих факторов, нарушающих нормальное функционирование системы. Например, возмущающими факторами в системе регулирования скорости вращения двигателя постоянного тока могут явиться изменения: нагрузки на валу двигателя, питающего напряжения, сопротивления щеточных контактов и т. п. Поэтому применяемые для целей управления разомкнутые системы могут удовлетворительно решать поставленные перед ними задачи лишь в двух случаях: когда возмущающие воздействия заранее заданы в виде определенной функции времени или статистически, либо в искусственных, редко встречающихся условиях, когда влияния возмущающих воздействий пренебрежимо малы или отсутствуют. Подобные условия на практике встречаются весьма редко, в связи с чем в области автоматического управления разомкнутые системы имеют ограниченное применение.

Реальные автоматические системы будут лучше выполнять свои функции в том случае, если они имеют возможность контролировать результаты своей работы и корректировать их. Например, если скорость вращения двигателя отклоняется в ту или иную сторону от заданного значения, то входное воздействие должно быть сформировано таким образом, чтобы компенсировать влияние возмущающих факторов. Для этого управляющее воздействие должно вырабатываться с учетом информации о результатах управления. С этой целью в автоматических системах применяется *обратная связь*.

Любой элемент системы, а также всякая группа ее элементов имеет *вход*, т. е. место приложения воздействия, и *выход*, т. е. место получения сигнала, посылаемого дальше. Обратная связь, охватывающая элемент, группу элементов или всю автоматическую систему в целом, означает, что выходной сигнал x_2 подается на вход (рис. 2.1), где складывается с входным сигналом x_1 или вычитается из него, что обозначается на схеме знаком \otimes . При сложении этих сигналов обратная связь называется *положительной*, а при вычитании — *отрицательной*. Система, имеющая контур обратной связи, называется замкнутой.

Функции обратной связи впервые установил Максвелл, рассматривая в 1868 г. машину для управления кораблем.

Понятие об обратной связи является одним из основных в кибернетике и означает управление на основе действительного, а не ожидаемого выполнения приказов [2.5]. Неисправность канала обратной связи приводит к тому, что автоматическая система оказывается дезорганизованной. Обратная связь противодействует тенденции к дезорганизации в работе автоматической системы, т. е., как заметил Винер, приводит к временной и местной перемене обычного направления изменения энтропии.

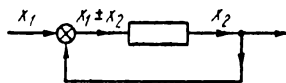


Рис. 2.1. Обратная связь

Пользуясь обратной связью, можно выяснить свойства и состояния объекта, оказывая на него пробные воздействия. Управляющие воздействия, служащие для этого, называют *изучающими*, в отличие от *направляющих*, предназначенных для управления объектом. При наличии тех и других управляющих воздействий имеет место *дуальное управление*.

Фактическое состояние системы, выявляемое с помощью обратной связи, зависит от ранее приложенного управляющего воздействия. Поэтому Винер определяет обратную связь как «свойство, позволяющее регулировать будущее поведение прошлым выполнением приказов». Она может быть, отмечает он, «столь проста, как обратная связь безусловного рефлекса, или она может быть связью более высокого порядка, когда прошлый опыт учитыва-

ется не только для регулирования специфических движений, но также формирования всей линии поведения. Подобная обратная связь может представлять — и часто действительно представляет — то, что, с одной стороны, известно как условный рефлекс, а с другой — как познание» [2.5]. Управление на основе информации, получаемой посредством обратной связи, является тем общим, что свойственно процессам управления в органической природе, в технике и в обществе. Все эти процессы протекают без нарушений функционирования управляемых объектов лишь при условии исправной обратной связи. Многочисленные обратные связи существуют и в человеческом организме.

Иногда в системе нельзя заметить никаких технических устройств, осуществляющих функции обратной связи. В действительности обратная связь существует и в этих системах, если они замыкаются через человека [2.11]. Такова, например, система водитель — автомобиль. Это имеет место и при управлении катером, курс и положение которого на волнах все время корректирует человек. Если бы человек этого не делал, то катер сбился бы с курса или даже опрокинулся. Подобным же образом обратная связь обеспечивает правильное действие всякой автоматической системы. За исключением указанных выше особых случаев, в которых пригодны разомкнутые системы, управляющая система всегда должна быть замкнутой: либо устройствами обратной связи, либо через человека, участвующего в процессе управления. В последнем случае систему называют *полуавтоматической*.

Действие замкнутой автоматической системы может быть пояснено на примере работы привода наведения зенитного орудия (рис. 2.2) [2.6]. Радиолокационная установка *РУ* определяет пространственное положение цели и параметры ее движения. На основе этой информации вычислительное устройство *ВУ* вырабатывает управляющий сигнал в виде угла наведения зенитного орудия с учетом всех факторов, влияющих на точность стрельбы (направление и скорость ветра, состояние атмосферы и т. п.). При этом необходимо иметь в виду, что за время полета снаряда после выстрела цель успеет переместиться в пространстве. Поэтому для обеспечения встречи снаряда с целью необходимо наводить орудие не на цель, а в некоторую точку *В*, в которой должна оказаться

цель ко времени встречи. Управляющий сигнал от вычислительного устройства подается на усилитель мощности $У$, выходной сигнал которого используется для управления мощным исполнительным двигателем $ИД$, осуществляющим перемещение орудия в пространстве до нужного положения. По каналу обратной связи $ОС$ на вычислительное устройство передается информация о текущем положении орудия, которое сопоставляется с требуемым положением, определяемым вычислительным устройством. Поскольку орудие должно быть установлено в определенное поло-

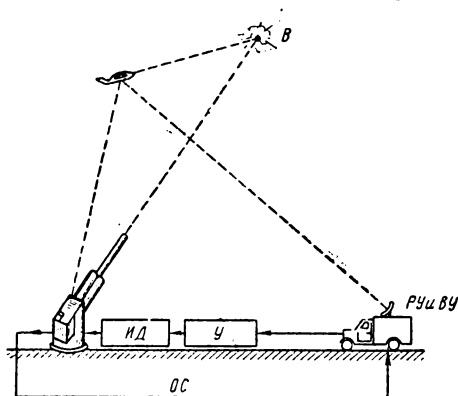


Рис. 2.2. Привод наведения зенитного орудия

жение в горизонтальной и вертикальной плоскости, применяют две автоматические системы.

В зависимости от назначения автоматической системы основой ее действия могут быть: программа, слежение, поиск.

В частном случае программа может требовать *стабилизации*, т. е. поддержания некоторого неизменного состояния в системе. В таком случае система называется *стабилизирующей*. Пример автоматической системы, действующей по программе, называемой *программной*, — поезд, управляемый автомашинистом, обеспечивающим заданное расписание движение по железной дороге. Пример системы, осуществляющей слежение и называемой *следающей*, — зенитное орудие, управляемое приводом, поворачивающим ствол орудия в соответствии с перемещением

цели. Пример системы, работающей методом поиска (называемой далее *поисковой*), — авиационный реактивный двигатель, управляемый автоматом, ищущим наиболее экономичный режим его работы. В отличие от систем *автоматического поиска* (САП), к которым относится последний пример, системы программные и следящие объединяются под названием систем *автоматического регулирования* (САР). Действия систем автоматического регулирования детерминированы *программой* или командами, тогда как у систем автоматического поиска они являются пробными, корректируемыми по их результатам.

Когда ствол зенитного орудия отклоняется от того положения, которое требуется для слежения за целью, информация об этом поступает по каналу обратной связи, перерабатывается в вычислительном устройстве и заставляет исполнительный двигатель поворотом ствола устранить отклонение. Так осуществляется *управление по отклонению*. Но отклонение возникает под действием какого-то возмущения, например выстрела, создающего момент, поворачивающий ствол. По специальному каналу может передаваться информация об этом возмущении, заставляющая исполнительный двигатель его компенсировать, развивая момент, противодействующий моменту от выстрела. Таким образом осуществляется *управление по возмущению*. При управлении по отклонению происходит исправление ошибок, возникающих в работе системы, при управлении по возмущению — их предотвращение. Применяется также соединение этих двух методов — *комбинированное управление*. Каждый из этих трех видов управления осуществим различными способами.

Передача всякого сигнала требует времени. Автоматическим системам, как и всем динамическим системам, свойственна инерционность. Поэтому нарушение, вызываемое в системе возмущающим воздействием, как и самое появление возмущения, не во всех случаях может быть предотвращено с помощью обратной связи. Если инерционность системы достаточно мала, а время передачи сигнала обратной связи несоизмеримо меньше времени нарастания возмущения, то система практически мгновенно реагирует на возмущение, компенсируя вызываемые им нарушения. Таковы *инвариантные системы*, в той или иной мере нечувствительные к возмущающим воздействиям [2.18]. Но при наличии заметной инерционности,

а также в тех случаях, когда скорости распространения сигналов и нарастания возмущений соизмеримы, обратная связь позволяет ликвидировать влияние возмущения лишь с некоторым опозданием. Такой метод управления может оказаться непригодным, приводя систему в состояние *самовозбуждения*, когда в ней возникают автоколебания, которые сами себя поддерживают и не допускают устранения последствий возмущения. Тогда для улучшения процесса управления система должна реагировать не на действующие, а на ожидаемые возмущения, которые она должна компенсировать с опережением или их предсказывать (прогнозировать).

Системы с предсказанием возмущений, или *прогнозирующие системы*, легко осуществить, если характер возмущений заранее известен. Это хорошо поясняет приводимый Винером пример управления автомобилем на обледенелой дороге. Если вести машину обычным способом, то она забуксует, прежде чем водитель успеет что-либо сделать. Поэтому надо действовать рулем рывками, не настолько сильными, чтобы вызвать большое скольжение, но достаточными, чтобы водитель ощущал, не грозит ли машине буксование. Это — пример компенсации возмущений. В случаях возмущений менее определенного характера необходимое предсказание может быть сделано лишь на основании изучения (статистической обработки) возмущений, возникавших в течение некоторого времени. Если при этом такому же изучению подвергаются меры компенсации возмущений и достигнутые результаты, после чего делаются выводы о целесообразном поведении системы в новых условиях, то система способна анализировать свой опыт и на его основе находить решения новых задач. Действующие так системы называют *обучающимися*. Если оценку решений, находимых системой на основе своего опыта, делает человек, то система *обучаемая*. Обучаемыми системами являются, в частности, перцептроны. Если ту же оценку делает автоматическое устройство, отключаемое по окончании обучения, то система *автоматически обучаемая*. И если такое устройство входит в состав системы, то она может быть названа *самообучаемой*.

Предсказание означает расширение функций управляющего автомата, позволяющее анализировать условия и обстановку работы системы и вырабатывать требования

к управлению на основе информации, получаемой и перерабатываемой достаточно долго. Учет опыта своей работы означает, что система пользуется не только текущей информацией, но и накопленной и обработанной информацией за более или менее длительное время. В этом случае обратная связь не только корректирует действие системы в каждый момент времени, но и вырабатывает линию поведения как продолжение предыдущей, это — *обратная связь линий поведения*.

Действие такой обратной связи можно пояснить на уже разобранном примере управления зенитным оружием. В этом случае управление должно обеспечить встречу снаряда и поражаемой им цели. Самолет маневрирует, и, чтобы успешно вести огонь, нужно прогнозировать его движение, учитывая его технические характеристики и наблюдаемое поведение. Повторение такого прогноза в ряде случаев показывает, когда он был правилен, а когда ошибочен, и позволяет учитывать этот опыт в очередном случае, каждый раз его пополняя. Система, работающая таким образом, управляет своим поведением. Запоминая опыт своей работы, она делает и долгосрочный прогноз. Учитывая такие условия, как погода, рельеф местности, тип и тактику цели, она вырабатывает закон преследования. Такого рода системы, предназначенные для противовоздушной обороны США во время второй мировой войны, создавал Винер, вводя в технику свойства и принципы действия живых организмов. Эта работа Винера стала важнейшим этапом развития кибернетики.

Выбирая линию поведения при вновь встретившихся условиях, автоматическая система может оказаться не приспособленной к наилучшему решению стоящей перед ней задачи. Наиболее совершенные системы в таком случае самонастраиваются, т. е. изменяют свои свойства таким образом, чтобы могло быть достигнуто наилучшее решение. Они вырабатывают оптимальную программу такого изменения или *стратегию управления* [2.17].

Самонастраивающиеся системы изменяют свое устройство в зависимости от изменения условий и течения процесса управления. Простейший вид таких систем — *экстремальные*, обеспечивающие экстремальное (максимальное или минимальное) значение какой-либо регулируемой величины. Задачей экстремального регулирования является, например, поддержание максимального коэффициента

полезного действия энергетической установки при изменении режима ее работы. Это может быть авиационный реактивный двигатель, к. п. д. которого поддерживается постоянным при изменении скорости полета и других влияющих факторов посредством изменения сечения сопла. Но может быть и так, что положение и уровень экстремума неизвестны, и система должна их находить. Изменяя положение регулирующего органа, система сравнивает последовательно получающиеся значения регулируемой величины. Если оказывается, что экстремум пройден, то происходит реверс регулирующего органа и т. д. Системы, ищущие экстремум, являющийся переменным по положению и величине, называют *оптимальными*. Такая система, например, может задавать самолету режим полета, обеспечивающий наибольшее значение расстояния, проходимого за счет расхода единицы горючего, в условиях обледенения, для которых режим полета нельзя задать заранее, так как изменение аэродинамических свойств самолета зависит от характера обледенения.

Заметим, что проходя экстремум и возвращаясь к нему, система его «запоминает». Здесь, следовательно, имеет место простейшее накопление информации.

Еще более совершенные системы — *самоорганизующиеся*. В этих системах по ходу процесса управления изменяется не только устройство, но и принцип действия. В них происходит, например, изменение знака обратной связи, ставящее систему в наиболее выгодные условия, причем важную роль получают положительные обратные связи [2.13]. Такие автоматические системы в наибольшей мере подобны живым организмам. С обучающимися и самонастраивающимися их называют самоприспосабливающимися системами или системами автоматической адаптации, т. е. автоматического приспособления.

Управление наиболее эффективно при наилучшем использовании энергии и всех других ресурсов. Такое управление называют *оптимальным* [2.12], а осуществляющие его системы — *автоматическими оптимизаторами*. К достижению оптимального управления следует стремиться при создании автоматических систем. Однако это не всегда возможно и не во всех существующих системах можно осуществить оптимальное управление. Поэтому задачей технической кибернетики должно быть изучение (проектирование, построение и исследование — синтез

и анализ) всевозможных автоматических управляющих систем, а не только обеспечивающих оптимальное управление.

Простейшими автоматическими системами являются системы автоматического регулирования. Одной из древнейших САР было устройство, которым снабжали ветряные мельницы, чтобы поворачивать их крылья против ветра.

Первой из известных в истории САР была следящая система, построенная в XVI в. Агостино Рамелли (1530—1590 гг.) — итальянцем, служившим во Франции в качестве военного инженера¹. Это была водяная мельница, в которой подача зерна на жернова автоматически регулировалась в зависимости от твердости зерна и количества воды, поступающей из реки, при этом на жернова поступало столько зерна, сколько они могли перемолоть (если зерна много, то жернова засоряются, если мало, то они разгоняются и нагреваются). Эта система уже имела, очевидно, обратную связь между жерновами и зерном. Первую электрическую следящую систему построил в 1867 г. А. П. Давыдов на батарее «Не тронь меня» в Ревеле. Наиболее совершенные АУС возникли в результате объединения достижений, полученных в развитии автоматических регуляторов от регулятора ветряной мельницы и вычислительных машин от механической машины Паскаля, построенной в 1642 г.

Рассмотренные принципы действия автоматических систем существуют также в организмах и социально-экономических системах управления. Действие органической системы легко наблюдать, на примере зрительной обратной связи, корректирующей наши движения. Это действие бывает то форсирующим, то демпфирующим, т. е. обратная связь меняет знак. Мы работаем с предсказанием новых условий, учитываем свой опыт, обучаясь на нем, изменяем характер своих действий в соответствии с изменением обстановки.

В социально-экономической системе обратная связь выражается в проверке выполнения, контроле качества и учете эффективности плановых работ и мероприятий. И здесь имеет место предсказание новой обстановки, учет опыта, обучение и изменение принципа действий.

¹ Учителем Рамелли был Мариньяно, один из лучших полководцев того времени, вышедший из школы Леонардо да Винчи.

Обратная связь, учет опыта и прогноз предстоящих, условий работы — вот то общее, что характеризует управляющие системы во всех областях, где они существуют, и что делает возможным их изучение с общих позиций кибернетики. Целью такого изучения должно быть управление всеми процессами, происходящими в мире, их подчинение задачам человеческого общества.

Примеры. 1. Человек, стоящий на платформе, которая движется с переменной скоростью, подвергается возмущающим воздействиям сил инерции. Чтобы удержаться на ногах, он должен управлять своим телом, сопротивляясь этим силам. Здесь имеет место управление, имеющее целью стабилизировать положение человека относительно платформы. Он может осуществлять это управление двояко: с помощью обратной связи или посредством компенсации. Обратную связь дает кинестетическое ощущение нарастающего воздействия, которое человек уравнивает противоположным усилием. При очень быстром нарастании воздействия человек такое усилие развить не успевает, т. е. обратная связь не обеспечивает управление. Компенсация — это усилие, которое человек развивает, держась, например, за поручни, когда он предвидит или ожидает, что появится возмущение.

2. Рыси в основном питаются зайцами. Количество рысей увеличивается с ростом количества зайцев, но это приводит к уменьшению последних, когда численность рысей достигает некоторого предела. Излишек рысей является возмущением, которое устраняется уменьшением числа зайцев. Таким образом, между численностью рысей и зайцев имеется отрицательная обратная связь. Но количество зайцев уменьшается не сразу — обратная связь действует недостаточно быстро. В результате в системе зайцы — рыси возникают автоколебания в численности [2.8]. Это *система «хищник — жертва»*.

3. Значительное запаздывание действия обратной связи имеет место в химической промышленности, когда обратная связь осуществляется движением жидких или газообразных продуктов по трубам. В этих случаях лучшие результаты дает управление по методу компенсации [2.9].

§ 2.3. ФУНКЦИОНАЛЬНЫЕ СХЕМЫ И БЛОК-СХЕМЫ

Всякая автоматическая система имеет две основные части, показанные на рис. 2.3: *объект управления* *O* и автоматическую управляющую систему АУС. Эти части соединяются *каналами связи* 1 и 2. По каналу 1 идет информация от АУС, воздействующая на объект; по каналу 2 поступает информация о состоянии объекта, с учетом которого вырабатываются управляющие сигналы, несущие информацию по каналу 1. АУС может еще получать информацию извне: по каналу 3 — задания, которые

она должна исполнять, управляя объектом; по каналу 4 — сведения о внешних условиях, в которых находится объект, получающий воздействия этих условий по каналу 5.

На рис. 2.3 представлена простейшая *функциональная схема* автоматической системы. Так называют схемы, на

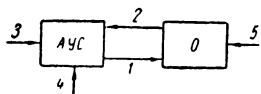


Рис. 2.3. Простейшая функциональная схема автоматической системы

которых показаны части системы, имеющие те или иные функции, необходимые для ее работы, и каналы связи, соединяющие эти части, а также направление передачи информации по этим каналам, указываемые стрелками, как на рис. 2.3. Несмотря на разницу в назначении и конструкции раз-

личных автоматических систем, в них всегда можно выделить части, имеющие одинаковые функции, т. е. исполняющие одинаковые роли, во время действия системы. На этом и основано составление функциональных схем. Их элементы могут иметь разнообразное устройство.

Развивая простейшую функциональную схему, изображенную на рис. 2.3, представим ее в виде, показанном на рис. 2.4. Здесь *ОО* — окружающая обстановка, влияющая на объект *О* по каналу 5, а по каналу 6 — на датчики обстановки *ДО*, от которых информация поступает в канал 4. В канал 2 информация поступает от *средств наблюдения СН*, получающих сведения о состоянии объекта по каналу 7. Канал 3 имеет здесь то же назначение, что и на рис. 2.3. На рис. 2.5 показана отдельно АУС. Она состоит из следующих главных частей: *счетно-решающего устройства СР*, *управляющего автомата УА* и *управляемого устройства УУ*, которым часто является некоторый *исполнительный механизм*. *СР* получает информацию по каналам 2, 3 и 4, а также по каналу 11, по которому поступают сведения о фактическом состоянии управляемого устройства. Пользуясь этой информацией, *СР* вырабатывает *командную информацию*, поступающую в *УА* по каналу 8. Автомат получает также информацию о состоя-

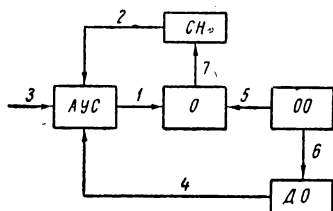


Рис. 2.4. Функциональная схема автоматической системы

нии управляемого устройства по каналу 10. На основе информации, получаемой по каналам 8 и 10, автомат выдает управляющую информацию, которая поступает в управляемое устройство по каналу 9.

В качестве счетно-решающего устройства может быть применена *управляющая машина* (например, электронная или пневматическая), перерабатывающая информацию посредством вычислительных и логических операций. Такая машина может управлять не только по данным текущей информации, поступающей по каналам 2 и 4, но и на основании опыта работы системы, т. е. по данным информации, накопленной в течение более или менее длительного времени. Пользуясь накопленной информацией, управляющая машина может предсказывать условия дальнейшей работы системы и управляющие воздействия, которые будут требоваться в этих условиях. В таком случае система является самообучаемой (см. § 2.2).

Если управляющая машина является цифровой, то информация на ее вход поступает в дискретные

моменты времени и управляющая информация появляется на выходе также в дискретные моменты времени. В этом случае автоматическая система является *дискретной*.

Если действия, требуемые от системы, сравнительно просты и немногочисленны, то самообучение достигается и при помощи более простых средств, чем управляющие машины. Более простые электрические и даже механические счетно-решающие приборы применяются также в случаях, когда не требуется самообучения системы или не столь высоки требования к ее быстродействию. Во всех случаях, однако, счетно-решающее устройство — будь это управляющая машина или простейшая конструкция — действует по программе, заданию или по вложенным в нее критериям. Программу и критерии задает человек, построивший или пустивший в ход систему; это и есть информация, поступающая по каналу 3.

В составе АУС могут быть части, изменяющие свойства и устройство управляющего автомата во время работы, в зависимости от информации, поступающей по ка-

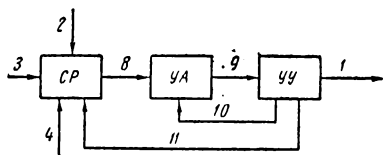


Рис. 2.5. Функциональная схема автоматической управляющей системы

стрелками, указывающими направление передачи информации по ним. УМ, получая информацию по каналам 2, 3, 4 и 11, вырабатывает команду, поступающую в К, где она сопоставляется с информацией, поступающей по каналу 2, и после корректировки передается в канал 8. По каналу 8 информация идет в УА и УС, куда подается также информация из канала 11. Сопоставляя команду с действительным состоя-

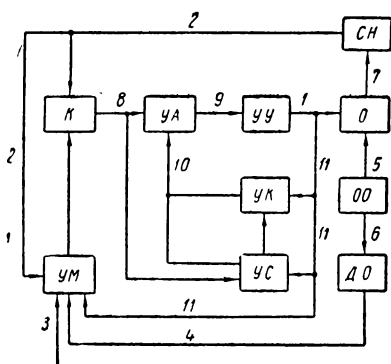


Рис. 2.6. Функциональная схема самонастраивающейся автоматической системы

нием $УУ$, $УС$ вырабатывает меры самонастройки, информацию о которых передает в $УК$, куда поступает также информация из канала 11 . $УК$ воздействует на $УА$ по каналу 10 и связано с $УС$ дополнительным контуром обратной связи.

В табл. 2.1 приведены примеры реализации функциональных схем, представленных на рис. 2.4 и 2.5, в задачах управления уличным движением и курсом корабля. Действие АУС в этих задачах должно обеспечивать возможно большую пропускную способность магистралей, на пересечении которых установлен управляемый светофор, и движение корабля по заданному курсу. Для этого в частности, датчики подъезжающего транспорта (например, фотоэлектрические или электромагнитные) подают информацию, по которой счетно-решающее устройство подсчитывает количество автомобилей, движущихся по всем

Таблица 2.1

Функциональный анализ применения двух систем автоматического управления

Решаемая задача Элемент функциональной схемы	Управление уличным движением	Управление курсом корабля
О — объект	Движущийся транспорт	Корабль
СН — средства наблюдения	Датчики подъезжающего транспорта	Гирокомпас
ОО — окружающая обстановка	Выделенный транспорт (машины скорой помощи, пожарные и т. п.)	Навигационные и метеорологические условия
ДО — датчики обстановки	Датчики подъезжающего транспорта	Приборы для навигационного и метеорологического наблюдения
УУ — управляющее устройство	Светофор	Рули поворота
УА — управляющий автомат	Автоматический переключатель сигнальных огней	Автоматический рулевой привод
СР — счетно-решающее устройство	Вычислительная машина или логическое устройство	Вычислительная машина или счетно-решающий прибор

направлениям, и по заложенным в него критериям решает, какому направлению дать зеленый свет. На перекрестках с небольшим движением эту задачу может решать сравнительно простой электронный прибор. Выделенные машины, которые нужно пропускать внеочередно, должны быть снабжены устройством для особого воздействия на те же датчики. В более сложных условиях система автоматического управления уличным движением может быть самообучаемой.

В простейших задачах управления ряд элементов рассмотренной выше общей функциональной схемы может отсутствовать. Могут отсутствовать датчики обстановки, если достаточно учитывать ее влияние на объект по изменению состояния последнего, т. е. с помощью информации от средств наблюдения. В этом случае отсутствуют каналы связи 4 и 6. Если ситуация, возникающие в процессе управления, немногочисленны, то излишне и счетно-решающее устройство. Тогда отпадают каналы связи 8 и 11, а информация из канала 2 поступает непосредственно в управляющий автомат. Получающаяся при этом функциональная схема изображена на рис. 2.7. В ней не может быть самообучения; самонастройка тоже отсутствует, может отсутствовать и канал связи 10. Канал 3 теперь подает информацию к средствам наблюдения, к которым переходят простейшие функции счетно-решающего устройства: сопоставление информации, поступающей по каналу 7, с заданными критериями (канал 3) и выработка командной информации, передаваемой по каналу 3, заменяющему, таким образом, канал 8.

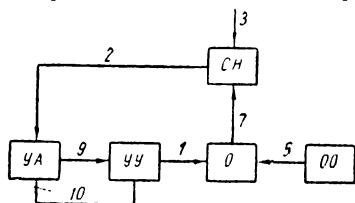


Рис. 2.7. Функциональная схема простейшей автоматической системы

По схеме, изображенной на рис. 2.7, осуществляются простейшие процессы *автоматического регулирования*. В этих случаях средством наблюдения служит некоторый *чувствительный элемент ЧЭ*, а управляемым устройством — *регулирующий орган РО*, как показано на рис. 2.8. Управляющий автомат УА при этом может состоять из *усилителя У* и *исполнительного органа ИО*, связанных каналом 12. Если перемещение исполнительного органа

не требует больших усилий или усилие, достаточное для этого перемещения, развивает чувствительный элемент, то усилитель не требуется и автоматическую регулируемую систему называют *регулятором прямого действия*. В противном случае, в *регуляторе непрямого действия*, необходимо, чтобы чувствительный элемент приводил в действие усилитель, управляющий энергией постороннего источника, за счет которой работает исполнительный орган. Эти схемы устройства УА относятся и к ранее рассмотренным системам. Чаще встречаются автоматы с усилителями. Вместо окружающей обстановки в анализе работы регуляторов рассматривают нагрузку H объекта, обусловленную условиями его эксплуатации.

Простейшие системы автоматического управления действуют, получая лишь текущую информацию, без ее запоминания для дальнейшей работы. Информация, получаемая в них по каналу обратной связи, относится к моменту време-

ни, отличающемуся от момента выработки управляющего сигнала лишь на малое время, занимаемое ее передачей и переработкой. Следовательно, это текущая информация, лишь поступающая с небольшим опозданием. Командная информация вырабатывается в таких системах несколько раньше ее получения управляющим устройством или регулирующим органом, но на основании данных о состоянии объекта, полученных с помощью обратной связи, т. е. это тоже текущая информация. В отличие от простейших автоматических устройств, действующих лишь на основе текущей информации, более сложные системы, работающие с помощью переработки запоминаемой информации, называют *кибернетическими*.

Функциональные схемы, представленные на рис. 2.3 — 2.8, замкнуты. Замкнутость схемы определяется тем, что в ней существует циркуляция информации по замкнутому контуру, образуемому каналами связи, при обходе которого направление передачи информации не меняется. Контур 1—2 на рис. 2.3, 1—7—2 на рис. 2.4, 9—10 и 8—9—11 на рис. 2.5, 9—1—7—2 на рис. 2.7, 12—9—1—7—2

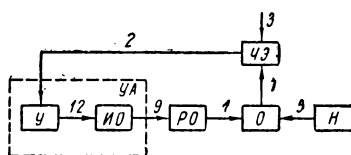


Рис. 2.8. Функциональная схема системы автоматического регулирования

на рис. 2.8. Это означает, что в рассмотренных схемах имеет место не только передача информации от управляющих элементов (по каналам 1, 9 и 12), но и обратная передача по каналам обратной связи 2, 7, 10 и 11. Обратная связь может охватывать систему в целом (канал 2) или некоторые части системы (канал 10 и 11). Заметим, что контур 1—5—6—4 на рис. 2.4 по указанным признакам не является замыкающим систему.

Функциональная схема может быть совершенно одинаковой в разных случаях, относящихся к работе весьма различных систем автоматического управления. Один и

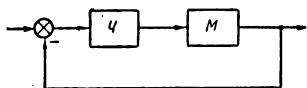


Рис. 2.9. Блок-схема полуавтоматической системы

тот же элемент функциональной схемы получает в каждом из таких случаев конструктивную реализацию, которая может быть совсем непохожей на реализации того же элемента в других случаях. Например, чувствительным элементом в одном из случаев, приведенных в табл.

2.1, является гирокомпас, а в другом — датчик подъезжающего транспорта. Элементами некоторых автоматических систем могут быть растения и животные.

Если в функциональной схеме какой-либо автоматической системы заменить элементы конструктивными блоками, то получится *блок-схема*. При этом некоторые элементы функциональной схемы могут быть представлены одним блоком блок-схемы. В других случаях несколько блоков блок-схемы могут представлять какой-либо элемент функциональной схемы. В литературе встречаются также смешанные схемы, содержащие элементы функциональных схем и блок-схем.

На блок-схемах полуавтоматических систем одним из блоков является человек. На рис. 2.9 показана блок-схема системы, замкнутой через человека, управляющего машиной, которая может быть катером, автомобилем, самолетом и т. п.

Примеры. 1. На рис. 2.10 схематически представлена система автоматического регулирования уровня воды в котле паровой машины, построенная И. И. Ползуновым в 1764 г. [2.20]. Питающая вода поступает в котел по водопроводной трубе *в*, а пар идет из котла в машину. При увеличении расхода пара уровень воды в котле понижается, плавающий на поверхности воды поплавки *п*

опускается и открывает заслонку 3, с которой он соединен тягой, вследствие чего увеличивается поступление воды в котел и уровень воды повышается. Регулятор Ползунова поддерживает требуемый уровень воды. В нем нет усилителя, т. е. он является регулятором прямого действия.

Представляя систему Ползунова функциональной схемой, изображенной на рис. 2.8, имеем: *О* — паровой котел, *Н* — расход пара, *ЧЭ* — поплавков, *ИО* — заслонка, *РО* — водопроводная труба. Критерий, задаваемый по каналу 3, фиксируется длиной тяги.

2. На рис. 2.11 представлена принципиальная схема термостата с простейшим регулятором температуры. В термостат *Т* вставлен ртутный контактный термометр *КТ*. Контакты термометра должны быть установлены на делении его шкалы, совпадающем с требуемой температурой. При действии электрического нагревателя *Н*, находящегося в термостате, температура повышается, и высота столбика ртути в термометре увеличивается. Когда температура достигает требуемого значения, ртуть замыкает контакты термометра. При этом замыкается цепь постоянного тока, питающего катушку контактора *К*, а контактор включает переменный ток, питающий нагреватель. Когда после этого температура начинает понижаться, контакты термометра будут разомкнуты и контактор вновь включит питание нагревателя. Здесь контактор служит усилителем, так как контактный термометр — слишком маломощный элемент, чтобы включать и выключать питание нагревателя. Следовательно, рассматриваемое устройство является регулятором непрямого действия. Пользуясь и в этом случае функциональной схемой 2.8, имеем: *О* — термостат, *Н* — температура окружающей среды, *ЧЭ* — термометр, *У* и *ИО* — контактор, *РО* — электрический нагреватель. Критерий, задаваемый по каналу 3, фиксируется контактами термометра.

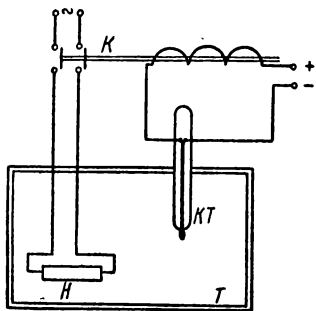


Рис. 2.11. Принципиальная схема термостата с простейшим регулятором температуры

что и на рис. 2.8, элементы *У* и *ИО* функциональной схемы объединены в блок *К* (контактор), в связи с чем отсутствует канал 12; элемент *РО* заменен блоком *ЭН* (электрический нагреватель), элемент *О* — блоком *Т* (термостат), *ЧЭ* — блоком *КТ* (контактный термометр), элемент *Н* заменен фактором *ТОС* (температура окружающей среды).

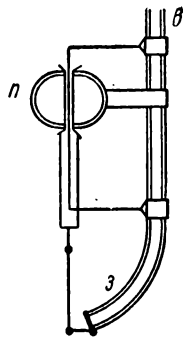


Рис. 2.10. Автоматическая система Ползунова

Блок-схема этой системы представлена на рис. 2.12. Здесь, как показывают каналы связи, обозначенные теми же цифрами,

На рис. 2.13,а представлена общая схема системы наведения ракеты-перехватчика, предназначенной для поражения воздушной цели. Следящий за целью радиолокатор *СЦЛ* предназначен для своевременного обнаружения цели и последующего ее сопровождения и позволяет получать информацию о координатах цели в пространстве. Вычислительная машина *ВМ* определяет по этим данным

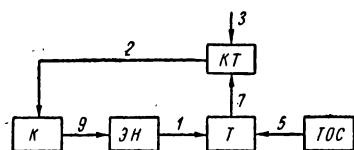


Рис. 2.12. Блок-схема простейшей системы автоматического регулирования температуры

будущую траекторию полета цели. Аналогичную информацию о координатах ракеты-перехватчика вырабатывает радиолокатор, следящий за ракетой *СРА*. Сравнивая траектории полета ракеты и цели, вычислительная машина определяет необходимые изменения в траектории ракеты с тем, чтобы обеспечить поражение цели. Соответствующие управляющие сигналы с помощью радиопередающего устройства *РП* направляются в ракету и используются для корректировки траектории ракеты, имеющей для этого специальную систему автоматического управления. Блок-схема такой системы изображена на рис. 2.13,б. Разница между действительными и требуемыми координатами ракеты усиливается усилителем *У*, а затем превращается в механическое перемещение рулевых устройств, которые перемещаются с помощью исполнительного двигателя *ИД*.

§ 2.4. АЛГОРИТМЫ

При подаче информации на вход любого элемента автоматической системы на выходе его появляется информация, воздействующая на какой-либо другой элемент, как это показывает функциональная схема системы. В соответствии с этим действие автоматической системы можно представить в виде некоторой последовательности *актов переработки информации*, происходящих в ее элементах. Процесс управления можно рассматривать, соответственно, как последовательность некоторых элементарных актов переработки информации. Эта последовательность может изменяться в зависимости от выполнения или невыполнения некоторых *логических условий*. Совокупность

предписаний, определяющих требуемую последовательность актов переработки информации в процессе управления, есть *алгоритм* этого процесса. Слово алгоритм взято из математики, где оно обозначает предписание о порядке

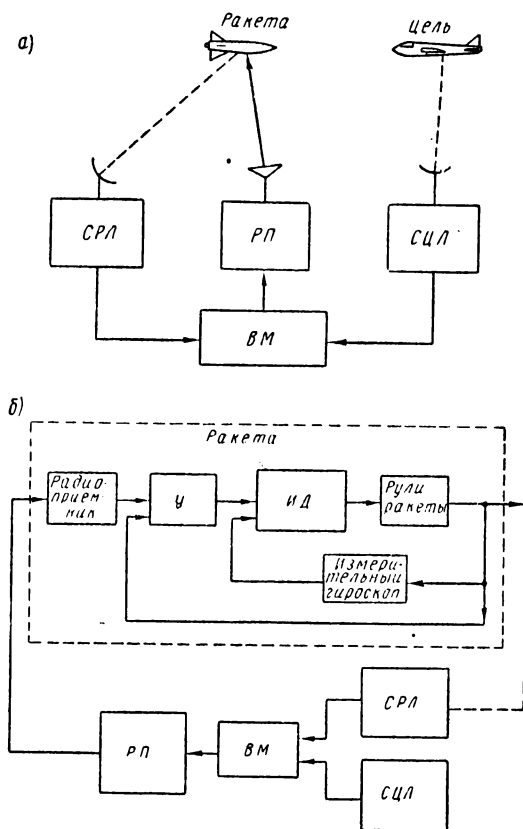


Рис. 2.13. Блок-схема системы наведения ракеты-перехватчика на воздушную цель

выполнения операций, необходимых для решения задач некоторого типа. Оно происходит от имени математика Аль-Хорезми (IX в.), составившего правила четырех арифметических действий в десятичной системе счисления. Алгоритм автоматической системы описывает ее работу,

показывая, как осуществляются происходящие в ней процессы. Он служит, таким образом, для исследования (или анализа) такой системы. Но может быть задан процесс управления, состоящий из некоторой последовательности актов переработки информации. Тогда алгоритм этого процесса является заданием для построения (или синтеза) реализующей такой процесс автоматической системы. В обоих случаях составление алгоритма означает *алгоритмизацию* рассматриваемой автоматической системы в первом случае и заданного процесса управления во втором.

Обозначим акт переработки информации, имеющий порядковый номер m , через A_m , а k -е по порядку логическое условие — l_k . Выражение, показывающее последовательность актов в зависимости от определяющих ее логических условий, называют *логической схемой алгоритма*, а акты и условия — членами этой схемы.

Рассмотрим два наиболее употребительных способа записи логических схем алгоритмов.

При записи схемы по первому способу после каждого логического условия ставится стрелка \uparrow с некоторым номером. Эта стрелка указывает, что в случае невыполнения условия, после которого она поставлена, происходит следующий по схеме акт, а другой, перед которым ставится стрелка \downarrow с тем же номером [2.4]. Когда логическое условие выполнено, пишут $l=1$, если же не выполнено, то $l=0$. Условие, которое никогда не выполняется, называют *тождественно ложным* и обозначают через ω . Поставленная после него стрелка \uparrow позволяет неограниченно повторять предшествующую часть алгоритма, начинающуюся после стрелки \downarrow с тем же номером. Если в некоторых случаях процесс прекращается после какого-либо логического условия, то это отмечается *условием остановки*, обозначаемым далее через s . Если после некоторого логического условия следует тот или иной из дальнейших актов, в зависимости от выполнения или невыполнения этого условия, то схему алгоритма удобно писать в две строки. Так, в схеме

$$\begin{array}{ccccccc} \overset{3}{\downarrow} A_1 & \overset{2}{\downarrow} A_2 & \overset{1}{\uparrow} A_3 & & & & \\ & & \downarrow A_4 & & \overset{2}{\uparrow} A_5 & A_6 & \overset{3}{\uparrow} s \\ & & & & l_1 & & \end{array} \quad (2.1)$$

после A_2 наступает A_3 , если $l_1=1$ и A_4 , если $l_1=0$, после чего процесс идет одинаково в обоих случаях.

Во втором способе записи логических схем алгоритмов стрелка заменяется *знаками перехода* [2.6]. Если в случае выполнения логического условия следует акт, записанный вслед за ним, то никакого знака не ставят, как не ставят стрелки и в первом способе. Если же в этом случае следует акт, записанный дальше, то после логического условия ставится знак \lceil с некоторым номером сверху, а перед актом, который должен наступить, — знак \rfloor (*называемый правым*) с таким же номером снизу. Такой же знак с некоторым номером ставится перед актом, наступающим при невыполнении логического условия, после которого в этом случае должен стоять знак \lfloor (*левый*) с тем же номером. Если после какого-либо логического условия необходимы оба знака, т. е. \lceil и \lfloor , то они объединяются в знак $\lceil \rfloor$. Знак \lfloor с некоторым номером ставится также после акта, за которым следует не акт, написанный рядом с ним справа, а стоящие в другом месте схемы акт или условие, предшествующие в таком случае знаком \rfloor с тем же номером. На последовательность соседних актов и условий не влияют стоящие между ними знаки перехода.

Пользуясь указанными правилами применения знаков перехода, легко представить схему алгоритма (2.1) в следующем виде:

$$\underset{4}{\rfloor} A_1 \underset{3}{\rfloor} A_2 \underset{1}{\rfloor} A_3 \underset{2}{\rfloor} \underset{1}{\lceil} A_4 \underset{2}{\rfloor} \underset{3}{\rfloor} A_5 A_6 \underset{4}{\rfloor} s. \quad (2.2)$$

Для составления логической схемы алгоритма необходимо рассмотреть процессы, происходящие в системе, и выявить, из каких элементарных актов переработки информации складываются эти процессы и какими логическими условиями определяется последовательность таких актов. Для примера составим логическую схему алгоритма работы регулятора температуры, рассмотренного в примере 2 предыдущего параграфа.

Если регулятор достаточно долго не был включен, то в термостате устанавливается температура, равная температуре окружающей среды. После включения ход процесса зависит от переменного соотношения двух тепловых потоков: потока Q_1 , получаемого термостатом от нагревателя, и потока Q_2 , теряемого термостатом в окружающую среду.

Пусть заданная температура выше температуры окружающей среды. Тогда при включении регулятора контакты термостата разомкнуты, и контактор замыкает цепь переменного тока. Включенный нагреватель повышает температуру до заданного значения. При этом столбик ртути перекрывает контакты термостата, и контактор разрывает цепь нагревателя. Последний начинает охлаждаться, но в течение некоторого времени остается $Q_1 > Q_2$, вследствие чего температура в термостате продолжает повышаться. Затем она начинает падать при выключенном

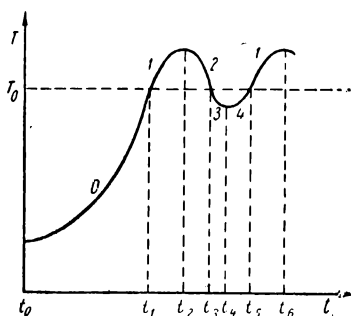


Рис. 2.14. Автоколебания в системе автоматического регулирования температуры

нагревателе, пока не достигнет заданного значения. При этом размыкаются контакты термостата, вновь включается нагреватель, но пока он разогреется $Q_1 < Q_2$, температура продолжает падать. Затем она начинает повышаться и повторяется предыдущий процесс.

Процесс регулирования состоит из участков 0, 1, 2, 3 и 4, показанных на рис. 2.14 и поясняемых табл. 2.2. Как видно, температура в термостате колеблется около

заданного значения с периодом, равным $t_3 - t_1$. Причина этих колебаний — тепловая инерция нагревателя, т. е. свойство системы, не зависящее от внешних воздействий. Это *автоколебания*, называемые так в отличие от колебаний, вызываемых внешними причинами, например изменением температуры окружающей среды.

Выясненная картина регулирования температуры показывает, что процесс состоит здесь из следующих элементарных актов переработки информации:

A_1 — температура \rightarrow высота h столбика ртути,

A_2 — $h \rightarrow$ выключение нагревателя,

A_3 — $h \rightarrow$ включение нагревателя.

Как видно из рис. 2.14 и табл. 2.2, акт A_2 может наступить после акта A_1 только при том условии, что температура T в термостате, повышаясь, достигла заданного значения T_0 . После этого должен опять происходить акт A_1 ; акт A_3 может наступить лишь тогда, когда темпе-

Таблица 2.2

Работа простейшего регулятора температуры

Время	Участок процесса	Состояние нагревателя	Тепловой баланс
t_0	—	Включается	Слева $Q_1 = Q_2$ Справа $Q_1 > Q_2$
$t_0 - t_1$	0	Включен	$Q_1 > Q_2$
t_1	—	Выключается	$Q_1 > Q_2$
$t_1 - t_2$	1	Выключен	$Q_1 > Q_2$
t_2	—	То же	$Q_1 = Q_2$
$t_2 - t_3$	2	«	$Q_1 < Q_2$
t_3	—	Включается	$Q_1 < Q_2$
$t_3 - t_4$	3	Включен	$Q_1 < Q_2$
t_4	—	То же	$Q_1 = Q_2$
$t_4 - t_5$	4	«	$Q_1 > Q_2$
t_5	—	Выключается	$Q_1 > Q_2$
$t_5 - t_6$	1	Выключен	$Q_1 > Q_2$
		и т. д.	

ратура достигнет значения T_0 в процессе понижения. Следовательно, последовательность элементарных актов переработки информации должна подчиняться двум логическим условиям, которые записаны в табл. 2.3: на участках 0, 3 и 4 $l_1=0$, $l_2=1$; на участках 1 и 2 $l_1=1$, $l_2=0$.

Таблица 2.3

Логические условия процесса регулирования температуры

l	0	1
l_1	$T < T_0$	$T \geq T_0$
l_2	$T \geq T_0$	$T < T_0$

Отсюда следует, что логическая схема алгоритма имеет вид

$$\begin{array}{c} \downarrow^{1,3} A_1 l_1 \uparrow^1 A_2 \downarrow^2 A_1 l_2 \uparrow^2 A_3 \omega \uparrow^3 \end{array}$$

или

$$\frac{1}{1} \mid \frac{1}{3} \mid A_1 l_1 \mid \frac{1}{1} \mid A_2 \mid \frac{2}{2} \mid A_1 l_2 \mid \frac{2}{2} \mid A_3 \omega \mid \frac{3}{3}.$$

Тождественно-ложное условие ω , поставленное в конце, означает, что процесс, записанный левее, при включенном регуляторе повторяется бесконечно.

С помощью логических схем алгоритмов можно описать и исследовать переработку информации и в значительно более сложных системах. Составление и анализ этих схем — основа применения вычислительных машин в АУС, а также моделирования процессов управления. Не всегда возможно построить *универсальный алгоритм*, описывающий все случаи действия сложной системы, но всегда можно пользоваться *частными алгоритмами*, определяющими часть таких случаев, практически важных для решения той или иной задачи [2.15].

Логические схемы алгоритмов можно подвергать *равносильным преобразованиям*, в результате которых получаются схемы разного вида, описывающие один и тот же процесс. Такие преобразования целесообразны, когда приводят к упрощению логических схем. Например [2.16], схему алгоритма

$$A_1 \downarrow_1 \downarrow_2 A_2 \downarrow_3 \downarrow_1 I_1 \uparrow_4 \downarrow_3 A_3 \downarrow_2 \downarrow_4 A_4 \downarrow_5 A_5 s \quad (2.3)$$

можно заменить двумя более простыми схемами

$$A_1 I_1 \uparrow_4 \downarrow_4 A_4 \downarrow_5 A_5 s \quad (2.4)$$

и

$$\downarrow_2 A_2 A_3 \downarrow_2. \quad (2.5)$$

Последняя схема означает бесконечное повторение поочередно следующих друг за другом актов A_2 и A_3 . Записанная с помощью первого способа (стрелок) схема алгоритма (2.5) выглядела бы так:

$$\downarrow A_2 A_3 \omega \uparrow.$$

Отсюда видно, что знаки перехода избавляют от необходимости вводить в схему тождественно-ложное условие.

Вводя правило записи, по которому

$$I \downarrow_i \downarrow_i = I \uparrow_i, \quad (2.6)$$

схему алгоритма (2.4) можно представить в более простом виде

$$A_1 I_1 \overset{5}{\vdash} A_4 \underset{5}{\vdash} A_5 s. \quad (2.7)$$

Схемы алгоритма (2.7) и (2.5) являются равносильными схеме (2.3). При составлении равносильных схем употребительны подобные (2.6) выражения равносильностей, как, например,

$$I \overset{k}{\vdash} \underset{k}{\vdash} = I \underset{i}{\vdash}, \quad (2.8)$$

$$I \overset{i}{\vdash} \underset{i}{\vdash} = \underset{i}{\vdash}. \quad (2.9)$$

Автоматические управляющие системы можно классифицировать по признакам, характеризующим алгоритмы их работы. В основном здесь различаются *детерминированные алгоритмы* САР и *недетерминированные алгоритмы* САП. Самоорганизующиеся системы работают по *самоорганизующемуся алгоритму*, изменяющемуся по ходу процесса управления. Это характерно и для ОУС.

Примеры. 1. Последовательность элементарных актов переработки информации по алгоритму (2. 1) имеет вид:

при $I_1 = 0, I_2 = 1, I_3 = 1$

$$A_1 A_2 A_4 A_5 A_6$$

— процесс окончен;

при $I_1 = 1, I_2 = 1, I_3 = 0$

$$A_1 A_2 A_3 A_5 A_6 A_1. \dots$$

— процесс продолжается бесконечно.

2. По алгоритму (2. 7) последовательность актов такова:

при $I_1 = 0$

$$A_1 A_4 A_5$$

— процесс окончен;

при $I_1 = 1$

$$A_1 A_5$$

— процесс окончен.

§ 2.5. БУЛЕВА АЛГЕБРА

В результате применения математической символики к формальной логике в прошлом веке возникла новая наука — *математическая логика*. Важнейшим для кибернетики разделом этой науки является *булева алгебра*.

называемая так по имени выдающегося английского математика Джорджа Буля. В книге «Исследование законов мышления», изданной в 1854 г., Буль писал: «Назначение настоящего трактата — исследовать основные законы тех операций ума, посредством которых производится рассуждение; выразить их на символическом языке некоторого исчисления и на этой основе установить науку логики и построить ее метод; сделать этот метод основой общего применения математической доктрины вероятностей; и, наконец, собрать из различных элементов истины, выявленных в ходе этих изысканий, некоторые правдоподобные указания относительно природы и строения человеческого ума» [2.1]¹.

Булева алгебра оперирует с *высказываниями*. Алгебраическому правилу математики «обратная величина обратной величины есть эта (последняя) величина» аналогично правило булевой алгебры: «отрицание отрицания высказывания есть само это высказывание». Для автоматических систем под высказываниями следует понимать состояния, в которых могут находиться системы или их части. Высказывание может быть либо *истинным*, либо *ложным*. Например:

«Катушка реле P возбуждена» — x_1 .

«Нормально открытый контакт реле P разомкнут» — x_2 .

«Нормально закрытый контакт реле P разомкнут» — x_3 .

Два высказывания x и y будем называть *равными* и писать $x=y$, если они одновременно истинные или одновременно ложные. Так, $x_1=x_3$, что следует из определения нормально закрытого контакта.

Высказывание x называется *отрицанием* высказывания y (это записывается $x=\bar{y}$), если x — истинно, когда y — ложно, и x — ложно, когда y — истинно. В приведенных примерах $x_2=\bar{x}_1$, что есть, по существу, определение нормально открытого контакта.

Таким образом, высказывание есть переменная, принимающая два значения: истина или ложь (сокращенно И или Л). В приведенных примерах значение высказываний x_2 и x_3 определялось значениями высказывания x_1 . Можно считать x_1 независимой переменной, а x_2 и x_3 зависимыми переменными.

¹ Цитировано по [2.7].

Из двух высказываний x и y можно образовать сложное высказывание z , значение которого (истинность или ложность) будет определяться значениями высказываний x и y .

Высказывание z называется *дизъюнкцией* высказываний x и y , если оно ложно тогда и только тогда, когда одновременно ложны высказывания x и y . Символически логическая зависимость z от x и y записывается так: $z = x \vee y$. Читается: « x или y ». Например, на схеме, представленной на рис. 2.15, может быть:

«контакт реле P_1 замкнут» — x_1 ;

«контакт реле P_2 замкнут» — x_2 ;

«контакт реле P_1 замкнут или контакт реле P_2 замкнут» — x_3 ;

«катушка контактора K возбуждена» — x_4 ;

«контакт K замкнут» — x_5 .
Здесь

$$x_5 = x_4 = x_3 = x_1 \vee x_2.$$

Дизъюнкция характеризуется табл. 2.4.

Высказывание z называется *конъюнкцией* высказываний x и y , если оно истинно тогда и только тогда, когда одновременно истинны высказывания x и y . Записывается это так: $z = x \cdot y$ и читается « x и y ». Например (рис. 2.16),

$$x_5 = x_4 = x_6 = x_1 \cdot x_2,$$

где x_6 обозначает высказывание:

«контакты P_1 и P_2 замкнуты»; x_1 , x_2 , x_4 и x_5 то же, что и в предыдущем примере.

Конъюнкция характеризуется табл. 2.5.

Высказывание z называется *импликацией* высказываний x и y , если оно ложно тогда и только тогда, когда x — истинно, а y — ложно. Записывается это так: $z = x \Rightarrow y$ и читается: «если x , то y ». Например (рис. 2.17):

«катушка реле P_1 возбуждена» — x_1 ;

«катушка реле P_2 возбуждена» — x_2 ;

«если катушка реле P_1 возбуждена, то и катушка реле P_2 возбуждена» — x_3 ;

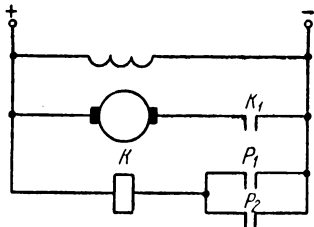


Рис. 2.15. Схема управления маломощным двигателем постоянного тока (к примеру дизъюнкции двух высказываний)

«катушка контактора K возбуждена» — x_4 .
Здесь

$$x_4 = x_3 = x_1 \Rightarrow x_2.$$

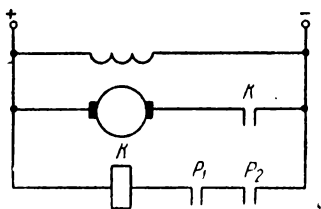


Рис. 2.16. Схема управления маломощным двигателем постоянного тока (к примеру конъюнкции двух высказываний)

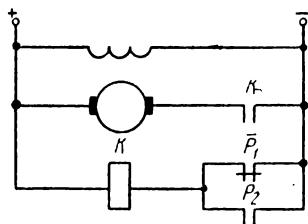


Рис. 2.17. Схема управления маломощным двигателем постоянного тока (к примеру импликации двух высказываний)

Таблица 2.4
Дизъюнкция

x	y	$x \vee y$
Л	Л	Л
И	Л	И
Л	И	И
И	И	И

Таблица 2.5
Конъюнкция

x	y	$x \cdot y$
Л	Л	Л
И	Л	Л
Л	И	Л
И	И	И

Таблица 2.6
Импликация

x	y	$x \Rightarrow y$
Л	Л	И
И	Л	Л
Л	И	И
И	И	И

Таблица 2.7
Равнозначность

x	y	$x \sim y$
Л	Л	И
И	Л	Л
Л	И	Л
И	И	И

Импликация характеризуется табл. 2.6.

Высказывание z называется *равнозначностью* высказываний x и y , если оно — истинно, когда x и y одновременно

менно истинны либо одновременно ложны. Записывается это так: $z = x \sim y$, читается: « x равнозначно y ». Например (рис. 2.18),

$$x_4 = x_1 \sim x_2,$$

где x_1 , x_2 и x_4 те же высказывания, что и в предыдущем примере. Равнозначность характеризуется табл. 2.7;

Дизъюнкция, конъюнкция, импликация и равнозначность являются *логическими функциями* двух переменных, принимающими значения вместе со своими аргументами из одного и того же двухэлементного множества $\{И, Л\}$. В

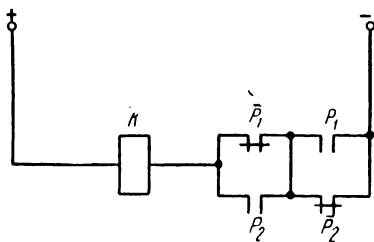


Рис. 2.18. Фрагмент релейно-контактной схемы (пример равнозначности двух высказываний)

технических приложениях элементы этого множества чаще обозначают через 0 и 1, а сами функции называются *переключательными*, или *булевыми функциями*. Другие переключательные функции двух аргументов — *двухместные функции* — могут быть представлены в виде суперпозиции вышеперечисленных и функции отрицания [2.14].

Непосредственной подстановкой всех возможных значений независимых переменных проверяется справедливость тождеств

$$x_1 \Rightarrow x_2 = \bar{x}_1 \vee x_2,$$

$$x_1 \sim x_2 = (x_1 \Rightarrow x_2) (x_2 \Rightarrow x_1),$$

которые показывают, что для записи всех переключательных двухместных функций достаточно трех функций: дизъюнкции, конъюнкции и отрицания. Непосредственно проверяются и следующие три тождества:

$$\bar{\bar{x}} = x,$$

$$x_1 \cdot x_2 = \bar{\bar{x}_1 \cdot x_2},$$

$$\bar{x}_1 \vee x_2 = \bar{\bar{x}_1 \cdot x_2}.$$

Алгебраическая операция, задаваемая конъюнкцией на множестве $\{0, 1\}$:

$$\begin{aligned} 0 \cdot 0 &= 0 \\ 0 \cdot 1 &= 0 \\ 1 \cdot 0 &= 0 \\ 1 \cdot 1 &= 1 \end{aligned}$$

называется *логическим умножением*. Алгебраическая операция, задаваемая дизъюнкцией на множестве $\{0, 1\}$:

$$\begin{aligned} 0 \vee 0 &= 0 \\ 0 \vee 1 &= 1 \\ 1 \vee 0 &= 1 \\ 1 \vee 1 &= 1 \end{aligned}$$

называется *логическим сложением*. Эти две операции обладают следующими свойствами:

1. Сложение и умножение коммутативны:

$$x \vee y = y \vee x, \quad x \cdot y = y \cdot x.$$

2. Сложение и умножение ассоциативны:

$$(x \vee y) \vee z = x \vee (y \vee z), \quad (x \cdot y) \cdot z = x \cdot (y \cdot z).$$

3. Справедливы два дистрибутивных закона:

$$(x \vee y) \cdot z = (x \cdot z) \vee (y \cdot z), \quad (x \cdot y) \vee z = (x \vee z) \cdot (y \vee z).$$

Обозначим через 1 *тождественно истинное высказывание*, т. е. истинное при всех значениях независимых пе-

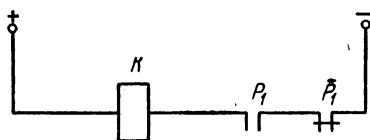


Рис. 2.19. Фрагмент релейно-контактной схемы (пример тождественно-ложного высказывания)

ременных, а через 0 — *тождественно-ложное высказывание* (ложное при всех значениях независимых переменных). Например, высказывание: «катушка K возбуждена» (рис. 2.19) тождественно ложно, так как катушка K обесточена при любом состоянии реле P_1 . Справедливы

следующие тождества:

$$\begin{array}{ll} x \cdot 1 = x & x \vee 1 = 1 \\ x \cdot 0 = 0 & x \vee 0 = x \\ x \cdot x = x & x \vee x = x \\ x \cdot \bar{x} = 0 & x \vee \bar{x} = 1 \end{array}$$

Устройства или элементы, выполняющие операции отрицания, конъюнкции и дизъюнкции, имеют названия, соответственно, «НЕ», «И» и «ИЛИ». Элемент «НЕ» выдает сигнал на выходе при отсутствии сигнала на входе (на-

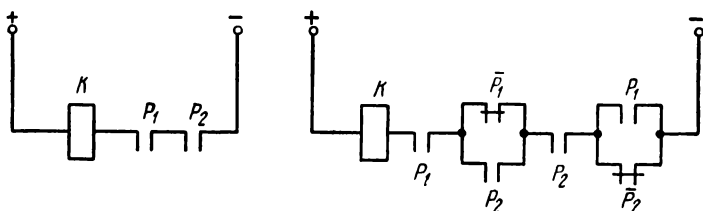


Рис. 2.20. Две функционально равнозначные схемы

пример, реле замыкает цепь при снятии возбуждения). Элемент «И» дает выходной сигнал при наличии сигналов на двух его входах. Элемент «ИЛИ» выдает сигнал при наличии сигнала на одном из входов. Более сложные элементы «И — НЕ» и «ИЛИ — НЕ». Первый из них выполняет операцию Шеффера, второй — операцию Пирса.

Примеры. 1. Дано: $z = P_1 \cdot (P_1 \sim P_2) \cdot P_2$. Записать высказывание z , пользуясь дизъюнкцией, конъюнкцией и отрицанием.

$$\begin{aligned} z &= P_1 \cdot (P_1 \Rightarrow P_2) \cdot (P_2 \Rightarrow P_1) \cdot P_2 = \\ &= P_1 \cdot (\bar{P}_1 \vee P_2) \cdot (\bar{P}_2 \vee P_1) \cdot P_2 = \\ &= (P_1 \cdot \bar{P}_1 \vee P_1 \cdot P_2) \cdot (\bar{P}_2 \cdot P_2 \vee P_1 \cdot P_2) = \\ &= (0 \vee P_1 \cdot P_2) \cdot (0 \vee P_1 \cdot P_2) = P_1 \cdot P_2 \cdot P_1 \cdot P_2 = P_1 \cdot P_2 \end{aligned}$$

Установленное тождество $P_1 \cdot (P_1 \sim P_2) \cdot P_2 = P_1 P_2$ можно проиллюстрировать функциональной равносильностью двух схем (рис. 2.20). Катушка контактора K в обоих случаях возбуждена только при одновременном возбуждении катушек реле P_1 и P_2 .

2. С помощью символов дизъюнкции и конъюнкции можно написать вдобавок к равносильностям (2.6), (2.8) и (2.9) следующие

равносильности [2.16], упрощающие запись логических схем алгоритмов:

$$l_1 \overset{i}{\sqcap} l_2 \overset{i}{\sqcap} = l_1 \vee l_2 \overset{i}{\sqcap},$$

$$l_1 \overset{i}{\sqcup} l_2 \overset{i}{\sqcup} = l_1 \cdot l_2 \overset{i}{\sqcup}.$$

§ 2.6. СИГНАЛЫ

Под влиянием управляющих и возмущающих воздействий в автоматической системе возникают *сигналы*, являющиеся материальными носителями информации об этих воздействиях. Сигнал представляет собой некоторый процесс, описываемый изменением какой-либо физической величины во времени. Такая величина описывает или изображает какой-либо параметр системы, т. е. передает информацию о нем. Например, напряжение на выходе термоэлемента описывает измеряемую температуру. Действие системы зависит от характера преобразований и условий передачи сигналов.

Употребляются сигналы, весьма различные по физической природе. Это показано в табл. 2.8, где перечислены виды сигналов в каналах связи функциональных схем, рассмотренных в табл. 2.2 (каналы занумерованы по рис. 2.4 — 2.8). Канал 3, по которому в систему вводятся критерии, определяющие ее действие, или программа, в таблицу не помещен. Другие каналы, как видно из таблицы, могут представлять собой разнообразные пути распространения сигналов. Вообще это могут быть световые лучи, электрические цепи, пути распространения тепла, акустических колебаний, механических, гидравлических, пневматических и метеорологических воздействий и т. п.

Сигналы каждого вида отличаются физическими характеристиками, свойственными этому виду, в зависимости от чего изменяется их *смысловое* (или семантическое) *значение*. Световые сигналы, например, имеют разный цвет. Совокупность их цветов позволяет однозначно выразить все сообщения или команды, которые могут быть необходимы в задаче управления движением, решаемой с помощью таких сигналов. Соответствие между множеством сигналов, получаемым варьированием физических характеристик сигнала какого-либо вида, и множеством передаваемых сообщений называют *изоморфизмом*.

Воздействие сигнала на какой-либо блок системы всегда требует некоторой работы. Следовательно, сигнал должен иметь достаточную *мощность*. При недостаточной, например, мощности электрических сигналов, которые должны вызывать какие-либо механические перемещения, этих перемещений не произойдет, и система перестанет действовать. При наличии усиления требуемая мощность сигнала определяется работой, необходимой для воздействия на усилитель.

Т а б л и ц а 2.8

Виды сигналов в каналах связи автоматических систем, рассмотренных в табл. 2.2

Канал	Система управления	
	уличным движением	курсом корабля
1	Световой	Гидравлический
2	Электрический	Электрический
4	То же	То же
5	Световой или звуковой	Гидравлические и аэродинамические
6	Световой или электрический	Те же, а также световые, звуковые или радио
7	То же	Механический
8	Электрический	Электрический
9	То же	Механический
10	—	Электрический
11	Электрический	То же
12	То же	„

Еще одно требование к сигналу состоит в том, что он должен приносить достаточную информацию. Информация должна быть достаточной, чтобы при приеме были

отличимы сигналы, различные по условиям изоморфизма. Пусть, например, нужно передавать лишь две команды: «вперед» и «назад». При передаче этих команд электрическими сигналами, изоморфизм может состоять в том, что одной команде соответствует положительное напряжение, а другой — отрицательное. Численное значение напряжения может быть в обоих случаях одинаковым и достаточным, чтобы сигналы были удовлетворительными по мощности. Если при передаче сигнала нет помех, то, вычисляя получаемую от него информацию I по формуле (1.27), следует положить $P_0 = 1/2$, $P = 1$. Тогда

$$I = \log_2 \frac{1}{1/2} = 1 \text{ дв. ед.}$$

Если же на принимаемый сигнал накладываются помехи, то $P < 1$. Это означает отсутствие уверенности в том, какой знак имеет принимаемое напряжение. При этом $I < 1$ дв. ед. Положение будет совершенно неудовлетворительным при полной неопределенности знака сигнала, когда $P = 1/2$ и $I = 0$. В этом примере информация, приносимая сигналом, может изменяться, в зависимости от помех, от нуля до единицы и она может быть достаточной, если ее значение будет ближе к верхнему пределу, когда достаточно близка к единице вероятность P .

Таким образом, сигналы должны удовлетворять требованиям изоморфизма и должны быть достаточными по мощности и информации. Несоблюдение хотя бы одного из этих трех условий ведет к нарушению действия системы.

Следует заметить, что требования достаточной мощности и информации могут быть совсем не связаны одно с другим. Например, наименьшая мощность сигналов, принимаемых по телефону, должна быть достаточной, чтобы они были различимы на фоне шумов. При этой мощности информация, получаемая по телефону, может быть в одних случаях очень мала, а в других — велика. Это зависит от того, что ожидается узнать и что узнается из телефонного разговора.

Согласно сказанному, вместо различных воздействий, в частности управляющих, будем рассматривать далее сигналы, несущие информацию о воздействиях.

Пример. Когда огни железнодорожного семафора различимы, они всегда дают одно и то же количество информации, равное 1 дв. ед. Сквозь туман зеленый огонь виден хуже красного. Если туман настолько плотен, что, подъезжая к семафору, машинист не может увидеть зеленого огня, то в случае, когда движение разрешено, никакого сигнала принято не будет. Не получая сигнала, машинист может с равной вероятностью предполагать, что имеет место любой из двух возможных сигналов. Помеха в виде тумана дает, таким образом, $P = 1/2$. Но и $P_0 = 1/2$, поскольку может быть лишь два сигнала. Следовательно, в рассматриваемом случае туман сводит полученную информацию к нулю.

ПРЕОБРАЗОВАНИЕ СИГНАЛОВ

§ 3.1. МАТЕМАТИЧЕСКОЕ ОПИСАНИЕ СИГНАЛОВ

Рассмотрим элемент системы, получающий информацию по каналу 1 и выдающий ее по каналу 2 (рис. 3.1). Воздействующий на этот элемент сигнал поступает на его вход (§ 2.2) в некоторый момент времени, до которого на входе сигнала не было. Примем за начало отсчета времени момент появления входного сигнала, т. е. положим, что сигнал появляется на входе, когда время $t=0$. Положим также, что физическая величина, характеризующая входной сигнал, является непрерывной функцией времени $f_1(t)$, называемой *описывающей*. Функцию $f_1(t)$ подчиним лишь тому условию, что при $t=0$ непрерывны и все ее производные, которые могут быть нужны для нахождения сигнала на выходе [3.8]. Тогда сигнал x_1 , появляющийся на выходе при $t=0$, можно представить следующим выражением:

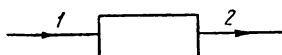


Рис. 3.1. Вход и выход элемента системы

$$x_1(t) = f_1(t) 1(t), \quad (3.1)$$

где $1(t)$ — *единичная функция*, определяемая как

$$1(t) = \begin{cases} 0 & t < 0 \\ 1 & t \geq 0 \end{cases} \quad (3.2)$$

(рис. 3.2). Единичная функция безразмерна.

Из выражения (3.1) следует, что функция $x_1(t)$ непрерывна при $t=0$ лишь в том случае, когда $f_1(0)=0$. Если же $f_1(0)=f_{10} \neq 0$, то функция $x_1(t)$ при $t=0$ имеет скачок, равный f_{10} (рис. 3.3), так как, согласно определению (3.2),

$$x_1(t) = 0 \text{ при } t < 0 \quad (3.3)$$

независимо от вида описывающей функции $f_1(t)$.

Передача сигнала от входа до выхода рассматриваемого элемента должна занимать некоторое время τ , обусловленное скоростью распространения сигнала внутри этого

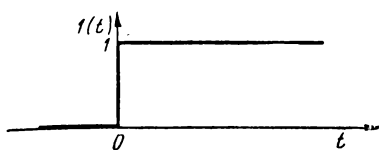


Рис. 3.2. Единичная функция

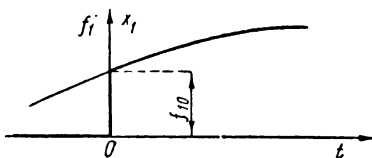


Рис. 3.3. Скачок при включении входного сигнала

элемента. Поэтому выходной сигнал, вызываемый сигналом $x_1(t)$, появится при $t = \tau$ и, аналогично (3.1), должен иметь выражение

$$x_2(t - \tau) = f_2(t) 1(t - \tau). \quad (3.4)$$

Здесь $f_2(t)$ — описывающая функция, представляющая собой зависимость физической величины, характеризующей

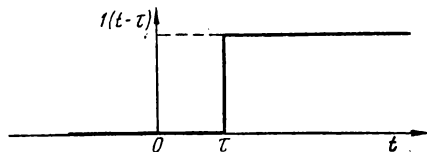


Рис. 3.4. Запаздывающая единичная функция

выходной сигнал, от времени. Запаздывающая единичная функция $1(t - \tau)$ имеет вид, показанный на рис. 3.4.

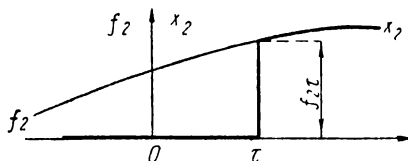


Рис. 3.5. Скачок при появлении выходного сигнала

Функция $x_2(t)$ имеет скачок $f_{2\tau}$ при $t = \tau$, если $f_2(\tau) = f_{2\tau} \neq 0$ (рис. 3.5). Величину τ называют *временем чистого или транспортного запаздывания*.

Выражение сигнала через описывающую и единичную функцию позволяет, таким образом, связать сигнал с действием того элемента автоматической системы, который его принимает или выдает. Описывающая функция может существовать при любых значениях времени. Сигнал, воздействующий на какой-либо элемент, существует для этого элемента лишь с момента включения. Выходной сигнал существует лишь с момента его появления на выходе того элемента, который его выдает. Именно такие

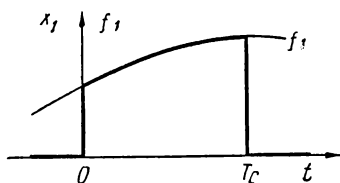


Рис. 3.6. Входной сигнал ограниченной длительности

сигналы, а не существующие неопределенно давно, рассматриваются при изучении процессов управления. Но выражения (3.2) и (3.4) представляют сигналы, длительность которых после включения не ограничена, тогда как в реальных процессах управления сигналы действуют

лишь в течение конечного времени. Поэтому в некоторых задачах описание сигналов выражениями указанного вида недостаточно, и необходимо иметь более полное математическое описание сигнала, показывающее не только его появление или включение, но и его исчезновение или выключение.

Рассмотрим входной сигнал $x_1(t)$, включаемый при $t=0$ и выключаемый при $t=T_c$:

$$x_1(t) = \begin{cases} 0 & t < 0 \\ f_1(t) & 0 \leq t \leq T_c \\ 0 & t > T_c \end{cases}$$

(рис. 3.6). Такой сигнал может быть выражен с помощью разности единичных функций

$$1(t) - 1(t - T_c),$$

т. е.

$$x_1(t) = f_1(t) [1(t) - 1(t - T_c)]. \quad (3.5)$$

Сигналы вида (3.1), (3.4) и (3.5) будем называть *непрерывными*, указывая этим на их характер в промежутке времени T_c от включения до выключения. Если время T_c мало по сравнению с длительностями, существенными

для рассматриваемого процесса, то сигнал будем называть *импульсивным*. В таком случае описывающая функция определяет *форму* импульсивного сигнала или, как часто сокращенно говорят, форму импульса. Когда T_c равно пренебрежимо малому промежутку времени Δ , импульсивный сигнал описывают не его зависимостью от времени в этом промежутке, а его *импульсом*, равным

$$\int_{\Delta} f(t) dt.$$

Такой импульс приближенно считают *мгновенным* и появляющимся в середине интервала Δ . Импульсивные сигналы часто следуют один за другим через одинаковые промежутки времени T . Время T называют *периодом повторения*. При этом получается *последовательность импульсивных сигналов* или же *последовательность мгновенных импульсов*, если таковые имеют место.

Интервал Δ может иметь любое значение, не превышающее периода повторения T . При $\Delta < T$ имеет место последовательность импульсов, которые можно считать мгновенными, если $\Delta \ll T$. При $\Delta = T$ вместо последовательности импульсов получается *кусочно-непрерывный сигнал*.

Если

$$f_1(t) = c = \text{const}, \quad (3.6)$$

т. е. описывающая функция $f_1(t)$ тождественно равна некоторой постоянной c , то

$$x_1(t) = c \cdot 1(t), \quad (3.7)$$

т. е. входной сигнал подобен единичной функции. При $c = 1$ такой сигнал называют *единичным скачком*. Не рассматривая квантования, которое обнаруживается при микроскопическом исследовании явлений, а ограничиваясь макроскопическим изучением, как это обычно имеет место в автоматике, следует иметь в виду, что единичная функция — лишь математическая идеализация реального процесса весьма быстрого возрастания воздействия от нуля до некоторого значения, которое затем остается неизменным [3.12]. Это поясняет рис. 3.7, на котором сплошной линией показано реальное воздействие x_1 , а пунктиром —

единичная функция, описывающая его приближенно. Замена реального процесса единичной функцией допустима, если время Δ , в течение которого x_1 возрастает от 0 до 1, пренебрежимо мало по сравнению с наименьшими длительностями, представляющими интерес в исследовании, при котором производится эта аппроксимация.

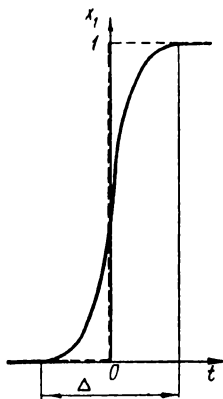


Рис. 3.7. Входной сигнал, аппроксимируемый единичным скачком

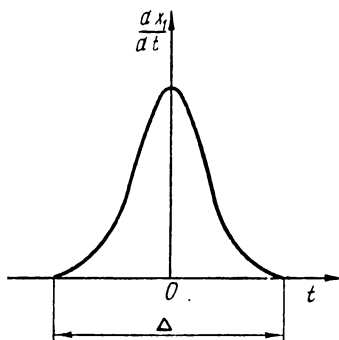


Рис. 3.8. Первая производная входного сигнала, аппроксимируемого единичным скачком

Рассмотрим первую производную сигнала x_1 , представленного сплошной линией на рис. 3.7, по времени. График этой производной имеет вид, показанный на рис. 3.8. Обозначим реальную зависимость сигнала от времени на участке Δ через $F(t)$ и вычислим импульс ее первой производной:

$$\int_{-\infty}^{\infty} \frac{dF}{dt} dt = F(t) \Big|_{-\infty}^{\infty} = 1,$$

т. е. этот импульс равен единице при любом виде функции $F(t)$, изменяющейся на участке Δ от 0 до 1 и любом значении Δ . То же имеет место и в случае предельного перехода, когда $\Delta \rightarrow 0$. При этом сигнал принимает вид единичного скачка (3.7), а его производная устремляется к бесконечности при $t=0$, но равна нулю при всех других значениях времени. При этом импульс первой производной функции $F(t)$ сохраняет равное единице значение

и становится *мгновенным единичным импульсом*, а производная $F'(t)$ превращается в *дельта-функцию*

$$\delta(t) = \begin{cases} 0 & t \neq 0 \\ \infty & t = 0 \end{cases}, \int_{-\infty}^{\infty} \delta(t) dt = 1. \quad (3.8)$$

Мгновенный единичный импульс безразмерен. Так как

$$\lim_{\Delta \rightarrow 0} F(t) = 1(t),$$

дельта-функция является первой производной единичной функции в обобщенном (по способу ее образования) смысле¹.

Рассмотрим теперь импульсивный сигнал в виде некоторой силы $F_1(t)$, действующей в течение времени $t_0, t_0 + \Delta$. Ее импульс

$$A = \int_{t_0}^{t_0 + \Delta} F_1(t) dt$$

можно представить также в виде

$$A = F_{\text{ср}} \Delta, \quad (3.9)$$

где $F_{\text{ср}}$ — среднее значение функции $F_1(t)$ в промежутке времени Δ . Потребуем, чтобы импульс A оставался неизменным, когда $\Delta \rightarrow 0$. Тогда силу, действующую в момент времени t_0 , можно записать как $A\delta(t - t_0)$, так как по (3.8) импульс такой силы равен A . Это — сила удара, имеющего импульс A . Таким образом, произведение $A\delta(t - t_0)$ выражает силу, с которой действует мгновенный импульс A , приложенный в момент времени t_0 . Предельная краткость этого выражения совмещается с большим содержанием: оно дает силу удара, показывает ее импульс и указывает время ее появления. При $A = 1$ имеет место мгновенный единичный импульс.

Положим, что удары, каждый из которых производится мгновенным импульсом (3.9), следуют друг за другом через одинаковые промежутки времени T , так что

$$t_0 = nT \quad (n = 0, 1, 2 \dots). \quad (3.10)$$

Тогда последовательность возникающих сил можно представить рядом

$$F_{\text{ср}} \Delta [\delta(t) + \delta(t - T) + \delta(t - 2T) + \dots].$$

¹ См. § 3.2.

Это и есть силы, так как каждый член ряда умножен на импульс, а дельта-функция имеет размерность, обратную времени. Силы, выражаемые членами этого ряда, появляются лишь в моменты времени (3.10), причем каждому значению n соответствует только один член ряда, отличающийся от нуля.

Предположим теперь, что импульс A изменяется от удара к удару, так что при ударе, происходящем в момент (3.10), импульс имеет значение $\Delta F_{1cp}(nT)$. Тогда последовательность возникающих сил примет вид

$$\Delta [F_{1cp}(0) \delta(t) + F_{1cp}(T) \delta(t-T) + F_{1cp}(2T) \cdot \delta(t-2T) + \dots].$$

Аналогично записывается всякая последовательность мгновенных импульсов, появляющихся через одинаковые интервалы времени T (период повторения), т. е. в моменты (3.10):

$$x(t) = \Delta [f(0) \delta(t) + f(T) \delta(t-T) + f(2T) \delta(t-2T) + \dots]. \quad (3.11)$$

Сигнал, который реально можно наблюдать в автоматической системе, конечно, не может быть образован

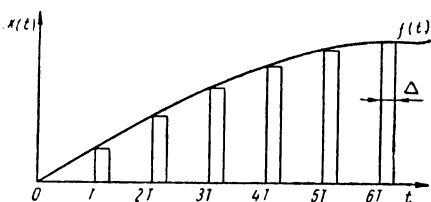


Рис. 3.9. Модулированная последовательность мгновенных импульсов

мгновенными импульсами. Он образован импульсами конечной длительности Δ и определенной формы, как, например, сигнал, изображенный на рис. 3.9. Но если длительность импульсов Δ достаточно мала, то для работы системы

она как и форма их, оказывается несущественной. Элементы системы одинаково реагируют на достаточно короткие импульсы разной формы, но одинаковой величины. На рис. 3.9 импульс, появляющийся в момент времени nT , равен $\Delta f(nT)$. В выражении (3.11) этот импульс имеет то же значение

$$\int_{nT - \frac{T}{2}}^{nT + \frac{T}{2}} \Delta \delta(t - nT) f(nT) dt = \Delta f(nT). \quad (3.12)$$

Форма записи сигнала (3.11) есть математическая идеализация реального процесса, как и выражение (3.7). Эти идеализации значительно упрощают анализ автоматических систем.

Непрерывные сигналы, поступающие на вход импульсных устройств, удобно описывать выборочными значениями, которые они принимают в определенные моменты времени, называемыми *дискретами*. С этой целью пользуются функциями, отличными от нуля лишь при целочисленных значениях аргумента n , называемыми *решетчатыми*. Аргумент таких функций записывают в квадратных скобках. График *единичной решетчатой функции* представлен на рис. 3.10. Сигналу (3.1) соответствует решетчатая функция

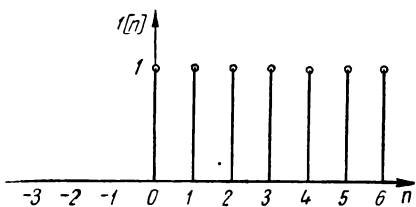


Рис. 3.10. Решетчатая единичная функция

$$x_1[n_1] = f_1(t) 1[n_1].$$

Вследствие фильтрующего действия функции $1[n]$ описывающую функцию здесь можно представить также в решетчатой форме. Тогда

$$x_1[n_1] = f_1[n_1] 1[n_1]. \quad (3.13)$$

Выходной сигнал, вызываемый поданным на вход сигналом $x_1[n_1]$ при наличии запаздывания τ ,

$$x_2[n_2] = f_2[n_2] 1[n_2], \quad (3.14)$$

где аргумент $n_2 = 0$ при $t = \tau$. Следовательно, выходной сигнал появляется в моменты времени

$$t - \tau = n_2 T.$$

Выходной сигнал смещен относительно входного на время τ или на *относительное смещение*

$$\varepsilon = \frac{\tau}{T}. \quad (3.15)$$

Полагая

$$n_1 = n,$$

можно рассматривать входной сигнал как решетчатую функцию $x_1[n]$ аргумента n , выходной — как *смещенную*

решетчатую функцию $x_2[n, \varepsilon]$ того же аргумента n , где ε является параметром:

$$x_2[n, \varepsilon] = f_2[n, \varepsilon] 1[n, \varepsilon]. \quad (3.16)$$

Шкалы аргументов n_1 и n_2 представлены на рис. 3.11. Рассматривая переменное смещение, непрерывную функцию можно полностью описать смещенной решетчатой при $0 \leq \varepsilon < 1$.

Входной сигнал, выключаемый при $n = N_c$,

$$x_1[n] = f_1[n] \{1[n] - 1[n - N_c - 1]\}. \quad (3.17)$$

Сигналы вида (3.13), (3.16) и (3.17) будем называть *решетчатыми*.

Замечая, что

$$f_1[n] = f_1(nT),$$

можно также рассматривать выражение вида (3.16)

$$x[n] = f[n] 1[n] \quad (3.18)$$

как описание сигнала, образуемого последовательностью мгновенных импульсов (3.11). Действительно, согласно

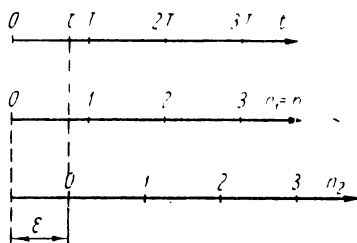


Рис. 3.11. Шкалы целочисленных аргументов входного и выходного сигналов

(3.12), импульс каждого члена последовательности (3.11) пропорционален соответствующей тому же аргументу n дискрете сигнала (3.18). Каждая дискрета представляет собой значение описывающей функции, модулирующей мгновенные единичные импульсы в (3.11). Поэтому график решетчатой функции (3.18) можно рассматривать

как графическое представление сигнала (3.11), не поддающегося непосредственно графическому описанию вследствие наличия дельта-функций в его выражении.

Все сигналы, значения которых изменяются лишь скачками, называются *дискретными*. Сюда относятся последовательности мгновенных импульсов, решетчатые сигналы, а также последовательности импульсивных сигналов (в частности, кусочно-непрерывные сигналы), сохраняющие постоянное значение в интервале Δ , т. е. образуемые

прямоугольными импульсами. Кусочно-непрерывные сигналы, образуемые импульсами другой формы, являются *дискретно-непрерывными*.

С помощью единичных функций можно составить, как показано выше, простое и адекватное физическим условиям описание сигналов, связанных с элементами автоматической системы в течение ограниченного времени. Для импульсивных сигналов такое описание можно составить также с помощью дельта-функций. Применение этих функций к задачам автоматического управления позволяет получать строгие решения наиболее простыми и изящными способами, как это будет показано далее.

Примеры. 1. Пусть описывающая функция непрерывного сигнала

$$f(t) = 2 \sin \left(100t + \frac{\pi}{2} \right),$$

где t — в секундах. Сигнал подается на вход некоторого элемента автоматической системы при $t = 10$ сек. Тогда его выражение следует записать в виде

$$x(t) = 2 \sin \left(100t + \frac{\pi}{2} \right) 1(t - 10).$$

Введя время

$$t' = t - 10,$$

запишем тот же сигнал иначе:

$$x(t') = 2 \sin \left[100(t' + 10) + \frac{\pi}{2} \right] 1(t').$$

Из обоих выражений видно, что в момент включения сигнал появляется скачком.

2. Пусть та же описывающая функция модулирует последовательность импульсов, имеющих длительность $\Delta = 0,5$ сек; период повторения $T = 1$ сек. С помощью (3.11) эту последовательность можно выразить в функции времени t' :

$$x[n] = [\delta(t') + \delta(t' - T) + \delta(t' - 2T) + \dots + \delta(t' - nT) + \dots] \sin \left[100(t' + 10) + \frac{\pi}{2} \right].$$

В форме решетчатой функции эта последовательность

$$x[n] = 2 \sin \left[100(n + 10) + \frac{\pi}{2} \right] 1[n].$$

§ 3.2. ПРОИЗВОДНЫЕ РЕАЛЬНЫХ СИГНАЛОВ

Изучение всякой непрерывной системы требует исследования сигнала $x_2(t)$, возникающего на выходе при появле-

нии какого-либо сигнала $x_1(t)$ на входе (см. рис. 2.2). В простейшем случае при этом можно пренебречь продолжительностью распространения сигнала от входа до выхода, полагая, что время чистого запаздывания (см.

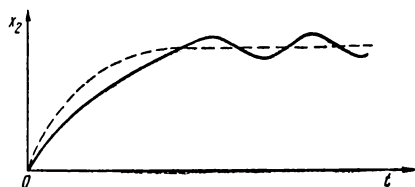


Рис. 3.12. Два вида установившегося процесса

§ 3.1) $\tau = 0$. Однако большинство автоматических устройств обладает *инерционностью*: при появлении сигнала x_1 на выходе начинается переходный процесс, по окончании которого наступает установившийся процесс (см. § 1.1). На рис. 3.12 показаны

случаи постоянного значения x_2 в установившемся процессе и гармонического установившегося процесса. Это явление называют *инерционным запаздыванием*. Если нельзя пренебречь чистым запаздыванием, то те же процессы имеют вид, изображенный на рис. 3.13.

Вследствие инерционности изучаемых систем в них возникают переходные процессы динамического характера. Для исследования таких процессов в непрерывных системах необходимо учитывать производные выходного сигнала, а во многих случаях и производные входного сигнала по времени, т. е. вычислять производные сигналов вида (3.1). При $t = 0$ такие функции могут иметь разрыв или угловую точку. Поэтому обычные производные здесь неприменимы [3.3] и должны быть заменены производными, называемыми *обобщенными* [3.8]. Обобщенную производную функции $x(t)$ будем обозначать символом Dx , а обычную производную функции $f(t)$ — штрихом, т. е. $f'(t)$.

Вычисляя обобщенную производную сигнала (3.1) как производную произведения, найдем [3.4]:

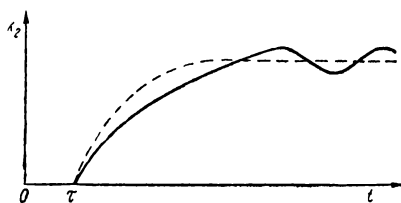


Рис. 3.13. Переходные и установившиеся процессы при наличии чистого запаздывания

$$Dx = f'(t) 1(t) + f(t) \delta(t).$$

Так как $\delta(t) = 0$ при $t \neq 0$, вместо $f(t)$ можно написать $f(0)$. Тогда находим:

$$Dx = f'(t) 1(t) + f(0) \delta(t). \quad (3.19)$$

Применяя операцию такого *обобщенного дифференцирования* [3.17] к выражению (3.19), получим вторую обобщенную производную

$$D^2x = f''(t) 1(t) + f'(0) \delta(t) + f(0) \delta'(t),$$

где $\delta'(t)$ — первая производная дельта-функции, понимаемая в обобщенном смысле. Так же найдем третью обобщенную производную

$$D^3x = f'''(t) 1(t) + f''(0) \delta(t) + f'(0) \delta'(t) + f(0) \delta''(t),$$

содержащую и вторую обобщенную производную дельта-функции. Обобщенная производная порядка k , как видно из предыдущих выражений, будет иметь вид

$$D^k x = f^{(k)}(t) 1(t) + f^{(k-1)}(0) \delta(t) + f^{(k-2)}(0) \delta'(t) + \dots + f(0) \delta^{(k-1)}(t), \quad (3.20)$$

где все производные дельта-функции следует понимать в обобщенном смысле [3.8].

Если $f^{(k-l)}(0) = 0$, то по правилам действий с дельта-функцией получаем:

$$f^{(k-l)}(0) \delta(t) = 0,$$

а $D^k x$ не содержит импульсов порядка l . В частности, при $f(0) = 0$

$$Dx = f'(t) 1(t). \quad (3.21)$$

Рассмотрим поведение обобщенных производных сигнала (3.1), непрерывного при $t=0$. В этом случае $x(0) = f(0) = 0$. Если и $f'(0) = 0$, то по (3.19) $Dx|_{t=0} = 0$, т. е. при $t=0$ непрерывна и производная Dx . Если же $f'(0) = f'_0 \neq 0$, то по той же формуле $Dx|_{t=0} = f'_0$, т. е. обобщенная производная имеет скачок при $t=0$. При этом, очевидно, функция $x(t)$ имеет угловую точку в начале координат, а D^2x — импульсивную составляющую $f''(0) \delta(t)$.

Пусть $f(0) = f'(0) = 0$, но $f''(0) = f''_0 \neq 0$. Применяя (3.19), находим, что

$$Dx|_{t=0} = 0;$$

вторая обобщенная производная по (3.20)

$$D^2x = f''(t) \mathbf{1}(t),$$

т. е.

$$D^2x|_{t=0} = f''_0.$$

Следовательно, при $t=0$ первая обобщенная производная имеет угловую точку, а вторая — скачок. Третья обобщенная производная содержит импульс по (3.20):

$$D^3x = f'''(t) \mathbf{1}(t) + f''_0 \delta(t).$$

Следующие обобщенные производные содержат импульсы все более высоких порядков, т. е. все более высокого порядка обобщенные производные δ -функции.

Если

$$f(0) = f'(0) = \dots = f^{(k)}(0) = 0,$$

но начальные значения производных порядка $k+1$ и выше не равны нулю, причем

$$f^{(k+1)}(0) = f^{(k+1)}_0 \neq 0,$$

то, как показывает (3.20), обобщенные производные порядка k и ниже при $t=0$ равны нулю, обобщенная производная порядка k имеет в начале координат угловую точку, обобщенная производная порядка $k+1$ имеет при $t=0$ скачок, равный $f^{(k+1)}_0$, а обобщенные производные порядка $k+2$ и выше содержат импульсы. В таком случае сигнал $x(t)$ имеет *гладкость* порядка k . Например, если сигнал первого порядка гладкости $k=1$, а

$$f(0) = f'(0) = 0,$$

то при $t=0$, по (3.19), $Dx=0$. Вторая производная ($k+1=2$)

$$\begin{aligned} D^2x &= D[f'(t) \mathbf{1}(t)] = \\ &= f''(t) \mathbf{1}(t) + f'(0) \delta(t) = f''(t) \mathbf{1}(t), \end{aligned}$$

откуда следует, что при $t=0$ эта производная имеет скачок, равный f'_0 . Третья обобщенная производная

$$D^3x = D[f''(t) \mathbf{1}(t)] = f'''(t) \mathbf{1}(t) + f''_0 \delta(t),$$

т. е. содержит импульс, так что нельзя говорить о ее конечном значении при $t=0$, хотя $f'''(0)$ имеет конечное значение.

Таким образом, производные сигнала (3.1) следует рассматривать как обобщенные, вычисляя их по формуле (3.20), и в том случае, когда при $t=0$ сигнал непрерывен и не имеет угловой точки, так как при этом обобщенные производные некоторых порядков могут иметь разрывы или содержать импульсы, отличаясь этим от обычных производных.

Нахождение функции по ее обобщенной производной — однозначная операция, не зависящая, в отличие от обычного интегрирования, ни от каких начальных условий [3.8]. Для доказательства этого важнейшего свойства обобщенной производной положим:

$$Dx \equiv 0. \quad (3.22)$$

Тогда из (3.19) следует, что

$$f'(t) 1(t) + f(0) \delta(t) \equiv 0, \quad (3.23)$$

а для этого необходимо, чтобы функция $f(t)$ удовлетворяла условиям

$$f'(t) = 0,$$

$$f(0) = 0,$$

т. е. чтобы $f(t) \equiv 0$, откуда

$$x(t) \equiv 0. \quad (3.24)$$

Таким образом, из тождества (3.22) следует тождество (3.24), что и доказывает указанное свойство обобщенной производной.

Обобщенная производная импульсивного сигнала, описываемого произведением дельта-функции на некоторую постоянную (см. § 3.1), выражается произведением той же постоянной на обобщенную производную дельта-функции. В общем случае сигнал может иметь *импульсивные составляющие* в виде обобщенных производных дельта-функции,

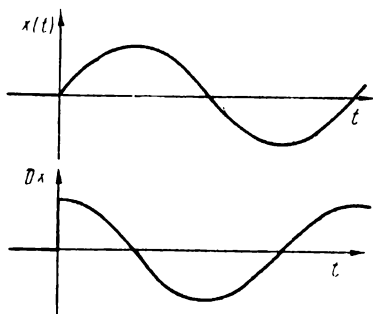


Рис. 3.14. Синусоидальный сигнал и его первая обобщенная производная

умноженных на некоторые коэффициенты, т. е.

$$x(t) = f(t) 1(t) + \lambda_0 \delta(t) + \lambda_1 \delta'(t) + \dots + \lambda_k \delta^{(k)}(t), \quad (3.25)$$

$$Dx = f'(t) 1(t) + f(0) \delta(t) + \lambda_0 \delta'(t) + \dots + \lambda_k \delta^{(k+1)}(t). \quad (3.26)$$

Аналогичную интегрированию операцию получения функции $x(t)$ по ее обобщенной производной будем обозначать через $D^{-1}x$ и рассматривать как *обобщенный интеграл* [3.15]. На основании доказанного выше свойства обобщенной производной

$$D^{-1}[Dx(t)] = x(t). \quad (3.27)$$

Примеры. 1. Если $f(t) = \sin t$, то $x(t) = \sin t \cdot 1(t)$ и $x(0) = 0$ — сигнал не делает скачка при $t=0$. Но $Dx = \cos t \cdot 1(t)$, а следова-

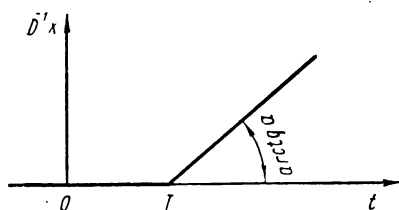


Рис. 3.15. Обобщенный интеграл сигнала $x(t) = a 1(t - T)$

тельно, Dx имеет скачок, $x(t)$ — угловую точку в начале координат (рис. 3.14), т. е. $x(t)$ — нулевого порядка гладкости [3.8].

2. Вычислим обобщенный интеграл сигнала

$$x(t) = a 1(t - T).$$

Рассматривая $x(t)$ как первую обобщенную производную сигнала

$$D^{-1}x = f(\tau) 1(\tau),$$

находим:

$$\tau = t - T,$$

$$\frac{d}{d\tau} f(\tau) = a,$$

$$f(0) = 0.$$

Следовательно (рис. 3.15),

$$f(\tau) = a\tau,$$

$$D^{-1}x = a(t - T) 1(t - T).$$

§ 3.3. ПРЕОБРАЗОВАНИЯ ЛАПЛАСА И ФУРЬЕ

Исследование связи между выходными и входными сигналами автоматических систем и их элементов весьма упрощается введением *изображений* этих сигналов *по Лапласу* и *по Фурье*. Не касаясь обоснований этих преобразований, изучаемых в математике, рассмотрим их применение к сигналам, описываемым непрерывными и решетчатыми функциями.

В табл. 3.1 сопоставлены некоторые сведения о *преобразованиях Лапласа* для сигналов обоих видов. Так как сигналы $x(t)$ и $x[n, \epsilon]$ равны нулю при отрицательных значениях t и n , их односторонние и двусторонние изображения совпадают. Поэтому далее рассматриваются только *односторонние изображения* этих сигналов, выражения которых приведены в таблице. Рассматриваемые сигналы являются функциями вещественных переменных t и n . В изображениях эти сигналы описываются с помощью, соответственно, функций комплексных переменных p и q . Комплексная переменная p , как и ее вещественная и мнимая части σ и ω , имеет размерность, обратную времени. Комплексная переменная q безразмерна, как и ее вещественная и мнимая части α и β . Величина ω является угловой частотой; это видно из выражения *ядра изображения* в тригонометрической форме. Для краткости величину ω будем называть далее просто *частотой*. Аналогичный смысл имеет величина β , называемая *относительной частотой*. Изображения и сигналы, т. е. *оригиналы*, будем связывать знаком соответствия \doteq . Изображения существуют, если сигналы мажорированы по модулю некоторыми показательными функциями *показателя роста* s и a . Как известно из операторного исчисления, каждому оригиналу соответствует определенное изображение, но данному изображению могут соответствовать оригиналы, имеющие различные значения в точках разрыва; если такие оригиналы не считать различными, то операторные соответствия взаимно однозначны.

Изображение непрерывного сигнала существует, если *интеграл Лапласа*, стоящий в правой части изображения, сходится. В зависимости от вида функции $x(t)$ сходимость интеграла Лапласа будет иметь место при тех или иных значениях p . Комплексные числа будем представлять на комплексной плоскости в прямоугольных коор-

Преобразования Лапласа

Сигнал (оригинал)	$x(t) = f(t) \mid (t)$	$x[n, \varepsilon] = f[n, \varepsilon] \mid [n, \varepsilon]$
Изображение	$X(p) = \int_0^\infty e^{-pt} x(t) dt$	$X(q, \varepsilon) = \sum_{n=0}^\infty e^{-qn} x[n, \varepsilon]$
Комплексная переменная	$p = \sigma + i\omega$	$q = \alpha + i\beta$
Ядро	$e^{-pt} = e^{-\sigma t} (\cos \omega t - i \sin \omega t)$	$e^{-qn} = e^{-\alpha n} (\cos \beta n - i \sin \beta n)$
Формула соответствия	$X(p) \doteq x(t)$	$X(q, \varepsilon) \doteq x[n, \varepsilon]$
Условие существования изображения	$ x(t) \leq M e^{st}, 0 < M < \infty, s < \infty$	$ x[n, \varepsilon] \leq M e^{an}, 0 < M < \infty, a < \infty$
Абсцисса абсолютной сходимости	$\sigma_0 \leq s$	$\alpha_0 \leq a$
Обратное преобразование	$x(t) = \frac{1}{2\pi i} \int_{\sigma - i\infty}^{\sigma + i\infty} e^{pt} X(p) dp$	$x[n, \varepsilon] = \frac{1}{2\pi i} \int_{\alpha - i\infty}^{\alpha + i\infty} e^{qn} X(q, \varepsilon) dq$
Свертка	$X_1(p) X_2(p) \doteq \int_0^t x_1(\tau) x_2(t - \tau) d\tau = \int_0^t x_1(t - \tau) x_2(\tau) d\tau$	$X_1(q) X_2(q) \doteq \sum_{m=0}^n x_1[m] x_2[n - m] = \sum_{m=0}^n x_1[n - m] x_2[m]$

динатах, которыми служат вещественная часть, откладываемая по оси абсцисс, и мнимая часть, откладываемая по оси ординат. Число p представляется точкой (σ, ω) такой плоскости (рис. 3.16). Из операторного исчисления известно, что если интеграл Лапласа сходится в некоторой точке p_1 , то он сходится во всех точках, для которых $\operatorname{Re} p > \operatorname{Re} p_1$. Для каждого сигнала существует абсцисса абсолютной сходимости, т. е. такое число σ_0 , что интеграл

$$\int_0^{\infty} |e^{-pt} x(t)| dt$$

существует, а следовательно, интеграл Лапласа абсолютно сходится при $\operatorname{Re} p > \sigma_0$ и не существует, т. е. расходится, при $\operatorname{Re} p < \sigma_0$.

Изображение решетчатого сигнала существует, если сходится ряд, через который оно выражается. В зависимости от вида функции $x[n, \varepsilon]$ это может иметь место при тех или иных значениях q . Если ряд сходится в некоторой точке q_1 комплексной плоскости, то он сходится во всех точках, для которых $\operatorname{Re} q > \operatorname{Re} q_1$. Для каждого сигнала существует абсцисса абсолютной сходимости α_0 : изображение существует, т. е. ряд сходится, при $\operatorname{Re} q > \alpha_0$ и не существует, т. е. ряд расходится, при $\operatorname{Re} q < \alpha_0$.

Абсциссы σ_0 и α_0 являются абсциссами самых правых (на комплексных плоскостях p и q) особых точек изображения. Эти особые точки лежат на прямых

$$\operatorname{Re} p = \sigma_0 \quad \text{и} \quad \operatorname{Re} q = \alpha_0.$$

Таким образом, абсциссы абсолютной сходимости могут быть найдены из исследования изображений. Они не больше показателей роста. В простейшем случае, когда описывающая функция $f[n] = 1$, т. е. сигнал, представляет собой единичную решетчатую функцию, его изображение

$$X(q) = \sum_{n=0}^{\infty} e^{-qn} 1[n] = \frac{1}{1 - e^{-q}}, \quad |e^{-q}| < 1$$

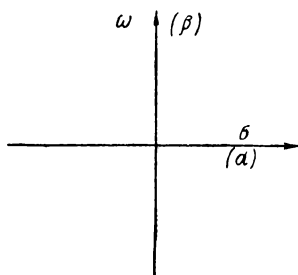


Рис. 3.16. Плоскость комплексных переменных p и q

(сумма членов бесконечной геометрической прогрессии), откуда $\alpha_0 = 0$, т. е. изображение существует во всей правой полуплоскости.

Изображения решетчатых сигналов по Лапласу всегда являются функциями экспоненты e^q . А так как $e^q = e^{q+i2k\pi}$, то изображение $2k\pi$ — периодическая функция вдоль мнимой оси. Таким образом, изображение достаточно задать при

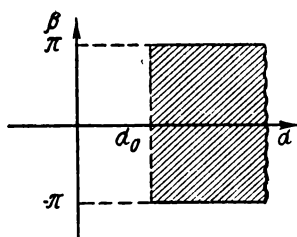


Рис. 3.17. Часть плоскости q , в которой полностью определяются изображения решетчатых сигналов

$\operatorname{Re} q > \alpha_0$ в любой полуполосе, параллельной вещественной оси и имеющей ширину 2π . На рис. 3.17 показана такая полуполоса, симметричная относительно вещественной оси. С этим свойством изображений дискретных сигналов связана возможность конформного отображения плоскости q на плоскость $z = e^q$. На этом основано так называемое z -преобразование.

Могут встретиться такие сигналы, которые не имеют свойств, необходимых для оригинала, т. е. не имеют изображения по Лапласу. Если, например, описывающая функция $f(t) = \frac{1}{t}$, т. е. представляет равнобокую гиперболу, то

$$x \cdot t = f(t),$$

а следовательно, сигнал не подчиняется условию мажорирования по модулю. При этом изображения нет, потому что, как легко видеть, интеграл Лапласа расходится при всех значениях p .

Условию мажорирования не удовлетворяет также дельта-функция. Оно выполняется, однако, для сигнала, создающего единичный импульс. Представим такой сигнал прямоугольником, показанным на рис. 3.18; он возникает при $t=0$, имеет длительность Δ и равен $1/\Delta$. Изображение такого сигнала равно

$$\frac{1 - e^{-p\Delta}}{p\Delta}.$$

При $\Delta \rightarrow 0$, когда рассматриваемый сигнал превращается в $\delta(t)$, это изображение имеет предел, равный единице.

Поэтому единица является изображением дельта-функции в обобщенном, по способу его образования, смысле [3.18].

В табл. 3.1 приведены также формулы свертки изображений кусочно-непрерывных и дискретных сигналов по Лапласу.

В тех случаях, когда абсциссы абсолютной сходимости σ_0 и α_0 отрицательны, изображения сигналов $x(t)$ и $x[n]$ существуют соответственно при $\operatorname{Re} p = 0$ и $\operatorname{Re} q = 0$. Иначе говоря, при отрицательных σ_0 и α_0 можно строить изображения непрерывных сигналов, полагая

$$p = i\omega, \quad (3.28)$$

а изображения решетчатых сигналов, полагая

$$q = i\beta. \quad (3.29)$$

При таких значениях комплексных переменных p и q из преобразований Лапласа получаются преобразования Фурье. В табл. 3.2 сопоставлены некоторые сведения о преобразованиях Фурье для сигналов $x(t)$ и $x[n]$.

Из условий существования изображения по Лапласу очевидно, что изображение по Фурье не может быть у сигналов, для которых абсциссы абсолютной сходимости σ_0 и α_0 положительны, так как в этом случае при $\sigma = 0$ интеграл Лапласа расходится. Следовательно, не все сигналы, имеющие изображение по Лапласу, имеют изображение по Фурье. Изображение по Фурье имеют лишь сигналы, для которых соблюдаются условия существования изображений, указанные в табл. 3.2. Это означает, что все особые точки изображений таких сигналов по Лапласу находятся в левой полуплоскости комплексных переменных p и q .

Изображения по Фурье всегда имеют абсолютно интегрируемые сигналы, т. е. удовлетворяющие условию

$$\int_0^{\infty} |e^{-i\omega t} x(t)| dt < +\infty, \quad (3.30)$$

так как

$$\int_0^{\infty} |e^{-i\omega t} x(t)| dt = \int_0^{\infty} |x(t)| dt < +\infty.$$

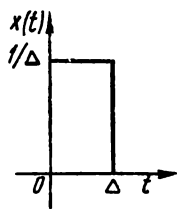


Рис. 3.18. К выводу изображения дельта-функции

Преобразования Фурье

Сигнал (оригинал)	$x(t) = f(t) 1(t)$	$x[n, \epsilon] = f[n, \epsilon] 1[n, \epsilon]$
Изображение	$X(i\omega) = \int_0^{\infty} e^{-i\omega t} x(t) dt$	$X(i\beta, \epsilon) = \sum_{n=0}^{\infty} e^{-i\beta n} x[n, \epsilon]$
Формула соответствия	$X(i\omega) \stackrel{*}{=} x(t)$	$X(i\beta, \epsilon) \stackrel{*}{=} x[n, \epsilon]$
Ядро	$e^{-i\omega t} = \cos \omega t - i \sin \omega t$	$e^{-i\beta n} = \cos \beta n - i \sin \beta n$
Условие существования изображения	$\sigma_0 < 0$	$\alpha_0 < 0$
Обозначение вещественной и мнимой частей изображения	$X(i\omega) = R(\omega) + iI(\omega)$	$X(i\beta, \epsilon) = R(\beta, \epsilon) + iI(\beta, \epsilon)$
Вещественная часть изображения	$R(\omega) = \int_0^{\infty} x(t) \cos \omega t dt$	$R(\beta, \epsilon) = \sum_{n=0}^{\infty} x[n, \epsilon] \cos \beta n$

Продолжение табл. 3.2

Сигнал (оригинал)	$x(t) = f(t) \text{ и } (t)$	$x[n, \epsilon] = f[n, \epsilon] \text{ и } [n, \epsilon]$
Мнимая часть изображения	$I(\omega) = -\int_0^{\infty} x(t) \sin \omega t \, dt$	$I(\beta, \epsilon) = -\sum_{n=0}^{\infty} x[n, \epsilon] \sin \beta n$
Интеграл Фурье	$x(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{it\omega} X(i\omega) \, d\omega$	$x[n, \epsilon] = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{i\beta n} X(i\beta, \epsilon) \, d\beta$
Вещественные формы интеграла Фурье	$x(t) = \frac{2}{\pi} \int_0^{\infty} R(\omega) \cos t\omega \, d\omega =$ $= -\frac{2}{\pi} \int_0^{\infty} I(\omega) \sin t\omega \, d\omega$	$x[n, \epsilon] = \frac{2}{\pi} \int_0^{\pi} R(\beta, \epsilon) \cos \beta n \, d\beta =$ $= -\frac{2}{\pi} \int_0^{\pi} I(\beta, \epsilon) \sin \beta n \, d\beta$

Заметим, что реально наблюдаемые сигналы имеют ограниченную длительность, т. е. отличны от нуля на промежутке времени конечной длины и ограничены по величине, а потому абсолютно интегрируемы.

Если же изображение по Лапласу имеет особые точки на мнимой оси или правой полуплоскости, то оригинал не удовлетворяет условию (3.30). В этих случаях интеграл Лапласа не сходится при $\sigma=0$, и преобразование Фурье невозможно.

Рассмотрим некоторый сигнал вида (3.1), для которого $\sigma_0=0$ (например, единичный скачок), и составим изображение по Лапласу этого сигнала, умноженного на $e^{-\mu t}$, где $\mu > 0$. По табл. 3.1 это будет

$$\int_0^{\infty} e^{-(p+\mu)t} x(t) dt = X(p+\mu).$$

Отсюда видно, что абсцисса абсолютной сходимости интеграла Лапласа от оригинала $e^{-\mu t} x(t)$ отрицательна: $\sigma_0 = -\mu$. Следовательно, из изображения

$$x(p+\mu) \doteq e^{-\mu t} x(t) \quad (3.31)$$

можно получить изображение по Фурье, полагая $\operatorname{Re} p=0$, при сколь угодно малом μ . Тогда предельным переходом при $\mu \rightarrow 0$ из (3.31) получается изображение по Фурье сигнала $x(t)$, для которого $\sigma_0=0$. Такое изображение, вводимое независимо от возможности его получения по табл. 3.2, называют изображением по Фурье в обобщенном смысле [3.18]. Оно существует при $\operatorname{Re} p=0$ только для таких оригиналов, для которых $\sigma_0=0$ [3.6]. Подобным же образом при $\sigma_0=0$ получается изображение в несобственном смысле для решетчатых сигналов.

Представляя ядро преобразования в тригонометрической форме, легко получить приведенные в таблице выражения вещественной и мнимой частей изображения. Как видно из этих выражений, вещественная часть изображения по Фурье — четная функция частоты ω , а мнимая — нечетная.

В табл. 3.2 приведены выражения обратных преобразований, называемых *интегралами Фурье*. Как и обратные преобразования Лапласа, их следует понимать в смысле Дирихле: интеграл, стоящий в правой части, равен значению оригинала в точках его непрерывности, а в точках

разрыва — полусумме значений оригинала слева и справа от точки разрыва [3.6]. Поэтому при $t < 0$ эти интегралы равны нулю.

Так как рассматриваемые сигналы — вещественные функции, выражающие их интегралы Фурье имеют вещественные формы. Некоторые из таких форм приведены в табл. 3.2. Покажем преобразование интеграла к этим формам для непрерывного сигнала [3.6]. Из приведенных в таблице выражений имеем:

$$e^{it\omega} X(i\omega) = (\cos t\omega + i \sin t\omega) [R(\omega) + iI(\omega)] = \\ = R(\omega) \cos t\omega - I(\omega) \sin t\omega + i [R(\omega) \sin t\omega + I(\omega) \cos t\omega].$$

Здесь вещественная часть — четная, а мнимая — нечетная функция частоты ω . Поэтому после подстановки последнего выражения в интеграл Фурье получаем:

$$x(t) = \frac{1}{\pi} \int_0^{\infty} [R(\omega) \cos t\omega - I(\omega) \sin t\omega] d\omega. \quad (3.32)$$

Принимая во внимание, что $x(t) = 0$ при $t < 0$, и полагая $t = -\tau$, $\tau > 0$, имеем:

$$0 = \frac{1}{\pi} \int_0^{\infty} [R(\omega) \cos \tau\omega + I(\omega) \sin \tau\omega] d\omega,$$

откуда, возвращаясь к обозначению параметра буквой t , находим, что при $t > 0$

$$\int_0^{\infty} R(\omega) \cos t\omega d\omega = - \int_0^{\infty} I(\omega) \sin t\omega d\omega.$$

Пользуясь этим соотношением, получаем из (3.32) выражения, приведенные в табл. 3.2.

В табл. 3.3 приведены некоторые непрерывные сигналы и их изображения по Лапласу и Фурье, а также абсциссы абсолютной сходимости интеграла Лапласа. В тех случаях когда $\sigma_0 = 0$, изображение по Фурье — в несобственном смысле. Табл. 3.4 содержит аналогичные данные для некоторых решетчатых сигналов. Существуют весьма подробные таблицы изображений по Лапласу [3.16]. В таких таблицах обычно опускают единичные функции в выражениях оригиналов, которые здесь необходимо писать в связи с принятым способом описания сигналов (см. § 3.1).

Таблица 3.3

Изображения некоторых непрерывных сигналов по Лапласу и Фурье

$x(t)$	$X(p)$	σ_0	$X(i\omega)$
$c * 1(t)$	$\frac{c}{p}$	0	$\frac{c}{i\omega}$
$e^{-bt} 1(t), b > 0$	$\frac{1}{p+b}$	$-b$	$\frac{1}{i\omega+b}$
$e^{bt} 1(t), b > 0$	$\frac{1}{p-b}$	b	—
$\sin \gamma t 1(t)$	$\frac{\gamma}{p^2 + \gamma^2}$	0	$\frac{\gamma}{-\omega^2 + \gamma^2}$
$\cos \gamma t 1(t)$	$\frac{p}{p^2 + \gamma^2}$	0	$\frac{i\omega}{-\omega^2 + \gamma^2}$

* Примечание: c — постоянная.

Таблица 3.4

Изображения некоторых решетчатых сигналов по Лапласу и Фурье

$x[n, \varepsilon]$	$X(q, \varepsilon)$	σ_0	$X(i\beta, \varepsilon)$
$c * 1[n, \varepsilon]$	$\frac{c}{1 - e^{-q}}$	0	$\frac{c}{1 - e^{-i\beta}}$
$e^{-b(n+\varepsilon)} 1[n, \varepsilon]$ $b > 0$	$\frac{e^{-b\varepsilon}}{1 - e^{-b-q}}$	$-b$	$\frac{e^{-b\varepsilon}}{1 - e^{-b-i\beta}}$
$e^{b(n+\varepsilon)} 1[n, \varepsilon]$ $b > 0$	$\frac{e^{b\varepsilon}}{1 - e^{b-q}}$	b	—
$\sin \gamma(n + \varepsilon) 1[n, \varepsilon]$	$\frac{e^q \sin \gamma \varepsilon + \sin \gamma(1 - \varepsilon)}{e^q - 2 \cos \gamma + e^{-q}}$	0	$\frac{e^{i\beta} \sin \gamma \varepsilon + \sin \gamma(1 - \varepsilon)}{e^{i\beta} - 2 \cos \gamma + e^{-i\beta}}$
$\cos \gamma(n + \varepsilon) 1[n, \varepsilon]$	$\frac{e^q \cos \gamma \varepsilon - \cos \gamma(1 - \varepsilon)}{e^q - 2 \cos \gamma + e^{-q}}$	0	$\frac{e^{i\beta} \cos \gamma \varepsilon - \cos \gamma(1 - \varepsilon)}{e^{i\beta} - 2 \cos \gamma + e^{-i\beta}}$

* c — постоянная.

Примеры. 1. Сигналы $x_1(t)$ и $x_2(t)$ имеют изображения по Лапласу соответственно

$$\frac{1}{p^2 + 8p + 15} \text{ и } \frac{1}{p^2 + 7p}.$$

Изобразимы ли сигналы $x_1(t)$ и $x_2(t)$ по Фурье?

Изображение по Лапласу сигнала $x_1(t)$ имеет две особые точки: $p_1 = -5$ и $p_2 = -3$, в которых знаменатель обращается в нуль. Абсцисса абсолютной сходимости для $x_1(t)$ равная, следовательно, -3 , меньше нуля. Сигнал $x_1(t)$ имеет изображение по Фурье

$$X_1(i\omega) = \frac{1}{(i\omega)^2 + 8i\omega + 15}.$$

Изображение по Лапласу сигнала $x_2(t)$ имеет особые точки $p_1 = -7$ и $p_2 = 0$. Абсцисса абсолютной сходимости $\sigma_0 = 0$. Изображение по Фурье для сигнала $x_2(t)$ не существует, но для него существует несобственное преобразование Фурье

$$X_2(i\omega) = \frac{1}{(i\omega)^2 + i7\omega}.$$

2. Найти изображение по Лапласу сигналов: $x_1[n] = e^{\alpha n} 1[n]$ и $x_2[n] = \cos \gamma n 1[n]$.

Изображение первого из них при

$$|e^{-(q-\alpha)}| < 1$$

$$\begin{aligned} X_1(q) &= \sum_{n=0}^{\infty} e^{-qn} e^{\alpha n} = \sum_{n=0}^{\infty} e^{-(q-\alpha)n} = 1 + e^{-(q-\alpha)} + \\ &+ e^{-2(q-\alpha)} + e^{-3(q-\alpha)} + \dots = \frac{1}{1 - e^{-(q-\alpha)}}, \end{aligned}$$

$$X_1(q) = \frac{e^q}{e^q - e^{\alpha}}$$

при $\operatorname{Re}(q - \alpha) > 0$ или $\operatorname{Re} q > \operatorname{Re} \alpha$.

Для нахождения $X_2(q)$ воспользуемся формулой Эйлера, линейностью преобразования и полученным выше изображением $X_1(q)$:

$$\cos \gamma n = \frac{e^{i\gamma n} + e^{-i\gamma n}}{2},$$

$$X_2(q) = \frac{1}{2} \cdot \frac{e^q}{e^q - e^{i\gamma}} + \frac{1}{2} \cdot \frac{e^q}{e^q - e^{-i\gamma}} = \frac{e^q (e^q - \cos \gamma)}{e^{2q} - 2e^q \cos \gamma + 1}.$$

3. Найти изображение по Лапласу сигнала

$$x(t) = \sin \gamma t 1(t - T).$$

Обозначив $t - T = \tau$, имеем:

$$x(t) = x(\tau + T) = \cos \gamma T \sin \gamma \tau 1(\tau) + \sin \gamma T \cos \gamma \tau 1(\tau).$$

Следовательно, искомое изображение

$$X(p) = \gamma \frac{\cos \gamma T}{p^2 + \gamma^2} + \frac{\sin \gamma T}{p^2 + \gamma^2} p.$$

§ 3.4. СЛУЧАЙНЫЕ СИГНАЛЫ

До сих пор предполагалось, что описывающие функции $f(t)$ и $f[n]$ могут быть предвычислены для любых значений t и n , т. е. рассматривались *детерминированные сигналы*. Если $f(t)$ и $f[n]$ — случайные функции, то такое предвычисление невозможно и имеют место *случайные* или *стохастические сигналы*. При повторении процессов, происходящих в системе, случайные сигналы $x(t)$ и $x[n]$ могут иметь различные значения при одних и тех же значениях t или n , отсчитываемых каждый раз от начала процесса, т. е. получаются различные *реализации* одного и того же случайного сигнала.

Таким образом, если непрерывный сигнал $x(t)$ случаен, то $x(t_0)$ есть *случайная величина* в каждый фиксированный момент времени t_0 . Соответственно для случайного дискретного сигнала $x[n]$ $x[n_0]$ есть случайная величина при фиксированном n_0 . Случайная величина $x(t_0)$ или $x[n_0]$ характеризуется своей *функцией распределения вероятностей* $F(y) = P\{x < y\}$, где $P\{x < y\}$ — вероятность того, что случайная величина x принимает значения, меньшие y . Если для $F(y)$ существует функция $w(x)$, называемая *плотностью вероятностей*:

$$F(y) = \int_{-\infty}^y w(x) dx,$$

то случайная величина x называется *непрерывной случайной величиной*. Последнее имеет место, например, когда $F(y)$ непрерывно дифференцируема, тогда

$$w(y) = F'(y).$$

Функцию распределения вероятностей $F(y)$ часто называют *интегральной функцией распределения*; плотность вероятностей называют еще *дифференциальным распределением*, или просто *распределением*.

Среди случайных величин, не являющихся непрерывными, т. е. не имеющих плотностей вероятностей, выделяют класс *дискретных случайных величин*. Такова, например, величина, принимающая значения y_1, y_2, \dots, y_n соответственно с вероятностями P_1, P_2, \dots, P_n . Ее функция распределения имеет график, показанный на рис. 3.19, а.

Как легко видеть, обобщенная производная такой функции

$$DF(y) = P_1 \delta(y - y_1) + P_2 \delta(y - y_2) + \dots + P_n \delta(y - y_n) \quad (3.33)$$

равна линейной комбинации δ -функций, сосредоточенных в точках y_1, y_2, \dots, y_n с коэффициентами P_1, P_2, \dots, P_n . Эту производную естественно назвать плотностью вероятностей дискретной случайной величины. Она условно пред-

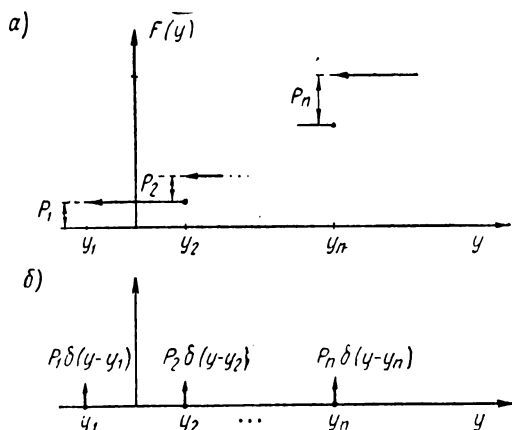


Рис. 3.19. Функции распределения дискретной случайной величины

ставлена на рис. 3.19, б. Дискретная случайная величина может иметь не только конечное, но и счетное множество значений $y_1, y_2, \dots, y_n, \dots$, принимаемых с вероятностями $P_1, P_2, \dots, P_n, \dots$, причем $\sum_{n=1}^{\infty} P_n = 1$. В этом случае функция плотности вероятностей

$$w(y) = \sum_{n=1}^{\infty} P_n \delta(y - y_n). \quad (3.34)$$

Случайные сигналы могут быть непрерывными или дискретными не только во времени, но также в вероятностном смысле. Если дискрета решетчатого случайного сигнала в различных реализациях может принимать любые значения, т. е. изменяться непрерывно, то такой сигнал.

дискретный во времени, является *непрерывным в вероятностном смысле*. Если же изменение дискретности от реализации к реализации происходит скачками, то этот сигнал *дискретен и в вероятностном смысле*. Подобно этому непрерывным или дискретным в вероятностном смысле может быть и непрерывный случайный сигнал.

Далее будем рассматривать непрерывные $x(t)$ и дискретные $x[n]$ во времени случайные сигналы, которые при фиксированном значении аргумента t или n являются непрерывной или дискретной (в вероятностном смысле) случайной величиной и, следовательно, имеют плотность вероятностей $w(x, t)$ или $w(x, n)$. Плотность вероятностей таких сигналов является функцией двух аргументов: x и t или x и n .

Важной характеристикой случайного сигнала является его *математическое ожидание*, иначе называемое *статистическим средним* или *средним по множеству*. Математическое ожидание непрерывного во времени сигнала

$$\overline{x(t)} = \int_{-\infty}^{\infty} x(t) w(x, t) dx. \quad (3.35)$$

Математическое ожидание дискретного во времени (решетчатого) сигнала

$$\overline{x[n]} = \int_{-\infty}^{\infty} x[n] w(x, n) dx. \quad (3.36)$$

Отклонение сигнала от математического ожидания называют *флюктуацией*. Математическое ожидание квадрата флюктуации или статистическое среднее этого квадрата есть *дисперсия*. Согласно этому определению дисперсия непрерывного сигнала

$$\begin{aligned} \sigma^2(t) &= \int_{-\infty}^{\infty} (x - \bar{x})^2 w(x, t) dx = \int_{-\infty}^{\infty} x^2 w(x, t) dx - \\ &- 2\bar{x} \int_{-\infty}^{\infty} x w(x, t) dx + \bar{x}^2 \int_{-\infty}^{\infty} w(x, t) dx. \end{aligned}$$

Замечая, что первый из трех последних интегралов равен математическому ожиданию квадрата сигнала, второй — математическому ожиданию сигнала, а третий — единице, получаем:

$$\sigma^2(t) = \overline{x^2(t)} - \{\overline{x(t)}\}^2, \quad (3.37)$$

т. е. дисперсия сигнала равна разности математического ожидания его квадрата и квадрата его математического ожидания. Для решетчатого сигнала аналогично

$$\sigma^2[n] = \overline{x^2[n]} - \{\overline{x[n]}\}^2. \quad (3.38)$$

Математическое ожидание сигнала в степени k называется моментом k -го порядка. Таким образом, в выражениях дисперсий (3.37) и (3.38) фигурируют моменты второго порядка.

Математическое ожидание произведения двух сигналов, один из которых сдвинут относительно другого во времени, называют *взаимной корреляционной функцией* этих сигналов. Математическое ожидание произведения двух значений одного и того же сигнала, сдвинутых на τ по оси времени в случае непрерывного сигнала или на ν по оси аргумента n в случае решетчатого сигнала, называют автокорреляционной, или *корреляционной функцией*. Случайный сигнал называют *стационарным*, если характеризующая его функция распределения не зависит от времени. Таковы, например, свойства записи качки корабля на ветровом волнении.

Математическое ожидание и дисперсия стационарных сигналов постоянны, т. е. не изменяются в течение времени существования сигнала, а корреляционные функции зависят соответственно только от τ или ν для непрерывных и решетчатых сигналов. По этим признакам нельзя, однако, утверждать, что сигналы стационарны в указанном выше смысле.

Стационарные сигналы с течением времени не могут ни затухать, ни стремиться к все большим значениям. В противных случаях имеют место *нестационарные сигналы*. Поэтому сигналы, ограниченные во времени, являются, строго говоря, нестационарными. Однако, если длительность сигнала (T_c или N_c) достаточно велика, он может с достаточной точностью удовлетворять условиям стационарности. Это и предполагается далее, когда рассматриваются только стационарные сигналы, или, что то же, стационарные случайные процессы изменения сигналов во времени.

Функция распределения всегда стремится к нулю как при уменьшении, так и при увеличении сигнала в конечном интервале. Поэтому к ней применимо преобразование Фурье. Это преобразование должно быть двусторонним,

так как сигнал, являющийся в данном случае аргументом оригинала, может принимать положительные и отрицательные значения. Таким образом, изображение функции распределения по Фурье, называемое *характеристической функцией*, имеет вид

$$W(i\xi) = \int_{-\infty}^{\infty} e^{-i\xi x} w(x) dx, \quad (3.39)$$

где ξ — аналог частоты, имеющий размерность, обратную размерности сигнала.

Для стационарных сигналов обычно справедлива *эргодическая гипотеза*, согласно которой средние значения по времени и множеству совпадают. Это означает, что исследование большого числа реализаций случайного сигнала

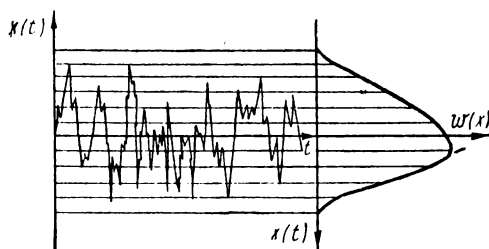


Рис. 3.20. Построение графика плотности вероятностей стационарных случайных сигналов

при одном и том же значении времени выявляет те же статистические свойства сигнала, что и исследование какой-либо одной его реализации на протяжении длительного времени.

Построение распределения по записи непрерывного во времени и в вероятностном смысле стационарного случайного сигнала, например, полученной на осциллограмме, может быть сделано следующим способом (рис. 3.20). Проводим линии, параллельные оси времени, через равные промежутки по оси x и подсчитываем число пересечений каждой линии с кривой записи сигнала. Это число откладываем как ординату кривой распределения. Масштаб ординат определяем из условия

$$\int_{-\infty}^{\infty} w(x) dx = 1, \quad (3.40)$$

приравнивая площадь, ограниченную кривой распределения и осью сигналов, единице.

Другой способ построения кривой распределения по реализации стационарного случайного сигнала показан на рис. 3.21. Промежуток от a до b , в котором принимает значения данная реализация $x(t)$, делится на n равных частей. Длина каждого частичного промежутка $\Delta x = \frac{b-a}{n}$. Приближенное значение ненормированной функции распределения $w_n(x)$ в i -ом частичном промежутке равно сумме длин промежутков оси времени $\Delta t_{i1} + \Delta t_{i2} + \Delta t_{i3}$, в которых реализация $x(t)$ принимает значения из этого

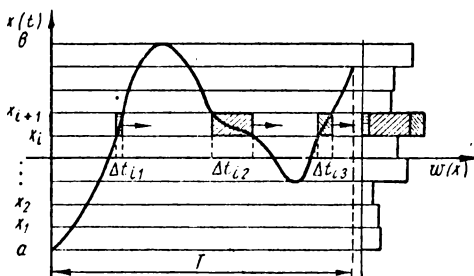


Рис. 3.21. К выводу формулы (3.41)

i -го промежутка оси абсцисс (i изменяется от 0 до $n-1$). Полученную ступенчатую кривую необходимо нормировать так, чтобы площадь под ней равнялась 1. Для этого все ординаты $w_n(x)$ нужно уменьшить в $\int_a^b w_n(x) dx$ раз. Этот интеграл

$$\begin{aligned} \int_a^b w_n(x) dx &= \sum_{i=0}^{n-1} \Delta x w_n(x_i) = \Delta x \sum_{i=0}^{n-1} w_n(x_i) = \\ &= \Delta x \sum_{i=0}^{n-1} (\Delta t_{i1} + \Delta t_{i2} + \Delta t_{i3}) = \Delta x \cdot T, \end{aligned}$$

где T — длительность реализации.

Итак, в точке x приближенное значение функции распределения ($x_i < x < x_{i+1}$)

$$w(x) \approx \frac{\Delta t_{i1} + \Delta t_{i2} + \Delta t_{i3}}{\Delta x T} = \frac{1}{T} \left(\frac{\Delta t_{i1}}{\Delta x} + \frac{\Delta t_{i2}}{\Delta x} + \frac{\Delta t_{i3}}{\Delta x} \right).$$

Делая предельный переход при $\Delta x \rightarrow 0$, получим:

$$\omega(x) = \frac{1}{T} \left(\frac{1}{x_1'} + \frac{1}{x_2'} + \dots + \frac{1}{x_k'} \right), \quad (3.41)$$

где x_j' ($j=1, 2, \dots, k$) есть значение производной по времени функции $x(t)$ на j -ом участке ее монотонности в момент, когда достигается значение x .

Статистические свойства стационарных случайных сигналов исследуют с помощью моментов первого и второго порядка и корреляционной функции. Выражения этих вероятностных характеристик непрерывного и решетчатого случайных сигналов, имеющих соответственно длительности T_c и N_c , приведены в табл. 3.5. Они получены на основании эргодической гипотезы, позволяющей приравнять средние по множеству средним по времени.

Из приведенной в таблице формулы момента k -го порядка при $k=1$ получим среднее по времени значение сигнала, а при $k=2$ — среднее по времени значение квадрата сигнала.

Корреляционная функция характеризует взаимозависимость значений сигнала, измеренных в моменты, отстоящие на τ или ν один от другого. Это легко уяснить из рассмотрения корреляционной функции $R_x[\nu]$. Чем ближе значения $x[n]$, отстоящие на ν друг от друга, и чем более сохраняется их близость при изменении n , тем больше $R_x[\nu]$. А это указывает на слабое изменение сигнала в зависимости от n , т. е. на взаимообусловленность его значений, разделенных тем или иным интервалом ν . Интервалы τ и ν могут быть как положительными, так и отрицательными, причем ν принимает лишь целочисленные значения.

Важнейшие свойства корреляционных функций указаны в табл. 3.5. Первое свойство непосредственно следует из выражений $R_x(\tau)$ и $R_x[\nu]$: значение корреляционной функции при $\tau=0$ (или $\nu=0$) равно моменту второго порядка и поэтому положительно. С помощью формул (3.37) и (3.38) оно выражается через математическое ожидание и дисперсию. В наличии второго свойства — свойства четности — легко убедиться с помощью тех же выражений. Третье свойство вытекает из уже рассмотренного физического смысла корреляционной функции; поэтому не будем приводить известного математического доказательства этого

Характеристики стационарных случайных сигналов

Сигнал	$x(t) = f(t) [1(t) - 1(t - T_c)]$	$x[n] = f[n] \{1[n] - 1[n - N_c - 1]\}$
Момент k -го порядка	$\overline{x^k(t)} = \lim_{T_c \rightarrow \infty} \frac{1}{T_c} \int_0^{T_c} x^k(t) dt$	$x^k[n] = \lim_{N_c \rightarrow \infty} \frac{1}{N_c + 1} \sum_{n=0}^{N_c} x^k[n]$
Корреляционная функция	$R_x(\tau) = \lim_{T_c \rightarrow \infty} \frac{1}{T_c} \int_0^{T_c} x(t) x(t - \tau) dt$	$R_x[\nu] = \lim_{N_c \rightarrow \infty} \frac{1}{N_c + 1} \sum_{n=0}^{N_c} x[n] x[n + \nu]$
Первое	$R_x(0) = \overline{x^2(t)} = \overline{\{x(t)\}^2} + \sigma^2$ $R_x(0) > 0$	$R_x[0] = \overline{x^2[n]} = \overline{\{x[n]\}^2} + \sigma^2$ $R_x[0] > 0$
Второе	$R_x(\tau) = R_x(-\tau)$	$R_x[\nu] = R_x[-\nu]$
Третье	$R_x(\tau) \leq R_x(0)$	$R_x[\nu] \leq R_x[0]$
Четвертое	При $T_c \rightarrow \infty$ $\lim_{\tau \rightarrow \infty} R_x(\tau) = \overline{\{x(t)\}^2}$ При $T_c < \infty$ $R_x(\tau) \parallel_{\tau > T_c} = 0$ $R_x(\tau) \parallel_{\tau > T_c} = 0$	При $N_c \rightarrow \infty$ $\lim_{\nu \rightarrow \infty} R_x[\nu] = \overline{\{x[n]\}^2}$ При $N_c < \infty$ $R_x[\nu] \parallel_{\nu > N_c} = 0$

Свойства корреляционной функции

свойства [3.17]. Для доказательства четвертого свойства заметим [3.11], что для сигналов неограниченной длительности при неограниченном увеличении τ и ν зависимость между значениями сигналов в моменты t и $t + \tau$ или n и $n + \nu$ ослабевает, так что в пределе эти значения становятся независимыми. А среднее значение произведения независимых случайных величин равно произведению средних значений сомножителей. Среднее же значение стационарного сигнала не зависит от времени, откуда и следуют предельные значения корреляционных функций,

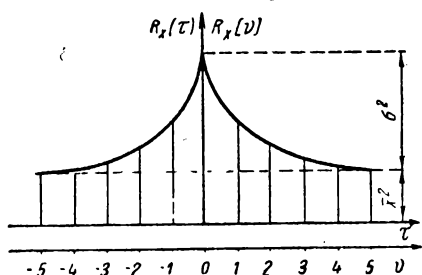


Рис. 3.22. Примерные графики корреляционных функций непрерывного и решетчатого случайных сигналов

указанные в табл. 3.5 при $T_c \rightarrow \infty$ и $N_c \rightarrow \infty$. В этих случаях корреляционные функции могут быть затухающими лишь у сигналов, математическое ожидание которых равно нулю. Но у сигналов ограниченной длительности корреляционные функции затухают, так как они равны нулю при $\tau > T_c$ и $\nu > N_c$.

Типичные графики корреляционных функций показаны на рис. 3.22. Убывание с ростом аргумента может быть, однако, немонотонным. Если сигнал имеет неслучайную составляющую, то построение корреляционной функции позволяет выявить эту составляющую по нарушению четвертого свойства.

Построение корреляционной функции по записи процесса облегчается применением приборов, называемых корреляторами, коррелометрами и коррелографами [3.7].

Рассмотрим непрерывный во времени сигнал конечной длительности T_c . Если такой сигнал удовлетворяет условиям, необходимым для того, чтобы он был оригиналом (§ 3.2), то к нему применимо преобразование Фурье, так как для него соблюдается условие (3.30). Корреляционная функция такого сигнала имеет выражение, приведенное в табл. 3.5 [3.13]. Вид этой корреляционной функции позволяет применить к ней преобразование Фурье. Состав-

вим двустороннее преобразование

$$R(i\omega) = \int_{-\infty}^{\infty} e^{-i\omega\tau} R_x(\tau) d\tau. \quad (3.42)$$

Подставляя сюда выражение $R_x(\tau)$, находим, что

$$R(i\omega) = \frac{1}{T_c} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-i\omega\tau} x(t) x(t+\tau) d\tau dt.$$

Вводя

$$\lambda = t + \tau,$$

следует положить, что

$$d\tau = d\lambda,$$

так как $t = \text{const}$ при интегрировании по τ . Поэтому

$$\begin{aligned} R(i\omega) &= \frac{1}{T_c} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-i\omega\lambda} x(\lambda) e^{i\omega t} x(t) dt d\lambda = \\ &= \frac{1}{T_c} \int_{-\infty}^{\infty} e^{-i\omega\lambda} x(\lambda) d\lambda \int_{-\infty}^{\infty} e^{i\omega t} x(t) dt = \frac{1}{T_c} X(i\omega) \cdot X(-i\omega). \end{aligned}$$

Пользуясь табл. 3.1, имеем:

$$R(i\omega) = \frac{1}{T_c} [R(\omega) + iI(\omega)] [R(-\omega) + iI(-\omega)],$$

откуда в силу четности вещественных и нечетности мнимых частей

$$\begin{aligned} R(i\omega) &= \frac{1}{T_c} [R(\omega) + iI(\omega)] [R(\omega) - iI(\omega)] = \\ &= \frac{1}{T_c} [R^2(\omega) + I^2(\omega)]. \end{aligned}$$

Следовательно,

$$R(i\omega) = \frac{1}{T_c} |X(i\omega)|^2, \quad (3.43)$$

т. е. изображение корреляционной функции равно квадрату модуля изображения сигнала, деленному на длительность сигнала. Таким образом, изображение корреляционной функции по Фурье — вещественная функция частоты ω .

Выражение (3.43) справедливо при любой длительности T_c , от которой не зависит его левая часть. Поэтому

$$R(i\omega) = \lim_{T_c \rightarrow \infty} \frac{1}{T_c} |X(i\omega)|^2. \quad (3.44)$$

Если $[x]$ — размерность сигнала, а $[t]$ — размерность времени, то вещественная и мнимая части, а также модуль изображения сигнала по Фурье имеют размерность $[x][t]$, величина же, стоящая в правой части (3.44), — размерность $[x]^2[t]$. Правую часть (3.44) называют *спектральной плотностью сигнала*

$$s(\omega) = \lim_{T_c \rightarrow \infty} \frac{1}{T_c} |X(i\omega)|^2. \quad (3.45)$$

Как видно, это — вещественная функция частоты ω . Когда мощность сигнала пропорциональна x^2 (как в случае, например, когда x — электрическое напряжение), спектральная плотность пропорциональна производной мощности по частоте, т. е. описывает распределение мощности в спектре частот: произведение $s(\omega) d\omega$ пропорционально мощности в спектральном интервале $d\omega$.

Из выражений (3.44) и (3.45) следует, что

$$s(\omega) = R(i\omega), \quad (3.46)$$

т. е. спектральная плотность непрерывного сигнала равна изображению его корреляционной функции по Фурье. Отсюда посредством обратного преобразования Фурье получаем *формулу Хинчина* [3.2]

$$R_x(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{i\omega\tau} s(\omega) d\omega, \quad (3.47)$$

т. е. корреляционная функция выражается интегралом Фурье от спектральной плотности.

Из (3.46) и (3.42) имеем:

$$\begin{aligned} s(\omega) &= \int_{-\infty}^{\infty} e^{-i\omega\tau} R_x(\tau) d\tau = \\ &= \int_{-\infty}^{\infty} R_x(\tau) \cos \omega\tau d\tau - i \int_{-\infty}^{\infty} R_x(\tau) \sin \omega\tau d\tau, \end{aligned}$$

откуда в силу четности корреляционной функции

$$s(\omega) = 2 \int_0^{\infty} R_x(\tau) \cos \omega \tau d\tau. \quad (3.48)$$

Отсюда следует, что спектральная плотность — четная функция частоты. Поэтому из (3.47) таким же путем находим:

$$R_x(\tau) = \frac{1}{\pi} \int_0^{\infty} s(\omega) \cos \omega \tau d\omega. \quad (3.49)$$

Формулы (3.48) и (3.49) также получил А. Я. Хинчин [3.2].

Полагая в (3.49) $\tau = 0$, находим выражения среднего квадрата сигнала через спектральную плотность:

$$\bar{x}^2 = \frac{1}{\pi} \int_0^{\infty} s(\omega) d\omega. \quad (3.50)$$

Это — одна из важнейших характеристик случайного сигнала.

Для стационарного случайного решетчатого сигнала нетрудно вывести аналогичные предыдущим зависимости между спектральной плотностью $s(\beta)$ и корреляционной функцией $R_x[\nu]$ [3.13]:

$$s(\beta) = \sum_{\nu=0}^{\infty} 2[\nu] R_x[\nu] \cos \beta \nu, \quad (3.51)$$

где

$$2[\nu] = \begin{cases} 1 & \nu = 0 \\ 2 & \nu \neq 0 \end{cases}; \quad (3.52)$$

$$R_x[\nu] = \frac{1}{\pi} \int_0^{\pi} s(\beta) \cos \beta \nu d\beta. \quad (3.53)$$

Полагая $\nu = 0$, из (3.53) находим по первому свойству корреляционной функции (см. табл. 3.5) выражение среднего значения квадрата решетчатого сигнала через спектральную плотность:

$$\overline{x^2[n]} = \frac{1}{\pi} \int_0^{\pi} s(\beta) d\beta. \quad (3.54)$$

Рассмотрим пример применения формулы (3.51) в случае, когда корреляционная функция задана в виде

$$R_x[\nu] = \begin{cases} 1 & \nu = 0 \\ 0 & \nu \neq 0 \end{cases}$$

(в единицах измерения квадрата сигнала). Тогда

$$s(\beta) = 2 [0] R_x[0] \cos 0 = 1,$$

т. е. спектральная плотность постоянна и равна единице квадрата сигнала при всех значениях относительной частоты (рис. 3.23).

Постоянная спектральная плотность характеризует случайные сигналы, имеющие спектр, называемый *белым*. Такой спектр могут иметь как решетчатые, так и непрерывные сигналы. В связи со сказанным выше о физическом смысле спектральной плотности непрерывного сигнала, легко видеть, что в этом случае мощность сигнала

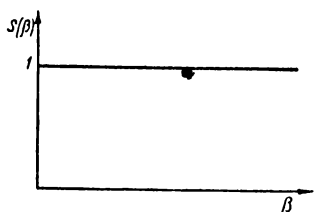


Рис. 3.23. Белый спектр

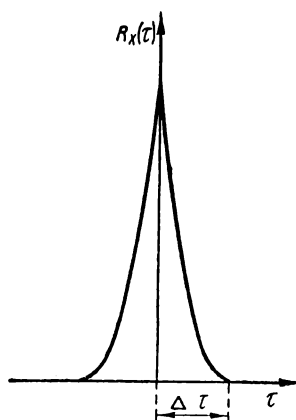


Рис. 3.24. Корреляционная функция непрерывного случайного сигнала с белым спектром

равномерно распределена по спектру частот. Для решетчатых сигналов спектральная плотность пропорциональна, аналогично, производной мощности сигнала по относительной (безразмерной) частоте, так что белый спектр имеет тот же смысл. В случае непрерывных сигналов белый спектр соответствует корреляционной функции, падающей до нуля за очень короткий промежуток времени $\Delta\tau$ (рис. 3.24). Это означает, как и в случае решетчатых сигналов, слабую корреляцию. Вообще, чем шире график спектральной плотности, тем уже график корреляционной функции, и наоборот.

Рассмотрим теперь информацию, получаемую от случайных сигналов. Информацию от решетчатого случайного сигнала, дискретного в вероятностном смысле, легко определить непосредственно по (1.32):

$$I = \sum_{n=0}^N \left(\sum_i P_{ni} \log P_{ni} - \sum_i P_{0ni} \log P_{0ni} \right), \quad (3.55)$$

где P_{0ni} и P_{ni} — априорная и апостериорная вероятности i -го значения n -й дискреты.

Эта формула применима для сигналов, обладающих следующим свойством: тот факт, что при $n = n_0$ дискрета приняла определенное значение, не изменяет функции распределения следующих дискрет.

Пользуясь (3.55), можно вывести и выражение информации от решетчатого случайного сигнала x , непрерывного в вероятностном смысле. Разобьем интервал изменения каждой дискреты на равные участки Δx , одинаковые для всех дискрет. Тогда

$$P_{0ni} = w_0(x_i, n) \Delta x,$$

$$P_{ni} = w(x_i, n) \Delta x,$$

если $w_0(x, n)$ и $w(x, n)$ — распределения, характеризующие n -ю дискрету до и после приема. Подставив последние выражения в (3.55), найдем искомую информацию предельным переходом при $\Delta x \rightarrow 0$:

$$\begin{aligned} I &= \lim_{\Delta x \rightarrow 0} \sum_{n=0}^N \left\{ \sum_i w(x_i, n) \Delta x \log_2 [w(x_i, n) \Delta x] - \right. \\ &\quad \left. - \sum_i w_0(x_i, n) \Delta x \log_2 [w_0(x_i, n) \Delta x] \right\} = \\ &= \sum_{n=0}^N \lim_{\Delta x \rightarrow 0} \left\{ \sum_i w(x_i, n) \Delta x \log_2 w_0(x_i, n) - \right. \\ &\quad \left. - \sum_i w_0(x_i, n) \Delta x \log_2 w(x_i, n) + \right. \\ &\quad \left. + \left[\sum_i w(x_i, n) \Delta x - \sum_i w_0(x_i, n) \Delta x \right] \log_2 \Delta x \right\}. \end{aligned}$$

Отсюда, учитывая, что

$$\lim_{\Delta x \rightarrow 0} \sum_i w(x_i, n) \Delta x = \int_{-\infty}^{\infty} w(x, n) dx = 1,$$

и раскрывая неопределенность, получаем:

$$I = \sum_{n=0}^N \left[\int_{-\infty}^{\infty} w(x, n) \log_2 w(x, n) dx - \int_{-\infty}^{\infty} w_0(x, n) \log_2 w(x, n) dx \right]. \quad (3.56)$$

Последняя формула справедлива для сигналов, обладающих тем свойством, что появление дискреты опреде-

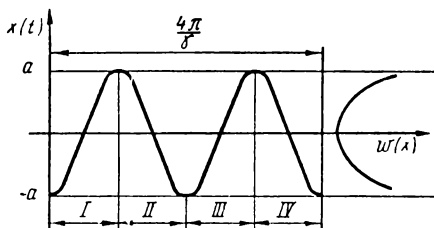


Рис. 3.25. К построению кривой распределения вероятностей стационарного случайного сигнала по реализации

$$x(t) = -a \cos \gamma t [1(t) - 1(t - T)]$$

ленного значения не изменяет функцию распределения следующих дискрет.

Примеры. 1. Пусть реализация стационарного случайного сигнала имеет вид (рис. 3.25)

$$x(t) = -a \cos \gamma t \left[1(t) - 1\left(t - \frac{4\pi}{\gamma}\right) \right].$$

Построим функцию распределения случайного сигнала по этой реализации. В записи реализации можно выделить четыре участка монотонности. В соответствии с (3.41)

$$w(x) = \frac{\gamma}{4\pi} \left(\frac{1}{x'_I} + \frac{1}{x'_{II}} + \frac{1}{x'_{III}} + \frac{1}{x'_{IV}} \right) = \frac{\gamma}{\pi} \frac{1}{x'_I},$$

так как в данном случае

$$x'_I = x'_{II} = x'_{III} = x'_{IV},$$

где x'_I — есть значение производной $x(t)$ в момент времени t из первого участка монотонности, в который реализация принимает значение x :

$$x'_I = \gamma a \sin \gamma t = \gamma a \sqrt{1 - \cos^2 \gamma t} = \gamma \sqrt{a^2 - x^2}.$$

Следовательно,

$$w(x) = \begin{cases} \frac{1}{\pi \sqrt{a^2 - x^2}} & -a \leq x \leq a \\ 0 & |x| > a. \end{cases}$$

График $w(x)$ приведен на рис. 3.25.

2. Вычислим корреляционную функцию для сигнала

$$x(t) = -\cos \gamma t \, 1(t).$$

Имеем:

$$R_x(\tau) = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T \cos \gamma(t - \tau) \cos \gamma t \, dt.$$

Принимая во внимание, что

$$\cos \gamma(t - \tau) \cos \gamma t = \frac{\cos \gamma \tau + \cos(2t - \tau) \gamma}{2},$$

получаем:

$$R_x(\tau) = \lim_{T \rightarrow \infty} \frac{1}{2T} \left(\cos \gamma \tau \int_0^T dt + \frac{1}{2\gamma} \sin \gamma(2t - \tau) \Big|_0^T \right) = \frac{1}{2} \cos \gamma \tau.$$

Будем считать рассматриваемый здесь сигнал реализацией стационарного случайного процесса. Тогда, как легко показать, функция распределения этого процесса будет такой же, что и в примере 1:

$$w(x) = \begin{cases} \frac{1}{\pi \sqrt{1 - x^2}} & -1 \leq x \leq 1 \\ 0 & |x| > 1. \end{cases}$$

Вычислим дисперсию этого случайного процесса, принимая во внимание, что для него имеет место очевидное равенство

$$\overline{x(t)} = 0.$$

Тогда выражение (3.37) примет вид

$$\sigma^2 = \int_{-\infty}^{\infty} x^2 w(x) \, dx,$$

откуда, после подстановки $w(x)$,

$$\sigma^2 = \frac{1}{\pi} \int_{-1}^1 \frac{x^2}{\sqrt{1 - x^2}} \, dx.$$

Делая замену переменной $x = \sin \vartheta$, получим:

$$\begin{aligned}\sigma^2 &= \frac{1}{\pi} \int_{-\pi/2}^{\pi/2} \frac{\sin^2 \vartheta \cos \vartheta}{\sqrt{1 - \sin^2 \vartheta}} d\vartheta = \frac{1}{\pi} \int_{-\pi/2}^{\pi/2} \sin^2 \vartheta d\vartheta = \\ &= \frac{1}{\pi} \int_{-\pi/2}^{\pi/2} \frac{1 - \cos 2\vartheta}{2} d\vartheta = \frac{1}{\pi} \left(\frac{1}{2} \vartheta \int_{-\pi/2}^{\pi/2} -\frac{1}{4} \sin 2\vartheta \int_{-\pi/2}^{\pi/2} \right) = \\ &= \frac{1}{\pi} \left(\frac{\pi}{2} - 0 \right) = \frac{1}{2}.\end{aligned}$$

Тот же результат можно получить, пользуясь соотношением

$$\sigma^2 = R_x(0),$$

которое при $\overline{x(t)} = 0$ следует из (3.37) и первого свойства корреляционной функции.

3. Для случайного сигнала $x[n]$ функция плотности вероятностей имеет вид

$$w(x, n) = P_1 \delta(x - x_1) + P_2 \delta(x - x_2),$$

причем $P_1 + P_2 = 1$; $P_1, P_2 \geq 0$.

Найдем математическое ожидание этого сигнала. Так как $w(x, n)$ не зависит от n , то сигнал $x[n]$ — стационарный и его математическое ожидание — константа. Согласно (3.36)

$$\begin{aligned}\bar{x}[n] &= \int_{-\infty}^{\infty} x \hat{w}(x) dx = \int_{-\infty}^{\infty} [x P_1 \delta(x - x_1) + x P_2 \delta(x - x_2)] dx = \\ &= P_1 \int_{-\infty}^{\infty} x \delta(x - x_1) dx + P_2 \int_{-\infty}^{\infty} x \delta(x - x_2) dx = P_1 x_1 + P_2 x_2.\end{aligned}$$

§ 3.5. КВАНТОВАНИЕ СИГНАЛОВ

Квантование означает преобразование сигнала, непрерывного в широком интервале времени, в кусочно-непрерывный, обладающий непрерывностью лишь в меньших интервалах времени, т. е. дискретный или дискретно-непрерывный. Квантование может быть произведено *по уровню*, *времени* или *уровню и времени одновременно* (рис. 3.26). В первом случае сигнал сохраняет некоторое значение, пока он не изменится (при непрерывном изменении) на Δx (сплошная линия на рисунке). Во втором — пока не пройдет время Δt , после чего он принимает новое значение, совпадающее с полученным к тому времени при непрерывном изменении (на рисунке пунктиром). Квантование по времени, в отличие от указанного, может не

сопровождаться запоминанием значения непрерывного сигнала на весь интервал времени Δt . При этом сигнал становится импульсивным. В третьем случае сигнал также принимает значения, отличающиеся на целое число шагов Δx , ближайшие к тем, какие он имеет при непрерывном изменении, через промежутки времени Δt , сохраняя их в течение этих промежутков (точечная линия на рисунке).

Величины Δx и Δt называют *шагами* (или интервалами) квантования. Их выбор должен быть подчинен некоторым условиям.

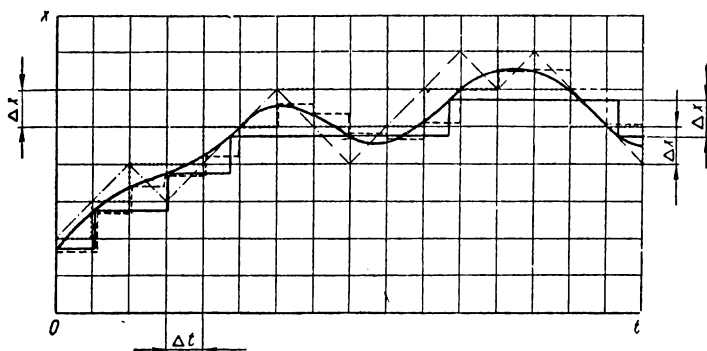


Рис. 3.26. Четыре способа квантования

Рассмотрим квантование по времени сигнала $x(t)$, имеющего изображение $X(i\omega)$ по Фурье. Практически спектр сигнала всегда ограничен некоторой *полосой пропускания*, характерной для системы, через которую сигнал проходит. Положим, что в пределах такой полосы частота изменяется от $-\omega_c$ до ω_c , так что ширина полосы равна удвоенной *частоте среза* ω_c . Так как спектр не существует вне этой полосы, рассматриваемый сигнал может быть выражен с помощью интеграла Фурье в следующей форме:

$$x(t) = \frac{1}{2\pi} \int_{-\omega_c}^{\omega_c} e^{it\omega} X(i\omega) d\omega.$$

Отсюда, положив

$$t = \frac{\pi}{\omega_c} n, \quad (3.57)$$

где $\frac{\pi}{\omega_c}$ — период, определяемый шириной полосы пропускания, имеем:

$$x\left(\frac{\pi}{\omega_c} n\right) = \frac{1}{2\pi} \int_{-\omega_c}^{\omega_c} e^{i\frac{\pi}{\omega_c} n\omega} X(i\omega) d\omega.$$

Интеграл, входящий в это выражение, представляет собой n -й коэффициент разложения функции $X(i\omega)$ по гармоникам с периодом, равным $\frac{\pi}{\omega_c}$. Таким образом [3.14], значения, которые сигнал $x(t)$ принимает в моменты времени (3.57), определяют коэффициенты Фурье в разложении его комплексного спектра, а стало быть и самый спектр. Но сигнал полностью определяется его спектром. Следовательно, сигнал полностью определяется своими значениями $x\left(\frac{\pi}{\omega_c} n\right)$. Поэтому интервал квантования по времени должен быть таким, чтобы соседние значения получаемого при этом дискретного сигнала отстояли друг от друга не более чем на время t по (3.57). Это будет иметь место, если интервал квантования

$$\Delta t \leq \frac{\pi}{\omega_c}. \quad (3.58)$$

Тогда квантование по времени не будет сопровождаться потерей информации, зависящей от спектра сигнала. Условие (3.58) выражает *теорему Котельникова* [3.1]. Оно следует также из более общего положения, согласно которому сигнал, имеющий длительность T_c , полностью определяется совокупностью его значений, число которых равно $\frac{\omega_c}{\pi} T_c$, причем они могут быть и неравноотстоящими [3.14].

В случаях постоянного интервала Δt удобно рассматривать определяемую им *частоту квантования*

$$\omega_k = \frac{2\pi}{\Delta t}.$$

Тогда условие (3.58) принимает вид

$$\omega_k \geq 2\omega_c. \quad (3.59)$$

Пусть модуль $|X(i\omega)|$ комплексного спектра квантуемого сигнала зависит от частоты (сплошная линия на рис. 3.27).

Условие (3.59), обеспечивающее сохранение информации, зависящей от спектра, показывает, что в спектре квантованного сигнала имеет место повторение спектра квантуе-

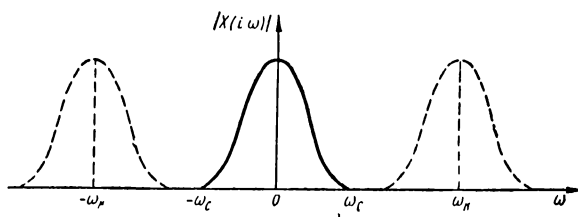


Рис. 3. 27. Спектр квантованного сигнала

мого сигнала с частотой ω_k (пунктиром на рис. 3.27). Это можно доказать и непосредственно [3.10]. При

$$\omega_k < 2\omega_c$$

повторяющиеся спектры налагаются друг на друга, вследствие чего спектры квантованного и квантуемого сигналов становятся различными.

Такое же наложение имеет место при $\omega_c \rightarrow \infty$, когда спектр квантуемого сигнала не ограничен. Если, однако, при этом существует полоса частот, за пределами которой значения модуля $|X(i\omega)|$ пренебрежимо малы, то ширина этой полосы определяет зна-

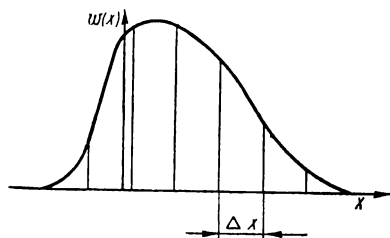


Рис. 3.28. Квантование функции распределения

чение ω_c , которым можно воспользоваться, чтобы применить теорему Котельникова. Эта теорема приближенно справедлива, таким образом, в случае неограниченного спектра, который имеют сигналы, ограниченные во времени.

Квантование сигнала по уровню удобно рассматривать как задачу квантования случайного непрерывного сигнала x , имеющего плотность вероятности $\omega(x)$. В результате квантования сигнала с интервалом Δx значения $\omega(x)$ будут отличаться от нуля лишь при $x = n\Delta x$ (рис. 3.28), где

записанные с помощью дельта-функций они имеют вид

$$\delta(x - n\Delta x) \int_{n\Delta x - \frac{1}{2}\Delta x}^{n\Delta x + \frac{1}{2}\Delta x} w(x) dx.$$

Следовательно, квантование сигнала по уровню преобразует непрерывную функцию $w(x)$ в импульсивную аналогично квантованию по времени [3.5]. Поэтому, заменяя время сигналом, а сигнал — функцией $w(x)$ в приведенном выше исследовании квантования по времени, можно получить для квантования по уровню результат, аналогичный теореме Котельникова. При этом спектр квантуемого сигнала заменяет характеристическая функция $W(i\xi)$ (3.39). Если ее модуль пренебрежимо мал при $|\xi| > \xi_c$, то квантование не сопровождается потерей информации, зависящей от характеристической функции сигнала, т. е. от его вероятностных свойств, когда

$$\Delta x \leq \frac{\pi}{\xi_c}. \quad (3.60)$$

Интервалы Δt и Δx могут быть выбраны в пределах, определяемых условиями (3.31) и (3.60), применительно к той или иной задаче квантования. Такие задачи возникают, например, при воздействии непрерывных сигналов на цифровые элементы АУС. Вводя дополнительные условия, связанные с решаемой задачей квантования, можно искать оптимальные значения Δt и Δx , которые однако должны удовлетворять указанным условиям.

Квантование по времени и уровню возможно также посредством линейного изменения квантованного сигнала в пределах каждого шага квантования по времени с угловым коэффициентом, равным ± 1 . На рис. 3.26 такой способ квантования сигнала показан пунктиром с точкой [3.9].

Пример. Найдём условия квантования по уровню для сигнала

$$x(t) = at [1(t) - 1(t - T)] \quad (3.61)$$

(рис. 3.29). В этом случае производная описывающей функции $f'(t) = a$. Рассматривая данный сигнал как одну из реализаций ста-

ционного случайного процесса, воспользуемся формулой (3.41)

$$w(x) = \begin{cases} 0 & x < 0 \\ \frac{1}{aT} & 0 \leq x \leq aT \\ 0 & x > aT \end{cases}$$

Таким образом,

$$w(x) = \frac{1}{aT} [1(x) - 1(x - aT)]$$

(рис. 3.30).

Функция $w(x)$ позволяет вычислить характеристическую функцию по формуле (3.39), заменяя бесконечные пределы граничными

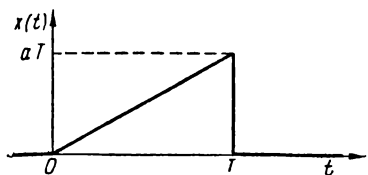


Рис. 3.29. Сигнал (3.61)

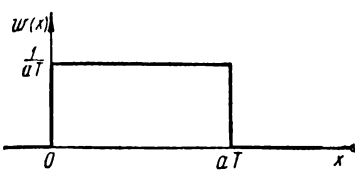


Рис. 3.30. Плотность вероятности сигнала, представленного на рис. 3.29

значениями сигнала, при которых плотность вероятности отличается от нуля. Поэтому

$$W(i\xi) = \frac{1}{aT} \int_0^{aT} e^{-i\xi x} dx = \frac{i}{aT\xi} (e^{-iaT\xi} - 1).$$

Отсюда находим вещественную и мнимую части характеристической функции:

$$\begin{aligned} \operatorname{Re} W(i\xi) &= \frac{\sin aT\xi}{aT\xi}, \\ \operatorname{Im} W(i\xi) &= \frac{\cos aT\xi - 1}{aT\xi}. \end{aligned}$$

Легко убедиться, что наибольшие экстремальные значения вещественной и мнимой частей равны:

$$\begin{aligned} \operatorname{Re} W(i0) &= 1, \\ \operatorname{Im} W\left(i \frac{\pi}{2aT}\right) &= -\frac{2}{\pi}. \end{aligned}$$

Поэтому примем, что при $\xi = \xi_c$ амплитуда вещественной части имеет абсолютное значение

$$\frac{1}{aT\xi_c} = \epsilon \ll 1.$$

Тогда

$$\xi_c = \frac{1}{aT\epsilon}$$

и по (3.60) интервал квантования

$$\Delta x \leq \pi aT\epsilon.$$

УРАВНЕНИЯ ДВИЖЕНИЯ И ДИНАМИЧЕСКИЕ ХАРАКТЕРИСТИКИ АВТОМАТИЧЕСКИХ СИСТЕМ

§ 4.1. УРАВНЕНИЯ ДВИЖЕНИЯ НЕПРЕРЫВНЫХ СИСТЕМ

Свойства автоматической системы определяются ее *параметрами*, т. е. физическими величинами, характеризующими ее элементы. Например, контур RC характеризуют электрическое сопротивление и электрическая емкость, пружину — ее жесткость; сопротивление, емкость и жесткость являются параметрами системы, в состав которой входят эти элементы. Параметры далее предполагаются сосредоточенными в некоторых пунктах системы, которая рассматривается как система с *сосредоточенными параметрами*.

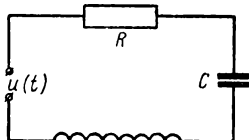


Рис. 4.1. Электрический контур с сосредоточенными параметрами

Пользуясь физическими законами, которыми описываются явления, происходящие в рассматриваемой системе, можно составить уравнение, связывающее входной и

выходной сигналы, их производные и интегралы через параметры системы. Такое уравнение называется *уравнением движения* этой системы, понимая движение, как изменение выходного сигнала во времени. Согласно выясненному в предыдущем параграфе производные и интегралы реальных сигналов должны быть обобщенными. Поэтому уравнение движения автоматической системы, написанное относительно выходного сигнала, должно быть *уравнением в обобщенных производных*.

Рассмотрим составление уравнения движения, взяв для примера электрический контур, в который последовательно включены сопротивление, конденсатор и катушка индуктивности (рис. 4.1). Пусть на контакты контура подается напряжение

$$u(t) = f_u(t) l(t), \quad (4.1)$$

которое будем рассматривать в качестве входного сигнала. В качестве выходного сигнала может быть взята любая величина, характеризующая процесс, вызываемый в контуре напряжением $u(t)$: ток, заряд конденсатора, напряжение на любом из трех элементов контура. Предположим, что нас интересует зависимость тока от времени. Так как ток появляется при включении напряжения (4.1), его зависимость от времени должна иметь вид

$$i(t) = f_i(t) 1(t),$$

где $f_i(t)$ — описывающая функция тока.

Заряд конденсатора, также появляющийся при включении напряжения (4.1), должен быть записан как

$$q(t) = f_q(t) 1(t),$$

где $f_q(t)$ — описывающая функция заряда.

Поэтому ток необходимо рассматривать как обобщенную производную заряда

$$i(t) = Dq. \quad (4.2)$$

Выражение заряда через ток запишем в форме обобщенного интеграла

$$q(t) = D^{-1}i. \quad (4.3)$$

Если C — емкость конденсатора, то напряжение на нем равно

$$\frac{1}{C} D^{-1}i.$$

Обозначая через $\Psi(i)$ потокосцепление катушки индуктивности, зависящее от тока, найдем напряжение на ней:

$$D\Psi = \Psi'(i) Di,$$

где $\Psi'(i) = \frac{d\Psi}{di}$ — обычная производная потокосцепления по току.

Применяя теперь второй закон Кирхгофа, при сопротивлении, равном R , имеем:

$$\Psi'(i) Di + Ri + \frac{1}{C} D^{-1}i = u(t). \quad (4.4)$$

Это есть уравнение движения рассматриваемого контура в обобщенных производных. Полученное уравне-

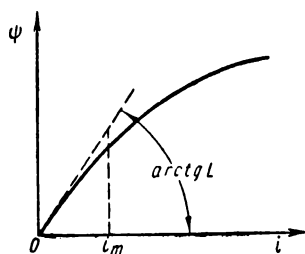


Рис. 4.2. К линейзации уравнения движения контура, показанного на рис. 4.1

ние (4.4) может быть нелинейным, если коэффициент при производной Di зависит от тока. В таком случае рассматриваемый контур является нелинейной системой. Вместо реальной нелинейной системы можно рассматривать ее *линейную модель*, описываемую *линейным уравнением движения (уравнением линейного приближения)*, если *нелинейное уравнение поддается линейзации*. В данном случае *линеаризация* обычно достигается заменой кривой $\Psi(i)$ касательной при $t=0$ (рис. 4.2). Это может быть допустимо, если ток не превышает некоторого значения i_m (рис. 4.2).

В этом случае

$$\Psi'(i)|_{i=0} \approx L,$$

где L — индуктивность цепи.

Тогда (4.4) преобразуется в *линейное уравнение в обобщенных производных с постоянными коэффициентами*

$$LDi + Ri + \frac{1}{C} D^{-1}i = u(t). \quad (4.5)$$

Отсюда видно, что уравнение движения является линейным с постоянными коэффициентами, если постоянны параметры системы.

В общем случае уравнение движения может содержать члены интегрального вида, как это имеет место в (4.4) и (4.5) (последние члены в левой части). Следует заметить, что подходящей подстановкой, т. е. переходом к другому выходному сигналу, от таких членов можно избавляться. Так, в рассматриваемом примере, пользуясь соотношениями (4.2) и (4.3), можно получить уравнение движения, составленное не относительно тока, а относительно заряда конденсатора. В таком уравнении уже не будет интегрального члена; в него будут входить (в левую часть) лишь члены, содержащие заряд и его первую и вторую обобщенные производные. Далее рассматриваются только уравнения такого вида (без интегральных членов).

Уравнение в обобщенных производных не является дифференциальным уравнением в обычном смысле, так как его члены могут быть разрывными, а также могут иметь угловые точки и содержать импульсы. Другое и важнейшее отличие этого уравнения от обычного дифференциального уравнения состоит в том, что его решение единственное. Это следует из доказанной выше однозначности операции нахождения функции по ее обобщенной производной. Таким образом, для нахождения выходного сигнала из уравнения движения в обобщенных производных не нужно никаких начальных условий, необходимых при интегрировании обычных дифференциальных уравнений [4.4].

В общем случае система может подвергаться воздействиям, зависящим от обобщенных производных входного сигнала x_1 . Поэтому уравнение движения линейной системы с постоянными параметрами в общей форме запишется

$$A_0 D^n x_2 + A_1 D^{n-1} x_2 + \dots + A_{n-1} D x_2 + A_n x_2 = \\ = B_0 D^m x_1 + B_1 D^{m-1} x_1 + \dots + B_{m-1} D x_1 + B_m x_1. \quad (4.6)$$

Порядок производных в левой и правой частях уравнения (4.6) может быть различным. Например, если в уравнении движения контура, рассмотренного в предыдущем параграфе, $x_1 = u$, а $x_2 = q$, то $n = 2$ и $m = 0$. Здесь не принимаем во внимание чистое запаздывание, полагая время его равным нулю.

Преобразуем уравнение (4.6), взяв за скобки A_n в левой части и в B_m в правой. Тогда, обозначая

$$\frac{A_0}{A_n} = T_n, \quad \frac{A_1}{A_n} = T_{n-1}, \quad \dots, \quad \frac{A_{n-1}}{A_n} = T_1, \\ \frac{B_0}{B_m} = \tau_m, \quad \frac{B_1}{B_m} = \tau_{m-1}, \quad \dots, \quad \frac{B_{m-1}}{B_m} = \tau_1, \quad \frac{B_m}{B_m} = K,$$

получим:

$$T_n D^n x_2 + T_{n-1} D^{n-1} x_2 + \dots + T_1 D x_2 + x_2 = \\ = K (\tau_m D^m x_1 + \tau_{m-1} D^{m-1} x_1 + \dots + \tau_1 D x_1 + x_1).$$

Последнее уравнение можно записать в сокращенной символической форме вида

$$P_2(D) x_2(t) = K P_1(D) x_1(t), \quad (4.7)$$

где

$$P_2(D) = 1 + \sum_{k=1}^n T_k D^k, \quad (4.8)$$

$$P_1(D) = 1 + \sum_{k=1}^m \tau_k D^k \quad (4.9)$$

— полиномы от оператора D . У линейных систем с постоянными параметрами коэффициенты T_k и τ_k этих полиномов постоянны. Постоянная с индексом k должна иметь размерность времени в степени k , так как умножается на D^k , т. е. на величину, имеющую размерность времени в степени $-k$, а такое произведение, суммируясь с единицей, должно быть безразмерным. K — коэффициент усиления, размерность которого, как видно из (4.7), равна отношению размерностей x_2 и x_1 . Величины $\sqrt[k]{T_k}$ и $\sqrt[k]{\tau_k}$ называют *постоянными времени*.

В большинстве случаев удобно пользоваться уравнением движения в операторной форме. Чтобы перейти к такой его форме, рассмотрим изображение обобщенной производной по Лапласу [4.4]. Если

$$X(p) \doteq x(t),$$

где $x(t)$ — сигнал вида (3.1), то, как известно из операторного исчисления,

$$pX(p) - x(+0) \doteq f'(t) 1(t) \quad (4.10)$$

или, так как $x(+0) = f(0)$,

$$pX(p) - f(0) \doteq f'(t) 1(t).$$

Это — изображение первого члена обобщенной производной (3.19). Так как изображение δ -функции равно единице, изображение второго члена той же производной равно $f(0)$. Следовательно, изображение обобщенной производной

$$pX(p) - f(0) + f(0) = pX(p),$$

т. е.

$$pX(p) \doteq Dx.$$

Это соответствие нетрудно распространить на обобщенную производную порядка k :

$$p^k X(p) \doteq D^k x. \quad (4.11)$$

Изображение первой обобщенной производной можно получить и непосредственным применением преобразования Лапласа к правой части выражения (3.19). При этом найдем, что искомое изображение равно сумме изображений членов $f'(t)1(t)$ и $f(0)\delta(t)$. Замечая, что в области интегрирования, выполняемого при вычислении изображения, т. е. при $t > 0$, $1(t) = 1$, найдем первое слагаемое изображения как изображение производной $f'(t)$:

$$\int_0^{\infty} e^{-pt} f'(t) dt = pF(p) - f(0), \quad (4.12)$$

где

$$F(p) \doteq f(t).$$

Второе слагаемое изображения равно, очевидно, произведению постоянной величины $f(0)$ на изображение δ -функции, равное единице (§ 3.3), т. е. равно $f(0)$. Прибавляя это к (4.12) и замечая, что в области существования сигнала (при $t > 0$) $F(p) = X(p)$, получаем (4.11).

В общем случае сигнала с импульсивными составляющими, имеющего вид (3.25), соответствие (4.11) также имеет место. В этом легко убедиться, принимая во внимание, что в силу того же соответствия

$$\delta^{(k)}(t) \doteq p^k.$$

Соответствие (4.11) позволяет представить уравнение движения (4.7) в *операторной* форме, получаемой заменой сигналов их изображениями по Лапласу, а символа обобщенной производной порядка k — комплексным числом p в степени k :

$$P_2(p) X_2(p) = K P_1(p) X_1(p). \quad (4.13)$$

Здесь $X_1(p)$ и $X_2(p)$ — изображения входного и выходного сигналов;

$P_1(p)$ и $P_2(p)$ — полиномы, называемые соответственно *оператором воздействия* и *собственным оператором* рассматриваемой системы.

Решая уравнение (4.13) относительно изображения выходного сигнала, получаем:

$$X_2(p) = W(p) X_1(p), \quad (4.14)$$

где

$$W(p) = K \frac{P_1(p)}{P_2(p)} \quad (4.15)$$

— *передаточная функция* рассматриваемой линейной системы. Из того же уравнения и формулы (4.15) следует, что передаточная функция равна отношению лапласовых изображений выходного и входного сигналов:

$$W(p) = \frac{X_2(p)}{X_1(p)}. \quad (4.16)$$

Пусть входной сигнал x_1 имеет составляющие $x_{11}, x_{12}, x_{13}, \dots, x_{1n}$:

$$x_1 = x_{11} + x_{12} + x_{13} + \dots + x_{1n}.$$

Тогда, если $X_{11}, X_{12}, X_{13}, \dots, X_{1n}$ — изображения этих составляющих, то изображение входного сигнала

$$X_1 = X_{11} + X_{12} + X_{13} + \dots + X_{1n},$$

а изображение выходного сигнала, согласно (4.14),

$$X_2 = W(X_{11} + X_{12} + X_{13} + \dots + X_{1n}).$$

Отсюда следует, что каждая составляющая входного сигнала вызывает появление некоторой составляющей выходного сигнала, имеющей такой же вид, как и при отсутствии других составляющих входного сигнала. В этом проявляется *суперпозиция реакций* — основное свойство линейных систем.

При наличии нескольких входов, каждый из которых связан с выходом передаточной функцией, реакции, вызываемые поданными на эти входы сигналами, также складываются. Таким образом, изучение реакции линейной системы на любой сигнал, поданный на любой ее вход, можно делать в предположении, что никаких других сигналов система не получает, т. е. что она находится в покое. При наличии же нескольких сигналов, поданных на один и тот же или на разные входы, реакции линейной системы на них складываются. На этом основании предыдущие результаты можно обобщить на случай линейной системы, имеющей ряд входов, на каждый из которых подано несколько сигналов.

Пусть K_i и $P_{1i}(p)$ — коэффициент усиления и оператор воздействия, характеризующие i -й вход, на который подаются сигналы x_{ij} . Тогда операторное уравнение движения принимает вид

$$P_2(p) X_2(p) = \sum_i K_i P_{1i}(p) \sum_j X_{ij}(p). \quad (4.17)$$

Отсюда

$$X_2(p) = \sum_i W_i(p) \sum_j X_{ij}(p), \quad (4.18)$$

где

$$W_i(p) = K_i \frac{P_{1i}(p)}{P_2(p)} \quad (4.19)$$

— передаточная функция, связывающая i -й вход с выходом. Возвращаясь от операторного уравнения движения (4.17) к уравнению в обобщенных производных, получаем:

$$P_2(D) x_2(t) = \sum_i K_i P_{1i}(D) \sum_j x_{ij}(t). \quad (4.20)$$

В частном случае системы с одним входом, на который подается лишь один входной сигнал, из (4.17), (4.18) и (4.20) следуют уравнения (4.13), (4.14) и (4.7).

Входной сигнал вида (3.4), т. е. при наличии чистого запаздывания, по теореме сдвига оригинала

$$x_2(t - \tau) \doteq X_2(p) e^{-p\tau}.$$

Поэтому он может быть найден из (4.13), если передаточные функции устройства, в котором имеет место чистое запаздывание, определить как

$$W_i(p) = K_i \frac{P_{1i}(p)}{P_2(p)} e^{-p\tau_i} \quad (4.21)$$

(запаздывание может быть различным для разных входов). Определяемые таким образом передаточные функции устройства с запаздыванием являются трансцендентными.

Приведенный выше вывод уравнений движения и передаточных функций линейной системы показывает, что характеризующие такую систему собственный оператор и оператор воздействия, а равно ее передаточные функции не зависят от вида входных сигналов, а определяются только свойствами системы. В отличие от этого оба оператора, а следовательно, и передаточная функция линейной модели нелинейной системы зависят от входных сигналов. Действительно, положение точки нелинейной характеристики, около которой произведена линеаризация (см. рис. 4.2), может зависеть от входного сигнала, влияющего поэтому и на коэффициенты полиномов P_1 и P_2 .

Приведенный вывод передаточной функции показывает также, что ее выражение (4.15) характеризует систему,

находящуюся в любых условиях, а не только при так называемых «нулевых начальных условиях», как иногда указывают. В действительности выходной сигнал полностью и однозначно определяется передаточной функцией и вызывающим его входным сигналом. Это ясно видно из уравнения (4.14). Не может быть, следовательно, никаких «начальных условий», определяемых из каких-либо других соображений: выходной сигнал однозначно определяется входным и передаточной функцией вместе со своими начальными значениями.

Приравняв нулю знаменатель передаточной функции, получаем *характеристическое уравнение* рассматриваемой системы. Коэффициенты характеристического уравнения реальной системы всегда вещественны. Поэтому его корни, называемые *полюсами* передаточной функции, могут быть либо вещественными, либо сопряженными комплексными. Корни полинома $P_1(p)$ называются *нулями* передаточной функции.

Примеры. 1. Составим уравнение движения электромеханической автоматической следящей системы, функциональная схема которой представлена на рис. 4.3.

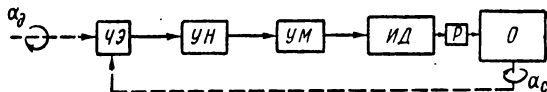


Рис. 4.3. Функциональная схема к примеру составления уравнения движения

Уравнение движения можно получить из условия равновесия действующих на выходном валу моментов

$$M_{\text{дв}} = M_{\text{д}} + M_{\text{пр}},$$

где $M_{\text{дв}}$ — момент, развиваемый исполнительным двигателем *ИД*;
 $M_{\text{д}}$ — динамический момент;

$M_{\text{пр}}$ — момент от противо-э. д. с. и сил трения.

На вход системы подается управляющее воздействие в виде угла поворота командной оси $\alpha_d = f(t) \text{ л } (t)$, а выходной вал системы, поворачивающий объект *О*, будет при этом перемещаться по закону $\alpha_c = \varphi(t) \text{ л } (t)$. Информацию об углах α_d и α_c воспринимает в виде электрических сигналов чувствительный элемент *ЧЭ*, который вырабатывает сигнал, определяемый углом рассогласования $\theta = \alpha_d - \alpha_c$. Этот сигнал через усилители напряжения и мощности *УН* и *УМ* передается исполнительному двигателю, который через редуктор *Р* сообщает требуемое движение объекту.

Момент $M_{\text{пр}}$ пропорционален скорости вращения системы:

$$M_{\text{пр}} = FD\alpha_c$$

(все параметры приведены к валу нагрузки), где F — коэффициент пропорциональности, зависящий от сил трения и противо-э. д. с. двигателя. Динамический момент зависит от ускорения системы и определяется выражением

$$M_d = JD^2\alpha_c,$$

где J — момент инерции системы.

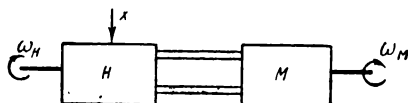


Рис. 4.4. Блок-схема гидрорегулятора скорости

С другой стороны, момент, развиваемый исполнительным двигателем замкнутой системы, должен быть пропорционален углу рассогласования $\theta = \alpha_d - \alpha_c$. Следовательно, $M_{\text{дв}} = K(\alpha_d - \alpha_c)$. Из условия равновесия моментов получаем выражение

$$K\alpha_d - K\alpha_c = JD^2\alpha_c + FD\alpha_c,$$

или

$$K\alpha_d = JD^2\alpha_c + FD\alpha_c + K\alpha_c,$$

которое преобразуем к виду

$$K\alpha_d = (JD^2 + FD + K)\alpha_c,$$

что является частным случаем общей формы записи уравнения движения (4.7). Уравнение движения системы можно записать и в несколько иной форме, если сделать в правой части подстановку $\alpha_d = \theta + \alpha_c$:

$$K\theta = JD^2\alpha_c + FD\alpha_c.$$

Такая форма записи оказывается полезной для проведения некоторых исследований, зависящих от специфики конкретных САУ.

2. Составим уравнения движения и передаточные функции объемного гидрорегулятора скорости, состоящего из насоса H , мотора M и соединяющих их труб [4.6], блок-схема которого представлена на рис. 4.4. Расход Q_H рабочей жидкости насосом, т. е. объем ее, подаваемый насосом в единицу времени, складывается из полезного расхода Q_M через мотор, расхода Q_Y , вызываемого утечкой в обеих машинах и в трубах, и расхода Q_C , обусловленного сжатием жидкости. Отсюда получается уравнение расходов

$$Q_H = Q_M + Q_Y + Q_C. \quad (4.22)$$

Момент M_m , развиваемый на валу гидромотора, уравновешивается динамическим моментом M_d и моментом статических сопро-

тивлений M_c , следовательно, уравнение моментов

$$M_m = M_d + M_c. \quad (4.23)$$

Поставленная задача требует совместного решения этих уравнений. Рассмотрим входящие в них величины.

Регулирование скорости осуществляется изменением расхода Q_n , происходящим при перемещении управляющего органа насоса. Обозначим через x относительное значение этого перемещения, определяемое так, что $-1 \leq x \leq 1$.

Тогда

$$Q_n = q_n \omega_n x, \quad (4.24)$$

где q_n — постоянная насоса, равная объему жидкости, подаваемому им при повороте его вала на единичный угол;

ω_n — угловая скорость вращения этого вала.

Вводя постоянную мотора q_m и угловую скорость ω_m его вала, имеем:

$$Q_m = q_m \omega_m. \quad (4.25)$$

Составляющую расхода Q_n , зависящую от утечки, определим как

$$Q_y = rP, \quad (4.26)$$

где r — ее значение на единицу давления P .

Составляющая же, обусловленная сжатием жидкости,

$$Q_c = D(\Delta V), \quad (4.27)$$

где ΔV — приращение объема V рабочей жидкости, находящейся под давлением вследствие сжатия.

Принимая, что приращение

$$\Delta V = \beta VP,$$

где β — коэффициент объемного сжатия жидкости (относительное приращение объема на единицу давления), находим из (4.27):

$$Q_c = \beta VDP. \quad (4.28)$$

Момент, развиваемый на валу мотора,

$$M_m = \eta q_m P, \quad (4.29)$$

где η — гидравлический коэффициент полезного действия мотора. Динамический момент

$$M_d = JD\omega_m. \quad (4.30)$$

Здесь момент инерции

$$J = J_m + J_r + J_n, \quad (4.31)$$

где J_m — момент инерции ротора гидромотора;

J_r — гидравлический момент инерции, учитывающий влияние поступательного движения жидкости в регуляторе [4.2];

J_n — момент инерции устройства, приводимого мотором, приведенный к валу последнего.

Момент статических сопротивлений предположим заданным в виде некоторой функции времени t :

$$M_c = m(t). \quad (4.32)$$

Подстановка (4.25), (4.26) и (4.28) в (4.22) приводит к преобразованному уравнению расходов

$$\beta VDP + rP + q_m \omega_m = q_n \omega_n x; \quad (4.33)$$

постановка же (4.29), (4.30) и (4.32) в (4.23) — к преобразованному уравнению моментов

$$JD\omega_m + m = \eta q_m P. \quad (4.34)$$

Находя P из (4.34) и подставляя в (4.33), исключим из этих уравнений давление. Из (4.34)

$$P = \frac{1}{\eta q_m} (JD\omega_m + m). \quad (4.35)$$

Здесь m и ω_m — переменные. От них зависит к. п. д. η , который изменяется в широких пределах при изменении нагрузки и скорости вращения гидромотора. А от η зависит гидравлический момент инерции [4.2]. Таким образом, подстановка (4.35) в (4.34) приводит к нелинейному уравнению из-за наличия в нем к. п. д. η , если даже пренебречь другими факторами, приводящими к нелинейности.

Обычный прием линеаризации в рассматриваемом случае состоит в допущении, что к. п. д. имеет некоторое постоянное среднее значение

$$\eta \approx \text{const}. \quad (4.36)$$

Это означает замену реального регулятора его линейной моделью.

Входящая в (4.33) угловая скорость ω_n определяется изменяющимся числом оборотов приводного двигателя насоса. Так, в электрогидроприводе число оборотов электродвигателя зависит от переходных процессов в электрической части привода, которые следует учитывать наряду с гидравлическими переходными процессами [4.1]. Предполагая, что мощность электродвигателя достаточно велика и что он имеет достаточно жесткую механическую характеристику, можно принять, как это делают,

$$\omega_n = \text{const}. \quad (4.37)$$

Тогда, вводя передаточное число гидрорегулятора

$$j = \frac{\omega_m}{\omega_n}, \quad (4.38)$$

напишем уравнение движения регулятора в форме

$$(T_2 D^2 + T_1 D + 1) j(t) = K_x x(t) - K_m (\tau_1 D + \tau_0) m(t). \quad (4.39)$$

Здесь переменные x , m и j записаны как функции времени.

Постоянные $T_1 = \frac{rJ}{\eta q_m^2}$, $T_2 = \frac{\beta VJ}{\eta q_m^2}$; $\tau_0 = r$, $\tau_1 = \beta V$. Коэффициенты

усиления $K_x = \frac{q_n}{q_m}$, $K_m = \frac{1}{\eta q_m^2 \omega_n}$.

Полученное уравнение (4.39) показывает, что передаточное число $j(t)$ зависит от двух воздействий: управляющего и возмущающего. Эти воздействия можно рассматривать как входные сигналы, под действием которых на выходе системы образуется функция $j(t)$. Уравнение (4.39) описывает линейную модель регулятора, вследствие чего процессы, вызываемые воздействиями $x(t)$ и $m(t)$, можно рассматривать порознь, учитывая лишь, что второй процесс может иметь место только при наличии первого. На выходе происходит суперпозиция этих процессов.

Оба сигнала начинают действовать на работающий регулятор при $t=0$. Поэтому их следует записать в виде

$$x(t) = f_x(t) 1(t), \quad (4.40)$$

$$m(t) = f_m(t) 1(t). \quad (4.41)$$

При $t=0$ возмущение $m(t)$ всегда появляется хотя бы в виде момента трения. Но и этот момент появляется скачком при начале движения; это ясно из того, что он меняет знак при перемене направления вращения. Наличие входных сигналов вида (4.40) и (4.41) соответственно ограничивает и функцию $j(t)$. Не рассматривая чистого запаздывания в регуляторе, следует, очевидно, представить передаточное число как

$$j(t) = f_j(t) 1(t).$$

В связи с указанным видом функций (4.40) и (4.41) уравнение движения (4.39) должно быть уравнением в обобщенных производных.

Применяя соответствие (4.11) к уравнению (4.39), получаем следующее операторное уравнение движения гидрорегулятора:

$$(T_2 p^2 + T_1 p + 1) J(p) = K_x X(p) - K_m (\tau_1 p + \tau_0) M(p), \quad (4.42)$$

где $X(p)$ и $J(p)$ — изображения функций $x(t)$ и $j(t)$.

Это уравнение можно записать иначе, введя передаточные функции

$$W_x(p) = \frac{K_x}{T_2 p^2 + T_1 p + 1}, \quad (4.43)$$

$$W_m(p) = K_m \frac{\tau_1 p + \tau_0}{T_2 p^2 + T_1 p + 1}, \quad (4.44)$$

первая из которых относится к управляющему, а вторая — к возмущающему воздействию. Тогда из (4.42) получаем изображение передаточного числа

$$J(p) = W_x(p) X(p) - W_m(p) M(p). \quad (4.45)$$

§ 4.2. ЧАСТОТНЫЕ ХАРАКТЕРИСТИКИ НЕПРЕРЫВНЫХ СИСТЕМ

Если изображение $X_2(p)$ выходного сигнала не имеет особых точек в правой полуплоскости, а входной сигнал изобразим по Фурье, то при подстановке (3.28) в правую часть уравнения (4.14) получим изображение по Фурье

(хотя бы в несобственном смысле — § 3.3) или *комплексный спектр* выходного сигнала:

$$X_2(i\omega) = W(i\omega) X_1(i\omega). \quad (4.46)$$

Здесь $X_1(i\omega)$ — комплексный спектр входного сигнала.

В уравнении (4.46) $W(i\omega)$ — *частотная характеристика* рассматриваемой системы. Из формулы (4.16) следует, что

$$W(i\omega) = \frac{X_2(i\omega)}{X_1(i\omega)}, \quad (4.47)$$

т. е. частотная характеристика равна отношению спектров выходного и соответственно входного сигналов.

При наличии чистого запаздывания частотная характеристика, как следует из формулы (4.21),

$$W(i\omega) = K \frac{P_1(i\omega)}{P_2(i\omega)} e^{-i\omega\tau}. \quad (4.48)$$

Частотную характеристику (4.47) всегда можно разложить на вещественную и мнимую части:

$$W(i\omega) = U(\omega) + iV(\omega). \quad (4.49)$$

Можно представить ее и в виде

$$W(i\omega) = M(\omega) e^{i\varphi(\omega)}, \quad (4.50)$$

где *модуль*

$$M(\omega) = \sqrt{U^2(\omega) + V^2(\omega)}, \quad (4.51)$$

а *фаза* удовлетворяет условию

$$\operatorname{tg} \varphi(\omega) = \frac{V(\omega)}{U(\omega)}. \quad (4.52)$$

Пользуясь вещественной и мнимой частями U и V как прямоугольными координатами, можно изобразить вектор (4.50) на плоскости этих координат. Кривая, проведенная через концы векторов, соответствующих различным значениям частоты ω , т. е. его годограф, называется *амплитудно-фазовой характеристикой* рассматриваемого устройства. Радиус-вектор, проведенный из начала координат в какую-либо точку этой кривой, равен по величине модулю M частотной характеристики, а угол φ , составляемый им с осью U , — ее фазе (рис. 4.5). Такую амплитудно-фазовую характеристику называют *характеристикой второго рода*. В отличие от нее у *характеристики первого рода* (называемой также *обратной* и *инверсной*) радиус-

вектор имеет величину, обратную модулю M , и составляет с осью U угол, отсчитываемый по часовой стрелке. При этом по оси абсцисс должна быть отложена величина

$$\frac{U}{U^2 + V^2},$$

а по оси ординат

$$-\frac{V}{U^2 + V^2}.$$

Заметим, что, как у всякого изображения по Фурье, вещественная часть частотной характеристики должна быть

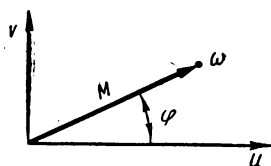


Рис. 4.5. К построению амплитудно-фазовой характеристики второго рода

четной функцией частоты, а мнимая — нечетной. Поэтому из двух ветвей амплитудно-фазовой характеристики, соответствующих положительным и отрицательным частотам, достаточно вычислить лишь одну; другая же является зеркальным изображением первой в вещественной оси.

Как показывает уравнение (4.46), частотная характеристика позволяет найти спектр выходного сигнала по спектру любого сигнала, поданного на вход. Вместе с тем она непосредственно определяет выходной сигнал, вызываемый *гармоническим воздействием*. Таким воздействием является входной сигнал

$$x_1(t) = a_1 \sin \gamma t \cdot 1(t). \quad (4.53)$$

Изображение вызываемого им выходного сигнала

$$X_2(p) = W(p) \frac{a_1 \gamma}{p^2 + \gamma^2}, \quad (4.54)$$

где $W(p)$ — передаточная функция.

Рассмотрим случай, в котором полюса изображения $X_2(p)$, зависящие от передаточной функции $W(p)$, находятся в левой полуплоскости (это означает, что рассматривается устойчивая система, см § 6.1). Но $X_2(p)$ имеет еще два сопряженных полюса на мнимой оси $p = \pm i\gamma$, обусловленных изображением входного сигнала. Поэтому выходной сигнал имеет затухающую (регулярную) составляющую $x_p(t)$, определяемую полюсами передаточной функции, и незатухающую (нерегулярную) составляющую $x_n(t)$,

определяемую полюсами изображения входного сигнала. Тогда

$$X_2(p) = X_p(p) + X_n(p),$$

где

$$X_p(p) \doteq x_p(t),$$

$$X_n(p) \doteq x_n(t).$$

При указанном расположении полюсов обе составляющие выходного сигнала изобразимы по Фурье. Поэтому из (4.54) найдем комплексный спектр выходного сигнала:

$$X_2(i\omega) = \frac{a_1 \gamma W(i\omega)}{\gamma^2 - \omega^2}.$$

Отсюда

$$\frac{a_1 \gamma W(i\omega)}{\gamma^2 - \omega^2} = \frac{A_1}{\gamma + \omega} + \frac{A_2}{\gamma - \omega} + X_p(i\omega), \quad (4.55)$$

или

$$a_1 \gamma W(i\omega) = A_1(\gamma - \omega) + A_2(\gamma + \omega) + X_p(i\omega)(\gamma^2 - \omega^2).$$

Полагая здесь $\omega = \gamma$, получаем:

$$a_1 \gamma W(i\gamma) = 2\gamma A_2$$

и находим

$$A_2 = \frac{a_1 W(i\gamma)}{2};$$

при $\omega = -\gamma$

$$a_1 \gamma W(-i\gamma) = 2\gamma A_1,$$

откуда

$$A_1 = \frac{a_1 W(-i\gamma)}{2}.$$

Таким образом, нерегулярная часть изображения из (4.55)

$$\begin{aligned} X_n(i\omega) &= \left(\frac{A_1}{\gamma + \omega} + \frac{A_2}{\gamma - \omega} \right) = \\ &= \frac{a_1}{2} \left[\frac{W(-i\gamma)}{\gamma + \omega} + \frac{W(i\gamma)}{\gamma - \omega} \right]. \end{aligned}$$

Согласно (4.49)

$$W(\pm i\gamma) = U(\gamma) \pm iV(\gamma),$$

так как $U(\gamma)$ — четная, а $V(\gamma)$ — нечетная функции частоты. Пользуясь этими выражениями, имеем:

$$\begin{aligned} X_n(i\omega) &= \frac{a_1}{2} \left[\frac{U(\gamma) - iV(\gamma)}{\gamma + \omega} + \frac{U(\gamma) + iV(\gamma)}{\gamma - \omega} \right] = \\ &= \frac{a_1 U(\gamma)}{2} \left(\frac{1}{\gamma + \omega} + \frac{1}{\gamma - \omega} \right) - \\ &- i \frac{a_1 V(\gamma)}{2} \left(\frac{1}{\gamma + \omega} - \frac{1}{\gamma - \omega} \right) = a_1 \frac{\gamma U(\gamma) + i\omega V(\gamma)}{\gamma^2 - \omega^2}. \end{aligned}$$

Отсюда, переходя к оригиналам, находим:

$$x_n(t) = a_1 [U(\gamma) \sin \gamma t + V(\gamma) \cos \gamma t] 1(t),$$

или

$$x_n(t) = a_2 \sin(\gamma t + \psi) 1(t), \quad (4.56)$$

где амплитуда выходного сигнала

$$a_2 = a_1 \sqrt{U^2(\gamma) + V^2(\gamma)},$$

а сдвиг его фазы определяется соотношением

$$\operatorname{tg} \psi = \frac{V(\gamma)}{U(\gamma)}.$$

Следовательно, нерегулярная составляющая выходного сигнала линейной системы, вызванная синусоидальным входным сигналом частоты γ , также синусоидальна и имеет ту же частоту, но другую амплитуду и сдвинута по фазе относительно входного сигнала.

Сопоставляя последние выражения с (4.51) и (4.52), находим, что отношение амплитуд выходного (нерегулярной составляющей) и входного сигналов равно модулю частотной характеристики при частоте входного сигнала:

$$\frac{a_2}{a_1} = M(\gamma); \quad (4.57)$$

разность фаз тех же сигналов равна фазе частотной характеристики при той же частоте:

$$\psi = \varphi(\gamma). \quad (4.58)$$

Зависимость модуля частотной характеристики от частоты называют *амплитудной характеристикой* системы, а зависимость аргумента частотной характеристики от частоты — *фазовой характеристикой*. Как видно из (4.57) и (4.58), значения амплитудной и фазовой характеристик

при некоторой частоте дают отношение амплитуд и разность фаз нерегулярной составляющей выходного и входного сигналов, получаемых при поданном на вход гармоническом сигнале той же частоты. Отсюда следует возможность экспериментального получения амплитудной и фазовой характеристик устойчивой линейной системы, для которой выходной сигнал через некоторое время после подачи на вход сигнала (4.53) будет содержать практически только нерегулярную составляющую.

Существует целый класс линейных систем, называемых *минимально-фазовыми*, для которых динамические свойства однозначно определяются только амплитудной или фазовой характеристикой (последней — с точностью до коэффициента усиления на нулевой частоте, т. е. по фазовой характеристике минимально-фазовой системы ее передаточная функция находится с точностью до постоянного множителя). Минимально-фазовая система имеет фазовую характеристику, наибольшую по сравнению со всеми другими системами, имеющими такую же амплитудную характеристику. Если учесть, что для большинства систем фазовая характеристика отрицательна, то минимально-фазовая система имеет наименьшую по абсолютной величине фазовую характеристику по сравнению со всеми другими, имеющими такую же амплитудную характеристику. К минимально-фазовым системам относятся системы, передаточная функция которых есть рациональная дробь с нулями и полюсами, расположенными в левой полуплоскости или на мнимой оси [4.3].

Все частотные характеристики определяют связь выхода системы с каким-либо ее входом и изменяются при переходе к другому входу.

Пример. Найдем частотные характеристики гидрорегулятора, передаточные функции которого по управляющему и возмущающему воздействиям имеют выражения (4.43) и (4.44). Из этих выражений получаем:

$$W_x(i\omega) = \frac{K_x}{1 - T_2\omega^2 + iT_1\omega},$$

$$W_m(i\omega) = K_x \frac{\tau_0 + i\tau_1\omega}{1 - T_2\omega^2 + iT_1\omega}.$$

Отделяя вещественную часть от мнимой, находим вещественную и мнимую составляющие частотных характеристик:

по управляющему воздействию —

$$U_x(\omega) = K_x \frac{1 - T_2 \omega^2}{(1 - T_2 \omega^2)^2 + T_1^2 \omega^2},$$

$$V_x(\omega) = -K_x \frac{T_1 \omega}{(1 - T_2 \omega^2)^2 + T_1^2 \omega^2};$$

по возмущающему воздействию —

$$U_m(\omega) = K_m \frac{\tau_0 + (\tau_1 T_1 - \tau_0 T_2) \omega_2}{(1 - T_2 \omega^2)^2 + T_1^2 \omega^2},$$

$$V_m(\omega) = K_m \frac{\tau_1 (1 - T_2 \omega^2) - \tau_0 T_1}{(1 - T_2 \omega^2)^2 + T_1^2 \omega^2} \omega.$$

§ 4.3. ЛОГАРИФМИЧЕСКИЕ ЧАСТОТНЫЕ ХАРАКТЕРИСТИКИ

Зависимость модуля частотной характеристики от частоты, представляемая в логарифмическом масштабе, называется *логарифмической амплитудной характеристикой* (ЛАХ), а зависимость аргумента частотной характеристики от логарифма частоты — *фазовой частотной характеристикой*.

При построении и применении этих характеристик пользуются акустическими единицами. Считают, что частоты ω_2 и ω_1 отличаются на *октаву*, если $\omega_2/\omega_1 = 2$, и на *декаду*, если $\omega_2/\omega_1 = 10$. Мощности P_2 и P_1 отличаются на ± 1 бел, если $\lg \frac{P_2}{P_1} = \pm 1$, и на ± 1 децибел (дб), если $\lg \frac{P_2}{P_1} = \pm 0,1$. Отличие мощностей на ± 1 дб означает, что $\frac{P_2}{P_1} = 10^{0,1} = 1,259$; при отличии мощностей на -1 дб $\frac{P_2}{P_1} = 10^{-0,1} = 0,7943$. Отношение амплитуд a_2 и a_1 , соответствующих мощностям P_2 и P_1 , отличающимся на ± 1 дб, должно удовлетворять условию

$$\lg \left(\frac{a_2}{a_1} \right)^2 = \pm 0,1,$$

или

$$20 \lg \frac{a_2}{a_1} = \pm 1.$$

При этом считают, что амплитуды a_2 и a_1 отличаются также на ± 1 дб. Поэтому отличие амплитуд на ± 1 дб

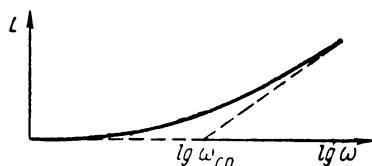
означает, что

$$\frac{a_2}{a_1} = \sqrt{1,259} = 1,122,$$

а отличие на -1 дБ , что $\frac{a_2}{a_1} = \frac{1}{1,122} = 0,8913$.

Так как модуль частотной характеристики определяет отношение амплитуд выходного и входного сигналов при гармоническом сигнале на входе, подставляя модуль $M = |W(i\omega)|$ в функции $\lg \omega$, уравнение ЛАХ записывают в виде

$$L = 20 \lg M. \quad (4.59)$$



Получающиеся при этом кривые аппроксимируют ломаными линиями, составленными из асимптот (рис. 4.6). Угловые точки этих ломаных определяют *сопрягающие частоты* $\omega_{сп}$.

Рис. 4.6. К построению логарифмической амплитудной характеристики

Построим ЛАХ системы, имеющей передаточную функцию

$$W(p) = \frac{K}{Tp + 1}. \quad (4.60)$$

Вводя частоту, определяемую постоянной времени T ,

$$\omega_T = \frac{1}{T},$$

приведем передаточную функцию к виду

$$W(p) = \frac{K\omega_T}{p + \omega_T}.$$

Тогда частотная характеристика рассматриваемого устройства

$$W(i\omega) = \frac{K\omega_T}{\omega_T + i\omega},$$

а ее модуль

$$|W(i\omega)| = \frac{K\omega_T}{\sqrt{\omega_T^2 + \omega^2}}.$$

Следовательно, по (4.59)

$$L = 20 \lg \frac{K \omega_T}{\sqrt{\omega_T^2 + \omega^2}},$$

или

$$L = 20 (\lg K + \lg \omega_T - \lg \sqrt{\omega_T^2 + \omega^2}).$$

Для построения и исследования ЛАХ достаточно рассмотреть случай, в котором $K = 1$, так как при всех других значениях коэффициента усиления характеристика получается параллельным переносом на отрезок $20 \lg K$ вдоль оси ординат. При $K = 1$

$$L = 20 \lg \omega_T - 20 \lg \sqrt{\omega_T^2 + \omega^2}. \quad (4.61)$$

Отсюда находим, что в области частот, где $\omega \ll \omega_T$, $L = 0$, т. е. характеристика асимптотически приближается к оси абсцисс.

В области же, где

$$\omega \gg \omega_T, \quad (4.62)$$

$$L = 20 \lg \omega_T - 20 \lg \omega, \quad (4.63)$$

т. е. кривая асимптотически приближается к прямой, представляемой этим уравнением.

Исследуем асимптоту, имеющую уравнение (4.63). Очевидно, что она пересекает ось абсцисс в точке $\omega = \omega_T$. Чтобы выяснить ее наклон, рассмотрим частоты ω_1 и ω_2 , находящиеся в области (4.62) и отличающиеся на октаву, так что

$$\omega_2 = 2\omega_1. \quad (4.64)$$

При $\omega = \omega_1$ из (4.72) имеем:

$$L_1 = 20 \lg \omega_T - 20 \lg \omega_1,$$

а при $\omega = \omega_2$ находим, учитывая (4.64):

$$L_2 = L_1 - 20 \lg 2.$$

Отсюда

$$L_2 - L_1 \approx -6 \text{ дб}.$$

Следовательно, исследуемая асимптота имеет наклон или *крутизну* — 6 *дб/окт* и расположена, как показано на рис. 4.7.

Чтобы измерить наклон асимптоты в децибелах на декаду, возьмем частоты ω_1 и ω_2 , отличающиеся на декаду

$$\omega_2 = 10\omega_1.$$

Этим частотам соответствуют следующие значения L :
при $\omega = \omega_1$

$$L_1 = 20 \lg \omega_T - 20 \lg \omega_1,$$

а при $\omega = \omega_2$

$$L_2 = 20 \lg \omega_T - 20 \lg 10 - 20 \lg \omega_1 = L_1 - 20.$$

Наклон асимптоты равен -20 дб/дек. Следовательно, $6 \text{ дб/окт} = 20 \text{ дб/дек.}$

Построение асимптот дает возможность получить кусочно-линейную аппроксимацию ЛАХ, т. е. придать амплитудной частотной характеристике простую форму. Аппроксимируя ЛАХ асимптотами, сопряженными при частоте

$$\omega = \omega_T, \quad (4.65)$$

получим при этой частоте наибольшую ошибку. Эта ошибка равна истинному значению ординаты в точке сопряжения, найденному подстановкой (4.65) в (4.61):

$$L = 20 \lg \omega_T - 20 \lg (\omega_T \sqrt{2}) = -20 \lg \sqrt{2} \approx -3 \text{ дб.}$$

Пример. Построим и исследуем ЛАХ по передаточной функции

$$W(p) = \frac{Tp}{Tp + 1}. \quad (4.66)$$

Полагая $p = i\omega$, находим из передаточной функции выражение частотной характеристики

$$W(i\omega) = \frac{iT\omega}{1 + iT\omega} = \frac{T^2\omega^2 + iT\omega}{1 + T^2\omega^2}.$$

Следовательно, вещественная часть частотной характеристики

$$U(\omega) = \frac{1}{1 + \frac{1}{T^2\omega^2}},$$

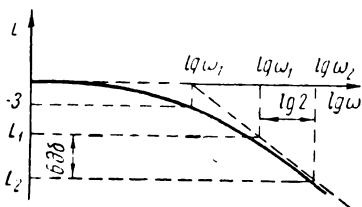


Рис. 4.7. ЛАХ системы с передаточной функцией (4.60)

мнимая часть

$$V(\omega) = \frac{1}{T\omega + \frac{1}{T\omega}},$$

модуль

$$M = \frac{1}{\sqrt{1 + \frac{1}{T^2\omega^2}}}.$$

Тогда

$$L = 20 \lg \frac{1}{\sqrt{1 + \frac{1}{T^2\omega^2}}} = 20 \lg \omega - 20 \lg \sqrt{\omega^2 + \omega_{ca}^2},$$

где

$$\omega_{ca} = \frac{1}{T}.$$

При $\omega \gg \omega_{ca}$ асимптотой является ось абсцисс:

$$L = 0.$$

При $\omega \ll \omega_{ca}$ получаем $L = 20 \lg \omega - 20 \lg \omega_{ca}$; в логарифмическом масштабе — это уравнение прямой, пересекающей ось $\lg \omega$ при

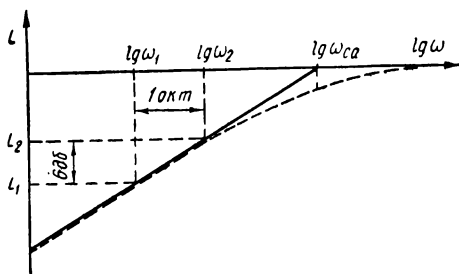


Рис. 4.8. ЛАХ системы с передаточной функцией (4.66)

$\omega = \omega_{ca}$, а ось L — при $L = -20 \lg \omega_{ca}$. Определим наклон второй асимптоты. Для этого возьмем частоты ω_1 и ω_2 , отличающиеся на октаву, так что $\omega_2 = 2\omega_1$. Этим частотам соответствуют следующие значения L :

$$L_1 = 20 \lg \omega_1 - 20 \lg \omega_{ca},$$

$$\begin{aligned} L_2 &= 20 \lg \omega_2 - 20 \lg \omega_{ca} = 20 \lg 2\omega_1 - 20 \lg \omega_{ca} = \\ &= 20 \lg 2 + 20 \lg \omega_1 - 20 \lg \omega_{ca} = 20 \lg 2 + L_1 \approx 6 + L_1. \end{aligned}$$

Следовательно, наклон асимптоты равен 6 дБ/окт (рис. 4.8). Наибольшая погрешность асимптотической аппроксимации будет, оче-

видно, при $\omega = \omega_{ca}$. Она равна истинному значению L при этой частоте, т. е.

$$20 \lg \omega_{ca} - 20 \lg \sqrt{2\omega_{ca}^2} = 20 \lg \omega_{ca} - 20 \lg \omega_{ca} - 20 \lg \sqrt{2} \approx -3 \text{ дб}.$$

§ 4.4. ВРЕМЕННЫЕ ХАРАКТЕРИСТИКИ НЕПРЕРЫВНЫХ СИСТЕМ

Передаточная функция и частотная характеристика линейного устройства зависят только от его динамических свойств и позволяют находить выходной сигнал по заданному входному, пользуясь уравнением движения в форме (4.14) или (4.46). Рассмотрим систему, имеющую лишь один вход, на который подан импульсивный сигнал $x_1(t) = \delta(t)$. Тогда (см. § 3.3) $X_1(p) = 1$; из (4.14) получаем:

$$X_2(p) = W(p). \quad (4.67)$$

Отсюда следует, что в случаях, когда полюса передаточной функции находятся в левой полуплоскости, комплексный спектр выходного сигнала

$$X_2(i\omega) = W(i\omega), \quad (4.68)$$

т. е. выражается частотной характеристикой системы. Тогда легко получить выражение выходного сигнала $x_2(t)$, пользуясь интегралом Фурье. Этот сигнал, представляющий собой реакцию рассматриваемого устройства на мгновенный единичный импульс, называют *импульсной характеристикой* (а также импульсной переходной функцией, или *весовой функцией*). Обозначая ее через $w(t)$, имеем:

$$w(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{i\omega t} W(i\omega) d\omega. \quad (4.69)$$

И наоборот, частотная характеристика является изображением импульсной характеристики по Фурье:

$$W(i\omega) = \int_0^{\infty} e^{-i\omega t} w(t) dt. \quad (4.70)$$

Последние выражения показывают, что импульсная характеристика системы, передаточная функция которой имеет полюса лишь в левой полуплоскости, — затухающая функция времени. Из них следует также, что передаточная функция — изображение импульсной характеристики по Лапласу:

$$W(p) \doteq w(t).$$

Заметим, что функция $w(t)$ является выходным сигналом вида (3.1), т. е. содержит множитель в виде единичной функции $1(t)$, исключающей ее существование при $t < 0$. Физически реакции на импульс не может быть до его появления, а для этого требуется, чтобы в математическом описании реакции присутствовал множитель $1(t)$.

Положим теперь, что в другом случае входной сигнал имеет вид единичного скачка: $x_1(t) = 1(t)$. При этом изображение входного сигнала

$$X_1(p) = \frac{1}{p},$$

а изображение выходного сигнала

$$X_2(p) = \frac{W(p)}{p}, \quad (4.71)$$

где $W(p)$ — передаточная функция системы.

Реакцию системы на входной сигнал в виде единичного скачка называют *переходной характеристикой*. Обозначим эту реакцию через $h(t)$, а ее изображение по Лапласу — через $H(p)$. Тогда передаточная функция

$$W(p) = pH(p), \quad (4.72)$$

т. е. является, согласно (4.11), лапласовым изображением первой обобщенной производной переходной характеристики.

Сопоставляя выражение (4.72) и соответствие (4.11), видим, что весовая функция есть первая обобщенная производная переходной характеристики

$$w(t) = Dh(t),$$

т. е.

$$w(t) = \dot{f}_h(t) 1(t) + f_h(0) \delta(t), \quad (4.73)$$

где $\dot{f}_h(t)$ — описывающая функция переходной характеристики.

Последнее выражение показывает, что в тех случаях, когда $\dot{f}_h(0) \neq 0$, т. е. переходная характеристика начинается скачком, весовая функция содержит импульс. Но весовая функция $w(t)$ имеет и составляющую $\dot{f}_w(t) 1(t)$, в которую входит ее описывающая функция $\dot{f}_w(t)$. Так как члены, содержащие единичную функцию, в обеих

частях уравнения (4.73) должны быть равны, получаем, что

$$\dot{f}_h(t) = \dot{f}_w(t).$$

Если полюса передаточной функции находятся в левой полуплоскости, то изображение выходного сигнала по Фурье в рассматриваемом случае имеет вид

$$H(i\omega) = \frac{W(i\omega)}{i\omega}. \quad (4.74)$$

Поэтому частотная характеристика

$$W(i\omega) = i\omega H(i\omega), \quad (4.75)$$

где

$$H(i\omega) \stackrel{\cdot}{=} h(t) \quad (4.76)$$

— изображение по Фурье переходной характеристики.

Импульсная и переходная характеристики являются функциями времени, определяющими свойства системы. Их называют поэтому *временными характеристиками*. По этим характеристикам можно судить и о реакции системы на другие воздействия, как будет показано в гл. VII.

Примеры. 1. Построим и исследуем переходную характеристику в случае, когда передаточная функция имеет вид (4.60). Согласно (4.71), изображение искомой характеристики

$$H(p) = \frac{K}{(Tp + 1)p}.$$

Переходя к оригиналу, получаем:

$$h(t) = K(1 - e^{-t/T}) 1(t). \quad (4.77)$$

Отсюда следует, что спустя время $t = T$ после включения входного сигнала, выходной сигнал

$$h(T) = K(1 - e^{-1}) = 0,632K. \quad (4.78)$$

С помощью этого соотношения можно найти постоянную времени T по переходной характеристике (рис. 4.9). Далее находим, что прямая $h = K$ является асимптотой рассматриваемой переходной характеристики, так как

$$\lim_{t \rightarrow \infty} h(t) = K.$$

При $t = 0$ имеем $h(0) = 0$ — характеристика проходит через начало координат. Поэтому обобщенная производная выходного

сигнала

$$Dh = \frac{K}{T} e^{-t/T} 1(t). \quad (4.79)$$

Иначе ее можно представить так:

$$Dh = \frac{K 1(t) - h(t)}{T}.$$

Замечая же, что

$$Dh = \lg \alpha 1(t),$$

где α — угол наклона касательной к характеристике; T — прилежащий к углу α катет прямоугольника (рис. 4.9). Отсюда вновь

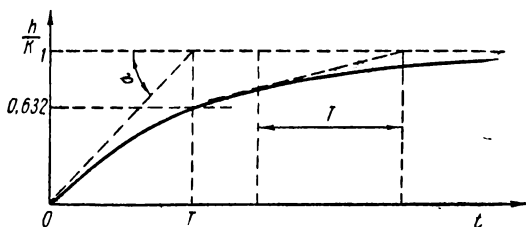


Рис. 4.9. Переходная характеристика в случае передаточной функции (4.60)

следует соотношение (4.78), а также уравнение касательной в начале координат $h = K \frac{t}{T}$, которое можно легко получить и из (4.77), полагая $t \approx 1 - \frac{t}{T}$.

Согласно (4.73), получили импульсную характеристику (4.79) рассматриваемого устройства.

2. Найдём переходную характеристику звена, имеющего передаточную функцию

$$W(p) = K \frac{1 + T_1 p}{1 + T_2 p}. \quad (4.80)$$

Изображение искомой характеристики

$$H(p) = K \frac{1 + T_1 p}{p(1 + T_2 p)}. \quad (4.81)$$

Здесь

$$\frac{1 + T_1 p}{p(1 + T_2 p)} = \frac{M}{p} + \frac{N}{1 + T_2 p},$$

где $M = 1$, $N = T_1 - T_2$; следовательно,

$$H(p) = K \left(\frac{1}{p} + \frac{\frac{T_1}{T_2} - 1}{\frac{1}{T_2} + p} \right),$$

откуда переходная характеристика

$$h(t) = K \left[1 - \left(1 - \frac{T_1}{T_2} \right) e^{-t/T_2} \right] 1(t). \quad (4.82)$$

3. Зная переходную характеристику звена, имеющего передаточную функцию (4.80), найдем его импульсную характеристику. По (4.82)

$$f_h(t) = K \left[1 - \left(1 - \frac{T_1}{T_2} \right) e^{-t/T_2} \right].$$

Вычислив производную этой функции, найдем согласно (4.81) описывающую функцию

$$f_w(t) = K \frac{T_2 - T_1}{T_2^2} e^{-t/T_2}.$$

Тогда по (4.73) получим искомую импульсную характеристику

$$w(t) = K \left[\frac{T_2 - T_1}{T_2^2} e^{-t/T_2} 1(t) + \frac{T_1}{T_2} \delta(t) \right].$$

§ 4.5. ХАРАКТЕРИСТИКИ ИМПУЛЬСНЫХ СИСТЕМ

В составе *импульсной системы* должны быть элементы, преобразующие непрерывные сигналы в импульсные. Такие элементы называют *импульсными*. Рассмотрим простейшие системы такого рода, содержащие лишь один импульсный элемент.

Пусть система состоит из последовательно соединенных импульсного элемента *ИЭ*, на который поступает непрерывный входной сигнал $x_0(t)$, и непрерывной части *НЧ*,

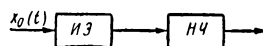


Рис. 4.10. Простейшая система с одним импульсным элементом

как показано на рис. 4.10. Импульсный элемент преобразует входной сигнал в импульсы того или иного вида. *Простейший импульсный элемент* дает мгновенные единичные импульсы, следующие с постоянным периодом повторения T_n . Если параметры импульсного элемента постоянны и можно пренебречь чистым запаздыванием, то импульс, появляющийся на его выходе в момент времени

$t = n T_n$, пропорционален значению входного сигнала в тот же момент времени. Таким образом, на выходе этого элемента получается последовательность мгновенных единичных импульсов, модулированных функцией $x_0(t)$. В частности таким элементом является *ключ* любой конструкции (например, электромагнитный или электронный), по-

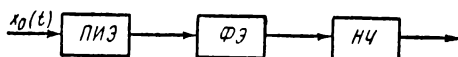


Рис. 4.11. Выделение простейшего импульсного элемента

средством которого осуществляется кратковременное включение сигнала.

Импульсный элемент, на выходе которого получают импульсы той или иной формы, можно заменить последовательным соединением простейшего импульсного элемента *ПИЭ* и *формирующего элемента ФЭ* непрерывного типа, преобразующего модулированные мгновенные еди-

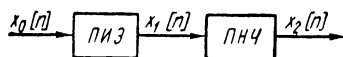


Рис. 4.12. Преобразованная схема простейшей системы с одним импульсным элементом

ничные импульсы в импульсы требуемой формы (рис. 4.11). Далее непрерывные элементы *ФЭ* и *НЧ* можно объединить в один элемент *ПНЧ*, называемый *приведенной непрерывной частью* [4.5]. Тогда рассматриваемая импульсная система, называемая *разомкнутой*, будет представлять собой последовательное соединение простейшего импульсного элемента *ПИЭ* и непрерывного элемента *ПНЧ* (рис. 4.12). *ПНЧ* будем предполагать линейной.

Так как импульсный элемент реагирует лишь на дискретные значения входного сигнала, то только на эти значения будет реагировать и рассматриваемая система. Поэтому вместо непрерывного входного сигнала $x_0(t)$ можно рассматривать решетчатый сигнал $x_0[n]$, который вызывает на выходе *ПИЭ* решетчатый сигнал $x_1[n]$, представляющий собой последовательность мгновенных единичных импульсов, модулированную функцией $x_0(t)$. Вследствие суперпозиции реакций *ПНЧ* сигнал $x_2(t)$ будет равен сумме реакций *a, б, в, г, д* на каждый импульс

$x_1[n]$ (рис. 4.13), определяемых импульсной характеристикой ПНЧ.

Пусть $\omega(t)$ — импульсная характеристика ПНЧ по (4.69). Реакция ПНЧ на единичный мгновенный импульс, приложенный при $n=m$, равна $\omega(t - mT_n)$. Иначе эту

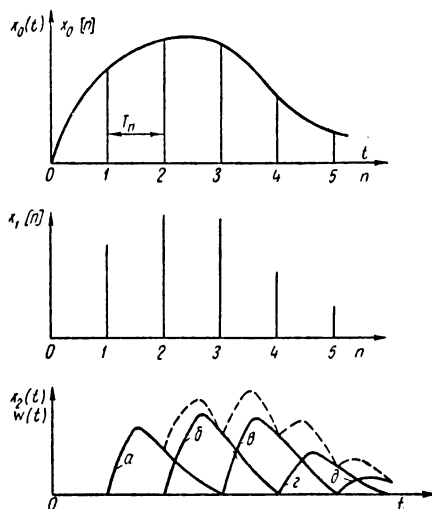


Рис. 4.13. Сигналы в простейшей импульсной системе

реакцию можно представить в виде смещенной решетчатой функции $\omega[(n-m)\epsilon]$, так как

$$t = (n + \epsilon) T_n.$$

Соответственно этому запишем

$$x_2(t) = x_2[n, \epsilon].$$

В интервале $0,1$, т. е. при $m=0$,

$$x_2[n, \epsilon] = x_0[0] \omega[n, \epsilon].$$

В интервале от $n=1$ до $n=2$, в котором $m=1$, к правой части предыдущего выражения добавляется $x_0[1] \times \omega[(n-1), \epsilon]$.

В интервале от $n=2$ до $n=3$, где $m=2$, добавляется еще $x[2] \omega[(n-2), \epsilon]$ и т. д. Таким образом, выход-

ной сигнал *ПНЧ* в интервале $n, n+1$

$$x_2[n, \varepsilon] = \sum_{m=0}^n x_0[m] w[(n-m), \varepsilon]. \quad (4.83)$$

Применяя к (4.83) дискретное преобразование Лапласа, имеем:

$$X_2(q, \varepsilon) = W(q, \varepsilon) \cdot X_0(q), \quad (4.84)$$

где

$$\begin{aligned} X_2(q, \varepsilon) &\doteq x_2[n, \varepsilon], \\ X_0(q) &\doteq x_0[m], \\ W(q, \varepsilon) &\doteq w[n, \varepsilon]. \end{aligned} \quad (4.85)$$

Уравнение (4.84) аналогично уравнению (4.14) непрерывной системы. В нем изображение $W(q, \varepsilon)$ является *передаточной функцией* рассматриваемой *импульсной системы*:

$$W(q, \varepsilon) = \frac{X_2(q, \varepsilon)}{X_0(q)}. \quad (4.86)$$

Согласно (4.85)

$$W(q, \varepsilon) = e^{-\varepsilon q} \sum_{n=0}^{\infty} e^{-qn} w[n, \varepsilon]. \quad (4.87)$$

Здесь

$$w[n, \varepsilon] = w[(n+1)\varepsilon] T_n = w(t)$$

— импульсная характеристика *ПНЧ*. Поэтому передаточная функция разомкнутой импульсной системы получается применением дискретного преобразования Лапласа к импульсной характеристике приведенной непрерывной части этой системы.

Получив передаточную функцию импульсной системы, можно вывести из нее и частотные характеристики. Как и для непрерывных систем, передаточная функция и все получаемые из нее характеристики зависят от того, с каким из входов системы рассматривается связь ее выхода.

Пример. Пусть *ПНЧ* простейшей импульсной системы имеет передаточную функцию (4.60). Импульсная характеристика *ПНЧ* по (4.79)

$$w(t) = \frac{K}{T} e^{-t/T} 1(t)$$

или, в форме смещенной решетчатой функции,

$$w(n, \epsilon) = \frac{K}{T} \exp\left(-\frac{n + \epsilon}{T} T_n\right) 1[n, \epsilon],$$

где T_n — период повторения, характеризующий ПИЭ. Тогда по (4.87) передаточная функция рассматриваемой простейшей импульсной системы

$$\begin{aligned} W(q, \epsilon) &= e^{-q\epsilon} \frac{K}{T} \sum_{n=0}^{\infty} \exp\left\{-\left(q + \frac{T_n}{T}\right)n - \epsilon \frac{T_n}{T}\right\} \times \\ &\times 1[n, \epsilon] = \frac{K}{T} e^{-\epsilon\left(\frac{T_n}{T} + q\right)} \sum_{n=0}^{\infty} \exp\left\{-\left(q + \frac{T_n}{T}\right)n\right\}, \end{aligned}$$

так как при $n \geq 0$ единичная функция $1[n, \epsilon] = 1$. Суммируя полученную геометрическую прогрессию, находим:

$$W(q, \epsilon) = \frac{K}{T} \frac{e^{-\epsilon\left(\frac{T_n}{T} + q\right)}}{1 - \exp\left(-q - \frac{T_n}{T}\right)}. \quad (4.88)$$

СТРУКТУРНЫЕ ПРЕОБРАЗОВАНИЯ АВТОМАТИЧЕСКИХ СИСТЕМ

§ 5.1. СТРУКТУРНЫЕ СХЕМЫ И СИГНАЛЬНЫЕ ГРАФЫ

В отличие от уже рассмотренных схем — функциональных и блок-схем — *структурные схемы* представляют собой автоматическую систему, состоящую из *элементов направленного действия*, определяемых их передаточными функциями и называемых звеньями. Элемент направленного действия — это такая часть автоматической системы, выходной сигнал которой зависит от входного, но не влияет на входной сигнал.

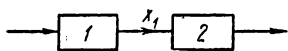


Рис. 5.1. Последовательное соединение

Структурная схема представляет систему, разделенную на звенья и показывает порядок их соединения и взаимодействия.

Существуют три основных способа соединения элементов структурных схем (рассмотренные ниже на примерах простейших систем, составленных из двух звеньев):

1) *последовательное* соединение, или *разомкнутая цепочка*, при котором выходной сигнал предыдущего элемента поступает на вход последующего (рис. 5.1). Сигнал x_1 является выходным для элемента 1 и входным для элемента 2;

2) *параллельное*, или *согласное параллельное*, соединение, при котором входы всех элементов получают один и тот же сигнал, а выходные сигналы их суммируются на общем выходе (рис. 5.2). Сигнал x , поданный на вход системы, поступает и на входы обоих элементов:

$$x_1 = x_2 = x.$$

Выходные сигналы y_1 и y_2 суммируются:

$$y = y_1 + y_2;$$

3) *обратная связь*, или *встречное параллельное* соединение двух элементов, при котором один из элементов

осуществляет обратную связь для другого. На рис. 5.3 элемент 1 охвачен обратной связью с помощью элемента 2. Выходной сигнал элемента 1 подается на выход системы и вход элемента 2. Выходной сигнал элемента 2 сумми-

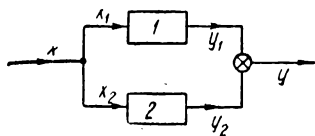


Рис. 5.2. Параллельное соединение

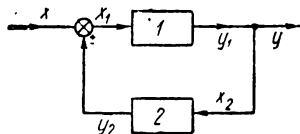


Рис. 5.3. Соединение обратной связи

руется с сигналом, подаваемым на вход системы. Обозначая входные сигналы через x , а выходные — через y , запишем:

$$\begin{aligned} x_2 &= y_1 = y, \\ x_1 &= x \pm y_2, \end{aligned} \quad (5.1)$$

где знаки «+» и «-» соответствуют случаям положительной и отрицательной обратным связям.

Зная передаточную функцию каждого элемента структурной схемы и изображение сигнала, поступающего на

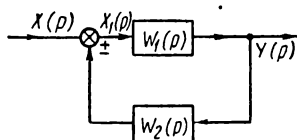


Рис. 5.4. Структурная схема с обратной связью

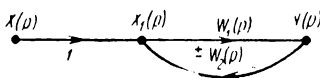


Рис. 5.5. Сигнальный граф, эквивалентный структурной схеме рис. 5.4

его вход, по (4.14) легко найти изображение сигнала, получаемого на выходе элемента. Для этого удобно обозначать элементы структурной схемы их передаточными функциями, а вместо сигналов указывать на схеме их изображения по Лапласу. Так, например, структурная схема, представленная на рис. 5.3, примет вид, изобраа-

женный на рис. 5.4, где

$$X(p) \doteq x(t),$$

$$X_1(p) \doteq x_1(t),$$

$$Y(p) \doteq y(t),$$

а $W_1(p)$ и $W_2(p)$ — передаточные функции элементов 1 и 2.

На структурных схемах такого вида, как показано на рис. 5.4, видны изображения сигналов и преобразую-

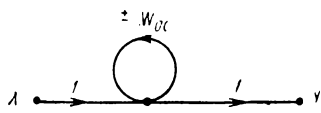


Рис. 5.6. Сигнальный граф, эквивалентный структурной схеме рис. 5.7

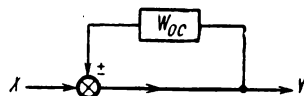


Рис. 5.7. Структурная схема системы с обратной связью

щие их передаточные функции. Те же сведения о рассматриваемой системе могут быть даны в форме *сигнального графа*, представленного на рис. 5.5. Сигнальный граф состоит из *узлов* и *ветвей*. В узлах появляются сигналы, представляемые их изображениями по Лапласу.

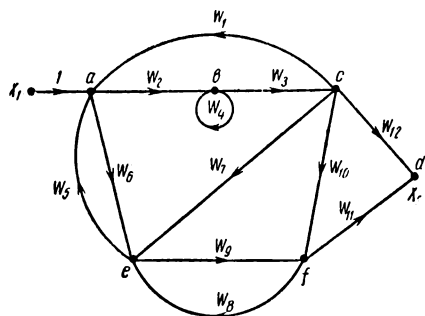


Рис. 5.8. Пример сигнального графа

Поэтому точки разветвления структурных схем, т. е. точки, в которых новые сигналы не образуются, узлами не являются. Ветви представляют собой передаточные функции. Таким образом, пройдя ветвь, начинающуюся в неко-

тором узле (по направлению, обозначенному стрелкой), сигнал получает изображение, определяемое формулой (4.14). Изображения сигналов, сходящихся в какой-либо узел, складываются, образуя изображение нового сигнала, выходящего из этого узла. Передаточная функция ветви обратной связи имеет знак этой связи. На рис. 5.5 и 5.6 изображены сигнальные графы, эквива-

лентные структурным схемам обратной связи, показанным на рис. 5.4 и 5.7. Как видно из сопоставления этих рисунков, сигнальный граф по начертанию проще структурной схемы. Его преимущество, однако, не только в этом, но и в упрощении исследования системы, которое он дает, как будет показано далее. Сигнальный граф представляет собой *топологическую структуру* рассматриваемой системы. Замена структурных схем сигнальными графами позволяет поэтому применять к исследованию автоматических систем топологические методы.

Пример. На рис. 5.8 представлен сигнальный граф системы [5.5], по которому на рис. 5.9 построена эквивалентная структурная схема.

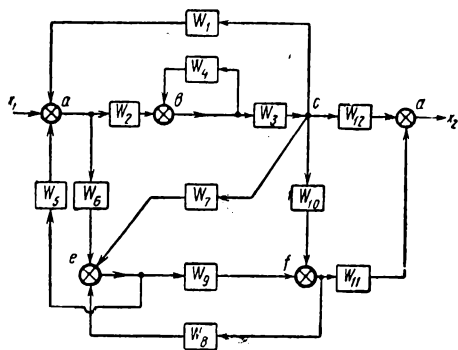


Рис. 5.9. Структурная схема, эквивалентная сигнальному графу рис. 5.8.

На схеме входной и выходной сигналы x_1 и x_2 — оригиналы изображения X_1 и X_2 . Строчными буквами на схеме обозначены точки, соответствующие также обозначенным узлам графа. Точка c схемы соответствует узлу c графа, так как в нем появляется сигнал, возникающий на выходе элемента W_3 .

§ 5.2. СИНТЕЗ ПЕРЕДАТОЧНЫХ ФУНКЦИЙ

Автоматическая система и ее блоки всегда являются соединением того или иного числа простейших элементов. Расчленяя ее на элементы с известными передаточными функциями и составляя структурную схему или сигнальный граф их соединений, легко найти передаточную функцию системы.

Составим передаточные функции систем, получаемых соединением элементов тремя способами, рассмотренными в предыдущем параграфе. При последовательном соеди-



Рис. 5.10. Сигнальный граф системы, состоящей из последовательно соединенных элементов

нении элементов с передаточными функциями W_i , где $i=1, 2, \dots, n$, из сигнального графа (рис. 5.10) непосредственно следует, что передаточная функция системы

$$W = \prod_{i=1}^n W_i, \quad (5.2)$$

т. е. равна произведению передаточных функций всех элементов. При параллельном соединении тех же элементов из сигнального графа (рис. 5.11) также непосредственно следует, что передаточная функция системы

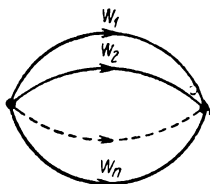


Рис. 5.11. Сигнальный граф системы, состоящей из параллельно соединенных элементов

$$W = \sum_{i=1}^n W_i, \quad (5.3)$$

т. е. равна сумме передаточных функций всех элементов. Для соединения обратной связи из сигнального графа, показанного на рис. 5.5, находим, что изображения X и Y входного и выходного сигналов связаны уравнением

$$Y = W_1 X \pm W_1 W_2 Y. \quad (5.4)$$

Из этого уравнения получаем отношение изображений выходного и входного сигналов, т. е. передаточную функцию системы

$$W = \frac{W_1}{1 \mp W_1 W_2}. \quad (5.5)$$

Заметим, что в этой формуле верхний знак относится к положительной обратной связи, а нижний — к отрицательной.

Уравнение (5.4) выражает суперпозицию реакций элемента, имеющего передаточную функцию W_1 , на входной сигнал и сигнал обратной связи. Поэтому этот элемент предполагается линейным. В дальнейшем будут предполагаться линейными все элементы, на вход которых поступает более одного сигнала, если только не будет рассматриваться особо оговоренный случай нелинейного устройства.

Передаточную функцию соединения обратной связи, представленного на рис. 5.6 и 5.7, получим из (5.5), положив $W_1 = 1$, а $W_2 = \pm W_{o.c.}$:

$$W = \frac{1}{1 \mp W_{o.c.}}. \quad (5.6)$$

Передаточная функция любой ветви сигнального графа, представленного на рис. 5.4, может принадлежать неко-

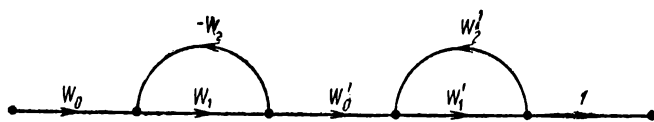


Рис. 5.12. Сигнальный граф, составленный из элементарных графов

торому соединению того или иного числа элементов. Заменяя ветви графами этих соединений, получим сигнальный граф более сложной системы. В ней будет, очевидно, n путей прохождения сигнала от входа к выходу, каждый из которых характеризуется передаточной функцией W_i . Подобно этому и в других сложных системах может быть несколько *прямых путей* прохождения сигнала от входа к выходу, на которых один и тот же узел встречается не более одного раза. Составив передаточные функции этих путей, легко получить передаточную функцию всей системы по (5.3). Если передаточную функцию какого-либо пути нельзя составить непосредственно пользуясь формулами (5.2), (5.3) и (5.5), то ее всегда можно найти из *узловых уравнений* этого пути, выражающих связь между изображениями сигналов, входящих в каждый узел, и изображением сигнала, выходящего из этого узла. Поясним примерами эти методы составления передаточных функций.

На рис. 5.12 представлен сигнальный граф системы, которая является последовательным соединением более простых систем. Ее сигнальный граф состоит, соответственно, из последовательного соединения двух *элементарных графов* вида, показанного на рис. 5.13, и ветви с передаточной функцией, равной единице. Последняя на передаточную функцию системы не влияет,



Рис. 5.13. Элементарный сигнальный граф

а передаточные функции элементарных графов могут быть легко написаны по (5.5) и (5.2). Эти передаточные функции имеют вид $W_0 W_1 / (1 + W_1 W_2)$.

Следовательно, передаточная функция рассматриваемой системы

$$W = \frac{W_0 W'_0 W_1 \bar{W}_1}{(1 + W_1 W_2)(1 + W'_1 W'_2)}.$$

На рис. 5.14 представлен сигнальный граф системы, передаточную функцию которой нельзя получить непо-

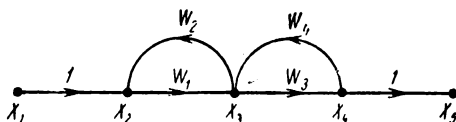


Рис. 5.14. К получению передаточной функции из узловых уравнений

средственным применением формул (5.2), (5.3) и (5.5). Сделаем это с помощью узловых уравнений, имеющих вид

$$\begin{aligned} X_2 &= X_1 + W_2 X_3, \\ X_3 &= W_1 X_2 + W_4 X_4, \\ X_4 &= W_3 X_3, \\ X_5 &= X_4. \end{aligned} \quad (5.7)$$

Эти уравнения, как и все последующие, написаны в общей форме, в которой знаки обратной связи скрыты в передаточных функциях. Из (5.7) легко находим, что отношение изображений X_5 и X_1 , т. е. передаточная функция рассматриваемой системы

$$W = \frac{W_1 \cdot W_3}{1 - W_1 W_2 - W_3 W_4}. \quad (5.8)$$

Сигнальный граф более сложной системы представлен на рис. 5.15 [5.6]. В ней имеется шесть прямых путей прохождения сигнала от входа к выходу, которые запишем, указывая передаточные функции образующих их ветвей:

- 1) W_1, W_3, W_4 ;
- 2) W'_1, W'_3, W'_4 ;
- 3) W_1, W_5, W'_4 ;
- 4) W'_1, W_6, W_4 ;
- 5) W_1, W_5, W'_3, W_6, W_4 ;
- 6) $W'_1, W_6, W_3, W_5, W'_4$.

Передаточную функцию первого и пятого путей получим, рассматривая часть сигнального графа, показанную на рис. 5.16, а передаточную функцию третьего пути — по рис. 5.17.

Составим передаточную функцию той части системы, которая представлена на рис. 5.16. Для этого будем постепенно упрощать сигнальный граф. Пользуясь (5.5), заменим сначала ветви W'_3 и W'_2 одной ветвью с передаточной функцией

$$W'_7 = \frac{W'_3}{1 - W'_2 W'_3}$$

(рис. 5.18). Затем с помощью (5.2) заменим ветви W_5 , W'_7 и W_6 ветвью с передаточной функцией

$$W_8 = W_5 W'_7 W_6$$

(рис. 5.19). Теперь заменим ветви W_2 и W_8 одной ветвью, имеющей находимую по (5.3) передаточную функцию

$$W_9 = W_2 + W_8$$

(рис. 5.20). Далее, вновь пользуясь (5.5), заменим ветви W_3 и W_9 ветвью с передаточной функцией

$$W_{10} = \frac{W_9}{1 - W_3 W_9}$$

(рис. 5.21). И, наконец, по (5.2) находим передаточную функцию рассматриваемых параллельно соединенных путей прохождения сигнала

$$W_{11} = W_1 \cdot W_{10} \cdot W_4.$$

Подставляя сюда выражения передаточных функций W_{10} , W_9 , W_8 и W'_7 , после упрощений находим сумму передаточных функций

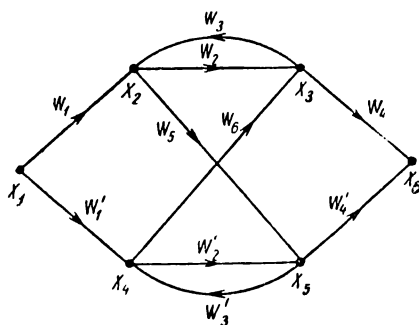


Рис. 5.15. Сигнальный граф сложной системы

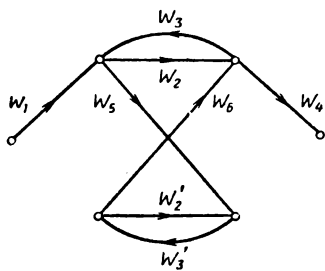


Рис. 5.16. Первый и пятый пути прохождения входного сигнала из графа рис. 5.15

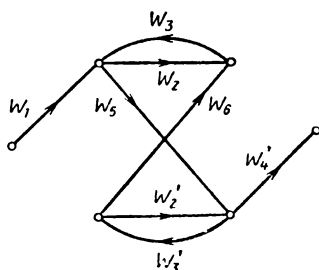


Рис. 5.17. Третий путь прохождения входного сигнала из графа рис. 5.15

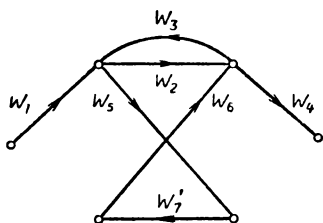


Рис. 5.18. Упрощение сигнального графа рис. 5.16

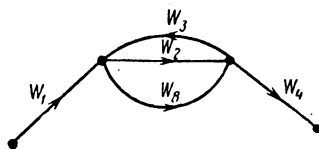


Рис. 5.19. Упрощение сигнального графа рис. 5.18

точных функций первого и пятого путей:

$$W_{11} = W_1 W_4 \frac{W_2 (1 - W_2' W_3') + W_3 W_5 W_6}{(1 - W_2 W_3) (1 - W_2' W_3') - W_3 W_3' W_5 W_6}.$$

Передаточная функция W'_{11} параллельно соединенных второго и шестого путей получается из последней формулы заменой передаточных функций, обозначенных штри-

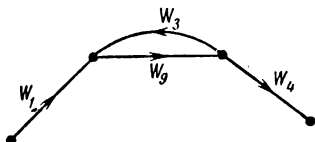


Рис. 5.20. Упрощение сигнального графа рис. 5.19

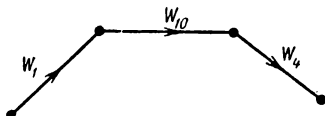


Рис. 5.21. Упрощение сигнального графа рис. 5.20

хами, функциями с одноименными индексами, но без штрихов, и наоборот. Это видно из сопоставления рис. 5.16 и 5.22. Таким образом,

$$W'_{11} = W'_1 W'_4 \frac{W_2 (1 - W_2 W_3) + W_3 W_5 W_6}{(1 - W_2 W_3) (1 - W_2' W_3') - W_3 W_3' W_5 W_6}.$$

Найдем теперь передаточную функцию третьего пути, постепенно упрощая сигнальный граф, представленный

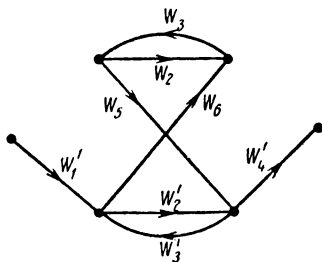


Рис. 5.22. Второй и шестой пути прохождения входного сигнала из графа рис. 5.15

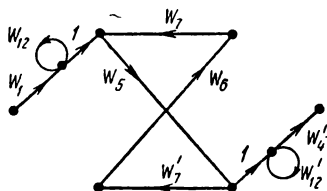


Рис. 5.23. Преобразование сигнального графа рис. 5.17

на рис. 5.17, подобно тому, как это сделано выше для первого и пятого путей. Сначала введем ветви с передаточными функциями: уже найденной W'_7 и $W_7 = \frac{W_3}{1 - W_2 W_3}$ (рис. 5.23). При этом следует ввести показанные на рисунке петли обратной связи, имеющие передаточные

функции

$$W_{12} = W_2 W_3,$$

$$W'_{12} = W'_2 W'_3.$$

Действительно, ветви W_7 и W'_7 находятся в линии обратной связи, охватывающей ветвь W_8 , и заменяют контуры, в которых сигнал последовательно проходит соответственно по ветвям W_3 , W_2 и W'_3 , W'_2 . Для прямого же сигнала, проходящего по рассматриваемому пути от

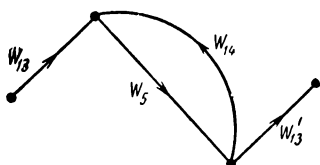


Рис. 5.24. Упрощение сигнального графа рис. 5.23

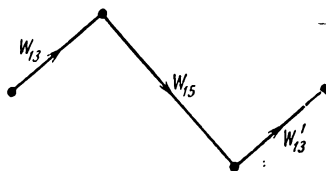


Рис. 5.25. Упрощение сигнального графа рис. 5.24

входа до выхода, в тех же контурах обратной связи сигнал последовательно проходит по ветвям W_3 , W_2 и W'_3 , W'_2 . Пользуясь (5.6), заменим теперь ветви W_1 и W'_4 с петлями W_{12} и W'_{12} ветвями с передаточными функциями

$$W_{13} = \frac{W_1}{1 - W_{12}},$$

$$W'_{13} = \frac{W'_4}{1 - W'_{12}}.$$

Наряду с этим заменим ветви W'_7 , W_6 и W_7 ветвью с передаточной функцией

$$W_{14} = W'_7 W_6 W_7$$

(рис. 5.24). Далее заменим ветви W_5 и W_{14} ветвью с передаточной функцией

$$W_{15} = \frac{W_5}{1 - W_5 W_{14}}$$

(рис. 5.25). Тогда передаточная функция рассматриваемого пути

$$W_{16} = W_{13} W_{15} W'_{13},$$

откуда после подстановки предыдущих выражений

$$W_{16} = \frac{W_1 W_5 W'_4}{(1 - W_2 W_3)(1 - W'_2 W'_3) - W_3 W'_3 W_5 W'_5}.$$

Заменяя в этом выражении передаточные функции, обозначенные штрихами, функциями с одноименными индексами, но без штрихов, и наоборот, получим передаточную функцию W'_{16} четвертого пути. Таким образом, как видно из сопоставления рис. 5.17 и 5.26,

$$W'_{16} = \frac{W'_1 W'_6 W_4}{(1 - W_2 W_3)(1 - W'_2 W'_3) - W_3 W'_3 W_5 W_6}.$$

Теперь легко написать передаточную функцию всей рассматриваемой системы (см. рис. 5.15)

$$W = W'_{11} + W_{11} + W'_{16} + W_{16}.$$

Подставляя сюда полученные выше выражения слагаемых правой части, находим:

$$\begin{aligned} W = & \{W_1 W_4 [W_2 (1 - W'_2 W'_3) + W'_3 W_5 W_6] + \\ & + W'_1 W'_4 [W'_2 (1 - W_2 W_3) + W_3 W_5 W_6] + \\ & + W_1 W_5 W'_4 + W'_1 W_6 W_4\} : [(1 - W_2 W_3)(1 - W'_2 W'_3) - \\ & - W_3 W'_3 W_5 W_6]. \end{aligned} \quad (5.9)$$

Тот же результат может быть получен из узловых уравнений. Вводя показанные на рис 5.15 изображения сигналов, легко составить следующую систему узловых уравнений:

$$X_2 = W_1 X_1 + W_3 X_3,$$

$$X_3 = W_2 X_2 + W_6 X_4,$$

$$X_4 = W'_1 X_1 + W'_3 X_5,$$

$$X_5 = W'_2 X_4 + W_5 X_2,$$

$$X_6 = W_4 X_3 + W'_4 X_5.$$

Передаточная функция (5.9) получается из этих уравнений как отношение X_6/X_1 .

В общем случае передаточная функция линейной системы определяется решением аналогичной системы линейных уравнений. Поэтому передаточную функцию можно записать формулой, подобной той, которой выражается решение системы линейных уравнений по способу Крамера. Если в системе имеется k прямых путей прохождения сигнала от входа до выхода, то по *формуле Мей-*

сона [5.1] ее передаточная функция

$$W = \frac{1}{\Delta} \sum_{i=1}^k G_i \Delta_i, \quad (5.10)$$

где G_i — передаточная функция i -го пути, составленная без учета имеющихся в нем обратных связей, т. е. произведения передаточных функций ветвей сигнального графа, образующих этот путь;

Δ — определитель сигнального графа системы;

Δ_i — адъюнкт прямого пути i .

Определители Δ и Δ_i выражаются через передаточные функции имеющихся в системе контуров, состоящих из

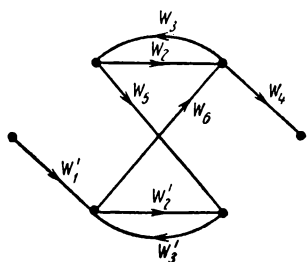


Рис. 5.26. Четвертый путь прохождения входного сигнала из графа рис. 5.15

ветвей, образующих замкнутый путь прохождения сигнала. Контур называется некасающимся, если он не содержит общих узлов. Определитель Δ равен единице минус сумма произведений передаточных функций всех возможных сочетаний по нечетным числам некасающихся контуров плюс сумма произведений передаточных функций всех возможных сочетаний по четным числам некасающихся контуров¹.

Определитель Δ_i составляется также для той части системы, которая не касается i -го прямого пути, т. е. не имеет с ним общих узлов.

Применяя (5.10) к системе, сигнальный граф которой представлен на рис. 5.15 для рассмотренных выше путей, имеем:

$$G_1 = W_1 W_2 W_4, \quad G_2 = W'_1 W'_2 W'_4, \quad G_3 = W_1 W_5 W'_4, \\ G_4 = W'_1 W_6 W_4, \quad G_5 = W_1 W_5 W'_3 W_6 W_4, \quad G_6 = W'_1 W_6 W_3 W_5 W'_4.$$

Система имеет два некасающихся контура с введенными выше передаточными функциями W_{12} и W'_{12} , а также контур с передаточной функцией $W_{17} = W_3 W_5 W'_3 W_6$. Поэтому

¹ В сумму членов, содержащих по одной передаточной функции, входят передаточные функции всех контуров.

в определителе Δ может быть три члена, представляющих собой сочетания по одному контуру (W_{12} , W'_{12} и W_{17}), и один член, представляющий сочетание двух контуров (W_{12} , W'_{12}); никаких других членов у него не может быть. Таким образом,

$$\Delta = 1 - W_2 W_3 - W'_2 W'_3 - W_3 W_8 W'_3 W_6 + W_2 W_3 W'_2 W'_3.$$

Легко видеть, что это — знаменатель выражения (5.9), в котором раскрыты скобки.

Для первого пути имеется некасающийся контур W'_{12} , а для второго — контур W_{12} . Следовательно, адьюнкты этих путей

$$\Delta_1 = 1 - W'_2 W'_3,$$

$$\Delta_2 = 1 - W_2 W_3.$$

Для остальных путей некасающихся контуров нет, так что

$$\Delta_3 = \Delta_4 = \Delta_5 = \Delta_6 = 1.$$

Подстановка полученных выражений передаточных функций G и определителей Δ в (5.10) дает уже выведенную передаточную функцию (5.9).

Таким образом, синтез передаточных функций сложных систем из известных передаточных функций элементов этих систем можно производить, пользуясь сигнальными графами, тремя способами:

- 1) последовательным упрощением графа по (5.2), (5.3) и (5.5);
- 2) составлением и решением узловых уравнений;
- 3) по (5.10).

Для первого способа требуется подробный анализ обратных связей, существующих в системе, но в некоторых случаях, применяя его, приходится частично пользоваться и вторым. Второй способ полезен для проверки результатов, получаемых двумя другими, однако для сложных систем громоздок. Третий способ наиболее прост и удобен, но формален.

Задача синтеза передаточных функций с помощью сигнальных графов существенно усложняется при наличии всех возможных связей между узлами графа. АУС, обладающие такой структурой, в общей форме описы-

ваются системой линейных уравнений вида

$$Dx_i = \sum_{j=1}^n a_{ij}x_j. \quad (5.11)$$

В подобную систему линейных уравнений может быть преобразовано уравнение движения системы. В этом случае все ветви графа характеризуются передаточными функциями вида $W_i = \frac{1}{p}$ либо коэффициентами передачи a_{ij} . Физически такую систему можно представить, например, в виде совокупности n интегрирующих блоков, всеми возможными способами связанных между собой. Наличие большого числа связей затрудняет даже для систем низкого порядка полное и правильное выделение всех прямых путей и контуров сигнального графа, необходимых для применения формулы (5.10), что снижает эффективность существующих методов выделения прямых путей и контуров, например, с помощью принципиальной контурной диаграммы [5.3].

Однако процесс выделения контуров и прямых путей в сигнальных графах систем описываемого типа может быть значительно облегчен и даже в некоторой степени формализован за счет введения некоторых дополнительных понятий и применения *ветвящихся диаграмм*. В качестве одного из таких понятий введем термин *порядок контура* (контур первого, второго и т. д. порядка), под которым понимается число ветвей с передаточной функцией $W_i = \frac{1}{p}$, входящих в исследуемый контур. По аналогии введем понятие *порядок пути*, вкладывая в него тот же смысл. Из определения следует, что максимальный порядок контуров и прямых путей равняется n .

Уравнение (5.11) не определяет места приложения входного и выходного воздействий, поэтому для синтеза передаточной функции W_{ij} будем считать, что вход и выход системы определяются свободным выбором исследователя (здесь i определяет место приложения входного воздействия, а j — выходного). В частном случае может быть $i = j$.

Совершенно очевидно, что сигнальный граф системы (5.11) содержит n контуров первого порядка, образуемых интегрирующими блоками и соответствующими собствен-

ными обратными связями. Число контуров второго порядка N_2 определяется числом сочетаний C_n^2 из n блоков по два:

$$N_2 = C_n^2 = \frac{n!}{2!(n-2)!}.$$

Сложнее подсчитать количество контуров более высокого порядка, поскольку комбинации из одних и тех же блоков системы могут образовывать различные контуры. Вместе с тем для подсчета количества таких контуров нельзя воспользоваться известными формулами для определения числа перестановок, так как не всякие последовательности блоков, различающиеся порядком следования,

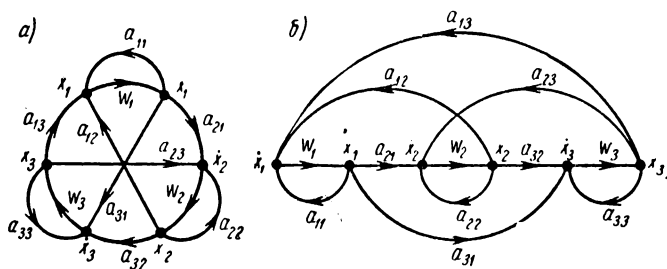


Рис. 5.27. Сигнальные графы системы (5.11) третьего порядка

являются новыми контурами. Например, последовательности $A-B-C$, $B-C-A$ и $C-A-B$ представляют собой один и тот же контур, а последовательность $A-C-B$ определяет новый контур. В дальнейшем перестановки, соответствующие различающимся контурам (например, $A-B-C$ и $A-C-B$), будем называть *контурными перестановками*.

При подсчете контуров третьего порядка следует иметь в виду, что совокупность из трех интегрирующих блоков может образовывать только два различных контура. Так, из рис. 5.27, где представлены сигнальные графы системы (5.11) при $n=3$ в двух видах, ясно, что могут быть выделены контуры третьего порядка $W_1 a_{21} W_2 a_{32} W_3 a_{13}$ и $W_1 a_{31} W_3 a_{23} W_2 a_{12}$. Кроме того, число таких совокупностей по три блока для системы n -го порядка определяется числом сочетаний C_n^3 из n элементов по три. Общее коли-

чество контуров третьего порядка

$$N_3 = C_n^3 \cdot 2 = \frac{n!}{3(n-3)!}.$$

Для определения общего числа контуров четвертого порядка целесообразно построить ветвящуюся диаграмму, более наглядно представляющую возможные пути передачи сигналов и позволяющую легко найти количество контурных перестановок. На рис. 5.28 показана ветвящаяся диаграмма системы (5.11) при $n=4$. Замечая, что

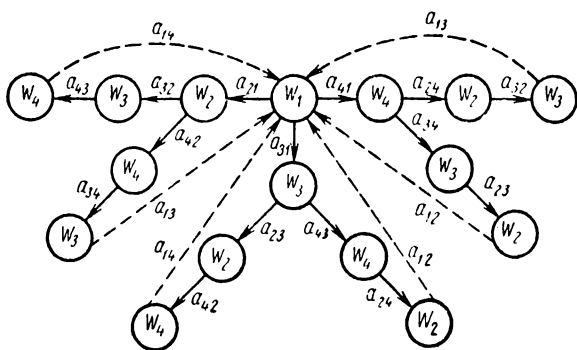


Рис. 5.28. Ветвящаяся диаграмма для выделения контуров

каждый блок рассматриваемой совокупности из четырех блоков должен входить в каждый контур и притом только один раз, можно на ветвящейся диаграмме рассмотреть все возможные пути прохождения сигнала по замкнутому контуру четвертого порядка по отношению к любому из четырех блоков, как к исходному (на рис. 5.28 это блок W_1). Из рис. 5.28 видно, что каждой совокупности из четырех блоков W_i соответствует шесть контурных перестановок (шесть контуров четвертого порядка). Следовательно, для системы n -го порядка число контуров четвертого порядка

$$N_4 = C_n^4 \cdot 6 = \frac{n!}{4(n-4)!}.$$

Аналогичным образом можно применить ветвящуюся диаграмму для определения контурных перестановок в совокупностях из m блоков: выходной сигнал произволь-

ного блока подается на $(m - 1)$ остальных блоков, с выхода каждого из которых сигнал передается на $(m - 2)$ блоков. Такое ветвление возможных путей передачи сигналов продолжается до тех пор, пока в каждой ветви не останется единственный путь — через последний оставшийся блок к исходной точке. Таким образом, количество контурных перестановок в совокупности из m блоков составляет $(m - 1)!$, откуда общее число контуров m -го порядка в системе (5.11) можно определить по формуле

$$N_m = C_n^m (m - 1)! = \frac{n!}{m(n - m)!}. \quad (5.12)$$

При этом из ветвящейся диаграммы легко получить передаточные функции контуров путем умножения произведения всех коэффициентов a_{ij} , встречающихся в данном контуре, на величину $\left(\frac{1}{p}\right)^m$.

Таким образом, из изложенного следует, что общее число контуров в сигнальном графе системы (5. 11)

$$\begin{aligned} N = n + C_n^2 + C_n^3 \cdot 2! + C_n^4 \cdot 3! + \dots \\ \dots + C_n^m (m - 1)! + \dots + (n - 1)! = n + \frac{n!}{2(n - 2)!} + \\ + \frac{n!}{3(n - 3)!} + \dots + \frac{n!}{m(n - m)!} + \dots + (n - 1)! \end{aligned} \quad (5.13)$$

Ветвящиеся диаграммы позволяют аналогичным образом выделить и все прямые пути между рассматриваемыми узлами. При этом совершенно очевидно, что при определении передаточной функции W_{ii} сигнальный граф имеет единственный прямой путь, характеризующийся преобразованием W_i . При отыскании прямых путей для передаточной функции W_{ij} ветвящаяся диаграмма строится таким образом, что каждый из путей, начинаясь в i -м блоке, продолжается до достижения j -го блока, причем из каждого очередного блока сигнал направляется на все блоки, не встречавшиеся до этого в рассматриваемом пути. Рис. 5.29 иллюстрирует процесс построения ветвящейся диаграммы для выделения прямых путей на примере системы пятого порядка, в которой входным является третий блок, выходным — пятый. Из рисунка видно, что при синтезе передаточной функции W_{ij} прямые пути первого порядка будут отсутствовать, а прямой путь второго порядка оказывается единственным. Из рис. 5.29

видно, что количество выходов любого из блоков диаграммы должно быть на единицу меньше количества входов предыдущего блока, с которым связан вход рас-

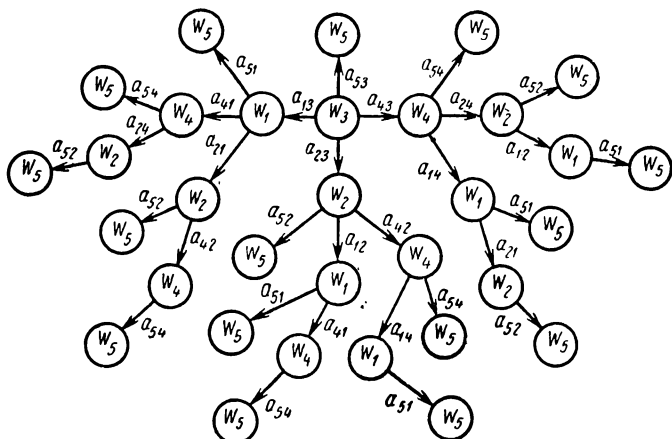


Рис. 5.29. Ветвящаяся диаграмма для выделения прямых путей

сматриваемого блока. Отсюда следует, что общее количество прямых путей M в сигнальном графе n -го порядка определяется формулой

$$M = 1 + (n-2) + (n-2)(n-3) + \dots + (n-2)(n-3)(n-4) + \dots + (n-2)(n-3) \dots [n - (n-2)] + (n-2)(n-3) \dots [n - (n-1)]. \quad (5.14)$$

Количество прямых путей n -го и $(n-1)$ -го порядков совпадают, поэтому выражение (5.14) принимает вид

$$M = 1 + (n-2) + (n-2)(n-3) + \dots + 2(n-2)(n-3) \dots [n - (n-2)]. \quad (5.15)$$

При выявлении некасающихся пар, троек и т. д. контуров полезным оказывается следующее необходимое условие непересечения контуров: чтобы контуры рассматриваемой пары, тройки, четверки и т. д. были некасающимися, необходимо, чтобы сумма порядков контуров рассматриваемой группы не превышала порядка исследуемой системы. Например, для двух контуров это условие

выражается неравенством

$$m + k \leq n, \quad (5.16)$$

где m и k — порядок контуров, входящих в рассматриваемую пару.

Все контуры первого порядка в рассматриваемых графах являются некасающимися, поэтому они могут образовывать некасающиеся группы из любого числа контуров, не превышающего порядок системы n . Количество пар, троек и т. д. некасающихся контуров определяется в этом случае числом сочетаний из n элементов соответственно по два, три и т. д.

При составлении групп некасающихся контуров признаком их несовпадения является условие несовпадения индексов в элементах передаточных функций отдельных контуров, составляющих группу.

Для иллюстрации изложенного составим передаточные функции W_{11} и W_{13} системы третьего порядка. Сигнальный граф (рис. 5.27, а) содержит следующие контуры:

- 1) $W_1 a_{11}$; 2) $W_2 a_{22}$; 3) $W_3 a_{33}$; 4) $W_1 a_{31} W_3 a_{13}$;
- 5) $W_1 a_{21} W_2 a_{12}$; 6) $W_3 a_{23} W_2 a_{32}$; 7) $W_1 a_{21} W_2 a_{32} W_3 a_{13}$;
- 8) $W_1 a_{31} W_3 a_{23} W_2 a_{12}$.

Находим определитель сигнального графа:

$$\begin{aligned} \Delta &= 1 - W_1 a_{11} - W_2 a_{22} - W_3 a_{33} - W_1 a_{31} W_3 a_{13} - W_1 a_{21} W_2 a_{12} - \\ &\quad - W_3 a_{23} W_2 a_{32} - W_1 a_{21} W_2 a_{32} W_3 a_{13} - W_1 a_{31} W_3 a_{23} W_2 a_{12} + \\ &\quad + W_1 a_{11} W_2 a_{22} + W_1 a_{11} W_3 a_{33} + W_2 a_{22} W_3 a_{33} + W_1 a_{11} W_2 a_{32} W_3 a_{23} + \\ &\quad + W_2 a_{22} W_1 a_{31} W_3 a_{13} + W_3 a_{33} W_1 a_{21} W_2 a_{12} - W_1 a_{11} W_2 a_{22} W_3 a_{33} = \\ &= 1 + (-a_{11} - a_{22} - a_{33}) \frac{1}{p} + (a_{11} a_{22} + a_{11} a_{33} + a_{22} a_{33} - \\ &\quad - a_{31} a_{13} - a_{21} a_{12} - a_{23} a_{32}) \left(\frac{1}{p}\right)^2 + (a_{11} a_{32} a_{23} + a_{22} a_{31} a_{13} + \\ &\quad + a_{33} a_{21} a_{12} - a_{21} a_{32} a_{13} - a_{31} a_{23} a_{12} - a_{11} a_{22} a_{33}) \left(\frac{1}{p}\right)^3. \quad (5.17) \end{aligned}$$

Поскольку при определении W_{11} имеем единственный прямой путь $W_1 = 1/p$, то адъюнкт Δ_1 будет:

$$\begin{aligned} \Delta_1 &= 1 - W_2 a_{22} - W_3 a_{33} - W_2 a_{32} W_3 a_{23} + W_2 a_{22} W_3 a_{33} = \\ &= 1 + (-a_{22} - a_{33}) \frac{1}{p} + (a_{22} a_{33} - a_{23} a_{32}) \left(\frac{1}{p}\right)^2. \quad (5.18) \end{aligned}$$

Тогда по (5.10)

$$W_{11} = \frac{W_1 \Delta_1}{\Delta} = [p^3 + (-a_{22} - a_{33})p + a_{22}a_{33} - a_{23}a_{32}] : [p^3 + (-a_{11} - a_{22} - a_{33})p^2 + (a_{11}a_{22} + a_{11}a_{33} + a_{22}a_{33} - a_{31}a_{13} - a_{21}a_{12} - a_{23}a_{32})p + a_{11}a_{32}a_{23} + a_{22}a_{31}a_{13} + a_{33}a_{21}a_{12} - a_{21}a_{32}a_{13} - a_{31}a_{23}a_{12} - a_{11}a_{22}a_{33}]. \quad (5.19)$$

Адьюнкт Δ_k получается из определителя Δ удалением k -го пути, а поскольку в рассматриваемом случае имеет место единственный прямой путь W_i , то порядок числителя передаточной функции W_{ii} всегда будет на единицу меньше порядка знаменателя. Следовательно, если найден определитель Δ , то числитель передаточной функции W_{ii} может быть получен из Δ формальным образом в результате понижения порядка Δ на единицу и удаления членов, содержащих сомножителями элементы с индексом i .

При синтезе передаточной функции W_{13} выделяем два прямых пути (см. рис. 5.27, б): $W_1 a_{31} W_3$ и $W_1 a_{21} W_2 a_{32} W_3$. По (5.10)

$$W_{13} = \frac{W_1 a_{31} W_3 (1 - W_2 a_{23})}{\Delta} + \frac{W_1 a_{21} W_2 a_{32} W_3}{\Delta} = [a_{31}(p - a_{22}) + a_{21}a_{32}] : [p^3 + (-a_{11} - a_{22} - a_{33})p^2 + (a_{11}a_{22} + a_{11}a_{33} + a_{22}a_{33} - a_{31}a_{13} - a_{21}a_{12} - a_{23}a_{32})p + a_{11}a_{32}a_{23} + a_{22}a_{31}a_{13} + a_{33}a_{21}a_{12} - a_{21}a_{32}a_{13} - a_{31}a_{23}a_{12} - a_{11}a_{22}a_{33}]. \quad (5.20)$$

Минимальный порядок прямых путей при синтезе передаточных функций W_{ij} равняется двум, поэтому порядок числителя W_{ij} будет $(n - 2)$.

Изложенный способ синтеза передаточных функций систем вида (5.11), хотя и требует значительного количества вычислений, носящих элементарный характер, позволяет с помощью формул (5.13) и (5.15) контролировать правильность выделения контуров и прямых путей сигнальных графов, причем сам процесс их выделения с помощью ветвящихся диаграмм и условий несоприкосновения оказывается достаточно простым.

Пример. Пользуясь формулой Мейсона, составим передаточную функцию системы, сигнальный граф которой представлен на рис. 5.8. Из рисунка видно, что сигнал может проходить от входа до выхода

по четырем прямым путям, передаточные функции которых

$$G_1 = W_2 W_8 W_{12},$$

$$G_2 = W_2 W_3 W_{10} W_{11},$$

$$G_3 = W_2 W_3 W_7 W_9 W_{11},$$

$$G_4 = W_6 W_9 W_{11}.$$

Адьюнкты этих путей

$$\Delta_1 = 1 - W_8 W_9,$$

$$\Delta_2 = 1,$$

$$\Delta_3 = 1,$$

$$\Delta_4 = 1 - W_4.$$

Определитель

$$\Delta = 1 - W_1 W_2 W_8 - W_4 - W_5 W_6 - W_8 W_9 - W_2 W_3 W_7 W_9 - \\ - W_2 W_3 W_5 W_8 W_{10} + W_1 W_2 W_3 W_8 W_9 + W_4 W_5 W_6 + W_4 W_8 W_9.$$

По формуле (5.10) [5.5] легко написать искомую передаточную функцию.

§ 5.3. ТОЖДЕСТВЕННЫЕ ПРЕОБРАЗОВАНИЯ СТРУКТУРНЫХ СХЕМ

Преобразования сигнальных графов, произведенные в предыдущем параграфе, являются примерами *структурного преобразования*, при котором изменяется схема, но не изменяется передаточная функция системы. При таком преобразовании возможно также переключение обратных связей и перенесение входов из одного места в другое.

В структурной схеме системы, состоящей из линейных элементов, можно переносить сигналы, подаваемые на вход тех или иных элементов, в другие точки схемы. Если, например, сигнал $x_1(t)$ подан на вход линейного элемента, имеющего передаточную функцию $W(p)$ (рис. 5.30, а), то его можно перенести на выход того же элемента (рис. 5.30, б), заменив сигналом

$$x_2(t) \doteq W(p) X_1'(p),$$

где

$$X_1(p) \doteq x_1(t).$$

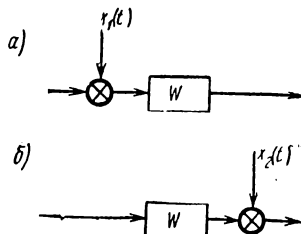


Рис. 5.30. Перенесение сигнала со входа на выход

Этот прием перенесения входных сигналов в структурной схеме приводит к возможности *переключения обратной*

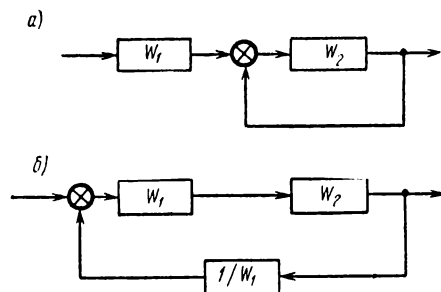


Рис. 5.31. Переключение единичной обратной связи

ных связей с преобразованием их передаточных функций. Так, например, обратную связь, охватывающую элемент W_2 и имеющую передаточную функцию, равную единице

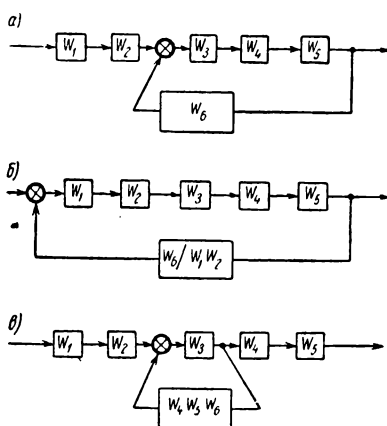


Рис. 5.32. Два случая переключения обратной связи

(рис. 5.31, а), можно переключить на вход элемента W_1 , введя в нее элемент с передаточной функцией $1/W_1$ (рис. 5.31, б). При этом, очевидно, на вход элемента W_2 будет поступать тот же сигнал, что и в предыдущем случае.

В общем случае переключение обратной связи приводит к увеличению или уменьшению количества охватываемых ею элементов прямого пути прохождения сигнала. Обобщая уже

сделанный вывод, легко видеть, что при этом передаточную функцию цепи обратной связи следует делить на произведение передаточных функций элементов прямого пути, вводимых в контур, или умножать на произведение передаточных функций элементов прямого пути, исклю-

чаемых из контура [5.2]. Применение этого правила поясняет рис. 5.32, где показано соединение с обратной связью (рис. 5.32, а) и ее переключение с увеличением (рис. 5.32, б) и уменьшением (рис. 5.32, в) числа охваченных ею элементов. На рисунке указано, как должна при этом изменяться передаточная функция цепи обратной связи, чтобы передаточная функция всего соединения в целом не изменялась.

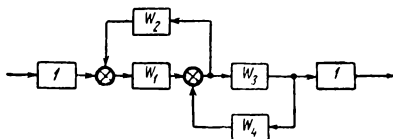


Рис. 5.33. Структурная схема, эквивалентная сигнальному графу по рис. 5.14

Таким образом можно преобразовать структурную схему с *перекрещивающимися обратными связями* к виду,

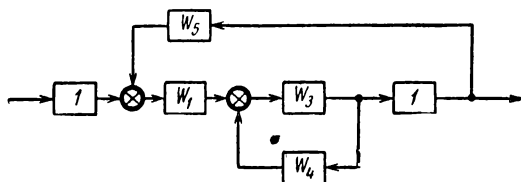


Рис. 5.34. Пересключение обратной связи в структурной схеме по рис. 5.30

в котором перекрещивающихся связей не будет. После такого преобразования передаточную функцию системы легко составить по структурной схеме. К нелинейным системам такое преобразование применять нельзя.

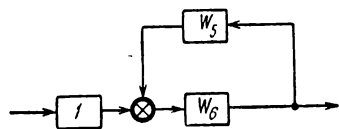


Рис. 5.35. Упрощение структурной схемы по рис. 5.31

На рис. 5.33 представлена структурная схема, эквивалентная сигнальному графу по рис. 5.14. На этой

схеме имеются перекрещивающиеся обратные связи. Устраним их, переключив одну из обратных связей, как показано на рис. 5.34. Согласно приведенному выше правилу, передаточная функция, которую должна иметь переключенная цепь обратной связи,

$$W_5 = \frac{W_2}{W_3}.$$

Заменяем теперь часть системы, охваченную этой обратной связью, элементом с передаточной функцией

$$W_6 = \frac{W_1 W_3}{1 - W_3 W_4}$$

(рис. 5.35). Тогда передаточная функция всей системы

$$W = \frac{W_6}{1 - W_5 W_6},$$

откуда после подстановки выражений W_5 и W_6 получим формулу (5.8).

Пример. Составим передаточную функцию системы, структурная схема которой представлена на рис. 5.36. Тем же способом,

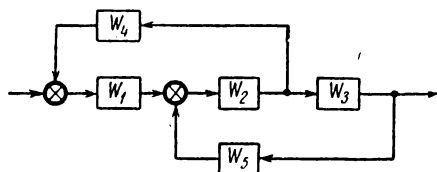


Рис. 5.36. Пример структурной схемы с перекрещивающимися обратными связями

что и для рис. 5.33, найдем выражение искомой передаточной функции:

$$W = \frac{W_1 W_2 W_3}{1 - W_1 W_2 W_4 - W_2 W_3 W_5}.$$

§ 5.4. ТИПОВЫЕ И ЭЛЕМЕНТАРНЫЕ ЗВЕНЬЯ

Каждый элемент автоматической системы характеризуется численными значениями некоторых параметров — физических величин, определяющих его свойства, как и свойства всей системы (см. § 4.2). Если стремиться к всестороннему описанию состояния элемента, не упуская никаких явлений, от которых оно зависит, то количество таких значений может стать бесконечно большим. Таким же станет и число обобщенных координат, определяющих состояние элемента. Рассматривая, например, скручиваемый вал, нужно было бы определять численное значение координаты в виде угловой деформации в каждой точке его длины, принимая во внимание, что жесткость вала (момент на единицу угловой деформации)

может быть параметром, изменяющимся по его длине. Это будет элемент с *распределенными параметрами*. Изменение его состояния, т. е. его движение, описывается уравнениями в частных производных. Обычно число параметров и обобщенных координат может быть уменьшено, если пренебречь второстепенными для решаемой задачи факторами и явлениями. Наименьшее число обобщенных координат, которое необходимо учитывать, при этом станет конечным, определяя число степеней свободы элемента в рассматриваемой задаче. Это будет элемент с сосредоточенными параметрами, движение которого описывается обыкновенными уравнениями в обобщенных производных (см. § 4.1). Например, можно рассматривать угловую деформацию между концами вала, учитывая жесткость вала в целом и сводя число степеней свободы к единице. То же может быть и в более сложных случаях. Так, иногда рассматривают угловую деформацию гидропередачи, пренебрегая всеми гидродинамическими явлениями, если искомый результат от них мало зависит. В этом состоит инженерный подход к исследованию движения автоматических систем. Он требует умения правильно выбирать число степеней свободы, которое следует принимать во внимание.

При исследовании АУС обычно необходимо рассматривать несколько степеней свободы. Но систему можно расчленить на *звенья*, состояние каждого из которых характеризуется одной обобщенной координатой на выходе, т. е. имеет одну степень свободы. Звено называют *элементарным*, если оно описывается уравнением движения не выше второго порядка и не может быть представлено в виде соединения других элементарных звеньев. Уравнение движения элементарного звена связывает входной и выходной сигналы x_1 и x_2 . Поскольку оно не выше второго порядка, собственный оператор звена имеет вид

$$P_2(p) = T_2 p^2 \pm T_1 p \pm 1, \quad (5.21)$$

а оператор воздействия —

$$P_1(p) = \tau_2 p^2 \pm \tau_1 p \pm 1, \quad (5.22)$$

где каждый член с двойным знаком может иметь любой из них, независимо от знаков других членов. Рассматривая звенья с постоянными параметрами, будем считать, что постоянные времени T и τ всегда положительны.

Звено называется *простым*, если $\tau_1 = \tau_2 = 0$, т. е. нет воздействия по производным входного сигнала.

Все многообразие звеньев, зависящее от их конструктивного выполнения и происходящих в них физических явлений, может быть описано небольшим числом операторов.

Наиболее часто встречаются следующие звенья, называемые *типовыми*:

простое *безынерционное* (или идеальное, пропорциональное, усилительное, безъемкостное)

$$P_1(p) = P_2(p) = 1;$$

простое *инерционное первого порядка* (или аperiodическое первого порядка, одноемкостное, релаксационное, статическое)

$$P_1(p) = 1,$$

$$P_2(p) = Tp + 1;$$

интегрирующее при $\tau < T$ и *дифференцирующее* при $\tau > T$

$$P_1(p) = \tau p + 1,$$

$$P_2(p) = Tp + 1;$$

идеальное интегрирующее

$$P_1(p) = 1,$$

$$P_2(p) = Tp;$$

идеальное дифференцирующее

$$P_1(p) = \tau p,$$

$$P_2(p) = 1;$$

aperiodическое второго порядка (или инерционное, двухемкостное)

$$P_1(p) = 1,$$

$$P_2(p) = T_2 p^2 + T_1 p + 1,$$

при $T_1^2 - 4T_2 \geq 0$. Корни характеристического уравнения в этом случае вещественные. Аperiodическое звено второго порядка можно представить последовательным соединением двух аperiodических звеньев первого порядка;

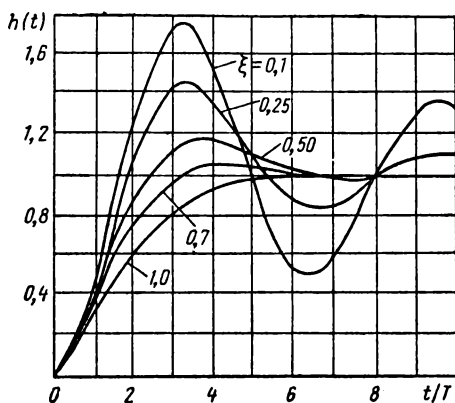


Рис. 5.37. Переходные характеристики колебательного звена

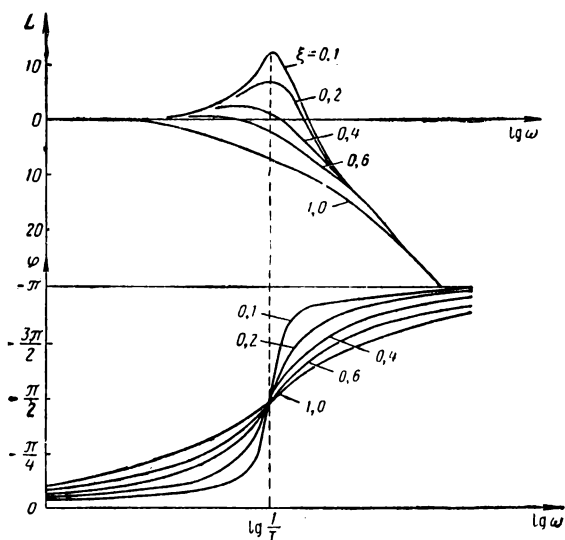


Рис. 5.38. Логарифмические частотные характеристики колебательного звена

колебательное

$$P_1(p) = 1,$$

$$P_2(p) = T^2 p^2 + 2\zeta T p + 1,$$

где T — постоянная времени;

ζ — параметр затухания.

Звено является колебательным, если $\zeta < 1$, т. е. корни характеристического уравнения комплексные:

$$P_{1,2} = -\frac{\zeta}{T} \pm i \frac{1}{T} \sqrt{1 - \zeta^2},$$

где ζ/T — коэффициент затухания, характеризующий скорость затухания переходного процесса, $\sqrt{1 - \zeta^2}/T$ — частота колебаний.

На рис. 5.37 представлены переходные характеристики колебательного звена в функции относительного времени t/T при $K=1$. Как следует из рисунка, уменьшение ζ ведет к увеличению амплитуды колебаний переходного процесса. При $\zeta=1$ колебания отсутствуют, звено является аperiodическим.

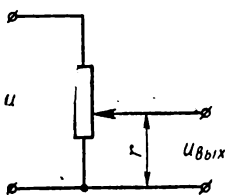


Рис. 5.39. Простое инерционное звено

ЛАХ колебательного звена при частоте $1/T$ имеет резонансный пик, причем $1/T$ является частотой собственных колебаний. На рис. 5.38. представлены логарифмические амплитудные и фазовые характеристики колебательного звена при различных значениях ζ и $K=1$;

консервативное,

$$P_1(p) = 1,$$

$$P_2(p) = T^2 p^2 + 1.$$

Консервативное звено является частным случаем колебательного звена при $\zeta=0$. ЛАХ консервативного звена имеет разрыв при частоте $1/T$ (табл. 5.1), что указывает на наличие незатухающих колебаний.

Собственный оператор инерционного звена второго порядка может быть представлен в виде произведения двух линейных операторов. Линейные операторы имеют вещественные коэффициенты лишь в случае вещественных корней характеристического уравнения звена второго

порядка. В этом случае звено второго порядка может быть заменено двумя инерционными звеньями первого порядка, т. е. не является элементарным. Но оно будет элементарным и называется колебательным, если его характеристическое уравнение имеет комплексные корни. Действительно, в таком случае линейные множители, на

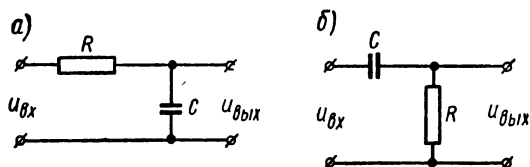


Рис. 5.40. Звенья:

а) идеальное интегрирующее; б) идеальное дифференцирующее

которые может быть разложен его собственный оператор, будут иметь комплексные коэффициенты, т. е. не будут являться операторами реальных звеньев первого порядка. Таким образом, колебательное звено не может быть заменено соединением других элементарных звеньев.

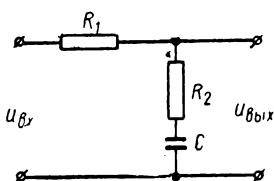


Рис. 5.41. Интегрирующее звено

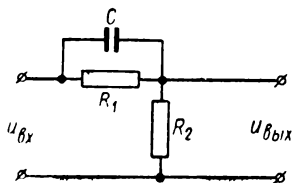


Рис. 5.42. Дифференцирующее звено

Все остальные типовые звенья, перечисленные выше, являются элементарными. Их характеристики приведены в табл. 5.1.

Примером простого безынерционного звена может быть делитель напряжения (рис. 5.39). Входной величиной здесь является положение движка потенциометра r , выходной — напряжение $U_{\text{вых}}$. Уравнение движения потенциометра

$$U_{\text{вых}} = Kr.$$

Электрический контур RC (рис. 5.40, а) является простым инерционным звеном, причем $x_1 = U_{вх}$, $x_2 = U_{вых}$, $T = RC$. Если постоянная времени T значительно больше времени действия сигнала, то этот контур является идеальным интегрирующим звеном.

На рис. 5.40, б представлен контур, тем лучше воспроизводящий идеальное дифференцирующее звено, чем меньше постоянная времени T .

На рис. 5.41 и 5.42 даны схемы интегрирующего и

Характеристики эле

№ п/п	Тип звена и его передаточная функция	частотная характеристика	Амплитудно-фазовая характеристика второго рода
1	Простое безинерционное $W(p) = K$	$U(\omega) = K$ $V(\omega) = 0$	
2	Апериодическое первого порядка $W(p) = \frac{K}{T_D + 1}$	$U(\omega) = \frac{K}{1 + T^2 \omega^2}$ $V(\omega) = -\frac{KT\omega}{1 + T^2 \omega^2}$	
3	Идеальная интегрирующая $W(p) = \frac{K}{T_D}$	$U(\omega) = 0$ $V(\omega) = -\frac{K}{T\omega}$	
4	Идеальное дифференцирующее $W(p) = K\tau p$	$U(\omega) = 0$ $V(\omega) = K\tau\omega$	

дифференцирующего контуров. Для интегрирующего $\tau = CR_2$, $T = \tau + CR_1$, для дифференцирующего $\tau = CR_1$, $T = \frac{\tau R_2}{R_1 + R_2}$. Примером колебательного звена является колебательная RLC цепь, уравнение движения которой рассмотрено в § 4.1.

Наиболее развитая структурная схема показывает, из каких элементарных звеньев состоит система и как эти звенья соединены.

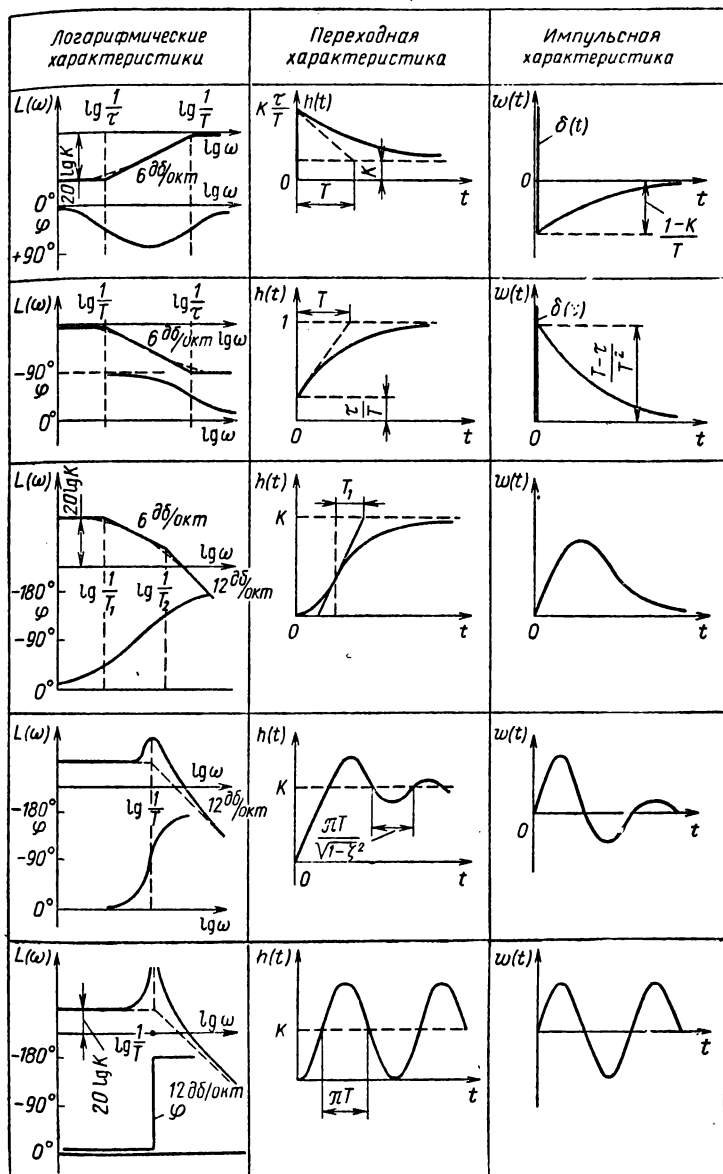
ментальных звеньев

Таблица 5.1

Логарифмические характеристики	Переходная характеристика	Импульсная характеристика

№ п/п	Тип звена и его передаточная функция	Частотная характеристика	Амплитудно-фазовая характеристика второго рода
5	Дифференцирующее $W(p) = K \frac{\tau p + 1}{T p + 1}$ $\tau > T$	$U(\omega) = K \frac{1 + \tau T \omega^2}{1 + T^2 \omega^2}$ $V(\omega) = K \frac{\tau - T}{1 + T^2 \omega^2} \omega$	
6	Интегрирующее $W(p) = \frac{\tau p + 1}{T p + 1}$ $T > \tau$	$U(\omega) = \frac{1 + \tau T \omega^2}{1 + T^2 \omega^2}$ $V(\omega) = -\frac{T - \tau}{1 + T^2 \omega^2}$	
7	Простое инерционное второго порядка $W(p) = \frac{K}{1 + T_1 p + T_2 p^2}$	$U(\omega) = K \frac{1 - T_2 \omega^2}{(1 - T_2 \omega^2)^2 + T_1^2 \omega^2}$ $V(\omega) = -K \frac{T_1 \omega}{(1 - T_2 \omega^2)^2 + T_1^2 \omega^2}$	
8	Колебательное $W(p) = \frac{K}{1 + 2\zeta T p + T^2 p^2}$	$U(\omega) = K \frac{1 - T^2 \omega^2}{(1 - T^2 \omega^2)^2 + (2\zeta T \omega)^2}$ $V(\omega) = -\frac{K 2\zeta T \omega}{(1 - T^2 \omega^2)^2 + (2\zeta T \omega)^2}$	
9.	Консервативное $W(p) = \frac{K}{1 + T^2 p^2}$	$U(\omega) = \frac{K}{1 - T^2 \omega^2}$ $V(\omega) = 0$	

Продолжение таблицы 5.1



Рассмотрим амплитудно-фазовые характеристики звена, имеющего передаточную функцию наиболее общего вида

$$W(p) = K \frac{\tau_2 p^2 + \tau_1 p + 1}{T_2 p^2 + T_1 p + 1}.$$

Частотная характеристика такого звена имеет вид

$$W(i\omega) = K \frac{1 - \tau_2 \omega^2 + i\tau_1 \omega}{1 - T_2 \omega^2 + iT_1 \omega},$$

откуда

$$W(i\omega) = K \frac{\mu (\cos \alpha + i \sin \alpha)}{m (\cos \beta + i \sin \beta)},$$

где

$$\left. \begin{aligned} \mu &= \sqrt{(1 - \tau_2 \omega^2)^2 + \tau_1^2 \omega^2}, \\ m &= \sqrt{(1 - T_2 \omega^2)^2 + T_1^2 \omega^2}, \\ \cos \alpha &= \frac{1 - \tau_2 \omega^2}{\mu}, \\ \sin \alpha &= \frac{\tau_1 \omega}{\mu}, \\ \cos \beta &= \frac{1 - T_2 \omega^2}{m}, \\ \sin \beta &= \frac{T_1 \omega}{m}. \end{aligned} \right\} \quad (5.23)$$

Иначе

$$W(i\omega) = K \frac{\mu}{m} e^{i(\alpha - \beta)}.$$

Сопоставляя последнее выражение частотной характеристики с (4.50), видим, что ее модуль

$$M = K \frac{\mu}{m},$$

а фаза

$$\varphi = \alpha - \beta.$$

Следовательно, координаты амплитудно-фазовой характеристики второго рода

$$\left. \begin{aligned} U_2 &= K \frac{\mu}{m} \cos(\alpha - \beta), \\ V_2 &= K \frac{\mu}{m} \sin(\alpha - \beta), \end{aligned} \right\} \quad (5.24)$$

а координаты характеристики первого рода

$$\left. \begin{aligned} U_1 &= \frac{m}{K_1 \mu} \cos(\alpha - \beta), \\ V_1 &= \frac{m}{K_1 \mu} \sin(\beta - \alpha), \end{aligned} \right\} \quad (5.25)$$

Подставляя в (5.24) и (5.25) выражения (5.23), получаем:

$$\left. \begin{aligned} U_2 &= K \frac{(1 - \tau_2 \omega^2)(1 - T_2 \omega^2) + \tau_1 T_1 \omega^2}{(1 - T_2 \omega^2)^2 + T_1^2 \omega^2}, \\ V_2 &= K \frac{\tau_1 (1 - T_2 \omega^2) - T_1 (1 - \tau_2 \omega^2)}{(1 - T_2 \omega^2)^2 + T_1^2 \omega^2} \omega; \end{aligned} \right\} \quad (5.26)$$

$$\left. \begin{aligned} U_1 &= \frac{1}{K} \frac{(1 - \tau_2 \omega^2)(1 - T_2 \omega^2) + \tau_1 T_1 \omega^2}{(1 - \tau_2 \omega^2)^2 + \tau_1 \omega^2}, \\ V_1 &= \frac{1}{K} \frac{T_1 (1 - \tau_2 \omega^2) - \tau_1 (1 - T_2 \omega^2)}{(1 - \tau_2 \omega^2)^2 + \tau_1 \omega^2} \omega. \end{aligned} \right\} \quad (5.27)$$

По (5.24—5.27) легко построить характеристики любого элементарного звена. Например, для простого инерционного звена первого порядка $\tau_1 = \tau_2 = 0$, $T_2 = 0$. При этом из (5.27)

$$\begin{aligned} U_1 &= \frac{1}{K}, \\ V_1 &= \frac{T_1 \omega}{K}, \end{aligned}$$

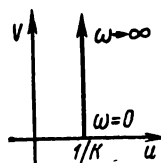


Рис. 5.43. Амплитудно-фазовая характеристика первого рода простого инерционного звена первого порядка

т. е. амплитудно-фазовая характеристика первого рода для рассматриваемого звена имеет вид, показанный на рис. 5.43.

В табл. 5.1 приведены частотные и временные характеристики элементарных звеньев.

Рассмотрим некоторые преобразования элементарных звеньев посредством охвата их обратной связью, осуществляемой также элементарными звеньями по схеме, изображенной на рис. 5.3, где звенья 1 и 2 — основное звено и звено обратной связи.

Пусть основное звено — простое инерционное первого порядка, а звено обратной связи — простое безынерционное. Передаточные функции этих звеньев

$$\begin{aligned} W_1(p) &= \frac{K_1}{T_1 p + 1}, \\ W_2(p) &= K_2. \end{aligned}$$

По формуле (5.22) передаточная функция звена 1, охваченного обратной связью через звено 2,

$$W(p) = \frac{K_1}{T_1 p + 1 \mp K_1 K_2},$$

или

$$W(p) = \frac{K}{T p + 1}, \quad (5.28)$$

где

$$K = \frac{K_1}{1 \mp K_1 K_2},$$

$$T = \frac{T_1}{1 \mp K_1 K_2}.$$

Если обратная связь отрицательная, то получается звено того же типа, что и основное, т. е. инерционное звено первого порядка, но с меньшей постоянной времени, а следовательно, менее инерционное, и с меньшим усилением. В пределе при $K_1 \rightarrow \infty$ $T = 0$, т. е. рассматриваемая обратная связь делает инерционное звено первого порядка безынерционным.

При положительной обратной связи могут быть следующие случаи. В первом случае $K_1 K_2 < 1$. Тогда охват рассматриваемой обратной связью не меняет типа звена, но увеличивает постоянную времени, т. е. инерционность, и коэффициент усиления. Во втором случае $K_1 K_2 = 1$, т. е. $K \rightarrow \infty$ и $T \rightarrow \infty$. При этом передаточная функция, получаемая из (5.28),

$$W(p) = \frac{K}{T p},$$

т. е. получается идеальное интегрирующее звено. В третьем случае $K_1 K_2 > 1$ и по (5.28)

$$W(p) = \frac{K'}{T' p - 1}, \quad (5.29)$$

где

$$K' = \frac{K_1}{K_1 K_2 - 1},$$

$$T' = \frac{T_1}{K_1 K_2 - 1}.$$

Получающееся звено с передаточной функцией (5.29) — неустойчиво, как будет выяснено в гл. VI.

Рассмотрим теперь два простых инерционных звена первого порядка, одно из которых служит звеном связи для другого. Передаточные функции этих звеньев

$$W_1(p) = \frac{K_1}{T_1 p + 1},$$

$$W_2(p) = \frac{K_2}{T_2 p + 1}.$$

По формуле (5.5) в этом случае

$$W(p) = \frac{K_1 (T_2 p + 1)}{(T_1 p + 1) (T_2 p + 1) \mp K_1 K_2},$$

или

$$W(p) = K_1 \frac{\tau_1 p + \tau}{T_2 p^2 + T_1' p + 1}, \quad (5.30)$$

где

$$\tau_1 = \frac{T_2}{1 \mp K_1 K_2},$$

$$\tau = \frac{1}{1 \mp K_1 K_2},$$

$$T_2' = \frac{T_1 T_2}{1 \mp K_1 K_2},$$

$$T_1' = \frac{T_1 + T_2}{1 \mp K_1 K_2}.$$

Таким образом, получается инерционное звено второго порядка с воздействием по первой производной.

Рассмотрим еще простое инерционное звено второго порядка, охваченное обратной связью с помощью дифференцирующего звена (рис. 5.44). Пусть

$$W_1(p) = \frac{K_1}{T_2 p^2 + T_1 p + 1}, \quad (5.31)$$

$$W_2(p) = K_2 \frac{\tau p + 1}{T p + 1}. \quad (5.32)$$

По формуле (5.5) имеем:

$$W(p) = \frac{K_1 (T p + 1)}{(T_2 p^2 + T_1 p + 1) (T p + 1) \mp K_1 K_2 (\tau p + 1)} \quad (5.33)$$

или

$$W(p) = K \frac{T p + 1}{T_2' p^2 + T_1' p + 1}, \quad (5.34)$$

где

$$K = \frac{K_1}{1 \mp K_1 K_2}, \quad T_3 = \frac{TT_2}{1 \mp K_1 K_2}, \quad (5.35)$$

$$T_2' = \frac{T_2 + T_1 T}{1 \mp K_1 K_2}, \quad T_1' = \frac{T_1 + T \mp K_1 K_2 T}{1 \mp K_1 K_2}.$$

Следовательно, элемент третьего порядка с воздействием по первой производной, имеющий передаточную функцию (5.34) и не являющийся элементарным звеном, может быть заменен соединением элементарных звеньев

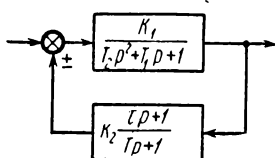


Рис. 5.44. Пример соединения элементарных звеньев по схеме обратной связи

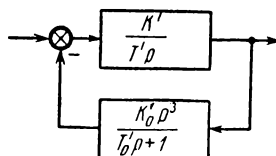


Рис. 5.45. Структурная схема, эквивалентная соединению по рис. 5.44. при знаке плюс

с передаточными функциями (5.31) и (5.32) по схеме рис. 5.44. Коэффициенты усиления и постоянные времени этих звеньев могут быть найдены из системы уравнений (5.35).

Пример. Идеальное интегрирующее звено, охваченное отрицательной обратной связью через инерционное звено первого порядка с воздействием по третьей производной (рис. 5.45), можно заменить только что рассмотренным соединением простого инерционного звена второго порядка с дифференцирующим звеном (рис. 5.44) по схеме положительной обратной связи.

Действительно, если

$$W_1(p) = \frac{K'}{T'p},$$

$$W_2(p) = \frac{K_0' p^3}{T_0' p + 1},$$

то передаточная функция рассматриваемого соединения

$$W(p) = K' \frac{T_0' p + 1}{T' p \left(\frac{K' K_0'}{T'} p^2 + T_0' p + 1 \right)}.$$

Такую же передаточную функцию получим из (5.33), приняв верхний знак и положив

$$K_1 = K', \quad T + T_1 - \tau = T', \quad T = T'_0, \\ K_1 K_2 = 1, \quad T T_2 = K' K'_0, \quad T T_1 + T_2 = T' T'_0,$$

откуда

$$T_2 = \frac{K' K'_0}{T'_0}, \quad T_1 = T' - T_2, \\ \tau = T_2 - T'_0, \quad K_2 = \frac{1}{K'}.$$

§ 5.5. ЗАМЫКАНИЕ ИМПУЛЬСНОЙ СИСТЕМЫ

На рис. 5.46 представлена схема простейшей замкнутой импульсной системы. Эта система получена из разомкнутой, рассмотренной в § 4.5, которая теперь охвачена обратной связью. В цепь обратной связи включено простое безынерционное звено с коэффициентом усиления K_{oc} .

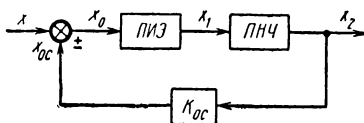


Рис. 5.46. Простейшая замкнутая импульсная система

Входной и выходной сигналы, предполагаемые непрерывными, обозначим через $x(t)$ и $x_2(t)$. При этом непрерывным должен быть и входной сигнал импульсного элемента

$$x_o(t) = x(t) \pm x_{oc}(t),$$

где $x_{oc}(t)$ — сигнал обратной связи.

Учитывая, что импульсный элемент воспринимает лишь дискретные значения этого сигнала и заменяя выражающие его непрерывные функции решетчатыми, имеем:

$$x_o[n] = x[n] \pm x_{oc}[n].$$

В операторной форме, получаемой с помощью дискретного преобразования Лапласа, последнее уравнение принимает вид

$$X_o(q) = X(q) \pm X_{oc}(q), \quad (5.36)$$

где

$$X(q) \doteq x[n],$$

$$X_{oc}(q) = K_{oc} X_2(q).$$

С другой стороны, полагая $\varepsilon = 0$, из (4.84) находим изображение выходного сигнала, воспринимаемого импульс-

ным элементом,

$$X_2(q) = W(q) X_0(q).$$

Подстановка этих выражений в (4.87) дает:

$$X_0(q) = X(q) \pm K_{o.c} W(q) X_0(q),$$

откуда

$$X_0(q) = \frac{X(q)}{1 \mp K_{o.c} W(q)}. \quad (5.37)$$

Из уравнений (4.84) и (5.37) находим операторное уравнение, аналогичное уравнению (4.67) непрерывной системы:

$$X_2(q, \epsilon) = \frac{W(q, \epsilon)}{1 \mp K_{o.c} W(q)} X(q). \quad (5.38)$$

Таким образом, передаточная функция рассматриваемой замкнутой импульсной системы

$$W_3(q, \epsilon) = \frac{W(q, \epsilon)}{1 \mp K_{o.c} W(q)}, \quad (5.39)$$

где $W(q, \epsilon)$ — передаточная функция той же системы в разомкнутом состоянии (т. е. при устранении обратной связи), а $W(q)$ — выражение $W(q, \epsilon)$ при $\epsilon = 0$.

Для более сложных замкнутых импульсных систем передаточные функции составляются аналогичным способом [5.4].

Пример. В случае, когда передаточная функция разомкнутой системы имеет вид (4.88), полагая в этом выражении $\epsilon = 0$, получаем:

$$W(q) = \frac{K}{T} \frac{1}{1 - \exp\left(-q - \frac{T_n}{T}\right)}.$$

Тогда по (5.39)

$$W_3(q, \epsilon) = \frac{K}{T} \frac{\exp\left(-\epsilon \frac{T_n}{T}\right)}{1 \mp K_{o.c} \frac{K}{T} - \exp\left(-q - \frac{T_n}{T}\right)}. \quad (5.40)$$

УСТОЙЧИВОСТЬ ЛИНЕЙНЫХ АВТОМАТИЧЕСКИХ СИСТЕМ

§ 6.1. НЕОБХОДИМОЕ И ДОСТАТОЧНОЕ УСЛОВИЕ УСТОЙЧИВОСТИ

Рассмотрим линейную систему с постоянными сосредоточенными параметрами, операторное уравнение движения которой приведено к виду (4.14). В таком случае изображение выходного сигнала, выраженное через передаточную функцию (4.15),

$$X_2(p) = K \frac{P_1(p)}{P_2(p)} X_1(p).$$

Предположим, что изображение входного сигнала имеет вид отношения двух полиномов от p :

$$X_1(p) = \frac{B_1(p)}{B_2(p)}.$$

Такое изображение имеют сигналы, практически встречающиеся в автоматических системах. Тогда

$$X_2(p) = K \frac{P_1(p) B_1(p)}{P_2(p) B_2(p)}.$$

Разлагая правую часть последнего выражения на простейшие дроби, можно написать, если числитель и знаменатель не имеют общих множителей:

$$X_2(p) = \frac{K_p(p)}{P_2(p)} + \frac{K_n(p)}{B_2(p)}. \quad (6.1)$$

Здесь первый член правой части имеет полюса передаточной функции рассматриваемой системы, а второй — полюса изображения входного сигнала. Полиномы $K_p(p)$ и $K_n(p)$ составляются с помощью известных приемов операторного исчисления.

Уравнение (6.1) показывает, что выходной сигнал имеет две составляющие: *регулярную* составляющую

$$x_p(t) \doteq \frac{K_p(p)}{P_2(p)}, \quad (6.2)$$

и *нерегулярную* составляющую

$$x_n(t) \doteq \frac{K_n(p)}{B_1(p)}.$$

Таким образом,

$$x_2(t) = x_p(t) + x_n(t). \quad (6.3)$$

Система является *устойчивой*, если регулярная составляющая выходного сигнала затухает, т. е.

$$\lim_{t \rightarrow \infty} x_p(t) \rightarrow 0. \quad (6.4)$$

Это значит, как следует из § 3.8, что все особые точки изображения $X_p(p)$ лежат в левой полуплоскости (рис. 3.16). Особыми точками изображения (6.2) являются его полюса, находимые из характеристического уравнения

$$P_2(p) = 0.$$

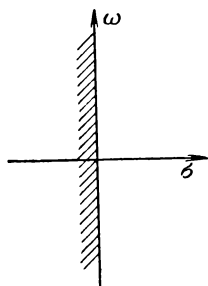


Рис. 6.1. Область устойчивости на плоскости

Чтобы полюса изображения $X_p(p)$ находились в левой полуплоскости, вещественные части всех корней характеристического уравнения должны быть отрицательными. Отсюда следует *необходимое и достаточное условие устойчивости* линейной системы, требующее,

чтобы все корни характеристического уравнения имели отрицательные вещественные части.

Таким образом, если двигаться на плоскости p по оси частот от $-\infty$ к $+\infty$, то все корни характеристического уравнения устойчивой линейной системы будут изображаться точками, находящимися слева от этой оси. Ось частот служит *границей* устойчивости, уравнение которой имеет, следовательно, вид

$$p = i\omega. \quad (6.5)$$

Область устойчивости принято отмечать штриховкой с соответствующей стороны ее границы, как показано на рис. 6.1. При $\sigma = 0$ получаются незатухающие колебания, представляющие собой регулярную составляющую выходного сигнала. Это — вид неустойчивого процесса.

Пример. Устойчива ли система, рассмотренная в примере § 5.4?

Передаточная функция показывает, что характеристическое уравнение этой системы имеет нулевой корень. Следовательно,

система находится на границе устойчивости. Небольшое изменение ее параметров, смещающее нулевой корень в ту или другую сторону с границы устойчивости, может превратить систему в устойчивую или неустойчивую.

§ 6.2. *D*-РАЗБИЕНИЕ

При изменении какого-либо параметра, входящего в уравнение движения, система может оказаться устойчивой или неустойчивой в зависимости от его значения. Если существует такое значение этого параметра, начиная с которого при изменении параметра в ту или другую сторону все корни характеристического уравнения оказываются в левой полуплоскости, то это значение является границей *области устойчивости* системы по рассматриваемому параметру. Границей области устойчивости, зависящей от одного параметра, является, таким образом, точка на оси значений этого параметра.

Если можно изменять одновременно два параметра, то границей области устойчивости является плоская кривая, если три — поверхность. При одновременном изменении N параметров границей области устойчивости является гиперповерхность в N -мерном пространстве параметров.

Ось, плоскость или пространство параметров можно разбить на области, в пределах каждой из которых некоторое число k корней имеет отрицательные вещественные части. Это — *D-разбиение*, предложенное Ю. И. Неймарком [6.4] и названное так в связи с выясненной в гл. IV возможностью замены комплексного числа p символом D . В области *D-разбиения*, определяющей значения параметров, при которых k корней находится слева (т. е. в левой полуплоскости корней), справа должно быть $n - k$ корней, если n — порядок системы. Такую область обозначают символом $D(k, n - k)$, служащим общим обозначением областей *D-разбиения*. В области устойчивости число k должно быть равно порядку системы, т. е. ею может быть только область $D(n, 0)$.

Рассмотрим систему, характеристическое уравнение которой имеет вещественный корень. Если система устойчива, то на плоскости корней этот корень, как и все остальные корни, находится слева. При этом состояние системы изображается некоторой точкой области $D(n, 0)$. Предположим, что вследствие изменения каких-то парамет-

ров системы рассматриваемый вещественный корень перемещается в правую полуплоскость. Это значит, что изображающая его точка на плоскости p , двигаясь по оси σ (см. рис. 6.1), пересекает границу устойчивости. Точка же, изображающая состояние системы, при этом должна перейти из области $D(n, 0)$ в область $D(n-1, 1)$, т. е. перейти границу D -разбиения. Следовательно, граница D -разбиения является отображением границы устойчивости в пространстве параметров, а граница устойчивости — отображением границы D -разбиения на плоскости корней.

Метод D -разбиения применим в тех случаях, когда параметры, в зависимости от которых нужно исследовать устойчивость, входят в характеристическое уравнение линейно. Рассмотрим этот метод в простейшем случае, когда устойчивость определяется в зависимости от одного лишь такого параметра, который обозначим через r . Решая характеристическое уравнение совместно с уравнением границы устойчивости (6.5) и отделяя вещественную часть от мнимой, получаем:

$$r = u(\omega) + iv(\omega). \quad (6.6)$$

Это выражение дает значения параметра r в различных точках границы устойчивости, определяемых различными частотами ω , т. е. значения r , при которых характеристическое уравнение имеет корни, изображаемые точками границы устойчивости.

Принимая u и v за прямоугольные координаты, можно построить в этих координатах кривую $v(u)$, каждой точке которой соответствует некоторое значение ω . В точках этой кривой параметр r имеет те же значения, соответствующие точкам границы устойчивости. Кривая $v(u)$ и является, следовательно, границей D -разбиения.

Таким образом получили границу D -разбиения на плоскости комплексного параметра r . В действительности, как уже было сказано в гл. IV, все параметры автоматической системы вещественны. Поэтому реальный смысл имеют только значения r в точках оси $v=0$ и представляет интерес лишь принадлежность отрезков этой оси к той или иной области D -разбиения.

Кривая $v(u)$ должна быть, очевидно, симметричной относительно оси $v=0$, так что достаточно построить одну из ее ветвей: ветвь отрицательных или ветвь положитель-

ных частот. При изменении частоты от $-\infty$ до 0 и от 0 до $+\infty$ область устойчивости на плоскости корней все время остается слева (см. рис. 6.1). На границе D -разбиения это отмечают штриховкой слева при движении в сторону возрастания частоты. Область устойчивости может быть только та часть плоскости r , которая окружена штриховкой, направленной внутрь ее. Это легко проверить для любой точки оси $v=0$, находящейся в пределах такой области. Если устойчивости здесь нет, то система неустойчива при любых значениях параметра r .

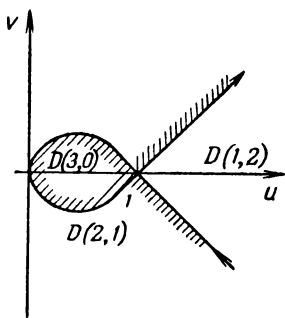


Рис. 6.2. Пример D -разбиения по одному параметру

При переходе из какой-либо области D -разбиения в соседнюю теряется один корень слева, если пересечь границу с заштрихованной стороны, и приобретает он при движении в противоположном направлении. Число корней изменяется на два при переходе через точку самопересечения границы D -разбиения. Это показано на рис. 6.2, где представлена граница D -разбиения в случае характеристического уравнения

$$p^3 + p^2 + p + r = 0 \quad (6.7)$$

[6.11] (направление возрастания частоты показано стрелками).

При изучении устойчивости в зависимости от двух переменных параметров граница D -разбиения строится в прямоугольных координатах, которыми служат эти параметры [6.12, 6.13].

Пример. Построим D -разбиение по параметру r системы второго порядка, имеющей характеристическое уравнение

$$p^2 + p + r = 0. \quad (6.8)$$

Подстановкой (6.5) это уравнение приводится к виду (6.6), причем

$$u(\omega) = \omega^2,$$

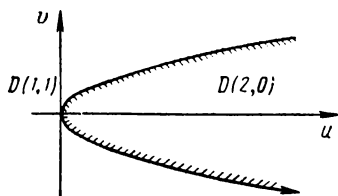
$$v(\omega) = -\omega.$$

Отсюда находим, что уравнение границы D -разбиения

$$v = \pm \sqrt{u},$$

а разбиение имеет вид, представленный на рис. 6.3. Стрелкой

на кривой показано направление возрастания частоты. Как видно из предыдущих выражений, в начале координат $\omega = 0$. Нанося штриховку по левой стороне при движении в направлении возрастания частоты, как сделано на рисунке, находим, что областью устойчивости может быть часть плоскости, охваченная кривой. Для проверки этого рассмотрим корни уравнения (6.8)



$$p_{1,2} = -\frac{1}{2} \pm \sqrt{\frac{1}{4} - r}.$$

Рис. 6.3. К примеру построения D -разбиения

Наличие в ней нулевого корня показывает, что эта точка принадлежит границе устойчивости. Вторая точка находится внутри области, охваченной штриховкой. Отрицательность обоих вещественных корней подтверждает, что здесь область устойчивости, на рисунке — $D(2,0)$. Третья точка лежит вне области устойчивости. Это подтверждается положительным знаком одного из вещественных корней. Поскольку же второй вещественный корень здесь отрицателен, это — область $D(1,1)$

В табл. 6.1 вычислены значения корней в трех точках вещественной оси. Первая из этих точек принадлежит границе D -разбиения.

Т а б л и ц а 6.1

К примеру построения D -разбиения

r	p_1	p_2
0	0	-1
1/4	-1/2	-1/2
-15/4	3/2	-5/2

§ 6.3. АЛГЕБРАИЧЕСКИЕ КРИТЕРИИ

Непосредственное применение необходимого и достаточного условия, требующее решения характеристического уравнения, просто лишь для систем первого и второго порядков, менее удобно в случаях третьего и четвертого порядков, а для систем более высоких порядков является довольно трудоемкой задачей. Однако, чтобы судить о том, удовлетворяет ли линейная система необходимому и достаточному условию устойчивости, нет надобности находить корни характеристического уравнения: достаточно убедиться в том, что их вещественные части отрицательны. Для этого служат *алгебраические критерии*,

Рассмотрим характеристическое уравнение вида

все коэффициенты которого вещественны. Предположим, что p_1, p_2, \dots, p_n — корни этого уравнения. По формулам Виета

Здесь в правой части k -го равенства стоит сумма произведений по k корней, взятая с «+», если k четное, и с «-», если k нечетное. Очевидно, что при отрицательных вещественных корнях коэффициенты положительны. То же имеет место при комплексно-сопряженных корнях с отрицательной вещественной частью, а также, когда корни того и другого рода.

Для систем первого и второго порядков условие одинаковости знаков коэффициентов является достаточным условием устойчивости. Если, например, $n=1$, то при $p+a=0$, где $a>0$, находим $p=-a$.

$$\begin{aligned} a_1 &= -(-K + 2\sigma), \\ a_2 &= \sigma(\sigma - 2K) + \omega^2, \\ a_3 &= -[-K(\sigma^2 + \omega^2)]. \end{aligned}$$

Как видно, коэффициент a_3 всегда положителен, a_1 может быть также положительным, если $K > 2\sigma$, а a_2 — если $\sigma^2 + \omega^2 > 2K$. Следовательно, все коэффициенты могут иметь одинаковые знаки, хотя система неустойчива, т. е. условие одинаковости знаков коэффициентов характеристического уравнения при $n \geq 3$ лишь *необходимо*, но не достаточно.

В конце XIX в. с необходимостью определения устойчивости систем высокого порядка встретился словацкий инженер Стодола — выдающийся конструктор паровых турбин, работавший в Швейцарии. По его предложению швейцарский математик Гурвиц в 1893 г. нашел способ, позволяющий достаточно просто определять по коэффициентам характеристического уравнения, отрицательны ли вещественные части его корней. Для этого составляется *определитель Гурвица*

$$D_n = \begin{vmatrix} a_1 & a_3 & a_5 & \dots & 0 \\ 1 & a_2 & a_4 & \dots & 0 \\ 0 & a_1 & a_3 & \dots & 0 \\ 0 & 1 & a_2 & \dots & 0 \\ 0 & 0 & a_1 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & a_n \end{vmatrix}, \quad (6.11)$$

по диагонали которого располагаются коэффициенты характеристического уравнения от a_1 до a_n . Выше коэффициента, стоящего на диагонали, в каждом столбце пишутся подряд старшие по индексу коэффициенты, а ниже — младшие. Затем составляются все диагональные миноры определителя:

$$D_1 = a_1,$$

$$D_2 = \begin{vmatrix} a_1 & a_3 \\ 1 & a_2 \end{vmatrix},$$

$$D_3 = \begin{vmatrix} a_1 & a_3 & a_5 \\ 1 & a_2 & a_4 \\ 0 & a_1 & a_3 \end{vmatrix}$$

и т. д. *Критерий Гурвица* требует, чтобы у системы, для которой соблюдается необходимое и достаточное условие устойчивости, выполнялось условие

$$D_k > 0, k = 1, 2, \dots, n.$$

Это легко проверить с помощью формул Виета, задаваясь корнями с отрицательной вещественной частью.

Несколько раньше (в 1877 г.) ту же задачу решил Раут в Шотландии. *Критерий Раута* имеет другую формулировку, но эквивалентен критерию Гурвица. Имея характеристическое уравнение (6.9), составим табл. 6.2 (таблицу Раута)..

В таблицу входят коэффициенты характеристического уравнения и коэффициенты с двойными индексами, вычисляемые по формулам

$$\left. \begin{aligned} a_{31} &= \frac{a_1 a_2 - a_3}{a_1}, \\ a_{32} &= \frac{a_1 a_4 - a_5}{a_1}, \\ a_{33} &= \frac{a_1 a_6 - a_7}{a_1}, \\ a_{34} &= \frac{a_1 a_8 - a_9}{a_1}, \\ &\dots \dots \dots \\ a_{41} &= \frac{a_{31} a_3 - a_1 a_{33}}{a_{31}}, \\ a_{42} &= \frac{a_{31} a_5 - a_1 a_{35}}{a_{31}}. \end{aligned} \right\} \quad (6.12)$$

и т. д. Для устойчивости системы критерий Раута требует, чтобы все элементы первого столбца таблицы были отличны от нуля и имели одинаковые знаки.

Критерий Гурвица менее удобен, чем критерий Раута, при больших значениях n , когда для его применения необходимо вычислять определители высоких порядков.

Применим критерий Гурвица к системе третьего порядка, для которой определитель (6.11)

$$D_3 = \begin{vmatrix} a_1 & a_3 & 0 \\ 1 & a_2 & 0 \\ 0 & a_1 & a_3 \end{vmatrix}.$$

Отсюда

$$D_1 = a_1,$$

$$D_2 = a_1 a_2 - a_3,$$

$$D_3 = a_3 D_2.$$

Из условия $D_1 > 0$ находим $a_1 > 0$. При этом из условия $D_2 > 0$ и $D_3 > 0$ получаем $a_3 > 0$. Тогда из условия $D_2 > 0$ следует, что $a_2 > 0$.

Т а б л и ц а 6.2

Строки	Столбцы				
	1	2	3	4
1	1	a_2	a_4	a_6
2	a_1	a_3	a_5	a_7
3	a_{31}	a_{32}	a_{33}	a_{34}
4	a_{41}	a_{42}	a_{43}
5	a_{51}	a_{52}
...					
$n+1$	a_{n1}				

Таким образом, определители D_1 , D_2 и D_3 будут положительными, как требует критерий Гурвица для устойчивой системы, если все коэффициенты характеристического уравнения положительны (необходимое условие устойчивости) и

$$a_1 a_2 - a_3 > 0. \quad (6.13)$$

Условие (6.13) было получено проф. И. А. Вышнеградским в 1876 г. [6.3], когда еще не был найден ни критерий Гурвица, ни критерий Раута, и называется *критерием Вышнеградского*. Система третьего порядка устойчива, если коэффициенты ее характеристического уравнения положительны и удовлетворяют критерию Вышнеградского (6.13).

Если система неустойчива, то во многих случаях можно добиться ее устойчивости, изменяя численные значения

коэффициентов характеристического уравнения посредством изменения параметров системы, от которых зависят эти коэффициенты. Например, в случае несоблюдения неравенства (6.13) систему третьего порядка можно сделать устойчивой, увеличивая коэффициенты a_1 и a_2 или уменьшая коэффициент a_3 . В таких случаях устойчивость называют *параметрической*. Но бывает и так, что неустойчивая система не приобретает устойчивости ни при каких значениях коэффициентов характеристического уравнения. Если, например, один из первых двух коэффициентов характеристического уравнения системы третьего порядка (a_1 или a_2) равен нулю, то критерию Вышнеградского не удовлетворяют никакие значения остальных коэффициентов. В этих случаях характеристическое уравнение *неполное*, в нем нет одного члена.

С помощью формул Виета нетрудно убедиться, что в случае неполного характеристического уравнения хотя бы один его корень должен иметь положительную вещественную часть, причем обусловленная этим неустойчивость системы не может быть устранена никаким изменением коэффициентов характеристического уравнения, не равных нулю. Система, обладающая такой неустойчивостью, называется *структурно неустойчивой*. Чтобы сделать структурно неустойчивую систему устойчивой, необходимо изменить ее структурную схему. Наоборот, система, устойчивость которой параметрическая, является *структурно устойчивой*.

Неполное характеристическое уравнение всегда указывает на структурную неустойчивость, но наличие всех членов характеристического уравнения еще не означает, что система структурно устойчива. Рассмотрим, например, цепочку, состоящую из двух звеньев, передаточные функции которых

$$W_1(p) = \frac{K_1}{T_1 p + 1}, \quad W_2(p) = \frac{K_2}{T_2 p^2 + 1},$$

и охваченную единичной отрицательной обратной связью. Характеристическое уравнение такой системы

$$1 + W_1(p) + W_2(p) = 0$$

приводится к виду (6.9), причем

$$a_1 = \frac{1}{T_1}, \quad a_2 = \frac{1}{T_2}, \quad a_3 = \frac{1 + K_1 K_2}{T_1 T_2}.$$

Применяя критерий Вышнеградского, получаем невыполнимое условие

$$-K_1 K_2 > 0,$$

а следовательно, рассматриваемая система, характеристическое уравнение которой содержит все члены, структурно неустойчива.

Пример. Составим таблицу Раута для системы третьего порядка. По (6.12)

$$a_{31} = \frac{a_1 a_2 - a_3}{a_1};$$

a_{33} и следующие элементы третьей строки равны нулю, вследствие чего $a_{41} = a_{31}$.

Следующие элементы четвертой строки, как и элементы следующих строк, равны нулю. Таким образом получаем табл. 6.3. Из первого столбца таблицы следует, что критерий Раута соблюдается при одинаковых знаках всех коэффициентов характеристического уравнения и соблюдении критерия Вышнеградского.

Т а б л и ц а 6.3
Таблица Раута при $n = 3$

Строки	Столбцы	
	1	2
1	1	a_3
2	a_1	a_3
3	a_{31}	0
4	a_3	0

§ 6.4. ДИАГРАММА ВЫШНЕГРАДСКОГО

Вышнеградский рассматривал переходные процессы в линейных системах третьего порядка с постоянными коэффициентами [6.3]. Если такая система может быть устойчивой, то ее характеристическое уравнение в общем случае получает вид

$$p^3 + a_1 p^2 + a_2 p + a_3 = 0. \quad (6.14)$$

Подстановка

$$p = q \sqrt[3]{a_3} \quad (6.15)$$

приводит характеристическое уравнение (6.14) к форме Вышнеградского

$$q^3 + xq^2 + yq + 1 = 0, \quad (6.16)$$

где коэффициенты x и y , называемые *характеристическими числами Вышнеградского*, определяются формулами

$$\left. \begin{aligned} x &= \frac{a_1}{\sqrt[3]{a_3}}, \\ y &= \frac{a_2}{\sqrt[3]{a_3^2}}. \end{aligned} \right\} \quad (6.17)$$

Введением безразмерных величин q , x и y Вышнеградский привел кубичное уравнение к двухпараметрическому виду¹. Это позволило ему графически представить свойства системы третьего порядка на плоскости, пользуясь характеристическими числами x и y как прямоугольными координатами.

Применяя критерий Вышнеградского к уравнению (6.16), получаем условие устойчивости

$$xy > 1. \quad (6.18)$$

Следовательно, граница устойчивости имеет уравнение

$$xy = 1, \quad (6.19)$$

т. е. является равнобокой гиперболой, называемой *гиперболой Вышнеградского* (рис. 6.4). В точках гиперболы вещественная часть корней характеристического уравнения (6.14) $\sigma = 0$. Принимая это во внимание и пользуясь формулами Виета, нетрудно получить из (6.14) и (6.17) уравнения границы устойчивости в параметрической форме:

$$\left. \begin{aligned} x &= \sqrt[3]{\frac{p_1^2}{\omega^2}}, \\ y &= \sqrt[3]{\frac{\omega^2}{p_1^2}}. \end{aligned} \right\} \quad (6.20)$$

Эти уравнения показывают, что $x \rightarrow \infty$ и $y = 0$ при $\omega = 0$. Наоборот, $x \rightarrow 0$ и $y \rightarrow \infty$ при $\omega \rightarrow \pm \infty$.

Ограничивая область устойчивости на плоскости параметров x и y , гипербола Вышнеградского представляет собой границу D -разбиения по двум параметрам. Двигаясь по этой границе в направлении возрастания частоты от $\omega = -\infty$ до $\omega = 0$, по общему правилу (см. § 6.2) наносим штриховку слева, где и находится область устой-

¹ Существуют и другие способы преобразования кубичного уравнения к двухпараметрическому виду [6.9].

чивости $D(3, 0)$. При этом проходим всю гиперболу как ветвь отрицательных частот. Продолжая затем движение в сторону возрастания частоты к $+\infty$, следует вновь

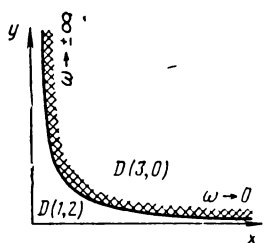


Рис. 6.4. Гипербола Вышнеградского

пройти всю гиперболу, но в обратном направлении. Чтобы правило штриховки не противоречило при этом известному уже расположению области устойчивости, следует представить себе движение по ветви положительных частот с обратной стороны чертежа. Гипербола штрихуется тогда дважды на стороне устойчивости. Переходя через гиперболу из области устойчивости, повсюду будем проходить через точки с двойной штриховкой, теряя, следовательно, два корня слева. Поэтому все точки плоскости параметров, в которых система неустойчива, принадлежат области $D(1, 2)$ [6.12], где $\sigma > 0$.

Дискриминант уравнения (6.16)

$$D = x^2 y^2 - 4(x^3 + y^3) + 18xy - 27. \quad (6.21)$$

Все три корня уравнения (6.16), а следовательно, и уравнения (6.14) вещественны, если $D > 0$. Поэтому граница области вещественных корней определяется уравнением

$$x^2 y^2 - 4(x^3 + y^3) + 18xy - 27 = 0. \quad (6.22)$$

Построив по точкам кривую, описываемую этим уравнением, Вышнеградский показал, что она имеет точку возврата при $x = y = 3$. Действительно, в точке возврата должно быть выполнено условие

$$\left(\frac{\partial^2 D}{\partial x \partial y}\right)^2 - \frac{\partial^2 D}{\partial x^2} \frac{\partial^2 D}{\partial y^2} = 0,$$

где D выражается формулой (6.21). Вычислив производные, легко привести это условие к виду

$$(2xy + 9)^2 - (y^2 - 12x)(x^2 - 12y) = 0.$$

Отсюда, при $x = y$

$$(2x^2 + 9)^2 = x^3(x - 12)^2,$$

т. е. уравнение имеет корень $x = 3$.

Для исследования точки возврата найдем касательную в ней, т. е. производную $y' = dy/dx$. Эта производная, как производная неявной функции (6.22) определяется уравнением

$$\frac{\partial D}{\partial x} + \frac{\partial D}{\partial y} y' = 0. \quad (6.23)$$

Однако в точке $x=y=3$ частные производные

$$\frac{\partial D}{\partial x} = 2xy^2 - 12x^2 + 18y,$$

$$\frac{\partial D}{\partial y} = 2x^2x - 12y^2 + 18x$$

равны нулю, вследствие чего в этой точке y' из уравнения (6.23) не определяется. Чтобы найти y' в рассматриваемой точке, дифференцируем (6.23) еще раз по x :

$$\frac{\partial^2 D}{\partial x^2} + \left(\frac{\partial^2 D}{\partial x \partial y} + \frac{\partial^2 D}{\partial y \partial x} \right) y' + \frac{\partial^2 D}{\partial y^2} y'^2 + \frac{\partial D}{\partial y} y'' = 0.$$

Принимая во внимание, что

$$\frac{\partial^2 D}{\partial x \partial y} = \frac{\partial^2 D}{\partial y \partial x},$$

а в исследуемой точке $\partial D / \partial y = 0$, для определения y' получаем уравнение

$$\frac{\partial^2 D}{\partial y^2} y'^2 + 2 \frac{\partial^2 D}{\partial x \partial y} y' + \frac{\partial^2 D}{\partial x^2} = 0.$$

Отсюда

$$y' = \frac{-\frac{\partial^2 D}{\partial x \partial y} \pm \sqrt{\left(\frac{\partial^2 D}{\partial x \partial y}\right)^2 - \frac{\partial^2 D}{\partial x^2} \frac{\partial^2 D}{\partial y^2}}}{\frac{\partial^2 D}{\partial y^2}},$$

где

$$\frac{\partial^2 D}{\partial x \partial y} = 4xy + 18,$$

$$\frac{\partial^2 D}{\partial x^2} = 2y^2 - 24x, \quad \frac{\partial^2 D}{\partial y^2} = 2x^2 - 24y.$$

При $x=y=3$

$$\frac{\partial^2 D}{\partial x^2} = \frac{\partial^2 D}{\partial y^2} = -54, \quad \frac{\partial^2 D}{\partial x \partial y} = 54 \quad \text{и} \quad y' = 1.$$

Следовательно, касательная в точке возврата совпадает с биссектрисой координатного угла (рис. 6.5). Замечая же, что исследуемая кривая симметрична относительно биссектрисы координатного угла, так как (6.22) не изменяется, когда x и y меняются местами, заключаем, что при $x=y=3$ кривая имеет точку возврата первого рода. Одна из ее ветвей обращена вогнутостью к оси абсцисс, а другая — к оси ординат.

Область вещественных корней находится между биссектрисой координатного угла и двумя ветвями кривой, представляемой уравнением (6.22), так как в этой части диаграммы $D > 0$, а вне ее $D < 0$. Это легко проверить. Например, при $x=y=1$ дискриминант $D = -16$. В области, где $D < 0$, характеристическое уравнение

имеет два комплексных сопряженных корня, вырождающихся на границе этой области в кратные вещественные корни. В точке возврата (6.16) принимает вид

$$q^8 + 3q^3 + 3q + 1 = 0$$

или

$$(q + 1)^8 = 0. \quad (6.24)$$

Отсюда следует, что в этой точке характеристическое уравнение имеет три равных вещественных корня.

Комплексные корни устойчивых систем находятся в области диаграммы Вышнеградского, расположенной между гиперболой и

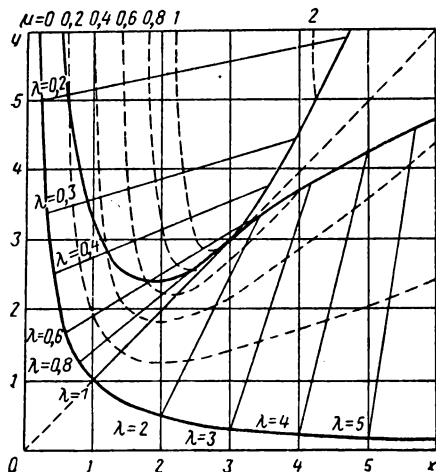


Рис. 6.5. Диаграмма Вышнеградского
с линиями

$\lambda = \text{const}$ и $\mu = \text{const}$

кривой, имеющей уравнение (6.22). Эти линии принадлежат семейству кривых равных затуханий гармонической составляющей переходного процесса [6.2].

Затухание гармонической составляющей определяется уменьшением ее амплитуды за период колебаний, равный $2\pi/\omega$. Отношение амплитуды некоторого периода к амплитуде предыдущего равно $\exp\left(2\pi \frac{\sigma}{\omega}\right)$. Следовательно, затухание зависит от отношения σ/ω .

При исследовании корней характеристического уравнения по диаграмме Вышнеградского удобно рассматривать корни q_1, q_2, q_3 уравнения (6.16) вместо корней p_1, p_2, p_3 уравнения (6.14). Пусть в общем случае это будет вещественный корень

$$q_1 = -\lambda$$

и комплексные корни

$$q_{2,3} = -\mu \pm i\nu,$$

где λ , μ и ν — безразмерные величины, принимающие только положительные значения. Перейти к корням уравнения (6.14) легко по (6.15):

$$\left. \begin{aligned} \rho_1 &= -\lambda \sqrt[3]{a_3}, \\ \sigma &= -\mu \sqrt[3]{a_3}, \\ \omega &= \nu \sqrt[3]{a_3}. \end{aligned} \right\} \quad (6.25)$$

Из этих выражений следует, что

$$\frac{\sigma}{\omega} = -\frac{\mu}{\nu},$$

т. е. затухание одинаково при одинаковых отношениях $|\sigma|/\omega$ и μ/ν . Обозначая через j затухание, выраженное в процентах, имеем:

$$j = 100 \left[1 - \exp \left(-2 \frac{\mu}{\nu} \right) \right]. \quad (6.26)$$

Гипербола Вышнеградского является линией нулевого затухания, так как в ее точках $\sigma = 0$. На границе области вещественных корней $\omega = 0$, и формула (6.26) показывает, что это — линия стопроцентного затухания. Между этими кривыми располагаются линии промежуточных значений j , постепенно трансформирующиеся с возрастанием затухания из равнобокой гиперболы в кривую с точкой возврата первого рода, имеющую касательной биссектрису координатного угла. Вместо затухания удобно пользоваться также коэффициентом демпфирования, нанося на диаграмму Вышнеградского линии его равных значений [6.10].

Применяя формулы Виета к уравнению (6.16), находим:

$$\left. \begin{aligned} x &= \lambda + 2\mu, \\ y &= 2\lambda\mu + \mu^2 + \nu^2, \\ 1 &= \lambda(\mu^2 + \nu^2). \end{aligned} \right\} \quad (6.27)$$

Исключая отсюда λ и вводя обозначение $x = 1 + \frac{\nu^2}{\mu^2}$, получаем [6.5]:

$$x^3 x^2 y^2 - 4x^2 (x^3 + y^2) - (2x^3 - 4x - 16) x y + x^3 - 12x^2 + 48x - 64 = 0. \quad (6.28)$$

При $x = 1$ получаем (6.22), а при $x \rightarrow \infty$ — уравнение гиперболы Вышнеградского. При промежуточных значениях x уравнение (6.28) представляет собой линии различных значений отношения ν/μ . Эти линии можно трактовать либо как линии соответственных постоянных значений затухания j , либо как линии постоянных значений колебательности системы, характеризуемой отношением ν/μ . Колебательность равна нулю на границе вещественных корней и бесконечно велика на границе устойчивости.

Зависимость (6.28) можно представить также в параметрическом виде, удобном для построений [6.5]:

$$\left. \begin{aligned} x &= \frac{1}{\mu^2 x} + 2\mu, \\ y &= \frac{2}{\mu x} + \mu^2 x. \end{aligned} \right\} \quad (6.29)$$

Отсюда при $x = 1$ получаются параметрические уравнения границы вещественных корней.

Из формул (6.27) получаем [6.8] также уравнения

$$y = \lambda x + \frac{1 - \lambda^3}{\lambda}, \quad (6.30)$$

$$y = 2\mu(x - 2\mu) + \frac{1}{x - 2\mu}. \quad (6.31)$$

Уравнение (6.30) представляет собой семейство прямых. Каждая прямая, соответствующая некоторому постоянному значению λ , имеет равный λ угловой коэффициент и пересекает ось ординат в точке

$$y = \frac{1 - \lambda^3}{\lambda},$$

а ось абсцисс — в точке

$$x = \frac{\lambda^3 - 1}{\lambda^2}.$$

Отсюда видно, что биссектриса координатного угла является линией $\lambda = 1$.

Уравнение (6.31) представляет собой семейство кривых $\mu = \text{const}$. Каждая кривая имеет минимум в точке с координатами

$$\begin{aligned} x &= 2\mu + \frac{1}{\sqrt{2\mu}}, \\ y &= 2\sqrt{2\mu}. \end{aligned}$$

Таким образом, геометрическое место минимумов этого семейства определяется уравнением

$$x = \frac{y^2}{4} + \frac{2}{y}.$$

Эта кривая имеет асимптоту, которой является ось абсцисс, и вершину в точке

$$\begin{aligned} x &= \frac{3}{\sqrt[3]{4}} \approx 1,89, \\ y &= \sqrt[3]{4} \approx 1,59. \end{aligned}$$

Каждая из кривых $\mu = \text{const}$ имеет две асимптоты: вертикальную

$$x = 2\mu$$

и наклонную

$$y = 2\mu(x - 2\mu).$$

Обе асимптоты пересекаются на оси абсцисс. При $\mu = 0$ вертикальная асимптота сливается с осью ординат, а наклонная — с осью абсцисс, т. е. с асимптотами гиперболы Вышнеградского, являющейся линией $\mu = 0$.

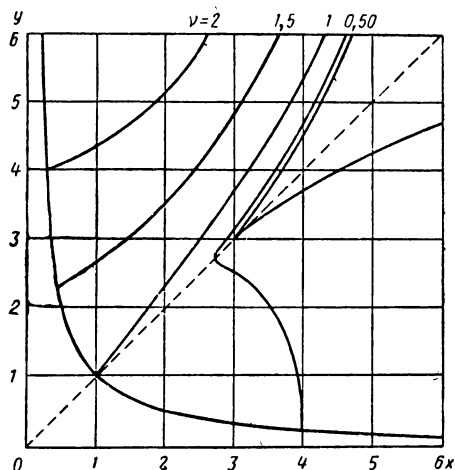


Рис. 6.6. Диаграмма Вышнеградского с линиями $\nu = \text{const}$

Для построения линий $\nu = \text{const}$ преобразуем выражения (6.27) в параметрические уравнения:

$$\left. \begin{aligned} x &= \frac{1}{\mu^2 + \nu^2} + 2\mu, \\ y &= \frac{2\mu}{\mu^2 + \nu^2} + \mu^2 + \nu^2. \end{aligned} \right\} \quad (6.32)$$

При $\nu = 0$ — это уравнения границы вещественных корней, совпадающие с получаемыми из (6.29). Координата x кривой $\nu = \text{const}$ имеет минимум, определяемый уравнением

$$\mu^2 = \sqrt{\mu} - \nu^2,$$

т. е. при $\mu > \nu^4$.

Семейства линий $\lambda = \text{const}$, $\mu = \text{const}$ и $\nu = \text{const}$ построены на рис. 6.5 и 6.6 [6.8]. В каждой точке их пересечения должны соблюдаться условия, выражаемые формулами (6.27). Пользуясь этими линиями, а также (6.29), на диаграмме можно строить различные характеристики системы.

Вышнеградский построил в рассматриваемой области диаграммы линию, в точках которой вещественный корень равен веществен-

ной части комплексных корней:

$$\rho_1 = \sigma. \quad (6.33)$$

Чтобы составить уравнение этой линии, введем условие (6.33) в выражения коэффициентов (6.10). Тогда

$$\begin{aligned} a_1 &= -3\sigma, \\ a_2 &= 3\sigma^2 + \omega^2, \\ a_3 &= -\sigma(\sigma^2 + \omega^2). \end{aligned}$$

Исключая отсюда σ и ω , находим:

$$2a_1^2 - 9a_1a_2 + 27a_3 = 0.$$

Далее, выражая a_1 и a_2 через x и y по формулам (6.17), искомое уравнение получаем в виде

$$y = \frac{2}{9}x^2 + \frac{3}{x}. \quad (6.34)$$

Кривая имеет минимум, координаты которого

$$\begin{aligned} x &= \frac{3}{\sqrt[3]{4}} \approx 1,89, \\ y &= \frac{2}{3}\sqrt[3]{4} \approx 2,38. \end{aligned}$$

Условие (6.33) показывает, что рассматриваемая кривая должна проходить через точку возврата границы области вещественных корней. Поэтому при $x = 3$ из (6.34) находим $y = 3$. Пройдя эту точку, кривая уходит в область вещественных корней, но здесь она теряет смысл. Ось ординат служит ее асимптотой, но кривая не пересекает границу устойчивости: из (6.34) видно, что ее ординаты больше ординат гиперболы (6.19) при одинаковых значениях абсцисс.

В точках исследуемой кривой соблюдается условие (6.32), т. е. $\mu = \lambda$. Между осями координат и линией, составленной из части этой кривой, лежащей вне области вещественных корней, и обращенной вогнутостью к оси абсцисс ветви границы вещественных корней, находится область, где $|p_1| > |\sigma|$. Это следует из того, что в этой области лежит граница устойчивости, в точках которой $\sigma = 0$. Область, где $|p_1| < |\sigma|$, находится между исследуемой кривой и другой ветвью границы вещественных корней (см. рис. 6.5).

Диаграмма Вышнеградского может быть применена также к исследованию устойчивости систем выше третьего порядка [6.7].

Пример. Исследуем по диаграмме Вышнеградского устойчивость системы, имеющей характеристическое уравнение (6.7), в котором $a_1 = a_2 = 1$, $a_3 = r$. По (6.17)

$$x = \frac{1}{\sqrt[3]{r}}, \quad y = \frac{1}{\sqrt[3]{r^2}}, \quad xy = \frac{1}{r}.$$

Следовательно, по (6.18) система устойчива при $r < 1$, что согласуется с D -разбиением, показанным на рис. 6.2.

§ 6.5. ЧАСТОТНЫЕ КРИТЕРИИ

Применение критериев Гурвица и Раута к системам высокого порядка требует длинных вычислений. В этих случаях имеют преимущества методы исследования устойчивости по частотным характеристикам. С помощью этих методов формулируются *частотные критерии устойчивости*.

Частотные характеристики можно получить, переходя от преобразования Лапласа к преобразованию Фурье. Если такой переход возможен, то он достигается подстановкой (3.28) в операторное уравнение движения (см. § 4.2). Такую подстановку нельзя делать в характеристическое уравнение, так как оно при этом потеряло бы смысл. В этом легко убедиться на примере простейшего характеристического уравнения первой степени

$$p + a = 0. \quad (6.35)$$

Однако можно рассматривать отдельно левую часть характеристического уравнения, представляющую некоторый вектор на комплексной плоскости корней, имеющий смысл и при $p = i\omega$. Этот вектор, называемый *характеристическим*, имеет n составляющих вида $p - p_k$, а при $p = i\omega$ — вида $i\omega - p_k$, т. е. выражается произведением

$$\prod_{k=1}^n (i\omega - p_k), \quad (6.36)$$

где n — степень характеристического уравнения.

Рассмотрим две составляющие характеристического вектора, соответствующие паре комплексно-сопряженных корней

$$p_k = \sigma_k + i\omega_k,$$

$$p_{k+1} = \sigma_k - i\omega_k,$$

где $\sigma_k < 0$, так что эти корни удовлетворяют необходимому и достаточному условию устойчивости. Нанеся их на плоскости корней (рис. 6.7), построим векторы $i\omega - p_k$ и $i\omega - p_{k+1}$. Исследуем, как изменяется положение вектора $i\omega - p_k$ при изменении ω от 0 до ∞ . Очевидно, что он повернется на угол $\pi/2 + \gamma$, где γ — показанный на рисунке угол между вектором p_k и отрицательным на-

правлением оси абсцисс. Вектор $i\omega - p_{k+1}$ повернется в том же случае на угол $\pi/2 - \gamma$. Эти повороты рассматриваемых векторов изменяют аргумент характеристического вектора на

$$\frac{\pi}{2} + \gamma + \frac{\pi}{2} - \gamma = \pi,$$

так что на каждый корень приходится изменение $\pi/2$. То же будет в случае вещественных корней, для которых $\gamma=0$, за исключением нулевых корней, которые, как легко видеть, не вызывают изменения аргумента характеристического вектора. Чисто мнимые корни изменяют аргумент характеристического вектора так, что при изменении частоты в указанных пределах положительный корень дает $\pm \pi$, а отрицательный — 0. Таким образом, на каждый корень приходится изменение аргумента на $\pm \frac{\pi}{2}$, но знак этого изменения оказывается неопределенным.

Рис. 6.7. Составляющие
характеристического
вектора

Рис. 6.7. Составляющие характеристического вектора

Чтобы построить характеристический вектор на его комплексной плоскости, необходимо отложить по осям координат его вещественную и мнимую части u и v , которые получим после подстановки (3.28) в левую часть характеристического уравнения, если отделим мнимые члены от вещественных. Тогда левая часть характеристического уравнения будет представлять собой вектор вида

$$M(i\omega) = u(\omega) + iv(\omega), \quad (6.37)$$

где

$$\begin{aligned} u &= a_n - a_{n-2}\omega^2 + a_{n-4}\omega^4 - \dots, \\ v &= a_{n-1}\omega - a_{n-3}\omega^3 + a_{n-5}\omega^5 - \dots \end{aligned} \quad (6.38)$$

Последние члены выражений u и v содержат коэффициент старшего члена характеристического уравнения a_0 .

Для каждой из составляющих рассмотренного выше характеристического вектора (6.36) u и v равны соответственно ее вещественной и мнимой компонентам. Например, корню характеристического уравнения (6.35) $p_1 = -a$ соответствует единственная в этом случае составляющая характеристического вектора, построенная на рис. 6.8. В (6.35) $a_n = a$, $a_{n-1} = 1$, так что $u = a$, $v = \omega$.

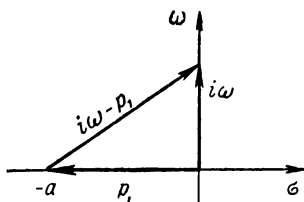


Рис. 6.8. Характеристический вектор системы первого порядка

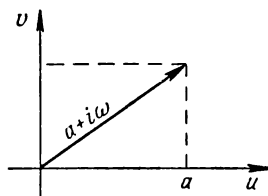


Рис. 6.9. Вектор Михайлова для системы первого порядка

Характеристический вектор, построенный на рис. 6.8, показан на плоскости u, v на рис. 6.9. В обоих случаях он одинаков:

$$i\omega - p_1 = a + i\omega.$$

Таким образом, при изменении частоты ω вектор (6.37) будет вести себя так же, как характеристический вектор (6.36), и его аргумент будет также изменяться за счет каждого корня характеристического уравнения, как это найдено выше для аргумента характеристического вектора.

Каждому значению частоты ω соответствует пара значений u и v . Принимая u и v за прямоугольные координаты, можно построить на их плоскости кривую, каждой точке которой будет соответствовать некоторое значение ω . Эта кривая будет годографом вектора (6.37), называемого далее *вектором Михайлова*. А. В. Михайлов

впервые применил частотные методы к исследованию автоматических систем [6.1].

Рассмотрим изменение аргумента вектора (6.37) при изменении частоты ω от 0 до ∞ . Если все n корней характеристического уравнения лежат в левой полуплоскости, то при таком изменении частоты аргумент вектора Михайлова, согласно выведенному выше, изменится на $n \frac{\pi}{2}$. Следовательно, необходимое и достаточное условие устойчивости системы можно интерпретировать так: при изменении частоты от 0 до $+\infty$ вектор Михайлова совершает поворот на угол $n \frac{\pi}{2}$, где n — порядок системы. Это первая формулировка частотного критерия устойчивости, называемого *критерием Михайлова*.

Рассматривая годограф, получаемый при указанном повороте вектора Михайлова, найдем, что при изменении частоты от 0 до ∞ годограф устойчивой системы должен окружать начало координат, пересекая n квадрантов; это — вторая формулировка критерия Михайлова.

Исследуем годограф вектора (6.37), называемый также *кривой Михайлова*. При $\omega=0$ из (6.37) находим $u=a_n$ и $v=0$. Следовательно, годограф начинается в точке $u=a_n$. Чтобы найти положение касательной в этой точке, вычислим производную

$$\frac{dv}{du} = \frac{dv/d\omega}{du/d\omega}.$$

Из выражений (6.38)

$$\frac{dv}{d\omega} = a_{n-1} - 3a_{n-3}\omega^2 + \dots,$$

$$\frac{du}{d\omega} = -2a_{n-2}\omega + 4a_{n-4}\omega^3 - \dots$$

Отсюда находим, что $dv/du \rightarrow \infty$ при $\omega=0$, т. е. в начальной точке годографа касательная параллельна оси ординат (рис. 6.10). Полагая $\omega \rightarrow \infty$, находим, что при четном n степень $du/d\omega$ выше степени $dv/d\omega$ и $dv/du \rightarrow 0$; следовательно, во втором и четвертом квадрантах годограф в пределе превращается в прямую, параллельную оси абсцисс. При нечетном n также находим, что $\frac{dv}{du} \rightarrow \pm \infty$ при $\omega \rightarrow \infty$. Следовательно, в первом и третьем

квадрантах годограф в пределе превращается в прямую, параллельную оси ординат. На рис. 6.10 показан примерный вид годографа систем различного порядка.

Как видно из рис. 6.10, координаты u и v годографа по очереди меняют знак, проходя через нуль. Отсюда третья формулировка критерия Михайлова: система устойчива, если при изменении частоты от 0 до ∞ координаты годографа поочередно проходят через нуль, в общем n раз. На рис. 6.11 показан примерный вид кривых $u(\omega)$ и $v(\omega)$ для устойчивой системы пятого порядка.

Если характеристическое уравнение имеет нулевой корень, то изменение аргумента вектора Михайлова при изменении частоты от 0 до ∞ на $\frac{\pi}{2}$ меньше требуемого для устойчивости системы. При этом $a_n = 0$ и годограф

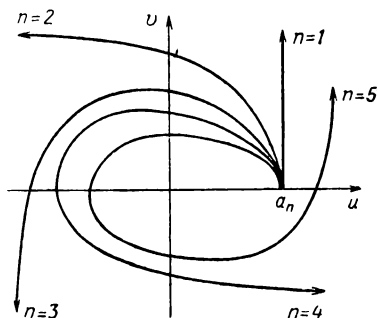


Рис. 6.10. Примерный вид кривых Михайлова для систем от первого до пятого порядка

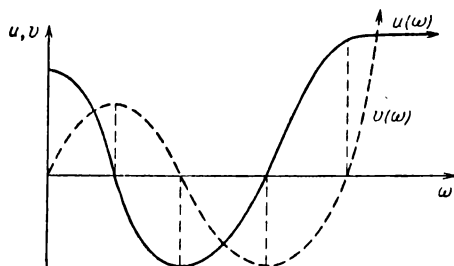


Рис. 6.11. Примерная зависимость координат точек кривой Михайлова от частоты для системы пятого порядка

начинается в начале координат. При наличии пары чисто мнимых корней годограф проходит через начало координат. В этих случаях, поскольку имеются корни, лежащие на мнимой оси, система находится на границе устойчи-

вости, если только все остальные корни лежат в левой полуплоскости.

Рассмотрим теперь влияние охвата отрицательной обратной связью на устойчивость системы. Пусть разомкнутая система (рис. 6.12) имеет передаточную функцию

$$W_p(p) = K \frac{P_1(p)}{P_2(p)} \quad (6.39)$$

и, следовательно, характеристическое уравнение

$$P_2(p) = 0. \quad (6.40)$$

Вектор Михайлова для разомкнутой системы имеет вид

$$M_p(i\omega) = P_2(i\omega). \quad (6.41)$$

Предположим, что характеристическое уравнение (6.40) имеет m корней, находящихся в правой полуплоскости.

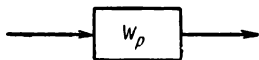


Рис. 6.12. Разомкнутая система

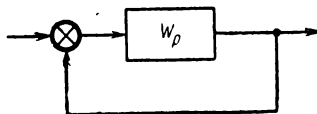


Рис. 6.13. Замкнутая система

Тогда при изменении частоты от 0 до ∞ аргумент вектора (6.41) изменяется за счет этих корней на $-m \frac{\pi}{2}$, а за счет корней, находящихся в левой полуплоскости, — на $(n - m) \frac{\pi}{2}$. Полное изменение аргумента вектора Михайлова при этом

$$(n - m) \frac{\pi}{2} - m \frac{\pi}{2} = (n - 2m) \frac{\pi}{2}.$$

После охвата (рис. 6.13) единичной отрицательной обратной связью передаточная функция системы примет вид

$$W_z(p) = \frac{W_p(p)}{1 + W_p(p)}, \quad (6.42)$$

а характеристическое уравнение —

$$1 + W_p(p) = 0, \quad (6.43)$$

откуда

$$KP_1(p) + P_2(p) = 0. \quad (6.44)$$

Теперь для замкнутой системы вектор Михайлова

$$M_3(i\omega) = KP_1(i\omega) + P_2(i\omega). \quad (6.45)$$

Чтобы замкнутая система была устойчивой, все n корней уравнения (6.44) должны находиться в левой полуплоскости. Тогда при изменении частоты от 0 до ∞ аргумент вектора (6.45) изменится на $n \frac{\pi}{2}$.

Введем вектор, представляющий собой после подстановки (3.28) левую часть уравнения (6.43):

$$N(i\omega) = 1 + W_p(i\omega). \quad (6.46)$$

Этот вектор называется *вектором Найквиста*. Принимая во внимание (6.39), имеем:

$$N(i\omega) = \frac{KP_1(i\omega) + P_2(i\omega)}{P_2(i\omega)}.$$

Отсюда, согласно (6.45) и (6.41),

$$N(i\omega) = \frac{M_3(i\omega)}{M_p(i\omega)}, \quad (6.47)$$

т. е. вектор Найквиста равен частному от деления вектора Михайлова замкнутой системы на вектор Михайлова разомкнутой системы. Следовательно, аргумент вектора Найквиста равен разности аргументов векторов Михайлова замкнутой и разомкнутой систем. При изменении частоты от 0 до ∞ изменение аргумента вектора Найквиста будет равно:

$$n \frac{\pi}{2} - (n - 2m) \frac{\pi}{2} = m\pi.$$

Отсюда следует первая формулировка *критерия Найквиста*: если характеристическое уравнение разомкнутой системы имеет m корней в правой полуплоскости, то аргумент вектора $N(i\omega)$ должен изменяться на $m\pi$ при изменении частоты от 0 до ∞ , чтобы замыкание делало систему устойчивой.

Для графической интерпретации этого критерия построим годограф вектора Найквиста по амплитудно-фазовой характеристике второго рода, данной для той же системы в разомкнутом состоянии. С этой целью представим частотную характеристику $W(i\omega)$ в форме (4.49)

и построим ее в прямоугольных координатах $U(\omega)$ и $V(\omega)$. Каждой точке U, V , получаемой при этом кривой, будет соответствовать некоторая частота ω , а вектор, имеющий

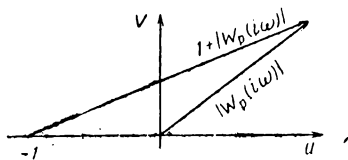


Рис. 6.14. Построение вектора Найквиста

начало в начале координат и конец в этой точке, будет равен по величине модулю частотной характеристики при частоте ω :

$$|W(i\omega)| = \sqrt{U^2(\omega) + V^2(\omega)}$$

(рис. 6.14). Теперь построим вектор, начинающийся в точке оси абсцисс, где $U = -1$, и кончающийся в рассмотренной точке U, V . Модуль этого вектора, как видно из рисунка, равен

$$\sqrt{(1+U)^2 + V^2} = |1 + W(i\omega)|,$$

откуда следует, согласно (6.46), что это — вектор $N(i\omega)$. Таким образом, частотная характеристика разомкнутой системы, построенная в координатах U, V , является годографом вектора Найквиста в координатах $U + 1, V$. При изменении частоты ω вектор $N(i\omega)$ замкнутой системы обегает своим концом частотную характеристику разомкнутой системы, имея начало в точке $(-1, 0)$.

Если разомкнутая система устойчива, то $m = 0$, и изменение аргумента вектора $N(i\omega)$ при изменении частоты от 0 до ∞ должно быть равно нулю, чтобы замкнутая система была также устойчивой. Это условие будет соблюдено, если частотная характеристика разомкнутой системы обходит точку $(-1, 0)$, как показано для примера на рис. 6.15. Если же $m > 0$, то частотная характеристика должна окружать точку $(-1, 0)$, чтобы при изменении частоты от 0 до ∞ вектор $N(i\omega)$ сделал поворот на угол $m\pi$, тогда замыкание сделает систему устойчивой. На рис. 6.16 показан пример, в котором $m = 2$, но частотная характеристика разомкнутой системы окружает точку

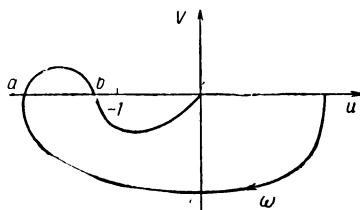


Рис. 6.15. Амплитудно-фазовая характеристика второго рода для устойчивой разомкнутой системы, сохраняющей устойчивость после замыкания

($-1, 0$) так, что при изменении частоты от 0 до ∞ вектор $N(i\omega)$ делает поворот на 2π ; это значит, что замкнутая система устойчива.

Рассмотрим систему, устойчивую в разомкнутом состоянии. По приведенной уже формулировке критерия Найквиста система будет устойчивой и в замкнутом состоянии, если частотная характеристика разомкнутой системы при $|W(i\omega)| > 1$ пересекает вещественную ось четное число раз, поочередно меняя знак производной частоты по углу или угла по частоте при этих пересечениях, либо таких пересечений (при модуле, большем единицы) не имеет. Это — вторая формулировка критерия Найквиста. Случай четного числа пересечений показан на рис. 6.15. Случай отсутствия пересечений левее точки $(-1, 0)$ показан на рис. 6.17, где представлена частотная характеристика устойчивой разомкнутой системы, сохраняющей устойчивость после замыкания

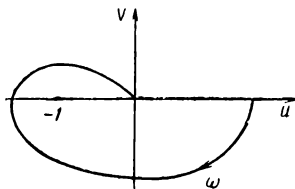


Рис. 6.16. Амплитудно-фазовая характеристика второго рода для неустойчивой разомкнутой системы ($m=2$), которая становится устойчивой после замыкания

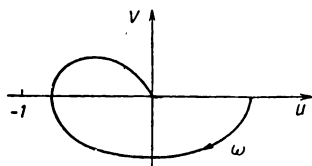


Рис. 6.17. Амплитудно-фазовая характеристика второго рода для устойчивой разомкнутой системы, сохраняющей устойчивость после замыкания

устойчивой разомкнутой системы, остающейся устойчивой и после замыкания.

Рассмотрим рис. 6.16, предполагая, что часть характеристики, проходящая левее точки $(-1, 0)$, приближается к этой точке. Пока эта часть характеристики остается левее точки $(-1, 0)$, замкнутая система сохраняет устойчивость; но система теряет устойчивость, как только

характеристика оказывается правее этой точки. Очевидно, поэтому, в случае характеристики, проходящей через точку $(-1, 0)$, замкнутая система будет находиться на границе устойчивости. Тот же вывод легко сделать и из рассмотрения случаев, представленных на рис. 6.15 и 6.17.

На этом основано понятие о *запасе устойчивости*, характеризующем удаление характеристики от точки $(-1, 0)$. Если замкнутая система находится на границе

устойчивости, то в точке $(-1, 0)$ модуль частотной характеристики равен единице, а фаза равна π . Поэтому

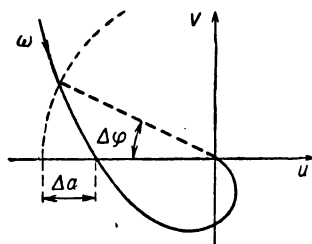


Рис. 6.18. Определение запасов устойчивости замкнутой системы по амплитуде и по фазе с помощью амплитудно-фазовой характеристики второго рода, построенной для разомкнутой системы

запас устойчивости по амплитуде Δa характеризуется отличием модуля $|W(i\omega)|$ от единицы при $\varphi = \pi$, а запас устойчивости по фазе $\Delta \varphi$ — отличием фазы φ от π при $|W(i\omega)| = 1$ (рис. 6.18). Пользуясь этими понятиями, можно исследовать запасы устойчивости системы [6.10].

Критерий Найквиста удобно применять, пользуясь логарифмическими частотными характеристиками. Покажем это на примере системы, устойчивой в разомкнутом состоянии. Амплитудно-фазовая характеристика

такой системы пересекает ось U при $|W(i\omega)| > 1$; значения фазы в точках пересечения кратны π ; на рис. 6.15 $\varphi = \pi$, но, может быть, вообще $\varphi = k\pi$, где k — целое число. Поэтому устойчивая разомкнутая система будет устойчивой в замкнутом состоянии, если в области частот, где $|W(i\omega)| > 1$, фазо-частотная характеристика четное число раз проходит каждое значение $k\pi$ с изменением знака $d\varphi/d\omega$ при каждом следующем прохождении одного и того же значения $k\pi$. Так как ЛАХ $L = 20 \lg |W(i\omega)|$, то условие $|W(i\omega)| > 1$ означает, что $L > 0$. Для применения этого критерия логарифмическая фазо-частотная характеристика строится под логарифмической амплитудной характеристикой с тем же масштабом по оси $\lg \omega$. На рис. 6.19 показано построение, соответствующее рис. 6.15: в области, где $L > 0$, фаза φ дважды принимает значение $-\pi$, меняя знак производной (точки a и b).

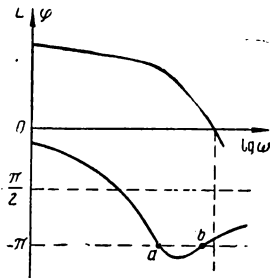


Рис. 6.19. Амплитудная и фазовая логарифмические частотные характеристики устойчивой разомкнутой системы, сохраняющей устойчивость после замыкания

Частотные критерии необходимы для исследования устойчивости систем с запаздыванием. Если передаточная функция такой разомкнутой системы имеет вид (4.21), то после замыкания единичной отрицательной обратной связью (см. рис. 6.13) характеристическое уравнение системы будет иметь вид

$$1 + K \frac{P_1(p)}{P_2(p)} e^{-p\tau} = 0,$$

или

$$P_2(p) + K P_1(p) e^{-p\tau} = 0. \quad (6.48)$$

Это — трансцендентное уравнение, имеющее бесконечное число корней, вследствие чего устойчивость характеризуемой им системы нельзя исследовать с помощью алгебраических критериев. Необходимое условие устойчивости, требующее одинаковости знаков всех членов уравнения, здесь не является достаточным даже для системы первого порядка.

Представляя левую часть уравнения (6.48) на комплексной плоскости и производя в ней подстановку (3.28), построим вектор

$$M(i\omega) = P_2(i\omega) + K e^{-i\omega\tau} P_1(i\omega). \quad (6.49)$$

Этот вектор можно рассматривать как вектор Михайлова, обобщенный на случай $\tau \neq 0$. Обобщение выражается в том, что составляющая $K P_1(i\omega)$ повернута на угол $-\omega\tau$. Отсюда следует, что здесь применим критерий Михайлова.

Теперь построим годографы векторов $P_2(i\omega)$ и $K P_1(i\omega)$, повернем последний годографа на $\tau\omega$ назад и сложим с векторами первого при одинаковых частотах (рис. 6.20). В результате этих действий получим годограф вектора (6.49), который должен удовлетворять критерию Михайлова, чтобы рассматриваемая система была устойчивой, имея порядок, равный степени полинома $P_2(p)$. Этот метод исследования устойчивости систем с запаздыванием предложен и подробно развит Н. Н. Мясниковым [6.6].

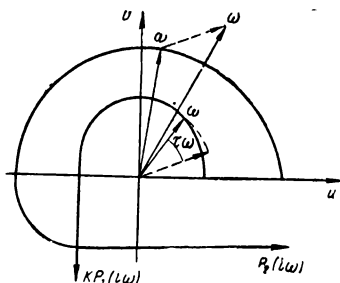


Рис. 6.20. Построение вектора Михайлова для системы с запаздыванием

Пример. Пользуясь критерием Найквиста, найдем значение параметра r , при котором система, имеющая в разомкнутом состоянии характеристическое уравнение (6.8) и коэффициент усиления, равный единице, после замыкания оказывается на границе устойчивости.

Из выражения передаточной функции разомкнутой системы

$$W_p(p) = \frac{1}{p^2 + p + r}$$

находим вещественную и мнимую части частотной характеристики:

$$U(\omega) = \frac{r - \omega^2}{r^2 - \omega^2 + \omega^4},$$

$$V(\omega) = - \frac{\omega}{r^2 - \omega^2 + \omega^4}.$$

Из этих выражений видно, что характеристика проходит через точку $(-1, 0)$ при $r = -1$ и $\omega = 0$. При таком значении параметра r замкнутая система находится на границе устойчивости, хотя в разомкнутом виде она неустойчива, как видно из D -разбиения (см. рис. 6.3), так как не соблюдается необходимое условие устойчивости. Тот же результат легко получить и из характеристического уравнения замкнутой системы при условии, что оно имеет нулевой корень.

§ 6.6. УСТОЙЧИВОСТЬ ИМПУЛЬСНЫХ СИСТЕМ

Устойчивость импульсной системы выражается в таком же затухании переходного процесса, как и у систем непрерывного действия (см. § 6.1). Необходимое и достаточное условие устойчивости состоит в требовании, аналогичном известному для непрерывных систем: вещественные части всех полюсов передаточной функции должны быть отрицательными. Для суждения о выполнении этого условия удобны частотные критерии.

Рассмотрим передаточную функцию замкнутой импульсной системы (см. § 5.5). Ее полюса определяются знаменателем выражения (5.30). В случае единичной отрицательной обратной связи этот знаменатель имеет вид

$$1 + W(q),$$

где $W(q)$ — передаточная функция той же системы при отсутствии обратной связи. Полагая в выражении $W(q)$

$$q = i\beta, \quad (6.50)$$

получим частотную характеристику $W(i\beta)$ разомкнутой импульсной системы. Тогда вектор Найквиста

$$N(i\beta) = 1 + W(i\beta) \quad (6.51)$$

будет определять устойчивость системы после замыкания подобно тому, как это имеет место для непрерывных систем (см. § 6.5). При этом в силу известной периодичности изображений, получаемых посредством дискретного преобразования Лапласа, изменение относительной частоты β следует рассматривать в пределах от 0 до π . Таким образом получаем следующую формулировку критерия Найквиста для импульсных систем: замкнутая система устойчива, если при возрастании относительной частоты от 0 до π годограф вектора Найквиста охватывает точку вещественной оси, имеющую абсциссу -1 , $m/2$ раз, где m — число полюсов передаточной функции разомкнутой системы, имеющих положительную вещественную часть. На рис. 6.21 показан пример годографа системы, которая становится устойчивой после замыкания при $m=2$. При построении годографа по осям координат откладываются вещественная и мнимая части частотной характеристики разомкнутой системы, представляемой в форме

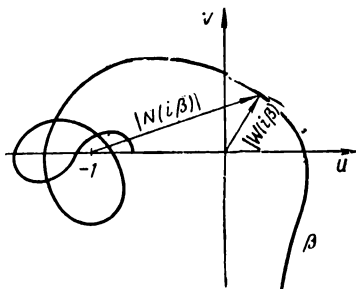


Рис. 6.21. Годограф вектора Найквиста импульсной системы, неустойчивой в разомкнутом состоянии, которая становится устойчивой после замыкания

$$W(i\beta) = U(\beta) + iV(\beta). \quad (6.52)$$

Пример. Исследуем устойчивость импульсной системы, имеющей в разомкнутом состоянии передаточную функцию (4.88). Единственный полюс этой передаточной функции

$$q = -\frac{T_n}{T}.$$

Следовательно, в разомкнутом состоянии рассматриваемая система устойчива. Ее частотную характеристику получим подстановкой (6.50) в (4.88) при $\epsilon = 0$:

$$W(i\beta) = \frac{K/T}{1 - \exp\left(-\frac{T_n}{T} - i\beta\right)}.$$

Выражая здесь экспоненциальный член знаменателя в тригонометрической форме и освобождаясь от комплексного знаменателя, находим:

$$U(\beta) = \frac{k}{T} \frac{1 - e^{-T_n/T} \cos \beta}{1 - 2e^{-T_n/T} \cos \beta + e^{-2T_n/T}},$$

$$V(\beta) = \frac{K}{T} \frac{e^{-T_n/T} \sin \beta}{1 - 2e^{-T_n/T} \cos \beta + e^{-2T_n/T}}.$$

Из последних выражений видно, что при изменении β от 0 до π амплитудно-фазовая характеристика лишь два раза пересекает ось u : $V(\beta) = 0$ при $\beta = 0$ и $\beta = \pi$. Абсциссы этих точек характеристики

$$U(0) = \frac{k/T}{1 - e^{-T_n/T}},$$

$$U(\pi) = \frac{k/T}{1 + e^{-T_n/T}}.$$

Очевидно, что $U(\pi) > 0$, но $U(0)$ может иметь знак минус, если $e^{-T_n/T} > 1$. Следовательно, в зависимости от численных значений параметров k , T и T_n , рассматриваемая система после замыкания может стать либо устойчивой, когда $U(0) > -1$, либо неустойчивой, когда $U(0) < -1$.

ПРОЦЕССЫ В ЛИНЕЙНЫХ СИСТЕМАХ

§ 7.1. СУПЕРПОЗИЦИЯ ВЫХОДНЫХ СИГНАЛОВ

Изучая линейную систему, на вход которой подано несколько сигналов, следует пользоваться присущим такой системе свойством суперпозиции, по которому воздействие каждого входного сигнала на нее можно исследовать независимо от остальных, считая систему находящейся в покое (см. § 4.1). Затем реакции, вызываемые всеми входными сигналами, складываются. Если при этом входные сигналы сдвинуты во времени, т. е. включаются не одновременно, то возникает задача описания сигналов, отображающего моменты их появления [7.8].

Рассмотрим непрерывный выходной сигнал

$$x_2(u) = f_2(u) 1(u - u_0), \quad (7.1)$$

представленный в функции времени, отсчитываемого по некоторой шкале времени u и существующий лишь при $u > u_0$. Этот сигнал будем описывать в зависимости от момента его появления по шкале времени t , так или иначе расположенной относительно шкалы u , но построенной в том же масштабе. Здесь возможны четыре случая, представленных на рис. 7.1, где моменты появления сигнала по шкале t обозначены через t_0 . В первом случае оси u и t расположены так, что их нулевые отметки совпадают. Во втором случае они сдвинуты так, что $u = 0$, когда $t = \tau$. В третьем случае оси u и t расположены так, что $u = 0$ при $t = -\tau$. То же имеет место в четвертом случае, но тут сигнал появляется при $t = -t_0$.

В табл. 7.1 приведены формулы перехода от шкалы t к шкале u , которые легко составить по рис. 7.1 для всех



Рис. 7.1. Шкалы времени u и t

Таблица 7.1

Переход от шкалы времени u к шкале t и выражения сигнала (7.1) в функции времени t

Случай по рис. 7.1	Формула перехода	Сигнал (7.1)
1	$u = t$	$x_2(t) = f_2(t) \cdot 1(t - t_0)$
2	$u = t - \tau$	$x_2(t - \tau) = f_2(t - \tau) \cdot 1(t - t_0)$
3	$u = t + \tau$	$x_2(t + \tau) = f_2(t + \tau) \cdot 1(t - t_0)$
4	$u = t + \tau$	$x_2(t + \tau) = f_2(t + \tau) \cdot 1(t - t_0)$

Таблица 7.2

Выражения сигнала, получаемые из таблицы 7.1 при $u_0 = 0$

Случай по рис. 7.1	Сигнал 7.1
1	$x_2(t) = f_2(t) \cdot 1(t)$
2	$x_2(t - \tau) = f_2(t - \tau) \cdot 1(t - \tau)$
3	$x_2(t + \tau) = f_2(t + \tau) \cdot 1(t + \tau)$
4	$x_2(t + \tau) = f_2(t + \tau) \cdot 1(t + \tau)$

четырёх случаев. Там же даны выражения сигнала в функции времени t , получаемые из (7.1) с помощью формул перехода. В табл. 7.2 приведены выражения того же сигнала, получаемые из табл. 7.1 при $u_0 = 0$.

Можно рассматривать линейную систему с двумя входами a и b , по отношению к которым она характеризуется передаточными функциями $W_a(p)$ и $W_b(p)$. Выходной сигнал будет при этом суперпозицией реакций на сигналы, подаваемые на входы a и b . Аналогично можно найти выходной сигнал и в случае большего числа входов, на которые воздействуют сигналы, сдвинутые во времени. Реакция на любой входной сигнал опять может быть исследована и построена независимо от других реакций.

Так же можно построить переходный процесс, возникающий вследствие выключения входного сигнала при $t = T_c$. Принцип суперпозиции позволяет рассматривать выключение входного сигнала x_1 как дополнительный сигнал x_{12} , подаваемый на вход. В этом случае, следовательно, при $t = 0$ на вход подается сигнал

$$x_{11}(t) = f_1(t) \cdot 1(t), \quad (7.2)$$

а при $t = T_c$ — сигнал

$$x_{12}(t) = -f_1(t) 1(t - T_c). \quad (7.3)$$

Рассматривая $x_{11}(t)$ и $x_{12}(t)$ как составляющие входного сигнала $x_1(t)$, имеем:

$$x_1(t) = x_{11}(t) + x_{12}(t),$$

откуда после подстановки (7.2) и (7.3) следует выражение (3.5) вида

$$x_1(t) = f_1(t) [1(t) - 1(t - T_c)]. \quad (7.4)$$

Пример. Если $u_0 = 0$ и $f_2(u) = \sin \gamma u$, то в третьем случае по табл. 7.2

$$x_2(t + \tau) = (\sin \gamma t \cos \gamma \tau + \cos \gamma t \sin \gamma \tau) 1(t + \tau).$$

§ 7.2. РЕШЕНИЕ ОПЕРАТОРНОГО УРАВНЕНИЯ ДВИЖЕНИЯ НЕПРЕРЫВНОЙ СИСТЕМЫ

Когда известны изображения входного и выходного сигналов, то по (4.14) можно найти передаточную функцию системы. Это значит, что, ничего не зная о системе, можно составить суждение о ее свойствах, рассматривая процесс на ее выходе, вызываемый некоторым процессом на входе. При такой постановке задачи рассматриваемая система является тем, что в кибернетике называют *черным ящиком*. Если же известны свойства системы, определяемые ее передаточной функцией, и входной сигнал, заданный его изображением, то из того же уравнения можно найти реакцию системы.

Рассмотрим процесс (3.4) на выходе непрерывной линейной системы, вызываемый сигналом (3.1), поданным на вход. Предположим, что при $t < 0$ никакого сигнала на входе не было, следовательно, до включения сигнала $x_1(t)$ на выходе не было никакого процесса. Таким образом, система, находящаяся в покое, подвергается внезапному воздействию сигнала $x_1(t)$, вызывающему появление сигнала $x_2(t)$ на выходе.

Считая заданным сигнал $x_1(t)$, будем считать известным его изображение $X_1(p)$. Тогда сигнал $x_2(t)$ можно найти из операторного уравнения движения (4.14), так как передаточная функция системы, реакция которой рассматривается, должна быть известной. Принципиально для этого нужно лишь перейти от изображения $X_2(p)$ к его оригиналу, пользуясь известными приемами опера-

торного исчисления. Согласно сказанному в § 3.2 найденное решение уравнения (4.13) равно нулю при отрицательных значениях времени и единственно. Физически это означает, что воздействие $x_1(t)$, приложенное в момент времени $t=0$ к системе, имеющей определенную передаточную функцию, вызывает в этой системе единственную, вполне определенную реакцию $x_2(t)$. Начальные значения реакции и ее производных определяются тем самым однозначно и отнюдь не могут быть заданы из каких-либо других соображений как произвольно выбираемые *начальные условия*. При исследовании реальных процессов начальные условия, как отметил Бриллюэн, «никогда не задаются» [7.16].

Упуская физическое обоснование единственности решения уравнения (4.13), часто полагают, будто это решение описывает процесс, начинающийся при *нулевых начальных условиях*. Это верно в том смысле, что исследуется процесс, вызываемый воздействием, приложенным к системе, находящейся в покое. Но поскольку воздействие имеет вид (3.1), такая интерпретация заставляет различать *начальные условия слева* и *справа* от точки $t=0$. Если известно, что начальные условия нулевые, то это относится, как правило, к условиям «слева». Таковы нулевые начальные условия для системы, находящейся в покое. Между тем, пользуясь односторонним преобразованием Лапласа, можно рассматривать только то, что происходит справа от $t=0$. Отсюда возникла задача перехода от начальных условий «слева» к начальным условиям «справа», получившая довольно громоздкое решение [7.15]. Из сказанного выше видно, что для решения уравнения (4.13), т. е. для построения переходного процесса, переход от начальных условий слева к начальным условиям справа вовсе не нужен [7.8].

Покажем, как непосредственно по изображению $X_2(p)$ можно, не переходя к оригиналу, выяснить вид переходного процесса в момент его возникновения при $t=0$, а во многих случаях и установившееся значение сигнала, т. е. $\lim_{t \rightarrow \infty} x_2(t)$.

По предельной теореме операторного исчисления начальное значение (при $t=0$) описывающей функции выходного сигнала

$$f_2(0) = \lim_{p \rightarrow \infty} pX_2(p). \quad (7.5)$$

Здесь изображение выходного сигнала, получаемое из уравнения (4.14), при заданном входном сигнале обычно имеет вид

$$X_2(p) = \frac{b_0 p^\mu + b_1 p^{\mu-1} + \dots + b_{\mu-1} p + b_\mu}{a_0 p^\nu + a_1 p^{\nu-1} + \dots + a_{\nu-1} p + a_\nu}, \quad (7.6)$$

так как передаточная функция $W(p)$ в большинстве случаев, которые далее и будем предполагать, — рациональная дробь, а изображение $X_1(p)$ в практических задачах бывает правильной дробью, когда входной сигнал непрерывен (3.1), и неправильной, когда он имеет импульсивные составляющие (3.11).

Если $\mu \geq \nu$, то выходной сигнал содержит импульсивные составляющие, т. е. имеет вид (3.11). При

$$\nu > \mu, \quad (7.7)$$

когда импульсивных составляющих на выходе нет, из (7.5) получаем следующие начальные значения описывающей функции выходного сигнала:

$$f_2(0) = \begin{cases} 0 & \nu > \mu + 1 \\ \frac{b_0}{a_0} & \nu = \mu + 1. \end{cases} \quad (7.8)$$

Таким образом, если $\nu > \mu + 1$, то выходной сигнал при $t = 0$ непрерывен; если же $\nu = \mu + 1$, то он появляется скачком, равным b_0/a_0 (рис. 7.2). Следовательно, по (4.17) его обобщенная производная в последнем случае

$$Dx_2 = f_2'(t) 1(t) + \frac{b_0}{a_0} \delta(t),$$

и нельзя говорить о ее начальном значении, так как она содержит импульс. В этом случае можно рассматривать лишь начальное значение производной $f_2'(t)$.

Из формул (4.11) и (7.8) имеем:

$$pX_2(p) - \frac{b_0}{a_0} \doteq f_2'(t) 1(t),$$

и по предельной теореме

$$f_2'(0) = \lim_{p \rightarrow \infty} p \left[pX_2(p) - \frac{b_0}{a_0} \right],$$

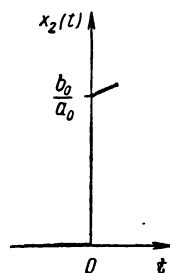


Рис. 7.2. Выходной сигнал, появляющийся скачком

откуда находим, что при $\nu = \mu + 1$ первая производная описывающей функции выходного сигнала имеет начальное значение

$$f'_2(0) = \frac{1}{a_0} \left(b_1 - a_1 \frac{b_0}{a_0} \right). \quad (7.9)$$

Легко убедиться, что при $\nu = \mu + \lambda$, $\lambda > 0$

$$f_2(0) = f'_2(0) = \dots = f_2^{(\lambda-2)}(0) = 0, \quad (7.10)$$

но

$$f_2^{(\lambda-1)}(0) = \frac{b_0}{a_0}. \quad (7.11)$$

Следовательно, x_2 , Dx_2 , ..., $D^{\lambda-2}x_2$ при $t=0$ непрерывны, причем $D^{\lambda-2}x_2$ имеет угловую точку, $D^{\lambda-1}x_2$ — скачок b_0/a_0 , а $D^\lambda x_2$ и высшие производные содержат импульсы. Это означает, что функция $x_2(t)$ имеет порядок гладкости $\lambda - 2$. Далее находим:

$$f_2^{(\lambda)}(0) = \frac{1}{a_0} [b_1 - a_1 f_2^{(\lambda-1)}(0)], \quad (7.12)$$

$$f_2^{(\lambda+1)}(0) = \frac{1}{a_0} [b_2 - a_2 f_2^{(\lambda-1)}(0) - a_1 f_2^{(\lambda)}(0)], \quad (7.13)$$

$$f_2^{(\lambda+2)}(0) = \frac{1}{a_0} [b_3 - a_3 f_2^{(\lambda-1)}(0) - a_2 f_2^{(\lambda)}(0) - a_1 f_2^{(\lambda+1)}(0)],$$

и т. д. Вообще

$$f_2^{(\lambda+x)}(0) = \frac{1}{a_0} [b_{x+1} - a_{x+1} f_2^{(\lambda-1)}(0) - a_x f_2^{(\lambda)}(0) - \dots - a_1 f_2^{(\lambda+x-1)}(0)], \quad (7.14)$$

где целое число

$$x \geq -1.$$

Выражение (7.14) представляет собой формулу Троицкого [7.12]. При отыскании по ней начального значения производной некоторого порядка при заданном λ определяется x и вычисляется столько членов в квадратных скобках, чтобы последний из них имел коэффициент a_1 и содержал начальное значение производной порядка $\lambda + x - 1$, т. е. на единицу меньшего, чем порядок производной, начальное значение которой является искомым. Например, при $\lambda = 1$ для вычисления $f'_2(0)$ следует положить

$$\lambda + x = 1,$$

откуда $x = 0$ и $\lambda + x - 1 = 0$.

Следовательно,

$$\dot{f}_2(0) = \frac{1}{a_0} [b_1 - a_1 f_2(0)],$$

откуда, принимая во внимание (7.6), получаем ранее введенную формулу (7.9) как частный случай формулы Тропского.

Другое предельное свойство изображения позволяет найти .

$$f_2(\infty) = \lim_{p \rightarrow 0} p X_2(p), \quad (7.15)$$

если только изображение $X_2(p)$ не имеет полюсов на мнимой оси и правее. Очевидно, $f_2(\infty) = 0$ во всех случаях, когда в (7.6) $a_0 \neq 0$.

Полученные таким образом значения функции $f_2(t)$ при $t=0$ и $t=\infty$, а также значения производных этой функции при $t=0$ позволяют определить вид переходного процесса в его начале и установившееся значение выходного сигнала. Для этого при заданном выходном сигнале нужно знать лишь передаточную функцию рассматриваемой системы.

Таким образом, начало и конец процесса, вызываемого на выходе линейной системы внезапно приложенным воздействием, однозначно определяются видом этого воздействия и передаточной функцией системы.

Рассмотрим теперь реакцию непрерывной линейной системы без запаздывания на последовательность импульсивных сигналов (3.11). В промежутке времени между подачей N -го и $(N+1)$ -го импульса, т. е. при

$$(N-1)T < t < NT, \quad (7.16)$$

выходной сигнал $x_2(t, N)$ определяется конечным числом n импульсов, поданных на вход, т. е. [7.14] входным сигналом

$$x_1(t, N) = \Delta \sum_{m=0}^{N-1} f_1(mT) \delta(t - mT). \quad (7.17)$$

Конечная последовательность импульсивных сигналов (7.17) имеет изображение

$$X_1(p, N) = \Delta \sum_{m=0}^{N-1} f_1(mT) e^{-mpT}. \quad (7.18)$$

Следовательно, в промежутке времени (7.16) выходной сигнал имеет изображение

$$X_2(p, N) = \Delta \sum_{m=0}^{N-1} W(p) f_1(mT) e^{-mpT}, \quad (7.19)$$

где $W(p)$ — передаточная функция рассматриваемой системы.

В случае системы, характеристическое уравнение которой имеет вид $\prod_{k=1}^n (p - p_k) = 0$ (кратных корней нет), передаточная функция

$$W(p) = \sum_{k=1}^n \frac{c_k}{p - p_k}. \quad (7.20)$$

В таком случае выходной сигнал найдем как суперпозицию n сигналов, имеющих изображения вида

$$Y(p, N) = \frac{c_k \Delta}{p - p_k} \sum_{m=0}^{N-1} f_1(mT) e^{-mpT} \quad (7.21)$$

и одинаковую шкалу отсчета времени.

Таким образом, в промежутке времени (7.16) общее выражение составляющих выходного сигнала имеет вид

$$y(t, N) = \Delta c_k \sum_{m=0}^{N-1} f_1(mT) e^{p_k(t-mT)}. \quad (7.22)$$

Легко видеть, что в следующем промежутке времени, когда $NT < t < (N+1)T$, т. е. после подачи еще одного импульса, та же составляющая

$$y(t, N+1) = y(t, N) + \Delta c_k f_1(NT) e^{p_k(t-NT)}. \quad (7.23)$$

Второе слагаемое правой части этого выражения есть реакция системы на N -й импульс. Отсюда следует, что при переходе с $(N-1)$ -го участка процесса на N -й, т. е. при появлении N -го импульса, рассматриваемая составляющая выходного сигнала имеет скачок, равный $\Delta c_k f_1(NT)$. Скачок выходного сигнала равен сумме скачков его составляющих, т. е. равен

$$\Delta f_1(nT) \sum_{k=1}^n c_k.$$

Замечая, что изображение (7.21) имеет лишь один простой полюс $p = -\theta$, и написав его в виде

$$Y(p, N) = \frac{c}{p + \theta} X_1(p, N),$$

можно найти его оригинал:

$$y(t, N) = c \operatorname{Res}_{p=-\theta} \left[\frac{1}{p + \theta} X_1(p, N) e^{pt} \right],$$

откуда [7.14]

$$y(t, N) = c X_1(-\theta, N) e^{-\theta t}. \quad (7.24)$$

Применим изложенную теорию к случаю гармонически модулированных импульсов (когда $f_1(t) = a \cos \gamma t$), воздействующих на систему с передаточной функцией (7.20) [7.14]. Чтобы получить изображение $X_1(p, N)$, представим входной сигнал (7.17) в форме

$$x_1(t, N) = \Delta f_1(t) \sum_{m=0}^{N-1} \delta(t - mT). \quad (7.25)$$

Подставляя сюда

$$f_1(t) = a \frac{e^{i\gamma t} + e^{-i\gamma t}}{2},$$

находим, что для получения искомого изображения следует воспользоваться соответствием, основанным на теореме смещения в комплексной плоскости:

$$e^{\pm i\gamma t} \sum_{m=0}^{N-1} \delta(t - mT) \doteq \sum_{m=0}^{N-1} e^{-m(p \mp i\gamma)T} = \frac{1 - e^{-N(p \mp i\gamma)T}}{1 - e^{-(p \mp i\gamma)T}}.$$

Тогда после преобразований имеем:

$$X_1(p, N) = \frac{a\Delta}{C^2(p)} \{A(p) - C(p) \cos[N\gamma T + \varphi(p)] e^{-NpT}\}, \quad (7.26)$$

где

$$A(p) = 1 - e^{-pT} \cos \gamma T, \quad B(p) = e^{-pT} \sin \gamma T,$$

$$C(p) = \sqrt{A^2(p) + B^2(p)},$$

$$\cos \varphi(p) = \frac{A(p)}{B(p)}, \quad \sin \varphi(p) = \frac{B(p)}{C(p)}.$$

Теперь по (7.26) находим:

$$X_1(-\theta, N) = \frac{a\Delta}{C^2} [A - C \cos(N\gamma T + \varphi) e^{N\theta T}],$$

где

$$A = 1 - e^{\theta T} \cos \gamma T, \quad B = e^{\theta T} \sin \gamma T, \quad C = \sqrt{A^2 + B^2},$$

$$\cos \varphi = \frac{A}{C}, \quad \sin \varphi = \frac{B}{C}.$$

Следовательно, по (7.24)

$$y(t, N) = \frac{ca\Delta}{C^2} [A - C \cos(N\gamma T + \varphi) e^{N\theta T}] e^{-\theta t}. \quad (7.27)$$

Пусть $c=1$, $a=1$, $\gamma = \frac{\pi}{4}$, $\theta = \ln \sqrt{2}$, $\Delta = T = 1$. Тогда $A=0$, $C=B=1$, $\varphi = \frac{\pi}{2}$. На рис. 7.3, а показан

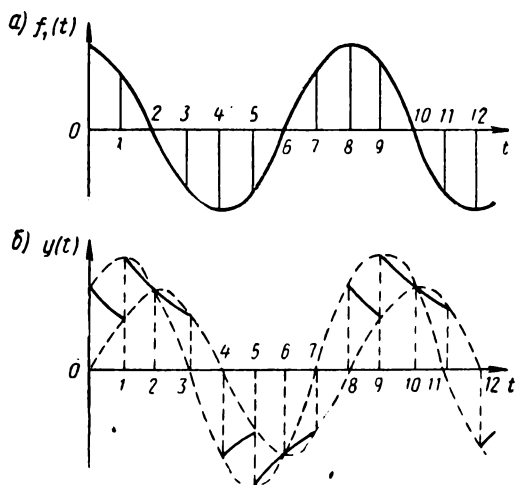


Рис. 7.3. К построению переходного процесса, возникающего в непрерывной системе под действием гармонически модулированной последовательности импульсных сигналов:

а) модулирующая функция; б) переходный процесс

график функции $f_1(t)$, ординаты которого при $t = NT$ в данном случае, согласно (7.23), равны скачкам выходного сигнала. Переходный процесс построен на рис. 7.3, б [7.14] по уравнению (7.27), принимающему вид:

$$y(t, N) = \sin \frac{\pi N}{4} \exp [(N - t) \ln \sqrt{2}]. \quad (7.28)$$

Скачки получаются при $t = N$, когда оканчивается N -й участок процесса и y имеет значения, определяемые нижней огибающей

$$y = \sin \frac{\pi}{4} t. \quad (7.29)$$

После скачка y принимает значение, определяемое верхней огибающей, уравнение которой получаем из (7.28) при увеличении N на единицу:

$$y = \sqrt{2} \sin \frac{\pi}{4} (t + 1). \quad (7.30)$$

В табл. 7.3 приведены выражения $x_1(t, N)$ и $y(t, N)$, полученные [7.14] аналогично еще для трех модулирующих функций: постоянной, показательной и линейной.

Таблица 7.3

К построению переходных процессов, вызываемых в системе с передаточной функцией $\frac{c}{p + \theta}$ последовательностью модулированных импульсивных сигналов

$f_1(t)$	a	$ae^{-\alpha t}$	at
$X_1(p, N)$	$a\Delta \frac{1 - e^{-NpT}}{1 + e^{-pT}}$	$a\Delta \frac{1 - e^{-N(p+\alpha)T}}{1 - e^{-(p+\alpha)T}}$	$\frac{a\Delta T e^{-pT}}{(1 - e^{-pT})^2} [1 + (N-1) \times \times e^{-NpT} - Ne^{-(N-1)pT}]$
$y(t, N)$	$\frac{ac\Delta}{e^{\theta T} - 1} (e^{N\theta T} - 1) e^{-\theta t}$	$\frac{ac\Delta}{1 - e^{(\theta-\alpha)T}} [1 - e^{N(\theta-\alpha)T}] \times \times e^{-\theta t}$	$\frac{ac\Delta T}{e^{\theta T} - 1} [1 + N(1 - e^{-\theta T}) \times \times e^{N\theta T} - e^{N\theta T}] e^{-\theta t}$

В общем случае системы с передаточной функцией $W(p)$ скачок выходного сигнала при подаче на вход N -го импульса, т. е. при $t = NT$, равен [7.14]:

$$\Delta f_1(NT) \lim_{p \rightarrow \infty} pW(p). \quad (7.31)$$

Отсюда начальное значение описывающей функции выходного сигнала

$$f_2(0) = \Delta f_1(0) \lim_{p \rightarrow \infty} pW(p), \quad (7.32)$$

что следует также из (7.5) после подстановки (7.19). Выражение (7.32) показывает, что порядок гладкости выходного сигнала, вызываемого входным (3.11), зависит только от передаточной функции $W(p)$ и не зависит от модулирующей функции $f_1(t)$. Поэтому сделанные выше заключения о характере выходного сигнала и его производных, а также формула (7.14) применимы и в случае входного сигнала (3.11), если μ и ν заменить степенями m и n оператора воздействия и собственного оператора системы, а под b и a (7.14) понимать коэффициенты этих операторов.

Примеры. 1. Как показывает выражение переходной характеристики (4.82), при $t=0$ эта характеристика имеет скачок:

$$h(0) = K \frac{T_1}{T_2}. \quad (7.33)$$

Это видно непосредственно из уравнения (4.81), в котором $\nu=2$, $\mu=1$, т. е. $\lambda=1$; при таком значении разности $\nu-\mu$, как было показано выше, $x_2(t)$ всегда делает скачок при $t=0$. Выражение $h(0)$ можно получить и из (7.6) [7.8].

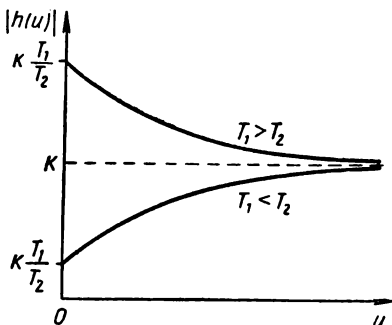


Рис. 7.4. Переходные характеристики звена, имеющего передаточную функцию (4.80)

2. Составим уравнение переходного процесса, возникающего, на выходе звена с передаточной функцией (4.80), если входной сигнал в виде единичного скачка, включенный при $t=0$, выключается при $t=T_c$, т. е. имеет выражение (7.4) при $f_1(t)=1$:

$$x_1(t) = 1(t) - 1(t - T_c). \quad (7.34)$$

Такой сигнал можно рассматривать как сумму двух составляющих вида $\pm 1(u)$.

Первая составляющая имеет верхний знак и соответствует первому случаю по табл. 7.2; вторая имеет нижний знак и соответствует второму случаю по той же таблице, причем $\tau=T_c$. Изображения этих составляющих равны, соответственно, $\pm 1/p$, а изображения вызываемых ими выходных сигналов $h(u)$, по (4.81), равны

$$\pm K \frac{1 + T_1 p}{(1 + T_2 p)p}. \quad (7.35)$$

Здесь $\mu=1$, $\nu=2$, $\lambda=1$. Следовательно, по (7.6)

$$f_2(0) = \frac{b_0}{a_0} = \pm K \frac{T_1}{T_2},$$

т. е. при $u=0$ соответствующие сигналы на выходе появляются скачком. Найдем установившееся значение каждого сигнала по (7.15):

$$f_2(\infty) = \pm K.$$

На рис. 7.4 по этим данным показан примерный график сигнала $h(u)$ при $T_1 > T_2$ и $T_1 < T_2$, получаемый таким образом без перехода от изображения к оригиналу.

Переходя теперь к оригиналу, воспользовавшись (4.82), получим:

$$h(u) = \pm K \left[1 + \left(\frac{T_1}{T_2} - 1 \right) e^{-u/T_2} \right] 1(u). \quad (7.36)$$

Здесь для реакции на первую составляющую входного сигнала $u=t$ знак верхний; для реакции на вторую составляющую $u=t-T_c$ и знак нижний. Производя суперпозицию сигналов (7.36), находим выходной сигнал:

$$x_2(t) = K \left\{ \left[1 + \left(\frac{T_1}{T_2} - 1 \right) e^{-t/T_2} \right] 1(t) - \left[1 + \left(\frac{T_1}{T_2} - 1 \right) e^{-\frac{t-T_c}{T_2}} \right] 1(t-T_c) \right\}, \quad (7.37)$$

т. е.

$$x_2(t) = \begin{cases} 0 & t < 0 \\ K \left[1 + \left(\frac{T_1}{T_2} - 1 \right) e^{-t/T_2} \right] & 0 < t < T_c \\ -K \left(\frac{T_1}{T_2} - 1 \right) (e^{T_1/T_2} - 1) e^{-t/T_2} & t > T_c \end{cases} \quad (7.38)$$

Составляя изображение сигнала (7.34), можно получить также этот результат:

$$X_1(p) = \frac{1}{p} - \frac{1}{p} e^{-pT_c},$$

откуда по (4.14)

$$X_2(p) = K \left[\frac{1 + T_1 p}{(1 + T_2 p) p} - \frac{1 + T_1 p}{(1 + T_2 p) p} e^{-pT_c} \right]$$

и после перехода к оригиналу — (7.37) [7.8].

Как видно из (7.37), при $t=0$ выходной сигнал имеет скачок, равный $K \frac{T_1}{T_2}$. Далее, вычитая из последнего выражения (7.38) предыдущее, находим, что при $t=T_c$ сигнал имеет скачок, равный $-K \frac{T_1}{T_2}$. Это показано на рис. 7.5, а для случая $T_1 < T_2$, а на рис. 7.5, б — для случая $T_1 > T_2$.

3. Найдем переходную характеристику системы четвертого порядка, имеющей передаточную функцию

$$W(p) = \frac{b_0 p^4 + b_1 p^3 + b_2 p^2 + b_3 p + b_4}{p^4 + a_1 p^3 + a_2 p^2 + a_3 p + a_4}. \quad (7.39)$$

Знаменатель этого выражения всегда может быть разложен на квадратичные сомножители [7.5]:

$$p^4 + a_1 p^3 + a_2 p^2 + a_3 p + a_4 = (p^2 + m_1 p + n_1)(p^2 + m_2 p + n_2).$$

Тогда, составляя изображение переходной характеристики и разлагая его на простейшие дроби, имеем:

$$\frac{b_0 p^4 + b_1 p^3 + b_2 p^2 + b_3 p + b_4}{p(p^2 + m_1 p + n_1)(p^2 + m_2 p + n_2)} = \frac{A}{p} + \frac{M_1 p + N_1}{p^2 + m_1 p + n_1} + \frac{M_2 p + N_2}{p^2 + m_2 p + n_2}.$$

Освобождаясь от знаменателей и приравнивая коэффициенты при одинаковых степенях p в левой и правой частях этого выра-

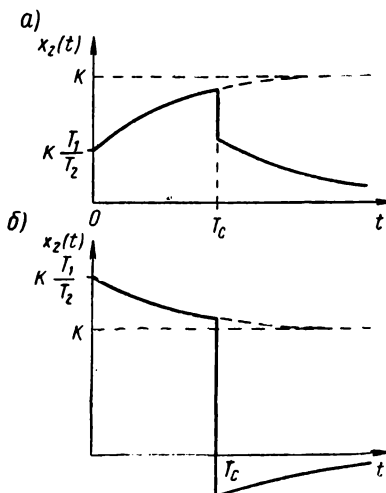


Рис. 7.5. Переходный процесс, вызываемый в звене с передаточной функцией (4.80) сигналом (7.34):

а) при $T_1 < T_2$; б) при $T_1 > T_2$.

жения, получаем систему из пяти линейных уравнений для определения неизвестных A , M_1 , M_2 , N_1 и N_2 :

$$\begin{aligned} A + M_1 + M_2 &= b_0, \\ (m_1 + m_2) A + m_2 M_1 + m_1 M_2 + N_1 + N_2 &= b_1, \\ (n_1 + m_1 m_2 + n_2) A + n_2 M_1 + n_1 M_2 + m_2 N_1 + m_1 N_2 &= b_2, \\ (m_1 n_2 + m_2 n_1) A + n_2 N_1 + n_1 N_2 &= b_3, \\ n_1 n_2 A &= b_4. \end{aligned}$$

Определитель этой системы

$$D = n_1 n_2 [(m_1 - m_2)(n_1 m_2 - n_2 m_1) - (n_1 - n_2)^2]$$

не равен нулю. По формуле Крамера

$$A = \frac{D_1}{D}, \quad M_1 = \frac{D_3}{D}, \quad M_2 = \frac{D_5}{D},$$

$$N_1 = \frac{D_4}{D}, \quad N_2 = \frac{D_6}{D},$$

где

$$D_1 = \frac{b_4 D}{n_1 n_2},$$

$$D_3 = n_2 \{n_1 [b_0 (m_1 m_2 n_1 - n_1^2 - m_1^2 n_2 + n_1 n_2) + b_1 (m_1 n_2 - m_2 n_1) + \\ + b_2 (n_1 - n_2) - b_3 (m_1 - m_2)] + b_4 (m_1^2 - n_1 - m_1 m_2 + n_2)\},$$

$$D_5 = n_1 \{n_2 [b_0 (m_1 m_2 n_2 - n_2^2 - m_2^2 n_1 + n_1 n_2) - b_1 (m_1 n_2 - m_2 n_1) - \\ - b_2 (n_1 - n_2) + b_3 (m_1 - m_2)] + b_4 (m_2^2 + n_1 - m_1 m_2 - n_2)\},$$

$$D_4 = n_2 \{n_1 [b_0 n_1 (m_2 n_1 - m_1 n_2) - b_1 n_1 (n_1 - n_2) + b_2 (m_1 - m_2) + \\ + b_3 (m_1 m_2 - n_2 - m_1^2 + n_1)] + \\ + b_4 (m_1^2 - m_1^2 m_2 + m_1 n_2 - 2m_1 n_1 + m_2 n_1)\},$$

$$D_6 = n_1 \{n_2 [b_0 n_2 (m_1 n_2 - m_2 n_1) + b_1 n_2 (n_1 - n_2) - b_2 n_2 (m_1 - m_2) + \\ + b_3 (m_1 m_2 - n_1 - m_2^2 + n_2)] + \\ + b_4 (m_2^2 + m_1 n_2 - m_1 m_2^2 + m_2 n_1 - 2m_2 n_2)\}.$$

Чтобы получить искомую характеристику как табличный оригинал, представим ее изображение в виде

$$H(p) = \frac{A}{p} + \frac{M_1 \left(p + \frac{m_1}{2}\right) + N'_1}{\left(p + \frac{m_1}{2}\right)^2 + n_1 - \frac{m_1^2}{4}} + \frac{M_2 \left(p + \frac{m_2}{2}\right) + N'_2}{\left(p + \frac{m_2}{2}\right)^2 + n_2 + \frac{m_2^2}{4}},$$

где

$$N'_1 = N_1 - \frac{1}{2} m_1 M_1,$$

$$N'_2 = N_2 - \frac{1}{2} m_2 M_2.$$

Тогда переходная характеристика

$$h(t) = \left[A + e^{-\frac{m_1}{2} t} \left(M_1 \cos \gamma_1 t + \frac{N'_1}{\gamma_1} \sin \gamma_1 t \right) + \right. \\ \left. + e^{-\frac{m_2}{2} t} \left(M_2 \cos \gamma_2 t + \frac{N'_2}{\gamma_2} \sin \gamma_2 t \right) \right] 1(t),$$

где

$$\gamma_1 = \sqrt{n_1 - \frac{m_1^2}{4}},$$

$$\gamma_2 = \sqrt{n_2 - \frac{m_2^2}{4}}.$$

Ее можно также представить в форме [7.7]:

$$h(t) = \left[A + A_1 e^{-\frac{m_1}{2}t} \sin(\gamma_1 t + \varphi_1) + A_2 e^{-\frac{m_2}{2}t} \sin(\gamma_2 t + \varphi_2) \right] l(t),$$

где

$$A_1 = \sqrt{M_1^2 + \frac{N_1'^2}{\gamma_1^2}},$$

$$A_2 = \sqrt{M_2^2 + \frac{N_2'^2}{\gamma_2^2}},$$

а начальные фазы φ_1 и φ_2 определяются формулами:

$$\sin \varphi_1 = \frac{M_1}{A_1},$$

$$\cos \varphi_1 = \frac{N_1'}{A_1 \gamma_1},$$

$$\sin \varphi_2 = \frac{M_2}{A_2},$$

$$\cos \varphi_2 = \frac{N_2'}{A_2 \gamma_2}.$$

4. Исследуем выходной сигнал, вызываемый в системе с передаточной функцией $W(p) = \frac{K}{p}$ последовательностью (3.11), модулированной функцией $f_1(t) = a$ [7.14]. Обращаясь к табл. 7.3, найдем, что в этом случае

$$X_1(0, N) = a\Delta \lim_{\theta \rightarrow 0} \frac{1 - e^{-NpT}}{1 - e^{-pT}} = a\Delta N$$

и по (7.24), полагая $c = K$,

$$x_2(t, N) = Ka\Delta N$$

Из выражения передаточной функции имеем:

$$\lambda = n - m = 1;$$

следовательно, порядок гладкости выходного сигнала

$$\lambda - 2 = -1,$$

т. е. он появляется скачком. По (7.29) начальный скачок, как и все последующие, равен $aK\Delta$. Таким образом, график выходного сиг-

нала — ступенчатая линия (рис. 7.6), ограниченная полупрямыми

$$y_1 = Ka \frac{\Delta}{T} t$$

и

$$y_2 = Ka \frac{\Delta}{T} (t + T).$$

§ 7.3. ДИФФЕРЕНЦИАЛЬНЫЕ УРАВНЕНИЯ ПЕРЕХОДНЫХ ПРОЦЕССОВ

Напишем уравнение движения (4.7) в виде

$$(T_n D^n + T_{n-1} D^{n-1} + \dots + T_1 D + 1) x_2(t) = \\ = K (\tau_m D^m + \tau_{m-1} D^{m-1} + \dots + \tau_1 D + 1) x_1(t). \quad (7.40)$$

Подставляя сюда выражения входного и выходного сигналов (3.1) и (3.4), раскрывая скобки и вычисляя обобщенные производные по (3.20), получим в обеих частях уравнения члены, содержащие импульсы разных порядков. Члены с импульсами одного и того же порядка в левой и правой частях должны быть равны. Это условие дает n алгебраических уравнений, определяющих начальные значения описывающей функции выходного сигнала и ее производных всех порядков до $(n-1)$ -го включительно. Так, член левой части уравнения (7.40), в который входит импульс высшего порядка, есть $T_n f_2(0) \delta^{n-1}$. Поскольку $m < n$ (см. § 4.1), в правой части такой импульс не присутствует. Поэтому уравнение, из которого определяется начальное значение описывающей функции, имеет вид

$$T_n f_2(0) = 0,$$

откуда $f_2(0) = 0$. Получаемые таким способом выражения начальных значений функции $f_2(t)$ и ее производных дает формула (7.14).

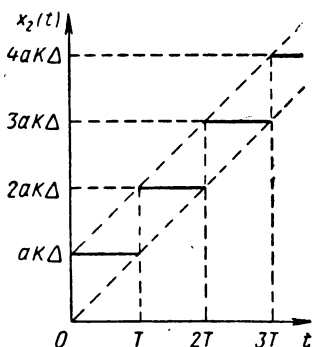


Рис. 7.6. К примеру построения переходного процесса, вызываемого последовательностью одинаковых импульсивных сигналов

Таким образом, импульсивные члены, появляющиеся в уравнении (7.40), после подстановки выражений обобщенных производных взаимно уничтожаются. Остаются лишь члены, содержащие единичную функцию, которая сокращается, и уравнение принимает вид

$$\sum_{k=1}^n T_k f_2^{(k)}(t) + f_2(t) = K \left[\sum_{k=1}^m \tau_k f_1^{(k)}(t) + f_1(t) \right].$$

Пользуясь введенными в § 4.1 обозначениями полиномов P_1 и P_2 , последнее уравнение можно сокращенно записать так:

$$P_2 \left(\frac{d}{dt} \right) f_2(t) = K P_1 \left(\frac{d}{dt} \right) f_1(t). \quad (7.41)$$

Это — неоднородное дифференциальное уравнение, из которого при заданной функции $f_1(t)$ может быть найдена функция $f_2(t)$.

Начальными условиями для интегрирования (7.41) являются начальные значения функции $f_2(t)$ и ее производных, определяемые по (7.14). Они однозначно определяются передаточной функцией и приложенным воздействием, но не могут быть ни заданы, ни выбраны как-либо иначе.

Таким образом, описывающую функцию выходного сигнала можно получить путем интегрирования дифференциального уравнения (7.41). Тем самым определяется и выходной сигнал.

Функцию $f_2(t)$ можно искать и как решение некоторого однородного дифференциального уравнения. Из (7.6) имеем:

$$(a_0 p^v + a_1 p^{v-1} + \dots + a_{v-1} p + a_v) x_2(p) = b_0 p^\mu + b_1 p^{\mu-1} + \dots + b_{\mu-1} p + b_\mu. \quad (7.42)$$

Каждый член этого уравнения как в правой части, так, если раскрыть скобки, и в левой представляет собой произведение одного из коэффициентов a, b на выражение вида $p^k X(p)$, являющееся, согласно (4.11), изображением обобщенной производной $D^k x(t)$, где $x(t)$ — оригинал изображения $X(p)$. В левой части (7.42)

$$X(p) = X_2(p) \doteq x_2(t).$$

В правой части —

$$X(p) = 1 \doteq \delta(t).$$

Поэтому уравнение (7.42) можно преобразовать в уравнение в обобщенных производных:

$$\begin{aligned} & (a_0 D^\nu + a_1 D^{\nu-1} + \dots + a_{\nu-1} D + a_\nu) x_2(t) = \\ & = (b_0 D^\mu + b_1 D^{\mu-1} + \dots + b_{\mu-1} D + b_\mu) \delta(t). \end{aligned} \quad (7.43)$$

Пользуясь (3.20) для вычисления обобщенных производных, вычислим члены левой части (7.43), начиная с последнего. При этом, полагая, что

$$\nu - \mu = \lambda, \quad (7.44)$$

учтем условие (7.10). Тогда

$$\begin{aligned} a_{\nu} x_2(t) &= a_{\nu} f_2(t) \mathbf{1}(t), \\ a_{\nu-1} D x_2(t) &= a_{\nu-1} f_2'(t) \mathbf{1}(t), \\ a_{\nu-2} D^2 x_2(t) &= a_{\nu-2} f_2''(t) \mathbf{1}(t), \end{aligned}$$

$$\begin{aligned} a_{v-\lambda} D^\lambda x_2(t) &= a_{v-\lambda} [f_2^{(\lambda)}(t) 1(t) + f_2^{(\lambda-1)}(0) \delta(t)], \\ a_{v-\lambda-1} D^{\lambda+1} x_2(t) &= a_{v-\lambda-1} [f_2^{\lambda+1}(t) 1(t) + f_2^\lambda(0) \delta(t) + \\ &\quad + f_2^{(\lambda-1)}(0) \delta'(t)], \end{aligned}$$

$$\begin{aligned} \dots\dots\dots \\ a_0 D^\nu x_2(t) = a_0 [f_2^{(\nu)}(t) - 1(t) + f_2^{(\nu-1)}(0)\delta(t) + \\ + f_2^{(\nu-2)}(0)\delta'(t) + \dots + f_2^{(\lambda-1)}(0)\delta^{(\nu-\lambda)}(t)]. \end{aligned}$$

Принимая во внимание, что

$$D^k \delta(t) = \delta^{(k)}(t).$$

сопоставим теперь члены обеих частей уравнения (7.43), содержащие импульсы одинаковых порядков. Начнем с импульса высшего порядка, которым как в правой части, так, согласно (7.44), и в левой является импульс $\delta^{(\mu)}(t)$. В левую часть этот импульс входит с коэффициентом $a_0 f_s^{(\lambda-1)}(0)$. По (7.11)

$$f_2^{(\lambda-1)}(0) = \frac{b_0}{a_0}.$$

Следовательно, в левой части коэффициент при импульсе $\delta^k(t)$ равен b_0 , т. е. такой же, как и в правой, так что члены, содержащие этот импульс, взаимно уничтожаются.

Следующий импульс $\delta^{(\mu-1)}(t)$ входит в левую часть умноженным на $a_0 f_2^{(\lambda)}(0) + a_1 f_2^{(\lambda-1)}(0)$, что равно b_1 , так

как по (7.12)

$$f_2^{(\lambda)}(0) = \frac{1}{a_0} \left(b_1 - a_1 \frac{b_0}{a_0} \right).$$

С таким же коэффициентом импульс $\delta^{(\mu-1)}(t)$ входит и в правую часть, а следовательно, содержащие его члены также взаимно уничтожаются. Подобным же образом, пользуясь (7.14), нетрудно убедиться, что в (7.43) взаимно уничтожаются члены, содержащие импульсы всех остальных порядков. В результате (7.43) превращается в обыкновенное однородное дифференциальное уравнение ν -го порядка

$$a_0 f_2^{(\nu)}(t) + a_1 f_2^{(\nu-1)}(t) + \dots + a_{\nu-1} \dot{f}_2(t) + a_\nu f_2(t) = 0.$$

Обозначая через $A_\nu(p)$ полином, стоящий в знаменателе изображения (7.6), последнее уравнение можно записать сокращенно как

$$A_\nu \left(\frac{d}{dt} \right) f_2(t) = 0. \quad (7.45)$$

Функцию $f_2(t)$ можно найти посредством интегрирования уравнения (7.45) при начальных условиях, заменяемых начальными значениями (7.14):

$$f_2^x(0) = \begin{cases} 0 & 0 \leq x \leq \lambda - 2 \\ (7.14) & \lambda - 1 \leq x \leq \nu - 1. \end{cases}$$

Как и (7.41), уравнение (7.45) является *дифференциальным уравнением* переходного процесса на выходе системы при заданном входном сигнале. Каков бы ни был входной сигнал, это дифференциальное уравнение всегда будет однородным, но коэффициенты и порядок уравнения изменяются при изменении входного сигнала.

Уравнения (7.41) и (7.45) эквивалентны: их решения одинаковы и описывают один и тот же переходный процесс. Однако при построении процесса с помощью аналоговой вычислительной машины [7.10] удобнее пользоваться однородным уравнением (7.45). Полезно заметить, что начальными условиями при интегрировании обоих дифференциальных уравнений процесса — неоднородного и однородного — должны быть начальные значения искомой функции $f_2(t)$, однозначно определяемые передаточной функцией системы и описывающей функцией входного сигнала. Это связано с тем, что дифференциальные урав-

нения (7.41) и (7.45) следуют из уравнений движения в обобщенных производных. Но уравнение движения составляется относительно выходного сигнала, тогда как дифференциальные уравнения, рассмотренные в этом параграфе, могут быть составлены лишь относительно описывающей функции этого сигнала¹.

Примеры. 1. Найдем переходную характеристику (4.93) из (7.41). В рассматриваемом случае, согласно (4.91), это уравнение принимает вид

$$\left(T_2 \frac{d}{dt} + 1\right) f_h(t) = K \left(T_1 \frac{d}{dt} + 1\right) c,$$

где $c = 1$. Отсюда

$$T_2 f_h'(t) + f_h(t) = K \quad (7.46)$$

и описывающая функция переходной характеристики

$$f_h(t) = K + C e^{-t/T_2},$$

где C — постоянная, подлежащая определению по начальному значению искомой функции. По (7.14)

$$f_h(0) = K \frac{T_1}{T_2}.$$

Следовательно,

$$C = K \left(\frac{T_1}{T_2} - 1 \right). \quad (7.47)$$

Тогда

$$f_h(t) = K \left[1 - \left(1 - \frac{T_1}{T_2} \right) e^{-t/T_2} \right],$$

откуда получаем (7.33).

Как видно из решения этой задачи, постоянная C , входящая в интеграл (7.46), не может быть произвольной. Правильное решение получается только в том случае, когда она имеет выражение (7.47).

2. Проследим появление и уничтожение импульсивных функций в уравнении движения в обобщенных производных на примере системы четвертого порядка, получающей входной сигнал в виде единичного скачка.

Передаточная функция рассматриваемой системы имеет вид (7.38), а уравнение движения в обобщенных производных

$$\begin{aligned} & (D^4 + a_1 D^3 + a_2 D^2 + a_3 D + a_4) h(t) = \\ & = (b_0 D^4 + b_1 D^3 + b_2 D^2 + b_3 D + b_4) 1(t), \end{aligned}$$

¹ В литературе часто пишут обычные дифференциальные уравнения для сигналов. Как уже выяснено в § 4.1, это некорректно. А некорректный метод естественно приводит к ошибочным решениям, уже отмеченным в [7.8]. См. также § 8.3.

где $h(t)$ — переходная характеристика, представляющая собой в данном случае реакцию рассматриваемой системы.

Раскрывая скобки в правой части этого уравнения, вычислим обобщенные производные единичной функции:

$$\begin{aligned}D^1 1(t) &= \delta(t), \\D^2 1(t) &= \delta'(t), \\D^3 1(t) &= \delta''(t), \\D^4 1(t) &= \delta'''(t).\end{aligned}$$

Раскрывая скобки в левой части уравнения движения, вычислим обобщенные производные функции $h(t)$. Первая обобщенная производная

$$Dh(t) = f_h'(t) 1(t) + f_h(0) \delta(t),$$

где, по формуле Тропского, начальное значение описывающей функции переходной характеристики $f_h(0) = b_0$. Следовательно,

$$Dh(t) = f_h'(t) 1(t) + b_0 \delta(t).$$

Вторая обобщенная производная

$$D^2 h(t) = f_h''(t) 1(t) + f_h'(0) \delta(t) + f_h(0) \delta'(t),$$

где $f_h'(0) = b_1 - a_1 b_0$.

Следовательно,

$$D^2 h(t) = f_h''(t) 1(t) + (b_1 - a_1 b_0) \delta(t) + b_0 \delta'(t).$$

Далее так же находим третью и четвертую обобщенные производные:

$$\begin{aligned}D^3 h(t) &= f_h'''(t) 1(t) + f_h''(0) \delta(t) + f_h'(0) \delta'(t) + f_h(0) \delta''(t), \\D^4 h(t) &= f_h^{IV}(t) 1(t) + f_h'''(0) \delta(t) + f_h''(0) \delta'(t) + f_h'(0) \delta''(t) + \\&\quad + f_h(0) \delta'''(t).\end{aligned}$$

Здесь по формуле Тропского

$$f_h''(0) = b_2 - a_2 b_0 - a_1 (b_1 - a_1 b_0),$$

$$\begin{aligned}f_h'''(0) &= b_3 - a_3 b_0 - a_2 (b_1 - a_1 b_0) - a_1 [b_2 - a_2 b_0 - \\&\quad - a_1 (b_1 - a_1 b_0)].\end{aligned}$$

Таким образом,

$$\begin{aligned}D^3 h(t) &= f_h'''(t) 1(t) + [b_2 - a_2 b_0 - a_1 (b_1 - a_1 b_0)] \delta(t) + \\&\quad + (b_1 - a_1 b_0) \delta'(t) + b_0 \delta''(t), \\D^4 h(t) &= f_h^{IV}(t) 1(t) + \{b_3 - a_3 b_0 - a_2 (b_1 - a_1 b_0) - \\&\quad - a_1 [b_2 - a_2 b_0 - a_1 (b_1 - a_1 b_0)]\} \delta(t) + [b_2 - a_2 b_0 - \\&\quad - a_1 (b_1 - a_1 b_0)] \delta'(t) + (b_1 - a_1 b_0) \delta''(t) + b_0 \delta'''(t).\end{aligned}$$

После подстановки полученных выражений обобщенных производных в уравнение движения коэффициенты при импульсах одина-

кового порядка в левой и правой частях оказываются одинаковыми. Так, при $\delta'''(t)$ слева и справа будет коэффициент b_0 . При $\delta''(t)$ слева, как и справа, получается коэффициент

$$b_1 - a_1 b_0 + a_1 b_0 = b_1,$$

При $\delta'(t)$ —

$$b_2 - a_2 b_0 - a_1 (b_1 - a_1 b_0) + a_1 (b_1 - a_1 b_0) + a_2 b_0 = b_2.$$

При $\delta(t)$, наконец, слева получается, как нетрудно проверить, коэффициент b_3 . Таким образом, импульсы, отсутствующие в действительности, но появляющиеся в уравнении движения при аппроксимации реального воздействия единичной функцией, взаимно уничтожаются. Это не зависит от значений коэффициентов b_0, b_1, b_2 и b_3 .

3. Найдем начальные условия для интегрирования уравнения движения (7.41) при $n=2$ и $m=1$. Подставляя в уравнение движения выражения входного и выходного сигналов, имеем:

$$\begin{aligned} T_2 [f_2''(t) 1(t) + f_2'(0) \delta(t) + f_2(0) \delta'(t)] + \\ = T_1 [f_2'(t) 1(t) + f_2(0) \delta(t)] + 1 = \\ = K \{ \tau_1 [f_1'(t) 1(t) + f_1(0) \delta(t)] + 1 \}. \end{aligned}$$

Приравняем члены левой и правой частей уравнения, содержащие импульсы одинакового порядка:

$$T_2 f_2(0) \delta'(t) = 0,$$

$$T_2 f_2'(0) \delta(t) + T_1 f_2(0) \delta(t) = K \tau_1 f_1(0) \delta(t).$$

Отсюда получаем начальные условия

$$f_2(0) = 0,$$

$$f_2'(0) = K \frac{\tau_1}{T_2} f_1(0).$$

Чтобы найти те же условия по формуле Тропского, необходимо задать функцию $f_1(t)$. Пусть $f_1(t) \equiv 1$, т. е. входной сигнал имеет вид единичного скачка. Тогда изображение выходного сигнала

$$X_2(p) = K \frac{\tau_1 p + 1}{(T_2 p^2 + T_1 p + 1)p}.$$

Следовательно,

$$f_2(0) = 0,$$

$$f_2'(0) = \frac{b_0}{a_0},$$

где $b_0 = K \tau_1$, $a_0 = T_2$, откуда и получается предыдущий результат.

§ 7.4. ПОСТРОЕНИЕ ПЕРЕХОДНОГО ПРОЦЕССА ПО ВРЕМЕННЫМ ХАРАКТЕРИСТИКАМ

Если на вход подан некоторый сигнал $x_1(t)$, имеющий изображение $X_1(p)$, то, воспользовавшись (4.14) и заменив передаточную функцию ее выражением через изображение

переходной характеристики (4.72), можно записать изображение выходного сигнала в виде

$$X_2(p) = pH(p) X_1(p).$$

По известной из операционного исчисления теореме о производной свертки

$$pH(p) X_1(p) \doteq \left\{ f_1(0) f_h(t) + \int_0^t f_1'(\tau) f_h(t - \tau) d\tau \right\} 1(t),$$

где $f_h(t)$ и $f_1(t)$ — описывающие функции переходной характеристики и входного сигнала.

Следовательно,

$$f_2(t) = f_1(0) f_h(t) + \int_0^t f_1'(\tau) f_h(t - \tau) d\tau. \quad (7.48)$$

Преобразуем входящий сюда *интеграл Дюамеля*, заменяя дифференциал $d\tau$ конечной разностью $\Delta\tau$ и производную $f_1'(\tau)$ — отношением конечных разностей $\Delta f_1/\Delta\tau$. Тогда

$$\int_0^t f_1'(\tau) f_h(t - \tau) d\tau \approx \sum_t f_h(t - \tau) \Delta f_1(\tau),$$

где суммирование должно быть выполнено в области изменения τ от 0 до t . Для этого рассмотрим $\Delta\tau$ как шаг квантования по времени, принимая во внимание лишь значения обеих функций, стоящих под знаком суммы, соответствующие началу каждого шага. Следовательно, обе эти функции можно рассматривать при суммировании как решетчатые функции целочисленного аргумента n . Таким образом, можно написать, что

$$\Delta f_1(t) = f_1[n] - f_1[n - 1]. \quad (7.49)$$

При этом n будет изменяться в области суммирования от 1 до $t/\Delta\tau$ и

$$f_h(t - \tau) = f_h\left[\frac{t}{\Delta\tau} - n\right]. \quad (7.50)$$

Теперь, учитывая (7.49) и (7.50), находим, что интеграл Дюамеля

$$\begin{aligned} & \int_0^t f_1'(\tau) f_h(t-\tau) d\tau \approx \\ & \approx \sum_{n=1}^{t/\Delta\tau} (f_1[n] - f_1[n-1]) f_h\left[\frac{t}{\Delta\tau} - n\right] \\ & f_2(t) \approx f_1[0] f_h\left[\frac{t}{\Delta\tau}\right] + \\ & + \sum_{n=1}^{t/\Delta\tau} (f_1[n] - f_1[n-1]) f_h\left[\frac{t}{\Delta\tau} - n\right]. \end{aligned} \quad (7.51)$$

Поясним применение этой формулы на простейшем примере вычисления $x_2(t)$ при $t = 2\Delta\tau$:

$$\begin{aligned} x_2(t) = & \{f_1[0]f_h[2] + (f_1[1] - f_1[0])f_h[1] + \\ & + (f_1[2] - f_1[1])f_h[0]\} 1(t). \end{aligned}$$

Это показано на рис. 7.7, где перемножаемые в каждом члене последние выражения разности ординат f_1 и f_h (те и другие отмечены двусторонними стрелками) соединены пунктирными линиями.

Такой способ построения переходного процесса с помощью переходной характеристики полезен в тех случаях, когда вид входного сигнала делает затруднительным построение процесса по передаточной функции или посредством решения дифференциальных уравнений, в частности когда входной сигнал задан графически.

Выражение (7.48) можно также записать в форме

$$f_2(t) = f_h(0) f_1(t) + \int_0^t f_h'(t-\tau) f_1(\tau) d\tau. \quad (7.52)$$

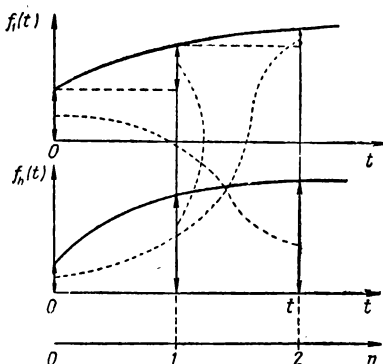


Рис. 7.7. К построению переходного процесса по переходной характеристике

Действительно, интегрируя по частям, имеем:

$$\int_0^t f_1(\tau) f_h(t-\tau) d\tau = f_1(\tau) f_h(t-\tau) \Big|_0^t + \int_0^t f_h(t-\tau) f_1(\tau) d\tau.$$

В силу соотношения (4.73) формула (7.52) принимает вид

$$f_2(t) = f_h(0) f_1(t) + \int_0^t f_w(t-\tau) f_1(\tau) d\tau. \quad (7.53)$$

Таким образом, переходный процесс можно построить и с помощью импульсной характеристики системы.

Пример. Зная переходную характеристику $h(t)$, найдем реакцию системы, имеющей передаточную функцию $W(p)$, на сигнал (7.34). Изображение этого входного сигнала

$$X_1(p) = \frac{1}{p} - \frac{1}{p} e^{-pT_c}.$$

Следовательно, изображение выходного сигнала

$$X_2(p) = \frac{W(p)}{p} - \frac{W(p)}{p} e^{-pT_c}$$

или, согласно (4.83),

$$X_2(p) = H(p) - H(p) e^{-pT_c}.$$

Отсюда выходной сигнал

$$x_2(t) = h(t) - h(t - T_c).$$

В частности, если передаточная функция имеет вид (4.80), из последнего уравнения легко получить результат (7.37).

§ 7.5. ЧАСТОТНЫЕ МЕТОДЫ ПОСТРОЕНИЯ ПЕРЕХОДНЫХ ПРОЦЕССОВ

Составив изображение выходного сигнала $X_2(p)$ и выяснив по расположению полюсов этого изображения, что выходной сигнал изобразим по Фурье, получаем комплексный спектр выходного сигнала:

$$X_2(i\omega) = W(i\omega) X_1(i\omega),$$

где $W(i\omega)$ — частотная характеристика системы;

$X_1(i\omega)$ — комплексный спектр входного сигнала.

Вводя вещественную и мнимую части $R_2(\omega)$ и $I_2(\omega)$ спектра выходного сигнала, представим его в виде

$$X_2(i\omega) = R_2(\omega) + i I_2(\omega).$$

Тогда по табл. 3.2 переходный процесс можно записать по одной из следующих формул:

$$x_2(t) = \frac{2}{\pi} \int_0^{\infty} R_2(\omega) \cos \omega t d\omega \quad (7.54)$$

или

$$x_2(t) = -\frac{2}{\pi} \int_0^{\infty} I_2(\omega) \sin \omega t d\omega. \quad (7.55)$$

На этих формулах основаны *частотные методы построения переходных процессов*.

Если входной сигнал $x_1(t)$ с течением времени стремится к некоторому постоянному значению a_1 , не равному нулю, то преобразование Фурье нельзя применить к нему непосредственно (см. § 3.3). Чтобы воспользоваться частотным методом построения переходного процесса и в этом случае, введем вспомогательный сигнал

$$\zeta_1(t) = a_1 1(t) - x_1(t). \quad (7.56)$$

Тогда

$$x_1(t) = a_1 1(t) - \zeta_1(t)$$

и выходной сигнал $x_2(t)$ получим как сумму реакций системы на сигналы $a_1 1(t)$ и $\zeta_1(t)$. Реакцию на сигнал $-\zeta_1(t)$, изобразимый по Фурье, можно найти по (7.54) и (7.55). Реакция на сигнал $a_1 1(t)$ равна $a_1 h(t)$. Дело сводится, таким образом, к построению переходной характеристики системы.

Переходная характеристика $h(t)$ с течением времени должна стремиться к некоторому постоянному $a \neq 0$. Поэтому она не изобразима по Фурье. Чтобы получить ее частотным методом, введем функцию

$$\zeta(t) = a 1(t) - h(t), \quad (7.57)$$

изобразимую по Фурье. Ее изображение

$$Z(i\omega) = \frac{a - W(i\omega)}{i\omega}.$$

Подставляя сюда частотную характеристику в форме (4.49), находим:

$$Z(i\omega) = \frac{a - [U(\omega) + iV(\omega)]}{i\omega}, \quad (7.58)$$

откуда следует, что это изображение имеет вещественную часть $-\frac{V(\omega)}{\omega}$ и мнимую часть $\frac{aU(\omega) - a}{\omega}$. Тогда по (7.54)

$$\zeta(t) = -\frac{2}{\pi} \int_0^{\infty} V(\omega) \frac{\cos \omega t}{\omega} d\omega. \quad (7.59)$$

Определяя теперь $h(t)$ из (7.57) и пользуясь выражением $\zeta(t)$ по (7.59), получаем:

$$h(t) = a \, 1(t) + \frac{2}{\pi} \int_0^{\infty} V(\omega) \frac{\cos \omega t}{\omega} d\omega. \quad (7.60)$$

На основании предельной теоремы и по (7.57) найдем величину a :

$$a = \lim_{\omega \rightarrow 0} i\omega \left[\frac{a}{i\omega} - Z(i\omega) \right].$$

Подставляя сюда (7.58), имеем:

$$a = \lim_{\omega \rightarrow 0} [U(\omega) + iV(\omega)],$$

откуда

$$a = U(0), \quad (7.61)$$

так как $V(0) = 0$.

Теперь из (7.60) получаем:

$$h(t) = U(0) \, 1(t) + \frac{2}{\pi} \int_0^{\infty} V(\omega) \frac{\cos \omega t}{\omega} d\omega. \quad (7.62)$$

Другой способ определения $h(t)$ дает формула (7.55), из которой по найденной выше мнимой части изображения (7.58)

$$\begin{aligned} \zeta(t) &= \frac{2}{\pi} \int_0^{\infty} [a - U(\omega)] \frac{\sin t\omega}{\omega} d\omega = \\ &= \frac{2}{\pi} \left[a \, 1(t) \int_0^{\infty} \frac{\sin t\omega}{\omega} d\omega - \int_0^{\infty} U(\omega) \frac{\sin t\omega}{\omega} d\omega \right]. \end{aligned}$$

Интеграл, входящий в первый член полученного выражения, равен $\pi/2$. Следовательно,

$$\zeta(t) = a \, l(t) - \frac{2}{\pi} \int_0^{\infty} U(\omega) \frac{\sin t\omega}{\omega} d\omega. \quad (7.63)$$

А так как, согласно (7.57),

$$h(t) = a \, l(t) - \zeta(t),$$

то по (7.63) получаем:

$$h(t) = \frac{2}{\pi} \int_0^{\infty} U(\omega) \frac{\sin t\omega}{\omega} d\omega. \quad (7.64)$$

Таким образом, переходную характеристику линейной системы можно найти по формуле (7.64), по смыслу сов-

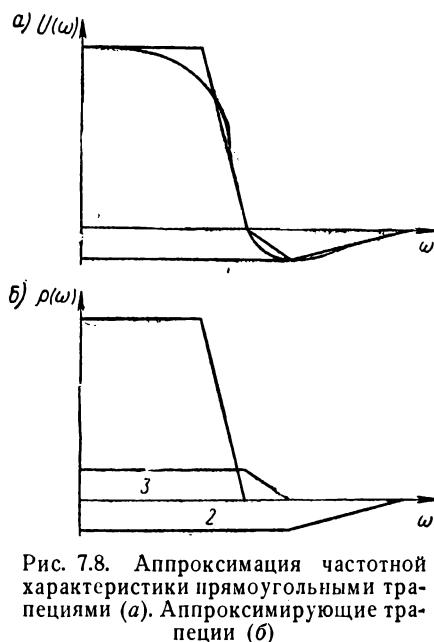


Рис. 7.8. Аппроксимация частотной характеристики прямоугольными трапециями (а). Аппроксимирующие трапеции (б)

падающей с (7.55), или же по формуле (7.62), отличающейся от (7.54) слагаемым $a \, U(0) \, l(t)$.

Для вычисления интегралов, входящих в выражения вида (7.54) и (7.55), существуют особые графические

приемы. Рассмотрим это на примере построения переходной характеристики $h(t)$. Воспользуемся формулой (7.64). По способу В. В. Солодовникова [7.3] для вычисления этой функции по (7.64)

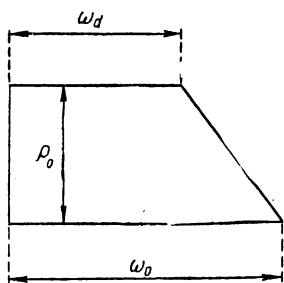


Рис. 7.9. Типовая трапецидальная частотная характеристика

нужно построить график вещественной части $U(\omega)$ частотной характеристики (рис. 7.8). Такой график аппроксимируется прямоугольными трапециями, располагаемыми так, чтобы прямой угол их совмещался с координатным углом, а большее основание — с осью частот. На рис. 7.8, а аппроксимация сделана тремя трапециями, представленными отдельно на рис. 7.8, б. Трапеция 1 имеет положительные ординаты $\rho(\omega)$, трапеция 2 — отрицательные, а трапеция 3, равновеликая излишней части трапеции 2, должна иметь поэтому положительные ординаты. Эти трапеции являются *трапецидальными частотными характеристиками*. С их помощью искомым процесс

$$x_2(t) = \frac{2}{\pi} \sum_i \int_0^{\infty} \rho_i(\omega) \frac{\sin t\omega}{\omega} d\omega, \quad (7.65)$$

где суммирование производится по всем трапециям, т. е. по индексу i .

Типовая трапецидальная характеристика показана на рис. 7.9. Она имеет высоту ρ_0 , меньшее основание ω_d и большее ω_0 . Ее вид характеризуется тремя числами: ρ_0 , ω_0 и $\chi = \frac{\omega_d}{\omega_0}$.

У единичной трапеции $\rho_0 = 1$, $\omega_0 = 1$. Для нее

$$\frac{2}{\pi} \int_0^{\infty} \rho(\omega) \frac{\sin t\omega}{\omega} d\omega = h(t, \chi).$$

Функция $h(t, \chi)$ подробно табулирована [7.4]. Поэтому вычисление $x_2(t)$ легко сделать, заменяя трапеции, полученные при аппроксимации кривой $U(\omega)$, единичными. При этом значение функции $h(t, \chi)$, найденное для каждой трапеции, необходимо умножить на ρ_0 . Кроме того, так

как переход к единичной трапеции означает уменьшение основания в ω_0 раз, то в ω_0 раз увеличивается время. Действительно, умножая в (7.65) время на m и обозначая $m\omega = \omega_m$, имеем:

$$x_2(m, t) = \frac{2}{\pi} \sum_i \int_0^{\infty} \rho_i \left(\frac{\omega_m}{m} \right) \frac{\sin t \omega_m}{\omega_m} d\omega_m.$$

Следовательно, чтобы найти значение выходного сигнала для некоторого значения t , необходимо брать табличное значение функции h для t/ω_0 . Таким образом,

$$x_2(t) = \sum \pm \rho_0 h \left(\frac{t}{\omega_0}, \chi \right), \quad (7.66)$$

где суммирование производится по всем трапециям, знаки ординат которых присваиваются членам суммы.

Этот способ построения процесса полезен в тех случаях, когда изображение процесса задано графиком $R_2(\omega)$. По графику $I_2(\omega)$ способ построения процесса аналогичен рассмотренному.

Построение переходного процесса по таким формулам как (7.54), (7.55), (7.62) и (7.64) сводится к вычислению интегралов вида

$$\int_0^{\infty} F(\omega) f(t, \omega) d\omega, \quad (7.67)$$

где $F(\omega)$ — функция, обычно задаваемая графически. Для вычисления такого интеграла можно воспользоваться *спектральным преобразованием* [7.18], введя новую переменную

$$S(\omega) = \int_0^{\omega} F(\vartheta) d\vartheta. \quad (7.68)$$

Построив график $S(\omega)$ (при графическом задании функции $F(\omega)$ это можно сделать графическим интегрированием, например, по формуле трапеций), аппроксимируем его подходящим образом и найдем частоту в функции новой переменной S :

$$\omega = \varphi(S). \quad (7.69)$$

Тогда интеграл (7.67) будет равен сумме вычисленных

для всех участков аппроксимации интегралов вида

$$I(t) = \int_{S_1}^{S_2} f[t, \varphi(S)] dS, \quad (7.70)$$

где S_1 и S_2 — границы участка, на котором пригодна аппроксимация (7.69).

Если некоторый участок графика $S(\omega)$, расположенный между точками (ω_1, S_1) и (ω_2, S_2) , аппроксимируется отрезком прямой

$$S - S_1 = k(\omega - \omega_1),$$

где

$$k = \frac{S_2 - S_1}{\omega_2 - \omega_1}, \quad (7.71)$$

то на этом участке

$$\varphi(S) = \omega_1 + \frac{S - S_1}{k}. \quad (7.72)$$

При аппроксимации параболой, имеющей вершину в точке (ω_1, S_1) ,

$$S_1 - S = k(\omega_1 - \omega)^2,$$

где

$$k = \frac{S_1 - S_2}{(\omega_1 - \omega_2)^2}, \quad (7.73)$$

и

$$\varphi(S) = \omega_1 + \sqrt{\frac{S_1 - S_2}{k}}. \quad (7.74)$$

При аппроксимации равнобокой гиперболой с асимптотами $\omega = \omega_a$ и $S = S_a$

$$(\omega - \omega_a)(S - S_a) = k. \quad (7.75)$$

Требуя, чтобы гипербола проходила через точки (ω_1, S_1) , (ω_2, S_2) и промежуточную точку (ω_3, S_3) ($\omega_1 < \omega_3 < \omega_2$), найдем:

$$\omega_a = \frac{A - C}{B - D}, \quad (7.76)$$

$$S_a = A - B\omega_a,$$

где

$$\begin{aligned} A &= \frac{\omega_3 S_3 - \omega_2 S_2}{\omega_3 - \omega_2}, \\ B &= \frac{S_3 - S_2}{\omega_3 - \omega_2}, \\ C &= \frac{\omega_1 S_1 - \omega_2 S_2}{\omega_1 - \omega_2}, \\ D &= \frac{S_1 - S_2}{\omega_1 - \omega_2}; \end{aligned} \quad (7.77)$$

k определяется из (7.75) при подстановке вместо ω и S , соответственно ω_1 , ω_2 или ω_3 и S_1 , S_2 или S_3 .

Формула (7.68) показывает, что линейная и параболическая аппроксимация функции $S(\omega)$ означает линейную аппроксимацию функции $F(\omega)$ соответственно параллельными оси частот и наклонными отрезками прямых. Такая аппроксимация функции $F(\omega)$ имеет место и при построении процесса с помощью трапецидальных частотных характеристик. Гиперболическая аппроксимация $S(\omega)$ означает более сложную и точную аппроксимацию функции $F(\omega)$.

Рассмотрим применение спектрального преобразования (7.68) к построению переходной характеристики по формуле (7.64) [7.18]. В этом случае

$$F(\omega) = U(\omega),$$

$$f(t, \omega) = \frac{\sin t\omega}{\omega}.$$

Тогда на участке линейной аппроксимации, вводя (7.72), находим:

$$I(t) = k(\sin t\omega_2 - \sin t\omega_1), \quad (7.78)$$

где k определяется по (7.71).

На участке параболической аппроксимации, вводя (7.74), получаем:

$$I(t) = 2k \left[\frac{1}{t} (\cos t\omega_2 - \cos t\omega_1) + \omega_1 (\sin t\omega_2 - \sin t\omega_1) \right], \quad (7.79)$$

где k определяется по (7.73).

На участке гиперболической аппроксимации, вводя (7.74), имеем:

$$I(t) = \frac{K}{\omega_a} \left\{ \frac{\sin t\omega_2}{\omega_2 - \omega_a} - \frac{\sin t\omega_1}{\omega_1 - \omega_a} + \left(t \sin t\omega_a + \frac{\cos t\omega_a}{\omega_a} \right) \times \right. \\ \times [\operatorname{si} t(\omega_2 - \omega_a) - \operatorname{si} t(\omega_1 - \omega_a)] + \left(\frac{\sin t\omega_a}{\omega_a} - t \cos t\omega_a \right) \times \\ \times [\operatorname{ci} t(\omega_2 - \omega_a) - \operatorname{ci} t(\omega_1 - \omega_a)] + \frac{1}{\omega_a} (\operatorname{si} t\omega_1 - \operatorname{si} t\omega_2) \left. \right\}, \quad (7.80)$$

где k определяется по (7.75).

Переходная характеристика

$$h(t) = \frac{2}{\pi} \sum_{i=1}^n I_i(t), \quad (7.81)$$

где n — число участков аппроксимации функции $U(\omega)$.

Формулы (7.78) — (7.80) пригодны и для вычисления переходного процесса по (7.55), если положить $F(\omega) = \omega I_2(\omega)$.

Таким образом, для построения переходных процессов с помощью спектральных преобразований достаточно иметь таблицы тригонометрических функций и таблицы интегральных синусов и косинусов.

Примеры. 1. Найдём комплексный спектр регулярной составляющей выходного сигнала, получаемого при подаче гармонического воздействия (4.53) на вход системы, имеющей частотную характеристику $W(i\omega)$.

По найденному в § 4.2 спектр выходного сигнала

$$X_s(i\omega) = a_1 \gamma \frac{U(\omega) + iV(\omega)}{\gamma^2 - \omega^2},$$

а спектр его нерегулярной составляющей

$$X_n(i\omega) = a_1 \frac{\gamma U(\gamma) + i\omega V(\gamma)}{\gamma^2 - \omega^2}.$$

Следовательно, спектр регулярной составляющей

$$X_p(i\omega) = X_s(i\omega) - X_n(i\omega) = a_1 \gamma \frac{U(\gamma) - U(\omega) + i \left[\frac{\omega}{\gamma} V(\gamma) - V(\omega) \right]}{\omega^2 - \gamma^2}.$$

2. Пользуясь методом спектральных преобразований, построим переходную характеристику простого инерционного звена первого порядка, имеющего передаточную функцию

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

В этом случае

$$U(\omega) = \frac{1}{1 - \omega^2},$$

$$S(\omega) = \int_0^{\omega} \frac{d\omega}{1 + \omega^2} = \operatorname{arctg} \omega.$$

Аппроксимируем эту функцию двумя линейными участками и одним гиперболическим, выбрав их границы, приведенные в табл. 7.4. В той же таблице приведены значения k , вычисленные для линейных участков по (7.71), а для гиперболического — по (7.75). Для гиперболического участка принято $\omega_3 = 2,6$. При этом по (7.77) $A = 1,57$, $B = 0,0362$, $C = 1,55$, $D = 0,0762$, по (7.76) $\omega_a = -0,448$, $S_a = 1,58$.

Таблица 7.4

**К примеру построения переходного процесса
с помощью спектральных преобразований**

Аппроксимация	ω_1	ω_2	k
Линейная	0	0,6	0,901
Линейная	0,6	1,0	0,613
Гиперболическая	1,0	10	-1,15

Результаты вычисления по (7.81) приведены в табл. 7.5. Там же для сравнения даны значения, полученные по (4.77), т. е.

$$h(t) = 1 - e^{-t}. \quad (7.82)$$

Таблица 7.5

К оценке точности метода спектральных преобразований

t	$h(t)$	
	по (7.81)	по (7.82)
0	0	0
0,1	0,0926	0,0952
0,2	0,181	0,181
0,4	0,330	0,330
0,6	0,450	0,451
0,8	0,551	0,551
1,0	0,633	0,632
1,5	0,778	0,777
2,0	0,861	0,865

Сопоставление данных, приведенных в табл. 7.5, показывает высокую точность метода спектральных преобразований, для выяснения которой и служит рассмотренный пример [7.18].

§ 7.6. КОРНЕВЫЕ ГОДОГРАФЫ

Частотные методы анализа и синтеза АУС имеют существенный недостаток, заключающийся в том, что связь между частотными свойствами системы и ее поведением во временной области выражается лишь в неявной форме. Исследование же систем третьего и более высокого порядка на базе обычных методов решения уравнений движения представляет собой весьма трудоемкий процесс, связанный с перебором нескольких вариантов решения, так как этот метод не позволяет непосредственно учитывать влияние отдельных параметров системы на характер ее работы. В связи с этими обстоятельствами за последнее время получил развитие *метод корневых годографов*, позволяющий осуществлять одно-временное исследование частотных и переходных характеристик. Этот метод весьма прост, удобен в применении и обладает большой наглядностью.

Метод корневых годографов, предложенный впервые К. Ф. Теодорчиком в СССР [7.1] и В. Р. Эвансом в США [7.2], базируется на связи нулей и полюсов передаточной функции разомкнутой системы с нулями и полюсами передаточной функции замкнутой системы (нулями и полюсами передаточной функции принято называть, как было указано в § 4.1, соответственно, корни полиномов, стоящих в ее числителе и знаменателе). Сущность метода заключается в графо-аналитическом определении на комплексной плоскости геометрического места точек, представляющих собой значения корней характеристического уравнения замкнутой системы при непрерывном изменении одного из ее параметров в пределах от 0 до ∞ . Совокупность получающихся при таком изменении *траекторий корней* характеристического уравнения называется *корневым годографом* системы. В качестве переменного параметра обычно выбирают коэффициент усиления разомкнутой системы.

Рассмотрим основные особенности метода корневых годографов на примере одноконтурной системы, пред-

ставленной на рис. 7.10 и имеющей в разомкнутом виде передаточную функцию

$$W_1(p) = K \frac{P_1(p)}{P_2(p)}, \quad (7.83)$$

где K — коэффициент усиления. Подставляя это выражение в (5.6), найдем, что передаточная функция замкнутой системы

$$W(p) = \frac{KP_1(p)}{P_2(p) + KP_1(p)}.$$

Из последних выражений очевидно, что нули передаточной функции замкнутой системы совпадают с нулями¹ разомкнутой системы, полюса же замкнутой системы отличаются от полюсов системы в разомкнутом состоянии (при неединичной отрицательной обратной связи нулями системы в замкнутом состоянии являются нули разомкнутой системы и полюса звена обратной связи).

Для определения полюсов замкнутой системы нужно решить характеристическое уравнение (6.44).

При $K=0$ полюса замкнутой системы определяются как решения уравнения $P_2(p)=0$, т. е. совпадают с полюсами разомкнутой системы. Если выражение (6.44) разделить на K , то при $K \rightarrow \infty$ полюса замкнутой системы будут определяться решениями уравнения $P_1(p)=0$, т. е. нулями разомкнутой системы. Отсюда следует, что траектории корней характеристического уравнения замкнутой системы, получающиеся при изменении K от 0 до ∞ , должны начинаться в полюсах разомкнутой системы и заканчиваться в нулях разомкнутой системы.

Согласно (5.6) характеристическое уравнение замкнутой системы

$$W_1(p) = -1. \quad (7.84)$$

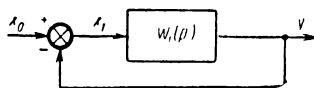


Рис. 7.10. Одноконтурная система, охваченная отрицательной обратной связью

¹ Здесь и далее для краткости будем опускать слова «передаточная функция», заменяя выражение «нули (полюса) передаточной функции разомкнутой (замкнутой) системы» словами «нули (полюса) разомкнутой (замкнутой) системы».

Поскольку $W_1(p)$ есть комплексная величина, она может быть представлена в виде $|W_1(p)|e^{i\varphi} = |W_1(p)|(\cos \varphi + i \sin \varphi)$, тогда (7.84) распадается на два уравнения:

$$|W_1(p)| = 1, \quad (7.85)$$

$$\arg W_1(p) = \pm (2k + 1)\pi \quad (k = 0, 1, 2 \dots). \quad (7.86)$$

В основе метода корневых годографов лежит графическое решение уравнений (7.85) и (7.86) на комплексной плоскости p . Эти уравнения обычно называются соответственно *амплитудным и фазовым критериями*.

Рассмотрим частный случай, в котором передаточная функция разомкнутой системы

$$W_1(p) = K \frac{1 + \tau_1 p}{p(1 + T_1 p)(1 + T_2 p)(1 + T_3 p)}. \quad (7.87)$$

Это выражение удобно для дальнейшего переписать в следующей форме:

$$W_1(p) = K \frac{\left(\frac{1}{\tau_1} + p\right) \tau_1}{T_1 T_2 T_3 p \left(\frac{1}{T_1} + p\right) \left(\frac{1}{T_2} + p\right) \left(\frac{1}{T_3} + p\right)}. \quad (7.88)$$

Любой из сомножителей $\left(p + \frac{1}{\tau_i}\right)$ и $\left(p + \frac{1}{T_i}\right)$ можно представить в виде

$$p + \frac{1}{T_i} = p - p_i \quad \text{и} \quad p + \frac{1}{\tau_i} = p - z_i,$$

где $p_i = -\frac{1}{T_i}$ и $z_i = -\frac{1}{\tau_i}$ — полюса и нули разомкнутой системы. Тогда из (7.88) получим:

$$W_1(p) = KC \frac{(p - z_1)}{p(p - p_1)(p - p_2)(p - p_3)}, \quad (7.89)$$

где

$$C = \frac{\tau_1}{T_1 T_2 T_3}.$$

Множители $(p - z_i)$ и $(p - p_i)$ на комплексной плоскости p представляют собой разность векторов p и z_i или p_i и изображаются векторами, направленными из точки z_i (p_i) в точку p (рис. 7.11). Если исследуемая точка p представляет собой корень характеристического уравнения замкнутой системы, то на основании (7.85), (7.86), (7.89)

и рис. 7.10 можно написать:

$$\frac{KC |p - z_i|}{|p| |p - p_1| |p - p_2| |p - p_3|} = 1, \quad (7.90)$$

$$\varphi_{z_1} - \varphi_0 - \varphi_{p_1} - \varphi_{p_2} - \varphi_{p_3} = \pm (2k + 1) \pi. \quad (7.91)$$

Аналогичные уравнения можно составить при исследовании систем любого порядка. Обобщение амплитудного

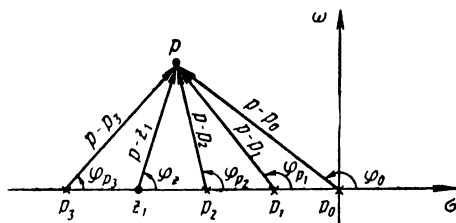


Рис. 7.11. К выводу амплитудного и фазового критериев

и фазового критериев на передаточную функцию, имеющую m нулей и n полюсов, дает:

$$\frac{KC \prod_{f=1}^m \rho_{z_f}}{\rho_{p_0}^i \prod_{l=1}^{n-i} \rho_{p_l}} = 1, \quad (7.92)$$

$$\sum_{f=1}^m \varphi_{z_f} - \left(i\varphi_0 + \sum_{l=1}^{n-i} \varphi_{p_l} \right) = \pm (2k + 1) \pi, \quad (7.93)$$

где ρ_{z_f} и ρ_{p_l} — модули векторов, соответствующих нулям и полюсам разомкнутой системы;

φ_{z_f} и φ_{p_l} — аргументы этих же векторов;

$$C = \frac{\prod_{f=1}^m \tau_f}{\prod_{l=1}^n \tau_l}.$$

Изменяя K в пределах $0 \leq K \leq \infty$, можно построить годограф полюсов замкнутой системы. При этом, руко-

водствуясь несколькими простыми правилами, можно с помощью обычных чертежных инструментов быстро отыскать на плоскости p геометрическое место точек, являющихся корнями замкнутой системы. Рассмотрим основные свойства корневых годографов.

1. Корневой годограф содержит n ветвей в виде отрезков кривых или прямых (n — порядок характеристического уравнения разомкнутой системы), начинающихся при $K=0$ в полюсах разомкнутой системы.

Это свойство — следствие теоремы высшей алгебры, согласно которой корень уравнения является непрерывной функцией его коэффициентов. Поскольку комплексные корни характеристического уравнения являются сопряженными, ветви годографа, не лежащие на вещественной оси, должны быть симметричны относительно этой оси.

2. При $K \rightarrow \infty$ ветви корневого годографа заканчиваются в нулях разомкнутой системы.

Для реальных автоматических систем обычно $m < n$. В этом случае m ветвей будут при $K \rightarrow \infty$ стремиться к нулям разомкнутой системы, а остальные $n - m$ ветвей уходят в бесконечность.

3. Ветви вещественных полюсов расположены на чередующихся участках вещественной оси. Чтобы определить, проходит ли годограф через некоторую точку этой оси, следует воспользоваться правилом: если общее число нулей и полюсов разомкнутой системы, расположенных справа от исследуемой точки вещественной оси, является нечетным, то через эту точку проходит ветвь корневого годографа.

Это правило следует из фазового критерия. Зная, что все разностные векторы, проведенные в исследуемую точку из нулей и полюсов, лежащих слева от него, имеют аргумент 0, можно без труда с помощью рис. 7.12 убедиться в том, что в точке p' фазовый критерий выполняется, а в точке p'' (как и в любой точке отрезка вещественной оси, заключенного между z_1 и p_2) — не выполняется, в связи с чем на отрезке $z_1 - p_2$ корневой годограф проходить не может. При применении этого правила учитываются только вещественные нули и полюса разомкнутой системы, так как при сложении аргументов φ_+ и φ_- (рис. 7.12) любой пары комплексных сопряженных

полюсов сумма всегда будет равна 2π и не повлияет на общий фазовый угол.

4. Уходящие в бесконечность ветви корневого годографа имеют асимптоты, число которых, как и количество таких ветвей, равно $(n-m)$, где n — число полюсов разомкнутой системы, m — число нулей ее.

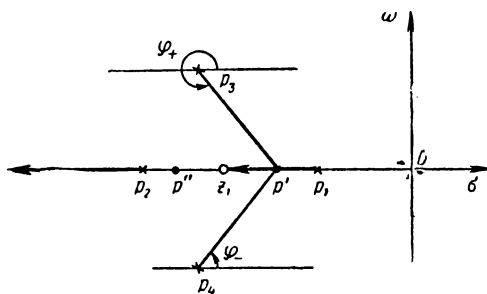


Рис. 7.12. К определению участков вещественной оси, по которым проходит корневой годограф

Положение асимптот на плоскости p можно определить с помощью элементарных формул. Согласно (7.84) и (7.89) характеристическое уравнение замкнутой системы запишем в виде

$$KC \frac{\prod_{i=1}^m (p - z_i)}{\prod_{l=1}^n (p - p_l)} = -1. \quad (7.94)$$

Совершая предельный переход при $p \rightarrow \infty$, получаем [7.6]:

$$\lim_{p \rightarrow \infty} \left[KC \frac{\prod_{i=1}^m (p - z_i)}{\prod_{l=1}^n (p - p_l)} \right] = \frac{KC}{p^{n-m}} = -1, \quad (7.95)$$

или

$$-KC = p^{n-m}. \quad (7.96)$$

Применяя к последнему выражению амплитудный и фазовый критерии, получаем:

$$|-KC| = |p^{n-m}|, \quad (7.97)$$

$$\arg(-KC) = \arg p^{n-m} = (2k+1)\pi. \quad (7.98)$$

Из уравнения (7.98) следует, что

$$(n-m) \arg p = (2k+1)\pi,$$

откуда

$$\theta_a = \arg p = \frac{2k+1}{n-m} \pi \quad (k=0, 1, 2, \dots, n-m-1). \quad (7.99)$$

Формула (7.99) служит для определения угла наклона асимптот к вещественной оси.

Асимптоты корневого годографа должны быть симметричными относительно вещественной оси и составляют $(n-m)$ — лучевую звезду, центр которой лежит на этой оси. Абсцисса центра звезды, т.е. точка пересечения асимптот, определяется выражением [7.6], [7.13]:

$$\sigma_a = \frac{\sum_{l=1}^n p_l - \sum_{i=1}^m z_i}{n-m}, \quad (7.100)$$

где $\sum_{l=1}^n p_l$ и $\sum_{i=1}^m z_i$ — соответственно сумма всех полюсов и сумма всех нулей разомкнутой системы.

5. Если ветви годографа расположены на отрезке вещественной оси, заключенном между двумя полюсами (рис. 7.13), то по свойству 2 на этом участке должна быть точка, в которой происходит «отрыв» годографа от вещественной оси. В этой точке имеют место кратные корни.

Ветви корневого годографа всегда пересекают вещественную ось под прямым углом в силу их симметричности и непрерывности производных. Для определения точки пересечения годографа с вещественной осью [7.11, 7.13] рассмотрим точку p корневого годографа, удаленную от точки пересечения на достаточно малое расстояние δ (рис. 7.13, а). При очень малых значениях δ приращения аргументов $\Delta\theta_{p_l}$ и $\Delta\theta_{z_i}$ можно заменить их тангенсами, определяемыми отношением δ к соответствующим расстояниям от точки отрыва σ_k до нулей и полюсов разомк-

нотой системы. Исследуемая точка p будет принадлежать корневому годографу лишь в том случае, если суммарное приращение углов $\Delta\theta_{p_l}$ и $\Delta\theta_{z_i}$ (при учете их знаков) будет равняться нулю. Для рассматриваемого примера получаем:

$$-\frac{\delta}{\sigma_k} + \frac{\delta}{\sigma_k - p_1} + \frac{\delta}{\sigma_k - p_2} = 0. \quad (7.101)$$

После сокращения на δ находим:

$$\frac{1}{\sigma_k - p_1} + \frac{1}{\sigma_k - p_2} = \frac{1}{\sigma_k},$$

откуда и определяем абсциссу точки отрыва годографа от вещественной оси.

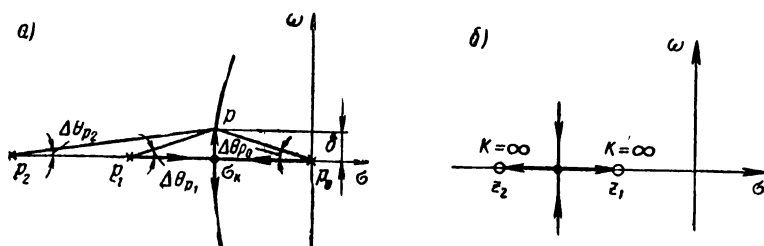


Рис. 7.13. Определение точки пересечения годографа с вещественной осью

В общей форме записи уравнение для определения абсциссы точки отрыва имеет вид [7.11]

$$\sum \frac{1}{\sigma_k - p_l} + \sum \frac{1}{\sigma_k - z_i} = 0, \quad (7.102)$$

где p_l — полюса разомкнутой системы;

z_i — нули разомкнутой системы.

Изложенный метод применяется также и в тех случаях, когда кратные корни расположены на участке вещественной оси, заключенном между двумя нулями (рис. 7.13, б) или нулем и полюсом разомкнутой системы.

6. Углы выхода θ_k годографа из комплексных полюсов и входа их в комплексные нули разомкнутой системы определяются на основании фазового критерия. Это свойство наглядно поясняется на примере рис. 7.14, где рассматривается годограф с комплексным полюсом, однако выводы распространяются и на случай с комплексным нулем.

Если ввести в рассмотрение точку p корневого годографа, весьма близко лежащую к полюсу p_1 , то можно с достаточной степенью точности считать, что аргументы $\varphi_{p_0}, \varphi_{z_1}, \varphi_{p_2}, \varphi_{p_3}$ сохранят свое значение, а аргумент вектора $p - p_1$ будет равняться углу выхода θ_1 . В этом случае фазовый критерий

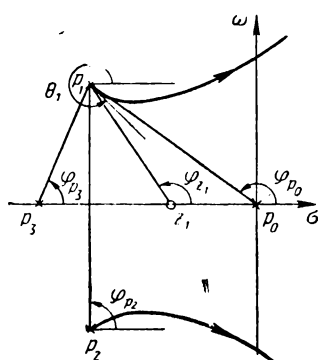


Рис. 7.14. К определению углов выхода годографа из комплексных полюсов

откуда

$$\varphi_{z_1} - (\varphi_{p_0} + \theta_1 + \varphi_{p_2} + \varphi_{p_3}) = \pm (2k + 1)\pi, \quad (7.103)$$

откуда

$$\theta_1 = \mp (2k + 1)\pi + \varphi_{z_1} - (\varphi_{p_0} + \varphi_{p_2} + \varphi_{p_3}), \quad (7.104)$$

что и требовалось получить.

7. Точка пересечения корневых годографов с мнимой осью может быть определена или с помощью одного из критериев устойчивости, или графически, методом подбора.

8. Значение переменной величины K , соответствующее исследуемой точке корневого годографа, т. е. известному корню замкнутой системы:

$$KC = \frac{r_{p_1} r_{p_2} \cdots r_{p_n}}{r_{z_1} r_{z_2} \cdots r_{z_m}}, \quad (7.105)$$

где r_{p_i} — длины векторов $(p - p_i)$;

r_{z_i} — длины векторов $(p - z_i)$.

На основании изложенных свойств корневых годографов разработаны простые правила их графического построения, которые в основном сводятся к следующему:

а) определить нули и полюса разомкнутой системы по передаточной функции $W(p)$, записанной в форме, аналогичной (7.89). Отметить полюса и нули на комплексной плоскости p , применяя одинаковый масштаб для абсцисс и ординат;

б) на основании свойства 3 определить участки вещественной оси, по которым проходит годограф;

в) определить число асимптот $(n - m)$ и их параметры (углы наклона θ_a к вещественной оси и точку пересечения σ_a асимптот) с помощью (7.99) и (7.100);

г) определить, имеются ли на годографе кратные корни, а если имеются, то найти точки пересечения σ_k ветвей годографа с вещественной осью;

д) если имеются комплексные полюсы или нули разомкнутой системы, то определить углы выхода (входа)

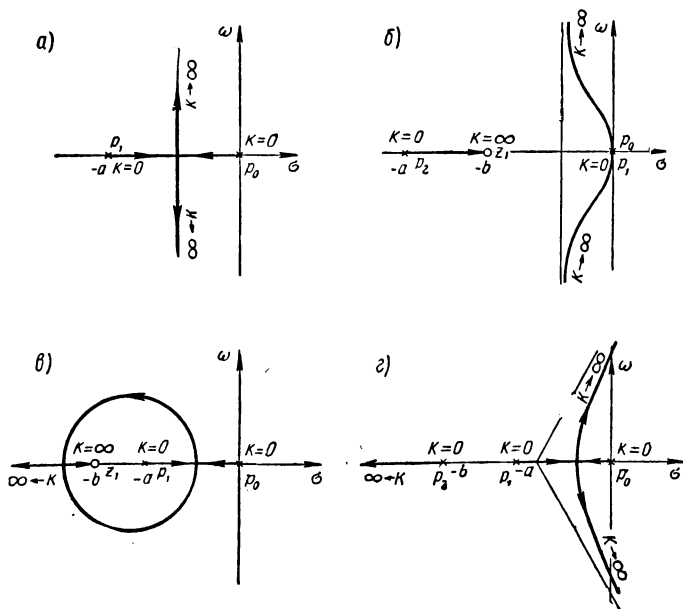


Рис. 7.15. Годографы простых систем:

а) $W_1(p) = \frac{K}{p(p+a)}$; б) $W_1(p) = K \frac{p+b}{p^2(p+a)}$; в) $W_1(p) = K \frac{p+b}{p(p+a)}$;
 г) $W_1(p) = \frac{K}{p(p+a)(p+b)}$

ветвей годографа из полюсов (в нули) на основании свойства б;

е) определить точку пересечения ветвей годографа с мнимой осью;

ж) пользуясь полученными параметрами асимптот и ветвей, построить на комплексной плоскости n ветвей годографа;

з) определить масштаб построенных линий по переменному параметру K . Для этой цели следует воспользо-

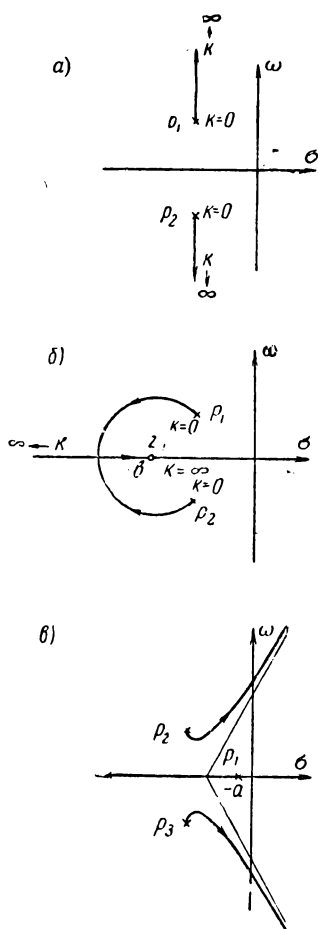


Рис. 7.16. Годографы систем с сопряженными комплексными полюсами:

$$a) W_1(p) = \frac{K}{p^2 + 2\xi\omega_n p + \omega_n^2};$$

$$б) W_1(p) = K \frac{p + b}{p^2 + 2\xi\omega_n p + \omega_n^2};$$

$$в) W_1(p) = \frac{K}{(p^2 + 2\xi\omega_n p + \omega_n^2)(p + a)}$$

ваться как уже известными из предыдущего построения точками (например, начальными точками ветвей σ_k и др.), так и найденными с помощью амплитудного критерия.

Изложенный порядок построения корневых годографов иллюстрируется примерами, приведенными в конце параграфа.

На рис. 7.15—7.19 представлены некоторые наиболее характерные и часто встречающиеся корневые годографы, соответствующие звеньям и системам с различными свойствами, определяемыми видом передаточных функций. Вид кривых на каждом из рисунков в применении к реальным системам может несколько измениться в деталях, так как масштаб по осям абсцисс и ординат не нанесен, однако приведенные графики наглядно иллюстрируют качественный характер корневых годографов для различных систем и соответствие их описанным выше свойствам. Рис. 7.15 и 7.16 дают возможность проследить за изменением корневых годографов при изменении исходного выражения передаточной функции.

Процессы построения корневых годографов и исследования с их помощью свойств АУС могут быть существенно облегчены за счет применения специального устройства — *спиральной линейки* (spirule), предназначенной для быстрого перемножения (и деления) комплексных чисел по-

средством перемножения модулей и сложения аргументов соответствующих векторов.

Спиральная линейка, внешний вид которой представлен на рис. 7.20, состоит из круглого прозрачного транспор-

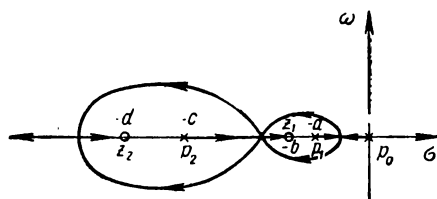


Рис. 7.17. Годограф системы $W_1(p) = \frac{K(p+b)(p+d)}{p(p+a)(p+c)}$

тира, скрепленного в центре заостренной осью с линейкой таким образом, что они могут вращаться около этой оси как независимо друг от друга, так и совместно (на трении). При этом верхний край линейки должен проходить через

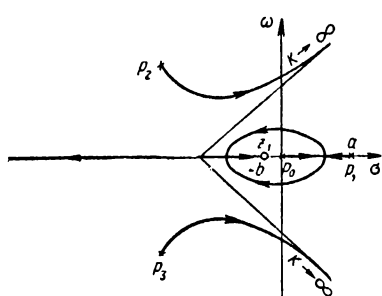


Рис. 7.18. Годограф системы

$$W_1(p) = \frac{K(p+b)}{p(p-a)(p^2 + 2\xi\omega_n p + \omega_n^2)}$$

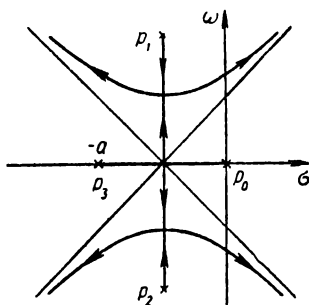


Рис. 7.19. Годограф системы

$$W_1(p) = \frac{K}{p(p+a)(p^2 + 2\xi\omega_n p + \omega_n^2)}$$

центр транспорта. Транспорт отградуирован по окружности через 5° , а на линейке нанесена логарифмическая спираль, служащая для перевода логарифмов длины векторов в угловые величины (угол между верхним краем линейки и линией, проходящей через центр транспорта и любую точку логарифмической спирали, пропорционален

логарифму расстояния от центра до этой точки). Логарифмическая спираль имеет градуировку в относительных единицах, причем за единицу измерения условно принят отрезок между центром транспортира и пересечением спирали с верхним краем линейки. Таким образом, изменению некоторой величины в 10 раз соответствует угол поворота 90° , в связи с чем транспортир имеет отметки умножения на 0,1; 1 и 10.

Спиральная линейка позволяет быстро найти на комплексной плоскости точки, в которых удовлетворяется фазовый критерий, определить значение переменного пара-

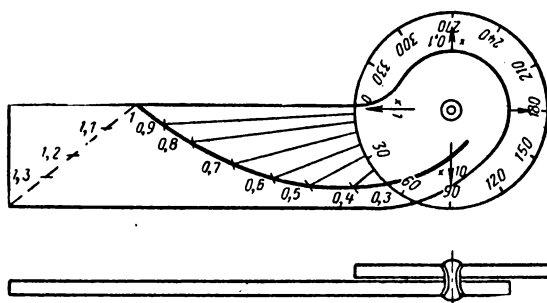


Рис. 7.20. Спиральная линейка

метра K в данной точке годографа, отыскать точки отрыва годографа от вещественной оси.

При отыскании точек корневого годографа, в которых должен выполняться фазовый критерий, с помощью линейки осуществляется суммирование аргументов векторов. С этой целью сначала верхняя кромка линейки совмещается с нулевой отметкой транспортира, ось линейки помещается в исследуемую точку плоскости и линейка располагается на комплексной плоскости горизонтально (рис. 7.21, а). Затем, придерживая пальцами диск, поворачивают линейку так, чтобы ее верхнее ребро прошло через первый полюс (рис. 7.21, б), зафиксировав тем самым аргумент первого вектора. Далее, освободив транспортир, снова поворачивают линейку (вместе с диском) в горизонтальное положение, после чего опять прижимают диск и направляют верхнее ребро линейки в следующий полюс, осуществляя тем самым суммирование двух аргументов.

Описанная процедура повторяется до тех пор, пока не будет получена сумма $\sum_{i=1}^n \varphi_{p_i}$ по всем полюсам.

Чтобы из полученной суммы вычесть аргументы векторов, проходящих через нули разомкнутой системы, следует совместным вращением линейки с транспортиром направить верхнее ребро линейки через соответствующий нуль, а затем, придерживая пальцами транспортир, повернуть линейку в горизонтальное положение.

Если по окончании этой процедуры суммарный аргумент, считываемый с транспортира, будет отличаться от

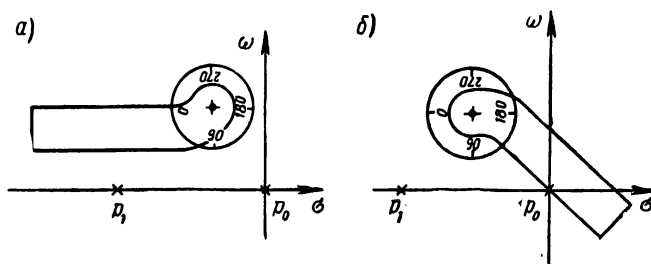


Рис. 7.21. Сложение углов с помощью спиральной линейки

180° , то исследуемая точка не принадлежит корневому годографу. Тогда следует выбрать другую точку и повторить описанный выше процесс. При наличии навыка этот метод «проб и ошибок» требует обычно не более двух-трех попыток.

Для определения местоположения кратных корней на оси абсцисс обычно исследуют, соблюдается ли фазовый критерий в точках комплексной плоскости, расположенных над соответствующим участком вещественной оси и имеющих малые ординаты. Применение спиральной линейки для определения переменного параметра k в данной точке годографа основывается на амплитудном критерии (7.92), из которого следует, что

$$k = KC = \frac{\rho_{p_0}^i \prod_{l=1}^{n-i} \rho_{p_l}}{\prod_{f=1}^m \rho_{z_f}}. \quad (7.106)$$

Сначала верхнее ребро линейки совмещается с индексом $\times 1$ транспортира, а ось последнего устанавливается в исследуемой точке годографа. Затем верхнюю грань линейки направляют в первый полюс (рис. 7.22, а) и, придерживая диск пальцами, поворачивают линейку до тех пор пока полюс не совпадет с какой-либо точкой логарифми-

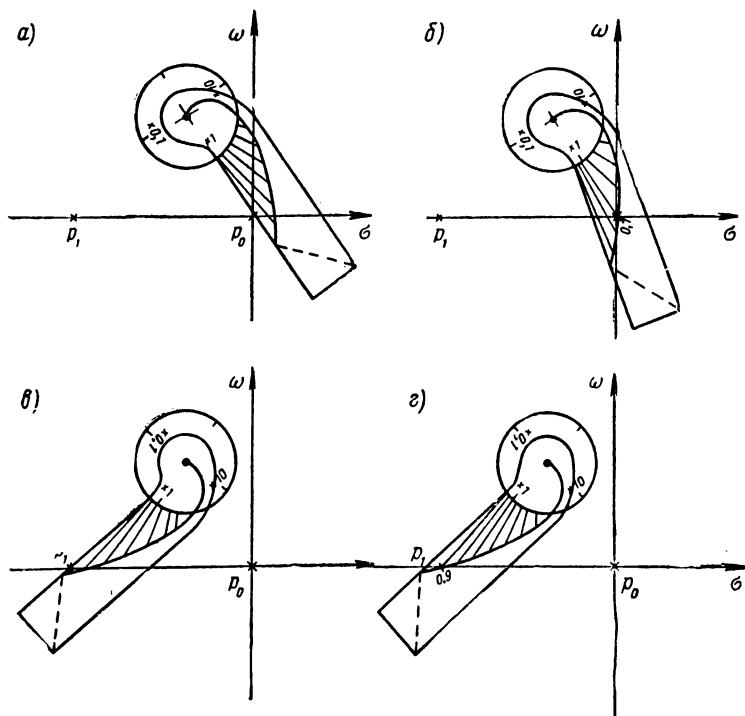


Рис. 7.22. Определение переменного параметра K с помощью спиральной линейки

ческой спирали (рис. 7.22, б). После этого диск освобождают, спиральное лекало совмещают со следующим полюсом и снова вращают линейку при фиксированном положении диска до совпадения этого полюса с логарифмической спиралью (рис. 7.22, в, г). Деление логарифмической спирали, с которым совпадает второй полюс, будет определять результат перемножения модулей двух векторов. Повторяя описанную процедуру по всем полюсам, получим в результате числитель выражения (7.106).

Чтобы полученное произведение разделить на знаменатель выражения (7.106), необходимо, освободив диск, последовательно устанавливать инструмент так, чтобы логарифмическая спираль пересекала каждый из нулей и, прижав транспортир пальцами, направить линейку в исследуемый нуль.

Искомое значение переменного параметра k определяется следующим образом: $k = (\text{отсчет по логарифмической спирали}) \times (\text{коэффициент по транспортиру}) \times (\text{масштабный коэффициент } n - m)$. Здесь масштабный коэффициент зависит от числа делений в общепринятых единицах длины, соответствующего единице длины логарифмической спирали.

Опираясь с векторами, длина которых превышает относительную единицу длины логарифмической спирали, следует пользоваться частью спирали, построенной в виде зеркального изображения действительной кривой (см. рис. 7.20). В этом случае порядок действий со спиральной линейкой при умножении и делении модулей необходимо поменять местами.

Корневые годографы являются эффективным инженерным аппаратом исследования автоматических систем, который с успехом применяется как при их анализе, так и при синтезе. Характер переходных процессов и запасы устойчивости линейной системы можно изменять, влияя на коэффициенты характеристического уравнения, зависящие от корней согласно формулам Виета (см. § 6.3). Реальные же значения корней определяются параметрами системы. Построение корневого годографа выявляет зависимость корней от варьируемого параметра и этим облегчает выбор значения этого параметра, при котором обеспечиваются требуемые характеристики работы системы.

Методу корневых годографов присущи и некоторые недостатки. Один из основных недостатков заключается в следующем: если на одной из ветвей годографа найдена или задана точка, определяющая положение корня характеристического уравнения замкнутой системы (например, ближайшего к мнимой оси корня), и с помощью выражения (7.106) вычислено соответствующее значение переменного параметра, то обычно нельзя сразу указать остальные корни. Для нахождения остальных корней приходится применять метод последовательных приближений к искомым точкам. Однако при наличии даже небольшого опыта

нужные точки корневых годографов можно отыскивать уже после двух-трех проб. В этом случае весьма полезным оказывается спиральное лекало.

Хотя чаще всего в качестве переменного параметра системы выбирают общий коэффициент усиления K , можно построить корневые годографы и по любому другому параметру.

При наличии корневого годографа можно построить простыми средствами частотную характеристику исследуемой системы. Для этого на ветвях годографа отмечают

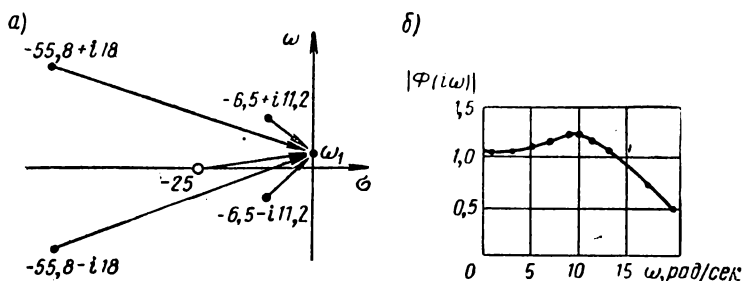


Рис. 7.23. К построению частотной характеристики системы по годографу

корни замкнутой системы, соответствующие выбранному значению переменного параметра. Затем в получаемом при таких корнях выражении передаточной функции замкнутой системы делают подстановку $p = i\omega$, в результате получается частотная характеристика замкнутой системы. На рис. 7.23, а [7.6] показано расположение нулей и полюсов замкнутой системы с передаточной функцией

$$W(p) = \frac{23,68(p + 25)}{(p + 6,5 - i11,2)(p + 6,5 + i11,2)(p + 55,8 - i18) \times (p + 55,8 + i18)}. \quad (7.107)$$

Из (7.107) находим:

$$W(i\omega) = \frac{23,68(i\omega + 25)}{(i\omega + 6,5 - i11,2)(i\omega + 6,5 + i11,2)(i\omega + 55,8 - i18) \times (i\omega + 55,8 + i18)}. \quad (7.108)$$

Каждый из сомножителей выражения (7.108) на комплексной плоскости можно представить вектором, проведенным из полюсов и нулей в точку, лежащую на мнимой

оси. Модули и аргументы этих векторов при данном значении частоты ω легко определяются из графика, после чего на основании равенства (7.108) можно вычислить координаты соответствующих точек амплитудной и фазовой частотных характеристик. Построенная на основании изложенного амплитудная частотная характеристика представлена на рис. 7.23, б.

Примеры. 1. Система автоматического регулирования числа оборотов дизельного двигателя [7.9] с центробежным регулятором имеет в разомкнутом состоянии передаточную функцию

$$W_1(p) = \frac{K(p+2)}{(p+7)(p^2+10p+41)}.$$

Построим корневые годографы замкнутой системы.

Найдем полюса и нули разомкнутой системы $p_1 = -7$, $p_{2,3} = -5 \pm i4$, $z_1 = -2$ и отложим их на комплексной плоскости (рис. 7.24). Корневой годограф будет состоять из трех ветвей, одна из которых начинается в полюсе и кончается в нуле разомкнутой системы, а две других — уходят в бесконечность (свойство 2).

В соответствии со свойством 3 годограф будет проходить по отрезку вещественной оси, заключенному между полюсом p_1 и нулем z_1 ; этот отрезок и является первой ветвью корневого годографа. Остальные ветви, начинающиеся в полюсах p_2 и p_3 , приближаются к асимптотам, имеющим наклон

$$\theta_{a1} = \frac{2k+1}{n-m} \pi = \frac{2 \cdot 0 + 1}{3-1} \pi = \frac{\pi}{2} \quad \text{и} \quad \theta_{a2} = \frac{2 \cdot 1 + 1}{2} \pi = \frac{3}{2} \pi.$$

Асимптоты пересекаются с вещественной осью в точке

$$\sigma_a = \frac{\sum_{i=1}^n P_i - \sum_{f=1}^m z_f}{n-m} = \frac{-7 - 2 \cdot 5 + 2}{2} = -7,5.$$

Таким образом, на графике (см. рис. 7.24) асимптоты примут вид прямой, параллельной оси ординат, пересекающей вещественную ось в точке $\sigma_k = -7,5$.

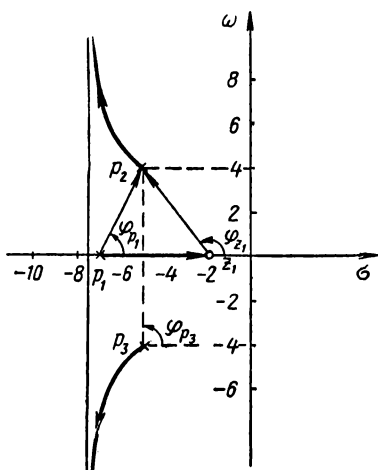


Рис. 7.24. Годограф системы

$$W_1(p) = \frac{K(p+2)}{(p+7)(p^2+10p+41)}$$

Угол выхода годографа из полюса p_3

$$\varphi_{p_3} = -180^\circ - \varphi_{p_1} - \varphi_{p_3} + \varphi_{z_1} = -180^\circ - 64^\circ - 90^\circ + 129^\circ = -205^\circ.$$

Угол φ_{p_3} будет равен 205° , так как обе ветви симметричны относительно вещественной оси.

Полученные величины дают возможность построить оставшуюся часть корневого годографа, общий вид которого показан на рис. 7.24. Масштаб кривых может быть определен в нескольких точках по переменному параметру K с помощью (7.105).

2. Дана передаточная функция разомкнутой системы автоматического управления

$$W_1(p) = \frac{K}{p(p+25)(p^2+100p+2600)}.$$

Построим корневой годограф замкнутой системы [7.6].

Определив полюса разомкнутой системы $p_0 = 0$, $p_1 = -25$, $p_{2,3} = -50 \pm i10$, отмечаем их на комплексной плоскости (рис. 7.25).

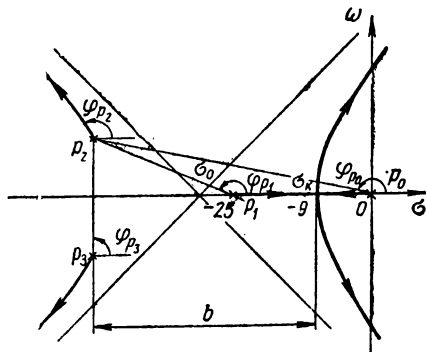


Рис. 7.25. Годограф системы $W_1(p) =$

$$= \frac{K}{p(p+25)(p^2+100p+2600)}$$

Годограф исследуемой системы имеет четыре ветви. Поскольку передаточная функция разомкнутой системы не имеет нулей, все четыре ветви с увеличением K стремятся в бесконечность, приближаясь к асимптотам с углами наклона

$$\theta_a = \frac{2k+1}{n-m} 180^\circ = \frac{2k+1}{4} 180^\circ = \pm 45^\circ; \pm 135^\circ.$$

Эти асимптоты пересекаются в точке

$$\sigma_k = \frac{0 - 25 - 2 \cdot 50}{4} = -31,25.$$

По полученным данным определяем часть вещественной оси, по которой проходят ветви корневого годографа (в данном случае она заключена между полюсами p_0 и p_1) и строим асимптоты.

Угол выхода годографа из полюса p_2

$$\begin{aligned}\varphi_{p_2} &= (1 + 2k)\pi - \varphi_{p_0} - \varphi_{p_1} - \varphi_{p_3} = \\ &= 540^\circ - (168,7^\circ + 158,2^\circ + 90^\circ) = 123,1^\circ.\end{aligned}$$

Угол выхода годографа из полюса p_3 равен $-123,1^\circ$. Это позволяет построить две ветви годографа, начинающиеся в полюсах p_2 и p_3 .

В соответствии со свойством 5, на вещественной оси имеется точка с кратными полюсами замкнутой системы. Эту точку σ_k можно определить из уравнения

$$-\frac{1}{\sigma_k} + \frac{1}{\sigma_k - p_1} + \frac{1}{\sigma_k - p_2} + \frac{1}{\sigma_k - p_3} = 0,$$

которое запишется в виде

$$-\frac{1}{\sigma_k} + \frac{1}{\sigma_k - p_1} + \frac{2b}{a^2 + b^2} = 0,$$

где $a = -10$ — мнимая часть полюсов p_2 и p_3 ;

$b = \sigma_k - (-50)$ — разность абсцисс кратной точки годографа и вещественной части полюсов p_2 и p_3 .

Решение полученного уравнения дает $\sigma_k = -9$.

Оставшиеся две ветви корневого годографа, беря начало в точках p_0 и p_1 , в точке σ_k отрываются под прямым углом от вещественной оси и асимптотически приближаются к прямым, построенным под углом $\pm 45^\circ$ к вещественной оси и пересекающим последнюю в точке σ_a . Общий вид корневого годографа представлен на рис. 7.25.

§ 7.7. СЛУЧАЙНЫЕ ПРОЦЕССЫ

Если входной сигнал линейной системы, действующий в течение времени T_{c1} , можно принять за стационарный, то выходной сигнал будет также стационарным и ограниченным во времени, хотя вследствие инерционности систем продолжительность его существования T_{c2} будет больше T_{c1} . К такому входному сигналу применимо преобразование Фурье хотя бы в несобственном смысле.

Обусловленный инерционностью системы выходной сигнал, существующий в течение времени $T_{c2} - T_{c1}$, не будет случайным. Поэтому случайный выходной сигнал существует лишь в течение времени $T_{c1} - T_c$, как и входной.

Пусть входной сигнал $x_1(t)$ задан его спектральной плотностью $s_1(\omega)$. Тогда спектральная плотность выходного сигнала $x_2(t)$

$$s_2(\omega) = \lim_{T_c \rightarrow \infty} \frac{1}{T_c} |X_2(i\omega)|^2. \quad (7.109)$$

Комплексные спектры рассматриваемых сигналов связаны обычным соотношением

$$X_2(i\omega) = W(i\omega) X_1(i\omega) \quad (7.110)$$

через частотную характеристику $W(i\omega)$. Подставляя (7.110) в (7.109), имеем:

$$s_2(\omega) = |W(i\omega)|^2 \lim_{T_c \rightarrow \infty} \frac{1}{T_c} |X_1(i\omega)|^2.$$

Но

$$\lim_{T_c \rightarrow \infty} \frac{1}{T_c} |X_1(i\omega)|^2 = s_1(\omega).$$

Следовательно, спектральная плотность выходного сигнала

$$s_2(\omega) = |W(i\omega)|^2 s_1(\omega), \quad (7.111)$$

т. е. может быть найдена по спектральной плотности входного сигнала и модулю частотной характеристики системы.

Определив спектральную плотность выходного сигнала, найдем корреляционную функцию:

$$R_{x_2}(\tau) = \frac{1}{\pi} \int_0^\infty s_2(\omega) \cos \tau\omega \, d\omega$$

и средний квадрат сигнала, характеризующий состояние системы:

$$\overline{x_2^2(t)} = \frac{1}{\pi} \int_0^\infty s_2(\omega) \, d\omega. \quad (7.112)$$

В случае нелинейной системы выходной сигнал будет нестационарным, несмотря на стационарность входного, так как нелинейность по-разному влияет на сигнал в зависимости от его значений, т. е. в различные периоды времени.

Пример. На вход линейной системы подан сигнал, обладающий белым спектром $s_1(\omega) = a$ в полосе частот $\omega_1 \leq \omega \leq \omega_2$. Найдем характеристики выходного сигнала, зная передаточную функцию системы

$$W(p) = \frac{K}{Tp + 1}.$$

По (7.111) спектральная плотность выходного сигнала в той же полосе

$$s_2(\omega) = \frac{K^2 a}{T^2 \omega^2 + 1}.$$

Средний квадрат выходного сигнала по (7.112)

$$\overline{x_2^2} = \frac{K^2 a}{\pi} \int_{\omega_1}^{\omega_2} \frac{d\omega}{T^2 \omega^2 + 1} = \frac{K^2 a}{\pi T} (\arctg T\omega_2 - \arctg T\omega_1).$$

§ 7.8. ПРОЦЕССЫ В ДИСКРЕТНЫХ СИСТЕМАХ

Дискретная система линейна, если параметры действующих в ней импульсов линейно зависят от модулирующего эти импульсы непрерывного сигнала. В таких системах выходной сигнал, вызываемый заданным входным сигналом $x_1[n]$, определяется операторным уравнением движения (5.28). Это уравнение выражает изображение $X_2(q, \epsilon)$ выходного сигнала замкнутой импульсной системы, описываемого смещенной решетчатой функцией $x_2[n, \epsilon]$. Однако переход от изображения $X_2(q, \epsilon)$ к оригиналу во многих случаях затруднителен, так как оно имеет бесконечное множество полюсов.

Полагая относительное смещение

$$\epsilon = 0, \quad (7.113)$$

можно исследовать начало и конец переходного процесса, не переходя от изображения выходного сигнала к оригиналу. Служащие для этого предельные соотношения имеют вид [7.17]

$$f_2[0] = \lim_{q \rightarrow \infty} X_2(q), \quad (7.114)$$

$$f_2[\infty] = \lim_{q \rightarrow 0} (e^q - 1) X_2(q). \quad (7.115)$$

Если достаточно найти выходной сигнал при условии (7.113), т. е. в виде несмещенной решетчатой функции, то переход от его изображения к оригиналу упрощается, когда передаточная функция рассматриваемой системы является дробно-рациональной функцией от $z = e^q$. Такие передаточные функции обычно встречаются на практике. Положим поэтому, что при условии (7.113) передаточная

функция замкнутой системы имеет вид

$$W(q) = \frac{b_0 z^\mu + b_1 z^{\mu-1} + \dots + b_{\mu-1} z + b_\mu}{a_0 z^\nu + a_1 z^{\nu-1} + \dots + a_{\nu-1} z + a_\nu},$$

причем $\nu > \mu$.

Деля числитель этого выражения на z^μ , а знаменатель на z^ν , получаем:

$$W_q = \frac{b_0 + b_1 z^{-1} + \dots + b_{\mu-1} z^{-(\mu-1)} + b_\mu z^{-\mu}}{a_0 + a_1 z^{-1} + \dots + a_{\nu-1} z^{-(\nu-1)} + a_\nu z^{-\nu}} e^{-q(\nu-\mu)}. \quad (7.116)$$

Дробь, стоящую здесь в правой части, разложим в бесконечный ряд по степеням z^{-1} :

$$\frac{b_0 + b_1 z^{-1} + \dots + b_\mu z^{-\mu}}{a_0 + a_1 z^{-1} + \dots + a_\nu z^{-\nu}} = A_0 + A_1 z^{-1} + A_2 z^{-2} + \dots$$

Умножая обе части этого выражения на знаменатель левой части и приравнявая коэффициенты при одинаковых степенях z^1 , получаем:

$$b_0 = a_0 A_0,$$

$$b_1 = a_1 A_1 + a_0 A_0,$$

$$\dots \dots \dots$$

$$b_k = a_0 A_k + a_1 A_{k-1} + \dots + a_k A_0.$$

Отсюда следует, что коэффициент

$$A_k = \frac{1}{a_0} \left(b_k - \sum_{r=1}^k a_r A_{k-r} \right), \quad (7.117)$$

где $b_k = 0$ при $k > \mu$ и $a_r = 0$ при $r > \nu$. Теперь передаточную функцию (7.116) можно записать в виде

$$W(q) = e^{-q(\nu-\mu)} \sum_{k=0}^{\infty} A_k e^{-qk}. \quad (7.118)$$

Изображение выходного сигнала при условии (7.113) получим, подставляя (7.118) в (5.28):

$$X_2(q) = \sum_{k=0}^{\infty} A_k e^{-q(\nu-\mu+k)} X_1(q). \quad (7.119)$$

Каждый член этой суммы является, очевидно, изображением оригинала, запаздывающего на $\nu - \mu + k$. Поэтому [7.17] искомый выходной сигнал

$$x_2[n] = \sum_{k=0}^{n-\nu+\mu} A_k x_1[n - (\nu + k - \mu)]. \quad (7.120)$$

Исследование импульсных систем может выполняться также с помощью логарифмических характеристик.

Пример. Найдем реакцию замкнутой импульсной системы, имеющей передаточную функцию (5.30), на входной сигнал $x_1[n] = a1[n]$ при единичной отрицательной обратной связи. В этом случае $\gamma = 1$,

$$\mu = 0, \quad b_0 = \frac{K}{T}, \quad a_0 = 1 + \frac{K}{T}, \quad a_1 = -\exp\left(-\frac{T_n}{T}\right).$$

По (7.117) находим:

$$\begin{aligned} A_0 &= \frac{1}{1 + \frac{K}{T}}, \\ &\dots\dots\dots \\ A_k &= \left(-\frac{a_1}{a_0}\right)^k A_0. \end{aligned}$$

Тогда по (7.120)

$$x_2[n] = aA_0 \sum_{k=0}^{n-1} \left(-\frac{a_1}{a_0}\right)^k 1[n] = aA_0 \frac{1 - \left(-\frac{a_1}{a_0}\right)^{n+1}}{1 + \frac{a_1}{a_0}} 1[n].$$

Отсюда следует, что начальное значение выходного сигнала не равно нулю. Это легко получить и из (7.114), пользуясь выражением (7.119) и находя изображение входного сигнала по табл. 3.4.

НЕПРЕРЫВНЫЕ НЕЛИНЕЙНЫЕ АВТОМАТИЧЕСКИЕ СИСТЕМЫ

§ 8.1. НЕЛИНЕЙНОСТИ

Как показывает пример электрического контура, рассмотренный в § 4.1, система становится нелинейной, если среди зависимостей между характеризующими ее физическими величинами появляется хотя бы одна *нелинейность*, которой нельзя пренебречь.

Нелинейности определяются их *статическими характеристиками*, представляющими собой зависимость между

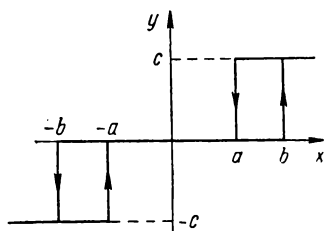


Рис. 8.1. Релейная характеристика с зоной нечувствительности и гистерезисом

сигналами на выходе и входе нелинейного элемента в установившемся режиме. Обозначая выходной сигнал через y , а входной через x , рассмотрим некоторые часто встречающиеся нелинейности. К ним относится *релейная характеристика*, в общем случае показанная на рис. 8.1. Особенности такой характеристики: *ограничение* выходной величины $|y| \leq c$; наличие *зоны нечувствительности*,

в пределах ширины которой $2a$ или $2b$ выходная величина не изменяется при любых значениях входной; *гистерезис*, обусловленный тем, что $a < b$. В частных случаях зоны нечувствительности или гистерезиса может не быть. Встречается также ограничение вследствие *насыщения* (рис. 8.2.) или ограничение линейной зависимости (рис. 8.3), например, при *движении до упора*. Характеристики, представленные на рис. 8.1—8.3, могут не иметь симметрии относительно осей координат: в этом случае их называют *несимметричными*. Встречаются также обычные петли гистерезиса.

Нелинейностью является и всякий зазор или люфт в механизме между ведущей и ведомой частями. На рис. 8.4 x — перемещение ведущей, y — ведомой части. Пока выбирается зазор, ведомая часть неподвижна. Затем она движется по линейному закону 1. При перемене направления движения она снова неподвижна, пока зазор выби-

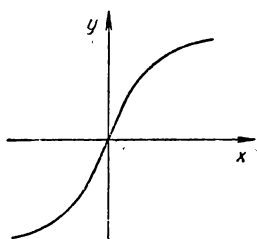


Рис. 8.2. Насыщение

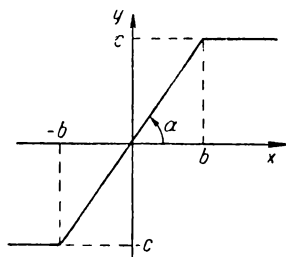


Рис. 8.3. Движение до упора

рается в обратном направлении, после чего движется по линейному закону 2. Это — упрощенная кинематическая картина, без учета динамических явлений, вследствие которых движение ведомой части будет происходить и при выборе зазора. Динамические явления не устраняют, но усложняют нелинейность.

Некоторые устройства имеют характеристику такого вида и при отсутствии люфта. В механических системах ее дает *сухое трение*. Другая характеристика сухого трения приведена на рис. 8.5. Здесь x — скорость движения, y — сила трения. При $x=0$, т. е. при перемене знака скорости, движение прекращается если сумма действующих сил при этом оказывается меньше c . На рис. 8.6 представлена нелинейная характеристика, обусловленная *жидким (или вязким) трением*.

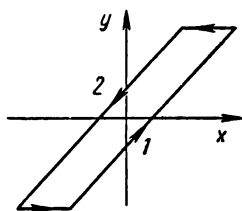


Рис. 8.4. Зазор

Последняя характеристика может быть линеаризована, если линейная часть ее занимает достаточно широкий участок. В отличие от этого характеристика сухого трения представляет собой *существенную нелинейность*, линеаризация которой привела бы к недоучету важных динамических явлений.

Наличие тех или иных нелинейностей практически имеет место во всех автоматических системах, при строгом рассмотрении всегда являющихся нелинейными. Линеаризация, а иногда также слабое влияние нелинейностей часто дают возможность рассматривать автоматические системы как линейные. К некоторым системам это не применимо и приходится их исследовать как нелинейные. К таким *существенно нелинейным системам* относятся, в частности, те, в которые некоторые нелинейности вводятся для требуемого изменения их свойств.

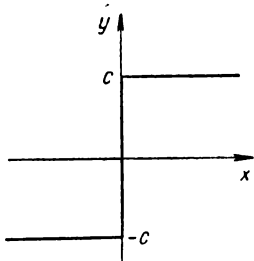


Рис. 8.5. Сухое трение

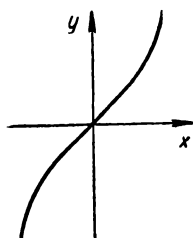


Рис. 8.6. Вязкое трение

Нелинейная зависимость $y(x)$ во многих случаях может быть линеаризована. При нелинейности такого характера, как показано на рис. 8.2, для линеаризации служит *метод малых возмущений* или малых колебаний. Сущность его в том, что нелинейная функция $y(x)$ — непрерывная и имеющая непрерывные производные в окрестностях точки x_0 — может быть разложена в ряд Тейлора:

$$y(x) = y(x_0) + \frac{y'(x_0)}{1!} \Delta x + \frac{y''(x_0)}{2!} (\Delta x)^2 + \dots$$

При достаточно малых Δx (таких, что их квадраты и более высокие степени пренебрежимо малы по сравнению с первой степенью) можно ограничиться двумя первыми членами ряда, полагая $y(x) \approx y(x_0) + a\Delta x$, где $a = y'(x_0)$. Условие применения этого метода — существование производной $y'(x)$, имеющей при $x = x_0$ единственное и конечное значение. На рис. 4.2 $x_0 = 0$.

Метод малых возмущений неприменим, когда $y(x)$ или ее производные имеют точки разрыва — например в случае, представленном на рис. 8.3. В таких случаях применяется *метод осреднения*. Данная зависимость $y(x)$ заме-

няется прямой, проводимой так, чтобы на рабочем участке отклонение $y(x)$ от прямой было наименьшим. Это применимо в ограниченных пределах изменения x , что, впрочем, относится и к методу малых возмущений: так, при достаточно большом значении i_m линеаризация, показанная на рис. 4.2, не может быть удовлетворительной.

Более точные результаты дает *кусочно-линейная аппроксимация* нелинейной характеристики. Когда характеристика ломаная, она получается непосредственно. При этом характеристика может быть описана с помощью единичных функций. Например, в случае движения до упора (см. рис. 8.3)

$$y = c[1(x+b) + 1(x-b) - 1] + kx[1(x+b) - 1(x-b)], \quad (8.1)$$

где

$$k = \operatorname{tg} \alpha. \quad (8.2)$$

Это означает, что

$$y = \begin{cases} -c & x < -b \\ kx & -b \leq x \leq b \\ c & x > b \end{cases}$$

Для криволинейных характеристик кусочно-линейная аппроксимация должна быть выбрана так, чтобы достиглось наилучшее приближение, например, по способу наименьших квадратов:

Метод *гармонической линеаризации* служит для замены нелинейной характеристики *описывающей функцией нелинейности*, определяющей реакцию нелинейного элемента на гармонический входной сигнал. Положим, что на вход симметричного нелинейного элемента подан сигнал (4.53). Тогда описывающая функция

$$h = \frac{a_2}{a_1}, \quad (8.3)$$

где a_2 — амплитуда основной гармоники (частоты γ) сигнала, возникающего на выходе того же элемента.

Представляя выходной сигнал разложенным в ряд Фурье, имеем:

$$a_2 = \frac{4}{\pi} \int_0^{\pi/2} f_n(\vartheta) \sin \vartheta \, d\vartheta, \quad (8.4)$$

где $\vartheta = \gamma t$, а $f_n(\vartheta)$ — описывающая функция сигнала, возникающего на выходе нелинейного элемента.

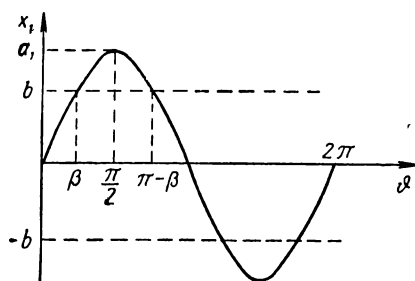


Рис. 8.7. К гармонической линеаризации движения до упора

После линеаризации этот сигнал

$$x_2(t) \approx a_2 \sin \gamma t. \quad (8.5)$$

Пример. Произведем гармоническую линеаризацию в случае движения до упора (рис. 8.3). Введя угол $\beta = \arcsin \frac{c}{a_1}$, находим (рис. 8.7):

$$f_n(\vartheta) = \begin{cases} k a_1 \sin \vartheta & 0 \leq \vartheta \leq \beta, \\ k a_1 \sin \beta & \beta \leq \vartheta \leq \frac{\pi}{2}, \end{cases}$$

где k определяется по (8.2). Подставляя выражения $f_n(\vartheta)$ в (8.4) получаем:

$$a_2 = \frac{4}{\pi} k a_1 \left(\int_0^{\beta} \sin^2 \vartheta d\vartheta + \sin \beta \int_{\beta}^{\pi/2} \sin \vartheta d\vartheta \right) = \frac{2}{\pi} k a_1 (\beta + \sin \beta \cos \beta).$$

Описывающую функцию рассматриваемой нелинейности находим по (8.3):

$$h = \frac{2}{\pi} k (\beta + \sin \beta \cos \beta).$$

§ 8.2. ВЛИЯНИЕ НЕЛИНЕЙНОСТИ НА СЛУЧАЙНЫЕ СИГНАЛЫ

Рассмотрим влияние нелинейного элемента на вероятностные характеристики проходящего через него стационарного случайного сигнала, непрерывного в вероятностном смысле, и на переносимую этим сигналом информацию.

Предположим, что статическая характеристика нелинейного элемента имеет вид, представленный на рис. 8.3, т. е. является линейной при изменении входного сигнала x от $-b$ до b . На линейном участке характеристики выходной сигнал

$$y = kx, \quad (8.6)$$

где $k = \operatorname{tg} \alpha$.

Линейный участок ограничен таким образом, что абсолютное значение выходного сигнала $|y| \leq c$ при любых значениях входного сигнала.

Найдем связь функций распределения вероятностей выходного и входного сигналов в области линейного преобразования, т. е. без учета ограничения статической характеристики рассматриваемого нелинейного элемента. Обозначим функции дифференциального и интегрального распределений для входного и выходного сигналов через $w_x(x)$, $F_x(x)$, $w_y(y)$ и $F(y)$.

Тогда, согласно (3.4),

$$w_x(x) dx = w_y(y) dy. \quad (8.7)$$

Но, по (8.6)

$$w_x(x) dx = \frac{1}{k} w_x\left(\frac{y}{k}\right) dy. \quad (8.8)$$

Из (8.7) и (8.8) следует, что

$$w_y(y) = \frac{1}{k} w_x\left(\frac{y}{k}\right). \quad (8.9)$$

Согласно сказанному в § 3.4,

$$F(y) = \int_{-\infty}^y w(v) dv.$$

Подставляя сюда (8.9), имеем:

$$F(y) = \frac{1}{k} \int_{-\infty}^y w_x\left(\frac{v}{k}\right) dv. \quad (8.10)$$

Но по (8.6)

$$\frac{1}{k} \int_{-\infty}^y w_x\left(\frac{v}{k}\right) dv = \int_{-\infty}^x w_x(u) du. \quad (8.11)$$

Сопоставляя 8.10) и (8.11), находим:

$$F(y) = F_x\left(\frac{y}{k}\right) \quad (8.12)$$

Соотношение (8.12) позволяет строить функцию $F(y)$ по заданной функции $F_x(x)$ графически, как показано на рис. 8.8.

При наличии ограничений вероятность появления выходного сигнала, определяемого неравенством $-\infty < y < -c$, равна нулю; если же $y \geq c$, то вероятность

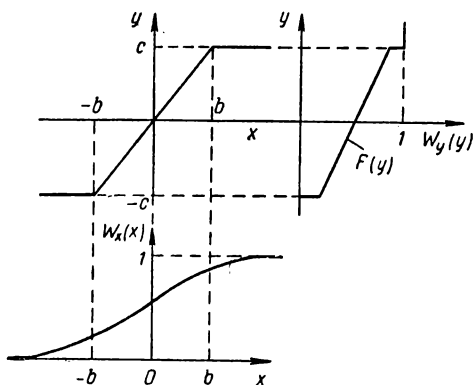


Рис. 8.8. Построение интегральной функции распределения выходного сигнала нелинейного элемента

существования выходного сигнала равна единице. Поэтому (рис. 8.9) интегральную функцию распределения вероятностей выходного сигнала можно записать с помощью единичных функций (см. § 8.1):

$$F_y(y) = F(y) [1(y+c) - 1(y-c)] + 1(y-c). \quad (8.13)$$

Эта функция непрерывна в области $-c < y < c$ и имеет разрывы первого рода при $y = \pm c$.

Дифференциальную функцию $w_y(y)$ распределения вероятностей выходного сигнала найдем как обобщенную производную функции $F_y(y)$:

$$w_y(y) = DF_y(y). \quad (8.14)$$

Подставляя (8.13) в (8.14), получаем:

$$\begin{aligned} \omega_y(y) = & F'(y) [1(y+c) - 1(y-c)] + \\ & + F(-c) \delta(y+c) + [1 - F(c)] \delta(y-c). \end{aligned} \quad (8.15)$$

Подобным образом функции распределения вероятностей выходного сигнала могут быть найдены для любого вида нелинейностей. Зная же функции распределения, нетрудно получить и другие вероятностные характеристики выходного сигнала (см. § 3.4).

Неоднозначность преобразования входного сигнала, имеющая место в рассматриваемом нелинейном элементе при $x < -b$ и $x > b$, приводит к потере информации. Это очевидно из физических соображений. Зная же дифференциальные функции распределения входного и выходного сигналов, эту потерю нетрудно вычислить.

В зоне линейного преобразования сигнала рассматриваемой потери информации нет. Поэтому искомая потеря

$$\Delta I = I_x - I_y, \quad (8.16)$$

где I_x и I_y — информация входного и выходного сигналов в зонах неоднозначного преобразования. Поскольку потери информации вследствие помех здесь не рассматриваются, по (3.56) для фиксированного n

$$I_x = - \int_{-\infty}^{\infty} \omega_x(x) \log_2 \omega_x(x) dx = - \int_b^{\infty} \omega_x(x) \log_2 \omega_x(x) dx. \quad (8.17)$$

Из (8.15)

$$I_y = - F(-c) \log_2 F(-c) - [1 - F(c)] \log_2 [1 - F(c)]. \quad (8.18)$$

Потеря информации, находящая подстановкой (8.17) и (8.18) в (8.16), имеет место лишь вследствие нелинейности преобразования сигнала и является дополнительной к потерям, обусловленным разного рода помехами.

При вычислении I_x и I_y бесконечные пределы интегрирования можно заменять конечными, учитывая вид функ-

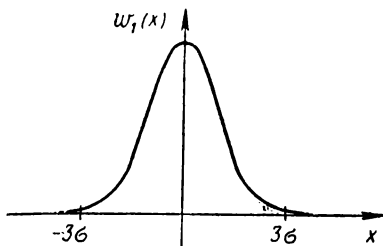


Рис. 8.9. Нормальное распределение

ции $w_x(x)$. Пусть, например, входной сигнал имеет *нормальное распределение*

$$w(x) = \frac{1}{\sigma_x \sqrt{2\pi}} \exp \left[-\frac{(x - \bar{x})^2}{2\sigma_x^2} \right], \quad (8.19)$$

где σ_x^2 — дисперсия, а \bar{x} — математическое ожидание входного сигнала (см. рис. 8.9). В этом случае можно считать, что вероятность появления сигнала $|x| > 3\sigma_x$ практически равна нулю. Поэтому при интегрировании по x в (8.17) и (8.18) вместо $\pm \infty$ можно брать пределы $\pm 3\sigma_x$, считая, что потеря информации (8.16) будет иметь место в зонах

$$y = \begin{cases} -c & -3\sigma < x < -b \\ c & b < x < 3\sigma. \end{cases}$$

Пример. На вход нелинейного элемента, имеющего статическую характеристику, показанную на рис. 8.3, подан непрерывный стационарный случайный сигнал, имеющий нормальное распределение при математическом ожидании, равном нулю. Найдём потерю информации, обусловленную нелинейностью преобразования этого сигнала.

По (8.17) и (8.19) при $m_x = 0$

$$\begin{aligned} I_x &= -\sqrt{\frac{2}{\pi}} \cdot \frac{1}{\sigma_x} \int_b^{3\sigma_x} \exp \left(-\frac{x^2}{2\sigma_x^2} \right) \log_2 \frac{\exp \left(-\frac{x^2}{2\sigma_x^2} \right)}{\sigma_x \sqrt{2\pi}} dx = \\ &= -0,016 + 0,56 \frac{b}{\sigma_x} \exp \left[-\frac{1}{2} \left(\frac{b}{\sigma_x} \right)^2 \right] + \log_2 \sigma_x + 2 \left[1 - \Phi \left(\frac{b}{\sigma_x} \right) \right], \end{aligned}$$

где интеграл вероятностей

$$\Phi(a) = \sqrt{\frac{2}{\pi}} \int_0^a e^{-\frac{t^2}{2}} dt. \quad (8.20)$$

По (8.18)

$$\begin{aligned} I_y &= -\sqrt{\frac{2}{\pi}} \frac{1}{\sigma_x} \int_b^{3\sigma_x} \exp \left[-\frac{x^2}{2\sigma_x^2} \right] dx \cdot \log_2 \left\{ \frac{1}{\sigma_x \sqrt{2\pi}} \int_b^{3\sigma_x} \exp \left[-\frac{x^2}{2\sigma_x^2} \right] dx \right\} dx = \\ &= \left[\Phi \left(\frac{b}{\sigma_x} \right) - 1 \right] \left\{ \log_2 \left[1 - \Phi \left(\frac{b}{\sigma_x} \right) \right] - 1 \right\}. \end{aligned}$$

Подстановка полученных выражений в (8.16) дает:

$$\Delta I = \left\{ \log_2 \sigma_x + 1 + \log_2 \left[1 - \Phi \left(\frac{b}{\sigma_x} \right) \right] \right\} \left[1 - \Phi \left(\frac{b}{\sigma_x} \right) \right] - 0,016 + 0,56 \frac{b}{\sigma_x} \exp \left[-\frac{1}{2} \left(\frac{b}{\sigma_x} \right)^2 \right].$$

§ 8.3. НЕЛИНЕЙНЫЕ СИСТЕМЫ. УРАВНЕНИЯ ДВИЖЕНИЯ В НОРМАЛЬНОЙ ФОРМЕ

Нелинейной называется система, для которой не выполняется принцип суперпозиции. Поэтому нелинейные системы нельзя исследовать теми методами, которые были рассмотрены в предыдущих главах. Процессы в них гораздо разнообразнее, чем в линейных системах, а анализ их сложнее.

При анализе линейных систем использовались следующие характеристики: а) переходная $h(t)$; б) импульсная $w(t)$; в) передаточная функция $W(p)$; г) частотная $W(i\omega)$, причем последнюю часто задавали в графической форме (амплитудно-фазовые характеристики или логарифмические частотные характеристики).

Каждая из этих четырех характеристик однозначно определяет динамические свойства линейной системы, т. е. знание одной из них позволяет вычислить реакцию линейной системы на любой входной сигнал. Этого нельзя сделать для нелинейной системы. Поясним этот важный факт примерами, для чего параллельно рассмотрим простейшие линейную и нелинейную системы.

В качестве линейной системы возьмем безынерционное звено, описываемое уравнением движения

$$x_2(t) = Kx_1(t),$$

где $K = \text{const}$, а в качестве нелинейной — так же безынерционное звено, описываемое уравнением движения

$$x_2(t) = \text{arctg } x_1(t). \quad (8.21)$$

Переходной характеристикой нелинейной системы, так же как и линейной, будем называть ее реакцию на единичный скачок. Если для линейного звена знание переходной характеристики $h(t) = kl(t)$ позволяет вычислить его реакцию на любой входной сигнал (свертка или переход к передаточной функции), то знание переходной характе-

ристики для рассматриваемого нелинейного звена $h_n(t) = \frac{\pi}{4} 1(t)$ не позволяет, например, вычислить даже его реакцию на сигнал $x_1(t) = 3 \cdot 1(t)$. Соответствующие построения выполнены на рис. 8.10. Аналогично обстоит дело и со второй временной характеристикой линейной системы — импульсной характеристикой. Изображение по

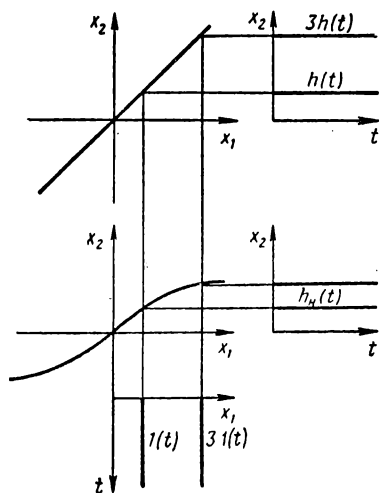


Рис. 8.10. Характеристики линейного и нелинейного безынерционных звеньев

Лапласу импульсной характеристики — передаточная функция — была определена в гл. IV для линейной системы как отношение изображения выходного сигнала к изображению соответствующего входного. Если формально распространить это определение на рассматриваемое нелинейное звено, то передаточная функция будет различной при разных входных сигналах. Например, если на вход подается сигнал $x_1(t) = 1(t)$, то $W_n(p) = \frac{\pi}{4}$, а при $x_1(t) = 3 \cdot 1(t)$ передаточная функция $W_n(p) = \text{arctg } 3$. Если на вход этого звена подать сигнал $x_1(t) =$

$= t 1(t)$ с изображением $\frac{1}{p^2}$, то на выходе будет сигнал $x_2(t) = \text{arctg } t \cdot 1(t)$ с изображением $\frac{ci p \sin p + si p \cos p}{p}$, т. е. формальная передаточная функция будет равна $— (ci p \sin p + si p \cos p) p$.

Аналогично не будет отражать динамических свойств нелинейной системы формально определенная частотная характеристика, в линейном случае равная отношению комплексного спектра выходного сигнала к комплексному спектру соответствующего входного.

В § 4.1 была рассмотрена нелинейная система в виде электрического контура. Уравнение движения в обобщен-

ных производных, составленное для этого контура, оказалось нелинейным. Тогда была произведена линеаризация, позволившая получить линейное уравнение движения в обобщенных производных, описывавшее процессы, протекающие в линейной модели рассматриваемой нелинейной системы. Такой же путь исследования непрерывных нелинейных систем обычно применяется и в более сложных случаях. При этом пользуются различными приемами линеаризации, а уравнения движения составляют в виде обычных дифференциальных уравнений.

Последнее обстоятельство означает, что при составлении уравнений движения не фиксируются моменты появления сигналов в системе, но предполагается, что эти сигналы действуют неопределенно давно. Согласно выясненному в § 3.1 это, конечно, не соответствует реальному положению, которое может быть в автоматической системе. Но следует иметь в виду, что такое представление сигналов допускается лишь в начальной стадии исследования системы. Затем после линеаризации и преобразования дифференциального уравнения движения к операторной форме можно найти изображение выходного сигнала по Лапласу и получить этот сигнал переходом к оригиналу. Поскольку при этом применяется одностороннее преобразование Лапласа, находимый так выходной сигнал отличается от нуля лишь в области $t \geq 0$, как и должно быть, если система начинает действовать при $t = 0$. Иначе говоря, если система без запаздывания, то найденный выходной сигнал содержит множитель $1(t)$. Следовательно, его нельзя подставить для проверки, например, в исходное дифференциальное уравнение движения. Таким образом, от неадекватных в физической картине предположений, сделанных в начале исследования, приходим к адекватному ей результату.

Заметим, что одним из преимуществ уравнений движения в обобщенных производных является возможность проверки результата исследования подстановкой получаемого выходного сигнала в уравнение движения. Кроме того, для преобразования уравнения движения в обобщенных производных к операторной форме не требуются начальные значения сигналов и их производных. Эти обстоятельства не существенны, однако, для исследования нелинейной системы с помощью дифференциальных уравнений движения в общем виде, к которому теперь и обращаемся.

Будем рассматривать нелинейные автоматические системы, которые с достаточной точностью описываются системами обыкновенных дифференциальных уравнений, записанных в *нормальной форме*:

$$\left. \begin{aligned} \dot{x}_1 &= f_1(t, x_1, x_2, \dots, x_n); \\ \dot{x}_2 &= f_2(t, x_1, x_2, \dots, x_n); \\ &\vdots \\ \dot{x}_n &= f_n(t, x_1, x_2, \dots, x_n), \end{aligned} \right\} \quad (8.22)$$

здесь t — независимая переменная (время);

x_1, x_2, \dots, x_n — неизвестные функции времени;

f_1, f_2, \dots, f_n — некоторые функции $n+1$ переменных, определяемые внутренним устройством автоматической системы и законами, которым подчиняются происходящие в ней процессы, а также идеализацией, применяемой при рассмотрении этих процессов.

Обычно одна из неизвестных функций $x_i(t)$ ($i = 1, 2, \dots, n$) является выходным сигналом рассматриваемой

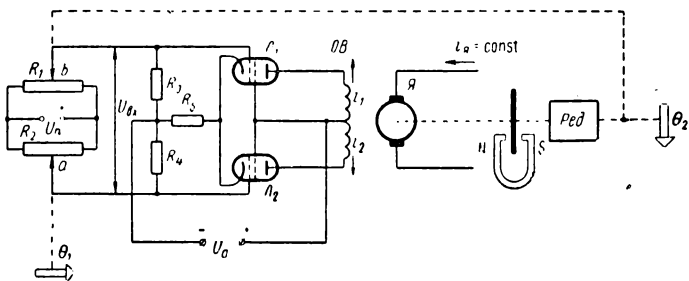


Рис. 8.11. Упрощенная принципиальная схема следящей системы

системы. Число уравнений n называется *порядком нормальной системы уравнений* и *порядком автоматической системы*.

Составим нормальную систему дифференциальных уравнений для следящей системы, представленной на рис. 8.11 упрощенной принципиальной схемой. Входным сигналом следящей системы является угол поворота

командного вала θ_1 . Этот сигнал $\theta_1(t)$ будем считать непрерывным. Командный вал связан с ротором задающего однооборотного потенциометра R_2 . Выходным сигналом привода является угол поворота выходного вала θ_2 , который связан с ротором принимающего потенциометра R_1 , идентичного R_2 . Сопротивления утечки сеток R_3 и R_4 электронного усилителя достаточно велики и, так как этот каскад работает без токов сетки за счет автоматического смещения R_3 , он не нагружает током мост, образованный задающим и принимающим потенциометрами. В анодные цепи ламп \mathcal{L}_1 и \mathcal{L}_2 включена обмотка возбуждения $ОВ$ двигателя постоянного тока, в цепи якоря $Я$ которого поддерживается неизменный во времени ток. Вал двигателя связан с выходным валом через редуктор $Ред$. На валу двигателя помещен демпфер: диск, вращающийся в постоянном магнитном поле, обеспечивающий тормозящий момент, пропорциональный скорости вращения диска, за счет вихревых токов, $M_m = K_v \frac{d\theta_{дв}}{dt}$.

При составлении уравнений движений системы из всех присущих ей нелинейностей будем считать существенной только одну нелинейную зависимость магнитного потока Φ от тока обмотки возбуждения двигателя i_b . Принимая во внимание высокое внутреннее сопротивление экранированных ламп, не будем учитывать инерционность усилителя.

Разобьем рассматриваемую следящую систему на следующие элементы:

а) задающий однооборотный потенциометр R_2 (вход — угол поворота командного вала θ_1 рад., выход — потенциал φ_1 точки a ;

б) принимающий потенциометр R_1 , идентичный R_2 (вход — угол поворота выходного вала θ_2 рад., выход — потенциал φ_2 точки b относительно той же точки, что и φ_1 ;

в) электронный усилитель с обмоткой возбуждения (вход — разность потенциалов $\varphi_1 - \varphi_2$, выход — ток возбуждения i_b ;

г) нелинейный элемент (вход — ток возбуждения i_b , выход — магнитный поток Φ , *вб*);

д) двигатель с приведенной нагрузкой (вход — поток возбуждения Φ , выход — угол поворота вала двигателя $\theta_{дв}$, рад.);

е) редуктор (вход — угол поворота двигателя $\theta_{\text{дв}}$, рад, выход — угол поворота выходного вала θ_2 , рад).

Уравнения движения этих элементов при соответствующей идеализации будут иметь соответственно вид:

$$\text{а) } \varphi_1 = K_1 \theta_1 \text{ при } -\pi < \theta_1 < \pi;$$

$$\text{б) } \varphi_2 = K_1 \theta_2 \text{ при } -\pi < \theta_2 < \pi,$$

$$\text{где } K_1 = \frac{U_{\text{н}}}{2\pi};$$

$$\text{в) } i_{\text{в}} = S(\varphi_1 - \varphi_2) \text{ при } |\varphi_1 - \varphi_2| < U_{\text{вх}}^*,$$

где S — крутизна каскада,
($-U_{\text{вх}}^*$, $U_{\text{вх}}^*$) — интервал линейности усилителя;

$$\text{г) } \Phi = f(i_{\text{в}}) \text{ при } |i_{\text{в}}| < i_{\text{в}}^*;$$

$$\text{д) } J_{\Sigma} \frac{d^2 \theta_{\text{дв}}}{dt^2} + K_v \frac{d \theta_{\text{дв}}}{dt} = C_M i_{\text{я}} \Phi \text{ при } |\Phi| < \Phi^*,$$

где J_{Σ} — приведенный момент инерции на валу двигателя;
 C_M — конструктивная постоянная двигателя;
 $i_{\text{я}}$ — постоянный ток якоря;

$$\text{е) } \frac{d \theta_2}{dt} = \frac{1}{i_p} \frac{d \theta_{\text{дв}}}{dt} \text{ при } \left| \frac{d \theta_{\text{дв}}}{dt} \right| < \omega_{\text{дв}}^*,$$

где i_p — передаточное число редуктора.

Исключая промежуточные координаты, получим:

$$J_{\Sigma} i_p \ddot{\theta}_2 + K_v i_p \dot{\theta}_2 = C_M i_{\text{я}} f[SK_1(\theta_1 - \theta_2)].$$

Обозначим теперь скорость выходного вала через ω и запишем уравнение движения привода в виде нормальной системы:

$$\left. \begin{aligned} \dot{\omega} &= -\frac{K_v}{J_{\Sigma}} \omega + \frac{C_M i_{\text{я}}}{J_{\Sigma} i_p} f[SK_1(\theta_1 - \theta_2)] \\ \dot{\theta}_2 &= \omega, \end{aligned} \right\}$$

при $|\omega| < \omega^*$, $|\theta_1| < \pi$ и $|\theta_2| < \pi$.

Последнюю систему можно переписать в виде

$$\left. \begin{aligned} \dot{\omega} &= -\frac{1}{T_M} \omega + \frac{K_A}{T_M} \cdot \frac{1}{i_p} f[SK_1(\theta_1 - \theta_2)] \\ \dot{\theta}_2 &= \omega, \end{aligned} \right\} \quad (8.23)$$

при $|\omega| < \omega^*$, $|\theta_2| < \pi$, $|\theta_1| < \pi$,

где

$$T_M = \frac{J_{\Sigma}}{K_V}, \quad K_{\Sigma} = \frac{C_M i_{\Sigma}}{K_V}.$$

Правая часть первого из уравнений (8.23)

$$f_1(t, \omega, \theta_2) = -\frac{1}{T_M} \omega + \frac{K_{\Sigma}}{T_M} \cdot \frac{1}{i_p} f[SK_1(\theta_1 - \theta_2)]$$

является нелинейной функцией трех переменных, заданной в параллелепипеде $|\omega| < \omega^*$, $|\theta_2| < \pi$, $|t| < +\infty$ (рис. 8.12). Заметим, что входной сигнал — известная функция времени $\theta_1(t)$, которую считаем непрерывной.

Правая часть второго из уравнений (8.23, $f_2(t, \theta_2, \omega) = \omega$ — линейная функция, которую будем считать заданной в том же параллелепипеде. Рассматриваемой нелинейной следящей системе можно поставить в соответствие и структурную схему, если динамические свойства линейных звеньев отражать, как обычно, передаточной функцией, а нелинейное звено характеризовать его уравнением $\Phi = f(i_{\Sigma})$ (рис. 8.13). Хотя эта структурная схема не позволяет получить, как в случае линейной системы, передаточную функцию, однако она полезна при составлении уравнений (8.23).

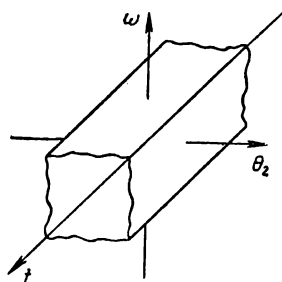


Рис. 8.12. Область задания правых частей уравнений движения следящей системы

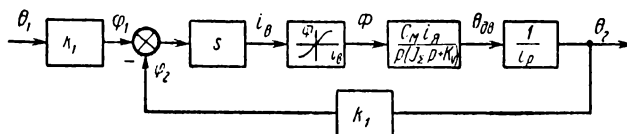


Рис. 8.13. Структурная схема нелинейной следящей системы

Выше было указано, что вид нормальной системы (8.22) дифференциальных уравнений зависит от идеализации, применяемой при рассмотрении процессов в авто-

матической системе. Действительно, при составлении (8.23) не только не учитывался ряд нелинейностей, но и учтенная нелинейность идеализирована. Считалось, что зависимость потока Φ от тока возбуждения может быть охарактеризована однозначной функцией f (рис. 8.14), хотя эта зависимость в действительности более сложна (рис. 8.15).

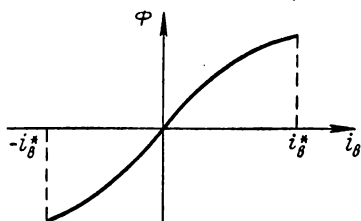


Рис. 8.14. График функции $\Phi = f(i_{\text{в}})$

Будем считать нелинейную функцию $f(i_{\text{в}})$ гладкой в начале координат, тогда по формуле Тейлора

$$\Phi(i_{\text{в}}) = f'(0) i_{\text{в}} + o(i_{\text{в}}), \quad (8.24)$$

или

$$\Phi(i_{\text{в}}) = K_{\Phi} i_{\text{в}} + o(i_{\text{в}}),$$

где $K_{\Phi} = f'(0)$. (рис. 8.16)

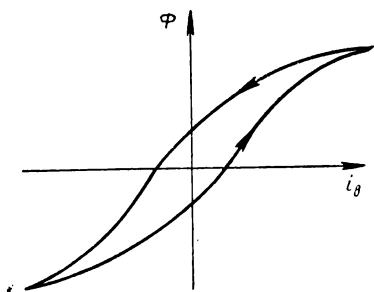


Рис. 8.15. Петля гистерезиса

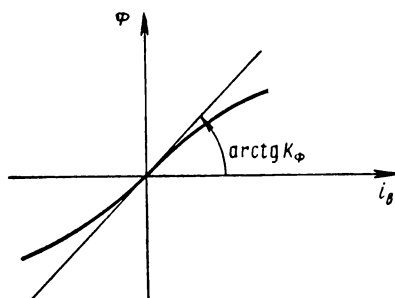


Рис. 8.16. Линеаризация нелинейной зависимости $\Phi = f(i_{\text{в}})$

Уравнения (8.23) с учетом (8.24) запишем в виде

$$\left\{ \begin{aligned} \dot{\omega} &= -\frac{1}{T_{\text{М}}} \omega + \frac{K_{\text{д}}}{T_{\text{М}}} \cdot \frac{1}{i_{\text{р}}} K_{\Phi} S K_1 (\theta_1 - \theta_2) + \\ &\quad + o[SK_1(\theta_1 - \theta_2)], \\ \dot{\theta}_2 &= \omega. \end{aligned} \right. \quad (8.25)$$

Вводя обозначение $K_1 S K_{\Phi} K_{\text{д}} \frac{1}{i_{\text{р}}} = K$ и пренебрегая бесконечно малой $o[SK_1(\theta_1 - \theta_2)]$ более высокого порядка, чем

$SK_1(\theta_1 - \theta_2)$, получим уравнения движения в нормальной форме для линейной модели рассматриваемой нелинейной системы:

$$\left. \begin{aligned} \dot{\omega} &= -\frac{1}{T_M} \omega - \frac{K}{T_M} \theta_2 + \frac{K}{T_M} \theta_1, \\ \dot{\theta}_2 &= \omega. \end{aligned} \right\} \quad (8.26)$$

Характеристическое уравнение линейной модели имеет вид

$$\begin{vmatrix} -\frac{1}{T_M} - p & -\frac{K}{T_M} \\ 1 & 0 - p \end{vmatrix} = 0, \quad (8.27)$$

где p — комплексное число (см. гл. III). Раскрывая определитель, получаем:

$$p^2 + \frac{1}{T_M} p + \frac{K}{T_M} = 0.$$

Уравнение (8.26) позволяет получить передаточную функцию линейной модели. Действительно, переходя к изображениям по Лапласу, имеем (при нулевых начальных значениях):

$$\left. \begin{aligned} p\Omega(p) &= -\frac{1}{T_M} \Omega(p) - \frac{K}{T_M} \Theta_2(p) + \frac{K}{T_M} \Theta_1(p), \\ p\Theta_2(p) &= \Omega(p), \end{aligned} \right\}$$

или

$$\left. \begin{aligned} \left(-\frac{1}{T_M} - p\right) \Omega(p) - \frac{K}{T_M} \Theta_2(p) &= -\frac{K}{T_M} \Theta_1(p), \\ \Omega(p) - p\Theta_2(p) &= 0; \end{aligned} \right\} \quad (8.28)$$

здесь

$$\Omega(p) \doteq \omega(t),$$

$$\Theta_2(p) \doteq \theta_2(t),$$

$$\Theta_1(p) \doteq \theta_1(t).$$

Решая систему (8.28) относительно $\Theta_2(p)$, получим:

$$\Theta_2(p) = \frac{\frac{K}{T_M}}{p^2 + \frac{1}{T_M} p + \frac{K}{T_M}} \Theta_1(p),$$

откуда передаточная функция

$$W(p) = \frac{K}{T_M p^2 + p + K}.$$

Наконец, передаточную функцию линейной модели можно получить, заменяя на структурной схеме, изображенной на рис. 8.13, нелинейное звено с уравнением

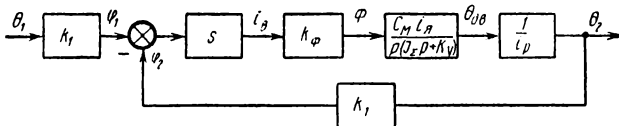


Рис. 8.17. Структурная схема линейной модели следящей системы

движения $\Phi = f(i_\theta)$ линейным с уравнением движения $\Phi = K_\Phi i_\theta$, где $K_\Phi = \left|_{i_\theta=0} \frac{df(i_\theta)}{di_\theta} \right|$ (рис. 8.17). В рассмотренном случае нелинейность следящей системы обусловлена безынерционным нелинейным звеном в ее составе.

Составим теперь уравнения движения той же следя-

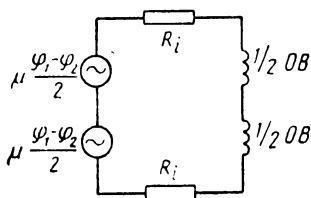


Рис. 8.18. Эквивалентная схема электронного усилителя следящей системы

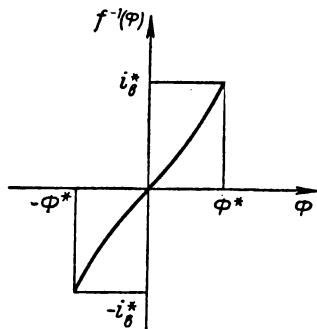


Рис. 8.19. График функции обратной к $f(i_\theta)$

щей системы с учетом инерционности усилителя. Расчленение на элементы остается тем же, но вместо элементов θ и g рассмотрим один нелинейный инерционный элемент (вход — разность потенциалов $\varphi_1 - \varphi_2$, выход — магнитный поток Φ). На рис. 8.18 приведена эквивалентная схема электронного каскада, нагруженного обмоткой

возбуждения OB двигателя, согласно которой

$$\omega \frac{d\Phi}{dt} + (2R_i + R_b) i_b = \mu (\varphi_1 - \varphi_2), \quad (8.29)$$

где ω — число витков обмотки возбуждения;

R_i — внутреннее сопротивление лампы;

μ — ее статический коэффициент усиления;

R_b — активное сопротивление обмотки возбуждения.

Обозначим через f^{-1} функцию, обратную f , выражающую зависимость потока Φ от тока возбуждения i_b (рис. 8.19), и положим $2R_i + R_b = R$. Тогда уравнение (8.29) принимает вид

$$\frac{d\Phi}{dt} = -\frac{R}{\omega} f^{-1}(\Phi) + \frac{\mu}{\omega} (\varphi_1 - \varphi_2) \quad \text{при} \quad |\varphi_1 - \varphi_2| < U_{\text{вх}}^*.$$

Уравнения движения остальных элементов остаются прежними. Исключая промежуточные координаты, получаем уравнения движения в нормальной форме:

$$\left. \begin{aligned} \frac{d\Phi}{dt} &= -\frac{R}{\omega} f^{-1}(\Phi) - \frac{\mu K_1}{\omega} \theta_2 + \frac{\mu K_1}{\omega} \theta_1, \\ \frac{d\omega}{dt} &= \frac{C_M i_p}{J_\Sigma i_p} \Phi - \frac{K_V}{J_\Sigma} \omega \end{aligned} \right\}$$

при

$$|\Phi| < \Phi^*, \quad |\omega| < \omega^*, \quad |\theta_2| < \pi, \quad |\theta_1| < \pi \quad (8.30)$$

или, вводя обозначения (8.23),

$$\left. \begin{aligned} \frac{d\Phi}{dt} &= -\frac{R}{\omega} f^{-1}(\Phi) - \frac{\mu K_1}{\omega} \theta_2 + \frac{\mu K_1}{\omega} \theta_1, \\ \frac{d\omega}{dt} &= \frac{K_A}{T_M} \cdot \frac{1}{i_p} \Phi - \frac{1}{T_M} \omega, \\ \frac{d\theta_2}{dt} &= \omega \end{aligned} \right\} \quad (8.31)$$

при (8.30).

По формуле Тейлора

$$i_b = f^{-1}(\Phi) = \frac{df^{-1}(0)}{d\Phi} \Phi + o(\Phi) = \frac{1}{K_\Phi} \Phi + o(\Phi), \quad (8.32)$$

так как $(f^{-1})' = \frac{1}{f'} = \frac{1}{K_\Phi}$.

Пользуясь последним соотношением, получим уравнения движения линейной модели:

$$\left. \begin{aligned} \frac{d\Phi}{dt} &= -\frac{R}{wK_{\Phi}} \Phi - \frac{\mu K_1}{w} \theta_2 + \frac{\mu K_1}{w} \theta_1, \\ \frac{d\omega}{dt} &= \frac{K_d}{T_M} \cdot \frac{1}{i_p} \Phi - \frac{1}{T_M} \omega, \\ \frac{d\theta_2}{dt} &= \omega. \end{aligned} \right\} \quad (8.33)$$

Характеристическое уравнение этой линейной системы с учетом $\frac{R}{wK_{\Phi}} = \frac{1}{T_B}$

$$\begin{vmatrix} -\frac{1}{T_B} - p & 0 & -\frac{\mu K_1}{w} \\ \frac{K_d}{T_M} \cdot \frac{1}{i_p} & -\frac{1}{T_M} - p & 0 \\ 0 & 1 & 0 - p \end{vmatrix} = 0.$$

Раскрывая определитель, получим:

$$-p \left(-\frac{1}{T_M} - p \right) \left(-\frac{1}{T_B} - p \right) - \frac{\mu K_1 K_d}{T_M w} \cdot \frac{1}{i_p} = 0$$

или

$$p^3 + \left(\frac{1}{T_M} + \frac{1}{T_B} \right) p^2 + \frac{1}{T_M T_B} p + \frac{K}{T_M T_B} = 0, \quad (8.34)$$

где $K = \frac{\mu K_1 K_d T_B}{w i_p}$.

Найдем передаточную функцию линейной модели, для чего в (8.33) перейдем к изображениям по Лапласу:

$$\left. \begin{aligned} \left(-\frac{1}{T_B} - p \right) \bar{\Phi}(p) - \frac{\mu K_1}{w} \Theta_2(p) &= -\frac{\mu K_1}{w} \Theta_1(p), \\ \frac{K_d}{T_M} \cdot \frac{1}{i_p} \bar{\Phi}(p) + \left(-\frac{1}{T_M} - p \right) \Omega(p) &= 0, \\ \Omega(p) - p \Theta_2(p) &= 0. \end{aligned} \right\}$$

Отсюда

$$\begin{aligned}
 W(p) &= \frac{\begin{vmatrix} -\frac{1}{T_B} - p & 0 & -\frac{\mu K_1}{\omega} \\ \frac{K_d}{T_M i_p} & -\frac{1}{T_M} - p & 0 \\ 0 & 1 & 0 \end{vmatrix}}{\begin{vmatrix} -\frac{1}{T_B} - p & 0 & \frac{\mu K_1}{\omega} \\ \frac{K_d}{T_M i_p} & -\frac{1}{T_M} - p & 0 \\ 0 & 1 & -p \end{vmatrix}} = \\
 &= \frac{\frac{\mu K_1 K_d}{T_M i_p \omega}}{p^3 + \left(\frac{1}{T_M} + \frac{1}{T_B}\right)p^2 + \frac{1}{T_M T_B}p + \frac{\mu K_1 K_d}{T_M i_p \omega}} = \\
 &= \frac{K}{T_M T_B p^3 + (T_B + T_M)p^2 + p + K}.
 \end{aligned}$$

Рассмотренные примеры показывают, как по уравнениям движения линейной модели в нормальной форме получить ее передаточную функцию. Рассмотрим обратную задачу. Пусть уравнение движения линейной системы для непрерывного входного сигнала имеет вид

$$T_M T_B \frac{d^3 \theta_2}{dt^3} + (T_B + T_M) \frac{d^2 \theta_2}{dt^2} + \frac{d \theta_2}{dt} + K \theta_2 = K \theta_1.$$

Запишем уравнения движения данной системы в нормальной форме. Для этого введем обозначения $\frac{d \theta_2}{dt} = \omega$ и $\frac{d^2 \theta_2}{dt^2} = \varepsilon$, после чего можно написать нормальную систему трех дифференциальных уравнений первого порядка, равносильную одному исходному уравнению третьего порядка:

$$\left. \begin{aligned}
 \frac{d \varepsilon}{dt} &= -\left(\frac{1}{T_M} + \frac{1}{T_B}\right) \varepsilon - \frac{1}{T_M T_B} \omega - \frac{K}{T_M T_B} \theta_2 + \frac{K}{T_M T_B} \theta_1, \\
 \frac{d \omega}{dt} &= \varepsilon, \\
 \frac{d \theta_2}{dt} &= \omega.
 \end{aligned} \right\} \quad (8.35)$$

Системы дифференциальных уравнений (8.33) и (8.35) описывают динамику одной и той же автоматической системы, но так как в первом случае в качестве обобщенных координат выбраны поток Φ , скорость выходного вала ω и его угол поворота θ_2 , а во втором — ускорение выходного вала ϵ , его скорость ω и угол поворота θ_2 , эти системы дифференциальных уравнений различны, хотя их характеристические уравнения совпадают. Действительно, характеристическое уравнение системы (8.35)

$$\begin{vmatrix} -\frac{1}{T_M} - \frac{1}{T_B} - p & -\frac{1}{T_M T_B} & -\frac{K}{T_M T_B} \\ 1 & 0 - p & 0 \\ 0 & 1 & 0 - p \end{vmatrix} = 0.$$

Раскрывая определитель, имеем:

$$\left(-\frac{1}{T_M} - \frac{1}{T_B} - p\right)p^2 - \frac{K}{T_M T_B} - \frac{1}{T_M T_B} p = 0,$$

что совпадает с (8.34).

Примеры. 1. Будет ли устойчива линейная система, если уравнения движения ее имеют вид

$$\left. \begin{aligned} \frac{dx}{dt} &= -x + z, \\ \frac{dy}{dt} &= -2y - z, \\ \frac{dz}{dt} &= y - z + s, \end{aligned} \right\}$$

где x, y, z — обобщенные координаты системы;

s — входной сигнал — непрерывная функция времени.

Характеристическое уравнение имеет вид

$$\begin{vmatrix} -1 - p & 0 & 1 \\ 0 & -2 - p & -1 \\ 0 & 1 & -1 - p \end{vmatrix} = 0$$

или

$$(-1 - p)(-2 - p)(-1 - p) - (1 + p) = 0.$$

Вынося общий множитель $(1 + p)$, получим:

$$(1 + p)(p^2 + 3p + 3) = 0.$$

$\pm \frac{\sqrt{3}}{2}$ имеют отрицательные вещественные части. Следовательно, данная линейная система устойчива.

2. Найдем передаточную функцию автоматической системы примера 1, считая выходным сигналом обобщенную координату x . Переходя к изображениям по Лапласу, получим систему трех алгебраических уравнений:

$$\left. \begin{aligned} (-1-p) X(p) + Z(p) &= 0, \\ (-2-p) Y(p) - Z(p) &= 0, \\ Y(p) + (-1-p) Z(p) &= -S(p). \end{aligned} \right\} \quad (8.36)$$

Отсюда

$$W_x(p) = \frac{X(p)}{S(p)} = \frac{\begin{vmatrix} 0 & 0 & 1 \\ 0 & -2-p & -1 \\ -1 & 1 & -1-p \end{vmatrix}}{\begin{vmatrix} -1-p & 0 & 1 \\ 0 & -2-p & -1 \\ 0 & 1 & -1-p \end{vmatrix}} = \frac{p+2}{p^3+4p^2+6p+3}.$$

3. Решим задачу, сформулированную в примере 2, но считая выходным сигналом обобщенную координату y .

Решая (8.36) по правилу Крамера, получим:

$$W_v(p) = \frac{Y(p)}{S(p)} = \frac{\begin{vmatrix} -1-p & 0 & 1 \\ 0 & 0 & -1 \\ 0 & -1 & -1-p \end{vmatrix}}{p^3 + 4p^2 + 6p + 3} = \frac{p+1}{p^3 + 4p^2 + 6p + 3}.$$

§ 8.4. ФАЗОВОЕ ПРОСТРАНСТВО АВТОНОМНОЙ СИСТЕМЫ

Автоматическая система называется *автономной*, если в ее уравнения движения явно не входит время t .

$$\left. \begin{aligned} \dot{x}_1 &= f_1(x_1, x_2, \dots, x_n) \\ \dot{x}_2 &= f_2(x_1, x_2, \dots, x_n) \\ &\vdots \\ \dot{x}_n &= f_n(x_1, x_2, \dots, x_n). \end{aligned} \right\} \quad (8.37)$$

Например, если для следящей системы, рассмотренной в § 8.3, входной сигнал $\theta_1(t)$ равен нулю, то привод автономен. Действительно, полагая в (8.31) $\theta_1(t) \equiv 0$,

есть $x = Ce^{-t}$. На рис. 8.20 показаны интегральные кривые, определяемые этим общим решением. Фазовой прямой является ось ординат. Если изображающая точка в какой-то момент времени находится в точке a , то в дальнейшем она движется по оси ординат, асимптотически приближаясь к началу координат. Причем характер этого движения изображающей точки по фазовой прямой не зависит от того, началось ли это движение в момент t_0 или в момент $t_0 + c$. Если движение изображающей точки началось в момент t_0 , то ее фазовая траектория соответствует интегральной кривой 1, если это движение началось в момент $t_0 + c$, то эта же фазовая траектория соответствует интегральной кривой 2. В обоих случаях изображающая точка из точки a в точку b попадает за промежуток времени k . Если изображающая точка в начальный момент будет помещена в точку O , то она будет находиться там и в дальнейшем. Точка O является целой фазовой траекторией, называемой *точкой покоя*. Этой фазовой траектории соответствует интегральная кривая, совпадающая с осью абсцисс. Таким образом, в данном случае фазовая прямая (рис. 8.21) разбивается на три фазовые траектории: два интервала $(-\infty, 0)$, $(0, +\infty)$ и точка O . Заметим, что эти фазовые траектории не пересекаются. Изображающая точка, двигаясь по траектории $(-\infty, 0)$, никогда не попадет в точку O ; это хорошо видно на рис. 8.20.

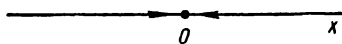


Рис. 8.21. Фазовая прямая уравнения $\dot{x} = -x$

Пусты теперь правая часть (8.39) является функцией, график которой изображен на рис. 8.22. На этом же рисунке проведено разбиение фазовой прямой на траектории. В точках O , x_1 , x_2 функция $f(x)$ обращается в нуль, т. е. в этих точках скорость изображающей точки — нуль. Поэтому точки O , x_1 , x_2 являются точками покоя. На интервале x_1 , x_2 функция $f(x)$ положительна, и, следовательно, скорость изображающей точки направлена здесь в сторону возрастания x . На интервале O , x_1 функция $f(x)$ отрицательна, и, следовательно, скорость изображающей точки направлена в сторону убывания x . Аналогично определяется направление движения изображающей точки на других участках фазовой прямой. На рис. 8.22 это направление указано стрелками. Здесь такие фазовые

траектории не пересекаются. Так, изображающая точка, двигаясь по участку траектории *I*, будет как угодно близко приближаться к участку *II* (точка *O*), но никогда не попадет в точку *O*. Это следует из теоремы существования и единственности решения, условия которой для

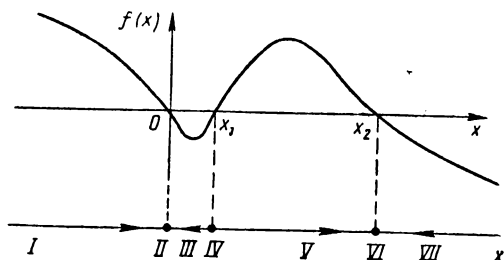


Рис. 8.22. Фазовая прямая дифференциального уравнения $\dot{x} = f(x)$

уравнения (8.39) выполнены, так как $f(x)$ непрерывна и обладает ограниченной производной (рис. 8.22). Заметим, что характер фазовых траекторий в окрестностях точек покоя O и x_1 различен. На основании этого в следующем параграфе будут сделаны важные выводы.

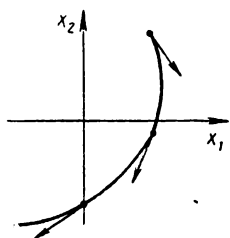


Рис. 8.23. Фазовая траектория

Рассмотрим теперь автономную автоматическую систему второго порядка и дадим ее уравнениям движения

$$\left. \begin{aligned} \dot{x}_1 &= f_1(x_1, x_2), \\ \dot{x}_2 &= f_2(x_1, x_2) \end{aligned} \right\} \quad (8.40)$$

кинематическую интерпретацию. Будем считать, что уравнения (8.40) определяют движение точки, которую также будем называть изображающей, на плоскости (x_1, x_2) (рис. 8.23). При этом первое уравнение в каждой точке с координатами x_1, x_2 этой плоскости, называемой *фазовой плоскостью*, определяет проекцию вектора скорости изображающей точки на ось абсцисс x_1 , а второе уравнение — на ось ординат x_2 .

Таким образом, система (8.40) на фазовой плоскости определяет поле скоростей, которое с течением времени не меняется, так как f_1 и f_2 от времени не зависят в силу

автономности системы. Каждому решению

$$\left. \begin{aligned} x_1 &= \varphi(t), \\ x_2 &= \dot{\varphi}(t) \end{aligned} \right\} \quad (8.41)$$

системы (8.40) ставится в соответствие движение изображающей точки, задаваемое уравнениями (8.41). Траектория, прочерчиваемая изображающей точкой при этом движении, называется фазовой траекторией, так что (8.41) являются параметрическими уравнениями фазовой траектории. Связь между кинематической интерпретацией системы (8.40) и ее решением (8.41) состоит в том, что скорость движения изображающей точки по фазовой траектории в каждый момент времени совпадает со скоростью, заданной системой (8.40) в том месте плоскости, где в этот момент находится изображающая точка. Характер движения изображающей точки по фазовой траектории не зависит от момента времени, в который

это движение началось в силу автономности системы. Заметим, что если выполнены условия теоремы существования и единственности решения для системы (8.40), то фазовые траектории не пересекаются. Фазовая плоскость, заполненная всей совокупностью фазовых траекторий, образующей наглядную картину возможных движений автономной системы (8.40), называется *фазовым портретом* этой системы.

Построим для примера фазовый портрет маятника, изображенного на рис. 8.24. Уравнение движения маятника без учета трения в оси имеет вид

$$J \frac{d^2 \alpha}{dt^2} = M,$$

где $J = ml^2$ — момент инерции маятника;

$M = -mgl \sin \alpha$ — момент силы тяжести относительно оси подвеса;

m — масса шара;

l — длина штанги подвеса;

α — угол, отсчитываемый от вертикали против часовой стрелки.

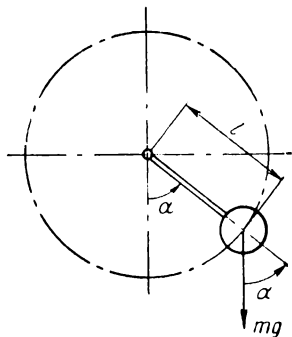


Рис. 8.24. Маятник

Вводя угловую скорость ω в качестве второй обобщенной координаты, получим уравнения движения в нормальной форме:

$$\left. \begin{aligned} \frac{d\alpha}{dt} &= \omega, \\ \frac{d\omega}{dt} &= -\frac{g}{l} \sin \alpha. \end{aligned} \right\} \quad (8.42)$$

Дифференциальные уравнения фазовых траекторий получим делением второго уравнения (8.42) на первое:

$$\frac{d\omega}{d\alpha} = -\frac{g}{l} \frac{\sin \alpha}{\omega}.$$

Разделяя переменные в последнем уравнении и интегрируя, имеем:

$$\omega^2 = \frac{2g}{l} \cos \alpha + C, \quad (8.43)$$

здесь C — постоянная интегрирования. Уравнение (8.43) описывает однопараметрическое семейство фазовых траекторий с параметром C . Чтобы представить себе ход этих кривых, запишем (8.43) в виде

$$\omega = \pm \sqrt{\cos \alpha + C}, \quad (8.44)$$

полагая для простоты $\frac{2g}{l} = 1$. Из (8.44) видно, что построение достаточно провести для верхней полуплоскости фазовой плоскости, учитывая перед корнем знак плюс. Картина в нижней полуплоскости является «зеркальным» отражением верхней полуплоскости в оси абсцисс. На рис. 8.25 проведено построение для верхней полуплоскости, а на рис. 8.26 приведен весь фазовый портрет. Направление движения изображающей точки определяется из (8.42). В верхней полуплоскости $\omega > 0$ и α возрастает, в нижней полуплоскости $\omega < 0$, а α убывает. Фазовые траектории ($2\pi k, 0$), где $k = 0, \pm 1, \pm 2, \dots$ являются точками покоя, причем фактически они отвечают одному и тому же состоянию маятника. Аналогично, точки покоя ($k\pi, 0$) по существу все совпадают. Избежать такой многозначности можно, рассматривая в данном случае движение изображающей точки не по плоскости, а по поверх-

ности прямого кругового цилиндра. Замкнутая траектория 3 соответствует периодическим колебаниям маятника около

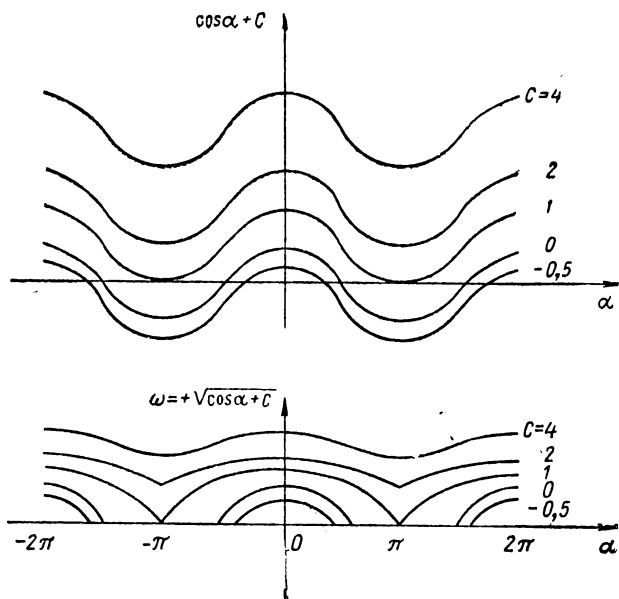


Рис. 8. 25. Построение фазового портрета маятника

положения равновесия $(0, 0)$. Фазовая траектория 1 соответствует вращению маятника против часовой стрелки,

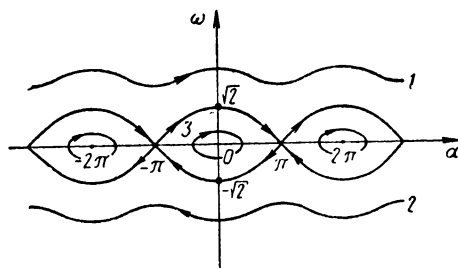


Рис. 8.26. Фазовый портрет маятника

а фазовая траектория 2 — вращению маятника по часовой стрелке. Заметим, что точку покоя $(0, 0)$ по виду фазовых

траекторий в ее окрестности называют точкой покоя типа «центр». Точку покоя $(\pi, 0)$ называют точкой покоя типа «седло».

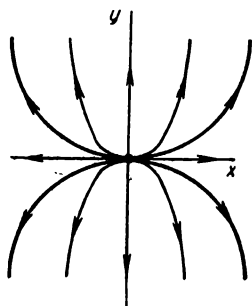


Рис. 8.27. Точка покоя типа «узел»

Фазовые траектории, приближающиеся к точке покоя типа «седло», называются *сепаратрисами*. Фазовые траектории в окрестности точки покоя имеют большое значение в вопросах, рассматриваемых в следующем параграфе. Рис. 8.26 показывает, что фазовый портрет дает возможность наглядно представить все многообразие движений автономной системы второго порядка, зависящее от начальных значений.

В случае автономной системы n -го порядка (8.37), ее кинематическая интерпретация может быть дана в n -мерном пространстве.

Примеры. 1. Известно, что в начальный момент t_0 угловое положение маятника (см. рис. 8.24) $\alpha = 0$. Какова должна быть скорость маятника при t_0 , чтобы он совершал вращательное движение против часовой стрелки (считать, что $l = \frac{1}{2g}$)?

Из рис. 8.26 получаем: $\omega > \sqrt{2}$.

2. Решим предыдущую задачу, если в момент t_0 угол $\alpha = \pi$.
 Ответ: $\omega > 0$.

3. Построим фазовый портрет линейной автономной системы:

$$\left. \begin{aligned} \frac{dx}{dt} &= x, \\ \frac{dy}{dt} &= 2y. \end{aligned} \right\} \quad (8.45)$$

Дифференциальное уравнение фазовых траекторий имеет вид

$$\frac{dx}{dy} = \frac{x}{2y}.$$

Разделяя переменные и интегрируя их, получим что фазовыми траекториями служат: начало координат, полуоси координат и полупараболы (рис. 8.27). Направление движения изображающей точки по фазовым траекториям определяется из (8.45).

§ 8.5. УСТОЙЧИВОСТЬ НЕЛИНЕЙНОЙ СИСТЕМЫ

Рассмотрим уравнения движения (8.40) автономной системы второго порядка. Решения этих уравнений, отвечающие состояниям равновесия автоматической системы, находятся из конечных уравнений

$$\left. \begin{aligned} f_1(x_1, x_2) &= 0, \\ f_2(x_1, x_2) &= 0. \end{aligned} \right\} \quad (8.46)$$

Система конечных уравнений (8.46) может быть решена графически (рис. 8.28). На этом рисунке показан случай, когда (8.46) имеет три решения:

$$\left. \begin{aligned} x_1 &= a, \\ x_2 &= b, \end{aligned} \right\} \quad \left. \begin{aligned} x_1 &= c, \\ x_2 &= d, \end{aligned} \right\} \quad \text{и} \quad \left. \begin{aligned} x_1 &= k, \\ x_2 &= l. \end{aligned} \right\}$$

Эти три решения определяют на фазовой плоскости (8.40) три точки (a, b) , (c, d) и (k, l) , в которых вектор

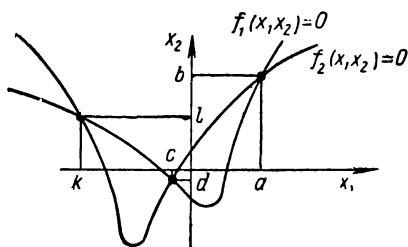


Рис. 8.28. Графическое решение системы уравнений для определения точек покоя

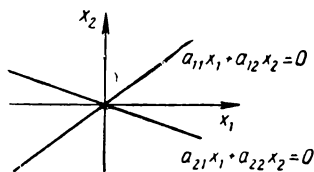


Рис. 8.29. Определение точки покоя линейной системы

скорости изображающей точки равен нулю, т. е. это — три точки покоя системы (8.40). Если система линейная и имеет вид

$$\left. \begin{aligned} \dot{x}_1 &= a_{11}x_1 + a_{12}x_2, \\ \dot{x}_2 &= a_{21}x_1 + a_{22}x_2, \end{aligned} \right\} \quad (8.47)$$

то (в практически интересном случае $\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} \neq 0$) она имеет только одну точку покоя $0, 0$ (рис. 8.29).

Рассмотрим, например, маятник (см. рис. 8.24), имеющий уравнения движения (8.42). Точки покоя опреде-

ляются из системы уравнений

$$\left. \begin{aligned} \frac{g}{l} \sin \alpha &= 0, \\ \omega &= 0. \end{aligned} \right\}$$

Отсюда $\alpha = \pi k$; $\omega = 0$ ($k = 0, \pm 1, \pm 2, \dots$).

В предыдущем параграфе было показано, что это множество решений отвечает, по существу, двум положениям равновесия маятника $(0, 0)$ и $(\pi, 0)$ (рис. 8.30). Из опыта известно, что первое положение равновесия будет наблюдаться в реальной системе, тогда как второе не будет наблюдаться даже приближенно. Этот пример показывает, что среди точек покоя, найденных из (8.46),

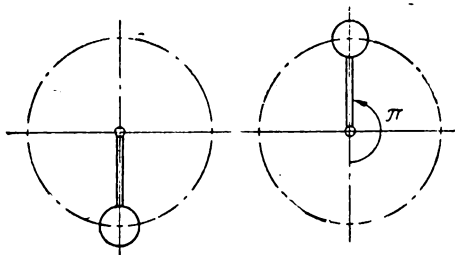


Рис. 8.30. Положения равновесия маятника

могут оказаться «фиктивные», т. е. ненаблюдаемые в реальной системе даже приближенно. Для выяснения «фиктивности» точки покоя не всегда возможно обращение к опыту, как в случае маятника. Желательно получить ответ на этот вопрос из анализа системы уравнений (8.40).

Рассмотрим фазовый портрет маятника (рис. 8.26) и выясним, какие свойства фазовой траектории $\pi, 0$ обуславливают ее физическую нереализуемость. Равновесие $\pi, 0$ отвечает начальным значениям $\alpha = \pi$, $\omega = 0$ при $t = t_0$. Если начальные значения будут незначительно отличаться от указанных, что всегда имеет место в реальной системе, то изображающая точка в момент $t = t_0$ будет находиться близко к точке $\pi, 0$, но, как видно из рис. (8.26), эта близость не сохранится при $t > t_0$. Характер фазовых траекторий в окрестности точки $(0, 0)$ таков, что если изображающая точка в момент $t = t_0$ находится близко к точке $0, 0$, то и при $t > t_0$ она будет вблизи этой точки. Точку покоя $0, 0$ называют *устойчивой*, а точку $\pi, 0$ — *неустойчивой*.

Дадим точное определение устойчивости точки покоя. Для этого заметим, что если рассматриваемая точка покоя расположена не в начале координат фазовой плоскости, то всегда можно перейти к новой системе координат с началом в этой точке. Действительно, если a, b точки покоя системы (8.40), то, сделав в ней замену переменной

$$x_1 = y_1 + a,$$

$$x_2 = y_2 + b,$$

получим:

$$\left. \begin{aligned} \dot{y}_1 &= f_1(y_1 + a, y_2 + b) \\ \dot{y}_2 &= f_2(y_1 + a, y_2 + b) \end{aligned} \right\},$$

или

$$\left. \begin{aligned} \dot{y}_1 &= g_1(y_1, y_2) \\ \dot{y}_2 &= g_2(y_1, y_2) \end{aligned} \right\}. \quad (8.48)$$

Система (8.48) имеет точку покоя $0, 0$. Движение $y_1 = 0, y_2 = 0$ назовем *невозмущенным движением*. Оно отвечает начальным значениям $y_1 = 0, y_2 = 0$ при $t = t_0$. Начальные значения $y_1 = \delta_1, y_2 = \delta_2$, из которых по крайней мере одно не равно нулю, назовем возмущениями, а движения (8.48), отвечающие этим начальным значениям, — *возмущенными движениями*.

Невозмущенное движение $y_1 \equiv 0, y_2 \equiv 0$ называется *устойчивым по Ляпунову*, если по любому $\varepsilon > 0$ можно подобрать $\delta > 0$ такое, что для всех возмущений, удовлетворяющих неравенствам

$$|\delta_1| \leq \delta,$$

$$|\delta_2| \leq \delta,$$

возмущенное движение $y_1 = y_1(t), y_2 = y_2(t)$ будет удовлетворять неравенствам

$$|y_1(t)| < \varepsilon, \quad |y_2(t)| < \varepsilon$$

при любом $t > t_0$. Если, кроме того, будут иметь место равенства

$$\lim_{t \rightarrow \infty} y_1(t) = 0,$$

$$\lim_{t \rightarrow \infty} y_2(t) = 0,$$

то говорят, что невозмущенное движение *асимптотически устойчиво*.

В соответствии с определением точки покоя II и VI на рис. 8.22 устойчивы, и эта устойчивость асимптотическая, точка покоя IV неустойчива. Точка покоя на рис. 8.27 неустойчива. Точка покоя 0, 0 на рис. 8.26 устойчива, но не асимптотически. Определение устойчивости точки покоя непосредственно распространяется на системы n -го порядка.

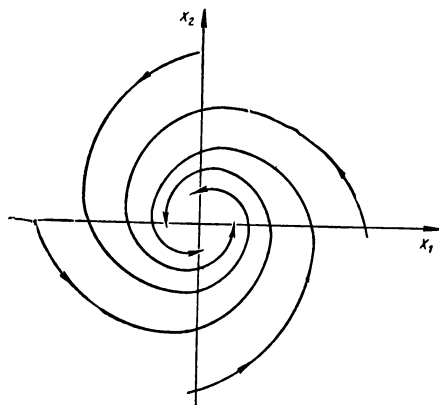


Рис. 8.31. Фазовый портрет системы, устойчивой в целом

Дадим теперь определение устойчивости системы. Автоматическая система называется устойчивой, если она обладает одним положением равновесия и это положение равновесия устойчиво.

Линейная автоматическая система имеет одно положение равновесия. Это положение равновесия асимптотически устойчиво, если все корни характеристического уравнения системы имеют отрицательную вещественную часть (см. § 6.1). Если нелинейная система такова, что для нее можно построить линейную модель (см. § 8.3), то асимптотическая устойчивость линейной модели свидетельствует об устойчивости нелинейной системы, и из неустойчивости линейной модели следует неустойчивость нелинейной системы. Если характеристическое уравнение линейной модели имеет часть корней на мнимой оси, а остальные — в левой полуплоскости, то суждение об устойчивости нелинейной системы по первому приближению невозможно.

Назовем область фазового пространства *областью притяжения точки покоя*, если изображающая точка из любой точки этой области асимптотически приближается к точке покоя. Автоматическая система, все фазовое пространство которой есть область притяжения единственной точки покоя, называется *устойчивой «в целом»*. На рис. 8.31 приведен фазовый портрет системы второго порядка, устойчивой «в целом». Если линейная система асимптотически устойчива, то это значит, что при достаточно малых возмущениях равновесие в системе восстанавливается, но тогда по принципу суперпозиции это положение равновесия восстанавливается при любых возмущениях. Таким образом, асимптотически устойчивая линейная система всегда устойчива «в целом». Для нелинейных систем устойчивость означает только то, что равновесие восстанавливается при достаточно малых возмущениях.

Примеры. 1. Исходя непосредственно из определения, выяснить устойчивость нулевого решения системы [8.2]:

$$\left. \begin{aligned} \frac{dx}{dt} &= -x, \\ \frac{dy}{dt} &= -y. \end{aligned} \right\}$$

Общее решение имеет вид

$$\left. \begin{aligned} x &= x_0 e^{-t}, \\ y &= y_0 e^{-t}. \end{aligned} \right\}$$

Ясно, что при всех $t \geq 0$ будут выполняться неравенства

$$|x_0 e^{-t}| < \varepsilon, \quad |y_0 e^{-t}| < \varepsilon,$$

если

$$|x_0| < \varepsilon, \quad |y_0| < \varepsilon.$$

Следовательно, нулевое решение устойчиво, и эта устойчивость асимптотическая, так как

$$\lim_{t \rightarrow +\infty} x_0 e^{-t} = 0,$$

$$\lim_{t \rightarrow +\infty} y_0 e^{-t} = 0.$$

2. Выяснить устойчивость точки покоя $\pi, 0$ системы

$$\left. \begin{aligned} \frac{d\alpha}{dt} &= \omega, \\ \frac{d\omega}{dt} &= -\sin \alpha. \end{aligned} \right\}$$

Выполним замену переменных:

$$\left. \begin{aligned} \alpha &= x + \pi, \\ \omega &= y, \\ \frac{dx}{dt} &= y, \\ \frac{dy}{dt} &= -\sin(x + \pi). \end{aligned} \right\}$$

Таким образом, замкнутой переменной задачей сведена к исследованию точки покоя $0, 0$ для системы

$$\left\{ \begin{aligned} \frac{dx}{dt} &= y, \\ \frac{dy}{dt} &= \sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \dots \end{aligned} \right.$$

Линейная модель последней системы

$$\left\{ \begin{aligned} \frac{dx}{dt} &= y, \\ \frac{dy}{dt} &= x \end{aligned} \right\}$$

имеет характеристическое уравнение

$$\begin{vmatrix} -p & 1 \\ 1 & -p \end{vmatrix} = 0$$

или

$$p^2 - 1 = 0.$$

Один из характеристических корней $p = 1$ расположен в правой полуплоскости, и, следовательно, точка покоя неустойчива.

3. Выяснить устойчивость точки покоя $0, 0$ системы примера 2.

Линейная модель будет иметь вид

$$\left\{ \begin{aligned} \frac{d\alpha}{dt} &= \omega, \\ \frac{d\omega}{dt} &= -\alpha. \end{aligned} \right\}$$

Характеристическое уравнение

$$\begin{vmatrix} -p & 1 \\ -1 & -p \end{vmatrix} = 0$$

имеет корни $p_1 = i, p_2 = -i$. Никакого заключения об устойчивости точки покоя $0, 0$ рассматриваемой нелинейной системы здесь сделать нельзя. Вид фазовых траекторий в окрестности точки $0, 0$ (см. рис. 8.26) говорит об устойчивости этой точки. Эта устойчивость не является асимптотической.

§ 8.6. АБСОЛЮТНАЯ УСТОЙЧИВОСТЬ

Рассмотрим автономную автоматическую систему, уравнения движения которой имеют вид

$$\left. \begin{aligned} \dot{x}_1 &= a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n + b_1\varphi(\sigma), \\ \dot{x}_2 &= a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n + b_2\varphi(\sigma), \\ &\vdots \\ \dot{x}_n &= a_{n1}x_1 + a_{n2}x_2 + \dots + a_{nn}x_n + b_n\varphi(\sigma), \\ \sigma &= c_1x_1 + c_2x_2 + \dots + c_nx_n, \end{aligned} \right\} \quad (8.49)$$

т. е. отличаются от линейных с постоянными коэффициентами наличием аддитивно входящей нелинейной функции $\varphi(\sigma)$ линейной комбинации обобщенных координат системы [8.1].

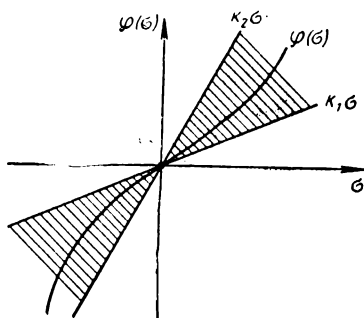
Будем считать, что график функции $\varphi(\sigma)$ непрерывной и определенной для всех вещественных σ лежит между прямыми, проходящими через начало координат с угловыми коэффициентами k_1 и k_2 (рис. 8.32), т. е.

$$k_1 \leq \frac{\varphi(\sigma)}{\sigma} \leq k_2. \quad (8.50)$$

Так как $\varphi(0) = 0$,

$$x_i = 0 \quad (8.51)$$

$$(i = 1, 2, \dots, n)$$



является решением системы (8.49), будем изучать устойчивость этого решения, определяющего положение системы.

Рис. 8.32. Графическая интерпретация неравенства (8.50)

Система (8.49) называется *абсолютно устойчивой* в определяемом (8.50) угле $[k_1, k_2]$, если ее решение (8.51) устойчиво «в целом» при любом выборе функции $\varphi(\sigma)$, удовлетворяющей условию (8.50). Покажем, что вопрос об абсолютной устойчивости системы (8.49) в угле $[k_1, k_2]$ при наперед заданных числах k_1 и k_2 можно свести к вопросу об абсолютной устойчивости системы в угле $[0, k]$, где $k = k_2 - k_1$. Для этого в (8.49) выполним подстановку

Рис. 8.34. Структурная схема нелинейной системы (8.49)

[illegible]

Решим систему (8.53) относительно $-\Sigma$:

$$-\Sigma(p) = \frac{P_1(p) Y(p)}{P_2(p)},$$

$$P_2(p) = \begin{vmatrix} p - a_{11} - a_{12} \dots - a_{1n} & 0 \\ -a_{21} & p - a_{22} \dots - a_{2n} & 0 \\ \dots & \dots & \dots \\ -a_{n1} & -a_{n2} \dots p - a_{nn} & 0 \\ c_1 & c_2 \dots c_n & 1 \end{vmatrix} = \begin{vmatrix} p - a_{11} - a_{12} \dots a_{1n} \\ -a_{21} & p - a_{22} \dots - a_{2n} \\ \dots & \dots & \dots \\ -a_{n1} & -a_{n2} \dots p - a_{nn} \end{vmatrix}$$

$$P_1(p) = \begin{vmatrix} p - a_{11} & -a_{12} & \dots & -a_{1n} & b_1 \\ -a_{21} & p - a_{22} & \dots & -a_{2n} & b_2 \\ \dots & \dots & \dots & \dots & \dots \\ -a_{n1} & -a_{n2} & \dots & p - a_{nn} & b_n \\ c_1 & c_2 & \dots & c_n & 0 \end{vmatrix}.$$

Таким образом $W(p) = P_1(p)/P_2(p)$. Если система (8.49) абсолютно устойчива в угле $[0, k]$, т. е. ее решение (8.51) устойчиво «в целом» при любой непрерывной $\varphi(\sigma)$, удовлетворяющей неравенству $0 \leq \frac{\varphi(\sigma)}{\sigma} \leq k$, то она устойчива при $\varphi(\sigma) \equiv 0$. Так как при этом (8.49) является линейной системой

$$\left. \begin{aligned} \dot{x}_1 &= a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n, \\ \dot{x}_2 &= a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n, \\ &\vdots \\ \dot{x}_n &= a_{n1}x_1 + a_{n2}x_2 + \dots + a_{nn}x_n \end{aligned} \right\} \quad (8.54)$$

с характеристическим уравнением $P_2(p)=0$, корни этого уравнения лежат в левой полуплоскости.

Рассмотрение абсолютно устойчивой системы (8.49) в угле $[0, k]$ при линейных характеристиках $\varphi(\sigma) = h\sigma$ и

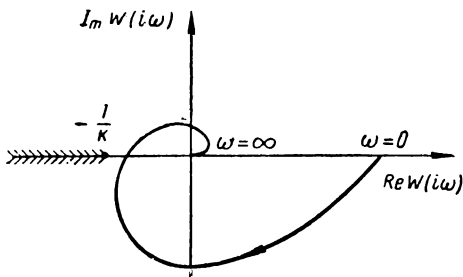


Рис. 8.35. Амплитудно-фазовая характеристика второго рода линейной части системы (8.49)

применение критерия Найквиста позволяет получить следующее необходимое условие абсолютной устойчивости.

Если система (8.49) абсолютно устойчива в угле $[0, k]$ (рис. 8.34), то амплитудно-фазовая характеристика второго рода ее линейной части не пересекает действительную ось между $-\infty$ и точкой $-\frac{1}{k}$, включая эту точку (рис. 8.35).

Действительно, из абсолютной устойчивости системы (8.49) в угле $[0, k]$ следует устойчивость линейной системы со структурной схемой, изображенной на рис. 8.34, при любом коэффициенте усиления h , удовлетворяющем

условию

$$0 \leq h \leq k. \quad (8.55)$$

Передаточная функция разомкнутой системы (8.53) равна $hW(p)$ и по критерию Найквиста годограф вектора $hW(i\omega)$ пересекает отрезок действительной оси левее точки $-1, 0$ четное число раз при любом h , лежащем между 0 и k (8.55). Далее ясно, что это четное число есть нуль, так как если при h_0 ($0 < h_0 \leq k$) годограф вектора $h_0W(i\omega)$ пересекает вещественную ось левее точки $-1, 0$, например, два раза, то при некотором h_1 ($0 < h_1 < h_0$) годограф вектора $h_1W(i\omega)$ будет иметь

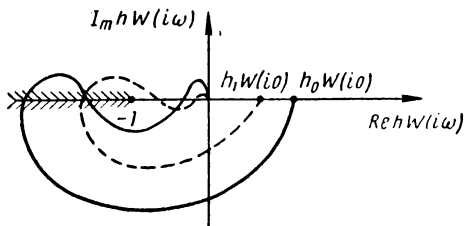


Рис. 8.36. К доказательству необходимого условия абсолютной устойчивости

одно такое пересечение (рис. 8.36). Из того что годограф вектора $kW(i\omega)$ не пересекает действительную ось левее точки $-1, 0$, следует, что годограф вектора $W(i\omega)$ не пересекает ее левее точки $-\frac{1}{k}, 0$.

Румынский ученый В. М. Попов получил достаточные условия абсолютной устойчивости, которые можно сформулировать в графической форме, аналогичной форме вышеизложенного необходимого условия. Введем в рассмотрение видоизмененную частотную характеристику линейной части системы (8.49) $W^*(i\omega)$, определив ее равенствами

$$\operatorname{Re} W^*(i\omega) = \operatorname{Re} W(i\omega),$$

$$\operatorname{Im} W^*(i\omega) = \omega \operatorname{Im} W(i\omega). \quad (8.56)$$

Годограф вектора $W^*(i\omega)$ назовем видоизмененной амплитудно-фазовой характеристикой линейной части системы (8.49). Так как видоизмененная амплитудно-фазо-

вая характеристика пересекает действительную ось в тех же точках, что и амплитудно-фазовая, то необходимое условие можно сформулировать так. Для абсолютной устойчивости системы (8.49) в угле $[0, k]$ необходимо, чтобы видоизмененная амплитудно-фазовая характеристика не пересекала отрезок действительной оси от $-\infty$ до $-\frac{1}{k}$. Если теперь предъявить более жесткие требо-

вания к виду видоизмененной амплитудно-фазовой характеристики, то получим следующие достаточные условия

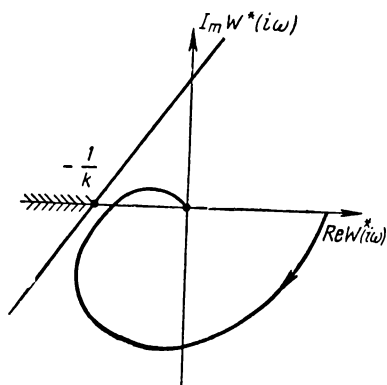


Рис. 8.37. Видоизмененная амплитудно-фазовая характеристика линейной части системы (8.49), абсолютно устойчивой в угле

абсолютной устойчивости системы (8.49) в угле $[0, k]$, которые приводятся здесь без доказательства. [8.1].

Чтобы имела место абсолютная устойчивость системы (8.49) в угле $[0, k]$, достаточно, чтобы ее линейная часть была устойчива и чтобы в плоскости $W^*(i\omega)$ можно было провести прямую через точку $-\frac{1}{k}, 0$ так, чтобы видо-

измененная амплитудно-фазовая характеристика линейной части вся лежала строго справа от этой прямой. Вышеуказанную

прямую будем называть *прямой Попова*, а точку $W^*(i\infty)$ не будем причислять к точкам видоизмененной амплитудно-фазовой характеристики, так что эта точка может быть расположена на прямой Попова.

На рис. 8.37 показан случай амплитудно-фазовой характеристики, в котором выполнены достаточные условия абсолютной устойчивости в угле $[0, k]$ в предположении, что полюса $W(p)$ лежат в левой полуплоскости. На рис. 8.38 показан случай, когда необходимое условие выполнено, а достаточное условие абсолютной устойчивости в угле $[0, k]$ не выполнено. Но если полюса $W(p)$ лежат в левой полуплоскости p , то выполнены достаточные условия Попова в меньшем угле $[0, k_1]$ ($k_1 < k$). Этот пример показывает, что не имеет смысла говорить просто

об абсолютной устойчивости, можно говорить об абсолютной устойчивости в каком-то угле.

Если передаточная функция $W(p)$ линейной части системы (8.49) имеет один нулевой полюс, а остальные полюса расположены в левой полуплоскости, то система (8.49) не является абсолютно устойчивой в угле $[0, k]$, так как она не является устойчивой при $\varphi(\sigma) \equiv 0$. В этом случае можно ставить вопрос об ее устойчивости в угле $[\varepsilon, k]$, где ε — сколь угодно малое положительное число.

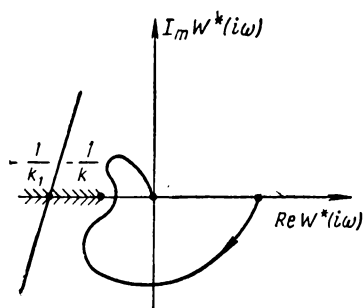


Рис. 8.38. Видоизмененная амплитудно-фазовая характеристика линейной части системы (8.49), не являющейся абсолютно устойчивой в угле $[0, K]$

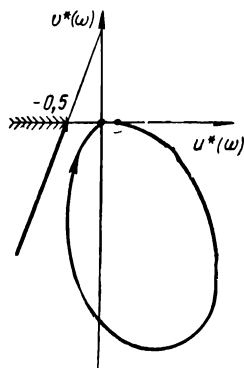


Рис. 8.39. Видоизмененная амплитудно-фазовая характеристика и прямая Попова для примера § 8.6

В случае когда характеристическое уравнение системы (8.49) имеет один нулевой корень, а остальные корни — в левой полуплоскости p , для абсолютной устойчивости этой системы достаточно, чтобы амплитудно-фазовая характеристика ее линейной части удовлетворяла следующим условиям:

$$\lim \operatorname{Im} W(i\omega) = -\infty$$

или

$$\operatorname{Im} W^*(i\omega) < 0,$$

видоизмененная амплитудно-фазовая характеристика линейной части лежит строго справа от прямой Попова, проведенной через точку $(-\frac{1}{k}, 0)$ в плоскости $W^*(i\omega)$.

Пример.

Является ли система

$$\left. \begin{aligned} \dot{x}_1 &= -x_1 + x_3, \\ \dot{x}_2 &= -2x_2 - x_3, \\ \dot{x}_3 &= x_1 + x_2 - x_3 - \varphi(x_1) \end{aligned} \right\} \quad (8.57)$$

абсолютно устойчивой в угле $[1, 3]$?

Выполним подстановку $\varphi_1(x_1) = \varphi(x_1) - x_1$, которая сводит задачу к исследованию в угле $[0, 2]$ абсолютной устойчивости системы

$$\left. \begin{aligned} \dot{x}_1 &= -x_1 + x_3, \\ \dot{x}_2 &= -2x_2 - x_3, \\ \dot{x}_3 &= x_2 - x_3 - \varphi_1(x_1). \end{aligned} \right\} \quad (8.58)$$

Передающая функция линейной части этой системы (от входа $y = \varphi_1(x_1)$ к выходу $-x_1$) имеет вид (§ 8.3)

$$W(p) = \frac{p+2}{p^3+4p^2+6p+3},$$

а частотная характеристика —

$$W(i\omega) = \frac{2+i\omega}{3-4\omega^2+i(6\omega-\omega^3)}.$$

Найдем вещественную $U(\omega)$ и мнимую $V(\omega)$ части частотной характеристики:

$$U(\omega) = \frac{6-2\omega^2-\omega^4}{9+12\omega^2+4\omega^4+\omega^6},$$

$$V(\omega) = \frac{-9\omega-2\omega^3}{9+12\omega^2+4\omega^4+\omega^6}.$$

Вещественная и мнимая части видоизмененной частотной характеристики имеют вид:

$$U^*(\omega) = \frac{6-2\omega^2-\omega^4}{9+12\omega^2+4\omega^4+\omega^6},$$

$$V^*(\omega) = \frac{-9\omega^3-2\omega^4}{9+12\omega^2+4\omega^4+\omega^6}.$$

На рис. 8.39 построена видоизмененная амплитудно-фазовая характеристика. Из построения видно, что система (8.58) является абсолютно устойчивой в угле $[0, 2]$, а следовательно, система (8.57) абсолютно устойчива в угле $[1, 3]$.

САМОПРИСПОСАБЛИВАЮЩИЕСЯ СИСТЕМЫ

§ 9.1. СИСТЕМЫ ЭКСТРЕМАЛЬНОГО И ОПТИМАЛЬНОГО РЕГУЛИРОВАНИЯ

Среди самоприспосабливающихся систем в настоящее время наибольшее развитие получили сравнительно простые, однако имеющие большое практическое значение виды этих систем — системы экстремального и оптимального регулирования (СЭР и СОР). Эти системы предназначены для автоматического поиска и поддержания оптимального управляющего воздействия, обеспечивающего работу объекта управления при экстремальном (максимальном или минимальном) значении принятого *критерия эффективности действия (критерия качества)*.

Для большинства объектов управления такая задача не может быть решена с помощью обычных САР, так как объекты обычно находятся под воздействием неконтролируемых возмущений, изменяющих регулируемую координату непредвиденным образом.

Устройство, предназначенное для автоматического поиска и поддержания экстремального значения показателя эффективности действия объекта, называется *автоматическим оптимизатором* (см. § 2.2). В зависимости от количества управляющих воздействий, используемых для получения экстремума функции качества, различают *одноканальные и многоканальные оптимизаторы*. В дальнейшем для большей простоты будем рассматривать, исключая особо оговоренные случаи, только экстремальные системы с одним регулирующим органом, т. е. с одноканальным оптимизатором.

Характер задач, решаемых СЭР и СОР, был рассмотрен в § 2.2 на конкретном примере регулирования самолетного двигателя. Характеристики, представленные на рис. 9.1 [9.17], показывают, что километровый расход Q топлива зависит от скорости V полета и при некотором оптимальном значении скорости $V_{\text{опт}}$ достигает минимума.

Из рис. 9.1 также видно, что с изменением веса G самолета оптимальное значение скорости полета также изменяется. Поскольку у сверхдальних самолетов вес топлива составляет значительную часть общего веса самолета, то

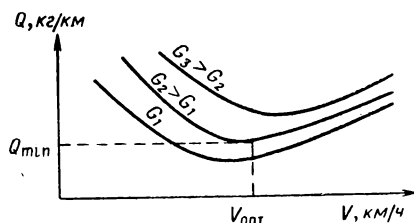


Рис. 9.1. Зависимость расхода топлива от скорости и веса самолета

по мере сгорания топлива оптимальный режим работы двигателей будет нарушаться. Следовательно, в процессе полета необходимо все время подбирать оптимальную скорость полета за счет изменения режима работы самолетных двигателей.

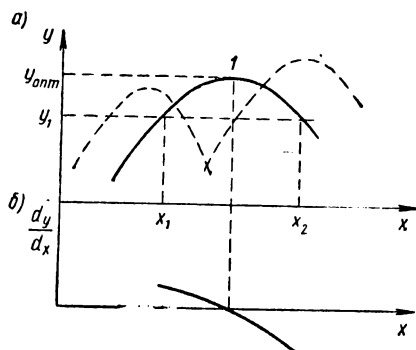


Рис. 9.2. Статическая характеристика объекта управления

Задачи аналогичного характера возникают при автоматизации работы паровых котлов, в химическом производстве, при эксплуатации двигателей внутреннего сгорания и т. п.

Основная особенность оптимального регулирования заключается в том, что заранее не известно требуемое

значение регулируемого параметра, и анализ состояния объекта регулирования в данный момент времени не может дать информации о направлении изменения управляющего воздействия для получения необходимого эффекта. Действительно, если известно только одно значение выходного сигнала y_1 , то по кривой $y = f(x)$, положение и форма которой могут изменяться с течением времени (рис. 9.2), такое значение выходной величины можно получить при $x = x_1$ и $x = x_2$, при этом направление изменения входного воздействия для приближения к экстремуму функции остается неизвестным.

Поэтому для решения задачи оптимизации работа СОР должна быть организована таким образом, чтобы в ней осуществлялся непрерывный поиск значения управляющего воздействия, при котором выходной сигнал обладал бы признаками экстремального значения. Такими признаками являются достижение регулируемой величиной наибольшего или наименьшего значения или изменение знака производной $\frac{dy}{dx}$ (чувствительности) при прохождении через экстремум.

Кривая $y = f(x)$ на рис. 9.2,а в точке 1 имеет экстремум, признаком которого может служить равенство нулю производной $\frac{dy}{dx}$. На рис. 9.2,б показан характер изменения чувствительности системы.

Работа СЭР основана на том, что первоначально входной сигнал принудительно изменяется в случайном направлении. Последующий анализ результата этих действий позволяет определить, приблизилось ли новое значение регулируемой величины к экстремальному значению (тогда необходимо продолжать изменение управляющего воздействия в прежнем направлении до достижения экстремума) или же регулируемая величина удалилась от экстремума и знак приращения входного сигнала надо изменить. После достижения экстремума процесс поиска обычно не прекращается, так как неконтролируемые возмущения могут изменить характеристику работы СЭР.

Наличие режима поиска экстремального значения регулируемой величины является характерной особенностью СЭР, поэтому их обычно классифицируют по признаку формирования поискового сигнала. Чаще всего выделяют следующие типы СЭР:

- 1) с запоминанием экстремума;
- 2) реагирующие на знак и величину производной $\frac{dy}{dx}$;
- 3) с непрерывным поисковым сигналом;
- 4) шагового типа.

СЭР с запоминанием экстремума. Действие СЭР с запоминанием экстремума основано на том, что в системе осуществляется непрерывное сравнение текущего значения регулируемого параметра с хранящимся в запоминающем устройстве экстремальным значением этого параметра. На рис. 9.3 представлена функциональная схема СЭР с запоминанием экстремума, а на рис. 9.4 — графики, поясняющие принцип ее работы. В такой системе значение сигнала на выходе запоминающего устройства ЗУ должно быть равно текущему значению регулируемого параметра до тех пор, пока выходной сигнал объекта управления возрастает. На входы устройства сравнения УС при этом подаются с запоминающего устройства ЗУ

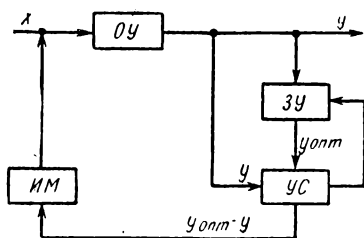


Рис. 9.3. СЭР с запоминанием экстремума

и объекта управления ОУ одинаковые сигналы, поэтому сигнал на выходе УС будет отсутствовать (участок $0 — t_1$ на рис. 9.4). После достижения максимума с дальнейшим изменением входного воздействия регулируемый параметр будет уменьшаться, при этом $y_{з.у}$ (выходной сигнал ЗУ) должен сохранять свое максимальное значение (участок $t_1 — t_2$). В этом случае на выходе УС появится разностный сигнал $y — y_{опт}$; когда эта разность достигнет величины срабатывания $u_{ср}$, определяемого зоной нечувствительности исполнительного механизма ИМ, последний осуществит реверс входного воздействия (участок $t_2 — t_4$). Одновременно с этим прежний сигнал ЗУ сбрасывается до уровня y_1 текущего значения регулируемого параметра, после чего весь цикл работы повторяется.

Один из возможных вариантов запоминающего устройства, предназначенного для СЭР с запоминанием экстремума, показан на рис. 9.5 [9.7]. Входное напряжение, пропорциональное регулируемому параметру объекта,

подается через диод D на запоминающий конденсатор C , который при увеличении входного напряжения U_1 будет заряжаться. При понижении напряжения U_1 выходной сигнал U_2 остается постоянным и пропорциональным экстремальному значению регулируемого параметра. При подаче в схему напряжения U_3 , вырабатываемого УС

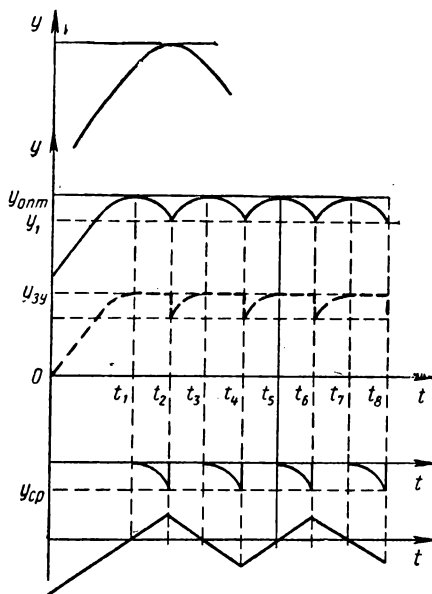


Рис. 9.4. График работы СЭР с запоминанием экстремума

и зависящего от разности сигналов $y_3, y - y$, конденсатор C разряжается через сопротивление R , в результате чего напряжение U_2 вновь становится равным напряжению U_1 , которое пропорционально величине y_1 .

СЭР с запоминанием экстремума весьма просты по своему устройству и обладают высокой помехоустойчивостью.

СЭР, реагирующие на изменение производной. СЭР, реагирующие на величину и знак производной, требуют измерения производных $\frac{dy}{dx}$ или $\frac{dy}{dt}$ или хотя бы определения знака производной $\frac{dy}{dx}$ (чувствительности), а сле-

довательно, и применения специальных дифференцирующих устройств.

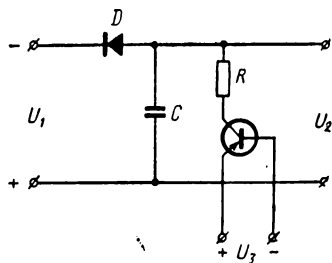


Рис. 9.5. Схема запоминающего устройства

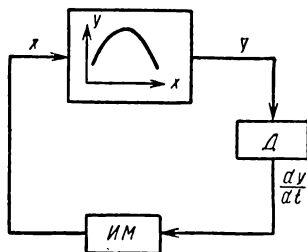


Рис. 9.6. СЭР с измерением производной:

Д — дифференцирующее устройство;
ИМ — исполнительный механизм

Функциональная схема простейшей СЭР с управлением по чувствительности приведена на рис. 9.6, а рис. 9.7

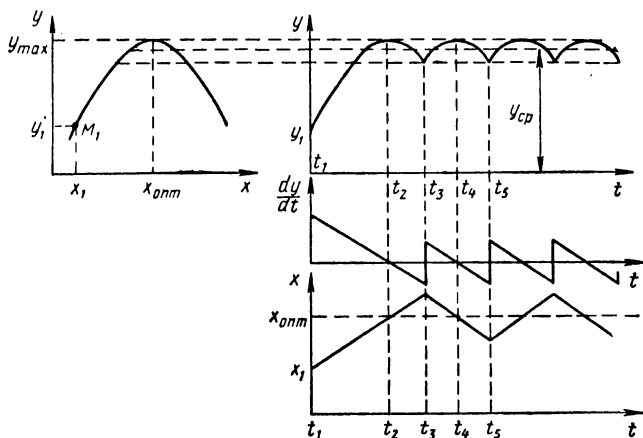


Рис. 9.7. График работы СЭР с измерением производной

поясняет действие такой системы [9.17]. Когда производная $\frac{dy}{dt}$ положительна, увеличение входного воздействия приводит к увеличению регулируемого параметра y (участок $t_1 — t_2$).

После достижения экстремума в момент времени t_2 входное воздействие x продолжает изменяться в прежнем направлении на протяжении отрезка времени $t_2 - t_3$, величина которого зависит от зоны нечувствительности исполнительного механизма ИМ. При этом знак производной $\frac{dy}{dt}$ изменится на отрицательный. В момент t_3 величина производной $\frac{dy}{dt}$ превысит зону нечувствительности, в результате чего будет осуществлен реверс ИМ и входное воздействие уменьшится, что приведет к тому, что выходной сигнал вновь приблизится к экстремуму. Такое движение системы в обратном направлении (участок $t_3 - t_6$) продолжается до нового реверса, после чего цикл работы будет повторяться. Оптимизатор строится таким образом, чтобы в процессе работы регулируемая величина все время удерживалась в пределах $y_{\text{опт}} - y_1$, определяемых величиной зоны нечувствительности ИМ.

При построении СЭР с измерением чувствительности $\frac{dy}{dx}$ можно воспользоваться двумя дифференцирующими устройствами, выходной сигнал которых пропорционален производным $\frac{dx}{dt}$ и $\frac{dy}{dt}$. Специальное делительное устройство осуществляет деление $\frac{dy}{dt} : \frac{dx}{dt}$, воспроизводя на выходе сигнал, пропорциональный чувствительности $\frac{dy}{dx}$. Однако такие СЭР обладают рядом недостатков, ограничивающих их практическое применение (недостаточная стабильность и надежность в работе, недостаточная помехоустойчивость, сложность).

Следует отметить также, что скорость изменения входного воздействия в режиме поиска не обязательно должна быть постоянной. На практике часто применяются СЭР, в которых скорость изменения входного воздействия в режиме поиска пропорциональна абсолютному значению производной $\frac{dy}{dt}$. Применение такого принципа работы позволяет уменьшить время выхода СЭР на оптимальный режим работы и добиться более точного приближения к экстремальному значению регулируемого параметра. Такие системы обычно называются СЭР с пропорциональным поисковым сигналом.

СЭР с непрерывным поисковым сигналом (непрерывные СЭР [9.9], СЭР с вспомогательной модуляцией). Принцип действия *СЭР с непрерывным сигналом* заключается в том, что на вход, помимо управляющего воздействия, подается дополнительный непрерывный сигнал небольшой амплитуды и фиксированной частоты. В качестве дополнительного воздействия удобнее всего пользоваться синусоидальным сигналом. Получающаяся при этом дополнительная синусоидальная составляющая выходного сигнала системы той же частоты ω выделяется с помощью фильтров и сравнивается с входным поисковым воздействием.

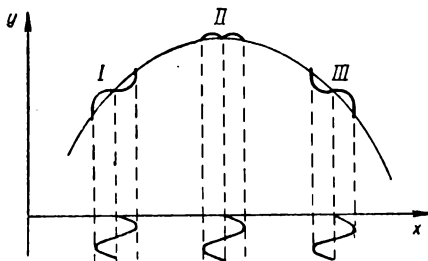


Рис. 9.8. Влияние поискового синусоидального воздействия на выходной сигнал системы

Признаком, определяющим направление изменения входного воздействия для достижения экстремума, является фаза поисковой составляющей выходного сигнала, которая инвертируется при переходе регулируемой величины через экстремум. Действительно, если в двух различных состояниях СЭР (точки I и III на рис. 9.8) подать на вход вместе с управляющим воздействием один и тот же поисковый сигнал $a \sin \gamma t$, то, как видно из рисунка, фаза поисковых составляющих выходного сигнала на участках I и III отличаются друг от друга на 180° . Таким образом, фаза поисковой составляющей выходного сигнала указывает на расположение рабочей точки СЭР относительно экстремума, а амплитуда ее, уменьшающаяся по мере приближения к экстремуму, определяет степень удаления рабочей точки от экстремума.

На рис. 9.9 представлена функциональная схема СЭР с непрерывным поисковым сигналом. Характер работы такой системы поясняется на рис. 9.10. Генератор низ-

кой частоты Γ предназначен для создания синусоидальных поисковых колебаний. Соответствующие им выходные колебания выделяются фильтром Φ_1 и поступают на множительное устройство МУ. При прохождении сигнала через реальную систему в нем появляются фазовые сдвиги, поэтому при сопоставлении входных и выходных колебаний в исследуемой системе такие фазовые сдвиги должны быть скомпенсированы. Эту функцию выполняет фазосдвиг

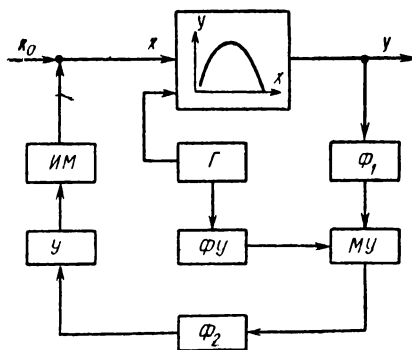


Рис. 9.9. СЭР с непрерывным поисковым сигналом

гающее устройство Φ_U , вырабатывающее сигнал $a \sin \times (\gamma t + \varphi)$, где φ — суммарный фазовый сдвиг, вызываемый реальной системой и фильтром Φ_1 . Этот сигнал и поисковая составляющая выходного сигнала $b \sin (\gamma t + \varphi)$ подаются на вход множительного устройства МУ. Произведение этих величин

$$a \sin (\gamma t + \varphi) \cdot b \sin (\gamma t + \varphi) = \frac{ab}{2} [1 - \cos 2 (\gamma t + \varphi)], \quad (9.1)$$

если рабочая точка расположена слева от максимума (участок I на рис. 9.8) и

$$\begin{aligned} a \sin (\gamma t + \varphi) \cdot b \sin (\gamma t + \varphi + \pi) &= \\ &= -\frac{ab}{2} [1 - \cos 2 (\gamma t + \varphi)], \end{aligned} \quad (9.2)$$

когда эта точка находится справа от максимума (участок III).

Выходной сигнал МУ поступает на сглаживающий фильтр Φ_2 , задерживающий переменную составляющую

сигнала МУ. Постоянная составляющая этого сигнала используется для управления исполнительным механизмом ИМ, осуществляющим изменение управляющего воздействия x .

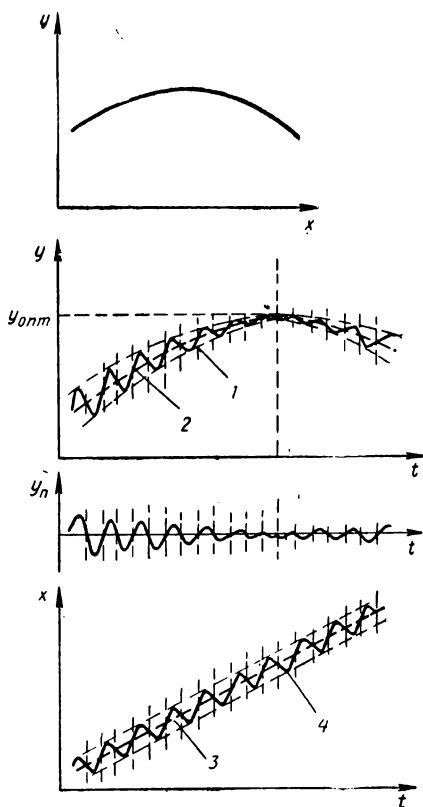


Рис. 9. 10. График работы СЭР с непрерывным поисковым сигналом:

1 — синусоидальная составляющая выходного сигнала; 2 — регулярная составляющая выходного сигнала; 3 — регулярный входной сигнал; 4 — синусоидальная составляющая входа; y_n — поисковая составляющая выхода

Схема, изображенная на рис. 9.9, позволяет получить переменную скорость изменения входного сигнала x . Действительно, чем больше удалена рабочая точка от максимума, тем больше амплитуда поисковой составляющей выходного сигнала y и тем больше выходной сигнал усилителя Y , а следовательно, тем быстрее ИМ изменяет входной сигнал.

СЭР с непрерывным поисковым сигналом обладают большей помехоустойчивостью по сравнению с системами с измерением производной, так как не содержат дифференцирующих устройств, увеличивающих влияние помех. Кроме того, такие системы позволяют осуществлять оптимизацию по нескольким параметрам одновременно, если применить поисковые воздействия с различными частотами.

Однако практическое применение таких систем ограничивается трудностью точной компенсации фазового сдвига φ , сравнительной их сложностью или наличием модулирующих колебаний, которые могут создать недопустимые в эксплуатации условия работы.

СЭР шагового типа. Функциональная схема СЭР шагового типа показана на рис. 9.11. Для функционирования такой системы в ней необходимо осуществлять квантова-

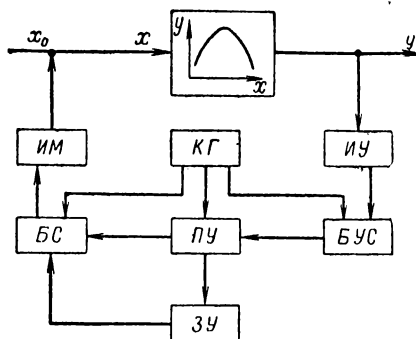


Рис. 9.11. СЭР шагового типа

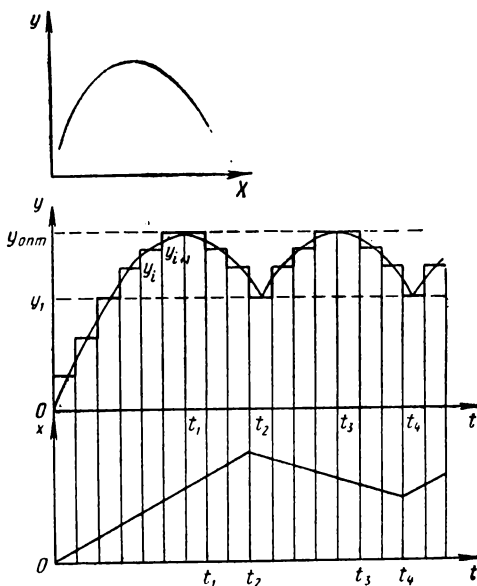


Рис. 9.12. График работы СЭР шагового типа

ние регулируемого сигнала по времени. Принцип действия СЭР шагового типа заключается в том, что в ней через равные интервалы времени осуществляется измерение регу-

лируемой величины с последующим анализом полученных значений с учетом предыдущих результатов. Например, если при поиске максимума функции $y = f(x)$ (рис. 9.12) разность $y_{i+1} - y_i$ окажется положительной, то входное воздействие x следует изменять в прежнем направлении. Если же эта разность станет отрицательной, то исполнительный механизм ИМ должен сделать очередной шаг в сторону уменьшения сигнала x . При этом можно применить непрерывный входной сигнал $x(t)$ или сигнал непре-

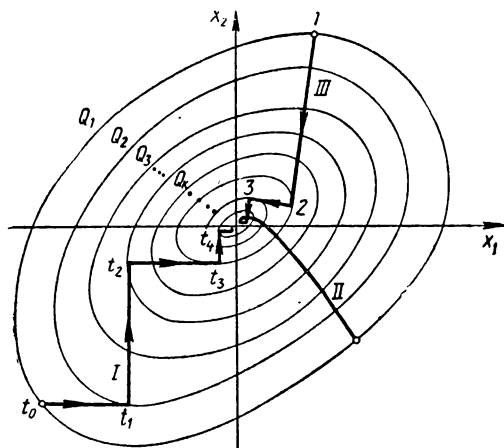


Рис. 9.13. Методы поиска экстремума функции двух переменных

рывного типа или в виде ступенчатой функции, изменяющейся при каждом шаге на Δx .

Вся последовательность работы системы координируется командным генератором КГ. Измерительное устройство ИУ подает сигнал на блок усреднения сигнала БУС, который вырабатывает среднее значение выходного сигнала за один шаг. Отметим, что при каждом i -м шаге БУС выдает среднее значение выходного сигнала за $(i - 1)$ -й шаг. Запоминающее устройство ЗУ предназначено для хранения информации о значении регулируемой величины y_{i-1} в предыдущем шаге. При очередном шаге измеренное текущее значение регулируемой величины y_i и считываемое в ЗУ значение y_{i-1} с помощью переключающего устройства ПУ подаются в блок сравнения БС, выходной сигнал которого пропорционален разности $y_i -$

— y_{i-1} . Этот выходной сигнал и служит для управления исполнительным механизмом. При $y_i - y_{i-1} > 0$ старая запись в ЗУ сбрасывается, запоминается значение y_i , направление изменения сигнала x сохраняется. При $y_i - y_{i-1} < 0$ осуществляется реверс ИМ. Достигнув максимума функции $y = f(x)$, система будет «рыскать» около экстремума.

Для нормального функционирования СЭР шагового типа важно правильно выбрать величину шага. Шаг должен быть достаточно большим, чтобы уменьшить влияние высокочастотных помех, способных вызвать ложные реверсы ИМ. В то же время увеличение шага ограничивается опасностью ложных срабатываний, вызванных дрейфом характеристики управляемой системы. Шаговые СЭР относятся к числу наиболее надежных.

Организация поиска. При решении практических задач чаще встречаются не простейшие системы, в которых оптимизируемая величина зависит только от одного управляющего воздействия, а СЭР, предназначенные для оптимизации выходной координаты или некоторого критерия качества, являющихся функцией нескольких независимых друг от друга величин. При этом особое значение приобретает вопрос о применении наиболее рационального метода поиска экстремума такой функции. Существует значительное количество возможных способов организации движения СЭР к экстремуму, из них наиболее перспективными по простоте и своим характеристикам являются следующие.

Метод поочередного изменения параметров (метод Гаусса — Зайделя). Пусть задачей многоканального оптимизатора является отыскание экстремума некоторого критерия качества $Q = F(x_1, x_2, \dots, x_n)$, который будем считать непрерывно дифференцируемой функцией при всех конечных значениях x_1, x_2, \dots, x_n . Рассматриваемый метод основан на поочередном изменении каждого из переменных параметров x_i до достижения частных экстремумов по этим переменным при фиксированных значениях остальных параметров. Сначала в СЭР изменяется параметр x_1 в направлении, обеспечивающем уменьшение абсолютного значения компоненты $\frac{\partial Q}{\partial x_1}$ градиента функции Q при $x_2 = \text{const}, x_3 = \text{const}, \dots, x_n = \text{const}$. После достижения $\frac{\partial Q}{\partial x_1} = 0$ осуществляется поиск второго

частного экстремума $\frac{\partial Q}{\partial x_2} = 0$, а затем аналогичным образом находят $\frac{\partial Q}{\partial x_3}$ и так далее. Такое функционирование СЭР продолжается по всем n координатам системы, после чего снова начинается изменение координаты x_1 до получения $\frac{\partial Q}{\partial x_1} = 0$ и весь цикл повторяется, пока оптимизатор не отыщет экстремум функции Q .

Поведение СЭР в режиме поиска экстремума методом Гаусса—Зайделя поясняется рис. 9.13, где для упрощения рассматривается функция двух переменных $Q = F(x_1, x_2)$.

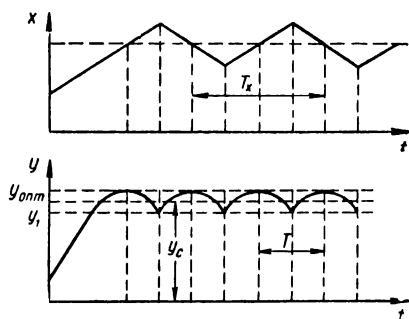


Рис. 9.14. Характеристики экстремальных систем

На этом графике построено семейство линий $Q_i = \text{const}$, причем точка экстремума функции Q совпадает с началом координат.

Если в начале работы СЭР находилась в точке t_0 (ломаная I), то изменение координаты x_1 при $x_2 = \text{const}$ будет осуществляться до получения $\frac{\partial Q}{\partial x_1} = 0$, что произойдет в точке t_1 . Затем фиксируется полученное значение переменной x_1 и изменяется величина x_2 (участок $t_1 - t_2$) и так далее до выхода в начало координат.

Достоинство метода Гаусса—Зайделя заключается в простоте устройств, реализующих такое движение к экстремуму, однако путь системы в область экстремума не является самым целесообразным.

Метод градиента. Особенность метода градиента заключается в одновременном изменении всех координат

x_1, x_2, \dots, x_n и определении всех компонент градиента функции Q . При этом изменение независимых параметров x_1, x_2, \dots, x_n может быть непрерывным или дискретным, причем скорость непрерывного изменения каждой из величин x_i следует выбирать пропорциональной соответствующим частным производным функции Q :

$$\frac{dx_1}{dt} = k \frac{\partial Q}{\partial x_1}, \quad \frac{dx_2}{dt} = k \frac{\partial Q}{\partial x_2}, \quad \dots, \quad \frac{dx_n}{dt} = k \frac{\partial Q}{\partial x_n}, \quad (9.3)$$

где $k = \text{const}$ — коэффициент пропорциональности ($k > 0$ при экстремуме-максимуме и $k < 0$ при экстремуме-минимуме).

При дискретном изменении координат x_i величина шага по каждой из этих координат должна быть пропорциональна соответствующим компонентам градиента:

$$\Delta x_1 = k \frac{\partial Q}{\partial x_1}, \quad \Delta x_2 = k \frac{\partial Q}{\partial x_2}, \quad \dots, \quad \Delta x_n = k \frac{\partial Q}{\partial x_n}. \quad (9.4)$$

Кривая II (рис. 9.13) иллюстрирует движение СЭР с непрерывным изменением параметров x_i в область экстремума. Эта кривая ортогональна к кривым $Q_i = \text{const}$.

При применении метода градиента движение системы к экстремуму представляет собой монотонный сходящийся процесс. Действительно, замечая, что производная функции качества Q по времени

$$\frac{dQ}{dt} = \frac{\partial Q}{\partial x_1} \frac{dx_1}{dt} + \frac{\partial Q}{\partial x_2} \frac{dx_2}{dt} + \dots + \frac{\partial Q}{\partial x_n} \frac{dx_n}{dt}, \quad (9.5)$$

с учетом (9.3) имеем

$$\frac{dQ}{dt} = k \left[\left(\frac{\partial Q}{\partial x_1} \right)^2 + \left(\frac{\partial Q}{\partial x_2} \right)^2 + \dots + \left(\frac{\partial Q}{\partial x_n} \right)^2 \right]. \quad (9.6)$$

Из (9.6) следует, что $\frac{dQ}{dt}$ везде сохраняет знак и в положении экстремума обращается в нуль.

Метод градиента характеризуется относительно быстрым и плавным выходом системы к экстремуму, а при шаговом движении — малой амплитудой колебаний относительно экстремума.

Метод наискорейшего спуска. Сущность этого метода заключается в том, что движение системы происходит в направлении вектора градиента в исходной точке системы (точка I на ломаной III, см. рис. 9.13) до тех

пор, пока частная производная функции Q по этому направлению не станет равной нулю (точка 2). В точке 2 снова определяется направление градиента и осуществляется движение вдоль нового вектора до точки 3, где производная функции Q по направлению этого вектора обращается в нуль. Такое движение системы повторяется до выхода ее в точку экстремума. Излом траектории движения происходит в точках касания направлений градиентов с линиями равного показателя качества.

Основным достоинством *метода наискорейшего спуска* является сравнительно малое время выхода системы в район экстремума.

В ряде случаев наилучшие результаты могут быть достигнуты за счет применения в системе комбинации нескольких методов движения. Например, рациональным можно считать объединение метода наискорейшего спуска, обеспечивающего наименьшее время выхода системы в район экстремума при больших начальных отклонениях, и метода градиента при движении системы в окрестностях экстремума.

Динамика и устойчивость СЭР. Большинство применяемых на практике СЭР после отыскания экстремума совершают колебания (*рыскания*) около оптимального значения регулируемой величины. Амплитуда этих колебаний зависит от величины зоны нечувствительности исполнительного механизма, управляющего изменениями входного воздействия x_1 , и определяет величину *зоны поиска* $y_1 - y_{\text{опт}}$.

Наличие периодических колебаний выходной величины приводит к тому, что среднее значение регулируемой величины y_c (рис. 9.14) будет отличаться от экстремального $y_{\text{опт}}$: при поиске максимума $y_c < y_{\text{опт}}$, а при минимизации функции $y_c > y_{\text{опт}}$. Разность $y_{\text{опт}} - y_c$ называется *потерей на поиск* и используется в качестве характеристики работы СЭР.

Как видно из рис. 9.14, период поисковых колебаний T выходного сигнала вдвое меньше периода колебаний входного воздействия T_x . Величина T называется *периодом рыскания*. Потеря на поиск за период

$$y_c - y_{\text{опт}} = \frac{1}{T} \int_{t_0}^T y dt - y_{\text{опт}}. \quad (9.7)$$

В СЭР, осуществляющих оптимизацию некоторой функции Q , зависящей от большого числа координат x_i , текущее значение потери на поиск можно определить с помощью степенного ряда.

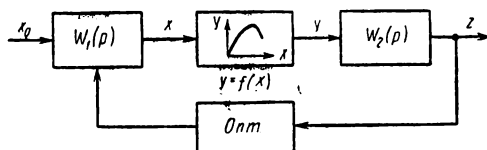


Рис. 9.15. Структурная схема инерционной СЭР

Изложенный выше материал относится к безынерционным системам без запаздывания, в которых реакция системы на входное воздействие является мгновенной. Однако все реальные устройства в той или иной степени обладают инерционностью, и для получения правиль-

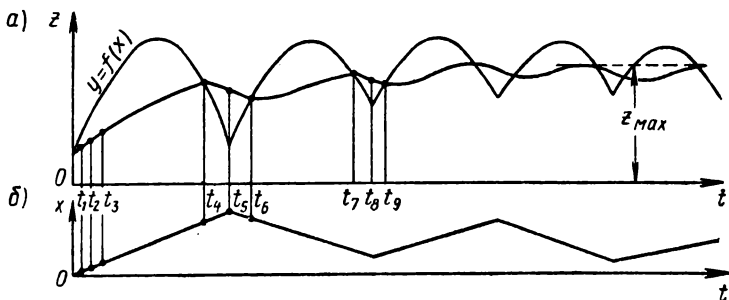


Рис. 9.16. Динамические процессы в инерционной СЭР

ных результатов при анализе и высоких показателей качества при синтезе таких СЭР необходимо учитывать их динамику.

Для удобства исследования динамики СЭР условно разбивают ее на несколько частей таким образом, чтобы все инерционные элементы были сосредоточены в выходной и входной линейных частях (рис. 9.15), а нелинейный блок с экстремальной характеристикой оказался безынерционным. В частных случаях в зависимости от свойств

исследуемой системы передаточная функция входной или выходной линейной части может оказаться равной единице. Наличие инерционных элементов в системе вызывает изменение выходного сигнала по форме (выходной сигнал сглаживается по сравнению, например, с рис. 9.14) и сдвиг по фазе его колебаний.

Рассмотрим методику исследования динамики СЭР в частном, но достаточно наглядном случае, когда $W_1(p) = 1$. При этом выходную линейную часть будем для определенности считать простым инерционным звеном первого порядка, а статическую характеристику $y = f(x)$ нелинейной части системы — параболой с экстремумом — максимумом. Таким образом

$$W_2(p) = \frac{1}{T_1 p + 1}. \quad (9.8)$$

Пользуясь методом, изложенным в § 8.1, напишем уравнение движения рассматриваемой системы в виде

$$T_1 \frac{dz}{dt} + z = y = f(x) \quad (9.9)$$

или

$$\frac{dz}{dt} = \frac{1}{T_1} [f(x) - z]. \quad (9.10)$$

При постоянной скорости изменения входного сигнала x

$$\frac{dx}{dt} = \pm a, \quad (9.11)$$

тогда (9.10) может быть преобразовано к виду

$$\frac{dz}{dx} = \pm \frac{1}{aT_1} [f(x) - z]. \quad (9.12)$$

Отсюда видно, что чем больше постоянная времени T_1 инерционной части системы, тем медленнее изменяется выходной сигнал z по отношению к входному воздействию x . В СЭР это обуславливает запаздывание выходного сигнала z .

Для большей простоты и наглядности анализ динамики СЭР часто осуществляется графическими методами. Пусть рассмотренная выше система, описываемая уравнениями (9.9) и (9.11), осуществляет оптимизацию параболической функции по способу с запоминанием экстремума [9.17]. Напомним, что в такой СЭР реверс исполни-

тельного механизма оптимизатора должен осуществляться в момент, когда разность между текущим и максимально достигнутым значениями выходного сигнала превысит зону нечувствительности. Для графического представления динамики работы инерционной СЭР может быть применен любой из численных методов решения дифференциальных уравнений, из которых наиболее простым является метод Эйлера. По этому методу плавная интегральная кривая $z(t)$ (рис. 9.16,а), представляющая собой

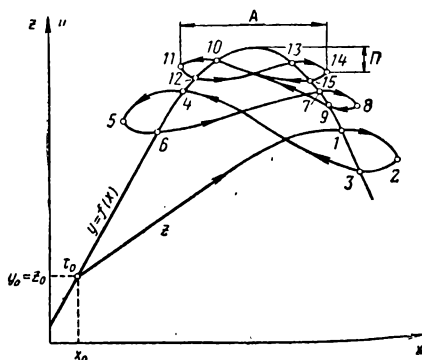


Рис. 9.17. Диаграмма работы СЭР в плоскости «вход — выход»:

A — амплитуда поисковых колебаний; Π — потеря на поиск

решение дифференциального уравнения, заменяется ломаной линией (ломаная Эйлера, рис. 9.16,б), отрезки которой имеют одинаковый наклон на участках от t_i до t_{i+1} , определяемый отношением

$$\frac{\Delta z_{i+1}}{\Delta t} = \varphi(z_i, t_i). \quad (9.13)$$

Здесь $\Delta z_{i+1} = z_{i+1} - z_i$, $\Delta t = t_{i+1} - t_i = \text{const}$.

Пусть СЭР включается в момент $t=0$, когда $z=z_0$ и $y=f(x_0)$ и оптимизатор увеличивает входной сигнал x , т. е. $x=x_0+at$. Разбивая ось абсцисс на ряд равных отрезков Δt , можно с помощью (9.10) найти приращение ломаной Эйлера на каждом из участков по формуле

$$\Delta z_i = \frac{\Delta t}{T_1} [f(x_i) - z_i], \quad (9.14)$$

которая характеризует динамический процесс поиска эк-

стремума. Из рис. 9.16 видно, что на участке от нуля до t_4 величина $z(t)$ увеличивается, поскольку здесь разность $f(x_{i+1}) - z_i > 0$. В момент времени t_4 кривая $z(t)$ совпадает со статической характеристикой $f(x_i) = z_4$ и $\frac{dz}{dt} = 0$. При дальнейшем увеличении x разность $f(x_i) - z_i$ становится отрицательной, т. е. производная $\frac{dz}{dt}$

меняет свой знак, а $z(t)$ начинает уменьшаться. При t_8 разность между текущим значением выходного сигнала z_8 и экстремальным z_4 превышает зону нечувствительности оптимизатора, исполнительный механизм которого реверсирует направление изменения входного воздействия.

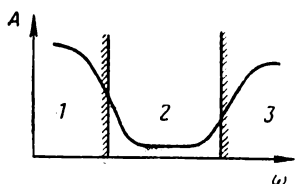


Рис. 9.18 Амплитудно-частотная характеристика помех:

1 — полоса дрейфа; 2 — диапазон поиска; 3 — высокочастотные помехи

Вследствие инерционности системы выходная координата еще некоторое время (до t_6) продолжает уменьшаться, а затем снова увеличивается. Процесс поиска повторяется до тех пор, пока в системе не установится режим периодических колебаний относительно среднего значения регулируемой величины, близкого к экстремальному (см. рис. 9.16).

Однако на практике для графического исследования процесса поиска экстремума инерционной СЭР чаще пользуются плоскостью с координатами «вход — выход», так как в этом случае кривая, характеризующая динамику СЭР, получается более наглядной и позволяет получить такую же информацию, как два графика во временной области.

Предположим, что поиск экстремума начат в момент времени t_0 при x_0 и $z_0 = y_0$, причем входное воздействие x изменяется в сторону увеличения (рис. 9.17). Правая часть уравнения (9.12) будет положительна, поэтому z возрастет до точки 1, где $\frac{dz}{dx} = 0$. При дальнейшем увеличении входного сигнала z становится меньше y , в связи с чем производная $\frac{dz}{dx}$ меняет знак на отрицательный. В точке 2 разность $z_2 - z_1$ превышает зону нечувствительности, исполнительный механизм оптимизатора

реверсирует направление изменения входного сигнала ($\frac{dx}{dt} < 0$). На участке от точки 2 до точки 3 сигнал z уменьшается в силу инерционных явлений, а затем снова возрастает до точки 4, после чего в точке 5 осуществляется новый реверс исполнительного механизма. Процесс продолжается до выхода системы в режим автоколебаний (рыскания) в области экстремума. На графике такой режим изображается замкнутой симметричной кривой 11—12—13—14—15—10—11, называемой *предельным циклом рыскания*. С помощью предельного цикла можно получить такие характеристики СЭР, как потеря на поиск Π , амплитуду поисковых колебаний A и т. д.

Исследование динамики СЭР может быть проведено и более точными аналитическими методами [9.9, 9.17], однако они являются и более трудоемкими.

Для получения правильных результатов при анализе и синтезе СЭР важное значение

имеет вопрос об устойчивости систем, т. е. об обеспечении нормального функционирования оптимальных систем при воздействии на них внешних и внутренних возмущений (изменение во времени экстремальной характеристики, влияние помех и др.). Реальные СЭР находятся обычно под воздействием монотонных возмущений, имеющих широкий частотный спектр (рис. 9.18). Низкочастотная часть спектра зависит от *дрейфа характеристик* элементов системы, приводящего к смещению положения экстремума, и отделена от области высокочастотных шумов и помех некоторой полосой частот, в которой амплитуда возмущений сравнительно мала. Выбирая частоту поисковых колебаний в пределах этой полосы частот, можно существенно уменьшить количество ложных срабатываний оптимизатора, являющихся следствием помех.

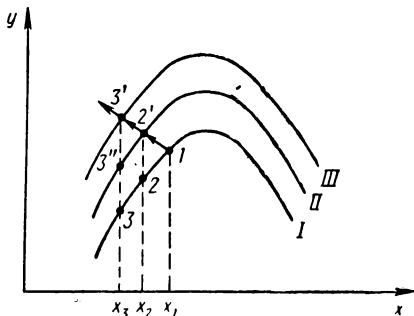


Рис. 9.19. К исследованию устойчивости работы СЭР

Однако даже и в этом случае при наличии дрейфа статической характеристики в СЭР возможна потеря устойчивости, в результате чего может произойти удаление СЭР от экстремума. Между тем по идее СЭР предназначены именно для оптимизации смещающихся непредвиденным образом функций, т. е. дрейф характеристики будет обычным явлением.

Пусть, например, возмущения вызывают смещение исходной статической характеристики I по вертикали в положения II и III (рис. 9.19) в моменты времени t_2 и t_3 . Если поиск начинается в точке I , причем исполнительный механизм будет сначала уменьшать управляющее

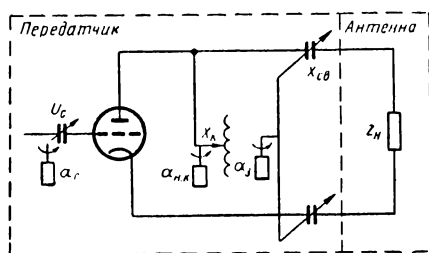


Рис. 9.20. Схема выходного каскада радиопередатчика

воздействие x , то при нормальном функционировании система должна прийти в точку 2 исходной характеристики I . При этом отрицательное приращение выходной координаты y послужило бы сигналом реверса исполнительного механизма, что обеспечивало бы движение к экстремуму. Но так как за время изменения x от x_1 до x_2 характеристика заняла положение II , то система придет в точку $2'$, которая соответствует положительному приращению y . Поскольку положительное приращение y в нормально функционирующей СЭР означает движение к экстремуму-максимуму, то направление изменения входного сигнала x сохраняется, а СЭР, которая из точки $2'$ должна была бы попасть в точку $3''$, при продолжающемся дрейфе придет в точку $3'$, еще больше удаленную от экстремума. Таким образом, исследуемая СЭР теряет устойчивость.

Для сохранения устойчивости СЭР может быть применен коммутатор, который через равные промежутки вре-

мени осуществляет принудительный реверс исполнительного механизма, делая движение системы к экстремуму немонотонным. При возникновении интенсивного дрейфа коммутатор удерживает величину управляющего сигнала в окрестностях значения, при котором начался дрейф, сохраняя устойчивость СЭР.

Примеры. 1. Для обеспечения достоверности получаемой информации в месте приема должно сохраняться максимальное отношение сигнал/шум, что имеет место при излучении передатчиком максимально допустимой мощности. На рис. 9.20 представлена электрическая схема выходного каскада радиопередатчика, в котором с помощью СЭР достигается и поддерживается на протяжении всей работы режим максимальной излучаемой мощности Q . Среди большого числа параметров x_k , $x_{св}$ и т. д., от которых зависит излучаемая мощность, следующие три выделены в качестве настраиваемых параметров: α_c — параметр, характеризующий амплитуду возбуждающего напряжения; $\alpha_{з.к}$ — параметр, определяющий загрузку контура; $\alpha_{н.к}$ — параметр, характеризующий расстройку контура с изменением загрузки. Для решения рассматриваемой задачи передатчик снабжен системой настройки контура в резонанс с задающей частотой возбуждателя и системой, обеспечивающей необходимую загрузку передатчика. Контроль за качеством настройки осуществляется по величине катодного тока лампы выходного каскада.

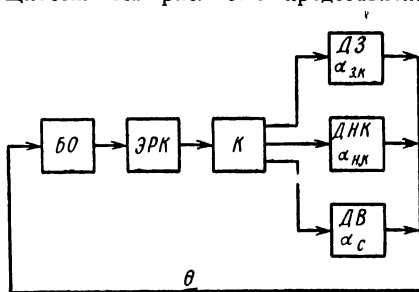


Рис. 9.21. Система оптимизации режима работы радиопередатчика:

БО — блок оптимизации; ЭРК — элемент разрешения коммутации; К — коммутатор; ДЗ — двигатель загрузки; ДНК — двигатель настройки контура; ДВ — двигатель возбуждения

Функциональная схема такого передатчика приведена на рис. 9.21. В основе работы СЭР лежит определение частных экстремумов оптимизируемой функции по каждой управляемой переменной α_i . Для определения этих частных экстремумов служит блок оптимизации БО. Сигналы, вырабатываемые БО, поступают на коммутатор К, осуществляющий переключение управления с одного исполнительного механизма на другой в зависимости от того, по какому параметру α_i отыскивается частный экстремум. В качестве исполнительных механизмов применяются двигатели, осуществляющие изменение параметров α_i . Элемент разрешения коммутации ЭРК посылает на К сигнал, разрешающий коммутацию, только в том случае, если последний шаг по предыдущей переменной был сделан в правильном направлении [9.16].

2. На рис. 9.22 приведена принципиальная схема электронного экстремального регулятора ЭЭР-1 с запоминанием экстремума. Ха-

ракторными особенностями такого регулятора являются наличие коммутатора поворочных реверсов и устройства самовыключения, прекращающего воздействие исполнительного механизма после достижения точки экстремума.

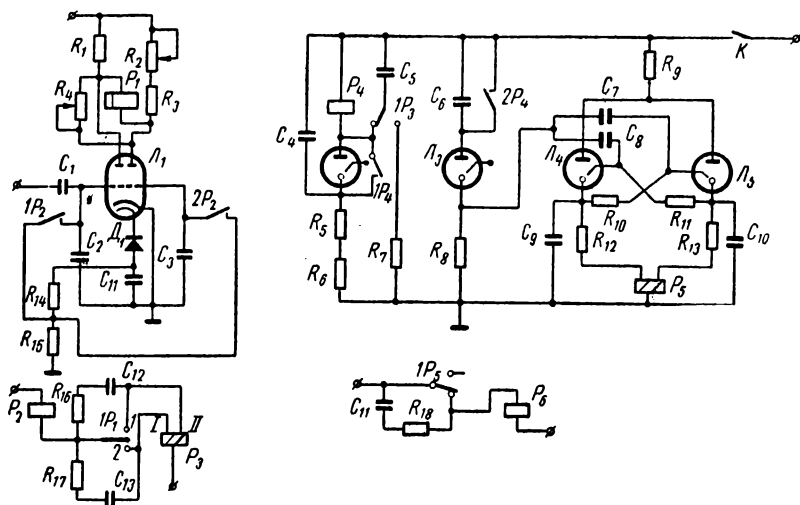


Рис. 9.22. Принципиальная схема электронного экстремального регулятора ЭЭР-1

Основной частью экстремального регулятора является *сигнум-реле*. Это логическое устройство, предназначенное для определения направления движения к экстремуму и обеспечения соответствующих изменений управляющих воздействий. В рассматриваемой схеме

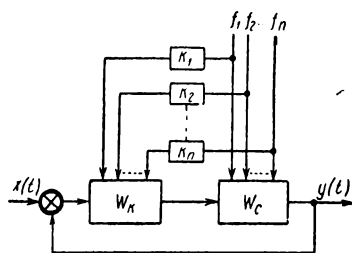


Рис. 9.23. СНС с разомкнутым циклом настройки

сигнум-реле, осуществляющее реверс направления изменения сигнала x , реализовано с помощью параллельно-балансного усилителя постоянного тока, собранного на двойном триоде L_1 (6Н15П). Оба триода лампы L_1 работают в одинаковых режимах, подгонка режимов осуществляется с помощью сопротивления R_2 . Между анодами лампы L_1 включена обмотка поляризованного реле P_1 , срабатывающего при возникновении достаточной по величине разности потенциалов между анодами. В зависимости от

знака приращения напряжения на сетке первого триода лампы L_1 анодный ток этого триода будет увеличиваться или уменьшаться,

изменяя тем самым анодное напряжение. В зависимости от полярности получаемой разности потенциалов перекидной контакт $1P$ реле P_1 будет занимать положение 1 или 2, подавая тем самым питание на катушку реле P_2 и одну из катушек реле P_3 . При срабатывании реле P_2 контакты $1P_2$ и $2P_2$ замыкаются, конденсаторы C_2 и C_3 разряжаются на землю, а запоминающая емкость C_1 заряжается до текущего значения напряжения источника входного сигнала. Такое положение соответствует исходному состоянию сигнум-реле, подготовленного к очередному шагу. При изменении входного напряжения через конденсатор C_2 потечет ток, потенциал сетки будет изменяться в соответствии с зарядом конденсатора C_2 ; по достижении зоны нечувствительности реле P_1 , а следовательно, и P_2 , P_3 сработают, возвращая сигнум-реле в исходное состояние. При положительном приращении входного напряжения, соответствующем движению системы в желаемом направлении, включается обмотка I реле P_3 , а при отрицательном — обмотка II. В первом случае контакты реле P_3 включают на заряд емкость C_5 , во втором же, т. е. при удалении системы от экстремума, реле P_3 подключает конденсатор C_5 к катушке реле P_4 , включенной в цепь разряда тиратрона L_3 , что в конечном итоге приводит к переключению выходного реле P_6 . Последнее предназначено для управления изменением входной координаты.

Коммутатор поперечных реверсов состоит из задающего генератора на тиратроне L_2 , промежуточного каскада (тиратрон L_3), релаксационного реле, собранного на тиратронах L_4 и L_5 . Когда заряд емкости C_4 достигает потенциала зажигания, тиратрон L_2 зажигается, включенное в цепь разряда тиратрона L_3 реле P_4 срабатывает, вызывая зажигание тиратрона L_3 и разряд конденсатора C_4 . С сопротивления R_8 потенциал подается на поджигающие электроды тиратронов L_4 и L_5 . В цепи релаксационного реле включены обмотки поляризованного реле P_5 , контакты которого включают или выключают катушку выходного реле P_6 . Частота зажигания тиратрона L_2 задающего генератора регулируется с помощью сопротивлений R_8 и R_9 [9.3].

§ 9.2. САМОНАСТРАИВАЮЩИЕСЯ СИСТЕМЫ

Интенсивное развитие, которое получили самонастраивающиеся системы (СНС, см. § 2.2) за последние годы привело к появлению большого числа их разновидностей. По принципу действия самонастраивающиеся системы можно разбить на следующие основные виды:

- 1) с разомкнутым циклом настройки;
- 2) с замкнутым циклом настройки;
- 3) аналитические.

В зависимости от общего назначения системы управления и условий ее работы контур самонастройки может быть предназначен для настройки корректирующих устройств, настройки отдельных параметров системы (напри-

мер, коэффициента усиления) или совокупности параметров, для оптимизации характеристик системы и т. д.

Системы с разомкнутым циклом настройки. Такие устройства являются довольно простыми и применяются на практике сравнительно давно. Структурная схема такой системы представлена на рис. 9.23 [9.41]. Система с разомкнутым циклом настройки состоит из объекта управления с передаточной функцией W_c , находящегося под воздействием внешних возмущений f_i , и настраиваемого корректирующего устройства W_k , предназначенного для компенсации влияния внешних условий на свойства объекта. При разомкнутом цикле самонастройки такая

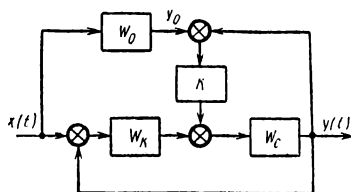


Рис 9.24. СНС с эталонным фильтром

компенсация возможна только при наличии достаточно полной текущей информации о характере возмущающего влияния внешних воздействий, поэтому в систему должны быть введены измерительные устройства с передаточными числами k_1, k_2, \dots, k_n .

В процессе работы под влиянием внешних воздействий свойства объекта изменяются, в связи с чем передаточная функция W_c будет отличаться от передаточной функции объекта $W_{c.0}$ в исходном состоянии. Следовательно, и передаточная функция корректирующего устройства W_k должна изменяться по отношению к ее первоначальному виду $W_{k.0}$ таким образом, чтобы общие свойства системы, определяемые ее передаточной функцией W в замкнутом состоянии, оставались неизменными, т. е.

$$W = \frac{W_k W_c}{1 + W_k W_c} = \text{const.} \quad (9.15)$$

Вполне очевидно, что (9.15) будет иметь место лишь при

$$W_k W_c = W_{k.0} W_{c.0} = \text{const},$$

откуда следует, что

$$W_k = \frac{W_{k.0} W_{c.0}}{W_c}. \quad (9.16)$$

В реальных устройствах обычно ограничиваются лишь приближенным воспроизведением условия (9.16), поскольку реализация корректирующих устройств, с высокой точ-

ностью удовлетворяющих этим условиям, а также получение полной информации о возмущающих воздействиях являются технически трудно выполнимой задачей.

Примером самонастраивающихся систем с разомкнутым циклом могут служить некоторые схемы автопилотов. В этом случае внешними условиями, влияющими на параметры летательного аппарата, являются давление и температура окружающего воздуха, скорость полета. В процессе полета, в зависимости от текущих значений внешних возмущений, контур самонастройки осуществляет целенаправленное изменение параметров корректирующего устройства для выполнения условия (9.15).

Недостаток систем с разомкнутым циклом самонастройки заключается в отсутствии самоанализа результатов осуществляемых ими изменений.

Системы с замкнутым циклом настройки [9.15]. Такие системы характеризуются наличием дополнительного замкнутого контура, позволяющего осуществлять анализ текущего состояния управляемого объекта, изменять параметры корректирующих элементов системы и оценивать эффективность предпринятых действий. Такой вид СНС является более универсальным и обладает большими возможностями, вследствие чего в последнее время получил преимущественное развитие.

Один из возможных способов построения *СНС с замкнутым циклом* самонастройки заключается в применении эталонного фильтра, передаточная функция которого соответствует передаточной функции W_0 оптимальной замкнутой системы¹. Действие таких систем основано на сравнении выходного сигнала объекта управления с желаемым выходным сигналом, вырабатываемым эталонным фильтром, и использовании разностного сигнала для компенсации отклонений динамических характеристик системы от заданных.

На рис. 9.24 изображена структурная схема такой системы. Здесь W_k и W_c — передаточные функции звеньев системы управления, W_0 — эталонный фильтр.

Разностный сигнал $y_0 - y$ с элемента сравнения поступает на усилитель с коэффициентом усиления K , после чего приходит на вход второго звена.

¹ Такие устройства называются системами с глубокой обратной связью.

Передаточная функция замкнутой системы по рис. 9.24

$$W = \frac{W_k W_c + K W_c W_0}{1 + W_k W_c + K W_c} \quad (9.17)$$

При достаточно больших значениях коэффициента усиления K имеет место приближенная зависимость $W \approx W_0$, означающая, что при изменении параметров объекта под влиянием внешних возмущений общая передаточная функция W замкнутой системы остается близкой к оптимальной передаточной функции W_0 . Такая система является в сущности инвариантной по отношению к изменениям ее параметров. Возможности таких систем в достижении высоких качественных показателей ограничиваются опасностью потери устойчивости при увеличении коэффициента усиления K .

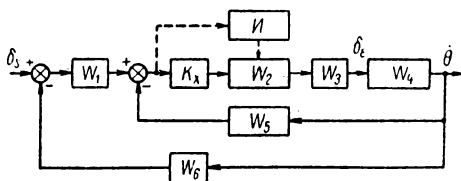


Рис. 9.25. Система управления самолетом по тангажу:

δ_s — перемещение органа управления; δ_e — отклонение руля высоты; $\dot{\theta}$ — скорость в плоскости тангажа; K_x — усилитель с переменным коэффициентом усиления

Принцип действия и основные особенности СНС с замкнутым циклом самонастройки во многом зависят от выбора управляемых характеристик объекта и цели самонастройки. Целью самонастройки является либо сохранение постоянства динамических характеристик объекта управления, свойства которого изменяются во времени, либо оптимизация системы по принятому критерию оптимальности за счет автоматической подстройки ее параметров. Для сохранения динамических свойств системы при изменяющихся условиях работы чаще всего применяется стабилизация расположения доминирующих нулей и полюсов системы на комплексной плоскости и стабилизация частотных и переходных характеристик.

Метод корневых годографов (см. § 7.6) оказался весьма полезным и удобным инструментом анализа и синтеза самонастраивающихся систем, благодаря чему был создан

целый ряд весьма ответственных и высококачественных СНС, применяемых, в частности, для управления полетом высокоскоростных объектов. Примером таких СНС может служить автопилот самолета типа В-25, структурная схема которого показана на рис. 9.25 [9.11]. В качестве управляемой переменной здесь принята угловая скорость $\dot{\theta}$ тангажа самолета, а в усилителе прямой системы предусматривается возможность изменения коэффициента усиления K_x в широких пределах.

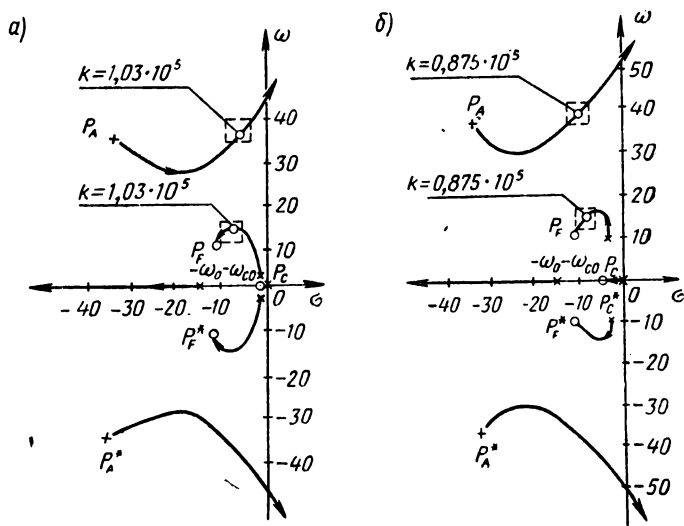


Рис. 9.26. Корневые годографы системы, изображенной на рис. 9.25, для крайних режимов полета

На рис. 9.26, а и б представлены корневые годографы исследуемой системы для двух крайних режимов полета, где символами P и ω обозначены полюса и нули передаточной функции разомкнутой системы, причем пунктирными квадратами выделены области, в которых доминирующие полюса и нули P_A и P_F могут перемещаться без существенного изменения свойств системы. Анализ показывает, что изменением коэффициента усиления можно при изменяющихся условиях полета сохранять расположение полюсов внутри пунктирных квадратов.

Отклонение параметров самолета от первоначальных значений вызывает изменения в расположении нулей и

полюсов системы, вследствие чего изменяется и собственная частота колебаний. Измеряя собственную частоту, можно по ее величине судить о расположении нулей и полюсов, т. е. сигнал на выходе устройства, измеряющего собственную частоту, может быть использован для управления потенциометром настройки коэффициента усиления. На рис. 9.25 пунктирной линией показана схема включения такого измерительного устройства *И* в систему управления, а на рис. 9.27 изображена схема этого устройства.

Принцип действия его заключается в том, что, например, при уменьшении собственной частоты выходной сиг-

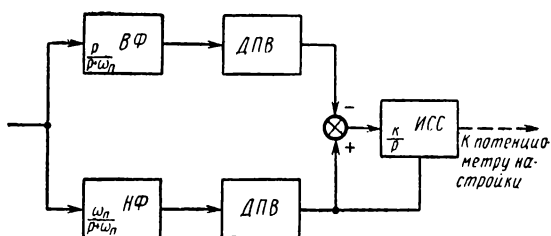


Рис. 9.27. Устройство для измерения собственной частоты колебаний системы:

ДПВ — двухполупериодный выпрямитель

нал низкочастотного фильтра *НФ* превышает выходную величину высокочастотного фильтра *ВФ* и двигатель интегрирующей следящей системы *ИСС* перемещает потенциометр настройки K_x в сторону увеличения. При повышении собственной частоты сигнал высокочастотного фильтра будет превышать сигнал низкочастотного и движок потенциометра будет перемещаться в противоположном направлении.

При разработке СНС с переменным коэффициентом усиления необходимо иметь в виду то обстоятельство, что чрезмерное увеличение коэффициента усиления в линейных системах может привести к потере устойчивости. Поэтому в системы такого типа обычно преднамеренно вводятся нелинейные звенья с ограничением или применяются специальные релейные схемы.

Во многих случаях объект управления бывает достаточно подробно изучен, что позволяет составить его мате-

матическое описание или построить физическую модель. Применение модели дает возможность построить СНС с большей точностью работы, скоростью выхода на оптимальный режим, позволяет иногда устранить поисковые колебания.

Методы построения СНС с моделью получили широкое развитие. Принцип построения таких систем и особенности их работы иллюстрируются далее конкретными схемами.

Среди разрабатываемых СНС с замкнутым циклом преобладают *поисковые системы* как экстремальные, так и неэкстремальные, которые производят настройку корректирующих элементов на основании анализа реакции системы на поисковые воздействия.

Поисковые системы с экстремальной самонастройкой являются одними из наиболее совершенных, но отличаются значительной сложностью; их реализация сталкивается с техническими трудностями, особенно при большом числе регулируемых параметров. Практически поисковые экстремальные системы представляют собой сочетание оптимизатора (обычно многоканального) одного из рассмотренных в § 9.1 видов с корректирующим устройством, осуществляющим за счет изменения параметров системы компенсацию влияния внешних и внутренних возмущений на свойства объекта. Таким образом, поисковые СНС содержат в качестве основных узлов уже известные устройства и не требуют специального изучения.

В задачах управления часто встречаются случаи, когда исключается возможность применения поисковых колебаний вследствие снижения качества конечного продукта, повышенного износа подвижных частей объекта управления и т. п. В таких случаях необходимо либо применять специальные беспоисковые способы самонастройки, либо осуществлять самонастройку с помощью моделей или вычислительных устройств.

Одним из удобных и достаточно простых методов построения беспоисковых СНС является оптимизация режима работы управляемого объекта с помощью так называемого дифференциального принципа или *дифференциальных СНС* [9.19]. Дифференциальными такие системы названы в силу того, что закон регулирования формируется в них с помощью анализа разности двух величин. Для получения такой разности статическая экстремаль-

ная характеристика сравнивается, как описано ниже, либо с аналогичной смещенной характеристикой, либо с деформированной характеристикой.

Дифференциальные СНС со смещенной экстремальной характеристикой применяются в тех случаях, когда объекты регулирования сравнительно просто поддаются

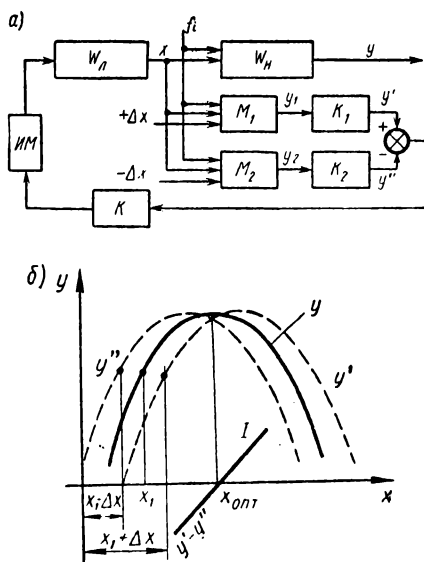


Рис. 9.28. Дифференциальная самонастраивающаяся система:

W_n , $W_н$ — линейная и нелинейная части системы;

K , K_1 , K_2 — усилители

моделированию. Обычно в таких системах применяют две модели, хотя работоспособность системы сохраняется и при действии одной модели, но с понижением точности настройки на экстремум.

На вход моделей M_1 и M_2 (рис. 9.28, а), помимо сигнала x , непрерывно подаются одинаковые по величине, но противоположные по знаку приращения $Δx$, в результате чего статические экстремальные характеристики моделей смещаются влево или вправо по отношению к статической характеристике регулируемого объекта

(рис. 9.28, б). Принцип действия такой системы заключается в следующем.

Из рис. 9.28, б видно, что экстремум статической характеристики $y=f(x)$ имеет место при $y'=y''$. Поэтому измеренные специальными датчиками, установленными на каждой модели, значения y' и y'' подаются на устройство вычитания. Величина $y'-y''$ будет изменяться по кривой I , причем при $y'-y''=0$ управляющее воздействие равняется $x_{\text{опт}}$, соответствующему максимуму статической характеристики. Следовательно, конечная цель работы контура самонастройки дифференциальной системы сводится к изменению с помощью исполнительного механизма $ИМ$ корректирующих параметров системы таким образом, чтобы поддерживалось условие $y'-y''=0$.

К достоинствам таких дифференциальных СНС следует отнести также инвариантность по отношению к возмущающим воздействиям f_i и к систематической ошибке измерения критерия оптимальности y_i .

Дифференциальные системы с деформацией статической характеристики обладают тем достоинством, что они не требуют применения моделей, но по своим качественным показателям менее совершенны, чем дифференциальные СНС со смещением характеристик. Деформация характеристики может быть достигнута за счет применения дополнительного усилителя, коэффициент усиления которого отличается от коэффициента усиления основного канала. В таких системах рабочая точка удерживается на некотором удалении от экстремума.

Аналитические самонастраивающиеся системы [9.15]. Контроль за изменениями динамических свойств объекта управления, возмущающих воздействий и других переменных факторов, а также определение характера последующих действий контура самонастройки с большим успехом осуществляется с помощью вычислительных машин. Вычислительные устройства аналитически решают задачу оптимизации исследуемой системы и вырабатывают управляющие сигналы для контура самонастройки. Такие системы стали называть *аналитическими СНС*.

Благодаря большому возможностям в улучшении качества работы таких систем они в последнее время интенсивно развиваются, в связи с чем появилось значительное количество их разновидностей. Рассмотрим наиболее характерные способы построения аналитических СНС.

СНС с подстройкой ко входному сигналу. Функциональная схема такой системы представлена на рис. 9.29, где O — объект, $УУ$ — управляющее устройство. Предположим, что эта система находится под воздействием входного сигнала, являющегося суммой двух составляющих: полезной составляющей $x(t)$ и случайной помехи $n(t)$, вероятностные характеристики которой известны. Вычислительное устройство $ВУ_{опт}$ служит для определения значения оптимальной импульсной переходной характеристики $\omega(t)_{опт}$. Текущее значение импульсной переходной функции $\omega(t)$, вырабатываемое вычислительным устройством $ВУ_c$, сравнивается с $\omega(t)_{опт}$, а результирующий сигнал используется

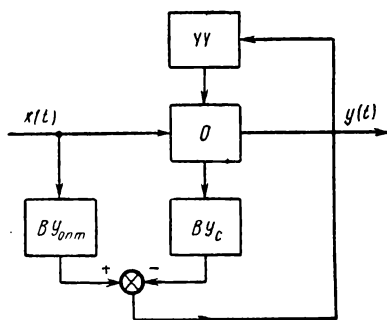


Рис. 9.29. Схема СНС с подстройкой ко входному сигналу

сигнал используется управляющим устройством $УУ$ для корректировки параметров системы с целью приблизить импульсную переходную характеристику системы к оптимальной.

С течением времени динамическая точность системы будет возрастать, приближаясь к оптимальным характеристикам.

Системы с определением динамических характеристик. Принцип действия таких систем заключается в определении с помощью вычислительных устройств тех или иных характеристик и сравнении их с эталонными или оптимальными характеристиками. На основании этого сравнения осуществляется подстройка параметров системы, целью которой является максимальное приближение динамических свойств системы к желаемым.

Оценка динамических свойств системы может быть сделана по ее частотным, переходным, статистическим и другим характеристикам. Для этого система должна содержать специальные устройства контроля характеристик, которые чаще всего строятся на основе вычислительных устройств. В подавляющем большинстве случаев автоматическое измерение характеристик системы осуществляется

с помощью искусственных поисковых составляющих выходного сигнала или естественных колебаний его, вызываемых случайными помехами на входе. Например, для определения частотных характеристик обычно используют естественные или искусственные гармонические составляющие выходного сигнала, импульсные характеристики определяют с помощью импульсного пробного сигнала.

На рис. 9.30 представлена функциональная схема аналитической СНС с определением частотных характеристик. Система строится в предположении, что изменения па-

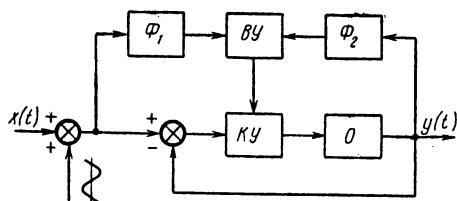


Рис. 9.30. СНС с определением частотных характеристик, управляющая объектом O

раметров под влиянием внешних возмущений являются медленными, вследствие чего коэффициенты частотной характеристики системы можно считать постоянными.

Частотная характеристика системы определяется ее вещественной и мнимой частями $U(\omega)$ и $V(\omega)$ (см. § 4.2). При наличии гармонической составляющей сигнала $a \sin \gamma t$ на входе системы ее выходной сигнал будет содержать компоненты $aU(\gamma) \sin \gamma t$, совпадающую по фазе с входным сигналом, и $aV(\gamma) \cos \gamma t$, фаза которой сдвинута на 90° . В соответствии с этим в состав схемы СНС входят узкополосные фильтры Φ_1 и Φ_2 , настроенные на частоту γ и предназначенные для выделения гармонических составляющих входного и выходного сигналов. Выделенные сигналы поступают на вычислительное устройство $ВУ$, определяющее соотношение их амплитуд и сопоставляющее фазы. Сигнал от $ВУ$ поступает на корректирующее устройство $КУ$.

В качестве фильтров Φ_1 и Φ_2 целесообразно применить синхронные детекторы с гармоническим опорным напряжением частоты γ , которые содержат множительные

ной, а отличаются они, в основном, способом измерения действительной импульсной характеристики.

Один из возможных [9.12] способов автоматического определения импульсной характеристики системы основан на применении принципов статистической корреляции. Он состоит в анализе взаимной корреляционной функции входного $x(t)$ и выходного $y(t)$ сигналов с помощью вычислительного и измерительного устройств ВУ и ИУ, когда входной сигнал представляет собой белый шум малой амплитуды, полученный от генератора белого шума ГБШ (рис. 9.32).

Если полоса частот белого шума на входе системы значительно превышает диапазон частот управляемой системы, то автокорреляционная функция является δ -функцией, а значит, взаимная корреляционная функция представляет собой импульсную характеристику системы. Необходимое множество таких точек может быть получено введением дополнительных каналов, соответствующих различным τ .

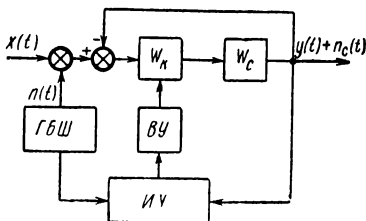


Рис. 9.32. СНС с определением импульсной характеристики

Оценку импульсной характеристики чаще всего производят на основании какого-либо критерия качества. Когда текущее значение критерия качества отклоняется от заданного, вычислительное устройство ВУ обеспечивает изменение параметров корректирующего устройства в требуемом направлении.

Системы, построенные на основе методов статистической динамики, обладают хорошими характеристиками, однако их применение затрудняется в тех случаях, когда время отклонения параметров системы слишком мало по сравнению с временем действия корреляторов, а также когда нежелательно присутствие широкополосного шума на выходе системы.

Аналитические СНС с экстремальной подстройкой параметров. При создании таких систем могут быть применены рассмотренные ранее методы оптимизации по принятому критерию оптимальности, т. е. способы поиска экстремума некоторой функции качества

Q за счет изменения переменных параметров системы. При этом в качестве составной части контура самонастройки могут быть применены как известные схемы оптимизаторов, так и вычислительные устройства. Среди экстремальных СНС большой интерес представляет система с подстройкой по методу градиента [9.8] и с вычислением компонент градиента с помощью вспомогательного оператора, которая достаточно полно характеризует основные особенности аналитических экстремальных СНС.

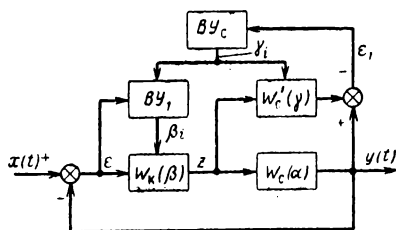


Рис. 9.33. Схема аналитической экстремальной СНС с подстройкой по методу градиента

Схема такой системы

представлена на рис. 9.33. Функция качества здесь выбрана зависящей от ошибки ϵ системы $Q = Nf(\epsilon)$. Задача контура самонастройки заключается в определении скорости изменения каждого из настраиваемых параметров β_i системы таким образом, чтобы эта скорость была пропорциональна соответствующим компонентам мгновенного вектора градиента функции качества

$$\dot{\beta}_i = a \frac{\partial Q}{\partial \beta_i}. \quad (9.18)$$

Дифференцируя выражение для функции качества, можно определить составляющие вектора градиента

$$-\frac{\partial Q}{\partial \beta_i} = N \frac{\partial f(\epsilon)}{\partial \epsilon} \cdot \frac{\partial \epsilon}{\partial \beta_i}. \quad (9.19)$$

Дифференцирование системы уравнений

$$y = W_c(\alpha)z; \quad z = W_k(\beta)\epsilon; \quad \epsilon = x - y \quad (9.20)$$

дает возможность определить $\partial \epsilon / \partial \beta_i$. Тогда из (9.20) получаем:

$$\frac{\partial y}{\partial \beta_i} = W_c(\alpha) \frac{\partial W_k(\beta)}{\partial \beta_i} \epsilon + W_c(\alpha) W_k(\beta) \frac{\partial \epsilon}{\partial \beta_i}. \quad (9.21)$$

Управляющее воздействие x не зависит от параметров системы β_i , поэтому из третьего равенства системы (9.20)

имеем:

$$\frac{\partial \varepsilon}{\partial \beta_i} = - \frac{\partial y}{\partial \beta_i}.$$

После простейших преобразований из (9.21) находим:

$$\frac{\partial \varepsilon}{\partial \beta_i} = - \frac{W_c(\alpha)}{1 + W_c(\alpha) W_k(\beta)} \cdot \frac{\partial W_k(\beta)}{\partial \beta_i} \varepsilon$$

или

$$\frac{\partial \varepsilon}{\partial \beta_i} = - W(\alpha, \beta) \varepsilon, \quad (9.22)$$

где

$$W(\alpha, \beta) = \frac{W_c(\alpha)}{1 + W_c(\alpha) W_k(\beta)} \cdot \frac{\partial W_k(\beta)}{\partial \beta_i}.$$

Вспомогательная модель $W_c(\gamma)$ функционирует аналогично контуру самонастройки самой системы. Реакция модели сравнивается с реакцией объекта, а сигнал рассогласования используется для настройки параметров модели таким образом, чтобы свести рассогласование к минимуму, т. е. чтобы поведение модели соответствовало протеканию процесса в объекте.

Поскольку параметры модели отражают параметры объекта управления, они могут быть использованы вычислительным устройством ВУ для определения параметров корректирующего устройства.

Преимущества рассмотренного метода заключаются в отсутствии пробных возмущений и в возможности ускорения и упрощения процесса вычисления компонент градиента с помощью сравнительно простого вспомогательного оператора.

§ 9.3. САМООРГАНИЗУЮЩИЕСЯ СИСТЕМЫ

Значение самоорганизующихся систем (см. § 2.2) трудно переоценить. В силу самого принципа действия такие системы обладают чрезвычайно высокой надежностью, гибкостью, способностью автоматически, без вмешательства человека исправлять случайные ошибки, допущенные при ее синтезе, осуществлять работу при минимальной априорной информации, т. е. в условиях начальной неопределенности.

Одна из характерных особенностей самоорганизующихся систем заключается в том, что, как уже было отмечено в § 2.2, они воспроизводят отдельные черты поведения живых организмов. В связи с этим при описании поведения и свойств технических самоорганизующихся систем весьма широко употребляются термины и понятия, заимствованные из биологии и физиологии.

Разработка принципов построения и практическая реализация самоорганизующихся систем представляют собой весьма сложную задачу, в связи с чем в настоящее время такие системы пока не получили широкого распространения среди технических систем автоматического управления. Однако самоорганизующиеся системы уже применяются в качестве моделей при изучении сложных биологических, экономических и других объектов, а их очевидные преимущества требуют настойчивой и квалифицированной работы с целью быстрее их развития.

Биологические самоорганизующиеся системы широко представлены в живых организмах многочисленными процессами регуляции, поддерживающими отдельные существенные характеристики организма (температуру тела, химический состав и давление крови и др.) в заданных, обычно довольно узких пределах. Это необходимо для обеспечения жизнеспособности организма при изменении внешних условий. Такое свойство живого организма получило название *гомеостазиса*, а соответствующие системы принято называть *гомеостатическими*. Их нормальное функционирование осуществляется с помощью отрицательных и положительных обратных связей.

У. Р. Эшби построил техническое устройство, обладающее свойством гомеостазиса и названное поэтому гомеостатом. Гомеостат Эшби вследствие своей наглядности, универсальности и общности принципа действия стал почти классическим примером самоорганизующейся системы, поэтому дальнейшее изучение таких систем будем проводить преимущественно на этом примере.

Гомеостат [9.6] представлял собой совокупность четырех одинаковых блоков, каждый из которых содержал подвижный электромагнит, поворачивающийся под воздействием токов, протекавших по его четырем обмоткам. Угол поворота каждого из магнитов от некоторого нулевого положения служил выходной координатой x_i соответствующего блока. На магнитах жестко закреплялись ука-

зательные стрелки, которые одновременно являлись движками потенциометров, что позволяло достаточно просто преобразовать механическое перемещение магнитов в электрический сигнал. Каждый из полученных электрических сигналов, пропорциональных угловым перемещениям магнитов, передавался на обмотки остальных трех магнитов через потенциометры с переключателями так, что передаваемый сигнал мог изменяться по величине и знаку. Кроме того, каждый блок имел цепь собственной обратной связи, коэффициент a_{ii} которой также мог меняться по величине и знаку.

Таким образом, на вход каждого из блоков (в обмотки электромагнита) подавались преобразованные выходные сигналы всех блоков, а угол поворота магнита определялся суммарным влиянием всех токов, протекающих в его обмотках. Другими словами, гомеостат характеризовался наличием всех возможных связей между блоками (рис. 9.34).

Поведение такой системы можно представить уравнениями

$$\left. \begin{aligned} \frac{dx_1}{dt} &= a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + a_{14}x_4, \\ \frac{dx_2}{dt} &= a_{21}x_1 + a_{22}x_2 + a_{23}x_3 + a_{24}x_4, \\ \frac{dx_3}{dt} &= a_{31}x_1 + a_{32}x_2 + a_{33}x_3 + a_{34}x_4, \\ \frac{dx_4}{dt} &= a_{41}x_1 + a_{42}x_2 + a_{43}x_3 + a_{44}x_4. \end{aligned} \right\} \quad (9.23)$$

Здесь коэффициенты a_{ij} определяют долю выходного сигнала j -го блока, передаваемого на вход i -го блока, причем $-1 \leq a_{ij} \leq 1$.

Система (9.23) может быть представлена в матричной форме

$$\dot{x}_i = Ax, \quad (9.24)$$

где

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{pmatrix}.$$

Цель работы гомеостата, который являлся демонстрационным прибором, заключалась в автоматическом поиске

устойчивого состояния, т. е. в обеспечении работоспособности системы с неопределенной исходной структурой за счет изменения коэффициентов a_{ij} . Однако даже такое сравнительно простое устройство обладало целым рядом принципиальных особенностей, качественно отличавших эту систему от существовавших ранее докибернетических систем.

Среди таких особенностей в первую очередь следует отметить избыточность структуры гомеостата. Исключительно высокая надежность и высокая способность к адап-

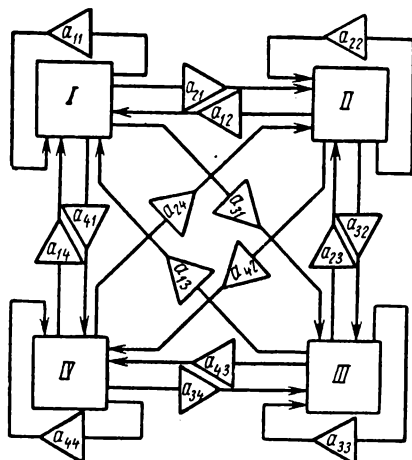


Рис. 9.34. Схема гомеостата Эшби

тации живых организмов главным образом обеспечиваются за счет избыточности в нервных и мышечных тканях. Избыточность в гомеостате достигается за счет того, что каждый из настраиваемых коэффициентов a_{ij} может принимать несколько значений, а общее количество всех возможных сочетаний таких значений коэффициентов составляет 390625.

Другая важная особенность гомеостата состоит в *статистическом характере поиска* устойчивого состояния. Скачкообразное изменение автоматически настраиваемых коэффициентов a_{ij} производилось по случайному закону до тех пор, пока случайным образом не было найдено удовлетворительное решение: сочетание их, обеспечивавшее устойчивость исследуемой системы. Применение вероятностного поиска придает кибернетическому устройству важ-

нейшее свойство — способность целенаправленно добиваться нужного результата при отсутствии априорной информации о свойствах системы и условиях ее работы. Это позволяет говорить о том, что гомеостатические системы действуют непредсказуемым образом, что порядок их работы заранее не определен конструктором. Оставляя в стороне вопрос о спорности этого утверждения (некоторые авторы, например Поль Косса [9.2], утверждают, что и в этом случае поведение машины целиком предопределено конструктором), отметим лишь универсальность свойств таких систем, позволяющих им решать поставленную задачу на основе лишь текущей информации о процессах управления. Это обстоятельство также роднит гомеостатические системы с высокоорганизованными живыми организмами.

Таким образом, основная качественная особенность гомеостата заключается в том, что он способен целеустремленно действовать в непредсказуемых условиях, автоматически добиваясь поставленной цели и не успокаиваясь вплоть до ее достижения. Поскольку целью гомеостата являлось отыскание устойчивых состояний, то указанное свойство целенаправленного поведения системы Эшби называл *ультраустойчивостью*. Это понятие является одним из важнейших для адаптивных систем. Термин «ультраустойчивость» является, очевидно, слишком узким для характеристики указанных свойств адаптивных систем, так как задачей подавляющего большинства таких систем является не просто сохранение устойчивости, а достижение заданных областей или значений самых различных критериев качества. Поэтому правильнее было бы применять более общий термин, однако в силу традиций в дальнейшем будем пользоваться словом «ультраустойчивость», вкладывая в него в необходимых случаях более широкий смысл.

Рассмотрим более подробно понятие ультраустойчивости. Для удобства и наглядности описания поведения гомеостатической системы воспользуемся методом фазового пространства (см. § 8.2) в применении к гомеостатической системе второго порядка, уравнения движения которой имеют вид

$$\left. \begin{aligned} \frac{dx_1}{dt} &= a_{11}x_1 + a_{12}x_2, \\ \frac{dx_2}{dt} &= a_{21}x_1 + a_{22}x_2. \end{aligned} \right\}. \quad (9.25)$$

Графически поведение системы, характер которой зависит от сочетаний коэффициентов a_{ij} , можно представить фазовыми траекториями на плоскости x_1x_2 .

При некоторых сочетаниях переменных коэффициентов a_{ij} система будет устойчивой, при других — неустойчивой. Система будет обладать свойством ультраустойчивости, если неудачные решения автоматически отбрасываются, а сочетания параметров с устойчивым протеканием процессов сохраняются.

В соответствии с определением устойчивости (см. § 8.5) система считается неустойчивой, если ее выходная координата возрастает при $t \rightarrow \infty$ и отсутствии входного сигнала. Для технической реализации системы требуется

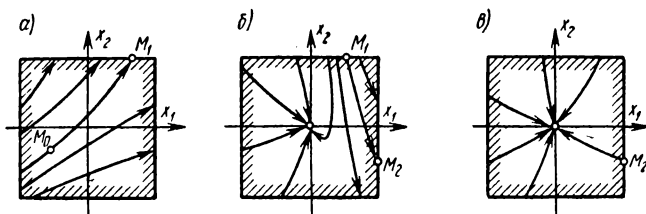


Рис. 9.35. К понятию ультраустойчивости

большая определенность, поэтому для рассматриваемых гомеостатических систем введем в эту общую формулировку ограничение, заключающееся в том, что система будет считаться неустойчивой, если хотя бы одна из координат x_i превысит некоторое заранее заданное и допустимое для этой системы значение.

Для придания свойства ультраустойчивости в систему (9.25) должно быть введено специальное переключающее устройство, осуществляющее скачкообразное переключение параметров a_{ij} каждый раз, когда фазовая траектория выйдет из допустимой области, ограниченной предельными значениями x_i . В устойчивом же состоянии система должна приходиться в положение равновесия, с которым в дальнейшем будет совпадать начало координат фазовой плоскости. Линия, ограничивающая область допустимых значений координат, называется *границей переключения* (заштрихована на рис. 9.35).

Предположим, что исследуется процесс, начинающийся в точке M_0 на рис. 9.35, а. Процесс этот, как видно из

рисунка, будет неустойчивым, вследствие чего фазовая траектория пересечет границу переключения в точке M_1 . В момент пересечения переключающее устройство изменит случайным образом сочетание параметров a_{ij} , в результате чего фазовый портрет системы примет вид, изображенный на рис. 9.35, б. Хотя в этом случае внутри контура имеется положение равновесия, новая фазовая траектория все же вновь пересечет границу переключения в точке M_2 , что приведет к очередному срабатыванию переключающего устройства. Рис. 9.35, в показывает, что система теперь становится устойчивой, изображающая точка стремится к положению равновесия, поиск прекращается.

Большое количество экспериментальных исследований, проведенных как самим Эшби с помощью гомеостата, так и другими авторами [9.10, 9.13] на моделях, позволило исследовать возможности гомеостатических систем, подтвердить их универсальность и исключительно высокую надежность. Гомеостат находил устойчивое состояние при различных дополнительных ограничениях, не учитываемых при его конструировании, например устойчивое состояние было найдено при жестком соединении двух подвижных магнитов друг с другом.

Однако статистический характер поиска, применяемый в гомеостатических системах, обладает и рядом недостатков, ограничивающих возможности практического применения таких систем. Одним из наиболее существенных недостатков является слишком большое время, затрачиваемое на поиск при большом количестве переменных x_i , т. е. при повышении сложности системы. Это время может достигать таких значений, что практически система не сможет найти устойчивое состояние в течение человеческой жизни.

Причина этого недостатка заключается в том, что с повышением сложности системы существенно снижается вероятность нахождения устойчивого состояния при очередном изменении коэффициентов a_{ij} . Опыты Эшби показывают, что вероятность нахождения устойчивого состояния в гомеостатической системе со статистическим поиском приближенно определяется экспериментальным соотношением

$$P \approx \left(\frac{1}{2}\right)^n, \quad (9.26)$$

где n — порядок системы, совпадающий с числом переменных x .

Вместе с тем можно показать, что обычно приводимые значения времени, требуемого на поиск устойчивого состояния, представляют собой верхнюю границу. На практике же они оказываются значительно меньше из-за различного рода ограничений и в силу других причин. Кроме того, время случайного поиска можно сократить за счет применения специальной методики, позволяющей увеличить вероятность нахождения устойчивого состояния. Одним из наиболее перспективных методов сокращения времени случайного поиска следует считать создание *мультиустойчивых* систем, сущность которых заключается в следующем [9.6].

В сложной ультраустойчивой системе время поискового режима оказывается очень большим в основном потому, что режим поиска по одной или нескольким переменным x_i приводит к потере имевшейся адаптации в остальных блоках. В результате такого взаимного влияния блоков друг на друга вышедший из состояния равновесия блок через несколько переключений находит устойчивое состояние, но к этому времени поиск уже начался еще в нескольких блоках, а это в свою очередь может привести к началу поиска в исходном блоке и т. д. Следовательно, сохранение адаптации в блоках, не подвергающихся влиянию внешних или внутренних (например, неисправность блока) возмущений, позволило бы значительно сократить время поискового режима.

В сложных системах, насчитывающих большое число переменных, возмущения или изменение условий работы непосредственно отражаются лишь на небольшом числе переменных. Если бы удалось осуществить поиск только по переменным, подвергшимся возмущениям, сохранив адаптацию остальных переменных, то время поиска оказалось бы сравнительно небольшим. Например, если в сложной системе, состоящей из 100 переменных, возмущение действует на переменные $x_2, x_{14}, x_{38}, x_{17}, x_{89}$, то требуется сгруппировать эти переменные в ультраустойчивую подсистему. Если же в дальнейшем возмущения влияют, допустим, на 7 других переменных x_i , то ультраустойчивая система должна быть образована именно из них. Такое последовательное перегруппирование подвергшихся воздействию возмущений переменных во временные ультра-

устойчивые подсистемы, в которых производится поиск устойчивого состояния, было названо *дисперсией режимов*. Системы, обладающие дисперсией режимов, называют мультиустойчивыми.

Известно [9.1], что физически дисперсию можно получить в случае, если все функции $\frac{dx_i}{dt}$, кроме входных функций блоков, образующих ультраустойчивую подсистему, обратить в нуль в момент начала случайного поиска. Тогда во время поиска все переменные x_i , кроме образующих ультраустойчивую подсистему, будут сохранять свое постоянное значение. Однако, несмотря на утверждение, что осуществить дисперсию таким путем довольно легко, до сих пор, насколько известно, такая мультиустойчивая система не была построена.

Более проста для технической реализации схема мультиустойчивой системы, основанная на развитии метода последовательного случайного поиска [9.21]. В сущности системы с последовательным случайным поиском представляют собой частный случай мультиустойчивой системы, в которой временные ультраустойчивые подсистемы состоят всего из одной переменной. На рис. 9.36 представлена схема модели такой мультиустойчивой системы для случая, когда ультраустойчивые подсистемы могут состоять из любого числа блоков, подвергающихся возмущениям, о которых может быть получена текущая информация.

Принцип работы такой системы заключается в том, что дисперсия режимов обеспечивается здесь не за счет обращения в нуль функций $\frac{dx_i}{dt}$, а за счет отключения

механизмов переключения параметров a_{ij} во всех блоках, не подвергающихся влиянию возмущений. Достоинством такой системы является то, что в процессе поиска не накладывается никаких ограничений ни на переменные x_i , ни на функции $\frac{dx_i}{dt}$ блоков, не участвующих в поиске.

Ультраустойчивая подсистема ищет устойчивое состояние в нормальных условиях работы системы в целом.

Потеря адаптации в одном или группе блоков будет иметь место в случае кратковременного воздействия со стороны входа достаточно сильных возмущений, выводящих соответствующие переменные x_i из области устойчивости. Случай долговременного влияния возмущений,

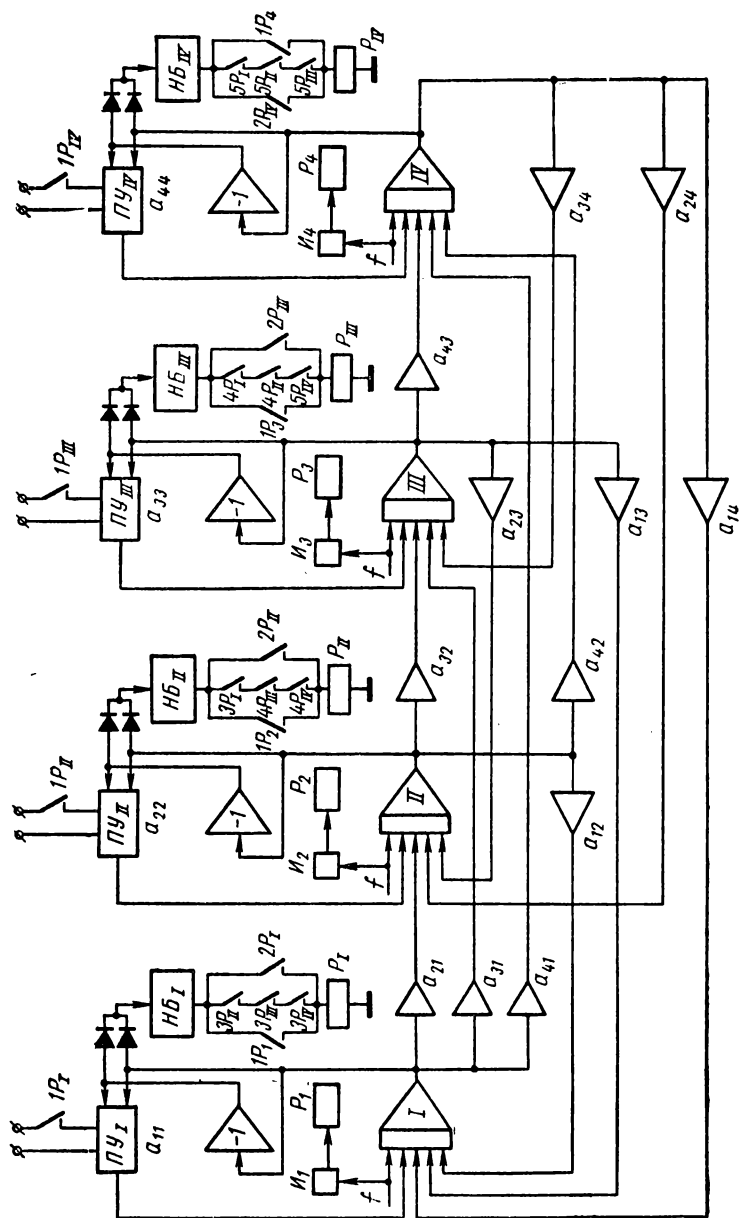


Рис. 9.36. Схема мультистадийной системы

выводящих соответствующие переменные из состояния устойчивости, рассмотрению не подлежит, поскольку при этом, сколько бы система ни осуществляла изменение настраиваемых параметров a_{ij} , неуправляемое возмущение будет удерживать x_i за пределами области устойчивости.

Модель гомеостатической системы может быть реализована с помощью известных способов построения таких систем на базе аналоговых вычислительных машин. Для простоты и наглядности схема, представленная на рис. 9.36, построена с применением электромеханических и электронных релейных устройств, но с таким же успехом она может быть воспроизведена и на основе бесконтактных переключающих устройств. Не останавливаясь подробно на описании работы модели в обычных режимах гомеостатической системы (такие описания имеются в литературе [9.10]), рассмотрим лишь принцип образования временных ультраустойчивых подсистем, предположив для определенности, что возмущение f первоначально выводит из области устойчивости блок I с переменной x_1 .

При построении модели принято, что автоматически настраиваемыми коэффициентами a_{ij} служат диагональные коэффициенты a_{ii} ($i=1, 2, 3, 4$), значения которых изменяются по случайному закону посредством переключающего устройства $ПУ_i$, управляемого с помощью реле P_k ($k=I, II, III, IV$). Границы контура переключения системы устанавливаются за счет изменения характеристик нелинейных блоков $НБ_K$, питающих реле P_k .

Для сохранения адаптации в блоках II, III, IV с началом поиска в блоке I реле P_I своими нормально замкнутыми контактами $3P_I, 4P_I$ и $5P_I$ размыкает цепи управления релё P_{II}, P_{III} и P_{IV} , вследствие чего параметры a_{22}, a_{33} и a_{44} будут сохранять свое значение независимо от состояния переменных x_2, x_3 и x_4 . Если же возмущающее воздействие влияет одновременно на блок I и блок II , то текущая информация о возмущениях, измеренная с помощью измерителей I_1 и I_2 , используется для управления срабатыванием релейных устройств P_1 и P_2 , которые контактами $2P_I$ и $2P_{II}$ шунтируют соответственно цепи $3P_{II}, 3P_{III}, 3P_{IV}$ и $3P_I, 4P_{III}, 4P_{IV}$, обеспечивая срабатывание реле P_I и P_{II} . В свою очередь, контакты $1P_1$ и $1P_2$ шунтируются контактами самоблокировки $2P_I$ и $2P_{II}$, чтобы поиск в выведенных из устойчивого состояния блоках не прерывался после прекращения

воздействия возмущений, если блоки еще не успели прийти в положение равновесия.

Нетрудно видеть, что аналогичным образом схема будет работать при любых комбинациях переменных x_i , подвергающихся влиянию возмущений и при любом их числе.

Отметим также, что принятый в схеме (см. рис. 9.36) способ построения позволяет аналогичными средствами осуществить и дисперсию режимов за счет обращения в нуль функций $\frac{dx_i}{dt}$ блоков, в которых необходимо сохранить адаптацию.

Таким образом, мультиустойчивые системы, обладая не худшей способностью к приспособлению по сравнению с обычными ультраустойчивыми системами, требуют вместе с тем гораздо меньшего времени поиска при большом числе переменных, что делает принцип ультраустойчивости практически осуществимым.

На основе исследования простейших гомеостатических систем разработаны чрезвычайно важные и сложные принципы построения *иерархичных ансамблей* систем. Принцип иерархичности управления заключается в многоэтапчатом построении управляющих систем с соподчиненными подгруппами систем с разными «уровнями» управления. С переходом в подсистему более низкого «уровня» информация конкретизируется, уточняется и употребляется для управления отдельными устройствами или для решения частных задач, промежуточных действий и т. д.

Большое научное и практическое значение имеют также работы по построению и исследованию самоорганизующихся систем, являющихся моделями нервных сетей. Для построения таких систем строятся электрические модели нейронов, применяются различные вероятностные элементы, из которых создаются системы условной вероятности, поведение которых во многом совпадает со свойствами высших организмов.

Пример. Гомеостатические системы были применены для автоматизации процессов конструирования АУС [9.13]. Как правило, структура и параметры части проектируемой системы строго детерминированы и не подлежат изменению (например, характеристики объекта управления, исполнительного двигателя и т. д.). Поэтому задача проектировщика сводится к синтезу стабилизирующих

и корректирующих устройств, которые в совокупности с детерминированной частью системы должны образовать систему автоматического управления, обладающую заданными динамическими и статическими свойствами. Задача определения передаточной функции стабилизирующих и корректирующих элементов может быть решена машинным методом с помощью гомеостатической системы; автоматически подбирающей параметры этих частей в соответствии с поставленными требованиями.

Уравнение детерминированной части системы набирается на аналоговой вычислительной машине, к которой присоединяется модель гомеостатической системы, перестраивающая случайным образом структуру и параметры на основе информации о регули-

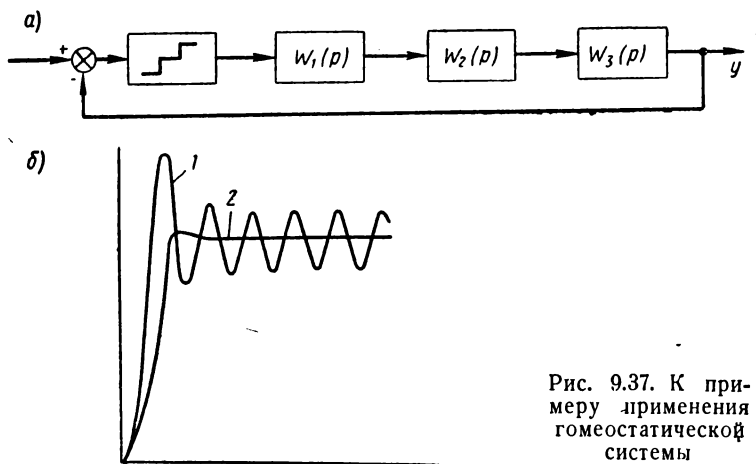


Рис. 9.37. К примеру применения гомеостатической системы

руемой величине. Следует отметить, что таким образом может быть синтезировано линейное корректирующее устройство, которое может компенсировать и влияние нелинейных элементов детерминированной части.

Пусть необходимо определить передаточную функцию W_k корректирующего устройства, имеющего назначением подавление автоколебания в релейной системе, структурная схема которой представлена на рис. 9.37, а. Передаточные функции отдельных звеньев этой системы определяются выражениями:

$$W_1(p) = \frac{K_1}{T_1 p + 1}; \quad W_2(p) = \frac{K_2}{T_2 p + 1}; \quad W_3(p) = \frac{K_3}{T_3 p + 1},$$

где $T_1 = T_2 = 1$ сек; $T_3 = 0,1$ сек; $K_2 = K_3 = 1$.

Автоматический синтез был осуществлен с помощью модели простейшей гомеостатической системы, состоящей из двух блоков и подключенной параллельно релейному элементу и звену $W_1(p)$. На вход второго блока был подан сигнал u_1 , а выходной сигнал этого же блока поступал на вход релейного элемента. Задача авто-

матического синтеза была решена в течение 3 мин 20 сек (за 5 циклов), в результате чего были получены значения коэффициентов $a_{11} = a_{22} = a_{21} = a_{12} = -1$.

Уравнение корректирующего устройства получает вид

$$\frac{dx_1}{dt} = x_1 + x_2,$$

$$\frac{dx_2}{dt} = x_1 + x_2 + y_1,$$

откуда передаточная функция

$$W_k(p) = \frac{p+1}{p(p+2)}.$$

На рис. 9.37, б кривой 2 представлен переходный процесс скорректированной релейной системы при скачкообразном изменении входного сигнала, а кривая 1 определяет характер переходного процесса в нескорректированной системе.

§ 9.4. ОБУЧАЮЩИЕСЯ СИСТЕМЫ

Системы, предназначенные для целенаправленного совершенствования структуры, параметров или алгоритма действия на основе анализа накапливаемой в процессе функционирования информации, составляют весьма важный и получающий все большее практическое значение класс обучающихся автоматов. Создание сложных обучающихся систем управления, способных автоматически вырабатывать наиболее оправданную линию своего поведения при изменяющихся внешних условиях, является важнейшим достижением технической кибернетики. Обучение можно рассматривать как вид организации систем, более совершенной по сравнению с адаптацией устройств, рассмотренных ранее.

Процесс обучения будет успешным лишь в случае, если на обучающуюся систему воздействуют сигналы внешней корректировки («поощрение» или «наказание»), фиксирующие целесообразность предпринятых действий. В самообучаемых системах такая внешняя корректировка отсутствует.

Характерная особенность поведения обучающихся систем заключается в том, что вначале машина обычно ничего «не знает» об условиях работы, но с течением времени, используя различные виды автоматического поиска с анализом его результатов, она накапливает информацию и применяет ее для корректировки своих

действий. Процесс обучения в большинстве случаев можно рассматривать как процесс последовательных приближений.

Наиболее важной и развитой группой обучающихся систем в настоящее время являются системы обучения распознаванию образов, которым посвящено большое количество трудов. Идея построения таких систем первоначально была выдвинута Розенблатом [9.4] и им реализована в виде модели, обладающей свойствами обучения и самообучения и названной перцептроном*. Это название было принято и для обозначения последующих аналогичных систем.

В основу действия перцептрона положены сформулированные Розенблатом теоремы, в соответствии с которыми функция памяти человека распределена случайным образом среди большого количества одинаковых ассоциирующих элементов и начальная неопределенность организации мозга при рождении очень велика, мыслящий организм вырабатывает понятия и алгоритм своих действий в окружающей обстановке в результате обучения и накопления опыта.

Перцептрон представлял собой статистическую самообучаемую систему со случайными связями между основными элементами. Схема простейшего перцептрона приведена на рис. 9.38 [9.14]. Он состоит из рецепторов P , ассоциативных элементов A , усилителей K с переменным коэффициентом усиления, сумматоров Σ и выходных элементов R («реагирующих» элементов).

Рецепторы выполняют функции датчиков устройства и служат для преобразования изображения в большое число электрических сигналов, поступающих в ассоциативные элементы. На практике в качестве рецепторов обычно применяют фотоэлементы.

Ассоциативное устройство состоит из набора одинаковых пороговых элементов, число которых близко к числу рецепторов. Эти элементы имеют по несколько входов и одному выходу, причем каждый рецептор может быть подключен к данному A -элементу со знаком плюс или минус, либо вообще не подключен. Характер связей между рецепторами и A -элементами является случайным, но в ходе эксперимента не изменяется. Задача ассоциа-

* Этот термин берет начало от латинского слова *perception* — понимание, познание.

тивных элементов здесь сводится к алгебраическому суммированию входных сигналов. Если полученная сумма оказывается больше некоторой величины Θ , то A -элемент возбуждается, вырабатывая выходной сигнал, равный

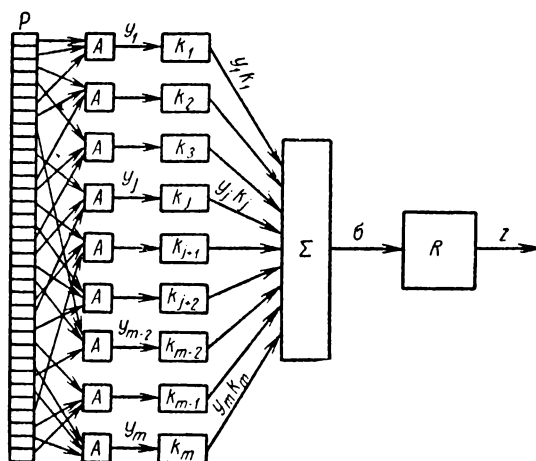


Рис. 9.38. Схема перцептрона

единице. В противном случае выходной сигнал равен нулю (пороговый элемент остается невозбужденным). Следовательно, выходной сигнал каждого из A -элементов

$$y_j = \begin{cases} 1 & \left(\sum_{i=1}^n r_{ij} x_i - \Theta \right) \geq 0, \\ 0 & \left(\sum_{i=1}^n r_{ij} x_i - \Theta \right) < 0, \end{cases} \quad (9.27)$$

где коэффициент r_{ij} , зависящий от характера связи между рецептором и j -м A -элементом, может принимать значения $+1$; 0 ; -1 . Выходные сигналы ассоциативных элементов через усилители с переменным коэффициентом усиления k_j (k_j может быть и меньше нуля) поступают в суммирующее устройство Σ , которое вырабатывает сигнал

$$\sigma = \sum_{j=1}^m k_j y_j,$$

где m — число ассоциативных элементов. Сигнал σ передается на вход выходного элемента R , реакция z которого отвечает условию

$$z = \begin{cases} 1 & \sigma \geq 0 \\ 0 & \sigma < 0. \end{cases} \quad (9.28)$$

Таким образом, задача обучения перцептрона состоит в таком подборе коэффициентов k_j , чтобы $z=1$ при появлении на входе фигур одного образа и $z=0$ при появлении фигур другого образа. (Для распознавания больше двух образов применяются несколько другие схемы, в основе которых лежит разбиение ассоциативных элементов на несколько групп.) Физически эта задача решается путем изменения смещений, подаваемых на усилитель k_j , причем в режиме обучения смещения изменяются человеком-учителем, а в режиме самообучения — с помощью специальной обратной связи.

Перцептрон может быть построен не только на основе случайных связей, но и в более специализированных вариантах. Возможности машин, обучающихся распознаванию образов, исключительно велики: известны исследования по применению таких машин для распознавания радиолокационных изображений, нефтеносных скважин, диагностики заболеваний, в читающих автоматах и т. д.

Процесс обучения машины распознаванию образов часто можно свести к определению поверхности, выделяющей в n -мерном пространстве области, соответствующие различным образам. В этом случае распознавание обученной машиной некоторого неизвестного изображения состоит в испытании этого изображения и в отыскании области, к которой относится данное изображение. Такой подход позволяет развить математическую теорию таких машин, выявляющую общность задач обучаемых систем и систем адаптации. В результате этого за последнее время были разработаны интересные и достаточно совершенные схемы обучаемых автоматов и систем, а также самосовершенствующихся программ управления, которые позволяют перейти к практическому применению таких устройств в системах автоматического управления.

Большую группу обучающихся и самообучаемых систем составляют устройства, имитирующие поведение живых организмов, выработку условных рефлексов. Представи-

телями систем этого типа являются широко известные «животные» электронного «зверинца» (черепахи Грея Уолтера, электронный «лис» Дюкрока, черепаха Стелла [9.15]), системы лабиринтного типа, модели для исследования сложных условных рефлексов и целый ряд других.

Широкие исследования существующих устройств этого типа и разработка новых принципов их построения имеют огромное значение для развития бионики.

Обучающиеся системы находятся еще в ранней стадии своего развития, практическое их применение носит ограниченный характер, однако они являются основой для решения в будущем наиболее сложных задач управления.

§ 9.5. ОПТИМАЛЬНОЕ УПРАВЛЕНИЕ

Система, которая является оптимальной (см. § 2.2) по некоторому признаку, характеризующему ее работу, может оказаться не оптимальной по другим признакам. Будем рассматривать лишь системы, оптимальные в принятом смысле, т. е. оптимальные по некоторому наиболее существенному для каждой из них критерию качества.

В связи с этим важное значение принимает вопрос о выборе *критерия оптимальности*, характеризующего эффективность управления. Такими критериями могут служить различные технические или экономические показатели, зависящие от условий работы системы. Например, точность ее работы, длительность переходного процесса, расход энергии, производительность объекта управления, себестоимость продукции или ее качество и т. п.

Принято различать две группы критериев оптимальности: *статистические критерии*, применяемые для оценки эффективности функционирования систем при наличии случайных воздействий, и *регулярные критерии*.

За последние годы широкое развитие, особенно в трудах советских ученых, получили статистические методы анализа и синтеза оптимальных систем. В тех случаях, когда сигнал на входе системы является случайной функцией времени или при регулярном управляющем воздействии система подвержена воздействию случайных помех с произвольной точкой их приложения (источниками таких случайных помех часто оказываются и элементы исследуемой системы), выходной сигнал и функционалы,

определяющие качество системы, также оказываются случайными. Поэтому под оптимальной системой в статистическом смысле следует понимать динамическую систему, обеспечивающую наилучшее приближение выходного сигнала к требуемому преобразованию полезного входного воздействия, подверженного влиянию случайных помех. В этом смысле система становится оптимальной тогда, когда принятый критерий оптимальности, в качестве которого обычно выбирают какую-либо статистическую характеристику, достигает экстремума.

Существует значительное количество статистических критериев оптимальности. Наибольшее распространение получил критерий минимума среднего квадрата ошибки ϵ_t^2 , в соответствии с которым система оптимальна, если ϵ_t^2 имеет минимум для каждого значения времени t . Очень часто на практике вместо среднего квадрата ошибки применяется тождественный ему критерий минимума среднеквадратичного значения ошибки (равного корню квадратному из ϵ_t^2).

Указанный критерий обладает большой общностью при сравнительной простоте, что позволило с успехом применять его для решения большого числа практических задач. Он принадлежит к классу так называемых *бейесовых критериев*, с точки зрения которых система считается оптимальной, если выполняется условие минимума математического ожидания

$$M[l(x_t, y_t)] = \min, \quad (9.29)$$

где l — некоторая известная неотрицательная функция, не зависящая от свойств системы;

x_t — случайный сигнал на входе системы, являющийся суммой полезного сигнала и шума;

y_t — случайный сигнал на выходе системы.

Функция l , называемая функцией потерь, или *функцией цены ошибки*, также является случайной. Левая часть выражения (9.29) называется *средним риском*.

В зависимости от выбора функции потерь могут быть получены различные критерии оптимальности, позволяющие оценить и оптимизировать те или иные характеристики системы. Например, критерий среднего квадрата ошибки получается из бейесова критерия, если за функцию потерь принять

$$l(x_t, y_t) = (x_t - y_t)^2. \quad (9.30)$$

Из байесовых критериев можно также получить критерий, позволяющий обнаружить полезный сигнал в смеси (критерий Неймана — Пирсона), критерий минимума математического ожидания числа выходов ошибки из заданных пределов в течение заданного интервала времени и ряд других [9.20].

Регулярные критерии оптимальности обычно принимаются в виде некоторой функции от параметров или функционала от управляющих воздействий и обобщенных координат системы. Среди них широкое распространение получили *интегральные критерии* различного вида. Если автоматическая система описывается системой обыкновенных дифференциальных уравнений

$$\frac{dx_i}{dt} = f_i(x_1, x_2, \dots, x_n, u_1, u_2, \dots, u_r),$$

где x_1, x_2, \dots, x_n — координаты объекта, а u_1, u_2, \dots, u_r — управляющие воздействия, то интегральным критерием может служить выражение [9.5]

$$I = \int_{t_0}^{t_1} f^0(x_1, x_2, \dots, x_n, u_1, u_2, \dots, u_r) dt. \quad (9.31)$$

Здесь $f^0(x_1, x_2, \dots, x_n, u_1, \dots, u_r)$ — известная функция, выбор которой зависит от назначения системы и от того, какая ее характеристика считается наиболее существенной.

В ряде случаев задачей оптимальной настройки системы является получение желательного переходного процесса. Одним из удобных и распространенных критериев для таких систем является *обобщенный интегральный критерий*

$$I = \int_0^{\infty} \sum a_{ij} x_i x_j dt, \quad (9.32)$$

где x_i ($i = 1, 2, \dots, n$) — координаты системы;
 a_{ij} — постоянные коэффициенты.

Величина I минимизируется за счет изменения настроечных параметров, в качестве которых могут быть приняты постоянные времени, коэффициент усиления.

Процесс функционирования любого управляемого объекта состоит, в сущности, в последовательном пере-

ходе от одного состояния к другому, причем каждый из таких переходов обычно может быть осуществлен многими способами. Действующая таким образом система считается оптимальной, если управляющие воздействия, переводящие систему в новое состояние, будут организованы так, что выбранный способ перехода окажется наиболее выгодным в принятом смысле. Оптимальное управление обеспечивает наивыгоднейшее значение определенного критерия, характеризующего эффективность управления при заданных ограничениях.

Таким образом, проблема оптимального управления сводится в большинстве случаев к решению задачи о нахождении экстремумов функций многих переменных или функционалов при наличии ограничений. Применение классических вариационных методов для решения этих задач затрудняется тем, что ограничения часто задаются в виде неравенств; кроме того, это требует чрезмерно большого количества элементарных операций. Более совершенным методом решения задач оптимального управления является математическое программирование.

Наиболее универсальным и важным разделом математического программирования является *принцип максимума*, развитый Л. С. Понтрягиным и его учениками [9.5] и составляющий ядро теории оптимальных процессов.

В теории оптимального управления часто встречается задача об *оптимальном быстродействии*, т. е. таком оптимальном процессе, при котором время перехода объекта управления из одного состояния в другое будет минимальным. Рассмотрим математическое решение задачи оптимального управления с помощью принципа максимума в применении к оптимальной по быстродействию системе [9.18].

Пусть исследуется некоторый объект, состояние которого в каждый момент времени задается фазовыми координатами x^1, \dots, x^n , являющимися функциями времени. Переход объекта из одного состояния в другое происходит под влиянием управляющих воздействий (управляющих параметров) u^1, \dots, u^r . Поведение объекта определяется системой обыкновенных дифференциальных уравнений вида

$$\dot{x}^i = f^i(x^1, \dots, x^n, u^1, \dots, u^r) = f^i(\bar{x}, \bar{u}), \quad (9.33)$$

где \bar{x} и \bar{u} — векторы: $\bar{x} = (x^1, \dots, x^n)$ и $\bar{u} = (u^1, \dots, u^r)$.

В реальных устройствах управляющие параметры не могут изменяться неограниченно, обычно для каждого из управляющих параметров u^1, \dots, u^r задаются некоторые пределы изменения. Это означает, что в пространстве управляющих параметров u^1, \dots, u^r задано некоторое множество U (*область управления*) и для реальной системы точка u должна принадлежать множеству $U (u \in U)^1$. При этом рассмотрению подлежат не только непрерывные, но и кусочно-непрерывные управления $u^1(t), \dots, u^r(t)$, т. е. управляющие параметры считаются безынерционными.

Будем считать, что исследуемая система, исходное состояние которой характеризовалось произвольной точкой x фазового пространства, должна перейти в некоторое новое состояние x_1 , причем необходимо найти такое оптимальное управление $u(t)$, при котором время $T(x)$ перехода из точки x в точку x_1 будет наименьшим. Проводя дальнейшие рассуждения, будем предполагать, что при любом x существует оптимальный процесс перехода в точку x_1 и что выполняется следующая гипотеза: функция $T(x) = T(x^1, x^2, \dots, x^n)$ непрерывна и имеет всюду, кроме точки x_1 , первые непрерывные частные производные $\frac{\partial T}{\partial x^1}, \frac{\partial T}{\partial x^2}, \dots, \frac{\partial T}{\partial x^n}$ и вторые непрерывные производные $\frac{\partial^2 T(x)}{\partial x^i \partial x^j}$ ($i, j = 1, 2, \dots, n$), а функции $f^i(x, u)$ — первые непрерывные производные $\frac{\partial f^i(x, u)}{\partial x^j}$ ($i, j = 1, 2, \dots, n$).

В начальный момент времени, когда начинается исследование объекта с начальным состоянием x_0 , он может уже находиться под воздействием управления $u = u_0$, вызывающим его перемещение в фазовом пространстве. Обозначая фазовую траекторию объекта через $y(t) = (y^1(t), y^2(t), \dots, y^n(t))$, при $t > t_0$ получим уравнение

$$\dot{y}^i(t) = f^i(y(t), u_0) \quad i = 1, 2, \dots, n \quad (9.34)$$

и начальное условие

$$y(t_0) = x_0. \quad (9.35)$$

Задача теперь будет заключаться в получении оптимального перехода объекта в точку x_1 из точки $y(t)$,

¹ Символ \in означает, что точка принадлежит множеству.

в которую объект попал под воздействием u_0 (рис. 9.39). Общее время движения объекта из точки x_0 в точку x_1 в этом случае будет складываться из времени $t - t_0$, за которое система приходит в точку $y(t)$, и времени $T(y(t))$ оптимального перехода из точки $y(t)$ в точку x_1 , т. е. будет равно $(t - t_0) + T(y(t))$. Обозначая время оптимального перехода из точки x_0 в точку x_1 (штриховая линия на рис. 9.39) через $T(x_0) = T(y(t_0))$, получим:

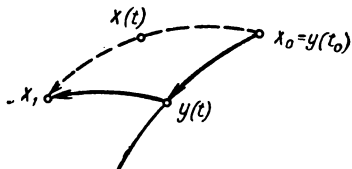


Рис. 9.39. Процесс перехода управляемого объекта из состояния x_0 в состояние x_1

$$T(y(t_0)) \leq (t - t_0) + T(y(t)). \quad (9.36)$$

Для удобства рассуждения вместо $T(x)$ введем функцию $\omega(x)$:

$$\omega(x) = -T(x), \quad (9.37)$$

на которую, конечно, распространяются рассмотренные ранее гипотезы. Подставив (9.37) в (9.36), находим:

$$\frac{\omega(y(t)) - \omega(y(t_0))}{t - t_0} \leq 1,$$

откуда с помощью предельного перехода при $t \rightarrow t_0$

$$\left. \frac{d}{dt} \omega(y(t)) \right|_{t=t_0} \leq 1. \quad (9.38)$$

По формуле полной производной находим:

$$\frac{d}{dt} \omega(y(t)) = \sum_{i=1}^n \frac{\partial \omega(y(t))}{\partial x^i} \dot{y}^i(t),$$

в соответствии с чем и с учетом (9.34), (9.35) можно (9.38) записать в виде

$$\sum_{i=1}^n \frac{\partial \omega(x_0)}{\partial x^i} f^i(x_0, u_0) \leq 1.$$

Последнее неравенство справедливо для произвольных x_0 и u_0 , вследствие чего для любой точки x фазового пространства, отличной от x_1 , и любой точки u области

управления U выполняется соотношение

$$\sum_{i=1}^n \frac{\partial \omega(x)}{\partial x^i} f^i(x, u) \leq 1. \quad (9.39)$$

Будем далее рассматривать оптимальный процесс $(u(t), x(t))$, переводящий объект из состояния x_0 в точку x_1 за отрезок времени $t_0 \leq t \leq t_1$.

Рассматривая некоторую промежуточную точку $x(t)$ оптимальной фазовой траектории (см. рис. 9.39), замечаем, что время перехода объекта из состояния $x(t)$ в состояние x_1

$$T(x(t)) = T(x_0) - (t - t_0),$$

откуда, на основании (9.37),

$$\omega(x(t)) = \omega(x_0) + t - t_0.$$

Взяв производную по x^i , находим аналогично с (9.39):

$$\sum_{i=1}^n \frac{\partial \omega(x(t))}{\partial x^i} \dot{x}^i(t) = 1,$$

или с учетом уравнения движения объекта для каждого оптимального процесса имеем:

$$\sum_{i=1}^n \frac{\partial \omega(x(t))}{\partial x^i} f^i(x(t), u(t)) = 1, \quad (9.40)$$

$$(t_0 \leq t \leq t_1).$$

Введем для удобства функцию

$$B(x, u) = \sum_{i=1}^n \frac{\partial \omega(x)}{\partial x^i} f^i(x, u), \quad (9.41)$$

которая в силу принятой выше гипотезы имеет непрерывные производные по переменным x^i :

$$\begin{aligned} \frac{\partial B(x, u(t))}{\partial x^k} &= \sum_{i=1}^n \frac{\partial^2 \omega(x)}{\partial x^i \partial x^k} f^i(x, u(t)) + \\ &+ \sum_{i=1}^n \frac{\partial \omega(x)}{\partial x^i} \cdot \frac{\partial f^i(x, u(t))}{\partial x^k}, \end{aligned} \quad (9.42)$$

$$k = 1, 2, \dots, n.$$

Из (9.39) и (9.40) вытекает, что для любого $x \neq x_1$

$$B(x, u) \leq 1, \quad (9.43)$$

для любого оптимального процесса $(x(t), u(t))$

$$B(x(t), u(t)) \approx 1. \quad (9.44)$$

Это означает, что при $x = x(t)$ функция $B(x, u(t))$ достигает максимума, вследствие чего ее частные производные по x^1, \dots, x^n обращаются в нуль:

$$\begin{aligned} & \sum_{i=1}^n \frac{\partial^2 \omega(x(t))}{\partial x^i \partial x^k} f^i(x(t), u(t)) + \\ & + \sum_{i=1}^n \frac{\partial \omega(x(t))}{\partial x^i} \cdot \frac{\partial f^i(x(t), u(t))}{\partial x^k} = 0, \quad (9.45) \\ & k = 1, \dots, n. \end{aligned}$$

Дифференцирование функции $\frac{\partial \omega(x(t))}{\partial x^k}$ по t с учетом уравнения движения объекта дает:

$$\begin{aligned} \frac{d}{dt} \left(\frac{\partial \omega(x(t))}{\partial x^k} \right) &= \sum_{i=1}^n \frac{\partial^2 \omega(x(t))}{\partial x^k \partial x^i} \dot{x}^i(t) = \\ &= \sum_{i=1}^n \frac{\partial^2 \omega(x(t))}{\partial x^k \partial x^i} f^i(x(t), u(t)). \end{aligned}$$

Это позволяет переписать соотношение (9.45) в виде

$$\begin{aligned} \frac{d}{dt} \left(\frac{\partial \omega(x(t))}{\partial x^k} \right) + \sum_{i=1}^n \frac{\partial \omega(x(t))}{\partial x^i} \cdot \frac{\partial f^i(x(t), u(t))}{\partial x^k} &= 0, \quad (9.46) \\ k &= 1, 2, \dots, n. \end{aligned}$$

Введем вспомогательные обозначения:

$$\frac{\partial \omega(x(t))}{\partial x^1} = \psi_1(t), \dots, \frac{\partial \omega(x(t))}{\partial x^n} = \psi_n(t). \quad (9.47)$$

Тогда вместо (9.41) будет

$$B(x(t), u(t)) = \sum_{i=1}^n \psi_i(t) f^i(x(t), u(t)).$$

Уравнение (9.44) при $t_0 \leq t \leq t_1$ для оптимального процесса $(x(t), u(t))$ теперь примет вид

$$\sum_{i=1}^n \psi_i(t) f^i(x(t), u(t)) \equiv 1, \quad (9.48)$$

а из (9.43) для любого $u \in U$ ($t_0 \leq t \leq t_1$) получаем:

$$\sum_{i=1}^n \psi_i(t) f^i(x(t), u) \leq 1. \quad (9.49)$$

Выражение (9.46) можно переписать в виде

$$\begin{aligned} \dot{\psi}_k(t) + \sum_{i=1}^n \psi_i(t) \frac{\partial f^i(x(t), u(t))}{\partial x^k} &= 0, \\ k &= 1, \dots, n. \end{aligned} \quad (9.50)$$

Чтобы сформулировать необходимое условие оптимальности, целесообразно ввести в рассмотрение вспомогательную функцию

$$\begin{aligned} H(\psi, x, u) &= \sum_{i=1}^n \psi_i f^i(x, u) = \\ &= \psi_1 f^1(x, u) + \dots + \psi_n f^n(x, u), \end{aligned} \quad (9.51)$$

число аргументов которой составляет $2n + r$. При этом соотношения (9.48) и (9.49) можно записать в следующем виде:

$$H(\psi(t), x(t), u(t)) \equiv 1 \quad (9.52)$$

— для оптимального процесса $(x(t), u(t))$ при $t_0 \leq t \leq t_1$ и

$$H(\psi(t), x(t), u(t)) \leq 1 \quad (9.53)$$

— для любой точки $u \in U$.

Сопоставляя (9.53) с (9.52), можно написать

$$\begin{aligned} \max_{u \in U} H(\psi(t), x(t), u) &= \\ &= H(\psi(t), x(t), u(t)), \quad t_0 \leq t \leq t_1. \end{aligned} \quad (9.54)$$

Выражения (9.50) могут быть переписаны в следующей форме:

$$\begin{aligned} \dot{\psi}_k(t) &= - \frac{\partial H(\psi(t), x(t), u(t))}{\partial x^k}, \\ k &= 1, \dots, n. \end{aligned} \quad (9.55)$$

Полученные результаты показывают, что если $x(t)$, $u(t)$ — оптимальный процесс, то существует такая функция $\psi(t) = (\psi_1(t) \dots \psi_n(t))$, что выполняются соотношения (9.52), (9.54), (9.55). Отсюда получается теорема, которая называется принципом максимума.

Пусть исследуется некоторый процесс $(x(t), u(t))$ при $t_0 \leq t \leq t_1$, переводящий объект с уравнением движения $\dot{x} = f(x, u)$, $u \in U$ из состояния x_0 в состояние x_1 . Введем

вспомогательную функцию $H(\psi, x, u) = \sum_{i=1}^n \psi_i f^i(x, u)$, с помощью которой можно определить переменные

$$\dot{\psi}_k = - \frac{\partial H(\psi, x(t), u(t))}{\partial x^k}, \quad (9.56)$$

$$k = 1, \dots, n.$$

Если процесс $(x(t), u(t))$ является оптимальным, то существует такое ненулевое решение $\psi(t) = (\psi_1(t), \dots, \psi_n(t))$, $t_0 \leq t \leq t_1$ системы уравнений (9.56), что для любого момента времени t при $t_0 \leq t \leq t_1$ выполнено условие максимума (9.54) и условие

$$H(\psi(t), x(t), u(t)) = 1. \quad (9.57)$$

Это и есть формулировка принципа максимума.

Полученные условия оптимальности, однако, не являются необходимыми, так как рассуждения, приводящие к этим условиям, строились на предположениях и гипотезах, выполнение которых не является необходимым для оптимальности процесса. Но если вместо условия (9.57) записать более слабое условие

$$H(\psi(t_1), x(t_1), u(t_1)) \geq 0, \quad (9.58)$$

то принцип максимума будет необходимым условием оптимальности (9.42).

Соотношения (9.43) и (9.44) составляют сущность *динамического программирования* — развитого за последние годы раздела математики, занимающегося вопросами оптимального многошагового управления. В общей форме сущность метода динамического программирования заключается в том, что если управление является оптимальным, то оно будет оптимальным и для процесса, остающегося после совершения первого шага.

Сопоставляя (9.44) с (9.43) для любого $x \neq x_1$, можно записать:

$$\max_{u \in U} B(x, u) = 1, \quad (9.59)$$

или с учетом (9.41)

$$\max \sum \frac{\partial \omega(x)}{\partial x^i} f^i(x, u) = 1. \quad (9.60)$$

Соотношение (9.60) называется *уравнением Беллмана*.

Пример. Движение управляемого объекта задано системой линейных дифференциальных уравнений [9.22]

$$\frac{dx_j}{dt} = \sum_{k=1}^n a_{jk} x_k + \sum_{l=1}^r b_{jl} u_l, \quad j = 1, 2, \dots, n, \quad (9.61)$$

причем в пространстве управлений U задана область управления в виде некоторого r -мерного куба неравенствами

$$|u_j| \leq 1; \quad j = 1, 2, \dots, r. \quad (9.62)$$

Требуется найти такое уравнение $\bar{u}^*(t)$, которое переводит систему из начального состояния x_0 в конечное x_1 за минимальное время.

Поставленная задача означает, что искомое уравнение $\bar{u}^*(t)$ должно минимизировать функционал

$$J = \int_{t_0}^{t_1} dt = t_1 - t_0 = T_0, \quad (9.63)$$

где T_0 — время переходного процесса.

В соответствии с (9.51) и (9.61) получаем:

$$H = \sum_{j=1}^n \psi_j \left(\sum_{k=1}^n a_{jk} x_k + \sum_{l=1}^r b_{jl} u_l \right) \quad (9.64)$$

или в векторной форме записи

$$H = (\bar{\psi}, A\bar{x}) + (\bar{\psi}, B\bar{u}), \quad (9.65)$$

где

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix}, \quad B = \begin{bmatrix} b_{11} & b_{12} & \dots & b_{1r} \\ b_{21} & b_{22} & \dots & b_{2r} \\ \dots & \dots & \dots & \dots \\ b_{r1} & b_{r2} & \dots & b_{rr} \end{bmatrix}. \quad (9.66)$$

Согласно (9.57) оптимальное управление максимизирует функцию H , но максимум функции H по переменной \bar{u} имеет место тогда, когда второе слагаемое в правой части выражения (9.65) максимально. Поэтому условие оптимальности искомого управления

можно представить в виде

$$(\bar{\psi}, \mathbf{B}\bar{u}) = \sum_{j=1}^n \left(\psi_j \sum_{l=1}^r b_{jl} u_l \right) = \sum_{l=1}^n \sum_{j=1}^n \psi_j b_{jl} u_l = \max. \quad (9.67)$$

Требования (9.67) выполняются при условии, что

$$\sum_{j=1}^n \psi_j b_{jl} u_l = \max, \quad l = 1, 2, \dots, r. \quad (9.68)$$

Поскольку каждая составляющая вектора $\bar{u} = (u_1, u_2, \dots, u_r)$ изменяется в заданной области независимо от остальных состав-

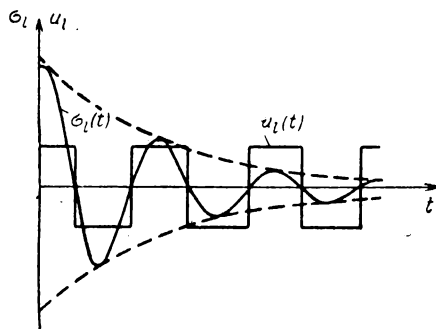


Рис. 9.40. К построению управления, оптимального по быстродействию

ляющих, то каждая составляющая u_l должна принимать максимальное значение. Поэтому условие (9.68) выполняется при выборе управления

$$u_l = \text{sign} \left(\sum_{j=1}^n b_{jl} \psi_j \right), \quad l = 1, 2, \dots, r. \quad (9.69)$$

Выражения (9.69) определяют оптимальное по быстродействию управление. Вводя обозначение $\sum_{j=1}^n b_{jl} \psi_j = \sigma_l$, получаем

$$u_l = \text{sign} \sigma_l, \quad l = 1, 2, \dots, r, \quad (9.70)$$

откуда видно, что искомое оптимальное управление является кусочно-постоянным (релейным). Характер изменения u_l во времени при заданном поведении σ_l представлен на рис. 9.40, из которого видно, что в точках, где $\sigma_l(t)$ обращается в нуль, управление $u_l(t)$ терпит разрыв, т. е. происходит скачок вектора \bar{u} из одной вершины r -мерного куба в другую. Таким образом, для получения максимального быстродействия управление объектом должно осуществляться по релейному закону.

НАДЕЖНОСТЬ АВТОМАТИЧЕСКИХ УПРАВЛЯЮЩИХ СИСТЕМ

§ 10.1 КОЛИЧЕСТВЕННЫЕ ХАРАКТЕРИСТИКИ НАДЕЖНОСТИ

Надежность — важнейшее свойство, которым должна обладать всякая автоматическая система. Она определяет способность системы правильно функционировать в течение достаточно длительного времени. Если система не обладает таким свойством, т. е. ненадежна, то при всем совершенстве в других отношениях она не может быть полезной. К пониманию этого привело появление сложных радиоэлектронных устройств, легко выходявших из строя при возникновении неисправностей отдельных элементов, входивших в их состав. Поэтому необходимы особые мероприятия, чтобы обеспечить надежность подобных устройств. Эту задачу решает теория надежности, получившая применение и к автоматическим системам.

Основное понятие учения о надежности — *отказ*. Это — нарушение нормального функционирования рассматриваемого устройства, приводящее к уходу его рабочих характеристик за допустимые пределы. Могут быть отказы вследствие внезапных повреждений (поломок, пробоев и т. п.) и вследствие постепенных изменений, обычно вызываемых *износом*. Произойдет ли отказ и к каким последствиям он приведет — все это можно определить, лишь оценивая вероятность подобных событий. Поэтому для количественного выражения надежности служат вероятностные характеристики работы рассматриваемых устройств.

Один из способов построения указанных характеристик состоит в изучении *продолжительности T исправной работы*, т. е. времени, протекающего от момента включения рассматриваемого устройства до возникновения первого отказа. Полагая, что в момент включения время $t=0$, будем считать первый отказ происходящим при $t=T$, рассматривая таким образом лишь положительные

значения времени. Время T — случайная величина, подчиненная условию

$$0 < T \leq \infty. \quad (10.1)$$

Вероятность того, что при каком-либо значении текущего времени t

$$T \geq t$$

называют *вероятностью безотказной работы* $P(t)$ в течение времени t . Вероятность отказа за то же время

$$Q(t) = 1 - P(t). \quad (10.2)$$

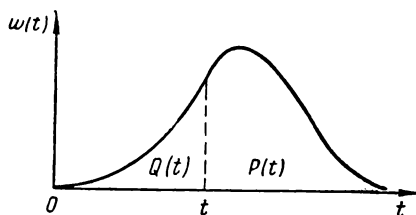


Рис. 10.1. Определение вероятности отказа и безотказной работы по кривой $w(t)$

В соответствии с (10.1)

$$\left. \begin{aligned} P(0) &= 1, \\ P(\infty) &= 0. \end{aligned} \right\} \quad (10.3)$$

Вводя соответствующую плотность вероятности $w(t)$, имеем:

$$\left. \begin{aligned} P(t) &= \int_t^{\infty} w(t) dt, \\ Q(t) &= \int_0^t w(t) dt. \end{aligned} \right\} \quad (10.4)$$

Эти соотношения поясняет рис. 10.1. Из них находим:

$$\dot{Q}(t) = w(t). \quad (10.5)$$

На практике по результатам испытаний может быть получена статистическая оценка вероятности безотказной

работы

$$P^*(t) = \frac{N - n(t)}{N} \quad (10.6)$$

и плотности вероятности

$$w^*(t) = \frac{n(t, \Delta t)}{N \Delta t}. \quad (10.7)$$

Здесь N — число элементов, поставленных на испытание;
 $n(t)$ — число элементов, отказавших за время от 0 до t ;
 $n(t, \Delta t)$ — число элементов, отказавших в интервале
 времени от

$$t - \frac{\Delta t}{2} \quad \text{до} \quad t + \frac{\Delta t}{2}.$$

Поскольку время T — случайная величина, можно определить лишь его математическое ожидание

$$\bar{T} = \int_0^{\infty} t w(t) dt.$$

Подставляя сюда

$$w(t) = -\frac{d}{dt} P(t), \quad (10.8)$$

что следует из (10.2) и (10.5), имеем:

$$\bar{T} = - \int_0^{\infty} t \frac{d}{dt} P(t) dt = \int_0^1 t dP(t).$$

Далее, интегрируя по частям, находим:

$$\bar{T} = tP(t) \Big|_0^1 + \int_0^{\infty} P(t) dt.$$

Здесь по (10.3) первое слагаемое равно нулю. Следовательно,

$$\bar{T} = \int_0^{\infty} P(t) dt, \quad (10.9)$$

что поясняет рис. 10.2.

Дисперсия, характеризующая разброс случайных значений времени T вокруг его среднего значения, по (3.37) равна

$$\sigma^2 = \overline{T^2} - (\bar{T})^2.$$

Иначе [10.3]

$$\sigma^2 = 2 \int_0^{\infty} t P(t) dt - (\bar{T})^2. \quad (10.10)$$

Предположим теперь, что на испытании находится большая группа однотипных элементов, число которых равно k , и что отказавшие элементы не заменяются новыми. Предположим также, что все эти элементы характеризуются одинаковой вероятностью безотказной работы $P(t)$. Тогда по истечении времени t после начала испыта-

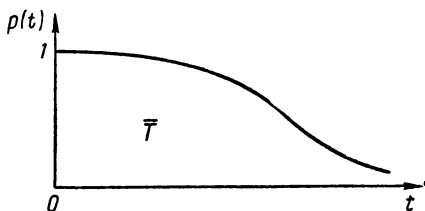


Рис. 10.2. Зависимость вероятности безотказной работы от времени

ния количество исправно действующих элементов данной группы будет $kP(t)$, а ко времени $t + \Delta t$ — $kP(t + \Delta t)$.

Составим отношение количества элементов, отказавших в интервале времени $t, t + \Delta t$, к числу исправно действующих в начале этого интервала. Это будет

$$\frac{P(t) - P(t + \Delta t)}{P(t)}.$$

Деля затем полученное выражение на Δt и переходя к пределу при $\Delta t \rightarrow 0$, получим отношение числа элементов, отказывающих в единицу времени, к числу исправно работающих

$$\lambda(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t) - P(t + \Delta t)}{\Delta t P(t)},$$

откуда

$$\lambda(t) = - \frac{1}{P(t)} \frac{d}{dt} P(t). \quad (10.11)$$

Подставив (10.8) в (10.11), получим:

$$\lambda(t) = \frac{w(t)}{P(t)}.$$

Величина $\lambda(t)$ называется *интенсивностью отказов*.

Учитывая (10.6) и (10.7), получаем выражение для определения $\lambda(t)$ по опытным данным:

$$\lambda^*(t) = \frac{n(t, \Delta t)}{\Delta t [N - n(t)]}. \quad (10.12)$$

Из (10.11)

$$\lambda(t) = -\frac{d}{dt} \ln P(t),$$

отсюда следует общее выражение *закона надежности*

$$P(t) = \exp \left[- \int_0^t \lambda(t) dt \right]. \quad (10.13)$$

После включения рассматриваемой группы элементов обычно имеет место приработка, во время которой отка-

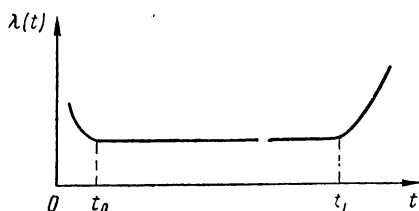


Рис. 10.3. Зависимость интенсивности отказов от времени

зывают дефектные элементы, вследствие чего интенсивность отказов нарастает. Затем при $t = t_0$ (рис. 10.3) наступает период нормальной работы, на подходе к которому $\lambda(t)$ снижается до некоторого значения, мало изменяющегося на этом участке, где имеют место лишь случайные отказы. Нормальная работа оканчивается при $t = t_1$, когда износ вновь приводит к росту $\lambda(t)$.

Пример.

Проводятся испытания $N = 500$ элементов, например реле времени. Через каждые 200 ч фиксируются отказы.

Построим по опытным данным графики функций $P^*(t)$, $w^*(t)$ и $\lambda^*(t)$. Числа отказов, регистрируемые при проверках, сведены в табл. 10.1.

Вероятность безотказной работы определяется по формуле (10.6) для каждого момента времени, совпадающего с окончанием фиксированного интервала $\Delta t = 200$ ч. Плотность вероятности и интенсивность отказов рассчитываются по (10.7) и (10.12) для каждого момента времени, совпадающего с серединой фиксированного интервала. Результаты расчетов сведены в табл. 10.1. Графики функций представлены на рис. 10.4.

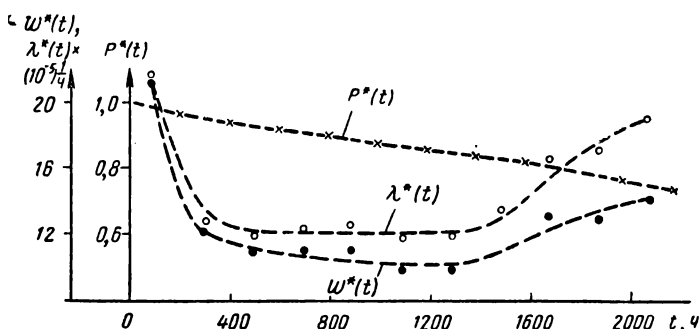


Рис. 10.4. Графики функций $P^*(t)$, $w^*(t)$ и $\lambda^*(t)$, построенные по данным табл. 10.1

Таблица 10.1

Характеристики надежности, полученные по данным испытаний

$t, \text{ч}$	$n(t), \text{отк}$	$P^*(t)$	$w^*(t) \cdot 10^{-5} \frac{1}{\text{ч}}$	$\lambda^*(t) \cdot 10^{-5} \frac{1}{\text{ч}}$
200	21	0,958	21	21,4
400	33	0,934	12	12,6
600	44	0,912	11	11,9
800	55	0,890	11	12,2
1000	66	0,868	11	12,5
1200	76	0,848	10	11,6
1400	86	0,828	10	11,9
1600	97	0,806	11	13,4
1800	110	0,780	13	16,3
2000	123	0,754	13	16,9
2200	137	0,726	14	18,9

§ 10.2. ЗАКОНЫ РАСПРЕДЕЛЕНИЯ ВРЕМЕНИ ВОЗНИКНОВЕНИЯ ОТКАЗА

В период нормальной работы, когда интенсивность отказов можно приближенно считать постоянной во времени, т. е.

$$\lambda(t) = \lambda = \text{const},$$

закон надежности (10.13) принимает вид

$$P(t) = e^{-\lambda t}. \quad (10.14)$$

Тогда по (10.8)

$$w(t) = \lambda e^{-\lambda t}. \quad (10.15)$$

Полученные выражения представляют собой *экспоненциальное распределение*.

Среднее время, получаемое из (10.9) и (10.14),

$$T = \int_0^{\infty} e^{-\lambda t} dt = \frac{1}{\lambda}. \quad (10.16)$$

При $t = T$ из (10.14) следует:

$$P(t) = \frac{1}{e} = 0,37.$$

На этом основании среднее время отказа определяется как время, в течение которого вероятность безотказной работы уменьшается до 37%.

Дисперсию, имеющую место при экспоненциальном законе надежности, можно найти из (10.10):

$$\sigma^2 = 2 \int_0^{\infty} t e^{-\lambda t} dt - \frac{1}{\lambda^2} = \frac{1}{\lambda^2},$$

т. е.

$$\sigma = T. \quad (10.17)$$

Рассмотренный экспоненциальный закон надежности (10.14) применим в условиях, при которых имеют место лишь случайные отказы, но нет отказов вследствие износа. В действительности отказы, обусловленные износом, происходят всегда и может быть лишь преобладание случайных отказов. В этом случае и применяется экспоненциальное распределение.

Если же преобладают отказы, вызываемые износом, то следует пользоваться нормальным распределением. Это распределение, часто называемое законом Гаусса, относится к случаям, в которых появление того или иного значения случайной величины зависит от большого числа случайных событий, каждое из которых оказывает на эту величину малое влияние, причем ни одно влияние не превалирует. Плотность вероятности нормаль-

ного распределения

$$w(t) = \frac{1}{\sigma \sqrt{2\pi}} \exp \left[-\frac{(t - \bar{T})^2}{2\sigma^2} \right], \quad (10.18)$$

где σ^2 — дисперсия;

\bar{T} — среднее время отказа.

Нормальный закон распределения применим для оценки надежности в том случае, если $\bar{T} \gg \sigma$, так как при этом с достаточной точностью выполняются условия

$$w(t) = \begin{cases} 0 & t < 0 \\ w(t) & t \geq 0, \end{cases} \quad (10.19)$$

$$\int_0^{\infty} w(t) dt = 1.$$

Вероятность безотказной работы при нормальном распределении определим по (10.4)

$$P(t) = \int_t^{\infty} \frac{1}{\sigma \sqrt{2\pi}} \exp \left[-\frac{(\tau - \bar{T})^2}{2\sigma^2} \right] d\tau,$$

откуда

$$P(t) = \frac{1}{\sqrt{2\pi}} \int_{\frac{t - \bar{T}}{\sigma}}^{\infty} e^{-\frac{x^2}{2}} dx.$$

Следовательно,

$$P(t) = \frac{1}{2} - \frac{1}{2} \Phi \left(\frac{t - \bar{T}}{\sigma} \right), \quad (10.20)$$

где $\Phi(x) = \frac{2}{\sqrt{2\pi}} \int_0^x e^{-\frac{t^2}{2}} dt$ — интеграл вероятности.

Принимая во внимание (10.2), получаем выражение вероятности отказа при нормальном распределении:

$$Q(t) = \frac{1}{2} + \frac{1}{2} \Phi \left(\frac{t - \bar{T}}{\sigma} \right). \quad (10.21)$$

Если условия (10.19) не выполняются, а статистические данные об отказах при $t > 0$ подчиняются закону, близкому к нормальному, то применяется усеченное нормальное рас-

пределение. В этом случае плотность вероятности [10.2]

$$w(t) = Ce^{-\frac{(t-\bar{T})^2}{2\sigma^2}}, \quad (10.22)$$

где C — постоянная, определяемая из условия (10.19):

$$C = \frac{1}{\sqrt{2\pi}\sigma \left[0,5 + \Phi\left(\frac{\bar{T}}{\sigma}\right) \right]}. \quad (10.23)$$

Вероятность безотказной работы при этом

$$P(t) = \frac{0,5 - \Phi\left(\frac{t-\bar{T}}{\sigma}\right)}{0,5 + \Phi\left(\frac{\bar{T}}{\sigma}\right)}. \quad (10.24)$$

Указанными законами распределения можно пользоваться лишь в те периоды, когда в основном имеют место отказы одного вида. Если же наблюдаются оба вида отказов и ни один из них не превалирует над другим, расчет надежности следует производить на основе *гамма-распределения* [10.3]. Тогда плотность вероятности можно взять в форме

$$w(t) = \frac{1}{\alpha! \beta} \left(\frac{t}{\beta}\right)^\alpha e^{-\frac{t}{\beta}}, \quad (10.25)$$

где α и β — положительные числа, причем α — целое. При помощи параметров α и β учитывают относительное влияние отказов по случайным причинам и вследствие износа. В случае гамма-распределения вероятность безотказной работы

$$P(t) = e^{-t/\beta} \sum_{j=0}^{\alpha} \frac{1}{j!} \left(\frac{t}{\beta}\right)^j; \quad (10.26)$$

среднее время

$$\bar{T} = \beta(\alpha + 1); \quad (10.27)$$

дисперсия

$$\sigma^2 = \beta^2(\alpha + 1). \quad (10.28)$$

При $\alpha = 0$ из последних выражений следует:

$$\bar{T} = \beta,$$

$$\sigma^2 = \bar{T}^2,$$

$$P(t) = e^{-t/\bar{T}},$$

т. е. гамма-распределение превращается в экспоненциальное, соответствующее слабому влиянию износа. Таким образом, рост влияния износа выражается в увеличении параметра α . При $T \gg \sigma$, что соответствует преобладанию отказов вследствие износа, гамма-распределение приближается к нормальному.

В ряде случаев интенсивность отказов не является постоянной даже в нормальный период работы. Так, функция $\lambda(t)$ может иметь вид

$$\lambda(t) = \alpha \lambda^{\alpha} t^{\alpha-1}. \quad (10.29)$$

При этом вероятность безотказной работы

$$P(t) = \exp [-(\lambda t)^{\alpha}], \quad (10.30)$$

плотность вероятности

$$w(t) = \alpha \lambda^{\alpha} t^{\alpha-1} \exp [-(\lambda t)^{\alpha}] \quad (10.31)$$

— *распределение Вейбулла*.

Численные значения α и λ определяются по данным испытаний. Экспоненциальный закон следует отсюда, как частный случай при $\alpha = 1$. При $\alpha = 2$ получается закон надежности, соответствующий *распределению Релея*:

$$\left. \begin{aligned} P(t) &= \exp [-(\lambda t)^2], \\ w(t) &= 2\lambda^2 t \exp [-(\lambda t)^2]. \end{aligned} \right\} \quad (10.32)$$

В эксплуатации управляющих систем приходится иметь дело с отказами того или иного вида в разные периоды работы, так как продолжительность применения таких систем обычно значительно больше среднего срока службы их элементов. Отсюда возникает необходимость получения такого закона распределения, который описывал бы распределение вероятности времени исправной работы элементов системы на всем протяжении времени ее работы.

Наиболее общей характеристикой, получаемой экспериментально и отражающей различные причины отказов, является зависимость $\lambda(t)$. Составим уравнение λ -характере-

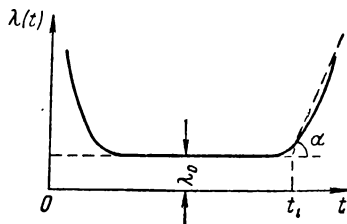


Рис. 10.5. Кусочно аппроксимированная λ -характеристика

ристики, для чего кусочно аппроксимируем ее, как показано на рис. 10.5. При этом подъем кривой в период работы учитывать не будем, так как обычно детали и устройства проходят стадию *тренировки*, во время которой неисправности и причины, вызвавшие их, устраняются. Аппроксимируя λ -характеристику двумя отрезками прямых, получаем [10.5]:

$$\lambda(t) = \lambda_0 1(t) + a(t - t_1) 1(t - t_1), \quad (10.33)$$

где λ_0 — интенсивность отказов в нормальный период работы;

$a = \operatorname{tg} \alpha$ — угловой коэффициент аппроксимирующего отрезка при $t > t_1$;

t_1 — момент времени, начиная с которого необходимо учитывать подъем характеристики.

Если зависимость $\lambda(t)$ является более сложной, необходимо аппроксимировать ее большим числом отрезков [10.6]. Подставляя (10.33) в (10.13), получаем выражение *обобщенного закона надежности*:

$$P(t) = \exp - \left[\lambda_0 t 1(t) + \frac{a}{2} (t - t_1)^2 1(t - t_1) \right], \quad (10.34)$$

откуда вероятность безотказной работы

$$P(t) = \begin{cases} e^{-\lambda_0 t} & \text{при } 0 \leq t < t_1 \\ e^{-\left[\lambda_0 t + \frac{a}{2} (t - t_1)^2 \right]} & \text{при } t \geq t_1. \end{cases} \quad (10.35)$$

При этом плотность вероятности

$$w(t) = [\lambda_0 1(t) + a(t - t_1) 1(t - t_1)] \exp - \left[\lambda_0 t 1(t) + \frac{a}{2} (t - t_1)^2 1(t - t_1) \right] \quad (10.36)$$

или

$$w(t) = \begin{cases} \lambda_0 e^{-\lambda_0 t} & \text{при } 0 \leq t < t_1 \\ [\lambda_0 + a(t - t_1)] \exp - \left[\lambda_0 t + \frac{a}{2} (t - t_1)^2 \right] & \text{при } t \geq t_1. \end{cases} \quad (10.37)$$

Уравнение (10.36) выражает *обобщенное распределение вероятностей времени безотказной работы*.

Исследование выражений (10.35) и (10.37) показывает [10.5], что $P(t)$ и $\omega(t)$ изменяются по экспоненциальному

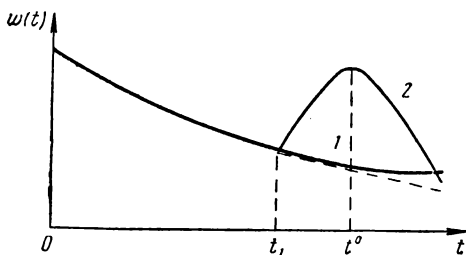


Рис. 10.6. Графики функций $\omega(t)$

закону при $0 \leq t < t_1$. Далее при $t \geq t_1$ с ростом времени убывают $P(t)$ и $\omega(t)$, если $\lambda_0^3 \geq a$ (рис. 10.6 и 10.7, кри-

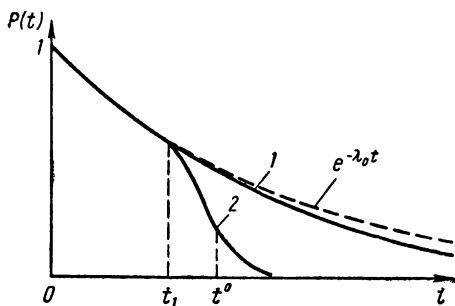


Рис. 10.7. Графики функций $P(t)$

вая 1). При $\lambda_0^3 < a$ функция $\omega(t)$, начиная с $t=t_1$, растет и имеет максимум в момент времени $t=t^0$ (см. рис. 10.6, кривая 2), причем

$$t^0 = t_1 - \frac{\lambda_0}{a} + \frac{1}{\sqrt{a}}. \quad (10.38)$$

Максимальное значение

$$\omega(t^0) = \sqrt{a} \exp - \left(\lambda_0 t_1 - \frac{1}{2} \frac{\lambda_0^3}{a} + \frac{1}{2} \right). \quad (10.39)$$

Функция $P(t)$ убывает и при $t=t^0$ имеет точку перегиба (см. рис. 10.7, кривая 2).

Среднее время \bar{T} при наличии обобщенного распределения вычисляем по (10.9):

$$\bar{T} = \frac{1}{\lambda_0} - \frac{1}{\lambda_0} e^{-\lambda_0 t_1} + \sqrt{\frac{\pi}{2a}} e^{-\lambda_0 \left(t_1 - \frac{\lambda_0}{2a}\right)} \left[1 - \Phi\left(\frac{\lambda_0}{\sqrt{a}}\right)\right]. \quad (10.40)$$

Введя обозначения:

$$\frac{1}{\lambda_0} = \bar{T}_{\text{экс}}, \quad \frac{\lambda_0}{\sqrt{a}} = \eta, \quad \frac{t_1}{\bar{T}_{\text{экс}}} = \theta, \quad (10.41)$$

получаем:

$$\bar{T} = \bar{T}_{\text{экс}} \left\{ 1 - e^{-\theta} + \sqrt{\frac{\pi}{2}} \eta e^{-\left(\theta - \frac{1}{2} \eta^2\right)} [1 - \Phi(\eta)] \right\}. \quad (10.42)$$

На рис. 10.8 представлены зависимости \bar{T} от η при различных значениях θ . Исследование этих кривых приводит к следующим выводам [10.5].

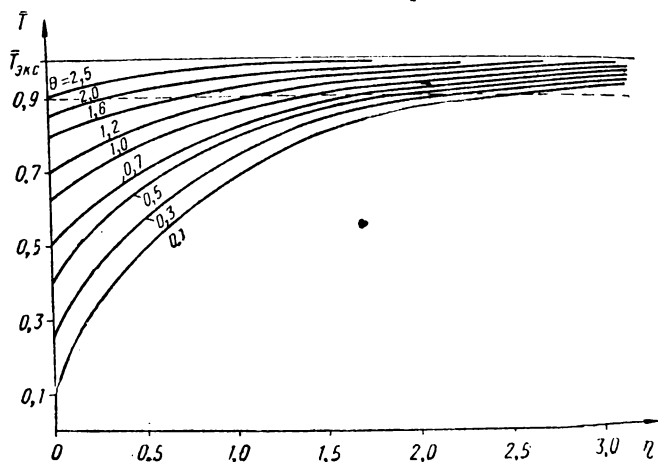


Рис. 10.8. Зависимости $\bar{T} = \bar{T}_{\text{экс}} \varphi(\theta, \eta)$

1. При $\theta \geq 2,5$ практически можно не учитывать отказы, возникающие вследствие износа, каково бы ни было отношение λ_0/\sqrt{a} , так как $0,9\bar{T}_{\text{экс}} \leq \bar{T} \leq \bar{T}_{\text{экс}}$. В этом случае можно приближенно считать, что закон распределения является экспоненциальным при любых значениях времени.

2. При $\eta \geq 2,7$ независимо от того, какое значение имеет t_1 , можно также не учитывать отказы, возникающие вследствие износа. Это утверждение справедливо уже для $t_1 \geq 0,1 \bar{T}_{\text{экс}}$. Таким образом, несмотря на то, что в этом случае отказы, вызванные износом, появляются достаточно рано, они настолько редки, что характеристики надежности можно определять по экспоненциальному закону. Дисперсия, определяемая по (10.10):

$$\sigma^2 = \frac{2}{a} e^{-\lambda_0 t_1} + \frac{a t_1 - \lambda_0}{a} \sqrt{\frac{2\pi}{a}} e^{-\left(\lambda_0 t_1 - \frac{\lambda_0^2}{2a}\right)} \left[1 - \Phi\left(\frac{\lambda_0}{\sqrt{a}}\right)\right] + \\ + \frac{2[1 - e^{-\lambda_0 t_1}(1 + \lambda_0 t_1)]}{\lambda_0^2} - \left\{ \frac{1}{\lambda_0} (1 - e^{-\lambda_0 t_1}) + \right. \\ \left. + \sqrt{\frac{\pi}{2a}} e^{-\left(\lambda_0 t_1 - \frac{\lambda_0^2}{2a}\right)} \left[1 - \Phi\left(\frac{\lambda_0}{\sqrt{a}}\right)\right] \right\}^2. \quad (10.43)$$

Это выражение значительно упрощается, если произведение $\lambda_0 t_1$ настолько мало, что приближенно можно принять:

$$\left. \begin{aligned} e^{-\lambda_0 t_1} &\approx 1 - \lambda_0 t_1, \\ \Phi\left(\frac{\lambda_0}{\sqrt{a}}\right) &\approx 0, \\ e^{-\left[\lambda_0 t_1 - \frac{\lambda_0^2}{2a}\right]} &\approx 1. \end{aligned} \right\} \quad (10.44)$$

Тогда из (10.43) получаем:

$$\sigma^2 \approx t_1^2 - 2 \frac{\lambda_0}{a} t_1 + \frac{2}{a} - \frac{\pi}{2a}. \quad (10.45)$$

Под воздействием внешних условий, например температуры, влажности и т. п., а также при различных режимах работы параметры законов распределения, которым подчиняется надежность элементов, изменяются. Это необходимо учитывать при расчете надежности элементов путем введения поправок, которые определяются экспериментально или подсчитываются по специальным, чаще всего эмпирическим, формулам.

Характеристики надежности реле обычно определяются в функции числа циклов работы n . Время t и число циклов работы n связаны через частоту f :

$$t = n \frac{1}{f}. \quad (10.46)$$

Пример. Контактная пара реле РЭС-6 работает с частотой 5 гц, коммутируя ток 2а. λ -характеристика, соответствующая этому режиму работы, приведена на рис. 10.9. Построим зависимости $w[n]$ и $P[n]$, являющиеся решетчатыми функциями числа циклов (α) работы. Число циклов работы n связано с временем t соотношением (10.46).

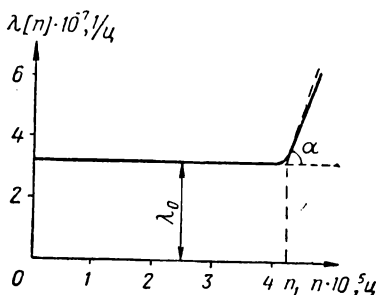


Рис. 10.9. λ -характеристика контактной пары реле РЭС-6

При расчете характеристик надежности контактной пары реле РЭС-6 необходимо применить обобщенное распределение. Параметры распределения, определенные по λ -характеристике:

$$\lambda_0 = 3,3 \cdot 10^{-7} \frac{1}{\alpha},$$

$$a = 7 \cdot 10^{-12} \frac{1}{\alpha^2},$$

$$n_1 = 4,1 \cdot 10^5 \text{ ц.}$$

Для данных значений параметров выполняется неравенство

$$\lambda_0^2 < a,$$

следовательно, $w[n]$, начиная с n_1 , растет, достигая максимума, согласно (10.38), при

$$n^* = 4,1 \cdot 10^5 + \frac{1}{\sqrt{7 \cdot 10^{-12}}} - \frac{3,3 \cdot 10^{-7}}{7 \cdot 10^{-12}} = 7,2 \cdot 10^5 \text{ ц.}$$

Максимальное значение функции $w[n]$ определяется из (10.39):

$$w[n^*] = \sqrt{7 \cdot 10^{-12}} e^{-\left(3,3 \cdot 10^{-7} \cdot 4,1 \cdot 10^5 - 0,5 \frac{10,8 \cdot 10^{-14}}{7 \cdot 10^{-12}} + 0,5\right)} = 14 \cdot 10^{-7} \frac{1}{\alpha}.$$

Значения $w[n]$ и $P[n]$ вычисляются по (10.35) и (10.37); они представлены на рис. 10.10.

Среднее время T найдем по графикам рис. 10.8. Для этого вычислим:

$$\theta = 4,1 \cdot 10^5 \cdot 3,3 \cdot 10^{-7} = 0,135; \quad \eta = \frac{3,3 \cdot 10^{-7}}{2,65 \cdot 10^{-6}} = 0,124.$$

При полученных θ и η имеем:

$$T = 0,25 T_{\text{экс}} = 0,25 \frac{1}{3,3 \cdot 10^{-7}} = 7,55 \cdot 10^6 \text{ ц}.$$

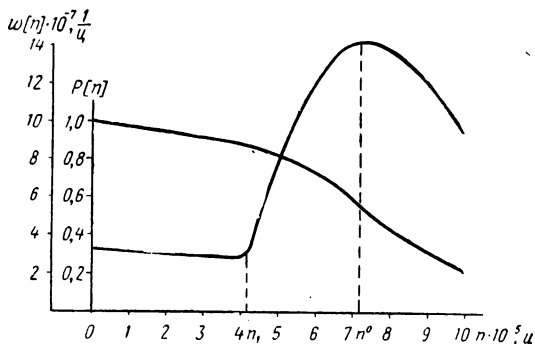


Рис. 10.10. Функции $P[n]$ и $w[n]$ контактной пары реле РЭС-6

Дисперсия

$$\sigma^2 = (4,1 \cdot 10^5)^2 - 2 \cdot 4,1 \cdot 10^5 \frac{3,3 \cdot 10^{-7}}{7 \cdot 10^{-12}} + \frac{1}{2 \cdot 7 \cdot 10^{-12}} = 18,49 \text{ ц}^2.$$

§ 10.3. ХАРАКТЕРИСТИКИ НАДЕЖНОСТИ СИСТЕМ

Надежность управляющей системы зависит от способа соединения элементов, из которых она состоит. Соединение, при котором отказ любого элемента вызывает отказ всей системы, называется *основным*. Если система правильно функционирует, несмотря на отказ ряда элементов, функции которых начинают выполняться другими элементами, то такое соединение называется *резервным*.

Рассмотрим расчет надежности систем по характеристикам надежности составляющих их элементов, находящихся в основном соединении. Предположим, что отказы всех элементов — случайные независимые события. Вероятность безотказной работы системы при этом [10.3]

$$P_c(t) = \prod_{i=1}^k P_i(t), \quad (10.47)$$

где k — число элементов системы;

$P_i(t)$ — вероятность безотказной работы i -го элемента.

Вероятность отказа такой системы

$$Q_c(t) = 1 - \prod_{i=1}^k P_i(t), \quad (10.48)$$

или

$$Q_c(t) = 1 - \prod_{i=1}^k [1 - Q_i(t)], \quad (10.49)$$

где $Q_i(t)$ — вероятность отказа i -го элемента.

Из (10.49) следует, что при малых Q_i , когда можно пренебречь произведениями этих величин,

$$Q_c(t) = \sum_{i=1}^k Q_i(t). \quad (10.50)$$

В нормальный период работы, когда имеет место экспоненциальный закон надежности, в соответствии с (10.14) и (10.47) получаем:

$$P_c(t) = \exp\left(-\sum_{i=1}^k \lambda_i t\right). \quad (10.51)$$

Величину

$$\lambda_c = \sum_{i=1}^k \lambda_i \quad (10.52)$$

называют *интенсивностью отказов системы* (или устройства, блока), состоящей из k элементов, находящихся в основном соединении. Среднее время \bar{T} такой системы по (10.9) и (10.52)

$$\bar{T}_c = \left(\sum_{i=1}^k \frac{1}{\bar{T}_i} \right)^{-1}, \quad (10.53)$$

где \bar{T}_i — среднее время \bar{T} i -го элемента.

Если закон распределения вероятности времени безотказной работы элементов, входящих в систему, имеет сложное выражение или надежность элементов характеризуется различными законами, удобно применять обобщенное распределение. Оно позволяет представить в одинаковой форме выражения характеристик надежности различных элементов, что значительно упрощает расчет

надежности систем [10.7]. По (10.34)

$$P_c(t) = \prod_{i=1}^k \exp - \left[\lambda_0 t \cdot 1(t) + \frac{a}{2} (t - t_1)^2 \cdot 1(t - t_1) \right]. \quad (10.54)$$

Если все элементы равнонадежны, то

$$P_c(t) = \exp - k \left[\lambda_0 t \cdot 1(t) + \frac{a}{2} (t - t_1)^2 \cdot 1(t - t_1) \right], \quad (10.55)$$

откуда

$$P_c(t) = \begin{cases} e^{-k\lambda_0 t}, & 0 \leq t < t_1 \\ e^{-k \left[\lambda_0 t + \frac{a}{2} (t - t_1)^2 \right]}, & t \geq t_1. \end{cases} \quad (10.56)$$

Среднее значение времени T , получаемое из (10.9) и (10.55),

$$\begin{aligned} \bar{T}_c = \frac{1}{k\lambda_0} \left\{ 1 - e^{-k\lambda_0 t_1} + \sqrt{\frac{\pi}{2ak}} \cdot k\lambda_0 \left[1 - \Phi \left(\sqrt{k} \frac{\lambda_0}{\sqrt{a}} \right) \right] \times \right. \\ \left. \times \exp - k \left(\lambda_0 t_1 - \frac{\lambda_0^2}{2a} \right) \right\}. \end{aligned} \quad (10.57)$$

Введем обозначения:

$$\frac{1}{k\lambda_0} = \bar{T}_{c, \text{экс}}; \quad \frac{t_1}{T_{c, \text{экс}}} = \theta_c; \quad \sqrt{k} \frac{\lambda_0}{\sqrt{a}} = \eta_c, \quad (10.58)$$

тогда

$$\begin{aligned} \bar{T}_c = \bar{T}_{c, \text{экс}} \left\{ 1 - e^{-\theta_c} + \sqrt{\frac{\pi}{2}} \eta_c \left[1 - \right. \right. \\ \left. \left. - \Phi(\eta_c) \right] \exp - \left(\theta_c - \frac{1}{2} \eta_c \right) \right\}. \end{aligned} \quad (10.59)$$

Выражение (10.59) аналогично (10.42), график которого представлен на рис 10.8. Следовательно, можно считать, что:

а) при $\theta_c \geq 2,5$ для расчета надежности системы можно применять экспоненциальное распределение и не учитывать отказы, возникающие вследствие износа;

б) при $\eta_c \geq 2,7$ можно также не учитывать отказы вследствие износа.

Из (10.41) и (10.58) видно, что

$$\theta_c = k\theta,$$

следовательно, уравнение

$$k\theta = 2,5 \quad (10.60)$$

описывает границу, отделяющую область Э (рис. 10.11), для которой справедливо экспоненциальное распределение. Например, если $\theta = 0,3$, то уже начиная с $k = 8$ при расчете характеристик надежности систем можно не учитывать отказы вследствие износа, т. е. можно при-

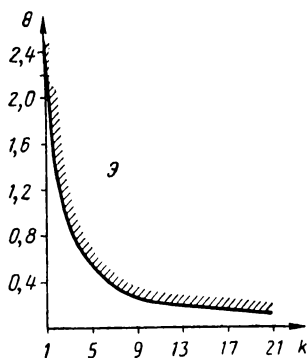


Рис. 10.11. Область значений T

менять экспоненциальное распределение. Следует еще раз заметить, что при расчете характеристик надежности систем необходимо учитывать условия эксплуатации элементов (см. § 10.2).

Вероятность безотказной работы $P_c(t)$ и среднее время T_c , вычисляемые по вышеприведенным формулам, являются основными характеристиками надежности *невосстанавливаемых систем*, не ремонтируемых после возникновения отказа.

В *восстанавливаемых системах* возникающие отказы устраняются, неисправные элементы и узлы заменяются новыми. Рассмотренные выше характеристики определяют надежность восстанавливаемых систем только до первого отказа. Последовательность отказов, возникающая в эксплуатации таких систем, образует *поток отказов*.

Интенсивность потока отказов системы, называемая также суммарной частотой отказов [10.2],

$$\Lambda_c(t) = \sum_{i=1}^k \Lambda_i(t), \quad (10.61)$$

здесь $\Lambda_i(t)$ — интенсивность потока отказов i -го элемента; k — число элементов в системе.

Интенсивность потока отказов элемента определяется отношением числа элементов, отказавших в единицу времени, к числу испытываемых элементов при условии, что все неисправные элементы заменяются новыми. Статистическая оценка $\Lambda_i(t)$ находится из выражения

$$\Lambda_i(t)^* = \frac{n(t, \Delta t)}{N_0 \Delta t}, \quad (10.62)$$

где $n(t, \Delta t)$ — число элементов, отказавших в интервале времени от $t - \frac{\Delta t}{2}$ до $t + \frac{\Delta t}{2}$;

N_0 — число испытываемых элементов.

При $t \rightarrow \infty$ интенсивность потока отказов стремится к постоянной

$$\Lambda_i = \frac{1}{\bar{T}_i}, \quad (10.63)$$

где \bar{T}_i — среднее время \bar{T} элемента. При экспоненциальном законе надежности $\Lambda_i = \lambda_i$.

Таким образом, для периода нормальной работы интенсивность потока отказов совпадает с интенсивностью отказов λ_i .

Принимая во внимание, что средний срок службы элемента \bar{T} , введем величину $\bar{\Lambda}_{\text{ст}}$, характеризующую усредненное значение интенсивности потока отказов в интервале $t_1 \leq t \leq \bar{T}$.

Рассмотрим группу, состоящую из k одинаковых элементов. За время от 0 до t_1 откажет в среднем k_1 элементов:

$$k_1 = k \lambda_0 t_1,$$

за время от t_1 до \bar{T} — k_2 элементов:

$$k_2 = k \bar{\Lambda}_{\text{ст}} (\bar{T} - t_1),$$

причем $k_1 + k_2 = k$. Следовательно,

$$1 = [\lambda_0 t_1 + \bar{\Lambda}_{\text{ст}} (\bar{T} - t_1)],$$

откуда

$$\bar{\Lambda}_{\text{ст}} = \frac{1 - \lambda_0 t_1}{\bar{T} - t_1}.$$

Принимая во внимание (10.41) и вводя обозначение $v = \frac{\bar{T}}{\bar{T}_{\text{экс}}}$, получаем:

$$\bar{\Lambda}_{\text{ст}} = \lambda_0 \left(\frac{1 - \theta}{v - \theta} \right). \quad (10.64)$$

Таким образом, в интервале от 0 до t_1 интенсивность потока отказов равна постоянной величине λ_0 , а при $t_1 \leq t \leq \bar{T}$ ее среднее значение характеризуется $\bar{\Lambda}_{\text{ст}}$. Сле-

довательно, для любого момента времени в интервале от 0 до T выражение интенсивности потока отказов $\Lambda(t)$ имеет вид (рис. 10.12)

$$\Lambda(t) = \lambda_0 \left[1(t) - 1(t - t_1) + \frac{1 - \theta}{1 - \theta} 1(t - t_1) \right]. \quad (10.65)$$

Значения *среднего числа отказов за время t и среднего времени между отказами (наработка на отказ)* в интервале от 0 до t определяются по формулам [10.2]

$$\bar{m} = t\Lambda_c(t), \quad (10.66)$$

$$\bar{t} = \frac{1}{\Lambda_c(t)}, \quad (10.67)$$

или

$$\bar{t} = \left[\sum_{i=1}^k \left(\frac{1}{\bar{T}_i} \right) \right]^{-1}, \quad (10.68)$$

где \bar{T}_i — среднее время i -го элемента.

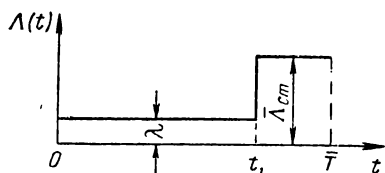


Рис. 10.12. Зависимость $\Lambda(t)$

Приведенные выше количественные характеристики позволяют оценивать надежность систем в процессе эксплуатации без учета времени, затрачиваемого на профилактические мероприятия и ремонт. Учитывать эти затраты времени можно с

помощью коэффициентов надежности. Наиболее часто применяется *коэффициент готовности*:

$$K_r = \frac{\bar{t}}{\bar{t} + \bar{t}_b}, \quad (10.69)$$

где \bar{t}_b — среднее время восстановления.

По физической сущности K_r — вероятность исправного состояния аппаратуры в любой момент времени.

Вследствие отклонений параметров реальных элементов от номинальных или расчетных данных при первом включении автоматическая система может дать отказ. Вероятность того что система окажется работоспособной в момент включения, характеризует *конструктивную надежность* системы.

Расчет конструктивной надежности заключается в определении вероятности соответствия системы ее обобщенной характеристике — *обобщенному функционалу*, в качестве которого удобно рассматривать передаточную функцию $W(p)$ или частотную характеристику [10.1]. Решение этой задачи получается наиболее простым, если вероятность соответствия определяется по модулю и фазе частотной характеристики [10.3]. Модуль $M(\omega)$ и фаза $\varphi(\omega)$ выражаются через параметры системы:

$$\left. \begin{aligned} M(\omega) &= f_1(x_1, x_2, x_3, \dots, x_n), \\ \varphi(\omega) &= f_2(x_1, x_2, x_3, \dots, x_n). \end{aligned} \right\} \quad (10.70)$$

Если параметры системы, как случайные величины, являются независимыми, то абсолютные значения отклонений

$$\left. \begin{aligned} \Delta M &= \sum_{i=1}^n \frac{\partial f_1(x_1, x_2, \dots, x_n)}{\partial x_i} \Delta x_i, \\ \Delta \varphi &= \sum_{i=1}^n \frac{\partial f_2(x_1, x_2, \dots, x_n)}{\partial x_i} \Delta x_i. \end{aligned} \right\} \quad (10.71)$$

Известно [10.3], что ΔM и $\Delta \varphi$ подчиняются нормальному распределению. Тогда вероятности соответствия модуля и фазы системы заданным значениям определяются по формулам

$$\left. \begin{aligned} P_M \{ \Delta M < \Delta M_{\text{доп}} \} &= \Phi \left(\frac{\Delta M_{\text{доп}} - \Delta \bar{M}}{\sigma_{\Delta M}} \right), \\ P_\varphi \{ \Delta \varphi < \Delta \varphi_{\text{доп}} \} &= \Phi \left(\frac{\Delta \varphi_{\text{доп}} - \Delta \bar{\varphi}}{\sigma_{\Delta \varphi}} \right). \end{aligned} \right\} \quad (10.72)$$

Здесь $\Delta M_{\text{доп}}$, $\Delta \varphi_{\text{доп}}$ — допустимые пределы изменения модуля и фазы;

$\Delta \bar{M}$, $\Delta \bar{\varphi}$ — средние значения отклонений модуля и фазы;

$\sigma_{\Delta M}$, $\sigma_{\Delta \varphi}$ — средние квадратичные отклонения.

Примеры. 1. Определим среднее значение времени до первого отказа и среднее значение времени между отказами при $t \rightarrow \infty$ узла аппаратуры, состоящего из пяти пар контактов реле типа РЭС-8. Контакты срабатывают с частотой 5 гц, коммутируя ток 0,3 а при напряжении 220 в. λ -характеристика контакта реле, работающего в этом режиме, приведена на рис. 10.13.

Для расчета характеристик надежности узла воспользуемся обобщенным распределением. Параметры распределения, полученные по λ -характеристике:

$$\lambda_0 = 0,95 \cdot 10^{-7} \frac{1}{\text{ч}},$$

$$a = 10^{-12} \frac{1}{\text{ч}^2},$$

$$n_1 = 8,7 \cdot 10^5 \text{ ч.}$$

По (10.41) находим:

$$\theta = 8,7 \cdot 10^5 \cdot 0,95 \cdot 10^{-7} = 0,08.$$

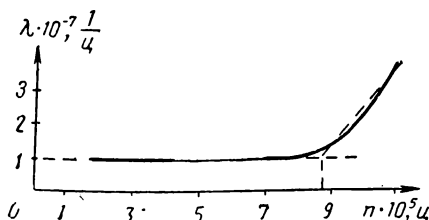


Рис. 10.13. λ -характеристика контактной пары реле РЭС-8

Так как точка k, θ не попадает в область Э, значение среднего времени отказа должно быть определено по общей формуле (10.59). По (10.58)

$$T_{\text{с.экс}} = \frac{1}{5 \cdot 0,95 \cdot 10^{-7}} = 21 \cdot 10^5 \text{ ч},$$

$$\theta_c = \frac{8,7 \cdot 10^5}{21 \cdot 10^5} = 0,41,$$

$$\eta_c = \frac{0,95 \cdot 10^{-7}}{\sqrt{10^{-12}}} \sqrt{5} = 0,21.$$

Тогда

$$T_c = 21 \cdot 10^5 \left\{ 1 - e^{-0,41} + \sqrt{\frac{3,14}{2}} 0,21 [1 - \Phi(0,21)] e^{-(0,41 - \frac{0,21}{2})} \right\} = 10,3 \cdot 10^5 \text{ ч.}$$

Среднее значение времени между отказами определяется по (10.68). Для этого при $T_{\text{экс}} = 105 \cdot 10^5 \text{ ч}$, $\theta = 0,08$, $\eta = 0,095$ по (10.42) найдем:

$$T = 105 \cdot 10^5 \left\{ 1 - e^{-0,08} + \sqrt{\frac{3,14}{2}} \cdot 0,095 e^{-(0,08 - \frac{0,095}{2})} \times [1 - \Phi(0,095)] \right\} = 19,6 \cdot 10^5 \text{ ч.}$$

Следовательно,

$$\bar{t} = \left[\sum_{i=1}^8 \left(\frac{1}{19,6 \cdot 10^8} \right) \right]^{-1} = 4 \cdot 10^8 \text{ } \mu.$$

2. Определим конструктивную надежность по модулю звена, имеющего передаточную функцию (4.69). Модуль должен находиться в пределах допуска $\delta M_{\text{доп}} = \pm 5\%$. Коэффициент усиления k звена имеет допуск $\delta k = \pm 20\%$. Постоянная времени $\tau = 0,07 \text{ сек}$ характеризуется допуском $\delta \tau = \pm 7\%$. Несущая частота $\omega = 50 \text{ гц}$ стабилизирована с точностью $\delta \omega = \pm 5\%$.

Модуль частотной характеристики данного звена

$$M(\omega) = \frac{K}{\sqrt{1 + \tau^2 \omega^2}}.$$

Изменение модуля в зависимости от изменений параметров по (10.71)

$$\Delta M = \frac{1}{\sqrt{1 + \tau^2 \omega^2}} \Delta K - \frac{K}{\sqrt{1 + \tau^2 \omega^2}} \cdot \frac{\omega^2 \tau}{(1 + \tau^2 \omega^2)} \Delta \tau - \\ - \frac{K}{\sqrt{1 + \tau^2 \omega^2}} \cdot \frac{\omega \tau^2}{(1 + \tau^2 \omega^2)} \Delta \omega$$

или в относительных значениях

$$\delta M = \frac{\Delta M}{M}, \quad \delta K = \frac{\Delta K}{K} \quad \text{и т. д.}$$

Отсюда

$$\delta M = \delta K - \frac{\omega^2 \tau^2}{1 + \tau^2 \omega^2} \delta \tau - \frac{\omega^2 \tau^2}{1 + \tau^2 \omega^2} \delta \omega.$$

Подставляя численные значения, получаем:

$$\delta M = \pm 20 - \frac{(2 \cdot 3,14 \cdot 50)^2 (0,07)^2}{1 + (2 \cdot 3,14 \cdot 50)^2 (0,07)^2} (\pm 7) - \\ - \frac{(2 \cdot 3,14 \cdot 50)^2 (0,07)^2}{1 + (2 \cdot 3,14 \cdot 50)^2 (0,07)^2} (\pm 5) = \pm 8,07\%.$$

Среднее значение

$$\delta M = \frac{\delta M_{\text{max}} + \delta M_{\text{min}}}{2}.$$

В рассматриваемом случае

$$\delta M = \frac{8,07 - 8,07}{2} = 0.$$

На основании правила, по которому

$$\text{Вер} \{ |\delta M_{\text{max}} - \delta \bar{M}| < 3\sigma_{\delta M} \} = 0,997,$$

для определения $\sigma_{\Delta M}$ можно воспользоваться приближенным соотношением

$$\sigma_{\Delta M} \approx \frac{\delta M_{\max} - \delta \bar{M}}{3}.$$

Следовательно,

$$\sigma_{\Delta M} \approx \frac{8,07}{3} = 2,69\%.$$

Теперь по (10.72) находим, что конструктивная надежность узла по модулю оценивается вероятностью

$$P_M \{ \delta M < \delta M_{\text{доп}} \} = \Phi \left(\frac{5 - 0}{2,69} \right) = 0,928.$$

§ 10.4. РЕЗЕРВИРОВАНИЕ

Несмотря на то, что техника развивается в направлении создания все более надежных устройств, нет элементов, работающих абсолютно надежно, т. е. таких, вероятность безотказной работы которых тождественно равна

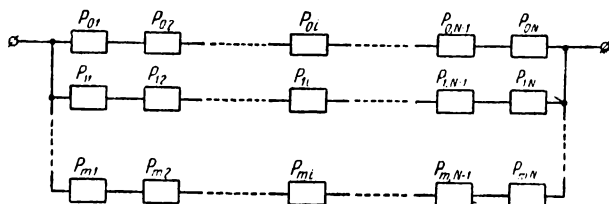


Рис. 10.14. Общее резервирование системы с кратностью m

единице. Для любого реального элемента $P(t)$ является убывающей функцией времени, и повышение надежности элемента приводит лишь к замедлению ее убывания.

Таким образом, система, состоящая из реальных элементов, не может быть абсолютно надежной. Однако важно, чтобы при эксплуатации систем вероятность их безотказной работы за определенный промежуток времени была достаточно высокой. Повышение надежности систем можно достигнуть посредством *резервирования*. Этот метод повышения надежности приводит к созданию *избыточных систем*, содержащих больше элементов, чем было бы нужно, если бы элементы были абсолютно надежными.

Если резервные элементы включены параллельно и работают, дублируя один другого, то такое резервирование называется *постоянным*. Если резервные элементы вступают в работу после отказа основных, то имеет место *резервирование замещением*. Такие способы резервирования можно осуществить путем *общего и раздельного соединения* резервных элементов. При общем соединении резервируется вся система в целом. Отношение числа резервных цепей к числу работающих называется *кратностью резервирования* и обозначается буквой m . На рис. 10.14 представлена система, состоящая из mN элементов, зарезервированная постоянно с кратностью m .

Вероятность отказа такого устройства

$$Q_y = \prod_{i=0}^m Q_{iц}, \quad (10.73)$$

где $Q_{iц}$ — вероятность отказа отдельной цепи, состоящей из N элементов.

Вероятность безотказной работы системы

$$P_y = 1 - \prod_{i=0}^m Q_{iц} \quad (10.74)$$

или

$$P_y = 1 - \prod_{i=0}^m (1 - P_{iц}). \quad (10.75)$$

Так как N элементов каждой цепи рассматриваемой системы находятся в основном соединении, то вероятность безотказной работы одной цепи

$$P_{iц} = \prod_{j=1}^N P_{ij}, \quad (10.76)$$

где P_{ij} — вероятность безотказной работы j -го элемента i -й цепи.

Подставляя (10.76) в (10.75), имеем:

$$P_y = 1 - \prod_{i=0}^m \left(1 - \prod_{j=1}^N P_{ij} \right). \quad (10.77)$$

Если все элементы равнонадежны, то

$$P_y = 1 - (1 - P^N)^{m+1}. \quad (10.78)$$

Резервирование системы по отдельным участкам называется *раздельным*. При этом резервироваться могут как сравнительно крупные блоки и узлы арматуры, так и отдельные элементы и даже внутриэлементные связи. На рис. 10.15 представлена избыточная система, состоя-

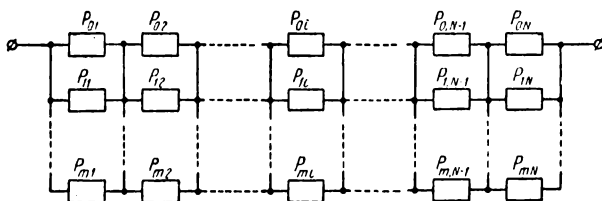


Рис. 10.15. Раздельное резервирование элементов системы с кратностью m

щая из N элементов, каждый из которых имеет m резервных. Если отказы всех элементов являются событиями случайными и независимыми, то вероятность безотказной работы такой системы

$$P_{\text{раз}} = \prod_{i=1}^N \left[1 - \prod_{j=0}^m (1 - P_{ij}) \right]. \quad (10.79)$$

В случае равнонадежных элементов

$$P_{\text{раз}} = [1 - (1 - P)^{m+1}]^N. \quad (10.80)$$

Сравнение эффективности общего и раздельного резервирования [10.3] показывает, что при одинаковом общем числе элементов mN и $Q \ll 1$ вероятность отказа системы при общем резервировании в n^{m-1} раз больше, чем при раздельном, т. е. раздельное резервирование позволяет получить более надежную систему, чем общее.

Следует заметить, что для повышения надежности систем нет необходимости резервировать все элементы с одинаковой кратностью. В реальных системах элементы по надежности различаются друг от друга и иногда очень сильно. Поэтому кратность резервирования подбирается так, чтобы все группы однотипных элементов обладали примерно одинаковой надежностью.

При введении постоянного резервирования следует учитывать, что отказы большинства элементов (например,

сопротивлений, диодов, реле и т. д.) могут быть двух видов: «обрыв» и «короткое замыкание». В зависимости от вида отказа резервные элементы присоединяются различно. Если элементы выходят из строя в основном из-за обрыва, а короткое замыкание — явление чрезвычайно редкое, то резервные элементы следует подключать параллельно. В противном случае, т. е. при отказах типа «короткое замыкание», необходимо резерв включать последовательно. Если же отказы элементов бывают того и другого вида, то применяются комбинированные, параллельно-последовательные соединения. Отдельный элемент заменяется совокупностью таких же элементов, функционирующих подобно ему. При этом следует учитывать, что параллельное соединение, являясь резервным для отказов типа «обрыв», приводит к увеличению относительного влияния отказов типа «короткое замыкание» и, наоборот, последовательное — к росту влияния отказов типа «обрыв».

Вероятность безотказной работы комбинированных схем по отношению к отказам определенного вида можно вычислять, применяя формулы (10.78) и (10.80). Однако по мере усложнения схем эти расчеты становятся чрезвычайно трудоемкими. Рассмотрим более общий прием, который применим к схемам любой сложности [10.3].

В комбинированной схеме, состоящей из k элементов, характеризуемых вероятностями безотказной работы P_1, P_2, \dots, P_k , имеется S путей передачи сигнала с входа на выход. Пути передачи сигнала, образуемые основным соединением элементов, соединены резервно. Для безотказной работы схемы достаточно безотказной работы элементов какого-либо одного пути при отказе всех остальных элементов: например, пути с номером j , который включает n элементов. Вероятность этого случая равна произведению вероятностей безотказной работы n элементов j -го пути, умноженному на произведения вероятностей отказа $(k - n)$ остальных элементов. Но передача сигнала по данному пути может быть выполнена и при безотказной работе одного, двух и т. д. до $(k - n)$ элементов, не находящихся на рассматриваемом пути. Вероятность каждого из этих случаев равна произведению вероятностей безотказной работы элементов, работающих исправно, умноженному на произведения вероятностей отказа остальных элементов. Вероятность же передачи сигнала по рассматриваемому пути равна сумме вероятностей всех возможных случаев

использования этого пути. Вероятность безотказной работы соединения равна сумме вероятностей прохождения сигнала по всем S путям:

$$P_c(t) = \sum_{j=1}^S \sum_{l=0}^{k-n} a_{lj} \prod_{i=1}^{n+l} P_i(t) \prod_{i=n+l+1}^k Q_i(t). \quad (10.81)$$

Здесь l — число безотказно работающих блоков, не находящихся на пути, для которого вычисляются произведения вероятностей.

Каждый элемент имеет номер i , изменяющийся от 1 до $n+l$ для каждого члена внутренней суммы. Коэффициент a_{lj} равен единице или нулю: $a_{lj} = 1$ для всех членов внутренней суммы при $j=1$, т. е. при суммировании по первому рассматриваемому пути; при суммировании по остальным путям следует полагать $a_{lj} = 0$ для всех слагаемых, которые встречались на предыдущих путях.

При $l = k - n$ для всех путей

$$a_{lj} \prod_{i=1}^{n+l} P_i(t) \prod_{i=n+l+1}^k Q_i(t) = a_{lj} \prod_{i=1}^k P_i(t).$$

В этом случае $a_{lj} = 1$ для одного пути, а для остальных необходимо брать $a_{lj} = 0$.

Для соединения, состоящего из элементов, имеющих одинаковую надежность, P_i и Q_i не зависят от номера элемента и могут быть обозначены P_s и Q_s . Тогда, вводя общее число безотказно работающих элементов $h = n + l$ из (10.81), получаем:

$$P_c(t) = \sum_{h=1}^k A_h P_s^h(t) Q_s^{k-h}(t), \quad (10.82)$$

здесь A_h — число возможных путей передачи сигналов, когда h элементов работают безотказно.

Вероятность отказа соединения

$$Q_c(t) = 1 - P_c(t).$$

Для вычисления $Q_c(t)$ можно получить другое выражение. Для этого необходимо рассмотреть условия, при которых сигнал не может пройти через схему из-за отказов некоторого количества элементов. Допустим, что в схеме имеется r групп элементов, причем группы образуют основное соединение, а элементы в каждой группе соеди-

нены резервно. Отказ соединения происходит при отказе всех u элементов, входящих в какую-либо группу. Следовательно, вероятность отказа соединения равна сумме вероятностей отказа каждой из r групп. Обозначая через j номер группы, а через y — число отказавших блоков, не входящих в рассматриваемую группу, получаем:

$$Q_c(t) = \sum_{j=1}^r \sum_{y=0}^{k-u} b_{yj} \prod_{i=1}^{u+y} Q_i(t) \prod_{i=u+y+1}^k P_i(t), \quad (10.83)$$

здесь коэффициент b_{yj} аналогично коэффициенту a_{ij} в (10.81) равен единице или нулю.

Для соединения, состоящего из элементов, характеризующихся одинаковыми значениями $P_g(t)$ и $Q_g(t)$, получаем:

$$Q_c(t) = \sum_{g=1}^k B_g P_g^{k-g}(t) Q_g^g(t), \quad (10.84)$$

где B_g — число возможных групп из g элементов, отказ которых вызывает отказ схемы.

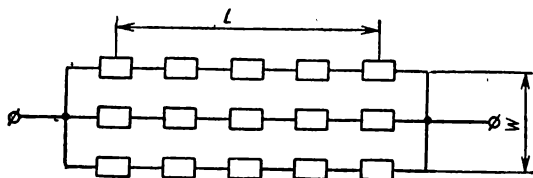


Рис. 10.16. Резервированная схема, имеющая длину $L=5$ и ширину $W=3$

Выражения (10.82) и (10.84) являются несколько видоизмененными формулами Мура и Шеннона [10.3]. В этих формулах некоторые члены иногда отсутствуют, так как при некоторых значениях h и g коэффициенты A_h и B_g равны нулю. Наименьшее количество элементов в пути, обеспечивающем прохождение сигнала через схему, называется длиной схемы и обозначается буквой L . Наименьшее количество элементов в группе, отказ которых вызывает отказ соединения, называется шириной схемы и обозначается буквой W . Для схемы, представленной на рис. 10.16, $L=5$, $W=3$.

Следует отметить, что формулы (10.81)—(10.84) определяют вероятность безотказной работы и вероятность

отказа схемы для отказов типа «обрыв». Для отказов типа «короткое замыкание» вероятности P и Q меняются местами. В этом случае по формулам (10.81) и (10.82) вычисляется вероятность отказа, а по (10.83) и (10.84) —

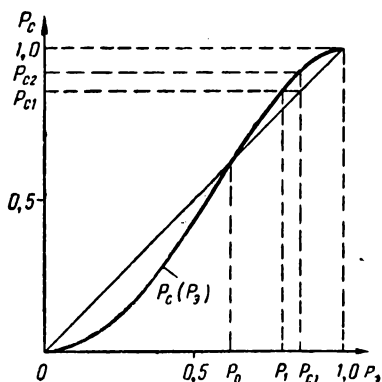


Рис. 10.17. Зависимость между вероятностью безотказной работы системы и элемента при комбинированном резервировании

вероятность безотказной работы, причем в правых частях этих выражений вероятности Q_i должны быть поставлены на место P_i , и наоборот.

Рассмотрим зависимость $P_c(t)$ от вероятности безотказной работы элемента $P_s(t)$ для комбинированных схем. Как показывают исследования выражения (10.82) [10.3], кривая $P_c(P_s)$ проходит через точки $(0, 0)$ и $(1, 1)$ и пересекает биссектрису координатного угла между ними (рис. 10.17). Если $P_s > P_0$, то схема об-

ладает более высокой надежностью. Дальнейшее увеличение надежности можно получить путем увеличения количества резервных элементов. Например, рассматривая резервное соединение как неделимый элемент можно построить из них подобную схему. Такой способ называется *итерацией*, т. е. подстановкой одной схемы в другую. Имея зависимость $P_c(P_s)$ для применяемой схемы, можно легко определить вероятность безотказной работы для любой по порядку итерации. Например, элемент, характеризуемый вероятностью безотказной работы P_1 , резервируется по схеме, представленной на рис. 10.18. Зависимость $P_c(P_s)$ для нее соответствует кривой, изображенной на рис. 10.17. Как видно из рисунка, вероятность безотказной работы схемы равна P_{c1} , причем $P_{c1} > P_1$. Следующая итерация повышает вероятность безотказной работы до величины P_{c2}

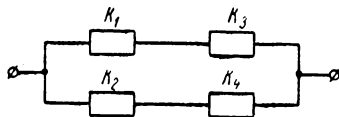


Рис. 10.18. Последовательно-параллельное соединение элементов

и т. д. Таким образом, повышая порядок итерации, можно получить вероятность безотказной работы схемы сколь угодно близкой к единице. Это достижимо тогда, когда рассматривается определенный промежуток времени, т. е. при неизменных значениях вероятности. Но вероятность безотказной работы — убывающая функция времени. Следовательно, всегда наступает момент, начиная с которого $P(t)$ элемента будет меньше P_0 . Таким образом, с течением времени эффективность резервирования уменьшается и для параллельно-последовательных схем при больших значениях t вероятность безотказной работы оказывается ниже, чем для одного реле.

Приведенные выше формулы можно применять для расчета надежности, если отказы элементов — события независимые. Это выполняется тогда, когда отсутствует *последствие отказов* или оно незначительно. Часто же это требование не выполняется, так как отказ одного из элементов в резервном соединении приводит к перераспределению нагрузок на других элементах, в результате чего интенсивность отказов их увеличивается. Если считать, что отказы одних элементов не меняют надежности других, то тем самым надежность всей системы будет завышена. Наоборот, если определять характеристики, исходя из наиболее тяжелых условий работы элементов, возникающих из-за наличия последствия, то тем самым надежность системы будет занижена. Для этих крайних случаев легко подсчитать вероятность безотказной работы системы. Пусть P_1 — вероятность в первом случае, P_2 — во втором. Тогда истинная вероятность безотказной работы P оценивается неравенством

$$P_2 < P < P_1.$$

Если приближенно положить, что $P = P_2$, то ошибка этого приближения будет меньше разности $P_1 - P_2$.

Для расчета надежности резервированных систем можно также применить матричный метод [10.2], который до некоторой степени позволяет учесть эффект последствия отказов.

При резервировании замещением необходимо применять переключающие устройства. Это усложняет систему и понижает ее общую надежность. Однако этот вид резервирования дает возможность применять, кроме *горячего резерва*, при котором избыточные элементы находятся

в рабочем состоянии с момента включения аппаратуры, *облегченный и ненагруженный резерв*. При облегченном или ненагруженном резерве избыточные элементы до момента включения их вместо отказавших элементов находятся соответственно в облегченном или ненагруженном режиме. Это приводит к повышению надежности устройства. Вместе с тем отсутствует изменение режимов работы элементов.

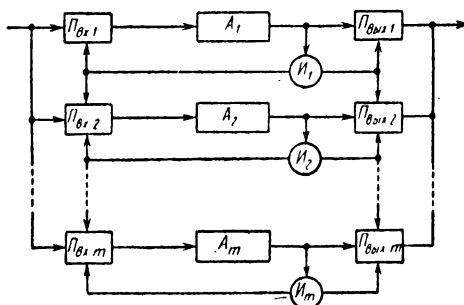


Рис. 10.19. Система с резервированием замещением

При горячем резерве расчет надежности производится аналогично расчету с постоянным резервированием, только учитывается надежность переключающих устройств. Вероятность безотказной работы резервированного узла с раздельными переключающими устройствами (рис. 10.19)

$$P_c(t) = 1 - [1 - P_{\text{и}}(t) \cdot P_{\text{вх}}(t) \cdot P_{\text{вых}}(t) \cdot P(t)]^m. \quad (10.85)$$

Здесь $P(t)$ — вероятность безотказной работы основного или резервного блоков A_1, A_2, \dots, A_m , предполагаемая для всех них одинаковой;

$P_{\text{и}}(t)$ — вероятность безотказной работы индикатора И;

$P_{\text{вх}}(t)$ — вероятность безотказной работы входного переключателя $\Pi_{\text{вх}}$;

$P_{\text{вых}}(t)$ — вероятность безотказной работы выходного переключателя $\Pi_{\text{вых}}$.

При облегченном или ненагруженном резерве расчет надежности систем следует выполнять с учетом того, что при отказе основного элемента в момент времени $\tau < t$ резервный работает в облегченном или ненагруженном режиме в течение времени τ и в нагруженном — в тече-

ние $t - \tau$. Если кратность резервирования единица, то при ненагруженном резерве вероятность безотказной работы системы за время t [10.3]

$$P_c(t) = P_{\text{осн}}(t) + \int_0^t P_{\text{рез}}(t - \tau) \omega_{\text{осн}}(\tau) d\tau, \quad (10.86)$$

где $P_{\text{осн}}(t)$ — вероятность безотказной работы основного элемента;

$P_{\text{рез}}(t - \tau)$ — вероятность безотказной работы резервного элемента;

$\omega_{\text{осн}}(\tau)$ — плотность вероятности безотказной работы основного элемента.

Если необходимо рассчитать надежность при большей кратности резервирования, то следует первоначально определить $P_c(t)$ при $m = 1$. Затем, принимая рассмотренное устройство за неделимый элемент, характеризуемый полученной вероятностью безотказной работы, определяем $P(t)$ для $m = 2$ и т. д. Вычисления проводятся до требуемой кратности резервирования.

При облегченном режиме отказ резервного элемента может наступить также до момента замещения основного. В этом случае формула (10.86) принимает следующий вид:

$$P_c(t) = P_{\text{осн}}(t) + \int_0^t P_1(\tau) P_{\text{рез}}(t - \tau) \omega_{\text{осн}}(\tau) d\tau, \quad (10.87)$$

где $P_1(\tau)$ — вероятность безотказной работы резервного элемента до включения его в работу.

Примеры. 1. Система, состоящая из трех последовательно соединенных устройств, дублируется другой, включенной постоянно. Вероятности безотказной работы устройств в течение времени t соответственно равны 0,82; 0,75; 0,68. Отказы устройств — события независимые. Во сколько раз повысится вероятность безотказной работы резервированной системы при переходе от общего к раздельному резервированию с той же кратностью?

Вероятность безотказной работы системы при общем резервировании по формуле (10.77)

$$P_{\text{об}} = 1 - \prod_{i=0}^1 \left(1 - \prod_{j=1}^3 P_{ij} \right) = 0,66.$$

Вероятность безотказной работы системы при раздельном резервировании по формуле (10.79)

$$P_{\text{раз}} = \prod_{i=1}^3 \left[1 - \prod_{j=0}^1 (1 - P_{ij}) \right] = 0,81.$$

Отношение вероятностей безотказной работы при раздельном и общем резервировании: $\frac{0,81}{0,66} \approx 1,3$.

Следовательно, в рассматриваемом случае эффективность раздельного резервирования в 1,3 раза выше, чем общего.

2. Контакты реле имеют отказы двух видов: «обрыв» и «короткое замыкание». Вероятность безотказной работы реле за время t по отношению к отказам типа «обрыв» $P_{обр} = 0,731$, а по отношению к отказам типа «короткое замыкание» $P_{к.з} = 0,852$. Для повышения надежности реле введено постоянное резервирование. Реле заменяется комбинацией из четырех реле ($K=4$), соединенных по схеме, представленной на рис. 10.20. Определим вероятность безотказной работы схемы и ее итерации первого порядка по отношению к отказам того и другого типа.

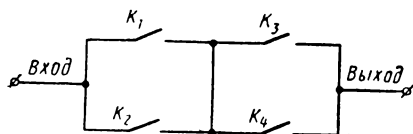


Рис. 10.20. Параллельно-последовательное соединение контактов реле

Для определения вероятности безотказной работы схемы по отношению к отказам типа «обрыв» воспользуемся общей формулой (10.81). В схеме имеется четыре пути передачи сигнала от входа к выходу:

$j=1$ — через реле K_1 и K_3 ,

$j=2$ — через реле K_2 и K_4 ,

$j=3$ — через реле K_1 и K_4 ,

$j=4$ — через реле K_2 и K_3 .

Следовательно, $m=4$, $n=2$, $l=1 \div 2$.

Определим члены внутренней суммы при $j=1$. Для всех них $a_{l1}=1$. В этом случае должны безотказно работать реле K_1 и K_3 , но могут правильно функционировать также реле K_2 и K_4 — одно из них или оба:

$$\left. \begin{array}{l} P_1 P_3 Q_2 Q_4 \\ P_1 P_3 P_2 Q_4 \\ P_1 P_3 Q_2 P_4 \end{array} \right\} \text{ при } l=1,$$

$$P_1 P_3 P_2 P_4 \quad \text{при } l=2.$$

При $j=2$ коэффициент $a_{l2}=1$ для следующих членов внутренней суммы:

$$\left. \begin{array}{l} P_2 P_4 Q_1 Q_3 \\ P_2 P_4 P_1 Q_3 \\ P_2 P_4 Q_1 P_3 \end{array} \right\} \text{ при } l=1.$$

Для члена $P_2P_4P_1P_3$ следует взять $a_{12}=0$, так как он был при рассмотрении первого пути. Для третьего $j=3$ и четвертого $j=4$ путей имеется лишь по одному члену с коэффициентом $a_{ij}=1$:

$$P_1P_4Q_2Q_3,$$

$$P_2P_3Q_1Q_4.$$

Суммируя все члены, получаем:

$$P_{с.обр} = P_1P_3(Q_2Q_4 + P_2Q_4 + P_4Q_3 + P_2P_4) + \\ + P_2P_4(Q_1Q_3 + P_1Q_3 + Q_1P_3) + P_1P_4Q_2Q_3 + P_2P_3Q_1Q_4.$$

Так как все элементы схемы одинаковы, считаем, что все они характеризуются одинаковой вероятностью безотказной работы $P_{обр}$. Подставляем в предыдущее выражение $P_{обр}$ и $Q_{обр} = 1 - P_{обр}$, тогда

$$P_{с.обр} = P_{обр}^4 - 4P_{обр}^3 + 4P_{обр}^2.$$

При $P_{обр} = 0,731$ $P_{с.обр} = 0,861$.

На рис. 10.21 кривой 1 представлена зависимость $P_{с.обр}(P_{обр})$ для рассматриваемой схемы. Графически определяем, что для итерации первого порядка вероятность безотказной работы $P_{ит.обр} = 0,962$.

Отказы типа «короткое замыкание» приводят к замыканию цепи между входом и выходом при разомкнутых контактах. Поэтому для определения вероятности безотказной работы в этом случае воспользуемся формулой (10.83), заменив Q на P . В рассматриваемой схеме может быть две группы, обеспечивающие замыкание цепи:

$$j=1 - K_1 \text{ и } K_2,$$

$$j=2 - K_3 \text{ и } K_4.$$

Следовательно, $r=2$, $u=2$, $y=0 \div 2$.

Члены внутренней суммы для первой группы $j=1$:

$$P_1P_2Q_3Q_4 \quad \text{при } y=0,$$

$$\left. \begin{array}{l} P_1P_3P_3Q_4, \\ P_1P_2Q_3P_4 \end{array} \right\} \quad \text{при } y=1,$$

$$P_1P_2P_3P_4 \quad \text{при } y=2.$$

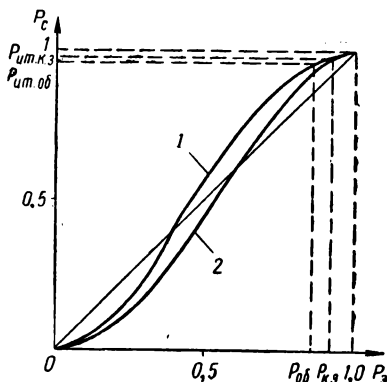


Рис. 10.21. Зависимость $P_c(P_3)$ для параллельно-последовательного соединения

Во второй группе $j=2$ коэффициент $b_{y2}=1$ для членов

$$P_3 P_4 Q_1 Q_2 \quad \text{при } y=0,$$

$$\left. \begin{array}{l} P_3 P_4 P_1 Q_2, \\ P_3 P_4 Q_1 P_2 \end{array} \right\} \quad \text{при } y=1.$$

Для повторяющегося члена $P_3 P_4 P_1 P_2$ следует взять $b_{y2}=0$. Суммируя все члены, получаем:

$$P_{\text{с. к. з}} = P_1 P_2 (Q_3 Q_4 + P_3 Q_4 + Q_3 P_4 + P_3 P_4) + \\ + P_3 P_4 (Q_1 Q_2 + P_1 Q_2 + Q_1 P_2).$$

Так как все реле одинаковы,

$$P_{\text{с. к. з}} = P_{\text{к. з}}^4 + 4P_{\text{к. з}}^3 Q_{\text{к. з}} + 2P_{\text{к. з}}^2 Q_{\text{к. з}}^2,$$

или

$$P_{\text{с. к. з}} = 2P_{\text{к. з}}^2 - P_{\text{к. з}}^4.$$

Подставляя значение $P_{\text{к. з}} = 0,852$, получаем $P_{\text{с. к. з}} = 0,925$. На рис. 10.21 кривой 2 показана зависимость $P_{\text{с. к. з}}$ от $P_{\text{к. з}}$. Для итерации первого порядка графически определяем вероятность безотказной работы.

$$P_{\text{ит. к. з}} = 0,979.$$

3. Для повышения надежности устройства применяется резервирование замещением с кратностью $m=2$. Резервные устройства до включения не нагружены. Интенсивность отказов $\lambda = 3,5 \cdot 10^{-3} \frac{1}{\text{ч}}$ постоянна. Определим вероятность безотказной работы резервированной системы в течение времени $t=100$ ч, без учета надежности переключающих устройств.

Вероятность безотказной работы системы, состоящей из основного и одного резервного устройства, при экспоненциальном распределении

$$P'_c(t) = e^{-\lambda t} + \int_0^t e^{-\lambda(t-\tau)} \lambda e^{-\lambda \tau} d\tau = e^{-\lambda t} + \lambda t e^{-\lambda t}.$$

Подставляя в эти выражения значения λ и t , получаем:

$$P'_c(t) = e^{-3,5 \cdot 10^{-3} \cdot 100} (1 + 3,5 \cdot 10^{-3} \cdot 100) = 0,945.$$

Рассмотрим систему, характеризуемую вероятностью безотказной работы $P'_c(t) = 0,945$, как основное устройство. Оно резервируется с кратностью $m=1$. Для определения вероятности безотказной работы всей системы необходимо найти выражение плотности вероятности рассматриваемого основного устройства, равное производной $P_c(t)$ по t :

$$w'_{\text{осн}}(\tau) = - \frac{dP'_c(\tau)}{d\tau} = \lambda^2 \tau e^{-\lambda \tau}.$$

Подставляя $w'_{\text{осн}}(\tau)$ в (10.81), получаем выражение вероятности безотказной работы всей системы:

$$P_c(t) = P'_c(t) + \int_0^t e^{-\lambda(t-\tau)} \lambda^2 \tau e^{-\lambda \tau} d\tau = P'_c(t) + \frac{\lambda^2 t^2}{2} e^{-\lambda t},$$

откуда

$$P_c(t) = 0,945 + \frac{(3,5 \cdot 10^{-8})^2 \cdot 100^2}{2} e^{-3,5 \cdot 10^{-8} \cdot 10^2} = 0,988.$$

§ 10.5. ЭКОНОМИКА НАДЕЖНОСТИ

Надежность является одним из важнейших факторов, определяющих эффективность применения управляющих систем. Действительно, ненадежная аппаратура, не обеспечивая управление, требует огромных затрат на ремонт и техническое обслуживание. Кроме того, любой отказ сопровождается экономическим ущербом, определяемым потерями продукции из-за простоя, производством брака из-за нарушения технологии и т. п. Поэтому при создании новых систем необходимо оценивать их эффективность с учетом надежности, рассматривая при этом возможные экономические потери, возникающие вследствие отказа.

Недостаточная надежность систем приводит к необходимости иметь специальные меры ее повышения. Это сопровождается экономическими затратами. Поэтому необходимо оценивать *экономический эффект* повышения надежности. Он является положительным, если затраты окупаются в период эксплуатации. В этом случае повышение надежности системы целесообразно.

Для невосстанавливаемых систем экономический эффект повышения надежности можно оценить *сроком окупаемости* τ .

Экономический эффект при увеличении надежности будет положительным, если

$$\tau < \bar{T}. \quad (10.88)$$

При этом следует учитывать, что значение среднего времени \bar{T} системы повышенной надежности должно быть меньше времени *морального износа*, иначе возможность получения экономического эффекта остается нереализованной.

Определение экономической эффективности повышения надежности восстанавливаемых систем можно ориентировочно выполнить следующим образом. Допустим, что до повышения надежности система характеризовалась средним числом отказов m_1 за время t_1 . При повышении надежности эта величина за то же время снизилась до m_2 . Если стоимость отказа, в которую входит стоимость ремонта и потеря в результате отказа, у менее надежной системы равна C_{o1} , а первоначальные затраты на изготовление системы равны $C_{н1}$, то затраты на изготовление и эксплуатацию этой системы

$$C_1 = C_{н1} + C_{o1}m_1. \quad (10.89)$$

Соответственно после введения мер повышения надежности общая сумма затрат

$$C_2 = C_{н2} + C_{o2}m_2, \quad (10.90)$$

где $C_{н2}$ — стоимость изготовления системы повышенной надежности;

C_{o2} — стоимость отказа системы повышенной надежности.

Обычно $C_{н1} < C_{н2}$, $C_{o1} \leq C_{o2}$. Однако при введении в систему повышенной надежности устройств, облегчающих поиск и устранение неисправностей, C_{o2} может быть меньше C_{o1} . Отношение

$$x = \frac{C_2}{C_1} \quad (10.91)$$

характеризует эффективность повышения надежности. При $x < 1$ экономический эффект положительный.

Одним из основных методов повышения надежности, как указывалось выше, является резервирование. Введение резерва, однако, увеличивает стоимость систем, их вес и габариты. В связи с этим возникает задача обеспечения *оптимального резервирования*, т. е. выбора способа резервирования и его кратности, обеспечивающих максимальное значение вероятности безотказной работы при каком-либо ограничении, например заданной стоимости или весе, или габаритах. Часто возникает и обратная задача, при которой необходимо определить способ резервирования и его кратность, обеспечивающие заданную ве-

роятность безотказной работы при минимальной стоимости (весе, габаритах) системы.

Задача определения оптимальной кратности резервных элементов решена для раздельного постоянного резервирования [10.2].

Пример. Интенсивность отказов системы $\lambda_0 = 0,531 \cdot 10^{-3} \frac{1}{ч}$. Меры повышения надежности позволяют понизить интенсивность ее отказов до величины $\lambda = 0,125 \cdot 10^{-3} \frac{1}{ч}$, при этом срок окупаемости затрат на повышение надежности $\tau = 2,7 \cdot 10^3$ ч.

Среднее значение времени T системы повышенной надежности

$$\bar{T} = \frac{1}{0,125 \cdot 10^{-3}} = 8,0 \cdot 10^3 \text{ ч.}$$

Меры повышения надежности дают положительный экономический эффект, так как $\tau < \bar{T}$. Следовательно, введение их целесообразно.

ЛИТЕРАТУРА

Основные пособия и дополнительная литература по главам указаны в хронологическом порядке, по годам издания, в пределах же каждого года — в алфавитном порядке по фамилиям авторов. Такой порядок позволяет легко находить литературу последнего из упомянутых и предыдущих годов.

0.1. Основы автоматического регулирования. Под ред. В. В. Солодовникова. Машгиз, 1954.

0.2. С. С. Цянь. Техническая кибернетика. ИЛ, 1956.

0.3. В. А. Тимофеев. Основы теории автоматического регулирования, вып. 1. Сев.-зап. заочн. политехн. ин-т, 1959.

0.4. Д. ж. Траксел. Синтез систем автоматического регулирования. Машгиз, 1959.

0.5. Д. ж. К. Ньютон, Л. А. Гулд, Д. Ф. Кайзер. Теория линейных следящих систем. Физматгиз, 1961.

0.6. А. А. Красовский, Г. С. Поспелов. Основы автоматики и технической кибернетики. Госэнергоиздат, 1962.

0.7. А. В. Солодов. Линейные системы автоматического управления с переменными параметрами. Физматгиз, 1962.

0.8. Э. А. Якубайтис. Основы технической кибернетики. Изд-во АН Латв. ССР, 1962.

0.9. А. А. Фельдбаум, А. Д. Дудыкин, А. А. Мановцев, Н. Н. Миролюбов. Теоретические основы связи и управления. Физматгиз, 1963.

0.10. Л. П. Крайзмер. Техническая кибернетика, изд. 2-е. Изд-во «Энергия», 1964.

0.11. В. И. Грубов, А. Г. Ивахненко, В. М. Мандровский-Соколов. Промышленная кибернетика. Справочник. Изд-во «Наукова Думка», 1966.

0.12. А. В. Солодов. Теория информации и ее применение к задачам автоматического управления и контроля. Изд-во «Наука», 1967.

Глава I

1.1. А. — М. Ампер. Essai sur la philosophie des sciences, Paris, 1843.

1.2. Н. А. Умов. Собрание сочинений, т. III, М., 1916.

1.3. Э. Шредингер. Что такое жизнь? ИЛ, 1947.

1.4. N. Wiener. Cybernetics or control and communication in the animal and the machine. N. Y., 1948., См. Н. Винер. Кибернетика. Изд-во «Советское радио», 1968.

1.5. J. Rothstein. Information, measurement and quantum mechanics. Science, v. 114, p. 171, 1951.

1.6. Ф. Крик. Структура наследственного вещества. Химич. наука и промышленность, т. I, стр. 472, 1956.

- 1.7. У. Р. Эшби. Схема усилителя мыслительных способностей. Сб. «Автоматы» под ред. К. Э. Шеннона и Дж. Маккарти. ИЛ, 1956.
- 1.8. Н. Винер. Кибернетика и общество. ИЛ, 1958.
- 1.9. П. Косса. Кибернетика. ИЛ, 1958.
- 1.10. У. Р. Эшби. Применение кибернетики в биологии и социологии. «Вопросы философии», 1958, № 12, стр. 110.
- 1.11. У. Р. Эшби. Введение в кибернетику. ИЛ, 1959.
- 1.12. Л. Бриллюэн. Наука и теория информации. Физматгиз, 1960.
- 1.13. Р. А. Сапожников. Основная задача кибернетики. Сб. трудов Ленингр. механич. ин-та, № 12 «Автоматическое управление», 1960.
- 1.14. А. М. Яглом, И. М. Яглом. Вероятность и информация. Физматгиз, 1960.
- 1.15. Автоматическое управление. Сб. статей под ред. В. В. Солодовникова. Изд-во АН СССР, 1961.
- 1.16. А. И. Берг. Кибернетика и некоторые технические проблемы управления народным хозяйством. «Вопросы философии», 1961, № 2, 11.
- 1.17. И. Я. Депман. Применение электронных счетных машин для отыскания совершенных чисел. Математическое просвещение, вып. 6, стр. 324, 1961.
- 1.18. Л. Бриллюэн. Термодинамика, статистика, информация. УФН, т. 77, стр. 337, 1962.
- 1.19. А. Г. Ивахненко. О значении принципа действия обучающихся систем. «Автоматика», 1962, № 5, стр. 81.
- 1.20. М. В. Келдыш. Развитие кибернетики — дело ученых различных специальностей. Изд. АН СССР, отд. техн. наук. «Техническая кибернетика», 1963, № 3, стр. 3.
- 1.21. Р. А. Сапожников. Информация и энтропия. Сб. трудов Ленингр. механич. ин-та, № 29, «Автоматическое управление», вып. 3, 1963.
- 1.22. К. Э. Шеннон. Работы по теории информации и кибернетике. ИЛ, 1963.
- 1.23. Н. Винер. Люди и машины. Наука и техника, 1964, № 5, стр. 21.
- 1.24. Р. А. Сапожников. О применении понятий энтропии, информации и негэнтропии. Сб. трудов Ленингр. механич. ин-та, № 41, «Техническая кибернетика», 1964.
- 1.25. У. Р. Эшби. Системы и информация. «Вопросы философии», 1964, № 3, стр. 78.
- 1.26. А. Эйнштейн. Творческая автобиография. В сб. «Физика и реальность». Изд-во «Наука», 1965.
- 1.27. Л. Бриллюэн. Научная неопределенность и информация. Изд-во «Мир», 1966.

Глава II

- 2.1. G. Boole. Investigation of the laws of thought. London, 1854.
- 2.2. Н. А. Романов. О возможности контакта между теорией вероятностей и учением академика И. П. Павлова об условных рефлексах. ДАН СССР, т. 1, стр. 193, 1935.

- 2.3. И. П. Павлов. Полное собрание сочинений, т. 3, кн. 1, 1951.
- 2.4. А. А. Ляпунов, Г. А. Шестопал. Об алгоритмическом описании процессов управления. Математическое просвещение, вып. 2, стр. 81, 1957.
- 2.5. Н. Винер. Кибернетика и общество. ИЛ, 1958.
- 2.6. J. J. D'Azzo, C. H. Houpsis. Control system analysis and synthesis. N. Y., 1960.
- 2.7. Э. Беркли. Символическая логика и разумные машины. ИЛ, 1961.
- 2.8. А. Тастин. Обратная связь. В сб. «Автоматическое управление». Изд-во АН СССР, 1961.
- 2.9. S. Calvert, G. Coulman. Feed forward control: its future role in the chemical industry. Chem. Engng. Progr., v. 57, p. 45.
- 2.10. Л. П. Крайзмер. Бионика. Госэнергоиздат, 1962.
- 2.11. В. А. Трапезников. Кибернетика и автоматическое управление. «Известия», 8 февр., 1962.
- 2.12. В. А. Трапезников. Кибернетика и автоматическое управление. Автоматика и телемеханика, т. 23, стр. 279, 1962.
- 2.13. А. Г. Ивахненко. Системы, що самоорганізуються, з додатними зворотними зв'язками. «Автоматика», 1962, № 3, стр. 33.
- 2.14. М. А. Айзерман и др. Логика. Автоматы. Алгоритмы. Физматгиз, 1963.
- 2.15. В. М. Глушков. Моделирование мыслительных процессов. «Природа», 1963, № 2, стр. 3.
- 2.16. Н. А. Криницкий, Г. А. Миронов, Г. Д. Фролов. Программирование (Справочная математическая библиотека под ред. Л. А. Люстерника и А. Р. Ямпольского). Физматгиз, 1963.
- 2.17. В. А. Трапезников. Основные направления развития технической кибернетики. Изв. АН СССР, отд. техн. наук. «Техническая кибернетика», 1963, № 3, стр. 5.
- 2.18. В. Н. Яворский, А. А. Бессонов, А. И. Кортаев, А. М. Потапов. Проектирование инвариантных следящих приводов. Изд-во «Высшая школа», 1963.
- 2.19. А. И. Берг. Бионика и ее значение для развития техники. Сб. «Бионика». Изд-во «Наука», 1965.
- 2.20. С. А. Эсмедляев. К двухсотлетию регулятора Ползунова. Сб. трудов Ленингр. механич. ин-та, № 51, «Техническая кибернетика», вып. 2, 1965.
- 2.21. Н. Винер. Кибернетика, второе изд. «Советское радио», 1968.

Глава III

- 3.1. В. А. Котельников. О пропускной способности «эфира» и проволоки в электросвязи. Материалы к I Всесоюзн. съезду по вопросам технич. реконструкции связи, 1933.
- 3.2. А. Я. Хинчин. Теория корреляции стационарных стохастических процессов. УМН, вып. 5, стр. 42, 1938.
- 3.3. Н. А. Лузин. К изучению матричной теории дифференциальных уравнений. «Автоматика и телемеханика», 1940. № 5, стр. 4.

3.4. Б. Вандер Поль, Х. Бреммер. Операционное исчисление на основе двустороннего преобразования Лапласа. ИЛ, 1952.

3.5. B. Widrow. A study of rough amplitude quantization by means of Nyquist sampling theory. Trans. IRE, v. PGCT — 3, p. 266, 1956.

3.6. Н. Д. Саперштейн. К вопросу о применении интеграла Фурье при исследовании переходных процессов. Сб. трудов Ленингр. механич. ин-та, № 12, «Автоматическое управление», 1960.

3.7. В. В. Солодовников. Статистическая динамика линейных систем автоматического управления. Физматгиз, 1960.

3.8. Р. А. Сапожников, Н. Д. Саперштейн. Расчет переходных процессов, возникающих в линейных системах при внешних воздействиях. Сб. трудов Ленингр. механич. ин-та, № 22, «Автоматическое управление», вып. 2, 1961.

3.9. Z. Frydrych. Kryteria prównawczej oceny przekształcania sygnałów ciągłych w sygnały dyskretne. Archiwum Automatyki i Telemechaniki, 7, 337, 1962.

3.10. Э. Джурри. Импульсные системы автоматического регулирования. Физматгиз, 1963.

3.11. Б. Р. Левин. Корреляция. Справочник «Автоматизация производства и промышленная электроника», т. 2, 1963.

3.12. Н. Д. Саперштейн. О решении уравнения в обобщенных производных на вычислительной машине непрерывного действия. Сб. трудов Ленингр. механич. ин-та, № 33, «Автоматическое управление», вып. 4, 1963.

3.13. Р. А. Сапожников. Стационарные случайные сигналы ограниченной длительности. Сб. трудов Ленингр. механич. ин-та, № 33, «Автоматическое управление», вып. 4, 1963.

3.14. К. Э. Шеннон. Работы по теории информации и кибернетике. ИЛ, 1963.

3.15. S. Seely. Dynamic systems analysis. New York, 1964.

3.16. В. А. Диткин, А. П. Прудников. Справочник по операционному исчислению. Изд-во «Высшая школа», 1965.

3.17. М. А. Айзерман. Теория автоматического регулирования. Изд-во «Наука», 1966.

3.18. Н. Д. Саперштейн, В. Л. Файншмидт. Начала операционного исчисления. Ленингр. механич. ин-т, 1966.

Глава IV

4.1. Л. П. Шиниберов. Электрогидропривод как объект регулирования. «Электричество», 1957, № 8, стр. 27.

4.2. Р. А. Сапожников, Б. М. Ларионов. Гидродинамический эффект в электрогидроприводах. НДВШ. «Электромеханика и автоматика», 1959, № 2, стр. 231.

4.3. Н. Т. Кузовков. Теория автоматического регулирования, основанная на частотных методах. Оборонгиз, 1960.

4.4. Р. А. Сапожников, Н. Д. Саперштейн. Расчет переходных процессов, возникающих в линейных системах при внешних воздействиях. Сб. трудов Ленингр. механич. ин-та, № 22, «Автоматическое управление», вып. 2, 1961.

4.5. Я. З. Цыпкин. Теория линейных импульсных систем. Физматгиз, 1963.

4.6. Р. А. Сапожников, Б. М. Ларионов. Переходные характеристики гидроприводов раздельного исполнения. В сб. «Пнеumo- и гидроавтоматика». Изд-во «Наука», 1964.

Глава V

5.1. S. J. Mason. Feedback theory-further properties of signal flow graphs. Proc. IRE, v. 44, p. 920, 1956.

5.2. С. А. Ковчин. Письмо в редакцию. Изв. ВУЗ. «Электро-механика», 1960, № 1, стр. 133.

5.3. Приспосабливающиеся автоматические системы, под ред. Э. Мишкина и Л. Брауна. М., ИЛ, 1963.

5.4. Я. З. Цыпкин. Теория линейных импульсных систем. Физматгиз, 1963.

5.5. J. Čermák, J. Závorka. Použití diagrama toku signálů y regulační technice. Automatizace, 6, 60, 1963.

5.6. S. Seely. Dynamic systems analysis. New York, 1964.

Глава VI

6.1. А. В. Михайлов. Метод гармонического анализа в теории регулирования. «Автоматика и телемеханика». 1938, № 3, стр. 27.

6.2. М. А. Айзерман. О затухании колебательного движения, характеризуемого линейным дифференциальным уравнением третьего порядка с постоянными коэффициентами. «Автоматика и телемеханика», 1940, № 1, стр. 55.

6.3. Д. К. Максвелл, И. А. Вышнеградский, А. Стодола. Теория автоматического регулирования. Изд-во АН СССР, 1949.

6.4. Ю. И. Неймарк. Устойчивость линеаризованных систем. ЛВВИА им. А. Ф. Можайского, 1949.

6.5. А. М. Рубинчик. Приближенный метод оценки качества регулирования в линейных системах. В сб. «Устройства и элементы теории автоматики и телемеханики». Машгиз, 1952.

6.6. Е. П. Попов. Динамика систем автоматического регулирования. ГИТТЛ, 1954.

6.7. В. И. Крутов. Исследование устойчивости систем высоких порядков при помощи диаграммы проф. И. А. Вышнеградского. В сб. «Двигатели внутреннего сгорания». Машгиз, 1955.

6.8. Р. А. Сапожников, Т. Н. Мокиенко. Исследование диаграммы Вышнеградского. Сб. тр. Ленингр. механич. ин-та, № 12, «Автоматическое управление», 1960.

6.9. E. Schwartz. Sechs «natürliche» Normierung der Stammgleichung dritten Grades. Arch. elektr. Übertrag, Bd. 15, S. 94, 1961.

6.10. С. М. Смирнов, П. В. Терентьев. Генераторы импульсов высокого напряжения. Изд-во «Энергия», 1964.

6.11. А. Ключев. Исследование систем автоматического регулирования с помощью амплитудно-фазовых критериев необходимой устойчивости по модулю и фазе. Изв. ВУЗ. Радиофизика. т. 8, стр. 203, 1965.

6.12. М. А. Айзерман, Теория автоматического регулирования. Изд-во «Наука», 1966.

6.13. Е. Д. Соломенцев. Теория функций комплексного переменного, операционное исчисление, устойчивость линейных систем. МЭИ, 1966.

Глава VII

7.1. К. Ф. Теодорчик. Траектории корней характеристического уравнения третьего порядка при непрерывном изменении свободного члена и максимальная достижимая при этом устойчивость. ЖТФ, т. 18, стр. 1394, 1948.

7.2. W. R. Evans. Graphical Analysis of control systems. Trans., AIEE, v. 67, p. 547, 1948.

7.3. В. В. Солодовников. О применении трапецеидальных частотных характеристик к анализу качества систем автоматического регулирования. «Автоматика и телемеханика», 1949, т. 5, стр. 362.

7.4. В. В. Солодовников, Ю. И. Топчиев, Г. В. Крутикова. Частотный метод построения переходных процессов. ГИТТЛ, 1955.

7.5. Т. Н. Мокиенко. Графическое представление некоторых свойств систем четвертого порядка. Сб. тр. Ленингр. механич. ин-та, № 12, «Автоматическое управление», 1960.

7.6. J. J. D' Azzo, C. H. Houpis. Feedback control system analysis and synthesis. New York, 1960.

7.7. Т. Н. Мокиенко. Переходная характеристика линейной системы четвертого порядка. Сб. тр. Ленингр. механич. ин-та, № 22, «Автоматическое управление», вып. 2, 1961.

7.8. Р. А. Сапожников, Н. Д. Саперштейн. Расчет переходных процессов, возникающих в линейных системах при внезапных воздействиях. Сб. трудов Ленингр. механич. ин-та, № 22, «Автоматическое управление», вып. 2, 1961.

7.9. М. А. Айзерман. Теория автоматического регулирования. Физматгиз, 1963.

7.10. Н. Д. Саперштейн. О решении уравнения в обобщенных производных на вычислительной машине непрерывного действия. Сб. трудов Ленингр. механич. ин-та, № 33, «Автоматическое управление», вып. 4, 1963.

7.11. В. Дель Торо, С. Р. Паркер. Принципы проектирования систем автоматического управления. Машгиз, 1963.

7.12. С. Д. Тропский. К вопросу о построении переходного процесса в линейных системах по передаточной функции. Сб. трудов Ленингр. механич. ин-та, № 33, «Автоматическое управление», вып. 4, 1963.

7.13. Э. Г. Удерман. Метод корневого годографа в теории автоматического управления. Госэнергоиздат, 1963.

7.14. Н. Д. Саперштейн. Построение переходных процессов в разомкнутых импульсных системах при помощи преобразования Лапласа. Сб. трудов Ленингр. механич. ин-та, № 51, «Техническая кибернетика», вып. 2, 1965.

7.15. В. А. Бесекерский, А. П. Попов. Теория систем автоматического регулирования. Изд-во «Наука», 1966.

7.16. Л. Бриллюэн. Научная неопределенность и информация. Изд-во «Мир», 1966.

7.17. К. В. Егоров. Основы теории автоматического регулирования. Изд-во «Энергия», 1967.

7.18. Р. А. Сапожников, Г. А. Балаболов. Построение переходных процессов с помощью спектральных преобразований. Ленинград. механич. ин-т. Тезисы доклада к научно-технич. конференции, посвященной 50-летию Советской власти, 1967.

Глава VIII

8.1. М. А. Айзерман, Ф. Р. Гантмахер. Абсолютная устойчивость регулируемых систем. Изд-во АН СССР, 1963.

8.2. Н. М. Матвеев. Методы интегрирования обыкновенных дифференциальных уравнений. Изд-во «Высшая школа», 1963.

8.3. Г. А. Татарникова. Влияние нелинейностей на случайный сигнал и переносимую им информацию. Сб. трудов Ленингр. механич. ин-та, № 62, «Техническая кибернетика», вып. 3, 1967.

Глава IX

9.1. С. С. Цянь. Техническая кибернетика. ИЛ, 1956.

9.2. П. Косса. Кибернетика. ИЛ, 1958.

9.3. В. В. Казакевич, Р. В. Корнилов, Н. Г. Христофоров. Электронный экстремальный регулятор. В сб. «Теория и применение дискретных автоматических систем». Изд-во АН СССР, 1960.

9.4. F. Rosenblatt. Perceptron simulation experiments. Proc. of the IRE, v. 48, N 3, 1960.

9.5. Л. С. Понтрягин, В. Г. Болтянский, Р. В. Гамкрелидзе, Е. Ф. Мищенко. Математическая теория оптимальных процессов. Физматгиз, 1961.

9.6. У. Р. Эшби. Конструкция мозга. ИЛ, 1962.

9.7. Э. А. Якубайтис. Основы технической кибернетики. Изд-во АН Латв. ССР, 1962.

9.8. Л. Г. Евланов. Самонастраивающаяся система с поиском градиента методом вспомогательного оператора. Сб. «Техническая кибернетика», № 1. Изд-во АН СССР, М., 1963.

9.9. А. А. Красовский. Динамика непрерывных самонастраивающихся систем. Физматгиз, 1963.

9.10. П. Н. Матвеев. К вопросу о моделировании гомеостатических систем. Сб. трудов Ленингр. механич. ин-та, № 29, «Автоматическое управление», вып. 3, 1963.

9.11. Приспосабливающиеся автоматические системы, под ред. Э. Мишкина и Л. Брауна. ИЛ, 1963.

9.12. В. Дель Торо, С. Р. Паркер. Принципы проектирования систем автоматического управления. Машгиз, 1963.

9.13. В. К. Чичинадзе, О. А. Чарквиани. Применение самонастраивающихся систем для автоматизации процессов конструирования систем автоматического управления. В сб. «Самонастраивающиеся автоматические системы». Изд-во «Наука», 1964.

9.14. Автоматизация производства и промышленная электроника (Энциклопедия современной техники), т. 4. Изд-во «Советская энциклопедия», 1965.

9.15. Аналитические самонастраивающиеся системы автоматического управления, под ред. В. В. Солодовникова. Изд-во «Машиностроение», 1965.

9.16. С. Э. Городецкий, Э. М. Саликов. Автоматическая оптимизация режима работы радиопередатчика. В сб. «Самонастраивающиеся системы». Изд-во «Наука», 1965.

9.17. М. Л. Либерзон, А. Б. Родов. Системы экстремального регулирования. Изд-во «Энергия», 1965.

9.18. В. Г. Болтянский. Математические методы оптимального управления. Изд-во «Наука», 1966.

9.19. В. И. Грубов, А. Г. Ивахненко, Б. Ю. Мандровский-Соколов. Промышленная кибернетика. Изд-во «Наукова думка», 1966.

9.20. Р. Л. Оптимальные оценки, определение характеристик и управление. Изд-во «Наука». 1966.

9.21. П. Н. Матвеев. О построении мультиустойчивых систем. Сб. трудов Ленингр. механич. ин-та, № 62, «Техническая кибернетика», вып. 3, 1966.

9.22. А. А. Павлов. Синтез релейных систем, оптимальных по быстрдействию. Изд-во «Наука», 1966.

Глава X

10.1. А. А. Бессонов. Расчет надежности систем автоматического управления. Сб. тр. Ленингр. механич. ин-та, № 22, «Автоматическое управление», вып. 2, 1961.

10.2. А. М. Половко. Основы теории надежности. Изд-во «Наука», 1964.

10.3. Р. А. Сапожников, А. А. Бессонов, А. Г. Шоломицкий. Надежность автоматических управляющих систем. Изд-во «Высшая школа», 1964.

10.4. Б. В. Гнеденко, Ю. К. Беляев, А. Д. Соловьев. Математические методы в теории надежности. Изд-во «Наука», 1965.

10.5. Н. А. Филадельфина. Оценка надежности с учетом внезапных и постепенных отказов. Сб. тр. Ленингр. механич. ин-та, № 51, «Техническая кибернетика», вып. 2, 1965.

10.6. Р. А. Сапожников, Н. А. Филадельфина. Обобщенный закон надежности. Сб. тр. Ленингр. механич. ин-та, № 62, «Техническая кибернетика», вып. 3, 1966.

10.7. Н. А. Филадельфина. Расчет надежности сложных устройств при учете внезапных и постепенных отказов. Изв. Ленингр. электротехнич. ин-та им. В. И. Ульянова (Ленина), вып. 56, 1, 1966.

ПРЕДМЕТНЫЙ УКАЗАТЕЛЬ

- Абсцисса абсолютной сходимости
- Автоколебания
- Автомат управляющий
- Автоматизация
 - комплексная
- Адьюнкт прямого пути
- Акт переработки информации
- Алгебра булева
- Алгоритм
 - детерминированный
 - недетерминированный
 - самоорганизующийся
 - универсальный
 - частный
- Алгоритмизация
- Ансамбль иерархий
- Аппроксимация кусочно-линейная
- Бел
- Бионика
- Бит
- Бинит
- Блок сообщения
- Блок-схема
- Быстродействие оптимальное
- Вектор Михайлова
 - Найквиста
 - характеристический
- Величина случайная
 - — дискретная
- Вероятность безотказной работы
 - апостериорная
 - априорная
- Ветвь графа
- Воздействие возмущающее
 - гармоническое
 - изучающее
 - направляющее
 - управляющее
- Время между отказами среднее
 - чистого запаздывания
- Вход
- Высказывание
 - истинное
 - ложное
 - равное
 - тождественно-истинное
- — ложное
- Выход
- Генерирование информации
- Гипербола Вышнеградского
- Гипотеза эргодическая
- Гистерезис
- Гладкость сигнала
- Годограф корневой
- Гомеостазис
- Гомеостат
- Граница переключения
 - устойчивости
- Граф сигнальный
 - элементарный
- Датчик обстановки
- Движение возмущенное
 - до упора
 - невозмущенное
 - устойчивое асимптотически
 - — по Ляпунову
- Дезинформация
- Декада
- Декодирование
- Дельта-функция
- Децибел
- Диаграмма ветвящаяся
- Дизъюнкция
- Дискрета
- Дисперсия
 - режимов
- Дифференцирование обобщенное
- Дрейф характеристик
- Единица двоичная
- Емкость кода
- Зазор
- Закон второй термодинамики
 - — — в кибернетическом выражении
- — — надежности
- — — обобщенный
- Запаздывание инерционное
 - транспортное
- Запас устойчивости
 - — по амплитуде
 - — — фазе
- Затухание
- Звено

- аperiодическое второго по-
рядка
- безынерционное
- дифференцирующее
- — идеальное
- Звено инерционного первого по-
рядка
- интегрирующее
- — идеальное
- колебательное
- консервативное
- простое
- типовое
- элементарное
- Знак перехода
- — левый
- — правый
- Значение смыслового сигнала
- Значения начальные сигнала и
его производных
- Зона нечувствительности
- поиска
- Избыточность кода
- Износ
- моральный
- Изображение корреляционной
функции
- по Лапласу
- — одностороннее
- — Фурье
- Изоморфизм
- Импликация
- Импульс
- мгновенный
- — единичный
- прямоугольный
- Инерционность
- Интеграл Дюамеля
- Лапласа
- обобщенный
- Фурье
- Интенсивность отказов
- — системы
- потока отказов
- Информация
- командная
- новая
- скрытая
- Итерация
- Канал связи
- —, пропускная способность
- — обратной
- Квантование
- по времени

- — уровню
- — — и по времени
- Кибернетика
- техническая
- Ключ
- Код
- оптимальный
- равномерный
- самокорректирующийся
- «четыре из семи»
- Шеннона—Фано
- Кодирование
- Колебания незатухающие
- Колебательность
- Количество информации на сим-
вол среднее
- Контур графа
- Контуры графа некасающиеся
- Конъюнкция
- Коэффициент готовности
- усиления
- Кратность резервирования
- Кривая Михайлова
- Критерии устойчивости алгеб-
раические
- — частотные
- Критерий амплитудный
- Вышнеградского
- Гурвица
- качества
- Михайлова
- Найквиста
- — для импульсных систем
- оптимальности
- — бейссов
- — интегральный
- — — обобщенный
- — регулярный
- — статистический
- Раута
- фазовый
- эффективности действия
- Линеаризация
- гармоническая
- Линейка спиральная
- Логика математическая
- Люфт
- Машина
- информационная
- самопрограммирующая
- управляющая
- — электронная (ЭУМ)
- цифровая вычислительная
(ЦВМ)

— — — универсальная
 — энергетическая
 Метод Гаусса — Зайделя
 — градиента
 — корневых годографов
 — малых возмущений
 — наискорейшего спуска
 — осреднения
 Методы построения переходных
 процессов частот-
 ные
 Механизм исполнительный
 Модель линейная
 Модуль частотной характери-
 сти-
 ки
 Мощность информационная
 — сигнала
 Надежность
 — конструктивная
 Нарботка на отказ
 Насыщение
 Негэнтропия
 Нелинейность
 — существенная
 Неопределенность
 Неупорядоченность
 Нуль
 Область притяжения точки
 покоя
 — управления
 — устойчивости
 Объект управления
 Ограничение
 Ожидание математическое
 Октава
 Оператор
 — воздействия
 — собственный
 Операция Пирса
 — Шеффера
 Определитель Гурвица
 — сигнального графа
 Оптимизатор автоматический
 — — многоканальный
 — — одноканальный
 Орган исполнительный
 — регулирующий
 Оригинал
 Основание
 Отказ
 Отрицание
 Параметр
 Параметры распределенные
 Переключение обратных связей

Перекодирование
 Перестановки контурные
 Период повторения
 — рыскания
 Перцептрон
 Плоскость фазовая
 Плотность вероятностей
 — спектральная
 Поиск случайный
 — статистический
 Показатель роста
 Полоса пропускания
 Полюс
 Помеха
 Портрет фазовый
 Порядок контура
 — пути
 — системы
 Последствие отказов
 Последовательность импульсив-
 ных сигналов
 — мгновенных импульсов
 Постоянная Больцмана
 Постоянные времени
 Потеря информации
 — на поиск
 Поток информации
 Поток отказов
 Преобразование Лапласа
 — структурное
 — Фурье
 Преобразования равносильные
 Принцип информации негэнтро-
 пийный
 — Карно
 — максимума
 — многократного применения
 информации
 Приспособление
 Программа
 Программирование динамическое
 Продолжительность исправной
 работы
 Производная обобщенная
 Процесс детерминированный
 — переходный
 — самоорганизующийся
 — установившийся
 Прямая Попова
 — фазовая
 Путь прохождения сигнала
 — — — замкнутый
 — — — прямой
 Равнозначность

- Равносильность
- Разряд
- Разум искусственный
- Разбиение D
- Распределение
 - Вейбулла
 - гамма
 - дифференциальное
 - нормальное
 - — усеченное
 - обобщенное
 - Релея
 - экспоненциальное
- Реализация случайного сигнала
- Регулирование автоматическое
- Регулятор непрямого действия
 - прямого действия
- Режим установившийся
 - переходный
- Резерв горячий
 - ненагруженный
 - облегченный
- Резервирование
 - замещением
 - оптимальное
 - постоянное
 - раздельное
- Риск средний
- Рыскание
- Самовозбуждение
- Связи обратные перекрещивающиеся
- Связь обратная
 - — линий поведения
 - — отрицательная
 - — положительная
- Седло
- Сепаратриса
- Сигнал
 - детерминированный
 - дискретный
 - — в вероятностном смысле
 - дискретно-непрерывный
 - импульсивный
 - кусочно-непрерывный
 - непрерывный
 - — в вероятностном смысле
 - нестационарный
 - решетчатый
 - случайный
 - стационарный
 - стохастический
- Сигнум-реле
- Символ информационный
 - кода
 - поверочный
- Система
 - автоматическая
 - — автономная
 - — помехоустойчивая
 - — программная
 - автоматического поиска (САП)
 - — регулирования (САР)
 - восстанавливаемая
 - гомеостатическая
 - динамическая
 - дискретная
 - замкнутая
 - избыточная
 - импульсная
 - инвариантная
 - кибернетическая
 - линейная
 - минимально-фазовая
 - мультиустойчивая
 - невозстанавливаемая
 - нелинейная
 - — существенно
 - непрерывная
 - обучаемая
 - — автоматически
 - обучающаяся
 - оптимального регулирования (СОР)
 - открытая
 - поисковая
 - полуавтоматическая
 - прогнозирующая
 - разомкнутая
 - с сосредоточенными параметрами
 - самонастраивающаяся (СНС)
 - — аналитическая
 - — дифференциальная
 - — поисковая
 - — с замкнутым циклом настройки
 - — — моделью
 - — — разомкнутым циклом самонастройки
 - самообучаемая
 - самоорганизующаяся
 - следящая
 - стабилизирующая
 - структурно неустойчивая
 - — устойчивая
 - счисления

- — оптимальная
- управляющая
- — автоматическая (АУС)
- — органическая (ОУС)
- устойчивая
- «хищник-жертва»
- экстремального регулирования (СЭР)
- — —, реагирующая на изменение производной
- — — с запоминанием экстремума
- — — — непрерывным поисковым сигналом
- — — шагового типа
- Скачок единичный
- Сложение логическое
- Смещение относительное
- Смысл сообщения
- Соединение основное
- параллельное
- последовательное
- резервное
- резервных элементов общее
- — — раздельное
- Составляющая импульсивная
- нерегулярная
- регулярная
- Состояние системы динамической
- Спектр белый
- комплексный
- Среднее по множеству
- статистическое
- Средство наблюдения
- Срок окупаемости
- Стабилизация
- Стратегия управления
- Структура топологическая
- Суперпозиция реакций
- Схема логическая алгоритма
- структурная
- функциональная
- Схемы равносильные
- Теорема Котельникова
- Теория автоматического регулирования
- — управления
- Точка изображающая
- покоя
- — неустойчивая
- — устойчивая
- Траектория корней
- фазовая

- Трапеция единичная
- Трение жидкое
- сухое
- Узел
- Ультраустойчивость
- Умножение логическое
- Упорядоченность
- Управление
- дуальное
- комбинированное
- оптимальное
- по возмущению
- — отклонению
- Уравнение Беллмана
- движения
- в обобщенных производных
- — — — с постоянными коэффициентами
- — линейное
- — нелинейное
- — операторное
- дифференциальное переходного процесса
- линейного приближения
- узловое
- характеристическое
- — в форме Вышнеградского
- — неполное
- Усилитель
- мыслительных способностей
- Условие логическое
- остановки
- тождественно-ложное
- устойчивости необходимое
- — — и достаточное
- Условия начальные
- нулевые
- — — слева
- — справа
- Устойчивость абсолютная
- «в целом»
- импульсных систем
- параметрическая
- Устройство счетно-решающее
- управляемое
- управляющее
- Фаза частотной характеристики
- Флюктуация
- Форма импульсивного сигнала
- нормальная системы дифференциальных уравнений
- Формула Мейсона
- Тронского

- Хинчина
- Шеннона
- Формулы Виета
- Функционал обобщенный
- Функция булева
- весовая
- двухместная
- единичная
- — запаздывающая
- — решетчатая
- корреляционная
- — взаимная
- логическая
- описывающая
- — нелинейности
- передаточная
- — импульсной системы
- переключательная
- распределения вероятностей
- — — интегральная
- решетчатая
- — смещенная
- характеристическая
- цены ошибок
- Характеристика амплитудная
- амплитудно-фазовая
- — — второго рода
- — — первого рода
- временная
- импульсная
- несимметричная
- переходная
- релейная
- статическая
- фазовая
- частотная
- — логарифмическая амплитудная (ЛАХ)

- — — —, крутизна
- — трапецеидальная
- — — типовая
- — фазовая
- Ценность сообщения
- Центр
- Цепочка разомкнутая
- Цикл рыскания предельный
- Частота
- квантования
- относительная
- сопрягающая
- среза
- Часть приведенная непрерывная (ПНЧ)
- Числа характеристические Выш-неградского
- Число отказов среднее
- Чувствительность
- Шаг квантования
- Шкала времени
- Шум
- Экономичность кода
- Элемент
- И
- ИЛИ
- импульсный
- — простейший (ПИЭ)
- направленного действия
- НЕ
- формирующий
- чувствительный
- Энтропия
- физическая
- Эффект экономический
- Ядро изображения
- Ящик черный

ОГЛАВЛЕНИЕ

Предисловие	3
-----------------------	---

Глава I

Введение в кибернетику

§ 1.1. Предмет и история кибернетики	5
§ 1.2. Энтропия	8
§ 1.3. Информация	15
§ 1.4. Информационные машины	24
§ 1.5. Кодирование информации	30

Глава II

Принципы действия и способы описания автоматических систем

§ 2.1. Задачи технической кибернетики	41
§ 2.2. Принципы действия автоматических систем	44
§ 2.3. Функциональные схемы и блок-схемы	53
§ 2.4. Алгоритмы	62
§ 2.5. Булева алгебра	70
§ 2.6. Сигналы	76

Глава III

Преобразование сигналов

§ 3.1. Математическое описание сигналов	80
§ 3.2. Производные реальных сигналов	89
§ 3.3. Преобразования Лапласа и Фурье	95
§ 3.4. Случайные сигналы	106
§ 3.5. Квантование сигналов	122

Глава IV

Уравнение движения и динамические характеристики автоматических систем

§ 4.1. Уравнения движения непрерывных систем	128
§ 4.2. Частотные характеристики непрерывных систем	140
§ 4.3. Логарифмические частотные характеристики	146
§ 4.4. Временные характеристики непрерывных систем	151
§ 4.5. Характеристики импульсных систем	155

Глава V

Структурные преобразования автоматических систем

§ 5.1. Структурные схемы и сигнальные графы	160
§ 5.2. Синтез передаточных функций	163
§ 5.3. Тожественные преобразования структурных схем	181
§ 5.4. Типовые и элементарные звенья	184
§ 5.5. Замыкание импульсной системы	199

Глава VI

Устойчивость линейных автоматических систем

§ 6.1. Необходимое и достаточное условие устойчивости	201
§ 6.2. D -разбиение	203
§ 6.3. Алгебраические критерии	206
§ 6.4. Диаграмма Вышнеградского	212
§ 6.5. Частотные критерии	224
§ 6.6. Устойчивость импульсных систем	232

Глава VII

Процессы в линейных системах

§ 7.1. Суперпозиция выходных сигналов	235
§ 7.2. Решение операторного уравнения движения непрерывной системы	237
§ 7.3. Дифференциальные уравнения переходных процессов	251
§ 7.4. Построение переходного процесса по временным характеристикам	257
§ 7.5. Частотные методы построения переходных процессов	260
§ 7.6. Корневые годографы	270
§ 7.7. Случайные процессы	289
§ 7.8. Процессы в дискретных системах	291

Глава VIII

Непрерывные нелинейные автоматические системы

§ 8.1. Нелинейности	294
§ 8.2. Влияние нелинейности на случайные сигналы	298
§ 8.3. Нелинейные системы. Уравнения движения в нормальной форме	303
§ 8.4. Фазовое пространство автономной системы	317
§ 8.5. Устойчивость нелинейной системы	325
§ 8.6. Абсолютная устойчивость	331

Глава IX

Самоприспосабливающиеся системы

§ 9.1. Системы экстремального оптимального регулирования	339
--	-----

§ 9.2. Самонастраивающиеся системы	363
§ 9.3. Самоорганизующиеся системы	377
§ 9.4. Обучающиеся системы	390
§ 9.5. Оптимальное управление	394

Глава X

Надежность автоматических управляющих систем

§ 10.1. Количественные характеристики надежности . . .	406
§ 10.2. Законы распределения времени возникновения отказа	411
§ 10.3. Характеристики надежности системы	421
§ 10.4. Резервирование	430
§ 10.5. Экономика надежности	443
Литература	446
Предметный указатель	454

Ростислав Алексеевич Сапожников
Павел Николаевич Матвеев
Борис Павлович Родин
Наталия Анатольевна
Филадельфина

**ОСНОВЫ
ТЕХНИЧЕСКОЙ
КИБЕРНЕТИКИ**

Редактор *А. М. Кравцова*
Художественный редактор *Т. А. Дурасова*
Художник *В. В. Кулешов*
Технический редактор *Э. М. Чижевский*
Корректор *М. И. Кряковкина*

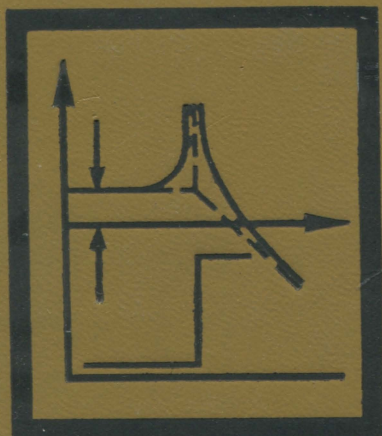
Т-11221. Сдано в набор 6/III-70 г. Подп.
к печати 30/VII-70 г. Формат $84 \times 108^{1/32}$.
Объем 14,5 печ. л., усл. п. л. 24,36. Уч.-
изд. л. 20,41. Изд. № СТД-56. Тираж 28 000
экз. Цена 89 коп. Зак. 819

БЗ—35/25 от 1/VI-70 г.

Москва, К-51, Неглинная ул., д. 29/14
Издательство «Высшая школа»

Ордена Трудового Красного Знамени Ленин-
градская типография № 1 «Печатный Двор»
им. А. М. Горького Главполиграфпрома
Комитета по печати при Совете Министров
СССР, г. Ленинград, Гатчинская, 26.

Р. А. САПОЖНИКОВ



ОСНОВЫ ТЕХНИЧЕСКОЙ КИБЕРНЕТИКИ

