Министерство образования и науки РФ

Федеральное государственное бюджетное образовательное учреждение высшего профессионального образования
ТОМСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ СИСТЕМ УПРАВЛЕНИЯ И РАДИОЭЛЕКТРОНИКИ (ТУСУР)

Радиотехнический факультет (РТФ)

Кафедра средств радиосвязи (СРС)

В.А. Кологривов

Основы автоматизированного проектирования радиоэлектронных устройств

Часть 2

Учебное пособие
для студентов радиотехнических специальностей

Рекомендовано Сибирским региональным отделением УМО высших учебных заведений РФ по образованию в области радиотехники, электроники, биомедицинской техники и автоматизации для межвузовского использования в качестве учебного пособия

2012

Кологривов В.А.

Основы АПР РЭУ: Учебное пособие. В 2-х частях. Часть 2. Учебное пособие для студентов направлений радиотехника и телекоммуникации. – Томск: ТУСУР. Образовательный портал, 2012. - 132 с.

Рассмотрены модели основных элементов электронных схем, современные методы формирования математических моделей, решения линейных и нелинейных систем алгебраических и дифференциальных уравнений и оптимизации характеристик устройств. Изложение методов и алгоритмов ориентировано на реализацию программ автоматизированного схемотехнического проектирования радиоэлектронных устройств.

Рекомендовано для студентов высших учебных заведений, обучающихся по направлениям подготовки дипломированных специалистов 654200 “Радиотехника” специальностей 200700, “Радиотехника” 201600 Радиоэлектронные системы; 654400 “Телекоммуникации” специальностей 071700 “Физика и техника оптической связи”, 201100 “Радиосвязь, радиовещание и телевидение”, 201200 “Средства связи с подвижными объектами”, 201800 “Защищенные телекоммуникационные системы”, а также специальности 075400 “Комплексная защита объектов информации”.

АННОТАЦИЯ

Данное учебное пособие представляет собой расширенный конспект лекций по основам автоматизированного схемотехнического проектирования радиоэлектронных устройств для студентов радиотехнических и связных специальностей очной и заочной форм обучения.

Особенностью данного пособия является систематическое использование модифицированных методов формирования математических моделей, позволяющих с единой позиции изложить методологию расчета рабочих режимов, частотных и временных характеристик, чувствительности их к изменению параметров устройств и внешних факторов.

Достаточно подробно рассмотрены модели основных элементов электронных схем, современные методы формирования математических моделей, решения линейных и нелинейных систем алгебраических и дифференциальных уравнений и оптимизации характеристик устройств.

Изложение методов и алгоритмов ориентировано на реализацию программ автоматизированного схемотехнического проектирования радиоэлектронных устройств.

Пособие может быть рекомендовано и студентам смежных специальностей по направлениям радиоэлектроники и телекоммуникаций, интересующимся вопросами автоматизированного схемотехнического проектирования радиоэлектронной аппаратуры.

СОДЕРЖАНИЕ

ПРЕДИСЛОВИЕ

Предлагаемое вниманию пособие представляет собой расширенный конспект лекций по основам автоматизированного схемотехнического проектирования радиоэлектронных устройств.

Пособие адресовано инженерно-техническим работникам, занимающимся вопросами автоматизации проектирования в радиоэлектронике, и студентам старших курсов радиотехнических специальностей, имеющих базовую подготовку по математике, программированию, теории цепей и сигналов и элементной базе радиоэлектронных устройств.

Имеющаяся литература по автоматизированному схемотехническому проектированию отражает в основном подход, базирующийся на методе переменных состояния, отличающегося повышенной сложностью формирования математической модели для цепей произвольного вида. В настоящее время при автоматизации схемотехнического проектирования предпочтение отдается подходу, использующему прямые методы формирования математической модели, в которых, соответствующая система уравнений, формируется непосредственно по схеме устройства. Суть данного подхода обстоятельно изложена в монографии: Влах И., Сингхал К. Машинные методы анализа и проектирования электронных схем: Пер. с англ. – М.: Радио и связь, 1988. - 560 с., вышедшей тиражом 20000 экземпляров.

В предлагаемом пособии сделана попытка адаптированного изложения подхода, развиваемого в упомянутой монографии, применительно к нашим условиям, в виде расширенного конспекта лекций по дисциплине «Основы автоматизированного проектирования радиоэлектронных устройств», читаемой студентам радиотехнического факультета ТУСУРа. При адаптации некоторые разделы исключены, другие – переработаны и дополнены, в соответствии с рабочей программой. В основном переработаны и дополнены разделы, посвященные моделям элементной базы, методам решения систем линейных алгебраических уравнений и методам оптимизации характеристик устройств.

Содержание данного пособия отражает опыт преподавания данной дисциплины для студентов специальности “Радиотехника” на радиотехническом факультете ТУСУРа с 1991 года, в соответствии с образовательным стандартом.

В заключение хотелось бы выразить слова благодарности всем сотрудникам кафедр РЗИ и СРС, оказавших помощь при постановке данной дисциплины и проведении занятий. Персонально слова благодарности выражаю коллегам, сотрудникам кафедры РЗИ М.Ю. Покровскому и А.С. Красько, которые в течение ряда лет помогали в проведении курсовых и лабораторных работ со студентами. Красько А.С. благодарен также за большую помощь, оказанную при подготовке электронного варианта конспекта лекций. Особые слова благодарности выражаю Г.Н. Глазову,

прочитавшему первый вариант рукописи и сделавшему целый ряд ценных замечаний.

ВВЕДЕНИЕ

В настоящее время инженер в процессе своей деятельности нередко использует ЭВМ для проведения различных вычислений, а в ряде случаев проектировщик современной аппаратуры просто не может обойтись без ЭВМ, как основного рабочего инструмента. При этом разработчик использует современные системы автоматизированного проектирования радиоэлектронных устройств (РЭУ) либо решает нестандартные задачи проектирования, опираясь на системы для инженерных и научных расчетов.

Автоматизированное проектирование позволяет существенно сократить финансовые затраты и время на разработку радиоэлектронной аппаратуры (РЭА), повышая точность расчетов и сокращая объем экспериментальных исследований. Продуктивное использование ЭВМ немыслимо без развитого прикладного программного обеспечения, позволяющего быстро и надежно моделировать и оптимизировать предлагаемые решения. В связи, с выше указанным, актуальна подготовка современных специалистов, владеющих основами автоматизированного проектирования.

Предлагаемое вниманию пособие представляет собой расширенный конспект лекций по дисциплине «Основы автоматизированного проектирования радиоэлектронных устройств» (Основы АПР РЭУ) и предназначено для подготовки студентов радиотехнических специальностей.

Цель данного пособия – раскрыть содержание, принципы и методологию современного состояния автоматизированного схемотехнического проектирования.

В задачи дисциплины входит – изучение: моделей элементной базы; современных методов и алгоритмов формирования математических моделей расчета режимов, частотных и временных характеристик, чувствительности к изменению параметров и оптимизация характеристик устройств.

Для достижения указанной цели и решения поставленных задач в конспекте лекций излагаются основные понятия и определения, модели элементной базы, совокупность современных методов и алгоритмов расчета основных характеристик, принципы построения программного обеспечения и методология автоматизированного схемотехнического проектирования.

При изложении материала, из всего многообразия тем, относящихся к автоматизированному схемотехническому проектированию, нами выбран минимум необходимый для овладения проблематикой, методологией, основными принципами, методами и алгоритмами, позволяющий решать широкий круг полезных задач.

Основное внимание при изложении дисциплины уделено расчету электронных схем, как электрическим моделям реальных узлов РЭУ. Однако

вопросы автоматизированного расчета и проектирования невозможно охватить в одном пособии, поэтому пришлось ограничиться в основном аналоговыми устройствами, хотя излагаемые методы могут быть, в большинстве своем, распространены и на дискретные или импульсные устройства. К сожалению, за пределами нашего внимания, кроме дискретных устройств, остаются: специфические моменты расчета и проектирования распределенных устройств СВЧ диапазона, спектральные задачи нелинейных устройств, электродинамический расчет конструкций и целый ряд других не менее важных вопросов. Частично с этими вопросами можно ознакомиться при изучении других дисциплин, а идеи и методы данного предмета помогут Вам успешно освоить перечисленные разделы моделирования, расчета и проектирования.

В пособии, для линейных и нелинейных аналоговых устройств, даны основные понятия и определения, рассмотрены модели основных элементов, изложены методы и алгоритмы формирования математических моделей, расчета частотных и временных характеристик, режима по постоянному току, чувствительности к изменению параметров и внешних факторов, методология и принципы автоматизированного проектирования с использованием методов параметрического синтеза.

В основу курса лекций вместо традиционного метода переменных состояния положены более простые, но не менее эффективные, прямые методы формирования математических моделей, совмещающие достоинства узлового и табличного методов, позволяющих реализовать идею сквозного проектирования. Суть данного подхода изложена в прекрасной монографии Влаха И. и Синкхала К. [1]. Этого подхода, как наиболее удачного, мы и придерживаемся в данном пособии. Структура конспекта лекций, методология и часть примеров заимствованы из этой монографии. Естественно, что, исходя из рабочей программы, собственного опыта и интересов, часть специфических разделов были исключены, сокращены или переработаны, но идеология изложения материала по возможности сохранена. Часть отсутствующего материала, разбросанного по разным учебникам, добавлена в качестве новых разделов.

Материал по разделам распределен следующим образом:

1. В первом разделе сформулированы цели, задачи и содержание автоматизированного проектирования РЭУ.
2. Во втором разделе даны основные элементы топологического описания электронных схем.
3. В третьем разделе дано обоснование классических методов (узлового и контурного) формирования математических моделей с позиций компонентного и топологического описания электронных схем.
4. В четвертом разделе дано развернутое содержание прямых методов формирования математических моделей, их сравнительная характеристика и основной идеи сквозного проектирования.

5. В пятом разделе рассмотрено содержание понятия модели, их классификация и описание моделей основных элементов РЭУ.
6. В шестом разделе излагаются основные методы решения систем линейных алгебраических уравнений общего вида, на которых собственно и базируются все вычислительные алгоритмы.
7. В седьмом разделе изложены методы расчета передаточных характеристик электронных схем и сопутствующие вопросы.
8. Восьмой раздел посвящен алгоритмам расчета чувствительности электронных схем к изменению параметров и их использованию для вычисления других характеристик.
9. Девятый раздел посвящен расчету режимов цепей по постоянному току и вопросам сходимости алгоритмов.
10. Десятый раздел посвящен вопросам численного интегрирования дифференциальных уравнений, способам их формирования и расчету временного отклика цепей.
11. В последнем разделе рассмотрены постановка задачи оптимизации, основные понятия, методы и алгоритмы оптимизации, а также вопросы автоматизации проектирования РЭУ с заданными характеристиками.

По краткому содержанию пособия следует заметить, что первые три раздела закладывают основные понятия описания электронных схем, как моделей реальных устройств. Содержание подхода сквозного проектирования базируется на прямых методах формирования математических моделей и в этом смысле четвертый раздел наиболее важен. Модели элементной базы РЭУ лежат в основе их компьютерного моделирования. Методы решения систем линейных алгебраических уравнений лежат в основе практически всех алгоритмов расчета характеристик. Численные методы интегрирования дифференциальных уравнений и проблема обеспечения их точности и устойчивости являются базовыми для расчета реакции устройств во временной области. Последующие разделы конкретизируют алгоритмы расчета основных характеристик в частотной и временной областях. Чувствительность характеристик к изменению параметров устройств важны, как на этапе производства, так и эксплуатации РЭУ. Расчет режимов цепей по постоянному току всегда предшествует расчету любых характеристик, так как режимы работы активных приборов (рабочие точки) в основном и определяет параметры реальных устройств. Оптимизация рассматривается в данной дисциплине, как основной прием автоматизированного проектирования узлов РЭУ с заданными характеристиками, именно параметрический синтез лежит в основе методологии автоматизированного проектирования.

Изучение всех перечисленных вопросов и их взаимосвязи и составляет основу автоматизированного проектирования РЭУ.

7 ПЕРЕДАТОЧНЫЕ ХАРАКТЕРИСТИКИ ЭЛЕКТРОННЫХ СХЕМ

7.1 Классический подход

Обобщенный узловой метод. Под передаточными характеристиками будем понимать совокупность характеристик определяемых отношениями токов и напряжений в различных частях схемы (чаще всего вход - выход). Поскольку понятие узла обобщенного узлового метода близко к понятию зажима (входа) он позволяет определить основные передаточные характеристики достаточно широкого класса схем.

Как известно система уравнений обобщенного узлового метода имеет вид

$$I = Y \cdot U, \tag{7.1}$$

где I - вектор задающих токов (источников сигнала); Y - матрица проводимости схемы; U - вектор искомых узловых напряжений. В общем, виде решение системы можно записать в виде

$$U = Y^{-1} \cdot I. \tag{7.2}$$

Откуда следует, что в основу определения передаточных характеристик должен быть положен вектор узловых напряжений. В этом смысле обобщенный узловой метод мало, чем отличается от других методов формирования математических моделей, и подход к вычислению передаточных характеристик которых рассмотрим позднее.

Начнем же изложение методов вычисления передаточных характеристик с классического метода, основанного на вычислениях алгебраических дополнений. Алгебраическое дополнение, как известно из линейной алгебры, это определитель матрицы образованный вычеркиванием соответствующих строк и столбцов, знак которого уточняется множителем $(-1)^{m+p}$, где m- сумма индексов вычеркнутых строк и столбцов, p - число перестановок индексов.

В соответствии с правилом Крамера компоненты вектора решений можно записать в виде

$$u_j = {}_j\Delta / \Delta, \tag{7.3}$$

где ${}_j\Delta$ - определитель, матрицы Y, в котором j - тый столбец, заменен вектором свободных членов I; Δ - определитель исходной матрицы.

При выводе соотношений полагаем, что вход многополюсника образован одним из узлов относительно общего, а выход - другим узлом относительно общего. Под передаточной характеристикой схемы (многополюсника) понимается ее реакция на входное воздействие при отсутствии других воздействий. В результате при расчете передаточной характеристики только одна компонента вектора тока с индексом соответствующим входному узлу отлична от нуля. Кроме того, при выводе соотношений будем различать определители матрицы проводимости

собственно схемы и матрицы проводимости с внесенными проводимостями источника сигнала и нагрузки.

Предположим, что источник тока (входного сигнала) действует на i-том входе (узле). Тогда напряжение

$$u_i = {}_i\Delta / \Delta,$$

или учитывая, что

$${}_i\Delta = I_i \cdot \Delta_{ii}$$

получаем

$$u_i = I_i \cdot \Delta_{ii} / \Delta.$$

Откуда входное сопротивление i-го входа равно

$$Z_i = \Delta_{ii} / \Delta. \qquad (7.4)$$

Напряжение на j-том узле (выходе) от источника тока подключенным к i-му узлу (входу) определится выражением

$$u_j = {}_j\Delta / \Delta$$

или учитывая, что

$${}_j\Delta = I_i \cdot \Delta_{ij}$$

имеем

$$u_j = I_i \cdot \Delta_{ij} / \Delta.$$

Откуда передаточное сопротивление со входа i на выход j равно

$$Z_{ij} = u_j / I_i = \Delta_{ij} / \Delta. \qquad (7.5)$$

Раскрыв отношение u_j / u_i из предыдущих выражений получаем коэффициент передачи по напряжению с i-го узла (входа) на j-ый узел (выход)

$$K_{Uij} = u_j / u_i = \Delta_{ij} / \Delta_{ii}. \qquad (7.6)$$

Обозначив сопротивление нагрузки на выходе (j-ый узел), через Z_L, запишем очевидное соотношение

$$u_j = I_j \cdot Z_L,$$

и заменяя u_j его выражением, получим

$$u_j = I_j \cdot Z_L = I_i \cdot \Delta_{ij} / \Delta.$$

Откуда получаем коэффициент передачи по току с i-го узла (входа) на j-ый узел (выход)

$$K_{Iij} = I_i / I_j = \Delta_{ij} / (Z_L \cdot \Delta). \qquad (7.7)$$

Получение выражений для других передаточных характеристик не представляет затруднений.

Отметим, что во всех предыдущих выражениях проводимость нагрузки подразумевалась внесенной в матрицу проводимости схемы. Используя свойства определителей, можно записать

$$\Delta = \hat{\Delta} + Y_L \cdot \Delta_{jj}, \qquad (7.8)$$

где $\hat{\Delta}$ - определитель матрицы проводимости без проводимости нагрузки; Y_L - проводимость нагрузки. Следует также подчеркнуть, что реальные нагрузки других узлов (зажимов), кроме входного и выходного, подразумеваются, внесенными в матрицу проводимости схемы.

Полученные выражения справедливы для матриц любого порядка, а также для матриц полученных в результате исключения переменных (токов и напряжений внутренних узлов) при понижении их порядка. Столь же просто можно получить передаточные выражения для случая, когда входная и выходная пары зажимов не имеют общего узла.

Аналогично можно получить выражения для передаточных характеристик схемы и в методе контурных токов, используя матрицу сопротивлений холостого хода Z.

Замечания относительно вычислений передаточных характеристик:

1. Прямое вычисление алгебраических дополнений нецелесообразно, т.к. вычисление каждого дополнения, как и определителя, по числу операций соизмеримо с решением исходной системы уравнений.

2. Предварительное обращение матрицы, в результате которого каждый элемент заменяется его алгебраическим дополнением, деленным на определитель, в некоторых случаях может оказаться избыточным, однако приемлемо для универсальных программ, содержащих расчет чувствительности, шумов, нелинейных эффектов, когда требуется определение реакции практически с любого узла.

3. Предварительное приведение схемы к внешним зажимам (вход - выход), путем исключения переменных (токов и напряжений внутренних узлов), наиболее предпочтительна с алгоритмической точки зрения в силу последовательного понижения порядка системы уравнений, что эквивалентно одновременному вычислению требуемого набора алгебраических дополнений, однако смена координат входа - выхода требует повторения вычислений.

Метод подсхем. Необходимо несколько подробнее остановиться на методе подсхем, позволяющем вести расчет сложных линейных электронных схем по частям. Подход к расчету сложных схем, основанный на методе подсхем, способствует понижению порядка решаемой на каждом этапе формирования систем уравнений, что сокращает требуемое время и память.

Подсхемой, как известно, называется независимая часть схемы. Независимость подсхемы подразумевает, например, выполнение таких требований, как принадлежность зависимых источников целиком одной из подсхем. Коррелированные шумовые источники также не могут принадлежать разным подсхемам.

В силу независимости подсхем, как было отмечено ранее (раздел 3, пункт 3.2), взаимные проводимости разных подсхем в общей матрице проводимости схемы равны нулю. Собственные проводимости общих узлов (соединений) равны алгебраической сумме собственных проводимостей

подсхем. Заметим, что схема по отношению к подсхемам выступает также как независимая часть.

Использование метода подсхем, в силу сделанных замечаний позволяет, с целью снижения порядка систем уравнений, предварительно составить уравнения подсхемы, исключить токи и напряжения, соответствующие внутренним узлам, и внести результирующую матрицу коэффициентов в общую матрицу схемы. Этот прием будем для краткости называть приведением подсхемы к внешним узлам (зажимам).

Действительно, если взять полную матрицу схемы до разбиения на подсхемы и применить исключение переменных, соответствующих внутренним узлам, то в соответствии с алгоритмом Гаусса можем записать

$$\hat{y}_{ij} = y_{ij} - y_{ik} \cdot y_{kj} / y_{kk}, \qquad (7.9)$$

где k - индекс исключаемого узла схемы, принадлежащего подсхеме. Если при этом узлы i и j не принадлежат подсхеме, то в силу ее независимости имеем

$$y_{ik} = y_{kj} = 0,$$

откуда получаем

$$\hat{y}_{ij} = y_{ij},$$

т.е. элементы матрицы проводимости, не принадлежащие подсхеме, при исключении ее внутреннего узла остаются без изменений. Для элементов матрицы проводимости схемы, соответствующих общим узлам (соединению подсхем между собой либо со схемой) i и j, и исключении узла k, принадлежащего подсхеме, соотношение (7.9) можно переписать в виде

$$\hat{y}_{ij} = y_{ext\_ij} + y_{ins\_ij} - y_{ik} \cdot y_{kj} / y_{kk}, \qquad (7.10)$$

где y_{ext_ij} - проводимость внешняя по отношению к данной подсхеме; y_{ins_ij} - проводимость собственно подсхемы. Отсюда следует, что элементы матрицы проводимости схемы, принадлежащие подсхеме, после исключения внутренних переменных, по прежнему входят в общую матрицу аддитивно. Таким образом, исключение внутренних переменных независимых подсхем можно провести предварительно, до внесения в общую матрицу проводимости.

В общем случае, можно строго доказать, что применение метода подсхем допустимо в тех системах параметров, где общие элементы матрицы коэффициентов можно представить линейным функционалом, для которого, как известно, выполняется свойство аддитивности.

Следует также отметить, что число подсхем и уровней подсхем может быть произвольно, необходимо лишь соблюдение условия независимости.

В качестве иллюстрации вычисления передаточных характеристик методом подсхем рассмотрим простую схему транзисторного каскада с комбинированной обратной связью (см. рисунок 7.1).

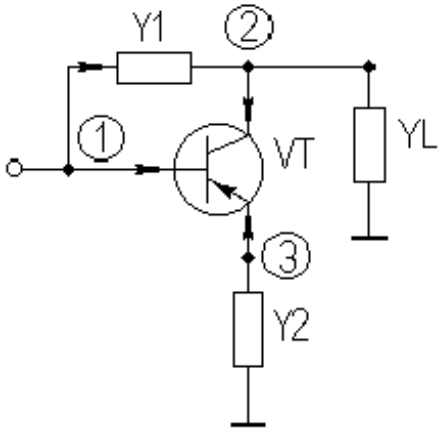


Рисунок 7.1- Схема транзисторного каскада

Пусть транзистор описан матрицей Y- параметров по схеме с ОЭ как четырехполюсника

$$Y = \begin{matrix} \\ b \\ c \end{matrix} \begin{matrix} \begin{matrix} b & c \end{matrix} \\ \begin{bmatrix} y_{11} & y_{12} \\ y_{21} & y_{22} \end{bmatrix} \end{matrix}.$$

Тогда на том основании, что неопределенная матрица проводимости любой схемы, не имеющей общего узла, имеет сумму элементов по любой строке и столбцу равную нулю, получим неопределенную матрицу проводимости транзистора через матрицу проводимости для схемы с ОЭ. Отсоединяя эмиттерный узел, от общего узла и дополняя матрицу строкой и столбцом, с элементами дающими сумму элементов по любой строке и столбцу равную нулю, получим неопределенную матрицу проводимости транзистора

$$Y = \begin{matrix} \\ b \\ c \\ e \end{matrix} \begin{matrix} \begin{matrix} b & c & e \end{matrix} \\ \begin{bmatrix} y_{11} & y_{12} & -(y_{11}+y_{12}) \\ y_{21} & y_{22} & -(y_{21}+y_{22}) \\ -(y_{11}+y_{21}) & -(y_{12}+y_{22}) & \sum y \end{bmatrix} \end{matrix},$$

где $\sum y = y_{11} + y_{12} + y_{21} + y_{22}$.

Запишем матрицу проводимости транзисторного каскада с комбинированной ОС, используя рассмотренные ранее (раздел.3, пункт 3.2) правила формирования

$$Y = \begin{matrix} \\ 1 \\ 2 \\ 3 \end{matrix} \begin{matrix} \begin{matrix} 1 & 2 & 3 \end{matrix} \\ \begin{bmatrix} y_{11}+Y_2 & -(y_{11}+y_{12}) & y_{12}-Y_2 \\ -(y_{11}+y_{22}) & \sum y + Y_1 & -(y_{12}+y_{22}) \\ y_{21}-Y_2 & -(y_{21}+y_{22}) & y_{22}+Y_2+Y_L \end{bmatrix} \end{matrix}.$$

Входное сопротивление каскада и коэффициенты передачи по напряжению и току определятся выражениями

$$Z_{in} = \Delta_{11} / \Delta;$$
$$K_u = \Delta_{13} / \Delta;$$
$$K_I = \Delta_{13} \cdot Y_L / \Delta.$$

Если же, в соответствии с алгоритмом Гаусса, из матрицы проводимости предварительно исключить переменные, соответствующие току и напряжению внутреннего узла 2, то получим матрицу эквивалентного четырехполюсника

$$\hat{Y} = \begin{bmatrix} \Delta_{22} / \Delta_{12,12} & \Delta_{21} / \Delta_{12,12} \\ \Delta_{12} / \Delta_{12,12} & \Delta_{11} / \Delta_{12,12} \end{bmatrix},$$

переобозначив выходной узел с 3 на 2. Определитель этой матрицы в соответствии с теоремой Якоби равен

$$\hat{\Delta} = (\Delta_{33} \cdot \Delta_{11} - \Delta_{13} \cdot \Delta_{31}) / \Delta^2_{13,13} = \Delta \cdot \Delta_{13,13} / \Delta^2_{13,13} = \Delta / \Delta_{13,13}.$$

Вновь определим названные передаточные характеристики, но теперь по матрице проводимости каскада $\hat{Y}$, приведенной к внешним зажимам (вход - выход)

$$Z_{in} = \hat{\Delta}_{11} / \hat{\Delta} = (\Delta_{11} / \Delta_{13,13}) / (\Delta / \Delta_{13,13}) = \Delta_{11} / \Delta;$$
$$K_u = \hat{\Delta}_{12} / \hat{\Delta}_{11} = (\Delta_{13} / \Delta_{13,13}) / (\Delta_{11} / \Delta_{13,13}) = \Delta_{13} / \Delta_{11};$$
$$K_I = \hat{\Delta}_{12} \cdot Y_L / \hat{\Delta}(\Delta_{13} \cdot Y_L / \Delta_{13,13}) / (\Delta / \Delta_{13,13}) = \Delta_{13} \cdot Y_L / \Delta.$$

Из полученных выражений видно, что передаточные характеристики совпадают с таковыми, полученными из полной матрицы проводимости. Тем самым, мы подтвердили правомерность идеи метода подсхем.

7.2 Функции цепи в современных методах

Передаточные характеристики в современной литературе называют иногда обобщенно функциями цепи либо схемы, понимая под этим зачастую не только передаточные характеристики, но и переходные, чувствительность и т.д.

Рассмотрим здесь альтернативный подход к определению функций цепи - передаточных характеристик по результатам решения соответствующих систем уравнений математической модели. Изложение, в основном, сориентируем на современные методы формирования математических моделей, такие как табличный, модифицированный табличный, модифицированный узловой, модифицированный узловой с проверкой. Однако данный подход в полной мере применим к классическим методам - обобщенных узловых потенциалов и контурных токов.

Математическая модель - алгебраическая система уравнений современных методов может быть записана в виде

$$T \cdot X = W, \qquad (7.11)$$

где T - матрица коэффициентов системы уравнений; W - вектор свободных членов системы; X - вектор неизвестных или вектор решений системы. В зависимости от метода формирования математической модели вектор неизвестных включает в себя в качестве переменных напряжения и токи ветвей, напряжения узлов.

Обозначим входные переменные как V_{in} или I_{in}, а выходные переменные как V_{out} или I_{out}. Функция цепи определится в этом случае, как отношение, какой либо выходной переменной к входной. Для цепи (схемы), имеющей один вход и один выход, можно определить наиболее известные передаточные функций: Z_{in}, Y_{in}, Z_{out}, Y_{out} - входные и выходные сопротивления и проводимости; Z_{con}, Y_{con} - переходные сопротивление и проводимость; K_u, K_I - коэффициенты передачи напряжения и тока. Эти функции, согласно определения передаточных характеристик, рассчитываются при нулевых начальных условиях. Для схем, имеющих несколько входов и или выходов необходимо использовать соответствующие индексы.

Предположим, что исходная система уравнений решена, тогда выходная переменная (величина) F, в простейшем случае, является линейной комбинацией компонент вектора решения X

$$F = d^t \cdot X \tag{7.12}$$

или

$$F = d^t \cdot T^{-1} \cdot W = d^t \cdot (adj(T) / det(T)) \cdot W, \tag{7.13}$$

где d^t - строка, состоящая из компонент равных 0, 1, -1; $adj(T)$ - присоединенная матрица, соответствующая транспонированию исходной матрицы и замене ее элементов алгебраическими дополнениями Δ_{ij}; $det(T) = \Delta_T$ - определитель исходной матрицы. Такое определение выходной переменной позволяет найти напряжение узла, напряжение ветви и ток ветви. Более сложная форма выходной функции, соответствующая, например, передаточной характеристике, пока не рассматривается.

Напоминаем, что исходная матрица в комплексной плоскости может быть представлена в виде

$$T = G + s \cdot C,$$

где G - действительная часть матрицы коэффициентов; C - мнимая часть матрицы коэффициентов; s - оператор Лапласа при чисто гармоническом воздействии равный $j \cdot \omega$. Отсюда следует, что определитель матрицы T и алгебраические дополнения матрицы T являются полиномами от переменной s, а функция F - рациональной функцией комплексной переменной s.

Таким образом, выходная функция F может быть представлена либо отношением полиномов

$$F(s) = \sum_{i=0}^{n} a_i \cdot s^i \; / \sum_{i=0}^{m} b_i \cdot s^i \,, \tag{7.14}$$

где a_i и b_i - коэффициенты полинома, или в факторизованной форме через нули и полюса

$$F(s) = K \cdot \prod_{i=1}^{n} (s - z_i) \,/\, \prod_{i=1}^{m} (s - p_i), \tag{7.15}$$

где K - постоянный множитель; z_i - нули, а p_i - полюса данной функции цепи.

Формирование символьного представление функции цепи с помощью ЭВМ. Для изложения алгоритма вычисления коэффициентов дробно-рационального представления функции цепи запишем ее в виде

$$F(s) = N(s) \,/\, D(s), \tag{7.16}$$

где $N(s)$ - полином числителя; $D(s)$ - полином знаменателя.

Как известно, для расчета функции цепи необходимо решить исходную систему уравнений, используя, например, LU - разложение матрицы T на треугольные сомножители, и выполнив прямую и обратную подстановки. Это потребует примерно $n^3 \,/\, 3$ операций, если не учитывать разряженность матрицы. Если же представить выходную функцию в виде отношения полиномов от переменной s, то это позволит:

1) быстро вычислять значение выходной функции на различных частотах;

2) используя преобразование Лапласа, можно получить временной отклик цепи. Для того чтобы такое представление стало возможным, необходимо организовать вычисление коэффициентов полиномов числителя и знаменателя.

В табличном, модифицированным табличном, модифицированным узловом и модифицированным узловом с проверкой методах частотно - зависимые элементы входящие в матрицу T всегда можно представить в форме, содержащей переменную s в числителе ($s \cdot C$, $s \cdot L$), что позволит в последующем воспользоваться преобразованием Лапласа для перехода из частотной во временную область. Для этого необходимо емкости всегда представлять проводимостями, а индуктивности - сопротивлениями.

Числитель и знаменатель $F(s)$ это, как показали в общем случае, полиномы от переменной s, причем знаменатель $D(s)$ - соответствует определителю системы. Выберем произвольно $s = s_i$ и произведем LU - разложение

$$L \cdot U \cdot X = W.$$

С помощью прямой и обратной подстановок найдем вектор решения $X(s_i)$. Допуская, что выходная переменная определена, как $d^t \cdot X(s_i)$ получим отсчет передаточной функции

$$F(s_i) = N(s_i) \,/\, D(s_i) = d^t \cdot X(s_i). \tag{7.17}$$

Повторяя подобные вычисления, определим серию отсчетов передаточной функции, которая при необходимости может быть аппроксимирована либо интерполирована в виде полинома от переменной s.

Можно поставить задачу найти отдельно полиномы числителя и знаменателя. Прежде всего, заметим, что в нашем определении выходной функции знаменатель представляет собой определитель исходной системы. Следовательно, после определения отсчета $F(s_i)$ необходимо вычислить определитель системы при $s = s_i$. Определитель найдем, как произведение диагональных элементов матрицы L

$$D(s_i) = det(T(s_i)) = det(L(s_i)) = \prod_{j=1}^{n} l_{jj} . \qquad (7.18)$$

Теперь отсчет значения полинома числителя можно получить как отношение

$$N(s_i) = F(s_i) \cdot D(s_i) . \qquad (7.19)$$

Таким образом, перебирая s_i, получим ряд отсчетов $N(s_i)$ и $D(s_i)$, по которым, используя интерполяцию, можно найти коэффициенты полиномов $N(s)$ и $D(s)$.

Можно и наоборот, определить полиномы $D(s)$ и $F(s)$, а полином числителя определить, как произведение полиномов

$$N(s) = F(s) \cdot D(s).$$

Задача нахождения коэффициентов полинома, по его значениям, при заданных s_i, известна в математике, как задача интерполяции и при этом точность интерполяции зависит от выбора значений s_i и вида интерполирующего полинома.

7.3 Интерполяция полиномов по точкам окружности

Классическая задача интерполяция. Допустим, что известны отсчеты функции и аргумента

$$y_i = f(x_i)$$

в $(n+1)$- ой отдельной точке, причем y_i и x_i, могут быть, как вещественными, так и комплексными. Требуется найти коэффициенты полинома

$$P_n(x) = \sum_{j=0}^{n} a_j \cdot x^j , \qquad (7.20)$$

проходящего через данные точки.

Подставив соответствующие значения x_i в полином (7.20), получим систему уравнений

$$a_0 + a_1 \cdot x_i + a_2 \cdot x_i^2 + \cdots + a_n \cdot x_i^n = y_i ,$$

где $i = 0,1,\cdots,n$; a_j - неизвестные коэффициенты. Матричная форма записи данной системы

$$\begin{bmatrix} 1 & x_0 & x_0^2 & \cdots & x_0^n \\ 1 & x_1 & x_1^2 & \cdots & x_1^n \\ 1 & x_2 & x_2^2 & \cdots & x_2^n \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 1 & x_n & x_n^2 & \cdots & x_n^n \end{bmatrix} \cdot \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ \cdots \\ a_n \end{bmatrix} = \begin{bmatrix} y_0 \\ y_1 \\ y_2 \\ \cdots \\ y_n \end{bmatrix} \qquad (7.21)$$

или

$$X \cdot A = Y. \qquad (7.22)$$

Решение уравнения (7.22) позволяет определить неизвестные коэффициенты компоненты вектора A.

Возникает вопрос о наилучшем выборе точек x_i с точки зрения точности интерполяции. Как правило, интерполяция с реальными величинами x_i численно нестабильна. Можно показать, что наилучший выбор значений x_i соответствует равноотстоящим точкам, лежащим на единичной окружности комплексной плоскости.

Отметим, что здесь речь идет не о каком - то нормировании отсчетов аргумента x, а о замене этих отсчетов равным количеством отсчетов взятых равномерно на единичной окружности комплексной плоскости. Этим отсчетам, равным по модулю единице, и отличающимся только фазой ставятся в соответствие отсчеты значений интерполируемой функции и ищутся коэффициенты интерполирующего полинома. С помощью такого приема добиваются численной стабильности и точности результатов интерполяции. Применяя затем найденный интерполирующий полином к реальным значениям аргумента, надеются получить более стабильные результаты.

Задача интерполяции по точкам единичной окружности. Выведем соотношения для интерполяции по точкам единичной окружности. Обозначим матрицу X из (7.22) следующим образом

$$X = \left[x_i^j \right],$$

где индекс i и показатель степени j пробегают значения от 0 до n. Точки, лежащие на единичной окружности, принимают значения

$$x_0 = 1; x_k = exp(\, j \cdot 2 \cdot k \cdot \pi /(n+1)), \qquad (7.23)$$

где $j = \sqrt{-1}$.

Введем обозначение

$$w = exp(\, j \cdot 2 \cdot \pi /(n+1)). \qquad (7.24)$$

Тогда

$$x_k = w^k, \qquad (7.25)$$

$$X = \left[w^{ij} \right]. \qquad (7.26)$$

Главное преимущество данного представления отсчетов x_i это простота и устойчивость вычислений. В частности покажем, что

$$X^{-1} = \left[w^{-ij} \right] / (n+1) = X^{+} / (n+1), \tag{7.27}$$

где X^{+}- транспонированная комплексно - сопряженная (эрмитово - сопряженная) матрица. При этом произведение

$$X \cdot X^{-1} = \left[w^{ij} \right] \cdot \left[w^{-ij} \right] / (n+1) = E$$

равно единичной матрице E.

Доказательство: Произвольный элемент произведения можно записать в виде

$$(1/(n+1)) \cdot \sum_{k=0}^{n} w^{ik} \cdot w^{-kj} .$$

Для диагонального элемента, при $i = j$, имеем

$$(1/(n+1)) \cdot \sum_{k=0}^{n} w^{ik} \cdot w^{-kj} = (1/(n+1)) \cdot \sum_{k=0}^{n} w^{0} = 1. \tag{7.28}$$

Для вне диагонального элемента, при $i \neq j$, получаем

$$(1/(n+1)) \cdot \sum_{k=0}^{n} w^{ik} \cdot w^{-kj} = (1/(n+1)) \cdot \sum_{k=0}^{n} (w^{i-j})^{k} .$$

Выражение в правой части представляет собой геометрическую прогрессию, сумма членов которой равна нулю

$$(1/(n+1)) \cdot \sum_{k=0}^{n} (w^{i-j})^{k} = (1/(n+1)) \cdot (1-(w^{i-j})^{n+1}) / (1-w^{i-j}) = 0 , \tag{7.29}$$

так как выражение в числителе может быть представлено в виде

$$(w^{i-j})^{n+1} = exp(j \cdot 2 \cdot \pi \cdot (n+1) \cdot (i-j) / (n+1)) = exp(j \cdot 2 \cdot \pi \cdot (i-j)) = 1.$$

Таким образом, соотношение (7.27) доказано и из него, в частности, следует ортогональность матрицы X, что, как известно, является гарантией стабильности вычислительного процесса.

Решение системы (7.22) для отсчетов, выбранных в соответствии с соотношениями (7.23), можно записать как

$$A = X^{-1} \cdot Y = (1/(n+1)) \cdot \left\lfloor w^{-ij} \right\rfloor \cdot Y , \tag{7.30}$$

или в координатной форме

$$a_j = \sum_{k=0}^{n} y_k \cdot w^{-jk} . \tag{7.31}$$

Исходный полином (7.20), определенный в точках x_k, представим в виде

$$y_k = \sum_{j=0}^{n} a_j \cdot w^{jk} . \tag{7.32}$$

Уравнения (7.31) и (7.32) являются взаимообратными и соответствуют дискретному преобразованию Фурье. При числе отсчетов, кратном степени $2^m = (n+1)$, где m - положительное целое число, мы пришли к известной модификации алгоритма быстрого преобразования Фурье.

7.4 Алгоритм формирования символьных функций

Остановимся немного подробнее на алгоритме вычисления символьных функций от частоты на основании предыдущего материала и приведем небольшой пример.

Алгоритм формирования выходной функции в символьном виде можно описать следующей последовательностью действий:

1. По числу реактивностей оценивается порядок функции цепи n_0, и выбираются точки s_i, равномерно распределенные на единичной окружности ($i = 0,1,\cdots,n_0$).

2. Полагаем $i = 0$.

3. Для текущего отсчета одним из методов формируем матрицу коэффициентов и вектор свободных членов. При этом компоненты, содержащие $1/s$, не допускаются. Источники входных воздействий тока либо напряжения полагаются единичными;

4. Решаем систему $T(s_i) \cdot X(s_i) = W$ методом LU - разложения.

5. Вычисляем текущее значение выходной функции $F(s_i) = d^t \cdot X(s_i)$.

6. Находим определитель матрицы $T(s_i)$ $D(s_i) = \prod_{k=1}^{n} l_{kk}$.

7. Определяем текущее значение числителя $N(s_i) = F(s_i) \cdot D(s_i)$.

8. Если $i < n_0$, то полагаем $i = i + 1$ и возвращаемся к пункту 2, иначе дальше к пункту 9.

9. Используя накопленные отсчеты аргумента s_i, числителя $N(s_i)$ и знаменателя $D(s_i)$, и применяя дискретное преобразование Фурье, определяем коэффициенты полиномов числителя и знаменателя.

Для иллюстрации алгоритма найдем передаточное сопротивление $Z_{trn} = V_2 / J$ простой цепи показанной на рисунке.7.2.

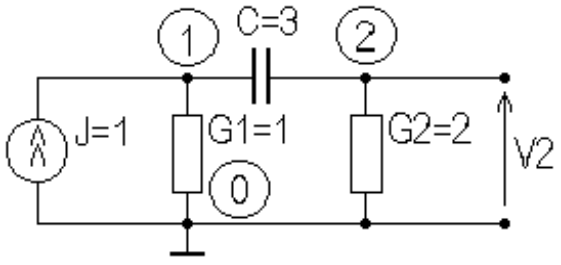


Рисунок 7.2 - Простая цепь

Для данной схемы, не содержащей индуктивностей, систему уравнений можно построить по методу узловых потенциалов

$$\begin{bmatrix} 1+3\cdot s & -3\cdot s \\ -3\cdot s & 2+3\cdot s \end{bmatrix}\cdot\begin{bmatrix} v_1 \\ v_2 \end{bmatrix}=\begin{bmatrix} 1 \\ 0 \end{bmatrix}.$$

Схема содержит один конденсатор, следовательно, порядок функции цепи должен быть $n_0=1$. Выбираем две равноотстоящие точки, лежащие на единичной окружности: $s_0=1,\ s_1=-1$. Подставим значение $s=1$ и проведем LU - разложение матрицы коэффициентов

$$\begin{bmatrix} 4 & 0 \\ -3 & 11/4 \end{bmatrix}\cdot\begin{bmatrix} 1 & -3/4 \\ 0 & 1 \end{bmatrix}\cdot\begin{bmatrix} v_1(s_0) \\ v_2(s_0) \end{bmatrix}=\begin{bmatrix} 1 \\ 0 \end{bmatrix}.$$

Используя прямую и обратную подстановки, получаем

$$\begin{bmatrix} v_1(s_0) \\ v_2(s_0) \end{bmatrix}=\begin{bmatrix} 5/11 \\ 3/11 \end{bmatrix}$$

откуда следует значение

$$F(s_0)=N(s_0)/D(s_0)=d^t\cdot V(s_0)=\begin{bmatrix} 0 & 1 \end{bmatrix}\cdot\begin{bmatrix} 5/11 \\ 3/11 \end{bmatrix}=3/11.$$

Определитель матрицы равен $D(s_0)=4\cdot(11/4)=11$, следовательно, значение
числителя в этой точке равно $N(s_0)=F(s_0)\cdot D(s_0)=11\cdot(3/11)=3$.

Выполнив аналогичные действия для $s_1=-1$, найдем следующее разложение на треугольные сомножители

$$\begin{bmatrix} -2 & 0 \\ 3 & 7/2 \end{bmatrix}\cdot\begin{bmatrix} 1 & -3/2 \\ 0 & 1 \end{bmatrix}\cdot\begin{bmatrix} v_1(s_1) \\ v_2(s_1) \end{bmatrix}=\begin{bmatrix} 1 \\ 0 \end{bmatrix}.$$

Решение системы дает

$$\begin{bmatrix} v_1(s_1) \\ v_2(s_1) \end{bmatrix}=\begin{bmatrix} 1/7 \\ 3/7 \end{bmatrix},$$

следовательно $F(s_1)=3/7$. Определитель системы $D(s_1)=-7$, откуда $N(s_1)=-3$.

По точкам $(1,3)$ и $(-1,-3)$ числителя и точкам $(1,11)$ и $(-1,-7)$ знаменателя, в соответствии с выражением (7.31), находим

$$a_0=0.5\cdot(N_0\cdot 1+N_1\cdot x_0)=0.5\cdot(3-3)=0,$$
$$a_1=0.5\cdot(N_0\cdot 1+N_1\cdot x_1)=0.5\cdot(3+3)=3,$$
$$b_0=0.5\cdot(D_0\cdot 1+D_1\cdot x_0)=0.5\cdot(11-7)=2,$$
$$b_1=0.5\cdot(D_0\cdot 1+D_1\cdot x_0)=0.5\cdot(11+7)=9.$$

Откуда, интерполирующие полиномы числителя и знаменателя, определяются как

$$N(s)=3\cdot s,\qquad D(s)=2+9\cdot s,$$

а передаточное сопротивление в символьном виде запишется

$$Z_{trn} = 3 \cdot s / (2 + 9 \cdot s),$$

в чем легко убедится прямым анализом схемы.

Остановимся на некоторых особенностях алгоритма формирования символьных функций:

1. Когда порядок цепи неизвестен, можно положить n_0, равным числу реактивных элементов. Вследствие высокой численной стабильности алгоритма интерполяции полиномов по точкам на единичной окружности можно пользоваться числом отсчетов больше или равным n_0. При завышении порядка коэффициенты при высших степенях полиномов будут равны нулю.

2. При простом нормировании параметров элементов схемы иногда может случиться, что полюс совпадает с одной из точек интерполяции на единичной окружности. В этом случае определитель матрицы T равен нулю и система уравнений в этой точке вырождается. Для устранения этого положения можно увеличить число точек интерполяции либо выбрать другую нормирующую частоту для данной схемы.

Следует также предупредить о некоторых особенностях символьного анализа:

1. В цепях высокой размерности коэффициенты полиномов могут отличаться на несколько порядков, и если не использовать нормировку, то может произойти потеря точности. Нормировать можно либо непосредственно матрицу системы уравнений, либо параметры элементов цепи.

2. Центральную частоту полосового фильтра и частоту среза фильтра нижних частот рекомендуется нормировать к величине $\omega = 1$ [рад./сек]. Импедансы цепи следует преобразовать таким образом, чтобы одно из сопротивлений стало равным 1 Ом.

При машинном формировании символьных функций необходимо принять во внимание следующее:

1. Можно использовать только точки, лежащие на и над вещественной осью, т.к. комплексно - сопряженные точки дают комплексно - сопряженные решения. Этот прием позволяет почти в два раза сократить объем вычислений.

2. Коэффициенты полиномов числителя и знаменателя по точкам $(s_i, N(s_i))$ и $(s_i, D(s_i))$ можно определять за один прием. Для этого необходимо сформировать комплексную функцию $c_i = N(s_i) + j \cdot D(s_i)$ и найти коэффициенты c_i. Вещественные части коэффициентов c_i будут соответствовать коэффициентам числителя, а мнимые части - коэффициентам знаменателя.

3. При вычислении определителя произведение диагональных элементов матрицы L следует искать в виде суммы логарифмов l_{kk} (в комплексной форме), а затем вычислять антилогарифм. Это позволит

предотвратить переполнение разрядной сетки ЭВМ при существенном различении порядков величин l_{kk} .

8 РАСЧЕТ ЧУВСТВИТЕЛЬНОСТИ ЭЛЕКТРОННЫХ СХЕМ

8.1 Определения чувствительности

Наиболее важной с точки зрения производства и эксплуатации радиоэлектронной аппаратуры характеристикой электронных схем является чувствительность их выходных функций к изменению параметров, как самих схем, так и параметров окружающей среды. Изменение параметров схемы может являться следствием технологического разброса параметров при производстве радиоэлектронных устройств (РЭУ) и элементной базы. При эксплуатации РЭУ также наблюдается изменение параметров элементов, как за счет старения, так и при изменении внешних факторов (температуры, радиации, влажности и т.д.).

Чувствительность определяется как производная дифференцируемой выходной функции F по параметру h

$$D_h^F = \partial F / \partial h. \tag{8.1}$$

Это определение удобно для вычисления на ЭВМ, однако, чаще используют безразмерное определение чувствительности - нормированную чувствительность.

Относительную или нормированную чувствительность определяют как

$$S_h^F = \partial ln(F) / \partial ln(h) = h \cdot \partial F / (F \cdot \partial h) = h \cdot D_h^F / F. \tag{8.2}$$

Иногда встречаются ситуации, когда номинальные значение функции F или параметра h равны нулю. В этом случае определение нормированной чувствительности неприемлемо и используют полуотносительные или полунормированные чувствительности

$$\tilde{S}_h^F = \partial F / \partial ln(h) = h \cdot \partial F / \partial h = h \cdot D_h^F, \tag{8.3}$$

$$\hat{S}_h^F = \partial ln(F) / \partial h = \partial F / (F \cdot \partial h) = D_h^F / F, \tag{8.4}$$

откуда следует, что

$$S_h^F = \tilde{S}_h^F / F, \quad S_h^F = \hat{S}_h^F \cdot h. \tag{8.5}$$

Чувствительность функций цепи. Часто выходная функция, например, в случае передаточной характеристики, определяется в виде отношения полиномов

$$F = T = N / D.$$

Логарифм этой функции продифференцируем по $\partial ln(h)$ и в результате получим

$$S_h^T = S_h^N - S_h^D. \tag{8.6}$$

Комплексная выходная функция может быть представлена в алгебраической форме

$$T = a + j \cdot b. \tag{8.7}$$

Нормированная чувствительность выходной функции в этом случае запишется

$$\partial ln(T) / \partial ln(h) = \partial ln(a + j \cdot b) / \partial ln(h)$$

или

$$(h / T) \cdot (\partial T / \partial h = (h / (a + j \cdot b)) \cdot (\partial a / \partial h + j \cdot \partial b / \partial h)$$

откуда

$$(Re(\partial T / \partial h) + j \cdot Im(\partial T / \partial h)) = (a / h) \cdot S_h^a + j \cdot (b / h) \cdot S_h^b,$$

$$S_h^a = (h / a) \cdot Re(\partial T / \partial h), \tag{8.8}$$

$$S_h^b = (h / b) \cdot Im(\partial T / \partial h), \tag{8.9}$$

где Re, Im - реальная и мнимая части комплексной функции.

Другую удобную форму выражения чувствительности можно получить, если записать

$$T = |T| \cdot exp(j \cdot \varphi), \tag{8.10}$$

где $|T|$ - модуль выходной функции, а φ - фаза (аргумент). Логарифмируя это выражение

$$ln(T) = ln(|T|) + j \cdot \varphi \tag{8.11}$$

и дифференцируя по h

$$\partial ln(T) / \partial h = S_h^F = \partial ln(|T|) / \partial h + j \cdot \partial \varphi / \partial h$$

с учетом (8.4) получаем

$$\hat{S}_h^{|T|} = Re(\hat{S}_h^T), \tag{8.12}$$

$$\hat{S}_h^{\varphi} = Im(\hat{S}_h^T) / \varphi, \tag{8.13}$$

т.к.

$$\hat{S}_h^{\varphi} = \partial ln(\varphi) / \partial h = (1 / \varphi) \cdot \partial \varphi / \partial h$$

либо

$$D_h^{|T|} = |T| \cdot Re(\hat{S}_h^T), \tag{8.14}$$

$$D_h^{\varphi} = Im(\hat{S}_h^T). \tag{8.15}$$

Еще одно представление комплексной выходной функции в алгебраической форме получается, когда в (8.11) $ln(|T|) = \alpha$ выражается в децибелах

$$ln(T) = \alpha + j \cdot \varphi. \tag{8.16}$$

Дифференцируя это выражение по параметру h, получаем

$$\partial \ln(T)/\partial h = \hat{S}_h^T = \partial \alpha/\partial h + j\cdot \partial \varphi/\partial h,$$

$$(\partial \alpha/\partial h)_{[непер]} = Re(\hat{S}_h^T), \qquad (8.17)$$

$$(\partial \varphi/\partial h) = Im(\hat{S}_h^T), \qquad (8.18)$$

где $\partial \alpha/\partial h$ - выражена в неперах, а $\partial \varphi/\partial h$ - выражена в градусах либо радианах. Для выражения чувствительности модуля в децибелах необходимо $\partial \alpha/\partial h$ умножить на $20\cdot ln(e)/ln(10) = 8.686$

$$(\partial \alpha/\partial h)_{[дБ]} = 8.686\cdot(\partial \alpha/\partial h) = 8.686\cdot(\partial |T|/\partial h). \qquad (8.19)$$

Очевидно, что чувствительность цепи является в общем случае функцией частоты.

Чувствительность нулей и полюсов. Одним из недостатков чувствительности функции цепи является ее зависимость от частоты. В результате для оценки поведения цепи при отклонении параметров необходимо рассчитать чувствительность в ряде частотных точек диапазона частот. Выбор этих точек не всегда очевиден.

С другой стороны, полюса и нули выходной функции представляют собой конечный ряд комплексных чисел, которые полностью определяют отклик цепи.

При расчете чувствительности нуля полинома необходимо иметь в виду, что положение нуля зависит от параметра. Следовательно, для любого нуля z полинома P (числителя либо знаменателя), можно записать

$$P(h, s(h))_{/s=z} = 0.$$

Дифференцируя это выражение по h, получаем

$$\partial P/\partial h + (\partial P/\partial h)/(\partial s/\partial h)_{/s=z} = 0,$$

или

$$\partial s/\partial h = \partial z/\partial h = -(\partial P/\partial h)/(\partial P/\partial s)_{/s=z}. \qquad (8.20)$$

Это выражение пригодно для вычисления чувствительности простых нулей. Нормированная чувствительность нуля запишется

$$S_h^z = (h/z)\cdot(\partial z/\partial h). \qquad (8.21)$$

Учитывая, что нуль полинома в общем случае является комплексным числом $z = a + j\cdot b$, чувствительность действительной и мнимой частей нуля, в соответствии с (8.8) и (8.9), можно определить следующим образом

$$S_h^a = (h/a)\cdot Re(\partial z/\partial h), \qquad (8.22)$$

$$S_h^b = (h/b)\cdot Im(\partial z/\partial h). \qquad (8.23)$$

Многопараметрическая чувствительность. Обычная чувствительность определяет изменение функции цепи при вариации одного из параметров. Однако в общем случае функция F зависит от нескольких параметров

$$F = F(h_1, h_2, \cdots, h_n) = F(h),$$

где h - вектор параметров.

Пусть необходимо оценить изменение функции F, когда некоторые или все параметры варьируются одновременно. Приращение функции F при бесконечно малых изменениях всех параметров определяется полной производной

$$dF = \sum_{i=1}^{m} (\partial F / \partial h) \cdot dh_i \ . \qquad (8.24)$$

Для перехода к нормированной чувствительности разделим обе части выражения (8.24) на F, а каждый элемент суммы умножим и поделим на h_i

$$dF / F = \sum_{i=1}^{m} ((\partial F / \partial h) / (h_i / F)) \cdot (dh_i / h_i) = \sum_{i=1}^{m} S_{h_i}^{F} \cdot (dh_i / h_i) . \qquad (8.25)$$

Очень часто, более удобным оказывается использование приращений

$$\Delta F / F \cong \sum_{i=1}^{m} S_{h_i}^{F} \cdot \Delta h_i / h_i \ , \qquad (8.26)$$

где $\Delta h_i / h_i$ - относительное изменение параметров, часто определяемое технологией изготовления элементов. Относительные изменения обычно таковы, что

$$|\Delta h_i / h_i| \le t_i \ ,$$

где t_i - допуск на i - ый элемент. Запишем отношение $\Delta F / F$ в наихудшем случае

$$|\Delta F / F| \le \sum_{i=1}^{m} \left| S_{h_i}^{F} \right| \cdot t_i \ . \qquad (8.27)$$

8.2 Алгоритмы расчета чувствительности

Расчет чувствительности, основанный на теореме о производной определителя по элементу. Как известно, в традиционных методах расчета электронных схем, таких как обобщенный узловой и контурный, все малосигнальные характеристики могут быть представлены выражениями, состоящими в основном из отношений алгебраических дополнений. С другой стороны известно, что определитель матрицы можно представить разложением по любой строке (столбцу) в виде

$$\Delta = \sum a_{ij} \cdot \Delta_{ij} \ . \qquad (8.28)$$

Дифференцируя это выражение по элементу определителя a_{ij}

$$\partial \Delta / \partial a_{ij} = \Delta_{ij} \qquad (8.29)$$

получаем, что производная определителя по элементу равна алгебраическому дополнению этого элемента. Используя этот факт, можно предложить подход к определению чувствительности малосигнальных характеристик электронных схем к их параметрам.

Удобнее всего, данный подход проиллюстрировать на примере параметров рассеяния, выраженных через алгебраические дополнения нормированной к проводимостям нагрузок матрицы проводимости

$$S_{ij} = 2 \cdot \Delta_{ji} / \Delta - \delta_{ij}, \qquad (8.30)$$

где S_{ij} - элемент матрицы рассеяния; δ_{ij} - символ Кронекера.

Определим функции чувствительности параметров рассеяния к проводимости пассивного двухполюсника y_0, включенному в схему, между произвольными узлами k и l, учитывая, что его проводимость войдет в элементы матрицы проводимости лишь на пересечении указанных строк и столбцов

$$D_{y_0}^{S_{ij}} = \partial S_{ij} / \partial y_0 = 2 \cdot (\Delta_{j(k+l),i(k+l)} \cdot \Delta - \Delta_{(k+l)(k+l)} \cdot \Delta_{ij}) \cdot \Delta^2 = \\ = -2 \cdot \Delta_{j(k+l)} \cdot \Delta_{(k+l)i} / \Delta^2 \qquad (8.31)$$

Для электронных схем, содержащих активные элементы, определим чувствительность параметров рассеяния к проводимости управляющей ветви y_c и коэффициенту передачи тока источника тока α, управляемого током ветви y_c и включенных между узлами k,l и p,q

$$D_{y_c}^{S_{ij}} = \partial S_{ij} / \partial y_c = 2 \cdot [(\Delta_{j(k+l),i(k+l)} + \alpha \cdot \Delta_{j(k+l),i(p+q)}) \cdot \Delta - \\ - (\Delta_{(k+l)(k+l)} + \alpha \cdot \Delta_{(k+l)(p+q)}) \cdot \Delta_{ji}] / \Delta^2 = \\ = 2 \cdot \Delta_{j(k+l)} \cdot (\Delta_{(k+l)i} + \alpha \cdot \Delta_{(p+q)i}) / \Delta^2, \qquad (8.32)$$

$$D_{\alpha}^{S_{ij}} = \partial S_{ij} / \partial \alpha = 2 \cdot (\Delta_{j(k+l),i(p+q)} \cdot \Delta - \Delta_{(k+l)(p+q)} \cdot \Delta_{ji}) \cdot \Delta^2 = \\ = -2 \cdot \Delta_{j(k+l)} \cdot \Delta_{(p+q)i} / \Delta^2 \qquad (8.33)$$

Отметим, что чувствительность к коэффициенту передачи по току α совпадает с чувствительностью по крутизне g_m источника тока управляемого напряжением на ветви y_c, в силу их одинаковой локализации в матрице проводимости.

При известных чувствительностях Y- параметров подсхемы y_{kl} к ее элементу y_0, чувствительность параметров рассеяния схемы к этому элементу подсхемы, в соответствии с определением сложной производной можно записать

$$D_{y_0}^{S_{ij}} = (\partial S_{ij} / \partial y_{kl}) \cdot (\partial y_{kl} / \partial y_0) = D_{y_{kl}}^{S_{ij}} \cdot D_{y_0}^{y_{kl}}. \qquad (8.34)$$

Последнее соотношение указывает на возможность применения метода подсхем при расчете чувствительности сложных электронных схем обобщенным методом узловых потенциалов.

Предлагаемый способ, столь же эффективно, может быть использован при определении чувствительности более высоких порядков и

чувствительности других характеристик электронных схем, однако требует развитого вычислительного аппарата детерминантной алгебры.

Отметим также, что вычисление кратных алгебраических дополнений на основании теорем детерминантной алгебры всегда можно свести к вычислению обычных одинарных алгебраических дополнений. В этом случае реализация данного алгоритма расчета чувствительности сводится к вычислению присоединенной матрицы (например, путем обращения матрицы, ее транспонирования и умножения на определитель) и перебору соответствующих алгебраических дополнений, что удачно сочетается с вычислением других малосигнальных характеристик электронных схем в составе универсальных программ.

Вычисление чувствительности на основе дифференцирования матрицы по ее элементам. Данный подход также ориентирован в основном на традиционные методы - обобщенных узловых потенциалов и контурных токов.

Как известно, любой двухполюсный элемент, включенный между узлами i и j, войдет в элементы матрицы проводимости на пересечении строк и столбцов с этими индексами. В связи с этим, производная матрицы проводимости Y по элементу y_0, включенному между узлами i и j равна

$$\partial Y / \partial y_0 = Y_{ij}, \tag{8.35}$$

где Y_{ij} - матрица с элементами $y_{ii} = y_{jj} = 1$ и $y_{ij} = y_{ji} = -1$, остальные элементы равны нулю.

Пусть требуется определить чувствительность передаточного импеданса Z_{pq}, являющегося элементом матрицы $Z = Y^{-1}$. Тогда, для определения $\partial Z / \partial y_0$, рассмотрим тождество $Z \cdot Y = 1$. Дифференцируя это тождество по y_0, включенному между узлами i и j, получим

$$(\partial Z / \partial y_0) \cdot Y + Z \cdot Y_{ij} = 0.$$

Откуда, учитывая, что $Z = Y^{-1}$, можем записать

$$\partial Z / \partial y_0 = -Z \cdot Y_{ij} \cdot Z. \tag{8.36}$$

Для конкретного элемента матрицы Z, раскрывая выражение (8.36), можно записать

$$\partial Z_{pq} / \partial y_0 = (z_{pi} - z_{pj}) \cdot (z_{iq} - z_{jq}). \tag{8.37}$$

Если речь идет об элементах матрицы проводимости α и g_m (коэффициент передачи по току и крутизна управляемого источника), причем источник включен между узлами k и l, а управляющие узлы i и j, то производная равна

$$\partial Y / \partial x_{kl}^{ij} = Y_{kl}^{ij}, \tag{8.38}$$

где Y_{kl}^{ij} - матрица, с отличными от нуля элементами $y_{ik} = y_{jl} = 1$ и $y_{il} = y_{kl} = -1$. Раскрывая соотношение (8.36), можно записать

$$\partial Z_{pq} / \alpha = (z_{pi} - z_{pj}) \cdot (z_{kq} - z_{lq}). \qquad (8.39)$$

Чувствительность решений линейных алгебраических систем уравнений. Чувствительность вектора решений линейной системы уравнений, как математической модели электронной схемы к изменению ее параметров, представляет несомненный интерес для разработчиков РЭУ, так как в конечном итоге определяет чувствительность всех характеристик.

Пусть имеем систему линейных уравнений

$$T \cdot X = W, \qquad (8.40)$$

где T- матрица коэффициентов и W- вектор свободных членов, могут быть функциями вектора параметров h с компонентами h_i. Формальное решение системы, как известно, имеет вид

$$X = T^{-1} \cdot W. \qquad (8.41)$$

Для оценки чувствительности вектора X к некоторому параметру h, продифференцируем выражение (8.40)

$$T \cdot (\partial X / \partial h) + (\partial T / \partial h) \cdot X = \partial W / \partial h.$$

Результат дифференцирования запишем в виде

$$T \cdot (\partial X / \partial h) = -((\partial T / \partial h) \cdot X - \partial W / \partial h). \qquad (8.42)$$

Анализ данного выражения показывает, что вектор X может быть определен из решения исходной системы, например, методом LU- факторизации. Производные матрицы $\partial T / \partial h$ и вектора $\partial W / \partial h$ по параметру h, как известно, определяются покомпонентным дифференцированием. В результате, вектор правой части системы (8.41) определяется достаточно просто. Далее необходимо на основе того же LU- разложения, т.к. матрица коэффициентов T остается прежней, найти решение с новой правой частью. В результате прямой и обратной подстановок найдем вектор $\partial X / \partial h$, определяющий чувствительность вектора X к изменению конкретного параметра h.

Если требуется определить чувствительности вектора X по отношению к нескольким параметрам h_i, то уравнение (8.42) необходимо составить и решить для каждого h_i.

Заметим, что методы LU- факторизации, как и QR- факторизации дают в этом случае существенное сокращение вычислительных операций за счет экономии на повторных разложениях той же матрицы коэффициентов.

Метод присоединенных систем уравнений. На практике часто требуется оценить чувствительность лишь отдельных компонент вектора X, определяющих выходную функцию $\boldsymbol{\Phi}$. При этом требуется обычно определить чувствительность $\partial \boldsymbol{\Phi} / \partial h_i$ выходной функции к изменению параметра h. При этом будем рассматривать простейшую выходную функцию $\boldsymbol{\Phi}(X)$, являющуюся линейной комбинацией компонент вектора X, определяемую выражением

$$\boldsymbol{\Phi} = d^t \cdot X, \qquad (8.43)$$

где d - вектор, выделяющий нужную комбинацию компонент и состоящий, в общем случае из $1,\ -1,\ 0$.

Формальное решение системы (8.42), определяющее чувствительность вектора X, запишется

$$(\partial X / \partial h) = -T^{-1} \cdot ((\partial T / \partial h) \cdot X - \partial W / \partial h). \qquad (8.44)$$

Для определения чувствительности

$$\partial \Phi / \partial h = d^t \cdot (\partial X / \partial h), \qquad (8.45)$$

воспользовавшись соотношением (8.44), получим

$$\partial \Phi / \partial h = -d^t \cdot T^{-1} \cdot ((\partial T / \partial h) \cdot X - \partial W / \partial h). \qquad (8.46)$$

Проанализируем полученное соотношение. Прежде всего, обратим внимание на вектор - строку $d^t \cdot T^{-1}$, которую можно интерпретировать, как решение транспонированной системы уравнений

$$Y^t = -d^t \cdot T^{-1}, \qquad (8.47)$$

т.к. иначе это соотношение можно записать в виде

$$T^t \cdot Y = -d\ . \qquad (8.48)$$

Это означает, что вектор $Y^t = -d^t \cdot T^{-1}$ можно определить вслед за вектором X из решения исходной, но транспонированной системы методом LU-факторизации. Подставляя в (8.46) Y^t, вместо $-d^t \cdot T^{-1}$, получим

$$\partial \Phi / \partial h = Y^t \cdot (\partial T / \partial h) \cdot X - Y^y \cdot (\partial W / \partial h). \qquad (8.49)$$

Анализируя записанное соотношение, видим, что для получения чувствительности выходной функции необходимо решить исходную и транспонированную системы уравнений, вычислить производные от матрицы и вектора свободных членов и результаты перемножить в соответствии с выражением (8.49). Для каждого параметра h_i заново формируются матрица $\partial T / \partial h$ и вектор $\partial W / \partial h$, а затем определяется правая часть. Векторы X и Y определяются один раз из исходной и транспонированной системы и не зависят от параметра h_i.

Таким образом, вычислительную процедуру метода присоединенной системы уравнений, можно представить в виде:

1) решаем исходную систему уравнений $T \cdot X = W$;

2) решаем транспонированную систему $T^t \cdot Y = -d$;

3) для каждого h_i формируем матрицу $\partial T / \partial h$ и вектор $\partial W / \partial h$ производных и, в соответствии с (8.49), вычисляем $\partial \Phi / \partial h_i$.

При решении исходной и присоединенной систем необходимо использовать результаты одного LU - разложения. Пункт 3 можно упростить, воспользовавшись специальным расположением нулевых элементов в матрице $\partial T / \partial h_i$ и векторе $\partial W / \partial h_i$.

Рассмотрим два примера на вычисление чувствительностей узловых потенциалов через чувствительность решений линейной системы уравнений и чувствительности потенциала второго узла методом присоединенной

системы уравнений для схемы с идеальным операционным усилителем изображенной на рисунке 8.1 к параметру G_1.

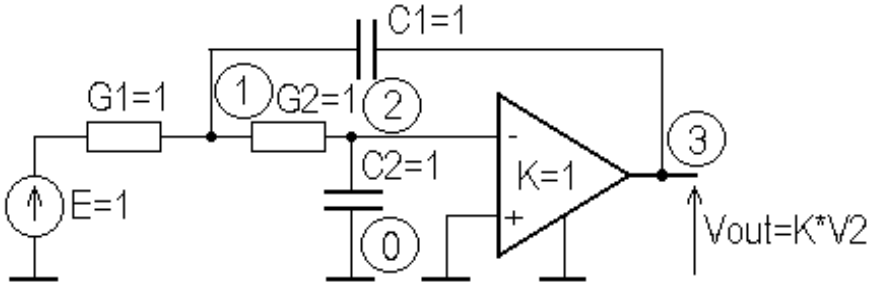


Рисунок 8.1 Схема с идеальным операционным усилителем

Составим систему узловых уравнений схемы. Для этого, вместо идеального операционного усилителя, возьмем вначале реальный операционный усилитель, моделируемый источником напряжения управляемым напряжением, с конечным выходным сопротивлением R. Далее, заменим его источником тока управляемым напряжением с выходной проводимостью $G = 1/R$. После составления узловой системы исключим, в соответствии с соотношениями Гаусса, ток и напряжение выходного узла и осуществим предельный переход элементов матрицы проводимости, при $G \to \infty$.

Узловая матрица проводимости схемы, учитывая, что крутизна источника тока управляемого напряжением равна $S = K \cdot G$, запишется

$$Y = \begin{bmatrix} G_1 + G_2 + s \cdot C_1 & -G_2 & -s \cdot C_1 \\ -G_2 & G_2 + s \cdot C_2 & 0 \\ -s \cdot C_1 & -K \cdot G & G + s \cdot C_1 \end{bmatrix},$$

где $s = j \cdot \omega$ - оператор Лапласа. Исключая напряжение и ток третьего узла, получим матрицу второго порядка

$$Y = \begin{bmatrix} G_1 + G_2 + s \cdot C_1 / (G + s \cdot C_1) & -G_2 - K \cdot G \cdot s \cdot C_1 / (G + s \cdot C_1) \\ -G_2 & G_2 + s \cdot C_2 \end{bmatrix}.$$

Осуществив предельный переход, при $G \to \infty$, и, заменив, источник напряжения на входе источником тока, получим следующую узловую систему уравнений для схемы с идеальным операционным усилителем

$$\begin{bmatrix} G_1 + G_2 + s \cdot C_1 & -G_2 - s \cdot C_1 \cdot K \\ -G_2 & G_2 + s \cdot C_2 \end{bmatrix} \cdot \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} G_1 \cdot E \\ 0 \end{bmatrix}.$$

Положим для определенности $s = 2$, и, подставив численные значения в систему, получим решение

$$X = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} 1/3 \\ 1/9 \end{bmatrix}.$$

Используя соотношение (8.42), вычислим вектор правой части

$$((\partial T / \partial h) \cdot X - \partial W / \partial h) = \begin{bmatrix} 2/3 \\ 0 \end{bmatrix},$$

откуда искомое решение имеет вид

$$\begin{bmatrix} \partial v_1 / \partial G_1 \\ \partial v_2 / \partial G_1 \end{bmatrix} = \begin{bmatrix} 2/9 \\ 2/27 \end{bmatrix}.$$

Для нахождения чувствительности $\partial v_2 / \partial G_1$, воспользуемся соотношением (8.49). Как и прежде, решение исходной системы равно

$$X = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} 1/3 \\ 1/9 \end{bmatrix}.$$

Т.к. выходной величиной является v_2, то $d = \begin{bmatrix} 0 & 1 \end{bmatrix}^t$ и сопряженная система $T^t \cdot Y = -d$ принимает вид

$$\begin{bmatrix} 4 & -1 \\ -3 & 3 \end{bmatrix} \cdot \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} 0 \\ -1 \end{bmatrix}.$$

Решение сопряженной системы запишется

$$Y = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} -1/9 \\ -4/9 \end{bmatrix}.$$

Используя соотношение (8.49), найдем чувствительность напряжения v_2 к изменению параметра G_1

$$\partial v_2 / \partial G_1 = \begin{bmatrix} -1/9 & -4/9 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} 1/3 \\ 1/9 \end{bmatrix} - \begin{bmatrix} 1/9 & -4/9 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 0 \end{bmatrix} = 2/27 .$$

Как видим, результаты расчета чувствительности $\partial v_2 / \partial G_1$ по соотношениям (8.42) и (8.49) совпадают.

Сделаем одно замечание общего характера, касающееся использования понятия сопряженной системы уравнений. Дело в том, что в целом ряде случаев при решении линейных систем уравнений с разными правыми частями W интересует не весь вектор решений X, а отдельные компоненты либо линейная суперпозиция компонент. В этом случае, наряду с исходной системой уравнений

$$T \cdot X_i = W_i, \tag{8.50}$$

целесообразно ввести выходную функцию вида

$$\Phi_1 = d^t \cdot X_i, \tag{8.51}$$

где $i = 1, \cdots, m$; m- число различных правых частей системы уравнений. Прямое решение уравнений (8.50) требует LU- разложения исходной матрицы коэффициентов T и последующей m- кратной прямой и обратной подстановок. Используя понятие присоединенной системы, формальное решение системы (8.50) запишется

$$\Phi_i = d^t \cdot T^{-1} \cdot W_i = (d^t \cdot T^{-1}) \cdot W_i = Y^t \cdot W_i, \tag{8.52}$$

где Y^t - решение присоединенной системы уравнений

$$T^t \cdot d = Y. \tag{8.53}$$

Чувствительность произвольной выходной функции. В отличие от рассмотренной линейной комбинации вектора решений выходная функция может иметь в общем случае произвольный вид

$$\boldsymbol{\Psi} = \boldsymbol{\Phi}(X,h), \tag{8.54}$$

где $\boldsymbol{\Phi}$ - дифференцируемая функция; h - вектор параметров.

Дифференцирование (8.54), как сложной функции, по компоненте вектора h дает

$$\begin{aligned} \partial\boldsymbol{\Psi}/\partial h = \partial\boldsymbol{\Phi}/\partial h + \sum(\partial\boldsymbol{\Phi}/\partial X_i)\cdot(\partial X_i/\partial h) = \\ = \partial\boldsymbol{\Phi}/\partial h + (\partial\boldsymbol{\Phi}/\partial X)^t \cdot (\partial X/\partial h) \end{aligned}, \tag{8.55}$$

где $(\partial\boldsymbol{\Phi}/\partial X)^t$ - вектор строка производных. Используя соотношение (8.44), для $(\partial X/\partial h)$ получаем

$$\partial\boldsymbol{\Psi}/\partial h = \partial\boldsymbol{\Phi}/\partial h - (\partial\boldsymbol{\Phi}/\partial X)^t \cdot T^{-1} \cdot ((\partial T/\partial h)\cdot X - \partial W/\partial h). \tag{8.56}$$

Сравнивая полученное соотношение с (8.46), видим, что здесь, вместо d^t, используется $(\partial\Phi/\partial h)^t$, и, присоединенная система уравнений, в данном случае, имеет вид

$$T^t \cdot Y = (\partial\Phi/\partial h). \tag{8.57}$$

Таким образом, для решения присоединенной системы, вначале необходимо решить исходную систему для определения X. В результате, подставив в (8.56) решение присоединенной системы (8.57), получим

$$\partial\boldsymbol{\Psi}/\partial h = \partial\boldsymbol{\Phi}/\partial h - Y^t \cdot (\partial T/\partial h)\cdot X - Y^y \cdot (\partial W/\partial h). \tag{8.58}$$

Чувствительность более высокого порядка. Определение чувствительности высокого порядка требует вычисления производных более высокого порядка. Формально вычисление производной более высокого порядка не сложнее вычисления первой производной, однако требует больших затрат машинного времени. Для выяснения понятия чувствительности более высоких порядков получим соотношения для чувствительности второго порядка.

Вначале продифференцируем исходную систему уравнений (8.40) по параметру h_1

$$T\cdot(\partial X/\partial h_1) + (\partial T/\partial h_1)\cdot X = \partial W/\partial h_1. \tag{8.59}$$

Для простоты пусть источники не зависят от параметров, т.е. $\partial W/\partial h_i = 0$. Продифференцируем соотношение (8.59) по параметру h_2

$$\begin{aligned} (\partial^2 T/(\partial h_1\cdot\partial h_2))\cdot X + (\partial T/\partial h_1)\cdot(\partial X/\partial h_2) + \\ + (\partial T/\partial h_2)\cdot(\partial X/\partial h_1) + T\cdot(\partial^2 X/(\partial h_1\cdot\partial h_2)) = 0 \end{aligned}. \tag{8.60}$$

Откуда следует

$$\begin{aligned} \partial^2 X/(\partial h_1\cdot\partial h_2) = -T^{-1}\cdot[(\partial^2 T/(\partial h_1\cdot\partial h_2))\cdot X + \\ + (\partial T/\partial h_1)\cdot(\partial X/\partial h_2) + (\partial T/\partial h_2)\cdot(\partial X/\partial h_1)] \end{aligned}. \tag{8.61}$$

Определим выходную функцию как

$$\boldsymbol{\Phi} = d^t \cdot X . \tag{8.62}$$

Тогда на основании предыдущего соотношения можно записать

$$\begin{aligned} &\partial^2 \boldsymbol{\Phi} / (\partial h_1 \cdot \partial h_2) = d^t \cdot \partial^2 X / (\partial h_1 \cdot \partial h_2) = \\ &= -d^t \cdot T^{-1} \cdot [(\partial^2 T / (\partial h_1 \cdot \partial h_2)) \cdot X + \\ &+ (\partial T / \partial h_1) \cdot (\partial X / \partial h_2) + (\partial T / \partial h_2) \cdot (\partial X / \partial h_1)] \end{aligned} . \tag{8.63}$$

Используя присоединенную систему уравнений

$$T^t \cdot Y = -d , \tag{8.64}$$

окончательно получим

$$\begin{aligned} &\partial^2 \boldsymbol{\Phi} / (\partial h_1 \cdot \partial h_2) = Y^t \cdot [(\partial^2 T / (\partial h_1 \cdot \partial h_2)) \cdot X + \\ &+ (\partial T / \partial h_1) \cdot (\partial X / \partial h_2) + (\partial T / \partial h_2) \cdot (\partial X / \partial h_1)] \end{aligned} . \tag{8.65}$$

Как видим, для определения чувствительности необходимо знать векторы $\partial X / \partial h_1$ и $\partial X / \partial h_2$. Допустим также, что выполнено LU-разложение $T = L \cdot U$. Требуемые производные вектора X можно получить двумя путями. Если число параметров, по отношению, к которым необходимо определять вторую производную велико (больше размерности матрицы T), то производные получают повторным применением метода, рассматривая каждый компонент вектора в качестве выходной функции и используя соотношение (8.49). Если число параметров не велико, то можно непосредственно определять $\partial X / \partial h_i$ из (8.42), учитывая, что $\partial W / \partial h_i = 0$, будем иметь

$$T \cdot (\partial X / \partial h_i) = -(\partial T / \partial h_i) \cdot X . \tag{8.66}$$

Число прямых и обратных подстановок в этом случае равно числу параметров.

Если рассчитывается чувствительность к параметрам элементов, то $\partial^2 T / (\partial h_i \cdot \partial h_j) = 0$ т.к. элементы матрицы T линейно зависят от параметров, поэтому первые производные будут равны константам при элементах, а вторые производные соответственно равны нулю. Если параметры входят в матрицу коэффициентов функционально, то вторая производная в общем случае не будет равной нулю.

Как видим, определение чувствительности второго порядка требует существенно больших вычислительных затрат, сводящихся к многократному решению исходной системы с разными правыми частями и решению транспонированной системы. Вычисление чувствительностей еще большего порядка потребует существенно больших вычислительных затрат.

Чувствительность второго порядка учитывает влияние на выходную функцию одновременного изменения двух параметров. В реальных условиях число параметров, подверженных изменению в силу различных факторов достаточно велико, что требует вычисления чувствительностей высоких порядков. Разумный объем вычислений заставляет с помощью

чувствительностей первого порядка выделять наиболее существенные параметры, дающие основной вклад в изменение характеристик.

8.3 Применение метода присоединенных систем

Рассмотрим наиболее полезные приложения метода присоединенных систем уравнений, позволяющего рассчитывать, как собственно чувствительности различных характеристик, так и ряд других характеристик на их основе.

Чувствительность по частоте. При расчетах в частотной области, вектор свободных членов W, обычно не зависит от частоты, а производная $\partial T / \partial \omega$, при $T = G + j \cdot \omega \cdot C$, становится равной $j \cdot C$. При этом соотношение (8.49) перепишется в виде

$$\partial \Phi / \partial \omega = j \cdot Y^t \cdot C \cdot X. \qquad (8.67)$$

При машинном формировании уравнений матрица C может существовать неявно. В этом случае чувствительность по частоте можно определить с помощью чувствительностей реактивных элементов

$$\partial \Phi / \partial \omega = (1/\omega) \cdot [\sum C_i \cdot (\partial \Phi / \partial C_i) + \sum L_i \cdot (\partial \Phi / \partial L_i)]. \qquad (8.68)$$

Так как Y и X, в общем случае, комплексные векторы, то выражение (8.67) будет иметь действительную и мнимую части. В соответствии с выражениями (8.14), (8.15), (8.17), (8.18), (8.19), можем записать

$$\partial |\Phi| / \partial \omega = |\Phi| \cdot Re((1/\Phi) \cdot (\partial \Phi / \partial \omega)), \qquad (8.68)$$

$$\partial |\varphi| / \partial \omega = Im((1/\Phi) \cdot (\partial \Phi / \partial \omega)), \qquad (8.69)$$

$$\partial \alpha / \partial \omega_{/\partial Б} \cong 8.686 \cdot Re((1/\Phi) \cdot (\partial \Phi / \partial \omega)). \qquad (8.70)$$

Соотношение (8.69), с точностью до знака определяет групповую задержку τ, так как

$$\tau = -\partial \varphi / \partial \omega. \qquad (8.71)$$

Таким образом, τ может быть определена, как результат расчета чувствительности методом присоединенной системы уравнений. Чувствительность амплитуды $\partial \alpha / \partial \omega$ полезна при поиске максимума и минимума АЧХ.

Чувствительность нулей. Пусть z_i является нулем некоторой неявной выходной функции $\Phi(s,h)_{/s=z_i} = 0$, определяющей ее вариации в окрестности нуля при изменении параметра h. Следовательно, h можно рассматривать как независимую переменную, а z_i - как зависимую переменную . Дифференцирование сложной функции, как и в случае нуля полинома (8.20) дает

$$(\partial \Phi / \partial s) \cdot (\partial s / \partial h)_{/s=z_i} + (\partial \Phi / \partial h) = 0,$$

откуда с учетом того, что $s = j \cdot \omega$ и соотношения (8.67), получаем

$$\partial s / \partial h_{/s=z_i} = \partial z_i / \partial h = \\ = -(\partial \Phi / \partial h) \cdot (\partial \Phi / \partial s)_{/s=z_i} = -(\partial \Phi / \partial h) / (Y^t \cdot C \cdot X)_{/s=z_i} \quad . \quad (8.72)$$

Таким образом, это соотношение позволяет определить чувствительность нуля выходной функции, не выражая его алгебраической функцией параметров.

Чувствительность полюсов. Определение чувствительности полюса эквивалентно определению чувствительности нуля знаменателя дробно - рационального представления выходной функции цепи к ее параметру. Обозначим i - тый полюс через p_i. Если $s = p_i$, то матрица T становится вырожденной и векторы X и Y вычислить нельзя. Это требует иного подхода к определению чувствительности полюса.

Рассмотрим LU- разложение матрицы коэффициентов $T = L \cdot U$ и продифференцируем его по параметру h

$$\partial T / \partial h = (\partial L / \partial h) \cdot U + L \cdot (\partial U / \partial h). \quad (8.73)$$

Умножив обе части уравнения, слева на Y^t и справа на X, получим

$$Y^t \cdot (\partial T / \partial h) \cdot X = Y^t \cdot (\partial L / \partial h) \cdot U \cdot X + Y^t \cdot L \cdot (\partial U / \partial h) \cdot X. \quad (8.74)$$

Поскольку, в общем случае, векторы X и Y произвольны, определим их из уравнений

$$U \cdot X = e_n, \quad (8.75)$$

$$L^t \cdot Y = l_{nn} \cdot e_n, \quad (8.76)$$

где e_n - вектор с компонентой n, равной единице, остальные равны нулю; l_{nn} - (n,n)- тый элемент нижней треугольной матрицы L. Так как в полюсе матрица T вырождена, то элемент $l_{nn} = 0$ и правая часть уравнения (8.76) будет нулевым вектором. (Для обеспечения этого условия может понадобиться преобразование подобия с полным либо частичным выбором элемента равного нулю и установки его на место l_{nn}. Выбор и перестановки производятся при LU- разложении матрицы T.) Поскольку система (8.76) вырождена, есть возможность произвольного выбора одного компонента вектора Y. Для удобства вычислений положим $y_n = 1$.

Теперь можно показать, что решение уравнений определяющих X и Y, позволяет рассчитать чувствительность полюсов, для чего подставим (8.75) и (8.76) в (8.74)

$$Y^t \cdot (\partial T / \partial h) \cdot X = Y^t \cdot (\partial L / \partial h) \cdot e_n + l_{nn} \cdot e_n^t \cdot (\partial U / \partial h) \cdot X. \quad (8.77)$$

Так как матрица L является нижней треугольной, то произведение $(\partial L / \partial h) \cdot e_n$ равно вектору, у которого все элементы нулевые, за исключением последнего, равного $(\partial l_{nn} / \partial h)$, т.е.

$$(\partial L / \partial h) \cdot e_n = (\partial l_{nn} / \partial h) \cdot e_n.$$

Этот вектор умножается слева на вектор Y^t, последний элемент которого $y_n = 1$. Кроме того, произведение $e_n^t \cdot (\partial U / \partial h)$ является нулевым вектором, поскольку матрица U является верхней треугольной с единичными диагональными элементами $u_{ii} = 1$, т.е.

$$e_n^t \cdot (\partial U / \partial h) = e_n^t \cdot (\partial u_{nn} / \partial h) = 0 .$$

В результате получаем

$$Y^t \cdot (\partial T / \partial h) \cdot X = \partial l_{nn} / \partial h . \qquad (8.78)$$

Заметим, что в точке полюса вместо равенства $\Phi(s,h) = 0$ можно использовать соотношение $l_{nn}(s,h) = 0$.

Так как полюс является нулем знаменателя дробно - рационального представления выходной функции, то, используя соотношение (8.70), можно определить чувствительность полюса

$$\begin{aligned} \partial p_i / \partial h &= -(\partial l_{nn} / \partial h) / (\partial l_{nn} / \partial s)_{/ s = p_i} = \\ &= -Y^t \cdot (\partial T / \partial h) \cdot X / (Y^t \cdot C \cdot X) \end{aligned} . \qquad (8.79)$$

Используя это уравнение, можно рассчитать чувствительность полюса выходной функции цепи.

Обобщая изложенное, представим алгоритм расчета чувствительности полюсов следующей последовательностью действий:

1) подставляя $s = p_i$ в матричное уравнение, проведем LU-факторизацию матрицы T;

2) из уравнения $U \cdot X = e_n$ обратной подстановкой определим вектор X;

3) из неопределенного уравнения $L^t \cdot Y = l_{nn} \cdot e_n$, полагая $y_n = 1$ обратной
подстановкой, определим остальные компоненты вектора Y;

4) с помощью соотношения (8.79), определим чувствительность текущего полюса p_i.

В тех случаях, когда необходимо произвести полный или частичный выбор ведущего элемента, чтобы l_{nn} был равен нулю, LU- факторизация запишется

$$P_1 \cdot T \cdot P_2 = L \cdot U ,$$

откуда

$$\partial l_{nn} / \partial h = Y^t \cdot P_1 \cdot (\partial T / \partial h) \cdot P_2 \cdot X .$$

где P_1 и P_2 - преобразующие матрицы, полученные из единичных перестановкой соответствующих строк и столбцов. Можно считать, что P_1 и P_2 преобразуют вектора Y и X, изменяя порядок следования их элементов.

Температурная чувствительность. Как уже отмечалось, чувствительность по отношению к параметрам, не входящим напрямую в матрицу коэффициентов системы, но элементы, которой зависят от этих

параметров, можно определить, пользуясь правилом дифференцирования сложных функций. Рассмотрим вопрос об определении чувствительности выходного напряжения к температуре $h = t^\circ$. Допустим, что m- тый элемент цепи имеет температурную зависимость, выражаемую соотношением

$$E_m = E_{m0} \cdot f_m(t^\circ). \tag{8.80}$$

Заменим функцию f_m несколькими первыми членами разложения ее в ряд Тейлора

$$E_m = E_{m0} \cdot (1 + r_m \cdot t^\circ + \cdots). \tag{8.81}$$

Применим правило дифференцирования сложных функций для получения $\partial \Phi / \partial t^\circ$, при условии, что температура действует только на m- тый элемент

$$\partial \Phi / \partial t^\circ = (\partial \Phi / \partial E_m) \cdot (\partial E_m / \partial t^\circ). \tag{8.82}$$

Здесь $\partial \Phi / \partial E_m$- чувствительность по отношению к параметру m- го элемента, которым может быть G, L, C или любой другой параметр, входящий в матрицу коэффициентов системы. Производную $\partial E_m / \partial t^\circ$ получаем из разложения в ряд Тейлора

$$\partial E_m / \partial t^\circ = E_{m0} \cdot [\partial f_m(t^\circ) / \partial t^\circ]. \tag{8.83}$$

Если использовать только линейный член разложения в ряд, то в соответствии с (8.81), получим

$$\partial E_m / \partial t^\circ = E_{m0} \cdot r_m. \tag{8.84}$$

Однако если от температуры зависят несколько элементов, то необходимо суммировать парциальные чувствительности по всем элементам

$$\partial \Phi / \partial t^\circ = \sum E_{m0} \cdot (\partial \Phi / \partial E_m) \cdot (\partial f_m(t^\circ) / \partial t^\circ). \tag{8.85}$$

В качестве примера, рассчитаем температурную чувствительность цепи рисунке 8.2.

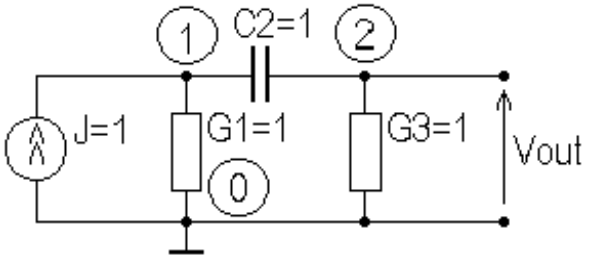


Рисунок 8.2 Простая цепь

Температурным коэффициентом емкости пренебрежем. Положим $s = j$. Система узловых уравнений цепи имеет вид

$$\begin{bmatrix} G_1 + s \cdot C_2 & -s \cdot C_2 \\ -s \cdot C_2 & C_3 + s \cdot C_2 \end{bmatrix} \cdot \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}.$$

Согласно (8.49), учитывая, что $\partial W / \partial t^\circ = 0$, можно записать

$$\partial \boldsymbol{\Phi} / \partial E_m = Y^t \cdot (\partial T / \partial E_m) \cdot X .$$

Подставив значения элементов цепи и решая исходную и транспонированную системы, получаем

$$X = T^{-1} \cdot W = V = \begin{bmatrix} (3-j)/5 \\ (2+j)/5 \end{bmatrix},$$

$$Y = -(T^t)^{-1} \cdot d = \begin{bmatrix} (-2-j)/5 \\ (-3+j)/5 \end{bmatrix},$$

где

$$T = T^t = \begin{bmatrix} 1+j & -j \\ -j & 1+j \end{bmatrix}, \quad T^{-1} = (T^t)^{-1} = \begin{bmatrix} (3-j)/5 & (2+j)/5 \\ (2+j)/5 & (3-j)/5 \end{bmatrix}, \quad d = \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

Так как $G_1 = G_1(t^\circ)$ и $G_3 = G_3(t^\circ)$, то

$$\partial T / \partial h_1 = \partial T / \partial E_1 = \partial T / \partial G_1 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix},$$

$$\partial T / \partial h_2 = \partial T / \partial E_2 = \partial T / \partial G_3 = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix},$$

откуда

$$\partial \boldsymbol{\Phi} / \partial G_1 = \partial v_2 / \partial G_1 = y_1 \cdot v_1 = (-7-j)/25,$$
$$\partial \boldsymbol{\Phi} / \partial G_3 = \partial v_2 / \partial G_3 = y_2 \cdot v_2 = (-7-j)/25,$$

и

$$\partial \boldsymbol{\Phi} / \partial t^\circ = \partial v_2 / \partial t^\circ =$$
$$= (\partial v_2 / \partial G_1) \cdot (\partial G_1 / \partial t^\circ) + (\partial v_2 / \partial G_3) \cdot (\partial G_3 / \partial t^\circ) .$$

Если проводимости $G_1 = G_3$ имеют одинаковый температурный коэффициент, то

$\partial G_1 / \partial t^\circ = \partial G_3 / \partial t^\circ = r_m$, то $\partial v_2 / \partial t^\circ = 2 \cdot r_m \cdot (-7-j)/25$.

Эквивалентные генераторы тока и напряжения. Поведение цепи относительно любой выделенной пары зажимов можно описать с помощью эквивалентного генератора тока либо напряжения - так называемые эквиваленты Нортона и Тевенина. Для этого предлагается выполнить следующие преобразования. Вначале необходимо найти ток короткого замыкания или напряжение холостого хода выделенных зажимов. На втором этапе необходимо исключить независимые источники тока, закоротить независимые источники напряжения и путем подключения к выделенным зажимам единичного источника тока или источника напряжения вычислить напряжение на этих зажимах или протекающий ток. Условно эти преобразования можно отобразить в виде рисунка 8.3, где из схемы N вынесены независимые источники и для простоты показана цепь с одним независимым источником каждого типа.

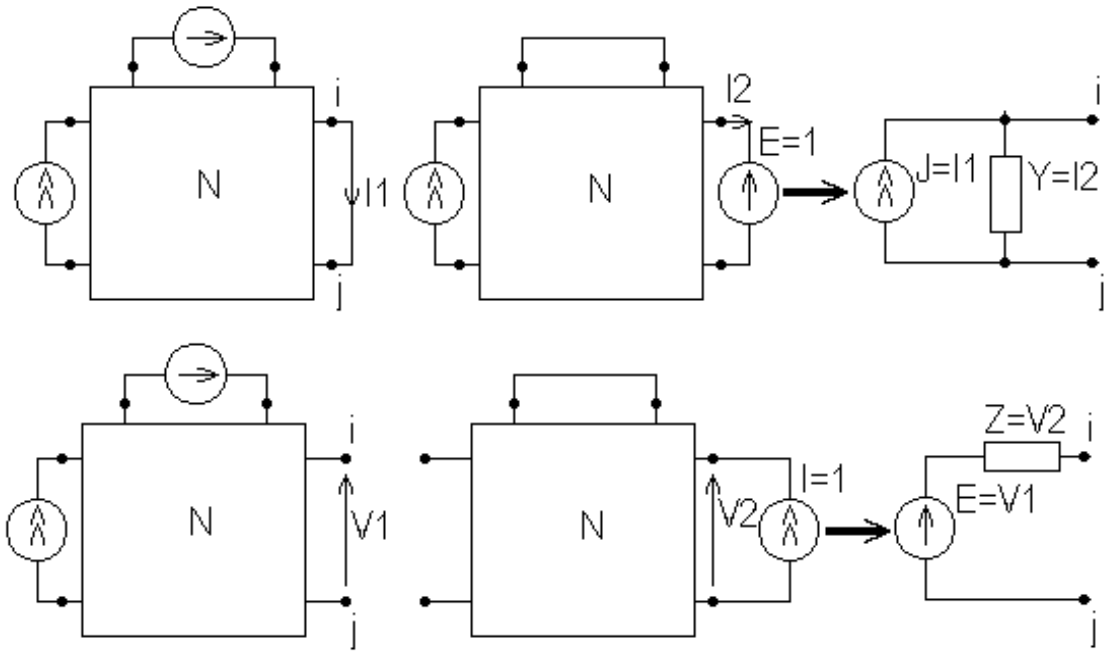


Рисунок 8.3 Представление эквивалентными генераторами

Прямое выполнение указанных преобразований неизбежно привело бы, по крайней мере, к двукратным вычислениям по двум различным схемам. Вначале по исходной схеме вычислили бы ток короткого замыкания или напряжение холостого хода, а затем, убрав внутренние независимые источники и, поставив на выделенные зажимы единичный генератор напряжения или тока, определили бы эквивалентную проводимость или сопротивление. Однако использование сопряженной системы уравнений позволяет выполнить необходимые вычисления гораздо эффективнее.

С математической точки зрения, требуемые преобразования можно представить уравнениями

$$T \cdot X_i = W_i, \tag{8.86}$$

$$\Phi_1 = d^t \cdot X_i, \tag{8.87}$$

где $i = 1,2$; W_i - два вектора соответствующих источникам; Φ_i - требуемые выходные величины. Здесь предполагается, что в W_1 включены первоначальные источники, а W_2 содержит требуемый единичный источник. Предполагается также, что выходная ветвь представлена соответственно проводимостью либо сопротивлением.

Подставив решение уравнения (8.86) в (8.87), получим

$$\Phi_1 = d^t \cdot T^{-1} \cdot W_i = Y^t \cdot W_i, \tag{8.88}$$

где Y^t - решение присоединенной системы

$$T^t \cdot Y = d. \tag{8.89}$$

Таким образом, обе требуемые величины - ток или напряжение и эквивалентную проводимость либо сопротивление можно рассчитать, решив

присоединенную систему уравнений и вычислив дважды произведение векторов, согласно уравнения (8.88).

Анализ шумов. Как уже отмечалось, наиболее важными шумовыми составляющими электронных схем являются - дробовой, тепловой и фликкер шумы. Два первых типа имеют вполне однозначную природу и выражения для их интенсивности. Третий вид шумов не поддается четкому описанию, и используются эмпирические соотношения.

Шумы обычно представляются в виде некоррелированных источников. Любую пару коррелированных источников шума, всегда можно представить набором некоррелированных источников, причем дополнительные источники включаются между исходными и имеют интенсивность, равную взаимной спектральной плотности источников.

Предметом анализа шумов является обычно определение вклада в выходной сигнал, как шумов источника сигнала, так и внутренних шумов устройства, т.е. речь, может идти о вычислении соотношения сигнал/шум на выходе устройства.

Интенсивность источников шума описывается спектральной плотностью, т.е. мощностью шумов, приходящейся на единицу полосы частот, таким образом, информация о фазе теряет смысл, что препятствует использованию принципа суперпозиции в обычном смысле. Поскольку спектральными плотностями в нашем подходе пользоваться неудобно, поэтому, учитывая, что спектральные плотности пропорциональны квадратам токов либо напряжений источников, опишем их интенсивность, как корень квадратный из спектральной плотности. Для учета независимости источников, т.е. исключения информации о фазе, вклад каждого источника будем рассматривать независимо.

Таким образом, необходимо последовательно рассчитать цепь с каждым из источников. Суммарная амплитуда источников на выходе равна корню квадратному из суммы квадратов каждого из вкладов. В сложных цепях число шумовых источников велико и могло бы потребоваться многократное решение системы уравнений. Однако, как было показано ранее, можно воспользоваться решением присоединенной системы уравнений, что позволит существенно сократить объем вычислений.

Формально наша задача сводится к решению систем

$$T \cdot X_i = t_i \cdot W_i, \tag{8.90}$$

где $i = 0,\cdots,m$; t_i - интенсивность i - го источника.

Выходная величина, как известно, есть линейная комбинация компонент вектора решений

$$\Phi_1 = d^t \cdot X_i. \tag{8.91}$$

Подставляя решение уравнения (8.90) в (8.91) получаем

$$\Phi_1 = d^t \cdot T^{-1} \cdot t_i \cdot W_i = t_i \cdot Y^t \cdot W_i, \tag{8.92}$$

где Y^t - решение присоединенной системы

$$T^t \cdot Y = d. \tag{8.93}$$

Индекс $i=0$ соответствует источнику входного сигнала, индексы $i=1,\cdots,m$ соответствуют источникам шума.

Таким образом, при расчете шумов вначале находим решение присоединенной системы, а затем определяем вклад каждого источника в выходную величину. Так как каждый вектор W_i содержит информацию об одном источнике и включает не более двух ненулевых компонент ± 1, то вычисление вклада сводится к одному вычитанию компонент y_i.

Амплитуду сигнала на выходе обозначим через $|\Phi_0|$, а амплитуду шумов представим выражением

$$A_N = (\sum_{i=1}^{m} |\Phi_i|^2)^{1/2}.$$

Интенсивности t_i зависят от типа элемента, так для тепловых шумов

$$t_i = \sqrt{4 \cdot k \cdot T \cdot \Delta f \cdot G_i},$$

где k - постоянная Больцмана; T - температура в градусах Кельвина; Δf - ширина полосы; G_i - проводимость.

В качестве примера, вычислим соотношение сигнал/шум схемы, изображенной на рисунке 8.4.

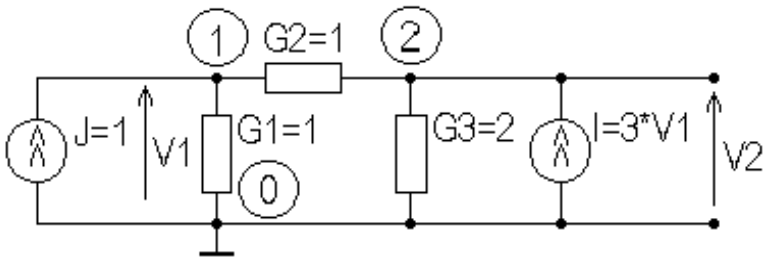


Рисунок 8.4 Схема для расчета шумов

Шумами зависимого источника пренебрежем, источник сигнала полагаем не шумящим. Узловая система уравнений для схемы имеет вид

$$\begin{bmatrix} G_1+G_2 & -G_2 \\ -G_2-g & G_2+G_3 \end{bmatrix} \cdot \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} J \\ 0 \end{bmatrix}.$$

Подставляя конкретные значения, запишем присоединенную систему уравнений $T^t \cdot Y = d$ в виде

$$\begin{bmatrix} 2 & -4 \\ -1 & 3 \end{bmatrix} \cdot \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix},$$

решив которую, получим

$$Y = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}.$$

Найдем сигнал на выходе цепи, используя соотношение (8.92)

$$\Phi_0 = t_0 \cdot Y^t \cdot W_0 = 2 \cdot \begin{bmatrix} 2 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 0 \end{bmatrix} = 4.$$

Амплитуды шумов, обусловленные G_1, G_2, G_3, соответственно равны

$$\Phi_1 = t_1 \cdot Y^t \cdot W_1 = t_1 \cdot \begin{bmatrix} 2 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 0 \end{bmatrix} = 2 \cdot t_1,$$

$$\Phi_2 = t_2 \cdot Y^t \cdot W_2 = t_2 \cdot \begin{bmatrix} 2 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ -1 \end{bmatrix} = t_2,$$

$$\Phi_3 = t_3 \cdot Y^t \cdot W_3 = t_3 \cdot \begin{bmatrix} 2 & 1 \end{bmatrix} \cdot \begin{bmatrix} 0 \\ 1 \end{bmatrix} = t_3.$$

Найдем амплитуду шума на выходе

$$A_N = (4 \cdot t_1^2 + t_2^2 + t_3^2)^{1/2} = \sqrt{4 \cdot k \cdot T \cdot \Delta f} \cdot (4 \cdot G_1 + G_2 + G_3)^{1/2} = \sqrt{28 \cdot k \cdot T \cdot \Delta f}$$

и соотношение сигнал/шум на выходе цепи

$$\rho = 4 / \sqrt{28 \cdot k \cdot T \cdot \Delta f \cdot G_i}.$$

На этом завершим раздел по расчету чувствительностей РЭУ и их приложений для вычисления других характеристик.

9 РАСЧЕТ ЦЕПЕЙ ПО ПОСТОЯННОМУ ТОКУ

9.1 Алгоритм Ньютона – Рафсона

Определение рабочей точки или расчет электрических цепей по постоянному току является обычно первым шагом при анализе нелинейных схем. Дело в том, что от режима работы, т.е. рабочей точки, существенно зависят параметры нелинейных элементов – диодов, транзисторов и так далее и, соответственно, характеристики нелинейных устройств. Математической моделью цепи по постоянному току является, в общем случае, система нелинейных алгебраических уравнений. При расчетах цепей по постоянному току используются нелинейные статические модели элементов. Реактивные элементы схемы в этом случае исключаются – конденсаторы, заменяются ветвями ХХ, а катушки индуктивности - ветвями КЗ. Аналитические решения нелинейных алгебраических систем, как правило, отсутствуют и используются итерационные методы, позволяющие определить приближенное решение с любой наперед заданной точностью.

Расчет по постоянному току включает в себя определение установившихся напряжений и токов цепи при включении источников питания и требует в общем случае решения систем нелинейных алгебраических уравнений. Наиболее распространенным алгоритмом решения систем нелинейных алгебраических уравнений является алгоритм Ньютона–Рафсона. В этом разделе предстоит рассмотреть данный алгоритм применительно к наиболее известным методам формирования математических моделей электронных схем – обобщенному узловому, табличному, модифицированному узловому и модифицированному узловому с проверкой.

Алгоритм Ньютона–Рафсона. Алгоритм Ньютона–Рафсона часто используется как один из методов отыскания корней полиномов и имеет квадратичную сходимость при хорошем начальном приближении.

В скалярном представлении, при решении в общем случае нелинейного уравнения вида f(x)=0, итерации вычисления очередного решения определяется выражением

$$X^{k+1} = X^k + \Delta X^k = X^k + f(X^k)/f'(X^k), \qquad (9.1)$$

где k - номер итерации.

Для некоторых простейших цепей возможно исключение промежуточных переменных и сведение задачи к поиску решения одного нелинейного уравнения. Для иллюстрации итерационной природы алгоритма рассмотрим подобный пример для схемы, изображенной на рисунке 9.1.

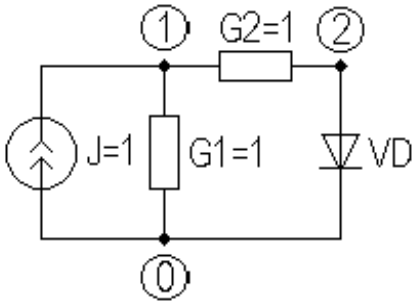


Рисунок 9.1 - Простая нелинейная схема

Пусть вольтамперная характеристика полупроводникового диода определяется упрощенным выражением

$$i_D = exp(40 \cdot V_D) - 1,$$

здесь использовано $V_D / \varphi_T \cong V_D / 25.6\ [мВ] \cong 40 \cdot V_D$.

Значения номиналов других ветвей приведены на рисунке 9.1.

Узловая система уравнений для данной схемы запишется в виде

$$3 \cdot v_1 - 2 \cdot v_2 = 1,$$

$$-2 \cdot v_1 + 2 \cdot v_2 + (exp(40 \cdot v_2) - 1) = 0.$$

Поскольку v_1 входит в оба уравнения линейно, исключим это напряжение, выразив из первого уравнения v_1, через v_2

$$v_1 = 1/3 + 2/3 \cdot v_2,$$

и, подставив его во второе уравнение, получим

$$f(v_2) = 2/3 \cdot v_2 + exp(40 \cdot v_2) - 5/3 = 0.$$

Производная от этой функции-выражения запишется

$$f'(v_2) = 2/3 + 40 \cdot exp(40 \cdot v_2) = 0.$$

Приняв в качестве начального значения $v_2^0 = 0.1\ B$ и, подставив полученные выражения в (9.1), в результате итераций с заданной точностью, получим установившееся значение $v_2 = 1.264388\,E-2\ B$, откуда однозначно следует $v_1 = 3.417626\,E-1\ B$.

Небольшое изменение цепи, например замена проводимости G_1 диодом, приведет уже к двум нелинейным уравнениям, поэтому целесообразно рассмотреть развитие метода Ньютона–Рафсона применительно к системе нелинейных алгебраических уравнений.

Рассмотрим систему n n нелинейных уравнений с n n переменными x_i

$$f_1(x_1, x_2, ..., x_n) = 0;$$

$$f_2(x_1, x_2, ..., x_n) = 0;$$

$$\ldots\ldots\ldots\ldots\ldots\ldots$$

$$f_n(x_1, x_2, ..., x_n) = 0.$$

Обозначим вектор переменных через X, а вектор функций через F, тогда, в общем виде, эту систему можно записать как

$$F(X) = 0. \qquad (9.2)$$

Опишем решение системы нелинейных уравнений (9.2) основанное на ее линеаризации. Линеаризация представляет собой достаточно распространенный прием преобразования нелинейной системы в линейную систему, путем разложения ее в окрестности предполагаемого решения в ряд Тейлора и удержания первых линейных членов ряда, включая первые производные.

Итак, предполагая, что система имеет решение X^*, разложим каждую функцию системы в ряд Тейлора в окрестности предполагаемого решения

$$f_1(x^*) = f_1(x) + \frac{\partial f_1}{\partial x_1}(x^* - x_1) + \frac{\partial f_1}{\partial x_2}(x^* - x_2) + \cdots + \frac{\partial f_1}{\partial x_n}(x^* - x_n) + \cdots;$$

$$f_2(x^*) = f_2(x) + \frac{\partial f_2}{\partial x_1}(x^* - x_1) + \frac{\partial f_2}{\partial x_2}(x^* - x_2) + \cdots + \frac{\partial f_2}{\partial x_n}(x^* - x_n) + \cdots;$$

……

$$f_n(x^*) = f_n(x) + \frac{\partial f_n}{\partial x_1}(x^* - x_1) + \frac{\partial f_n}{\partial x_2}(x^* - x_2) + \cdots + \frac{\partial f_n}{\partial x_n}(x^* - x_n) + \cdots.$$

Предположив, что X близко к $X^* = X + \Delta X$, пренебрежем членами выше первого порядка и запишем систему в линеаризованной форме

$$F(X^*) \cong F(X) + M(X) \cdot (X^* - X),$$

где

$$M(X) = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_n} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{\partial f_3}{\partial x_1} & \frac{\partial f_3}{\partial x_2} & \cdots & \frac{\partial f_3}{\partial x_n} \end{bmatrix}$$

матрица Якоби.

Если приравнять к нулю полученную систему уравнений, то решение не будет точно равно X^* из-за пренебрежения членами более высокого порядка и будет равно некоторому новому значению X. Отклонение от точного решения зависит от того, насколько хорошо многомерная поверхность, соответствующая нелинейной системе, аппроксимируется многомерной плоскостью в окрестности решения соответствующей линеаризованной системы. Кроме того, известно, что при соблюдении ряда условий и, в частности, при наличии хорошего начального приближения, повторное решение линеаризованной системы, при использовании предыдущего решения, в качестве нового начального приближения, обеспечивает снижение погрешности решения.

Таким образом, мы пришли к понятию итерации, основанной на повторном решении системы и понятию сходимости решения, т.е.

уменьшении ошибки при использовании предыдущего решения для вычисления нового.

Используя верхние индексы для обозначения последовательности итераций, можем записать линеаризованную систему в виде

$$F(X^k) + M(X^k) \cdot (X^{k+1} - X^k) = 0.$$

Формально решение этого уравнения на текущей итерации запишется

$$X^{k+1} = X^k - M^{-1}(X^k) \cdot F(X^k). \quad (9.3)$$

На практике стараются обойтись без явного обращения матрицы Якоби.

Так некоторые авторы, предлагают использовать итерационные соотношения для вычисления обратной матрицы Якоби текущей итерации через известную обратную матрицу на предыдущей итерации. Если новую матрицу Якоби представить как

$$M^{k+1} = M^k + \Delta M^k = M^k \cdot (E + (M^k)^{-1} \cdot \Delta M^k), \quad (9.4)$$

Тогда с определенным приближением можно записать

$$(M^{k+1})^{-1} \cong (E - (M^k)^{-1} \cdot \Delta M^k) \cdot (M^k)^{-1}, \quad (9.5)$$

где E - единичная матрица; ΔM^k - матрица приращения компонент матрицы Якоби на k - той итерации. Для малых приращений выражение (9.5) можем переписать в виде

$$(M^{k+1})^{-1} \cong (M^k)^{-1} - (M^k)^{-1} \cdot \Delta M^k \cdot (M^k)^{-1}. \quad (9.6)$$

Таким образом, получив однажды обратную матрицу и при условии малости приращений на очередной итерации, можно воспользоваться соотношением (9.6) для нахождения приближенного значения обратной матрицы следующей итерации. Использование этого соотношения, однако, ограничивается требованием обеспечения малости приращений. Обозначив $\Delta X^k = X^{k+1} - X^k$, перепишем уравнение (9.3) в виде

$$M(X^k) \cdot \Delta X^k = -F(X^k). \quad (9.7)$$

Решение уравнения, т.е. вектор приращений ΔX, найдем, например, с помощью LU- факторизации, а новое значение вектора переменных определим из уравнения

$$X^{k+1} = X^k + \Delta X^k. \quad (9.8)$$

Совокупность уравнений (9.7) и (9.8) есть запись алгоритма Ньютона–Рафсона.

Отметим также, что если в соотношении (9.7) убрать знак минус в первой части, тогда знак минус появиться в соотношении (9.8) перед вторым слагаемым. Алгоритм имеет довольно быструю сходимость – квадратичную вблизи точки решения. Недостаток алгоритма заключается в необходимости вычисления матрицы Якоби на каждой итерации.

Можно показать, что цель алгоритма заключается в уменьшении нормы ошибки от итерации к итерации

$$\left|F(X^{k+1})\right| \le \left|F(X^k)\right|. \quad (9.9)$$

Для обеспечения сходимости зачастую используют модифицированную форму уравнения (9.8)

$$X^{k+1} = X^k + t^k \cdot \Delta X^k, \qquad (9.10)$$

где t^k- параметр, выбираемый обычно в интервале $(0 \le t^k \le 1)$ для обеспечения сходимости, таким образом, чтобы выполнялось соотношение (9.9).

Проиллюстрируем применение алгоритма Ньютона-Рафсона, на примере решения системы нелинейных уравнений для двух диодной цепи, изображенной на рис.9.2.

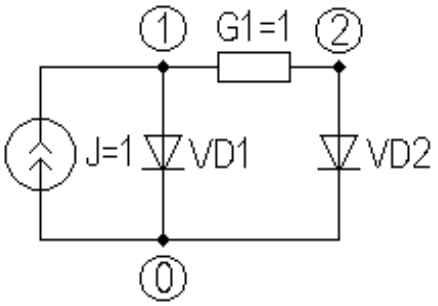


Рисунок 9.2 - Нелинейная цепь на двух диодах

Пусть каждый диод представлен упрощенной вольтамперной характеристикой

$$i_D = exp(40 \cdot V_D) - 1,$$

а начальные значения для напряжений на диодах, совпадающие с узловыми потенциалами, примем равными $v_1^0 = v_2^0 = 0.1 \ \ B$.

Метод узловых потенциалов дает следующую систему уравнений

$$i_{D1} + G \cdot (v_1 - v_2) = J,$$
$$-G \cdot (v_1 - v_2) + i_{D2} = 0.$$

Раскрывая выражение для токов диодов, и подставляя численные значения, получаем

$$f_1(v_1, v_2) = exp(40 \cdot v_2) + v_1 - v_2 - 2 = 0,$$
$$f_2(v_1, v_2) = -v_1 + v_2 + exp(40 \cdot v_2) - 1 = 0.$$

Вектор нелинейных функций и Якобиан системы определяется выражениями

$$f_1(v_1, v_2) = -J + (exp(40 \cdot V_{D1}) - 1),$$
$$f_2(v_1, v_2) = (exp(40 \cdot V_{D2}) - 1),$$
$$M(v_1, v_2) = \begin{bmatrix} 40 \cdot exp(40 \cdot v_1) + 1 & -1 \\ -1 & 40 \cdot exp(40 \cdot v_2) + 1 \end{bmatrix}.$$

При заданных начальных значениях, нелинейные функции и Якобиан, равны

$$f_1 = 52.59815; \quad f_2 = 53.59815;$$
$$M = \begin{bmatrix} 2184.926 & -1 \\ -1 & 2184.926 \end{bmatrix}.$$

Решение исходной нелинейной системы дает

$$\Delta v_1 = -0.0240844; \quad \Delta v_2 = -0.0245419.$$

Прибавляя полученные значения к начальным приближениям, получим

$$v_1^1 = v_1^0 + \Delta v_1^0 = 0.0759156\ ;$$
$$v_2^1 = v_2^0 + \Delta v_2^0 = 0.0754581\ .$$

Расчеты значений на ЭВМ, с точностью до пятого знака после запятой, дают следующие результаты для напряжений: $v_1 = 0.01712\ \ B$; $v_2 = 0.00041\ \ B$. Хотя начальное приближение далеко отстояло от полученного решения, алгоритм сошелся за 7 итераций.

Модификация Бройдена. Как уже отмечалось, алгоритм Ньютона–Рафсона имеет хорошую сходимость, однако требует вычисления матрицы Якоби, либо решения линеаризованной системы уравнений на каждом шаге итераций, что естественно ведет к большим затратам машинного времени. Бройденом была предложена модификация алгоритма Ньютона–Рафсона, лишенная этого недостатка. Модификация Бройдена имеет два следующих отличия:

1) на каждой итерации не формируют матрицу Якоби и не вычисляют обратную, не вычисляют дополнительные функции для получения численных оценок отклонения, а используют лишь функции, определяемые из постоянной матрицы схемы;
2) на каждой итерации рассчитывают коэффициент затухания, указывающей на сходимость и коэффициент позволяющий оценить ошибку вычисления до окончания решения.

Суть метода Бройдена заключается в использовании ранее упоминаемого весового коэффициента t^k, который, в модификации Бройдена, может быть больше единицы для обеспечения большей скорости сходимости

$$\Delta X^k = -M^{-1}(X^k)\cdot F(X^k)\,, \tag{9.11}$$
$$X^{k+1} = X^k + t^k \cdot \Delta X^k\,. \tag{9.12}$$

Вместо $M^{-1}(X^k)$, используется приближение к ней, вычисляемое на каждой итерации, в соответствии с выражением

$$H^{k+1} = H^k - \frac{\left[t^k \cdot \Delta X^k + H^k \cdot \left(F(X^{k+1}) - F(X^k)\right)\right]\cdot \left(\Delta X^k\right)^t \cdot H^k}{\left(\Delta X^k\right)^t \cdot H^k \cdot \left(F(X^{k+1}) - F(X^k)\right)}\,, \tag{9.13}$$

где $H^0 = (M^0)^{-1}$ – обратная матрица Якоби при начальном значении.

Таким образом, модификацию Бройдена, алгоритма Ньютона–Рафсона, можно представить следующей последовательностью действий.

1. Задание начального значения вектора переменных X^0.

2. Вычисление начального значения H^0 путем обращения матрицы Якоби $(M^0)^{-1}$.

3. Вычисление $F(X^k)$.

4. Вычисление $\Delta X^k = H^k \cdot F(X^k)$.

5. Выбор t^k, при котором $\left|F(X^{k+1})\right| \le \left|F(X^k)\right|$.

6. Расчет $X^{k+1} = X^k + t^k \cdot \Delta X^k$.

7. Проверка нормы вектора $\left|F(X^{k+1})\right|$ на сходимость.

8. Расчет $F(X^{k+1}) - F(X^k)$.

9. Вычисление H^{k+1} по соотношению (9.13).

10. Повторение вычислений, начиная с этапа 4.

9.2 Формирование нелинейных математических моделей

Обобщим, ранее изложенные методы формирования математических моделей линейных схем, на нелинейные схемы. Как и прежде, рассмотрим наиболее распространенные прямые методы формирования математических моделей – обобщенный метод узловых потенциалов, табличный, модифицированный табличный, модифицированный узловой и модифицированный узловой с проверкой.

При отыскании решения по постоянному току в цепи все катушки индуктивности закорачиваются (ветвь КЗ), а все конденсаторы исключаются, т.е. заменяются ветвью холостого хода (ХХ).

Обобщенный метод узловых потенциалов. Прежде всего, заметим, что уравнения для узловых потенциалов требуют, чтобы резисторы с нелинейным сопротивлением описывались в форме

$$i_b = g(V_b). \tag{9.14}$$

Индекс b будет обозначать напряжения и токи ветвей, а индекс n используется для обозначения узловых переменных. Предполагается также, что все независимые источники представлены источниками тока.

Запишем закон Кирхгофа для токов ветвей и выразим напряжения ветвей через напряжения узлов

$$A \cdot i_b = 0, \tag{9.15}$$

$$V_b = A^t \cdot V_n. \tag{9.16}$$

Разделив все ветви на две группы – ветви независимых источников тока и другие, можем переписать соотношение (9.15) в виде

$$A \cdot i_b = -A_J \cdot i_J, \tag{9.17}$$

или, обозначив узловые токи J_n

$$J_n = -A_J \cdot i_J, \tag{9.18}$$

окончательно получим

$$A \cdot i_b = J_n. \tag{9.19}$$

Подставив линейное уравнение ветви (9.14) в (9.19), получим

$$A \cdot g(V_b) = J_n, \tag{9.20}$$

а, используя (9.16), можем записать обобщенную форму узловых уравнений нелинейной цепи

$$A \cdot g(A^t \cdot V_n) = J_n . \quad (9.21)$$

Для представления алгоритма Ньютона-Рафсона перепишем узловую систему нелинейной цепи в виде

$$F(V_n) \equiv A \cdot g(A^t \cdot V_n) - J_n = 0 . \quad (9.22)$$

Дифференцированием сложной функции получаем матрицу Якоби узловой системы нелинейной цепи

$$M(V_n) = \partial F(V_n) / \partial V_n = A \cdot (\partial g(V_b) / \partial V_b) \cdot (\partial V_b / \partial V_n) . \quad (9.23)$$

Вводя обозначение

$$G_b(V_b) = \partial i_b / \partial V_b = \partial g(V_b) / \partial V_b , \quad (9.24)$$

и учитывая, что дифференцирование по ∂V_n, соотношений для напряжений ветвей (9.16), дает

$$\partial V_b / \partial V_n = A^t , \quad (9.25)$$

и выражение для Якобиана примет вид

$$M(V_n) = A \cdot G_b(V_b) \cdot A^t . \quad (9.26)$$

Как видим, вектор функций представляет собой систему узловых компонентных уравнений ветвей, а правило формирования Якобиана из проводимостей ветвей соответствует рассмотренному ранее правилу формирования матрицы проводимостей узловой системы уравнений.

В частном случае, линейных сопротивлений соотношение (9.14) примет вид

$$i_b = F(V_b) = G \cdot V_b . \quad (9.27)$$

Вектор узловых токов, с учетом (9.19), можно, как известно, записать

$$J_n = A \cdot i_b = A \cdot G \cdot V_b = A \cdot G \cdot A^t \cdot V_n , \quad (9.28)$$

откуда вектор правой части системы уравнений Ньютона–Рафсона можно представить

$$F(V_n) = A \cdot G \cdot A^t \cdot V_n - J_n . \quad (9.29)$$

Выражение для производной тока ветви по напряжению ветви, представленное через разность узловых напряжений, запишется

$$M(V_n) = \partial F(V_n) / \partial V_n = \partial i_b / \partial V_n = A \cdot G \cdot A^t . \quad (9.30)$$

Откуда следует вывод, что компоненты вектора функции, правой части системы уравнений Ньютона–Рафсона для линейных ветвей, совпадают с линейными узловыми компонентными уравнениями ветвей, а Якобиан узловой системы уравнений для линейной цепи совпадает с дифференциальной матрицей проводимости.

Таким образом, формирование математической модели цепи по постоянному току узловым методом, в виде линеаризованной итерационной системы Ньютона-Рафсона, полностью совпадает с ранее рассмотренным случаем линейных цепей. Отличие заключается лишь в том, что формирование повторяется на каждом шаге итерации при новых уточненных значениях напряжений и токов. Компонентные уравнения нелинейных ветвей вектора функций и их производные по узловым напряжениям, как

проводимости нелинейных ветвей, включаемые в Якобиан, задаются аналитическими выражениями, вычисляемыми на каждом шаге итераций.

Обозначив через V^0 начальное приближение вектора решений, распишем основные пункты алгоритма Ньютона-Рафсона в терминах обобщенного метода узловых потенциалов.

1. Установить $k = 0$ и вычислить вектор узловых токов $J_n = -A_j \cdot i_j$.
2. Определить напряжения на ветвях $V_b^k = A^t \cdot V_n^k$.
3. Найти токи нелинейных $i_b^k = g(V_b^k)$ и линейных $i_b^k = G \cdot V_b^k$ ветвей.
4. Вычислить компонентную матрицу $G_b(V_b^k) = \partial g(V_b^k) / \partial V_b^k$. Для линейных ветвей элементы матрицы $G_b(V_b^k)$ совпадают со значениями их проводимостей.
5. Вычислить $M(V_n^k) = A \cdot G(V_b^k) \cdot A^t$ и $F(V_n^k) = A \cdot i_b^k - J_n$.
6. Решить уравнение Ньютона-Рафсона $M(V_n^k) \cdot \Delta V_n^k = -F(V_n^k)$.
7. Уточнить вектор решений $V_n^{k+1} = V_n^k + \Delta V_n^k$.
8. Если точность не достигнута, то установить $k = k + 1$ и перейти к пункту 2.

Проиллюстрируем использование алгоритма Ньютона-Рафсона в обобщенном узловом методе, на примере простой нелинейной цепи, используемой нами в начале раздела (рисунок 9.3).

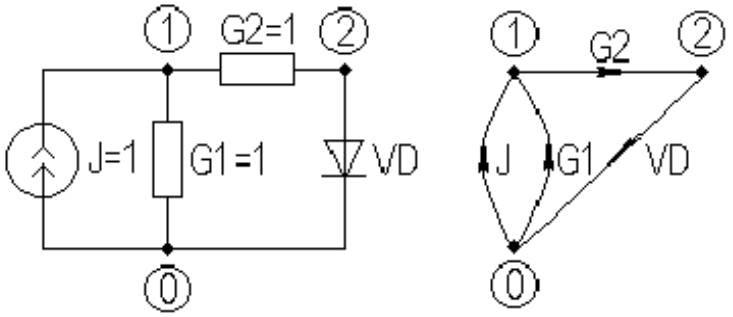


Рисунок 9.3 - Простая нелинейная цепь и ее граф

Пусть вектор начальных узлов напряжений определен как $V_n^0 = [0.3 \quad 0.02]^t$. Уравнение диода опишем простейшей вольт-амперной характеристикой $i_D = exp(40 \cdot V_D) - 1$. Значения номиналов других ветвей приведены на рисунке 9.3.

Дополненная матрица инциденций ветвей схемы, согласно рисунку 9.3, имеет вид

$$A_d = [A \quad A_J] = \begin{matrix} G1 & G2 & VD & J \end{matrix} \begin{bmatrix} -1 & 1 & 0 & -1 \\ 0 & -1 & 1 & 0 \end{bmatrix}.$$

Через матрицу инциденций ветвей независимых источников можно определить вектор узловых токов

$$J_n = -A_j \cdot J_b = -\begin{bmatrix} -1 \\ 0 \end{bmatrix} \cdot [1] = \begin{bmatrix} 1 \\ 0 \end{bmatrix}.$$

По матрице инциденций ветвей и начальному значению вектора узловых напряжений определим вектор начальных напряжений ветвей

$$V_b^0 = A^t \cdot V_n^0 = \begin{bmatrix} -1 & 0 \\ 1 & -1 \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 0.3 \\ 0.02 \end{bmatrix} = \begin{bmatrix} -0.3 \\ 0.28 \\ 0.02 \end{bmatrix}.$$

Вектор токов ветвей, согласно компонентным уравнениям, определится

$$I_b = \begin{bmatrix} G_1 \cdot v_{b1} \\ G_2 \cdot v_{b2} \\ exp(40 \cdot v_D) - 1 \end{bmatrix} = \begin{bmatrix} -0.3 \\ 0.56 \\ 1.22554093 \end{bmatrix}.$$

Матрица дифференциальных проводимостей ветвей равна

$$\frac{\partial I_b}{\partial V_b} = G_b = \begin{bmatrix} G_1 & 0 & 0 \\ 0 & G_2 & 0 \\ 0 & 0 & 40 \cdot exp(40 \cdot v_D) \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 89.02163712 \end{bmatrix}.$$

Через матрицу инциденций и линеаризованную матрицу дифференциальных проводимостей ветвей определим линеаризованную узловую матрицу проводимости

$$M(V_n) = A \cdot G_b \cdot A^t = \begin{bmatrix} 3 & -2 \\ -2 & 91.02163712 \end{bmatrix}.$$

Соответственно определяется вектор функций правой части системы Ньютона-Рафсона

$$F(V_n) = A \cdot I_b - J_n = \begin{bmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix} \cdot \begin{bmatrix} -0.3 \\ 0.56 \\ 1.22554093 \end{bmatrix} - \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} -0.14 \\ 0.66554093 \end{bmatrix}.$$

В результате система уравнений Ньютона-Рафсона, на основе обобщенного узлового метода, при заданных начальных значениях, примет вид

$$M(V_n^0) \cdot \Delta V_n^0 = F(V_n^0) = \begin{bmatrix} 3 & -2 \\ -2 & 91.02163712 \end{bmatrix} \cdot \begin{bmatrix} \Delta v_{n1}^0 \\ \Delta v_{n2}^0 \end{bmatrix} = -\begin{bmatrix} -0.14 \\ 0.66554093 \end{bmatrix}.$$

Решение системы на первой итерации равно

$$\Delta V_n^0 = \begin{bmatrix} \Delta v_{n1}^0 \\ \Delta v_{n2}^0 \end{bmatrix} = \begin{bmatrix} 0.042413361455 \\ -0.006379957841 \end{bmatrix}.$$

Соответственно, уточненные значения узловых напряжений после первой итерации при весовом коэффициенте $t^0 = 1$, будут равны

$$V_n^1 = V_n^0 + t_0 \cdot \Delta V_n^0 = \begin{bmatrix} 0.3424133614 \\ 0.013620042116 \end{bmatrix}.$$

Расчет на ЭВМ, с точностью до пятого знака после запятой, уже на четвертой

итерации привел к результатам

$$\Delta v_{n1}^{3} = -0.00001\ ;\ \Delta v_{n2}^{3} = -0.00002\ ;\ v_{n1}^{4} = 0.34176\ ;\ v_{n2}^{4} = 0.01264\ .$$

Таким образом, получен тот же результат, что и в первом примере данного раздела.

При выводе соотношений алгоритма Ньютона–Рафсона, применительно к узловому методу, был использован классический подход, основанный на совокупности компонентных и топологических уравнений. При этом топологические соотношения отображались матрицей инциденций. Однако в обобщенном узловом методе, можно, как ранее отмечалось при изложении метода, использовать и формальный подход. При этом матрицу Якоби и вектор функций правой части системы Ньютона–Рафсона можно сформировать напрямую по информации о ветвях схемы.

Матрица Якоби имеет такую же структуру, что и матрица проводимости. Проводимости линейных ветвей включаются в матрицу Якоби методом добавления, в соответствии с узлами подключения. Нелинейная проводимость $i_b = g(v_b)$, включенная между узлами k и l, вызовет узловые токи $i_k = g(v_k - v_l);;\ \ i_l = -g(v_k - v_l)$, добавляемые в вектор функций. Дифференцируя эти соотношения по узловым напряжениям, получаем следующий фрагмент матрицы Якоби

$$\begin{array}{c} \\ k \\ \cdot \\ l \end{array} \begin{array}{c} \begin{array}{ccc} k & \cdot & l \end{array} \\ \begin{bmatrix} \partial g / \partial v_k & \ldots & -\partial g / \partial v_l \\ \ldots & \ldots & \ldots \\ -\partial g / \partial v_k & \ldots & \partial g / \partial v_l \end{bmatrix} \end{array}.$$

Эти производные вычисляются при напряжениях, полученных на предыдущей итерации. Аналогично вносятся и другие нелинейные проводимости. Структура Якобиана фиксирована и дает возможность использовать алгоритмы, предназначенные для разряженных матриц при расчете сложных схем.

Правая часть системы уравнений Ньютона-Рафсона, определяемая соотношением $f(v_n^k) = A \cdot i_b^k - j_n$, также может быть сформирована напрямую из компонентных уравнений ветвей, методом добавления, в соответствии с узлами подключения. Компонентные уравнения представляют собой токи линейных и нелинейных ветвей, подключенных к узлу, а также токи независимых источников тока, подключенных к узлу.

Так диод, изображенный на схеме рисунка 9.4, определит следующие элементы Якобиана, и вектора правой части

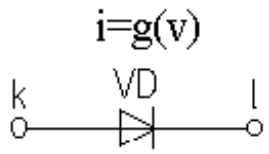


Рисунок 9.4 - Диод в качестве ветви схемы

$$M(V_n) = \begin{array}{c} \\ k \\ l \end{array} \begin{array}{c} \begin{array}{ccc} k & \cdot & l \end{array} \\ \begin{bmatrix} \partial g / \partial v_k & \dots & -\partial g / \partial v_l \\ \dots & \dots & \dots \\ -\partial g / \partial v_k & \dots & \partial g / \partial v_l \end{bmatrix} \end{array}; \quad F(V_n) = \begin{array}{c} k \\ l \end{array} \begin{vmatrix} + g(v) \\ - g(v) \end{vmatrix}.$$

Таким образом, обобщенный узловой метод может рассматриваться, как метод формирования математической модели нелинейной цепи.

Табличный метод. Перейдем к рассмотрению табличного метода формирования системы уравнений Ньютона–Рафсона.

Табличную систему, как совокупность компонентных и топологических уравнений, можно записать в общем виде

$$V_b - A^t \cdot V_n = 0, \tag{9.31}$$

$$P(V_b, I_b) = W_b, \tag{9.32}$$

$$A \cdot I_b = 0. \tag{9.33}$$

Компонентные уравнения (9.32) определяют связь между токами и напряжениями ветвей в неявной форме. Для линейных ветвей компонентные уравнения, как известно, принимают обобщенную линейную форму

$$Y_b \cdot V_b + Z_b \cdot I_b = W_b. \tag{9.34}$$

Соответственно, правая часть уравнения Ньютона-Рафсона может быть записана в виде

$$F(X) = \begin{vmatrix} V_b - A^t \cdot V_n \\ P(V_b, I_b) - W_b \\ A \cdot I_b \end{vmatrix} = 0,$$

где $X^t = \begin{bmatrix} V_b & I_b & V_n \end{bmatrix}^t$. Матрица Якоби на k - той итерации, как производная вектора $F(X)$ по компонентам вектора X, будет иметь вид

$$M(X) = \begin{bmatrix} 1 & 0 & -A^t \\ G^k & R^k & 0 \\ 0 & A & 0 \end{bmatrix},$$

где $G_k = \partial P / \partial V_b^k$; $R^k = \partial P / \partial I_b^k$.

Как видим, структура Якобиана совпадает с блочной формой табличной системы. Более того, в случае линейных цепей, вместо компонентного уравнения (9.32), можно записать уравнение (9.34) и в

результате дифференцирования вектора $F(X)$ убедимся, что Якобиан линейной цепи совпадает с традиционной матрицей коэффициентов табличной системы уравнений.

Система уравнений Ньютона–Рафсона построенная на основе табличного метода, как обычно, имеет вид

$$M(X^k)\cdot\Delta X^k = -F(X^k),$$

где $\Delta X^k = [\Delta V_b \quad \Delta I_b \quad \Delta V_n]^t$. После определения вектора приращений уточняем вектор неизвестных

$$X^{k+1} = X^k + t^k\cdot\Delta X^k.$$

Таким образом, табличный метод также может рассматриваться, как метод формирования математической модели нелинейной цепи.

Модифицированный табличный метод. В модифицированном табличном методе по сравнению с табличным методом с целью сокращения размерности системы уравнений (9.31-9.33) из рассмотрения исключается уравнение связи напряжений ветвей и узлов. При этом блочное уравнение (9.31) подставляется в уравнение (9.32) в результате чего получаем модифицированную табличную систему уравнений

$$P(A^t\cdot V_n, I_b) = W_b, \tag{9.35}$$

$$A\cdot I_b = 0. \tag{9.36}$$

Следовательно, вектор правой части системы уравнений Ньютона–Рафсона, запишется

$$F(X) = \begin{vmatrix} P(A^t\cdot V_n, I_b) - W \\ A\cdot I_b \end{vmatrix} = 0,$$

где $X = [I_b \quad V_n]^t$ - вектор неизвестных. Взяв производную на k - той итерации от вектора функций $F(X)$, по компонентам вектора неизвестных X, получим Якобиан

$$M(X) = \begin{bmatrix} G^k\cdot A^t & R^k \\ 0 & A \end{bmatrix},$$

где $G^k\cdot A^t = \partial P/\partial V_n^k$; $R^k = \partial P/\partial I_b^k$. Как видим, структура Якобиана совпадает со структурой матрицы коэффициентов модифицированной табличной системы уравнений. Более того, в случае линейных цепей, воспользовавшись компонентным уравнением (9.34), вместо (9.32), после подстановки в него (9.31), получим

$$Y_b\cdot A^t\cdot V_n + Z_b\cdot I_b = W_b, \tag{9.37}$$

вместо уравнения (9.35), и, дифференцируя полученную систему (9.37) и (9.36), убеждаемся, что Якобиан линейной цепи вырождается в матрицу коэффициентов модифицированной табличной системы уравнений.

Таким образом, система уравнений Ньютона–Рафсона, как математическая модель нелинейной цепи, может быть сформирована модифицированным табличным методом.

Модифицированный узловой метод. В модифицированном узловом методе, объединяющем достоинства узлового и табличного методов, как известно, все ветви цепи делят на три группы:

1) группа ветвей представимых проводимостью, причем токи этих ветвей не определяются в результате решения;
2) группа ветвей, не представимых проводимостью и, либо представимых проводимостью, но необходимо определить токи этих ветвей;
3) группа ветвей независимых источников тока вносимых в вектор узловых токов.

В обобщенном виде, узловые уравнения ветвей первой и третьей групп и компонентные уравнения ветвей второй группы, можно записать как

$$P_1(V_{n1}) - A_2 \cdot I_{b2} = J_{n1}, \tag{9.38}$$

$$P_2(V_{b2}, I_{b2}) = W_{b2}, \tag{9.39}$$

где $P_1(V_{n1}) = A_1 \cdot g(V_{b1}) \cdot A_1^t$; $J_{n1} = -A_J \cdot I_J$; $V_{b2} = A_2^t \cdot V_{n1}$.

Это позволяет записать вектор функций правой части системы уравнений Ньютона–Рафсона в виде

$$F(X) = \begin{bmatrix} P_1(V_{n1}) - A_2 \cdot I_{b2} - J_{n1} \\ P_2(A_2^t \cdot V_{n1}, I_{b2}) - W_{b2} \end{bmatrix} = 0,$$

где $X = \begin{bmatrix} V_{n1} & I_{b2} \end{bmatrix}^t$ - вектор неизвестных. Дифференцируя на k - той итерации вектор функций, по компонентам вектора неизвестных, получаем Якобиан системы уравнений Ньютона–Рафсона

$$M(X) = \begin{bmatrix} G_{n1}^k & A_2 \\ G_2^k \cdot A_2^t & R_2^k \end{bmatrix},$$

где $G_{n1}^k = \partial P_1 / \partial V_{n1}$; $G_2^k \cdot A_2^t = \partial P_2 / \partial V_{n1}$; $R_2^k = \partial P_2 / \partial I_{b2}$.

Как видим, структура Якобиана аналогична структуре матрицы коэффициентов модифицированного узлового метода.

Для линейных цепей узловые уравнения ветвей первой и второй групп и компонентные уравнения второй группы, вместо (9.38) и (9.39), как известно, запишутся в виде

$$Y_{n1} \cdot V_{n1} + A_2 \cdot I_{b2} = J_{n1}, \tag{9.40}$$

$$Y_{b2} \cdot V_{b2} + Z_{b2} \cdot I_{b2} = W_{b2}, \tag{9.41}$$

где $V_{n1} = A_1 \cdot Y_{b1} \cdot A_1^t$; $J_{n1} = -A_J \cdot I_J$; $V_{b2} = A_2^t \cdot V_{n1}$.

Дифференцированием этих уравнений можно убедиться, что Якобиан линейной цепи совпадает с обычной матрицей коэффициентов модифицированной узловой системы уравнений.

Таким образом, модифицированный узловой метод можно рассматривать, как метод формирования математической модели нелинейной цепи.

Модифицированный узловой метод с проверкой. Модифицированный узловой метод с проверкой, как известно, отличается от модифицированного узлового метода тем, что с целью снижения порядка, из системы уравнений исключаются те переменные, значения которых заранее известны. Речь идет, в основном, о токах ветвей холостого хода и напряжениях ветвей короткого замыкания, встречающихся в идеальных управляемых источниках. При этом ветви также разбиваются на три группы, но ветви второй группы вносимые в дополнение матрицы проводимости предполагается заносить в соответствии с таблицей. При этом, однако, структура матрицы коэффициентов остается аналогичной структуре матрицы коэффициентов модифицированного узлового метода.

В обобщенном виде узловые уравнения первой и второй групп и компонентные уравнения ветвей второй группы можно записать как

$$P_1(V_{n1}) - A_2' \cdot I_{b2}' = J_{n1}, \tag{9.42}$$

$$P_2(V_{b2}', I_{b2}') = W_{b2}', \tag{9.43}$$

где $P_1(V_{n1}) = A_1 \cdot g(V_{b1}) \cdot A_1^t$; $J_{n1} = -A_J \cdot I_J$; $V_{b2}' = (A_2')^t \cdot V_{n1}$. Это позволяет записать вектор функций правой части системы уравнений Ньютона–Рафсона в виде

$$F(X) = \begin{bmatrix} P_1(V_{n1}) - A_2' \cdot I_{b2}' - J_{n1} \\ P_2((A_2')^t \cdot V_{n1}, I_{b2}') - W_{b2}' \end{bmatrix} = 0,$$

где $X = \begin{bmatrix} V_{n1} & I_{b2}' \end{bmatrix}^t$ - вектор неизвестных. Дифференцируя на k - той итерации вектор функций по компонентам вектора неизвестных, получаем Якобиан системы уравнений Ньютона–Рафсона

$$M(X) = \begin{bmatrix} G_{n1}^k & A_2' \\ G_2^k \cdot (A_2')^t & R_2^k \end{bmatrix},$$

где $G_{n1}^k = \partial P_1 / \partial V_{n1}$; $G_2^k \cdot (A_2')^t = \partial P_2 / \partial V_{n1}$; $R_2^k = \partial P_2 / \partial I_{b2}'$.

Как видим, структура Якобиана аналогична структуре матрицы коэффициентов модифицированного узлового метода с проверкой.

Для линейных цепей узловые уравнения ветвей первой и второй групп и компонентные уравнения ветвей второй группы вместо (9.42) и (9.43), как известно, запишутся в виде

$$Y_{n1} \cdot V_{n1} + A_2' \cdot I_{b2}' = J_{n1}, \tag{9.44}$$

$$Y_{b2}' \cdot V_{b2}' + Z_{b2}' \cdot I_{b2}' = W_{b2}', \tag{9.45}$$

где $Y_{n1} = A_1 \cdot Y_{b1} \cdot A_1^t$; $J_{n1} = -A_J \cdot I_J$; $V_{b2}' = (A_2')^t \cdot V_{n1}$.

Дифференцированием этих уравнений можно убедиться, что Якобиан линейной цепи совпадает с обычной матрицей коэффициентов модифицированной узловой системы с проверкой.

Таким образом, модифицированный узловой метод с проверкой можно рассматривать как метод формирования математической модели нелинейной цепи.

Сходимость в диодно-транзисторных схемах. Применение метода Ньютона–Рафсона к расчету режимов диодно-транзисторных схем может привести к переполнению разрядной сетки ЭВМ. Это связано с тем, что в процессе итерационного поиска решения значения переменных, в частности напряжений, претерпевают значительные отклонения от начального значения. В тоже время вольт-амперные характеристики диодов и транзисторов описываются в основном экспоненциальными зависимостями, небольшие изменения аргументов которых (напряжений на переходах), могут привести весьма к большим значениям функций (токов через переход), что и является основной причиной срыва итерационного процесса и переполнения разрядной сетки ЭВМ.

Поясним этот факт с помощью рисунка 9.5, на котором изображена вольт-амперная характеристика (ВАХ) диода.

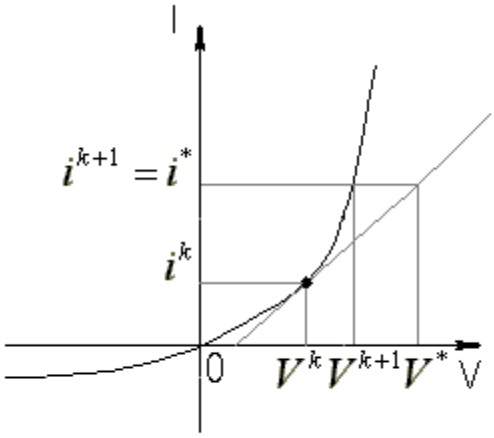


Рисунок 9.5 - ВАХ диода

Пусть ток диода описывается вольтамперной характеристикой диода вида $i_D = I_S \cdot (exp(40 \cdot v_D) - 1)$. Предположим, что на k- той итерации получена точка (v_k, i_k) и алгоритм предсказал на линеаризованной характеристике новую точку (v^*, i^*). Согласно алгоритму, необходимо определить $i^{k+1} = g(v^*)$, что может, в случае экспоненциальной характеристики, повлечь переполнение разрядной сетки ЭВМ. В качестве альтернативного шага определения $i^{(k+1)}$, предлагается брать горизонтальную проекцию на вольт - амперную характеристику точки пересечения вертикальной проекции точки v^*. Тогда $i^{k+1} = i^*$, а инверсия вольтамперной характеристики дает $v^{k+1} = 1/40 \cdot ln(1 + i^*/I_S)$. Горизонтальную проекцию на вольтамперную характеристику можно использовать для всех напряжений на диоде превышающих $0.7\ B$. Напряжение $0.7\ B$ соответствует $exp(40 \cdot 0.7) = 1.4463E + 12\ A$, что вполне представимо в современных ЭВМ.

Другое эвристическое правило рекомендует использовать горизонтальную проекцию, при $v^* - v^k > 0$, а вертикальную, в противном случае.

Трудности могут возникнуть и при отрицательных смещениях на диоде, так как при итерациях вычисляются производные от вольтамперной характеристики $i_D' = 40 \cdot I_S \cdot exp(40 \cdot v_D)$, которые, при $v < 0$, могут оказаться слишком малыми. Во избежание этой ситуации, при отрицательных напряжениях на диоде, предлагается заменить вольтамперную характеристику касательной, при $v_D = (-0.3 \div -0.5)\ B$.

Аналогично, в качестве альтернативного подхода, при прямом смещении на диоде, превышающем $0.7\ B$, можно предложить заменять вольт-амперную характеристику, на ее производную в точке $0.7\ B$.

Подобные способы преодоления переполнения разрядной сетки ЭВМ, в процессе итерационного поиска решения нелинейных систем уравнений, можно рекомендовать и в схемах на биполярных транзисторах, так как вольтамперные характеристики переходов, также имеют экспоненциальную зависимость от напряжений на переходах.

10 РАСЧЕТ ПЕРЕХОДНЫХ ПРОЦЕССОВ ЭЛЕКТРОННЫХ СХЕМ

10.1 Исходные определения

Переходный процесс определяется, как реакция цепи на входное воздействие во времени. В качестве входного воздействия используют как реальные сигналы, так и идеальные тестовые воздействия – единичный скачок или единичную дельта функцию, позволяющие проводить сравнение реакций различных цепей. Так реакция цепи на единичный скачок называется переходной характеристикой, а реакция на единичную дельта функцию носит название импульсной характеристики. Причем переходная и импульсная характеристики определяются при отсутствии других воздействий на схему. Реакцию на реальное воздействие будем называть обобщенно переходным процессом.

Как известно, для определения реакции цепи во времени необходимо иметь математическую модель цепи в виде системы дифференциальных уравнений, описывающих состояние цепи во временной области. Причем лучше всего иметь математическую модель в виде системы дифференциальных уравнений первого порядка в нормальной форме Коши – разрешенных относительно производных. Дело в том, что именно для системы дифференциальных уравнений в нормальной форме Коши разработаны методы аналитического и численного интегрирования, позволяющие находить систему функций, удовлетворяющих дифференциальным уравнениям. Аналитические решения имеют, в основном, лишь линейные системы дифференциальных уравнений. Уравнения с периодическими и нелинейными коэффициентами имеют решения в отдельных частных случаях. Кроме того, аналитическое решение, как правило, представляется функцией от матрицы коэффициентов системы, т.е. требует нахождения корней характеристического уравнения, либо решения полной проблемы собственных значений. Это, как известно, весьма трудоемкая задача, поэтому для их интегрирования используются численные методы, пригодные для решения по единым алгоритмам различных типов уравнений.

В данном разделе будут рассмотрены лишь численные методы интегрирования систем дифференциальных уравнений, пригодные, как для линейных, так и для нелинейных систем дифференциальных уравнений.

В качестве метода формирования математической модели цепи во временной области в виде системы дифференциальных уравнений первого порядка в нормальной форме Коши широко применяется так называемый метод переменных состояния, позволяющий на основе построения дерева графа цепи выбрать систему независимых переменных состояния. Однако метод переменных состояния весьма трудоемок, и существенно усложняется при наличии особенностей, управляемых источников и нелинейностей.

В связи с этим в данном разделе будут рассмотрены известные нам методы формирования математических моделей линейных и нелинейных цепей – табличный, модифицированный табличный, модифицированный узловой и модифицированный узловой с проверкой, которые позволяют управлять представлением ветвей в виде сопротивлений, либо проводимостей.

На этапе формирования математической модели этими методами в виде системы линейных либо нелинейных алгебраических уравнений накладываются ограничения на представление емкостных ветвей в виде проводимостей и индуктивных ветвей в виде сопротивлений. Сформированные таким образом системы алгебраических уравнений позволяют вынести оператор Лапласа перед мнимой частью матрицы коэффициентов системы и, используя формально преобразование Лапласа, перейти от системы алгебраических к системе дифференциальных уравнений, которые интегрируются затем численными методами. Заметим, что линейные системы алгебраических уравнений преобразуются в линейные обыкновенные дифференциальные уравнения, а нелинейные алгебраические уравнения – в нелинейные дифференциальные уравнения.

В случае нелинейных реактивностей в качестве переменных вводят дополнительно заряды на емкостях и магнитные потоки на индуктивностях, которые, в отличие от напряжений на емкостях и токов на индуктивностях, меняются непрерывно. При численном интегрировании систем нелинейных дифференциальных уравнений используют итерационные алгоритмы решения систем нелинейных алгебраических уравнений типа Ньютона–Рафсона.

Численные методы интегрирования основаны на конечно разностном представлении системы дифференциальных уравнений. Широко известны два класса методов – это методы Рунге–Кутта и линейные многошаговые формулы. Методы Рунге–Кутта нашли широкое применение во многих областях науки и техники, однако, при интегрировании дифференциальных уравнений электронных схем используются реже в связи с особой жесткостью этих систем уравнений.

Под жесткостью систем дифференциальных уравнений понимают большой разброс корней характеристических уравнений связанных с постоянными времени цепей и, приводящий к выбору минимального шага интегрирования из соображений устойчивости и, соответственно, увеличению числа шагов и времени счета.

Линейные многошаговые формулы, особенно так называемые формулы дифференцирования назад, обеспечивают большую устойчивость и позволяют проводить интегрирование с большим шагом по времени. Для сокращения времени интегрирования можно использовать численные методы с адаптирующим шагом. Линейные многошаговые методы, сочетающие прямые и обратные формулы интегрирования – “прогноз” и “коррекцию”,

позволяют находить компромисс между быстродействием и точностью при соблюдении устойчивости решения.

Кроме подхода, основанного на формировании и интегрировании системы дифференциальных уравнений, получил распространение подход, основанный на конечно-разностном представлении компонентных уравнений реактивных элементов. Метод, основанный на конечно-разностном представлении дифференциальных соотношений реактивных ветвей, получил название метода конечно-разностных моделей реактивных элементов.

При таком подходе каждый реактивный элемент цепи интерпретируется как сопротивление или проводимость с номиналом, определяемым номиналом реактивности и шагом интегрирования, с включенными, соответственно, последовательно либо параллельно источниками напряжения, либо тока определенной величины и зависящими от предыдущих значений этих переменных на реактивностях. Сложность модели зависит от способа представления производных конечными разностями.

В результате такого подхода, цепи, содержащие реактивные элементы, преобразуются в цепи с резисторами и источниками. По существу, при данном подходе вычисление переходного процесса сводиться к расчету резистивной цепи с источниками. Вычисление переходного процесса представляет процесс итерационного решения алгебраической системы уравнений с подстановкой на каждой итерации текущего решения в качестве предыдущего до тех пор, пока не получим установившееся решение.

Для линейных цепей это сводится к переформированию на каждом шаге итераций вектора правой части, зависящего от предыдущего решения и повторному решению системы уравнений с неизменной матрицей коэффициентов. В этой ситуации предпочтение, как известно, имеют методы основанные на факторизации матрицы коэффициентов системы – методы LU- и QR- факторизации, позволяющие почти вдвое быстрее получить повторное решение.

Данный подход применим и к нелинейным цепям, в том числе и с нелинейными реактивными элементами. При этом на каждом этапе итерации, кроме переформирования вектора правой части системы, переформированию подлежит и матрица коэффициентов, зависящая от переменных, найденных на предыдущей итерации. Для решения таких систем на каждом шаге итерации по времени следует применять итерационные методы типа Ньютона–Рафсона, используемые для решения нелинейных алгебраических систем.

Важно также подчеркнуть, что в данном случае снимаются ограничения на представление реактивных элементов на этапе формирования математической модели цепи. Таким образом, кроме выше перечисленных методов формирования математической модели цепи, можно применять и обобщенный узловой метод, который не позволяет представлять индуктивности в виде сопротивлений и использовать подход основанный на

переходе от алгебраической системы к системе дифференциальных уравнений с последующим интегрированием.

10.2 Простые методы интегрирования

Дифференциальное уравнение в нормальной форме Коши – разрешенное относительно производной можно записать в следующем виде

$$x' = \partial x / \partial t = f(x,t), \qquad (10.1)$$

где $x = x(t)$- функция времени t. Соответствующий этому уравнению интеграл запишется

$$x(t) = x(a) + \int_a^b f(x,t) \cdot \partial t, \qquad (10.2)$$

где $a,\ b$ – нижний и верхний пределы интегрирования.

Задача численного интегрирования дифференциального уравнения на каждом шаге интегрирования сводиться к нахождению отсчета функции $x_{n+1} = x_{n+1}(t_{n+1})$ в момент времени t_{n+1}, при известном значении функции $x_n = x_n(t_n)$ в предыдущий момент времени t_n, и заданном шаге интегрирования $h = \Delta t = t_{n+1} - t_n$. Естественно, что задача интегрирования предполагает задание начального значения функции $x_0 = x(t_0) = x(a)$ в начальный момент времени $t_0 = a$.

Начальные значения в задачах интегрирования называются начальными условиями или, при использовании других переменных интегрирования, граничными условиями и позволяют однозначно определить функцию, удовлетворяющую этим условиям.

При численном методе интегрирования и заданных начальных условиях, найденное значение функции $x_{n+1} = x_{n+1}(t_{n+1})$ в момент времени t_{n+1}, естественно, будет отличаться от истинного значения. Причем ошибка интегрирования будет зависеть от размера шага и метода интегрирования.

Проиллюстрируем на простейшем рисунке 10.1 вывод простых формул численного интегрирования и происхождение присущих им погрешностей.

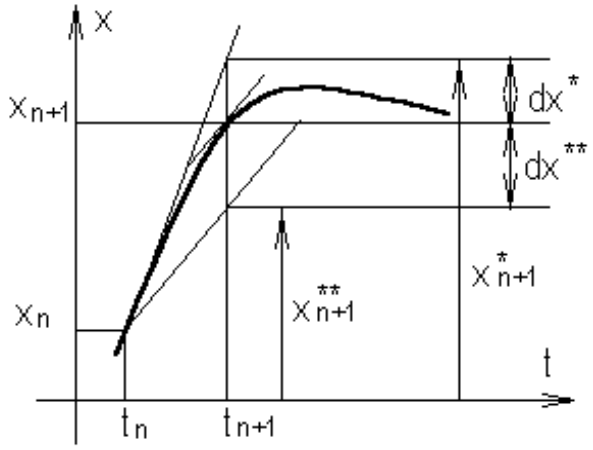


Рисунок 10.1 - Иллюстрация простых формул численного интегрирования

Прямая формула Эйлера. По известному значению функции x_n в момент времени t_n, найдем приближенное значение x_{n+1} в момент времени t_{n+1}, предполагая, что на малом интервале $h = t_{n+1} - t_n$ наклон функции остается неизменным. Тогда в соответствии с рисунком 10.1, учитывая, что $x' = tg\alpha$ можем записать

$$x_{n+1} = x_n + h \cdot x'_n. \qquad (10.3)$$

Это выражение известно, как прямая формула Эйлера. Здесь значение функции в следующей точке вычисляется через значение функции и ее производную в предыдущей точке. Как видно из рисунка 10.1, ошибка вычисления тем больше, чем больше шаг интегрирования.

Обратная формула Эйлера. Можно попытаться выразить значение функции в следующей точке через ее значение в предыдущей точке и значение производной в искомой точке. Как и прежде, полагаем, что наклон функции на интервале h остается неизменным. Тогда в соответствии с рисунком 10.1, учитывая, что $x' = tg\alpha$ можем записать

$$x_{n+1} = x_n + h \cdot x'_{n+1}. \qquad (10.4)$$

Это выражение известно как обратная формула Эйлера. Обратим внимание, что здесь значение функции x_{n+1} в точке t_{n+1} входит в правую и левую части выражения (10.4), поскольку $x'_{n+1} = f(x_{n+1}, t_{n+1})$. Трансцендентный характер формулы (10.4) требует для вычислений итерационные методы расчета, причем значение в предыдущей точке должно быть известно, а приближенное значение функции в искомой точке предсказано, например, по прямой формуле Эйлера. Тогда в результате итераций, по известному приближенному значению функции в искомой точке x_{n+1} и выражению для производной $x'_{n+1} = f(x_{n+1}, t_{n+1})$, находиться ее значение с заданной точностью.

Обратную формулу Эйлера можно интерпретировать как выражение значения функции в предыдущей точке через неизвестные значения функции и ее производной, в последующей точке

$$x_n = x_{n+1} - h \cdot x'_{n+1}.$$

Прямую и обратную формулы Эйлера можно записать также в виде

$$(x_{n+1} - x_n)/h = x'_n,$$

$$(x_{n+1} - x_n)/h = x'_{n+1}.$$

Из этих выражений следует, что приращение искомой функции на шаге интегрирования определяется значением производной этой функции в начале либо в конце интервала $h = t_{n+1} - t_n$. Такая интерпретация, прямой и обратной формул Эйлера, подводит к выводу о возможности получения других конечно –разностных формул интегрирования.

Формула трапеций. Так, определив приращение искомой функции на шаге интегрирования линейной комбинацией производных x'_n и x'_{n+1}

$$(x_{n+1} - x_n)/h = b_0 \cdot x'_n + b_1 \cdot x'_{n+1},$$

и полагая $b_0 = b_1 = 1/2$, получаем известную формулу трапеций

$$x_{n+1} = x_n + 0.5 \cdot h \cdot x'_n + 0.5 \cdot h \cdot x'_{n+1}. \qquad (10.5)$$

Формула трапеций является простейшей многошаговой формулой, когда для определения значения функции в текущей точке используются значения функции и ее производных в предыдущих точках. Подробнее на многошаговых формулах интегрирования остановимся позже, а сейчас рассмотрим вопросы, касающиеся точности и устойчивости алгоритмов численного интегрирования.

Иллюстрацию алгоритмов численного интегрирования проведем на примере простого линейного неоднородного дифференциального уравнения $x' = x + t^2$. Определим начальное значение $x_0 = 1$, при $t = 0$ и шаге $h = 0.025$. Точное решение этого уравнения имеет вид $x = 3 \cdot e^t - t^2 - 2 \cdot t - 2$, что позволит оценить точность численных методов интегрирования.

Вначале воспользуемся для интегрирования прямой формулой Эйлера (10.3). В начальный момент времени $t_0 = 0$, $x_0 = 1$ и $x'_0 = x_0 + t^2 = 1$. В следующий момент времени $t_1 = 0.025$,

$$x_1 = x_0 + h \cdot x'_0 = 1 + 0.025 \cdot 1 = 1.025$$

и

$$x'_1 = x_1 + t_1^2 = 1.025 + 0.000625 = 1.025625.$$

Результаты численного интегрирования на ЭВМ, включая ошибку ε, сведены в таблицу 10.1.

Таблица 10.1

Результаты интегрирования по прямой формуле Эйлера

t	x	ε
0.000	1.0000000	0.0000000
0.025	1.0250000	-0.0003204
0.050	1.0506406	-0.0006727
0.075	1.0766991	-0.0010583
0.100	1.1040340	-0.0014788
0.125	1.1318848	-0.0019355
0.150	1.1605726	-0.0024301
0.175	1.1901494	-0.0029642
0.200	1.2206688	-0.0035395

В обратной формуле Эйлера и формуле трапеций для вычисления значения функции в текущей точке используются значения производных в искомой точке. Для нахождения решения в данной ситуации используют итерационные методы, которые требуют начальное приближение значения функции.

Метод прогноза–коррекции. В качестве начального значения можно использовать либо значение функции в предыдущей точке, либо определить ее по прямой формуле Эйлера. Таким образом, мы приходим к понятию **прогноза**, когда предсказывается значение искомой функции в следующей точке.

На основе предсказания значения функции по исходному соотношению $x'_{n+1} = f(x_{n+1}, t_{n+1})$ можно определить грубое значение производной в этой точке и уточнить – скорректировать, это значение, используя обратную формулу Эйлера или формулу трапеций. Таким образом, мы приходим к понятию **коррекция**. По уточненному значению функции можно найти уточненное значение производной и уточнить значение функции еще раз т.д. Уточнение продолжают до тех пор, пока значение искомой функции не установиться с заданной точностью. После этого по прямой формуле Эйлера предсказывается значение функции в следующей точке и вновь уточняется и т.д.

Описанная методика численного интегрирования, когда для предсказания используются прямые формулы, а для коррекции обратные получила название **метода прогноза–коррекции**. Итерации уточнения позволяют фиксировать ошибку и адаптировать шаг численного интегрирования.

Проиллюстрируем метод прогноза–коррекции, используя линейное неоднородное дифференциальное уравнение из предыдущего примера. Для прогноза воспользуемся прямой формулой Эйлера, а для коррекции – обратной, причем уточнение будем производить в три этапа.

При $t=0$, $x_0=1$, $h=0.025$, и $x_0'=1$, предсказанное по прямой формуле Эйлера значение x_1^p, составляет $x_1^p = x_0 + h\cdot x_0' = 1 + 0.025\cdot 1 = 1.025$. Первая итерация, в соответствии с заданным уравнением и обратной формулой Эйлера, дает

$$x_1'^{(0)} = x_1^p + t_1^2 = 1.025 + 0.000625 = 1{,}025625;$$

$$x_1^{(1)} = x_0 + h\cdot x_1'^{(0)} = 1.02564063.$$

Вторая итерация:

$$x_1'^{(1)} = x_1^{(1)} + t_1^2 = 1.02626563;\ x_1^{(2)} = x_0 + h\cdot x_1'^{(1)} = 1.02565664.$$

Третья итерация:

$$x_1'^{(2)} = x_1^{(2)} + t_1^2 = 1.02628164;\ x_1^{(3)} = x_0 + h\cdot x_1'^{(2)} = 1.02565704.$$

Результаты численного интегрирования на ЭВМ, включая ошибку интегрирования ε, сведены в таблицу 10.2

Таблица10.2

Результаты интегрирования методом прогноз-коррекция.

t	x^p	x^c	ε
0.000	1.0000000	1.0256410	0.0000000
		1.0256567	
		1.0256570	
0.025	1.0256570	1.0250036	0.0003367
		1.0520196	
		1.0520200	
0.050	1.0520200	1.0791222	0.0007067
		1.0791387	
		1.0791391	
0.075	1.0791391	1.1070483	0.0011117
		1.1070653	
		1.1070658	
0.100	1.1070658		0.0015530

Сравнение ошибок интегрирования по прямой формуле Эйлера и методу прогноз–коррекция на основе прямой и обратной формулы Эйлера показывает, что для прямой формулы Эйлера ошибки отрицательны, тогда как для обратной формулы Эйлера они положительны. По интуиции комбинация этих двух формул, например, формула трапеций, может дать меньшую ошибку.

Проиллюстрируем результат интегрирования по формуле трапеций того же дифференциального уравнения, используя для предсказания прямую формулу Эйлера (таблица 10.3).

Таблица 10.3

Результаты интегрирования методом трапеций

t	x^p	x^c	ε
0.000	1.0000000	1.02532031	0.0000000
		1.02532432	
		1.02532437	
0.025	1.02532437	1.05131715	0.0000040
			0
		1.05132145	
		1.05132150	
0.050	1.05132150	1.07803542	0.0000082
			1
		1.07804002	
		1.07804008	
0.075	1.07804008	1.10552504	0.0000126
			3
		1.10552996	
		1.10553002	
0.100	1.10553002		0.0000172
			6

Ошибка действительно уменьшилась, что будет обсуждено позже при изложении ошибок усечения.

Полученные формулы численного интегрирования пригодны как для линейных, так и для нелинейных дифференциальных уравнений.

Получим выражения этих формул интегрирования для систем линейных дифференциальных уравнений.

Пусть имеем систему линейных дифференциальных уравнений в нормальной форме Коши и векторно-матричном представлении

$$X' = A \cdot X + W, \qquad (10.6)$$

где X - вектор искомых функций времени t; A – матрица коэффициентов системы; W – вектор воздействий, известных функций времени.

Применив прямую формулу Эйлера (10.3), к системе (10.6), в конечно-разностном представлении, получим

$$X_{n+1} = X_n + h \cdot X'_n = X_n + h \cdot (A \cdot X_n + W_n).$$

Откуда, приводя подобные, можем записать матричную форму прямой формулы

Эйлера

$$X_{n+1} = (1 + h \cdot A) \cdot X_n + h \cdot W_n. \qquad (10.7)$$

Формула (10.7) имеет итерационный характер, т.е. по известному на предыдущем шаге вектору неизвестных, находим его текущее значение, затем, подставляя в правую часть, находим следующее значение и т.д.

Используя обратную формулу Эйлера (10.4), получаем

$$X_{n+1} = X_n + h \cdot X'_{n+1} = X_n + h \cdot (A \cdot X_{n+1} + W_{n+1}).$$

Приведя подобные, запишем матричную форму обратной формулы Эйлера

$$(1 - h \cdot A) \cdot X_{n+1} = X_n + h \cdot W_{n+1}. \qquad (10.8)$$

Как видим в случае линейных систем дифференциальных уравнений, трансцендентность обратной формулы Эйлера исчезла, хотя, как и предыдущая формула (10.7), имеет итерационный характер. Решая на каждом шаге систему (10.8) находим очередное значение вектора неизвестных, затем, подставляя его в правую часть, находим новое значение и т.д.

Для нахождения решения системы (10.8) лучше воспользуемся методами, основанными на LU- или QR- факторизации исходной матрицы коэффициентов $(1 - h \cdot A)$. При этом, выполнив один раз разложение исходной матрицы на сомножители и, зафиксировав их, при повторных итерациях, будем решать факторизованные системы с новым вектором свободных членов, экономя время на повторных факторизациях.

В случае формулы трапеций матричное уравнение имеет вид

$$X_{n+1} = X_n + 0.5 \cdot h \cdot (A \cdot X_n + W_n + A \cdot X_{n+1} + W_{n+1}).$$

Перегруппировывая компоненты, получим уравнение

$$(1 - 0.5 \cdot h \cdot A) \cdot X_{n+1} = (1 + 0.5 \cdot h \cdot A) \cdot X_n + 0.5 \cdot h \cdot (W_n + W_{n+1}), \qquad (10.9)$$

которое решается аналогично (10.8), т.е. путем решения системы при заданных начальных условиях X_n, подстановки найденного решения X_{n+1} в правую часть на место X_n, нахождения нового решения и т.д.

10.3 Порядок метода интегрирования и ошибки усечения

Под порядком метода интегрирования, в общем случае, будем понимать число предыдущих отсчетов функции, используемых при вычислении текущего значения функции. Ошибка, заложенная при выводе формулы интегрирования соответствующего порядка за счет отбрасывания старших членов разложения функции в ряд Тейлора, называется ошибкой усечения.

Как уже отмечалось, в формулу трапеций значения производных в предыдущей и текущей точках входят с одинаковыми весами. Очевидно, что эти производные можно взять с разными весами, точно также, можно поступить и со значениями функций. В общем случае, формулу интегрирования можно записать

$$a_0 \cdot x_0 + a_1 \cdot x_1 - h \cdot (b_0 \cdot x'_0 + b_1 \cdot x'_1) = 0. \qquad (10.10)$$

При изложении вопроса ограничимся случаем, когда значение искомой функции x_1, в точке $t_1 = t_0 + h$, определяется через ее значение в предыдущей точке и значения ее производных в предыдущей и текущей точках. Значение функции в текущей точке можно в принципе определить через значения функции и производных в текущей и нескольких предыдущих точках, что

будет соответствовать так называемым многошаговым формулам интегрирования.

Используя выражение (10.10), при соответствующих коэффициентах $a_{0,1}$ и $b_{0.1}$, можно получить все три раннее рассмотренные формулы интегрирования. Рассмотрим некоторые свойства этих формул, переписав соотношение (10.10), в виде

$$a_0 \cdot x(t_0) + a_1 \cdot x(t_0 + h) - h \cdot (b_0 \cdot x'(t_0) + b_1 \cdot x'(t_0 + h) = 0.$$

Разложим функции $x(t_0 + h)$ и $x'(t_0 + h)$ в ряды Тейлора

$$a_0 \cdot x(t_0) + a_1 \cdot [x(t_0) + \frac{h}{1!} \cdot x'(t_0) + \frac{h^2}{2!} \cdot x''(t_0) + \frac{h^3}{3!} \cdot x'''(t_0) + ...] -$$

$$- h \cdot b_0 \cdot x'(t_0) - h \cdot b_1 \cdot [x'(t_0) + \frac{h}{1!} \cdot x''(t_0) + \frac{h^2}{2!} \cdot x'''(t_0) + \cdots = 0.$$

Перенесем слагаемые второго и более высоких порядков в правую часть и запишем

$$[a_0 + a_1] \cdot x(t_0) + [a_1 - b_0 - b_1] \cdot h \cdot x'(t_0) =$$

$$= -\left[\frac{a_1}{2!} - b_1\right] \cdot h^2 \cdot x''(t_0) - \left[\frac{a_1}{3!} - \frac{b_1}{2!}\right] \cdot h^3 \cdot x'''(t_0) - \cdots.$$

Это выражение будет удовлетворяться тождественно, при любых значениях $x(t)$ и ее производных в точке $t = t_0$, в случае обращения в нуль сомножителей в квадратных скобках. Это приводит к следующим равенствам:

1) для левой части - $a_0 + a_1 = 0$; $a_1 - b_0 - b_1$;

2) для правой части - $\frac{a_1}{2!} - b_1 = 0$; $\frac{a_1}{3!} - \frac{b_1}{2!} = 0$.

Из анализа следует:

1. Выбором четырех различных коэффициентов $a_{0,1}$, b_{01}, нельзя обратить в нуль все сомножители в квадратных скобках.

2. Наибольшее число первых сомножителей, которое может быть обращено в нуль, равно трем и это соответствует следующему выбору коэффициентов – $a_0 = -1$; $a_1 = 1$; $b_0 = b_1 = 0.5$. Подстановка их в исходное уравнение (10.10) приводит к формуле трапеций. Из уравнения для коэффициентов находим, что множители находящиеся перед производными, начиная с третьего порядка и выше, не равны нулю. Обозначив выражения в квадратных скобках через c_n, где n- порядок производной, получим значение первого множителя, отличного от нуля - $c_3 = -0.5$.

3. Если взять набор коэффициентов $a_{0,1}$, $b_{0,1}$, приводящий к формуле трапеций и принять, что искомая функция $x(t)$ является полиномом второй степени, то уравнение разложенное в ряд будет удовлетворяться точно, поскольку в этом случае все производные, начиная с третьей, равны нулю.

Из сказанного следует, что в формуле трапеций искомая функция аппроксимирована полиномом второй степени и говорят, что метод численного интегрирования, опирающийся на метод трапеций, имеет порядок $p = 2$. Погрешность описания функции $x(t)$ является полиномом второй степени и определяется отброшенными членами ряда Тейлора, содержащими производные третьего и более высоких порядков.

В связи с этим первый, не равный нулю сомножитель в уравнении разложенным в ряд Тейлора обозначают c_{p+1} и называют ошибкой усечения. Для формулы трапеций: $c_{p+1} = c_3 = -0.5$. Определив значения коэффициентов $a_{0,1}$, $b_{0,1}$, для прямой и обратной формул Эйлера, найдем, что порядок интегрирования у этих формул равен 1, а ошибка усечения c_2 составляет, соответственно, -0.5 и 0.5. Вспомнив, что в рассматриваемом примере, формула трапеций дала меньшую погрешность и это, как видим, связано с более высоким порядком метода и соответственно с меньшей ошибкой усечения. Более высокое значение порядка p и меньшая ошибка усечения c_{p+1} предпочтительны при оценке формул численного интегрирования.

10.4. Устойчивость методов интегрирования

Для характеристики метода численного интегрирования недостаточно знать его порядок и ошибку усечения. Важно еще одно свойство - устойчивость метода. Устойчивость метода численного интегрирования характеризует поведение ошибки интегрирования во времени. Нарастание ошибки интегрирования свидетельствует о неустойчивости метода. Устойчивость метода интегрирования, как будет изложено далее, зависит от шага интегрирования и значения корней характеристического уравнения. Поскольку точность также зависит от шага интегрирования, выбор размера шага часто является результатом компромисса таких характеристик, как точность и устойчивость метода интегрирования.

Изложим проблему устойчивости на примере простейшего дифференциального уравнения

$$x' = \lambda \cdot x, \qquad (10.11)$$

имеющего аналитическое выражение решения, в виде

$$x = x_0 \cdot exp(\lambda \cdot t),$$

где λ - корень характеристического уравнения, действительная, либо комплексная константа. Отметим также, что отклик линейной цепи на единичный скачок на входе, как результат решения системы дифференциальных уравнений в случае простых полюсов описывается соотношением вида

$$x(t) = \sum A_i \cdot exp(\lambda_i \cdot t),$$

где λ_i - корни характеристического уравнения. Таким образом, решение простейшего уравнения соответствует одной компоненте этого отклика, а результаты его исследования можно экстраполировать на общий случай.

Устойчивость прямой формулы Эйлера. Исследование устойчивости численных методов начнем с прямой формулы Эйлера

$$x_1 = x_0 + h \cdot x_0'.$$

Используя (10.11) и, подставляя в формулу, вместо x_0', его значение $\lambda \cdot x_0$, получим

$$x_1 = (1 + \lambda \cdot h) \cdot x_0.$$

На следующем шаге интегрирования, соответственно, получим

$$x_2 = x_1 + h \cdot x_1' = (1 + \lambda \cdot h) \cdot x_1 = (1 + \lambda \cdot h)^2 \cdot x_0.$$

Продолжив процедуру интегрирования в пределах n шагов, получим

$$x_n = (1 + \lambda \cdot h)^n \cdot x_0.$$

Предположим, что время интегрирования $t \Rightarrow \infty$ и, соответственно, $n \Rightarrow \infty$, тогда, для того чтобы x_n было ограниченным, для устойчивого дифференциального уравнения, когда $Re\lambda < 0$, необходимо выполнение условия

$$|1 + \lambda \cdot h| \le 1,$$

где h - размер шага (действительное число); λ - корень характеристического уравнения (возможно комплексный).

Найдем области устойчивости, удовлетворяющие неравенству. Введя обозначение

$$\lambda \cdot h = q = u + j \cdot v,$$

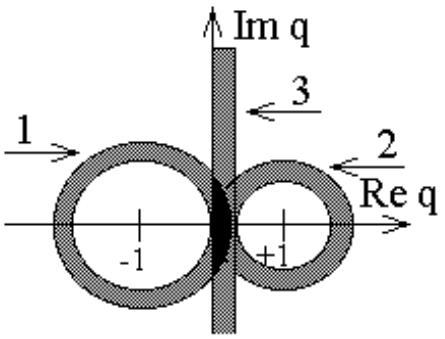
и подставив его в неравенство, получим

$$|1 + u + j \cdot v| \le 1$$

или

$$(1 + u)^2 + v^2 \le 1.$$

Это выражение описывает область устойчивости решения внутри единичного круга с центром в точке $(-1, 0)$, изображенную на рисунке 10.2.



1) прямая формула Эйлера, устойчивые решения внутри единичной окружности
2) обратная формула Эйлера, устойчивые решения снаружи единичной окружности
3) формула трапеций, устойчивые решения в левой полуплоскости

Рисунок 10.2 - Области устойчивости формул интегрирования

Полученный результат используется следующим образом. При $Re\lambda < 0$, шаг выбирается, исходя из условия $q = \lambda \cdot h < 1$, что соответствует точке внутри единичной окружности. В этом случае прямая формула Эйлера дает устойчивые решения, т.е. значение функции $x(t)$ при увеличении n будет оставаться конечной величиной. Величина шага определяет и устойчивость, и точность численного интегрирования и из соображений устойчивости может оказаться меньше, чем того требует точность. Так, при больших $|\lambda|$, шаг выбирается малым для обеспечения устойчивости.

$Re\lambda > 0$ соответствует физически неустойчивой схеме, однако при выборе малого шага и удовлетворения условия устойчивости решений будем получать конечное решение не соответствующее реальному поведению схемы.

В качестве иллюстрации, применим прямую формулу Эйлера к простейшему дифференциальному уравнению вида $x' = -x,$ при $x_0 = 1,$ соответствующему уравнению вида (10.11), при $\lambda = -1$. Результаты численного интегрирования, при разных h, сведены в таблицу 10.4.

Таблица 10.4

Результаты интегрирования уравнения $x' = -x,$ при различных h, по прямой формуле Эйлера.

Шаг	Значения	при	
n	h=0.1	h=2	h=3
0	1	1	1
1	0.9	-1	-2
2	0.81	1	4
3	0.729	-1	-8
4	0.65.61	1	16
5	0.59049	-1	-32
6	0.53144	1	64

Первая колонка решений, при $h = 0.1$, соответствует $q = \lambda \cdot h = -0.1$, т.е. области внутри единичной окружности и устойчивому решению. Вторая колонка соответствует решению, при $h = 2$, $q = -2$, т.е. на границе области устойчивости. Решение в этом случае осцилирует, но не нарастает. В третьей колонке решение, при $h = 3$, $q = -3$, находится вне области устойчивости. Решение нарастает и осцилирует, хотя, в соответствии с аналитическим решением равным $exp(-t)$, должно уменьшаться.

Устойчивость обратной формулы Эйлера. Продолжим исследование проблемы устойчивости для обратной формулы Эйлера

$$x_1 = x_0 + h \cdot x_1'.$$

Используя (10.11), и подставляя в формулу, вместо x_1', его значение $\lambda \cdot x_1$, получим

$$x_1 = x_0 + h \cdot \lambda \cdot x_1,$$

или

$$x_1 = x_0 /(1 - \lambda \cdot h) = x_0 /(1 - q).$$

На следующем шаге интегрирования, соответственно получим

$$x_2 = x_1 + h \cdot x_2' = x_1 /(1 - \lambda \cdot h) = x_0 /(1 - \lambda \cdot h)^2.$$

Продолжив процедуру интегрирования в пределах n шагов, получим

$$x_n = x_0 /(1 - \lambda \cdot h)^n = x_0 /(1 - q)^n.$$

Для устойчивого решения, устойчивого дифференциального уравнения, когда $Re\,\lambda < 0$, при $t \Rightarrow \infty$ и, соответственно, при $n \Rightarrow \infty$, необходимо выполнение условия

$$|1/(1 - q)| \le 1.$$

Найдем области устойчивости, удовлетворяющие неравенству. Раскрывая обозначение q, получаем

$$1 \le (1 - u)^2 + v^2.$$

Этому выражению при равенстве соответствует окружность единичного радиуса с центром в точке $(1,0)$, изображенной на рисунке 10.2. Неравенство удовлетворяется вне окружности. Таким образом, обратная формула Эйлера устойчива для устойчивых схем, когда $Re\,\lambda < 0$ при любом шаге h. Размер шага при этом должен выбираться лишь из соображений точности интегрирования.

При $Re\,\lambda > 0$, схема неустойчива, однако при шаге h, удовлетворяющем неравенству, получим устойчивое решение, хотя реальный отклик безгранично растет.

Устойчивость формулы трапеций. Рассмотрим, в заключение, формулу трапеций

$$x_1 = x_0 + 0.5 \cdot h \cdot (x_0' + x_1').$$

В соответствии с (10.11) подставив значения $x_0' = \lambda \cdot x_0$ и $x_1' = \lambda \cdot x_1$, получим

$$x_1 = x_0 + 0.5 \cdot h \cdot \lambda \cdot (x_0 + x_1),$$

или

$$x_1 = x_0 \cdot (1 + 0.5 \cdot h \cdot \lambda)/(1 - 0.5 \cdot h \cdot \lambda).$$

Соответственно для n - го шага, получим

$$x_n = [(2 + q)/(2 - q)]^n \cdot x_0.$$

В предельном случае, при $n \Rightarrow \infty$, условие устойчивости имеет вид

$$|(2 + q)/(2 - q)| \le 1|,$$

или

$$|(2 + u + j \cdot v)/(2 - u - j \cdot v)| \le 1.$$

Преобразуя это условие, получим его в виде $4 \cdot u \le 0$, показывающем, что границей устойчивости в этом случае является мнимая ось, а областью устойчивости - левая полуплоскость рисунок 10.2. Таким образом, формула трапеций устойчива для устойчивых схем, когда $Re\,\lambda < 0$, при любом шаге интегрирования h.

Обратим внимание на то, что, по сравнению с обратной формулой Эйлера, область устойчивости уменьшилась, а также на то, что метод интегрирования, основанный на формуле трапеций, имеет наибольший порядок из методов, использующих значения функции и ее производных в предыдущей и текущей точках.

Кроме того, формула трапеций дает устойчивый отклик для устойчивых цепей и нестабильный отклик для неустойчивых цепей. Это ценное свойство метода интегрирования, когда поведение цепи заранее не известно.

Следует также осознавать, что выполнение условий устойчивости численного метода не подразумевает правильности расчетов. Это лишь означает, что любая ошибка не увеличивается при последующих шагах. На величину ошибки, как уже отмечалось, влияют остаточные члены ряда Тейлора, которыми пренебрегают

$$\sum_{p+1}^{\infty} h_i \cdot c_i \cdot x^{(i)}(t_0),$$

так называемая ошибка усечения, пропорциональная шагу интегрирования h .

Одним из путей обеспечения точности при выбранном шаге h, является использование формулы с наивысшим порядком p. Можно так же показать, что увеличение порядка метода интегрирования может сопровождаться уменьшением устойчивости, что мы и отмечали для формулы трапеций по сравнению с обратной формулой Эйлера.

С другой стороны, при выбранном методе интегрирования необходимо выбрать такой шаг интегрирования h, чтобы выполнялось условие устойчивости, и обеспечивалась требуемая точность. В первую очередь, конечно необходимо, обеспечить условие устойчивости, иначе бессмысленно выполнять вычисления. При этом может оказаться размер шага h настолько малым, что потребуется значительное число шагов интегрирования. В частности, это наиболее вероятно для прямых методов интегрирования и менее вероятно для обратных методов интегрирования.

Следует, наконец, отметить, что полученные нами для простых методов интегрирования границы областей устойчивости справедливы лишь для рассматриваемого дифференциального уравнения вида (10.11). Для других дифференциальных уравнений и других методов интегрирования границы областей будут другими. Исследование границ устойчивости, для простых методов интегрирования и частного вида дифференциального уравнения, предпринято с целью, обозначить основные тенденции и дать сравнительную характеристику методов.

Дадим неформальное объяснение, некоторым часто используемым терминам и понятиям.

Формулы интегрирования, основанные на значениях функции и ее производных в предыдущие моменты времени, подобные прямой формуле Эйлера называют иногда явными. Явные формулы, в силу их слабой устойчивости используются главным образом для предсказания начальных значений при использовании других, неявных формул интегрирования.

Неявными формулами интегрирования, аналогичными обратной формуле Эйлера и формуле трапеций, называются формулы, связывающие значение

функции в следующей точке с ее производной в этой точке, а также значениями функции и ее производными в предыдущих точках. Неявные формулы, как правило, более устойчивые, используются, в основном, для коррекции решения, и в силу трансцендентного характера, когда неизвестное значение находится в правой и левой частях уравнения, решаются итерационными методами, используемыми при решении нелинейных алгебраических уравнений.

Объединение явных и неявных формул интегрирования приводит, как уже отмечалось, к методу прогноз-коррекция. При этом за счет итераций уточнения решения неявных формул, требуется большее число операций, однако, большая устойчивость неявных формул может позволить увеличить шаг интегрирования, в пределах обеспечения точности, и, следовательно, позволяет наоборот уменьшить число требуемых операций.

Формулы интегрирования, использующие для нахождения значения функции либо производной, в текущий момент времени, значения функции и ее производных в нескольких предшествующих моментах времени, носят название линейных многошаговых методов интегрирования (ЛММ).

Линейные многошаговые формулы, интегрирования, содержащие значение функции и ее производной в искомый момент времени, и произвольное число значений функции без значений производных, в предшествующие моменты времени**, называются формулами дифференцирования назад (ФДН).** Формулы дифференцирования назад обладают повышенной устойчивостью решений и зачастую используются в качестве основных алгоритмов численного интегрирования дифференциальных уравнений.

Формулу интегрирования называют A **- устойчивой,** если она дает ограниченное решение тестового дифференциального уравнения $x' = \lambda \cdot x$, для произвольных размеров шага и любого числа шагов, при $Re\,\lambda < 0$. Обратная формула Эйлера и формула трапеций обладают этим свойством.

Область абсолютной устойчивости, какой-либо формулы интегрирования, это часть плоскости $q = h \cdot \lambda$, в которой интегрирование дифференциального уравнения $x' = \lambda \cdot x$, при $Re\,\lambda < 0$, дает ограниченный результат при любом числе шагов. Так прямая формула Эйлера абсолютно устойчива внутри единичной окружности с единичным радиусом и центром в точке $(-1,0)$. Обратная формула Эйлера и формула трапеций абсолютно устойчивы во всей левой полуплоскости.

Большинство цепей встречающихся на практике, являются устойчивыми. Для линейной цепи это означает, что действительные части

корней характеристических уравнений отрицательны. В связи с этим, говоря об интегрировании уравнения $x' = \lambda \cdot x$, имеем в виду решение, при $Re\,\lambda < 0$.

Однако встречаются ситуации, например автоколебательные системы, когда схема в рабочей точке неустойчива и в установившемся режиме в ней наблюдаются периодические колебания. При численном интегрировании дифференциальных уравнений таких цепей следует применять методы, учитывающие эти свойства. Обратная формула Эйлера, например, в этой ситуации не подходит, т.к. дает затухающие решения. В этом случае предпочтительны, формула трапеций и другие методы интегрировании.

Жесткая система дифференциальных уравнений, это система имеющая несколько полюсов вблизи начала координат и часть полюсов весьма удаленных, причем все полюсы расположены в левой полуплоскости. Такая система устойчива, а компоненты решения, соответствующие далеким полюсам быстро затухают.

При интегрировании такой системы по прямой формуле Эйлера для обеспечения устойчивости потребуется очень малый шаг, чтобы $q = h \cdot \lambda$ попала для удаленных полюсов в область устойчивости. Выбор малого шага потребует огромного числа операций, хотя компоненты, обусловленные далекими полюсами, быстро затухают и сказываются лишь на начальном этапе.

С другой стороны, обратная формула Эйлера не вызывает таких проблем, т.к. она устойчива при любых h. Для обеспечения точности в обратной формуле Эйлера можно начать с малого значения шага h и перейти на большие значения, как только быстрые компоненты затухнут.

Для интегрирования жестких систем дифференциальных уравнений разработаны и другие эффективные методы, не требующие малого шага. Для этих целей широко применяются линейные многошаговые формулы, имеющие повышенную точность и специфическую форму границ областей устойчивости, отражающих особенности жестких систем. Для повышения эффективности методов интегрирования - точности и устойчивости используют адаптивные формы алгоритмов интегрирования, при которых в процессе интегрирования меняются шаг и порядок метода.

10.5 Расчет переходных процессов цепей

Как уже отмечалось, рассмотренные нами формулы численного интегрирования применимы как к линейным, так и нелинейным дифференциальным уравнениям.

Уравнения переменных состояния. Если цепь описана линейной системой дифференциальных уравнений в нормальной форме Коши, относительно переменных состояния

$$x'(t) = A \cdot x(t) + B \cdot w(t), \qquad (10.12)$$

то вычисление $x(t)$ не вызывает проблем. Так вычисление переменных состояния может быть осуществлено на основании полученных нами выражений в матричной форме (10.7 - 10.9).

Непосредственно метод переменных состояния нами не рассматривался ввиду того, что этот метод в настоящее время используется редко из-за сложности алгоритма формирования системы дифференциальных уравнений в нормальной форме Коши для произвольного вида цепей. Заметим только, что метод переменных состояния достаточно широко освещен в литературе. В качестве переменных состояния выступают независимые токи в индуктивностях и напряжения на емкостях. Для формирования системы дифференциальных уравнений в нормальной форме Коши в методе переменных состояния используются топологические уравнения, описывающие дерево графа через матрицу главных сечений и компонентные уравнения ветвей цепи.

Однако уравнения состояния, вида (10.12) для частного вида цепей, можно получить и на основе других методов. В качестве примера формирования и решения уравнений переменных состояния рассмотрим уравнения RC- цепи (рисунок 10.3), полученные обобщенным методом узловых потенциалов.

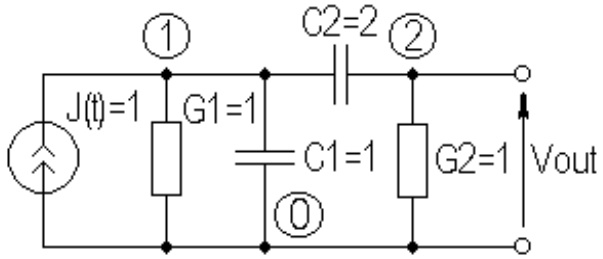


Рисунок 10.3 - Простая RC - цепь

Найдем временной отклик цепи на единичный скачок тока. Примем начальные напряжения на конденсаторах нулевыми, а шаг интегрирования $h = 0.05$. Так как в этой цепи нет индуктивностей, и узловые напряжения определяют переменные состояния, т.е. напряжения на емкостях, то для формирования уравнений состояния воспользуемся обобщенным узловым методом.

Система узловых уравнений, при заданных номиналах, имеет вид

$$\begin{bmatrix} 1+s\cdot 3 & -s\cdot 2 \\ -s\cdot 2 & 2+s\cdot 2 \end{bmatrix}\cdot\begin{bmatrix} v_1 \\ v_2 \end{bmatrix}=\begin{bmatrix} J(s) \\ 0 \end{bmatrix}.$$

Разделяя матрицу проводимости, на действительную и мнимую части и, вынося, оператор Лапласа s, как общий множитель, получим

$$\begin{bmatrix} 3 & -2 \\ -2 & 2 \end{bmatrix}\cdot\begin{bmatrix} s\cdot v_1 \\ s\cdot v_2 \end{bmatrix}=\begin{bmatrix} -1 & 0 \\ 0 & -2 \end{bmatrix}\cdot\begin{bmatrix} v_1 \\ v_2 \end{bmatrix}+\begin{bmatrix} J(s) \\ 0 \end{bmatrix}.$$

Умножая систему на обратную матрицу мнимой части, получаем

$$\begin{bmatrix} s\cdot v_1 \\ s\cdot v_2 \end{bmatrix}=\begin{bmatrix} -1 & -2 \\ -1 & -3 \end{bmatrix}\cdot\begin{bmatrix} v_1 \\ v_2 \end{bmatrix}+\begin{bmatrix} J(s) \\ J(s) \end{bmatrix}.$$

Применив преобразование Лапласа, получаем систему линейных дифференциальных уравнений с действительными коэффициентами, описывающих состояние цепи

$$\begin{bmatrix} v_1'(t) \\ v_2'(t) \end{bmatrix}=\begin{bmatrix} -1 & -2 \\ -1 & -3 \end{bmatrix}\cdot\begin{bmatrix} v_1(t) \\ v_2(t) \end{bmatrix}+\begin{bmatrix} j(t) \\ j(t) \end{bmatrix},$$

где $j(t)=1$.

В соответствии с обратной формулой Эйлера (10.8) для системы линейных дифференциальных уравнений с действительными коэффициентами

$$(1-h\cdot A)\cdot X_{n+1}=X_n+h\cdot W_{n+1}$$

определяем

$$(1-h\cdot A)=\begin{bmatrix} 1+h & 2\cdot h \\ h & 1+3\cdot h \end{bmatrix}.$$

Соответственно отклик цепи определяется системой уравнений

$$\begin{bmatrix} 1+h & 2\cdot h \\ h & 1+3\cdot h \end{bmatrix}\cdot\begin{bmatrix} v_{1,n+1} \\ v_{2,n+1} \end{bmatrix}=\begin{bmatrix} v_{1,n} \\ v_{2,n} \end{bmatrix}+\begin{bmatrix} h \\ h \end{bmatrix}.$$

Результат расчета нескольких шагов по времени приведен в таблице 10.5.

Таблица. 10.5

Результат интегрирования уравнений состояния по обратной формуле Эйлера

t	v_1	v_2
0.050	0.04366	0.04158
0.100	0.08195	0.07607
0.150	0.11571	0.10460
0.200	0.14562	0.12810
0.250	0.17227	0.14738
0.300	0.19615	0.16311
0.350	0.21768	0.17585

Подобный вывод уравнений состояния из узловых уравнений возможен лишь для RC- цепей, так как в этом случае s входит в числитель и, используя преобразование Лапласа, легко перейти к системе дифференциальных уравнений. Для перехода к системе дифференциальных уравнений в нормальной форме Коши необходимо также существование обратной матрицы C^{-1}.

Расчет временных характеристик прямыми методами. Расчет временных характеристик возможен на основе табличного и модифицированного узлового методов. Как известно, в табличном, модифицированном табличном, модифицированном узловом и модифицированном узловом с проверкой - методах, существует возможность управлять представлением реактивных ветвей, таким образом, чтобы оператор Лапласа можно было вынести перед мнимой частью матрицы коэффициентов, как общий множитель. Для соблюдения этого условия необходимо емкостные ветви представлять через адмитанс, а индуктивные ветви через импеданс, тогда оператор Лапласа окажется во всех компонентах реактивной части матрицы в числителе. Преобразование Лапласа, примененное к, таким образом, сформированной системе алгебраических уравнений, трансформирует ее в систему дифференциальных уравнений с матрицей коэффициентов, наиболее просто связанной с исходной матрицей алгебраической системы.

Итак, матрица коэффициентов, алгебраических систем уравнений, сформированная выше названными методами, при соблюдении указанных условий представления реактивных ветвей, может быть представлена в виде

$$T = G + s \cdot C,$$

где G- матрица действительной части; C- матрица мнимой части; s- оператор Лапласа. Соответственно алгебраическая система уравнений может быть записана

$$(G + s \cdot C) \cdot X = W. \qquad (10.13)$$

Применяя к такой алгебраической системе, преобразование Лапласа, получим систему дифференциальных уравнений

$$G \cdot X + C \cdot X' = W,$$

или

$$C \cdot X' = W - G \cdot X. \qquad (10.14)$$

Поскольку матрица мнимой части C, в общем случае, может быть вырождена, необходимо найти переход к системе дифференциальных уравнений в нормальной форме Коши, не требующий существования C^{-1}.

Рассмотрим обратную формулу Эйлера, предварительно помноженную на матрицу C

$$C \cdot X_{n+1} = C \cdot X_n + h \cdot C \cdot X'_{n+1},$$

и, подставляя, вместо $C \cdot X'_{n+1}$, значение, из дискретизированного, для этого случая, уравнения (10.14), запишем

$$C \cdot X_{n+1} = C \cdot X_n + h \cdot (W_{n+1} - G \cdot X_{n+1}),$$

или, группируя компоненты, окончательно получим

$$(C + h \cdot G) \cdot X_{n+1} = C \cdot X_n + h \cdot W_{n+1}. \qquad (10.15)$$

Это уравнение представляет собой обратную формулу Эйлера для интегрирования систем дифференциальных уравнений при матрице коэффициентов, заданной в комплексной форме. Это уравнение также показывает, что нет необходимости описывать цепь дифференциальными уравнениями в нормальной форме Коши, если известны действительные и мнимые части G и C, матрицы коэффициентов T. Матрицы G и C по отдельности могут быть вырождены, в то время как матрица $h \cdot G + C$, не вырождена.

Получим аналогичное соотношение, опираясь на метод трапеций

$$C \cdot X_{n+1} = C \cdot X_n + 0.5 \cdot h \cdot C \cdot X_n + 0.5 \cdot h \cdot C \cdot X_{n+1}.$$

Используя исходное уравнение (10.14) для замены $C \cdot X'_n$ и $C \cdot X'_{n+1}$, получаем

$$C \cdot X_{n+1} = C \cdot X_n + 0.5 \cdot h \cdot (W_n - G \cdot X_n) + 0.5 \cdot h \cdot (W_{n+1} - G \cdot X_{n+1}).$$

Группируя члены, содержащие X_{n+1} в правой части, получаем

$$\begin{aligned} &(C + 0.5 \cdot h \cdot G) \cdot X_{n+1} = \\ &= (C - 0.5 \cdot h \cdot G) \cdot X_n + 0.5 \cdot h \cdot (W_n + W_{n+1}). \end{aligned} \qquad (10.16)$$

Это уравнение представляет собой формулу трапеций для интегрирования систем дифференциальных уравнений при матрице коэффициентов, заданной в комплексной форме.

Рассмотрим пример использования обратной формулы Эйлера для систем дифференциальных уравнений, коэффициенты которых заданы в комплексной форме, для схемы, изображенной на рисунке 10.4.

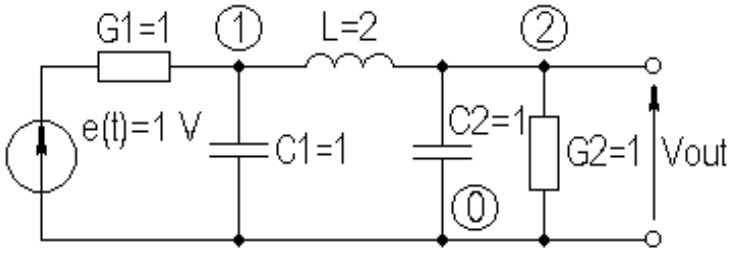


Рисунок 10.4 - Схема для расчета переходного процесса

Начальные условия положим нулевые, размер шага $h = 0.1$, источник напряжения представляет единичный скачок. Запишем составляющие системы уравнений, сформированной модифицированным узловым методом

$$G \cdot X + C \cdot X' = W,$$

где

$$G = \begin{bmatrix} G_1 & -G_1 & 0 & 0 & 1 \\ -G_1 & G_1 & 0 & 1 & 0 \\ 0 & 0 & G_2 & -1 & 0 \\ 0 & 1 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \end{bmatrix}; \quad C = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & C_1 & 0 & 0 & 0 \\ 0 & 0 & C_2 & 0 & 0 \\ 0 & 0 & 0 & -L & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix};$$

$$W = \begin{bmatrix} 0 & 0 & 0 & 0 & e(t) \end{bmatrix}^t.$$

Хотя матрица C является особой, суммарная матрица $(C + h \cdot G)$, не вырождена и решение (10.15) существует. Результаты расчета нескольких шагов интегрирования на ЭВМ сведены в таблицу 10.6.

Таблица 10.6.

Результаты интегрирования по обратной формуле Эйлера

t	V_{out}
0.0	0.0
0.1	0.09500·D-04
0.2	1.55650·D-03
0.3	3.69602·D-03
0.4	7.01808·D-03
0.5	1.16553·D-02
0.6	1.76897·D-02

Заметим, что в случае линейных цепей формирование матрицы коэффициентов производится один раз, и на каждом шаге интегрирования необходимо лишь переформировывать вектор правой части, с учетом предыдущего решения, а также входящий в него вектор свободных членов исходной алгебраической системы W, если входное воздействие является функцией времени.

В таких ситуациях, когда матрица коэффициентов системы уравнений остается постоянной, а меняется лишь вектор правой части, предпочтительно, как отмечалось ранее, применение алгоритмов решения линейных систем уравнений, основанные на факторизации матрицы коэффициентов (методы LU- и QR- факторизации).

Как видим, методы интегрирования или вычисления переходных процессов цепей свелись к последовательному решению систем линейных алгебраических уравнений.

В случае нелинейных цепей, матрицы коэффициентов G, C и вектор свободных членов W, в общем случае, являются функциями вектора переменных X, т.е. система алгебраических уравнений становится нелинейной. Применение преобразования Лапласа для перехода к системе алгебраических уравнений становится проблематичным, однако, с определенной степенью условности возможно и его формальное применение приведет к системе нелинейных дифференциальных уравнений. Так, обратную формулу Эйлера и формулу трапеций для интегрирования систем нелинейных дифференциальных уравнений, с комплексной матрицей коэффициентов исходной нелинейной алгебраической системы уравнений, можно записать

$$[C(X_n) + h \cdot G(X_n)] \cdot X_{n+1} = C(X_n) \cdot X_n + h \cdot W_{n+1}(X_n). \qquad (10.17)$$

$$[C(X_n) + 0.5 \cdot h \cdot G(X_n)] \cdot X_{n+1} = \\ = [C(X_n) - 0.5 \cdot h \cdot G(X_n)] \cdot X_n + h \cdot [W_n(X_n) + W_{n+1}(X_n)]. \qquad (10.18)$$

Для решения таких систем дифференциальных уравнений можно применять те же самые численные методы интегрирования. Однако, если, в случае линейных систем, достаточно один раз сформировать исходную алгебраическую систему уравнений и на каждом шаге переформировывать вектор свободных членов в формуле интегрирования с учетом предыдущего решения, то, в случае нелинейных цепей, необходимо на каждом шаге интегрирования переформировывать, как исходную алгебраическую систему, так и систему дифференциальных уравнений, с учетом решения на предыдущем шаге.

Прямое применение формул численного интегрирования для нелинейных систем дифференциальных уравнений не нашло, однако, широкого применения в силу низкой устойчивости и точности получаемых решений.

Более корректный подход к вычислению переходных процессов нелинейных цепей, основан на том, что система нелинейных дифференциальных уравнений, представленная в нормальной форме Коши, интерпретируется как система нелинейных уравнений, и к ней применяются известные итерационные методы решения типа Ньютона-Рафсона. Формулы интегрирования в этом случае есть результат применения итерационных алгоритмов к традиционным численным методам. Такой подход обеспечивает необходимую устойчивость и точность.

10.6 Метод дискретных моделей реактивных элементов

Кроме подхода, изложенного в предыдущем разделе и основанного на формировании и численном интегрировании системы дифференциальных уравнений, получил распространение альтернативный подход, использующий так называемые дискретные или сопровождающие модели реактивных элементов. При этом подходе компонентные дифференциальные уравнения реактивных элементов аппроксимируются соотношениями, соответствующими одной из формул численного интегрирования, а результаты аппроксимации интерпретируются, как резистивные дискретные модели реактивных элементов, значения которых зависят от шага интегрирования и результатов решения на предыдущем шаге.

После замены реактивных элементов резистивными моделями, известными методами формирования математических моделей формируется система алгебраических уравнений, в вектор свободных членов которой входят источники резистивных моделей, зависящие от результатов решения на предыдущем шаге. Решая систему алгебраических уравнений, находим очередное решение, подставляя найденное решение в правую часть, как предыдущее решение, можем найти решение в следующей точке и т.д.

Рассмотрим дискретные модели конденсатора, описываемого компонентным уравнением вида

$$i = C \cdot dv / dt . \qquad (10.19)$$

Заменяя производную конечной разностью

$$dv / dt = v'_{n+1} = (v_{n+1} - v_n) / (t_{n+1} - t_n) = (v_{n+1} - v_n) / h ,$$

и используя компонентное уравнение (10.19), получим соотношение

$$i_{n+1} = C \cdot (v_{n+1} - v_n) / h ,$$

которое можно переписать в виде

$$i_{n+1} = C / h \cdot v_{n+1} - C / h \cdot v_n , \qquad (10.20)$$

или

$$v_{n+1} = h / C \cdot i_{n+1} + v_n . \qquad (10.21)$$

Те же соотношения можно получить, подставляя в компонентное уравнение (10.19), вместо производной, ее выражение из обратной формулы Эйлера

$$v'_{n+1} = 1 / h \cdot v_{n+1} - 1 / h \cdot v_n .$$

Уравнениям (10.20, 10.21) соответствуют резистивные модели, изображенные на рисунке 10.5.

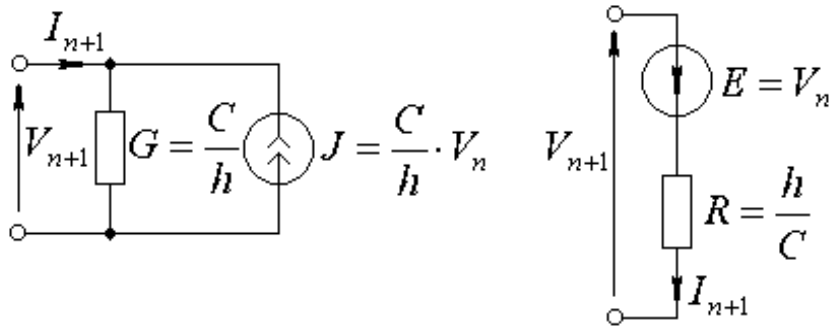


Рисунок 10.5 - Резистивные модели емкости на основе обратной формулы Эйлера

Заменяя производную, полусуммой производных в текущей и предыдущей точках, или конечной разностью половинного шага

$$dv/dt = (v'_{n+1} + v'_n)/2 = (v_{n+1} - v_{n+1/2})/h + (v_{n+1/2} - v_n)/h = \\ = (v_{n+1} - v_n)/h,$$

или

$$v'_{n+1} = -v'_n + 2/h \cdot (v_{n+1} - v_n),$$

и, используя компонентное уравнение (10.19), получим соотношение

$$i_{n+1} = -i_n + 2 \cdot C/h \cdot (v_{n+1} - v_n),$$

которое можно переписать в виде

$$i_{n+1} = 2 \cdot C/h \cdot v_{n+1} - 2 \cdot C/h \cdot v_n - i_n, \qquad (10.22)$$

или

$$v_{n+1} = h/(2 \cdot C) \cdot i_{n+1} - h/(2 \cdot C) \cdot i_n + v_n. \qquad (10.23)$$

Те же соотношения можно получить, подставляя в компонентное уравнение (10.19), вместо производной, ее выражение из формулы трапеций

$$v'_{n+1} = -v'_n + 2/h \cdot v_{n+1} - 2/h \cdot v_n.$$

Уравнениям (10.22, 10.23) соответствуют резистивные модели, изображенные на рисунке 10.6.

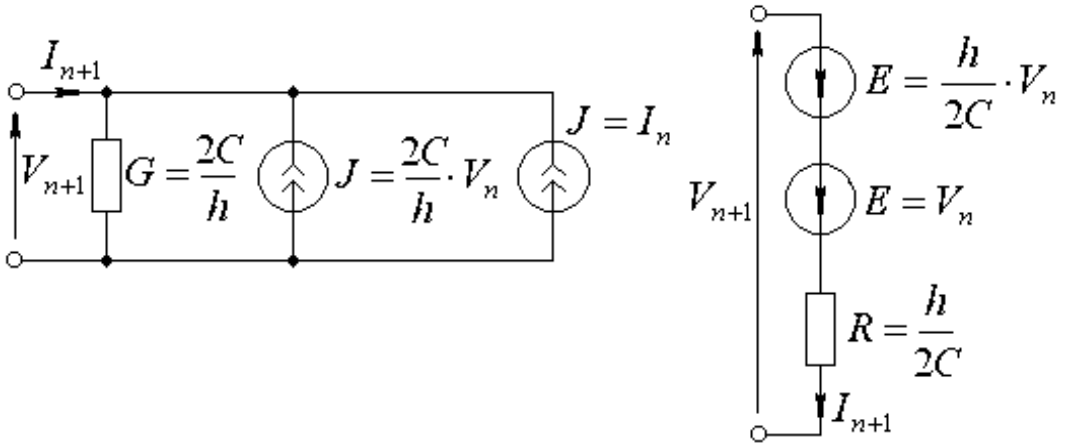


Рисунок 10.6 - Резистивные модели емкости на основе формулы трапеций

Перейдем к рассмотрению дискретных моделей катушки индуктивности, описываемой компонентным уравнением вида

$$v = L \cdot di/dt. \qquad (10.24)$$

Заменяя производную конечной разностью

$$di/dt = i'_{n+1} = (i_{n+1} - i_n)/(t_{n+1} - t_n) = (i_{n+1} - i_n)/h,$$

и используя компонентное уравнение (10.24), получим соотношение

$$v_{n+1} = L \cdot (i_{n+1} - i_n)/h,$$

которое можно переписать в виде

$$v_{n+1} = L/h \cdot i_{n+1} - L/h \cdot i_n, \qquad (10.25)$$

или

$$i_{n+1} = h/L \cdot v_{n+1} + i_n. \qquad (10.26)$$

Те же соотношения можно получить, подставляя в компонентное уравнение (10.24) вместо производной, ее выражение из обратной формулы Эйлера

$$i'_{n+1} = 1/h \cdot i_{n+1} - 1/h \cdot i_n.$$

Уравнениям (10.25, 10.26) соответствуют резистивные модели, изображенные на рисунке 10.7.

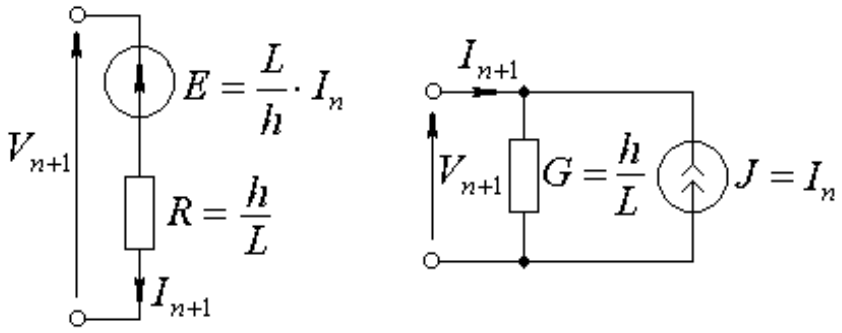


Рисунок 10.7 - Резистивные модели индуктивности на основе обратной формулы Эйлера

Заменяя производную, полусуммой производных в текущей и предыдущей точках или конечной разностью половинного шага

$$di/dt = (i'_{n+1} + i'_n)/2 = (i_{n+1} - i_{n+1/2})/h + (i_{n+1/2} - i_n)/h = \\ = (i_{n+1} - i_n)/h,$$

или

$$i'_{n+1} = -i'_n + 2/h \cdot (i_{n+1} - i_n),$$

и, используя компонентное уравнение (10.24), получим соотношение

$$v_{n+1} = -v_n + 2 \cdot L/h \cdot (i_{n+1} - i_n),$$

которое можно переписать в виде

$$v_{n+1} = 2 \cdot L/h \cdot i_{n+1} - 2 \cdot L/h \cdot i_n - v_n, \qquad (10.27)$$

или

$$i_{n+1} = h/(2 \cdot L) \cdot i_{n+1} - h/(2 \cdot L) \cdot v_n + i_n. \qquad (10.28)$$

Те же соотношения можно получить, подставляя в компонентное уравнение (10.24), вместо производной, ее выражение из формулы трапеций

$$i'_{n+1} = -i'_n + 2/h \cdot i_{n+1} - 2/h \cdot i_n.$$

Уравнениям (10.27, 10.28) соответствуют резистивные модели, изображенные на рисунке 10.8.

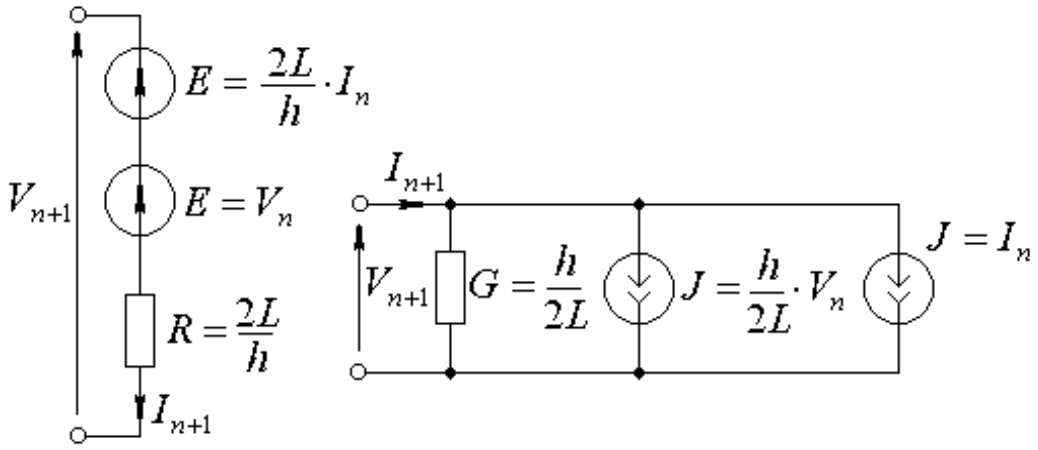


Рисунок 10.8 - Резистивные модели индуктивности на основе формулы трапеций

Используя другие конечноразностные представления производной, или подставляя в компонентные уравнения реактивных ветвей, ее выражение из других формул интегрирования, можем получить более сложные резистивные модели реактивных элементов.

Таким образом, применение дискретных моделей к реактивным элементам приводит к следующему:

1. Цепь становится резистивной, каждый реактивный элемент заменяется резистором либо проводимостью с номиналами, определяемыми номиналом реактивности и шагом интегрирования.
2. Последовательно с резистором или параллельно проводимости подключаются дополнительные источники, соответственно, напряжения или тока, число и номиналы которых зависят от используемого конечно-разностного представления, значений токов и напряжений в предыдущие моменты времени и шага интегрирования.

В результате описываемого подхода цепь становится резистивной с источниками, зависящими от значения переменных в предыдущие моменты времени. Математическая модель такой цепи представляет собой систему алгебраических уравнений с дополнительным вектором свободных членов, зависящим от состояния цепи в предыдущие моменты времени

$$T \cdot X_{n+1} = W + W(X_n). \qquad (10.29)$$

Для формирования математической модели цепи, реактивные элементы которой заменены резистивными моделями, можно воспользоваться любым из известных методов в зависимости от используемых моделей, причем ограничений на представление реактивных элементов не накладывается.

Для определения переходного процесса необходимо задаться начальными значениями вектора переменных, сформировать систему уравнений, решая которую, находим значение вектора неизвестных в текущий момент времени. Подставляя найденное значение вектора неизвестных, в качестве начального значения, ищем следующее значение и т.д.

Необходимо отметить, что в случае линейных цепей матрицу коэффициентов системы достаточно сформировать лишь один раз перед итерационным циклом по времени, в котором вектор свободных членов системы переформировывается каждый раз с учетом предыдущего решения.

В качестве иллюстрации рассматриваемого подхода рассмотрим формирование системы уравнений (10.29) для цепи изображенной на рисунке 10.9.

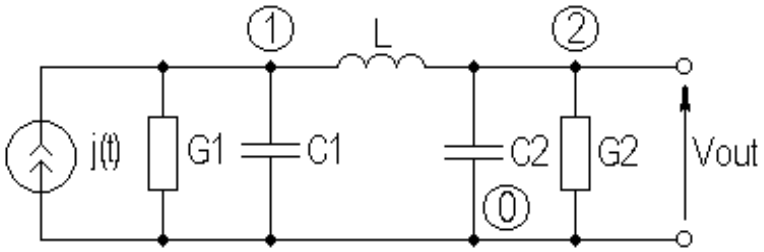


Рисунок 10.9 - Цепь для иллюстрации конечно-разностных моделей реактивных элементов

Используя конечно - разностные или сеточные модели, соответствующие обратной формуле Эйлера, и представляя конденсаторы

проводимостями и источниками тока (10.20), а катушку индуктивности резистором и источником напряжения (10.25), составим систему алгебраических уравнений резистивной цепи модифицированным методом узловых потенциалов

$$\begin{bmatrix} G_1 + C_1/h & 0 & 1 \\ 0 & G_2 + C_2/h & -1 \\ 1 & -1 & -L/h \end{bmatrix} \cdot \begin{bmatrix} v_{1,n+1} \\ v_{2,n+1} \\ v_{3,n+1} \end{bmatrix} = \begin{bmatrix} j_{n+1} \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} C_1/h \cdot v_{1,n} \\ C_2/h \cdot v_{2,n} \\ -L/h \cdot i_{L,n} \end{bmatrix}$$

Рассмотренный пример позволяет сделать несколько важных для первоначального усвоения замечаний:

1. Источники резистивных моделей не порождают дополнительных ветвей и узлов, а входят как начальные условия в вектор свободных членов системы.
2. Данный пример не совсем удачен в том плане, что конденсаторы подключены вторым зажимом к общему проводу, поэтому вместо разности узловых потенциалов $(v_{i,n} - v_{j,n})$ использованы просто узловые потенциалы.
3. Систему, аналогичную (10.29) можно получить, используя обратную формулу Эйлера для интегрирования систем дифференциальных уравнений (10.15) $(C + h \cdot G) \cdot X_{n+1} = C \cdot X_n + h \cdot W_{n+1}$, разделив ее на h, но это чисто внешняя связь.

Итак, для вычисления временных характеристик цепей, на основе сеточных или сопровождающих моделей, необходимо многократно решать систему линейных уравнений с разными правыми частями. Для решения систем в такой ситуации целесообразно использование методов решения, основанных на факторизации матрицы коэффициентов (методы LU- и QR-факторизации).

В случае определения реакции отдельных компонент вектора решений можно воспользоваться идеей решения сопряженной системы уравнений изложенной ранее. Для этого перепишем систему (10.29) в виде

$$T \cdot X_{n+1} = W(X_n). \qquad (10.30)$$

Вводя выходную функцию от вектора решения

$$\Phi(X_{n+1}) = d^t \cdot X_{n+1},$$

где d - вектор выбора компонент, и сопряженную систему уравнений в виде

$$T^t \cdot Y = -d,$$

можем записать

$$Y^t = -d^t \cdot T^{-1}.$$

Выходную функцию из (10.30) можем определить как

$$\Phi(X_{n+1}) = d^t \cdot T^{-1} \cdot W(X_n),$$

откуда, используя предыдущее соотношение, получаем

$$\Phi(X_{n+1}) = -Y^t \cdot W(X_n).$$

Т.е. решая один раз сопряженную систему, находим вектор Y^t, каждое новое значение выходной функции определяем, как скалярное произведение этого вектора на новый вектор правой части.

В случае нелинейных цепей матрица коэффициентов системы (10.29) также зависит от вектора решений

$$T(X_n) \cdot X_{n+1} = W(X_n). \qquad (10.31)$$

В результате, на каждом шаге итерации по времени, необходимо заново формировать матрицу коэффициентов и вектор свободных членов.

Для решения нелинейной системы с малым шагом по времени, можно попытаться решать ее как линейную на каждом шаге итерации, предполагая, что нелинейность слабо проявляется в пределах шага интегрирования. Однако в общем случае необходимо использовать итерационные методы решения нелинейных систем алгебраических уравнений, например, метод Ньютона-Рафсона.

11 ОПТИМИЗАЦИЯ ЭЛЕКТРОННЫХ СХЕМ

11.1 Введение в теорию оптимизации

Теория оптимизации находит широкое применение при проектировании электронных схем по заданному набору требований к характеристикам и ограничений на изменение параметров. Ее используют для максимизации или минимизации некоторой скалярной целевой функции нескольких переменных, на которые наложены дополнительные ограничения.

Оптимизация является по существу основным универсальным методом проектирования РЭУ, поскольку синтез развит в основном для пассивных цепей фильтрации и согласующе - трансформирующих цепей, а любые методики проектирования конкретного класса устройств носят частный характер.

Под оптимизацией или параметрическим синтезом, в общем случае, понимают процедуру целенаправленного перебора параметров заданного схемного решения РЭУ, с целью удовлетворения набора характеристик заданным требованиям. Критерием достижения заданных требований выступает значение ошибки или отклонение заданного набора характеристик от требуемых, являющаяся функцией внутренних и внешних параметров устройства, и принимающая нулевое значение при достижении заданных требований. Функцию ошибки называют чаще целевой функцией, и оптимизация сводится к задаче минимизации целевой функции в заданном пространстве параметров и заданных ограничениях на их значения.

Разработчику схем не требуется знать все аспекты теории оптимизации, но он должен знать основные понятия, например о градиенте, иметь представление о возможностях методов и уметь формировать целевую функцию. Чтобы сформировать целевую функцию, разработчик должен иметь представление об основах теории оптимизации, быть знаком с техническими терминами, используемыми в этой области, знать набор основных характеристик и их взаимообусловленность, а также основные практические ограничения на параметры.

Под пространством параметров оптимизации будем понимать основной набор варьируемых параметров схемы - номиналов элементов, режима работы и внешних параметров среды, например, температуры. Пространство параметров описывается вектором варьируемых параметров X.

Под целевой функцией оптимизации $F(X)$, будем понимать скалярную функцию, содержащую, в общем случае, информацию об отклонении требуемых характеристик (выходных функций) от заданных, и стремящуюся, при достижении оптимума, к нулю. Чтобы избавиться от влияния знака отклонения, в целевой функции берется сумма квадратов или четных степеней отклонений, из которой вычисляется корень четной степени. Для исключения влияния абсолютных значений характеристик их

отклонения нормируются относительно требуемых значений. Влияние отдельных характеристик в целевой функции подчеркивается с помощью весовых коэффициентов. Таким образом, целевую функцию, в общем виде, можно записать

$$F(X)=\left[\sum_{i=1}^{m}\alpha_i\cdot\left[\frac{f_{0i}(X)-f_i(X)}{f_{0i}(X)}\right]^p\right]^{1/p}, \qquad (11.1)$$

где m - число выходных функций (характеристик); p - показатель четной степени; α_i - весовой коэффициент; $f_{0i}(X)$ - требуемое значение выходной функции; $f_i(X)$ - текущее значение выходной функции. Кроме того отклонения функций нормированы относительно требуемых значений.

Достижение минимума целевой функции $F(X)$ должно выполнятся при соблюдении ряда ограничений на параметры либо функций от параметров в виде равенств

$$e_i(X)=0, \qquad (11.2)$$

или неравенств

$$g_i(X)\geq 0. \qquad (11.3)$$

Функции ограничения могут быть, как линейными, так и нелинейными. Вектор решения X должен удовлетворять этим ограничениям.

Типичным примером линейного ограничения типа неравенства является условие

$$x_i-l_i\geq 0, \qquad (11.4)$$

которое определяет, что переменная x_i, должна быть больше, чем нижний предел l_i. Диапазон изменения x_i может быть ограничен и сверху

$$u_i-x_i\geq 0. \qquad (11.5)$$

В частности, нельзя реализовать отрицательные параметры элементов схемы и их следует ограничить с помощью верхнего и нижнего пределов, чтобы получить технически приемлемое решение.

Для ограничения пространства параметров удовлетворяющего ограничениям иногда используют так называемую штрафную функцию, зависящую от степени нарушения ограничений. Штрафная функция входит множителем в целевую функцию, вызывая ее увеличение при нежелательном наборе параметров.

Почти все современные методы оптимизации основаны на определении такой последовательности векторов X, при которой выполняются условия

$$F(X^0)>F(X^1)>\cdots>F(X^k). \qquad (11.6)$$

Эта последовательность будет сходиться к минимуму, если функция выпукла и такой минимум будет глобальным. Если функция не выпукла, то, в общем случае, можно говорить лишь о локальном минимуме. В большинстве случаев ничего не известно о виде функции, в особенности, если она зависит от многих переменных.

Формально, под оптимизацией понимают процедуру

целенаправленного подбора параметров, приводящую к минимизации целевой функции $F(X)$, при соблюдении ограничений $E(X)$ и $G(X)$. Различные методы оптимизации отличаются друг от друга стратегией и тактикой поиска оптимума, способом получения информации о направлении поиска, способом определения оптимума в заданном направлении и критерием останова при достижении заданных требований.

Поиск оптимума всегда начинают с какой-то начальной точки X^0 пространства параметров, называемой нулевым приближением. Выбор нулевого приближения в многоэкстремальном пространстве существенно влияет на найденное решение, поэтому выбор нулевого приближения требует определенного опыта в данной области. Часто для обеспечения поиска глобального оптимума, нулевое приближение выбирают, используя многомерные датчики случайных последовательностей.

Формирование целевой функции $F(X)$ требует понимания сути задачи. Например, задачу отыскания решения X системы m, в общем случае, нелинейных уравнений с n неизвестными, где $m \le n$, $f_j(x_1, x_2, ..., x_n) = 0$, $j = 1, 2, \ldots, m$, можно сформулировать как задачу нахождения минимума скалярной функции нескольких переменных

$$F(X) = \sum_{j=1}^{m} f_j^2(X). \qquad (11.7)$$

Минимум этой функции, для которого $F(X^*) = 0$, является, очевидно, решением системы нелинейных уравнений.

Максимизация или минимизация проводится по одним и тем же алгоритмам, так как, если функция $F(X)$ имеет минимум в точке X, то функция $-F(X)$ имеет максимум в той же точке.

Большинство современных оптимизационных алгоритмов является итерационными, использующими на каждом шаге расчет требуемых характеристик. Два важных момента оптимизации связаны с выбором направления поиска оптимума, задаваемое вектором S, и отыскание самого оптимума X^*, в заданном направлении пространства параметров.

Для того чтобы двигаться вниз по склону многомерного пространства в окрестности оптимума, требуется информация о направлении убывания целевой функции. Это направление можно определить с помощью пробных шагов и оценок целевой функции, либо рассчитав производную функции. В многомерном пространстве информация о функции многих переменных содержится в градиенте. Градиент функции $F(X)$ от n переменных представляет, собой многомерный вектор вида

$$\nabla F = \nabla F(X) = \left[\partial F / \partial X_1, \partial F / \partial X_2, ..., \partial F / \partial X_n\right]^t. \qquad (11.8)$$

Градиент указывает направление возрастания функции и, следовательно, вектор $-\nabla F$ направлен по склону. Если рассмотреть другое направление S, то производная функции в этом направлении определяется выражением

$S^t \cdot \nabla F,$

и функция убывает в этом направлении при выполнении условия

$$-S^t \cdot \nabla F > 0. \qquad (11.9)$$

11.2 Классическая теория оптимизации

Из математики известно, что для определения максимума или минимума функции одной переменной нужно найти ее производную, приравнять ее нулю и найти точку X^*, являющуюся решением этого уравнения. Данный подход справедлив и для функций нескольких переменных, только в этом случае уже градиент приравнивается нулю

$$\nabla F(X) = 0 \qquad (11.10)$$

и вектор решения системы уравнений (12.10) определяет точку оптимума в многомерном пространстве X^*.

Для примера определим минимум функции

$$F(X) = (x_2 - x_1)^2 + (1 - x_1)^2.$$

Приравняем градиент функции нулю

$$\nabla F(X) = \begin{bmatrix} 4 \cdot x_1 - 2 \cdot x_2 - 2 \\ 2 \cdot x_2 - 2 \cdot x_1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix},$$

что можно переписать в виде

$$\begin{bmatrix} 4 & -2 \\ -2 & 2 \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 2 \\ 0 \end{bmatrix}.$$

Решение этой системы уравнений дает вектор

$$X = [1 \quad 1]^t.$$

Заметим, что данная задача приводит к решению системы линейных уравнений. В общем же случае возникает необходимость решения системы нелинейных алгебраических уравнений, которые, как известно, решаются итерационными методами.

Метод множителей Лагранжа. Более сложным случаем является задача нахождения минимума при определенных ограничениях, метод решения которой, впервые был предложен Лагранжем. Ставится задача нахождения минимума функции $F(X)$, при условии, что $e_j(X) = 0$, где $j = 1,\ldots,k$. При этом формируется новая, объединяющая функция, называемая лагранжианом

$$L(X, \Lambda) = F(X) - \sum_{j=1}^{k_1} \lambda_j \cdot e_j(X), \qquad (11.11)$$

где $\Lambda = [\lambda_1, \lambda_2, ..., \lambda_{k_1}]^t$ - вектор дополнительных переменных, называемых множителями Лагранжа. Для нахождения оптимума с ограничениями, выразим производные от функции $L(X, \Lambda)$ по переменным X и Λ, и, приравнивая их нулю, получим систему уравнений

$$\begin{aligned} \partial L / \partial x_i &= \partial F / \partial x_i - \sum_{j=1}^{k_1} \partial e_j(X) / \partial x_i = 0; \\ \partial L / \partial \lambda_j &= -e_j(X) = 0, \end{aligned} \tag{11.12}$$

где $i = 1, \ldots, n$; $j = 1, \ldots, k_1$. Решение данной системы соответствует оптимуму с ограничениями.

В качестве примера, найдем минимум функции

$$F(X) = (x_2 - x_1)^2 + (1 - x_1)^2,$$

при условии

$$E(X) = x_1 + x_2 - 4 = 0.$$

Сформируем Лагранжиан

$$L(X, \Lambda) = (x_2 - x_1)^2 + (1 - x_1)^2 - \lambda_1 \cdot (x_1 + x_2 - 4),$$

и, продифференцировав его, по x_1, x_2 и λ_1, получим систему линейных уравнений

$$\begin{aligned} \partial L / \partial x_1 &= 4 \cdot x_1 - 2 \cdot x_2 - \lambda_1 - 2 = 0, \\ \partial L / \partial x_2 &= -2 \cdot x_1 + 2 \cdot x_2 - \lambda_1 = 0, \\ \partial L / \partial \lambda_1 &= -x_1 - x_2 + 4 = 0. \end{aligned}$$

Решение системы дает точку оптимума $x_1 = 9/5;\ x_2 = 11/5;\ \lambda_1 = 4/5.$

Метод множителей Лагранжа имеет в основном теоретическое значение. Современные методы оптимизации в процессе итераций ищут последовательные приближения оптимального решения.

Основной итерационный алгоритм. Пусть заданы функция n переменных $F(X)$ и произвольная начальная точка

$$X^0 = [x_1^0, x_2^0, \ldots, x_n^0]^t.$$

Желательно двигаться от этой точки к другим, так чтобы выполнялись неравенства

$$F(X^0) > F(X^1) > \ldots > F(X^k). \tag{11.13}$$

Вначале следует определить направление поиска оптимума. Произвольное направление в n - мерном пространстве определяет вектор $S = [s_1, s_2, \ldots, s_n]^t$. Предположим, что сделан шаг d в каком-либо направлении. Для k - той итерации следующей точкой будет вектор

$$X^{k+1} = X^k + d_k \cdot S^k, \tag{11.14}$$

где d - шаг, вещественная константа. Шаг в процессе поиска может изменяться, в соответствии с определенной стратегией, отражающей успешность предыдущего поиска. Приращение вектора параметров отсюда равно

$$\Delta X^k = X^{k+1} - X^k = d_k \cdot S^k . \qquad (11.15)$$

Для эффективности поиска, такие шаги следует проводить по линейно независимым направлениям, т.е. необходимо построить ортогональную систему координат в исходной точке пространства параметров. Векторы, соответствующие n направлениям в пространстве параметров, можно объединить в матрицу размера $n * n$

$$S = \left\lfloor S^0, S^1, \ldots, S^{n-1} \right\rfloor, \qquad (11.16)$$

где S^i - вектор столбец, соответствующий i - му направлению. Эта матрица преобразуется в процессе поиска в соответствии с некоторой стратегией.

В качестве примера, рассмотрим процедуру поиска минимума функции

$$F(X) = x_1^2 + x_2^2 + 3 \cdot x_3^2 ,$$

с использованием матрицы направлений. Для простоты выберем матрицу направлений единичной

$$S = \begin{matrix} s^0 & s^1 & s^2 \end{matrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

и в качестве начальной точки возьмем

$$X^0 = \begin{bmatrix} 1 & 2 & 1 \end{bmatrix}^t .$$

Очевидно, что функция имеет глобальный минимум, при $x_1 = x_2 = x_3 = 0$. Начальное значение функции $F(X^0) = 8$. Выберем S^0, в качестве первого направления, и сделаем шаг длиной $d_0 = -0.5$ (пока выбор d произволен). В соответствии с (11.14), получаем следующую точку

$$X^1 = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} - 0.5 \cdot \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0.5 \\ 2 \\ 1 \end{bmatrix} .$$

В точке X^1 значение функции $F(X^1) = 7.25 < F(X^0)$ и отметим этот результат, как успешный. Делая шаг во втором направлении S^1 и, выбирая его длину большей, скажем, $d_1 = -1$, получаем

$$X^2 = \begin{bmatrix} 0.5 \\ 2 \\ 1 \end{bmatrix} - 1 \cdot \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0.5 \\ 1 \\ 1 \end{bmatrix}$$

и $F(X^2) = 4.25 < F(X^1)$. Можно сделать шаг и в третьем направлении, а затем повторить процедуру поиска сначала, до тех пор, пока не будет найден минимум.

Обсудим, однако, процедуру поиска на предмет усовершенствования. Несколько возможных усовершенствований видны, с первого взгляда. Так можно отказаться от принятия любого решения уменьшающего значение целевой функции, а попробовать найти минимум при движении в заданном направлении. При этом число шагов, необходимых для достижения оптимума, может быть сокращено. Следовательно, необходима теория минимизации функции при движении вдоль линии (одной переменной). Можно использовать производные целевой функции для определения наилучших направлений. Наконец, можно использовать вторые производные или их аппроксимации для улучшения направлений поиска. Желательно также определить стратегии изменения шага поиска в зависимости от его предшествующих результатов.

В случае сложного рельефа многомерного пространства и областей оптимумов можно предложить стратегию изменения ориентации направлений независимого поиска.

Методы, не требующие знания производных целевой функции, называются **методами прямого поиска.** Методы, использующие значения производных для определения направления поиска называются **градиентными методами.** Мы не будем на них пока останавливаться, так как в разделе, посвященном расчету чувствительностей, были рассмотрены эффективные алгоритмы вычисления градиента. Однако минимизация функции при движении в заданном направлении весьма актуальна и будет рассмотрена несколько позднее.

Если градиент в точке известен, то эта информация может быть использована для ускорения процесса минимизации. Действительно, так называемые **методы наискорейшего спуска** основаны на предположении, что наискорейшее убывание функции достигается в направлении, противоположном направлению градиента. Однако эти методы не учитывают информации, полученной на предыдущих шагах. Если предварительная информация используется для определения следующего направления поиска, то можно надеяться на ускорение сходимости. Такие методы объединены под названием **методов сопряженного градиента.**

В некоторых случаях используют информацию о вторых производных, обычно их непосредственно не получают, а вычисляют приближенно, используя различные аппроксимации. Такие методы известны под названием

методов с переменной метрикой или иногда их еще называют **квазиньютоновскими.**

Итерационный алгоритм поиска минимума. Обобщим материал, представив общий алгоритм, пригодный для многих итерационных методов минимизации. Различия появляются только в выборе направлений поиска S^k, а также в некоторых деталях критерия окончания.

1. Полагаем $k = 0$ и выбираем точку начального приближения X^0.

2. Вычислим $F(X^k)$ и $\nabla F(X^k)$, если метод требует градиента.

3. Определяем направление поиска S^k и нормируем вектор направления поиска к единичной длине.

4. В направлении поиска S^k найдем длину шага d_k такую, что $F(X^k + d_k \cdot S^k) < F(X^k)$ или функция $F(X^k + d_k \cdot S^k)$, минимальна в направлении S^k.

5. Вычислим $\Delta X^k = d_k \cdot S^k$ и $X^{k+1} = X^k + \Delta X^k$.

6. Если $\left|F(X^{k+1}) - F(X^k)\right| < \varepsilon_1$ и $\left\|\Delta X^k\right\| < \varepsilon_2$, то процесс сошелся, иначе положим $k = k + 1$ и перейдем к шагу 2.

Напомним, что норма вектора S определяется выражением $\|S\| = \sqrt{\sum S_i^2}$. Для нормировки вектора к единичной длине необходимо каждую компоненту вектора разделить на его норму.

Поиск вдоль заданного направления. Общий алгоритм поиска, изложенный выше, требует определения длины шага d_k вдоль направления поиска S^k. Там же была упомянута процедура выбора d_k для минимизации $F(X)$ вдоль направления $X^k + d_k \cdot S^k$. Рассмотрим два возможных способа представления направления поиска.

Первый способ не требует информации о градиенте, а использует параболическую интерполяцию по трем точкам, в направлении поиска, второй основывается на применении кубического интерполирующего полинома и его производных.

Параболическая интерполяция. Поясним первый метод поиска, на примере задачи с одной действительной переменной u. Обозначим функцию символом v. Рассмотрим рисунок 11.1.

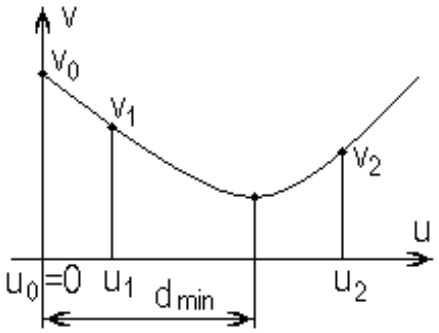


Рисунок 11.1 - Параболическая аппроксимация по трем точкам

Первое значение функции v_0 является результатом предыдущего шага минимизации. Установим начало координат таким образом, что $u_0 = 0$. Выберем произвольно величины двух шагов u_1 и u_2, рассчитаем соответствующие им значения функции v_1 и v_2. Таким образом получаем три точки $(0, v_0)$, (u_1, v_1) и (u_2, v_2), которые используются для интерполяции параболой $v = C \cdot u^2 + B \cdot u + A$. Подставив в это уравнение координаты трех точек, получим три уравнения относительно неизвестных A, B и C:

$$\begin{aligned} v_0 &= A; \\ v_1 &= C \cdot u_1^2 + B \cdot u_1 + A; \\ v_2 &= C \cdot u_2^2 + B \cdot u_2 + A. \end{aligned} \qquad (11.17)$$

Решение этой системы имеет вид

$$C = \frac{u_2 \cdot (v_1 - v_0) - u_1 \cdot (v_2 - v_0)}{u_1^2 \cdot u_2 - u_2^2 \cdot u_1}; \qquad (11.18)$$

$$B = \frac{-u_2^2 \cdot (v_1 - v_0) + u_1^2 \cdot (v_2 - v_0)}{u_1^2 \cdot u_2 - u_2^2 \cdot u_1}; \qquad (11.19)$$

$$A = v_0. \qquad 11.20)$$

Минимум параболы найдем, положив производную $v' = 2 \cdot C \cdot u + B = 0$, при условии, что $v'' = 2 \cdot C > 0$. Откуда расстояние до минимума параболы определится

$$d_{min} = \frac{-B}{2 \cdot C} = \frac{u_2^2 \cdot (v_1 - v_0) - u_1^2 \cdot (v_2 - v_0)}{2 \cdot [u_2 \cdot (v_1 - v_0) - u_1 \cdot (v_2 - v_0)]}, \qquad (11.21)$$

при условии, что $C > 0$.

В качестве иллюстрации, рассмотрим процедуру одномерного поиска минимума функции из предыдущего примера $F(X) = x_1^2 + x_2^2 + 3 \cdot x_3^2$, с начальной точкой $X^0 = [1 \quad 2 \quad 1]^t$ и значением функции в этой точке $F(X^0) = 8$. Первая точка соответствует $u_0 = 0,\ v_0 = 8$. Выберем длину шага

$d_0 = -0.5$ вдоль направления S^0. Это также было сделано в предыдущем примере и $F(X^0 + 0.5 \cdot S^0) = 7.25$. Вторая точка имеет координаты $u_1 = -0.5,\ v_1 = 7.25$. Необходимо сделать еще один шаг, скажем, при $d = 0.5$, для которого $F(X^0 + 0.5 \cdot S^0) = 9.25$. Третья точка соответствует $u_2 = 0.5,\ v_2 = 9.25$. Откуда, в соответствии с (11.21), имеем $d_{min} = -1$. Эта величина используется для перехода от точки X^0 к точке X^1 и, в соответствии с (11.14), $X^1 = [1 \quad 2 \quad 1]^t + (-1) \cdot [1 \quad 0 \quad 0]^t = [0 \quad 2 \quad 1]^t$, для которой $F(X^1) = 7$.

Так как в примере используется квадратичная функция трех переменных, то минимум в направлении S^0 получен точно. В более сложных случаях шаги можно продолжить вновь до достижения минимума. С другой стороны, поиск можно вести только до получения $F(X^1) < F(X^0)$, а затем перейти к поиску по другому направлению.

Интерполяция кубическим полиномом. Если градиенты определяются относительно просто, то вместе со значением функции можно рассматривать и ее производную. Это дает информацию для интерполяции кубическим полиномом

$$v = D \cdot u^3 + C \cdot u^2 + B \cdot u + A. \tag{11.22}$$

Его производная равна

$$v' = 3 \cdot D \cdot u^2 + 2 \cdot C \cdot u + B. \tag{11.23}$$

Два экстремума находятся приравниванием производной к нулю

$$u_{opt} = \frac{-2 \cdot C \pm \sqrt{4 \cdot C - 12 \cdot B \cdot D}}{6 \cdot D}. \tag{11.24}$$

При выборе конкретного экстремума заметим, что для минимума вторая производная должна быть положительной

$$6 \cdot D \cdot u + 2 \cdot C > 0.$$

Подставив в это выражение u_{opt} можно показать, что знак плюс в этой формуле соответствует минимуму.

Коэффициенты $A,\ B,\ C,\ D$ пока неизвестны, но, используя интерполяционный полином и его производную, для двух точек отсчета функции, получаем

$$\begin{aligned} v_0 &= A; \\ v_0' &= B; \\ v_1 &= D \cdot u_1^3 + C \cdot u_1^2 + B \cdot u_1 + A; \\ v_1' &= 3 \cdot D \cdot u_1^2 + 2 \cdot C \cdot u_1 + B. \end{aligned} \tag{11.25}$$

Так как коэффициенты A и B по существу известны, подставим их в последние два уравнения и получим систему

$$\begin{bmatrix} u_1^3 & u_1^2 \\ 3\cdot u_1^2 & 2\cdot u_1 \end{bmatrix} \cdot \begin{vmatrix} D \\ C \end{vmatrix} = \begin{vmatrix} v_1 - B\cdot u_1 - A \\ v_1' - B \end{vmatrix}, \tag{11.26}$$

решение которой, даст значения коэффициентов C и D. Тогда, в соответствии с (11.24), имеем

$$d\,min = \frac{-C + \sqrt{C^2 - 3\cdot B\cdot D}}{3\cdot D}, \tag{11.27}$$

если $D \neq 0$. При $D = 0$, интерполирующая кривая становится квадратичной параболой и, в соответствии с этим, модифицируются уравнения.

В качестве иллюстрации рассмотрим процедуру одномерного поиска минимума функции из предыдущего примера

$$F(X) = x_1^2 + x_2^2 + 3\cdot x_3^2,$$

с начальной точкой $X^0 = [1 \quad 2 \quad 1]^t$, значением функции в этой точке $F(X^0) = 8$ и направлением $S^0 = [1 \quad 0 \quad 0]^t$.

Градиент этой функции запишется $\nabla F(X) = [2\cdot x_1 \quad 2\cdot x_2 \quad 6\cdot x_3]^t$, и в начальной точке он равен $\nabla F(X_0) = [2 \quad 4 \quad 6]^t$. Выберем произвольный размер шага, например $d_0 = -3$.

Используя (11.14), найдем следующею точку

$$X^1 = X^0 - 3\cdot S^0 = [1 \quad 2 \quad 1]^t - 3\cdot [1 \quad 0 \quad 0]^t = [-2 \quad 2 \quad 1]^t.$$

Значения функции и ее градиента в этой точке равны

$$F(X^0 - 3\cdot S^0) = 11;$$

$$\nabla F(X^0 - 3\cdot S^0) = [-4 \quad 4 \quad 6]^t.$$

Производные функции в направлении поиска рассчитываются с помощью произведений векторов $S^t \cdot \nabla F(X)$, следовательно,

$$(S^0)^t \cdot \nabla F(X^0) = [1 \quad 0 \quad 0]\cdot [2 \quad 4 \quad 6]^t = 2;$$

$$(S^0)^t \cdot \nabla F(X^0 - 3\cdot S^0) = [1 \quad 0 \quad 0]\cdot [-4 \quad 4 \quad 6]^t = -4.$$

Теперь определим значения переменных интерполирующего полинома и его производной

$u_0 = 0,$ $\qquad u_1 = d_0 = -3,$

$v_0 = F(X^0) = 8,$ $\qquad v_1 = F(X^0 - 3\cdot S^0) = 11,$

$v_0' = (S^0)^t \cdot \nabla F(X^0) = 2,$ $\qquad v_0' = (S^0)^t \cdot \nabla F(X^0 - 3\cdot S^0) = -4.$

В соответствии с (11.25) и (11.26), получим $D = 0$ и $C = 1$. Интерполирующая кривая является параболой второй степени, что и следовало ожидать для данной задачи. Вместо (11.27) используем аналогичное выражение для параболы (11.21), откуда, как и в предыдущем примере, получим $d_{min} = -1$.

11.3 Квадратичные функции многих переменных

Выбор направления поиска, является наиболее сложной частью теории алгоритмов минимизации, при котором необходимо учитывать множество моментов. С одной стороны, желательно, чтобы алгоритм требовал расчета производных только первого порядка, так как производные более высокого порядка в общем случае рассчитать слишком сложно. С другой стороны, первые производные не содержат полной информации о кривизне функции. По этим причинам многие современные алгоритмы минимизации используют аппроксимацию вторых производных с помощью специальных формул. Для вывода таких формул необходимо понимать свойства квадратичных функций n переменных. В связи с этим, обсудим основные свойства квадратичных функций, что поможет разобраться и понять специфические моменты излагаемые в работах по оптимизации.

Произвольную дифференцируемую функцию можно разложить в ряд Тейлора и ограничиться квадратичными членами

$$F(X+\Delta X)=F(X)+(\Delta X)^t\cdot\nabla F(X)+1/2\cdot(\Delta X)^t\cdot G(X)\cdot\Delta X+\cdots. \quad (11.28)$$

Здесь $\nabla F(X)$ - градиент функции, определяемый выражением (11.8), а $G(X)$ - симметричная квадратная матрица называемая матрицей Гессе

$$G(X)=\begin{bmatrix} \frac{\partial^2 F}{\partial X_1\cdot\partial X_1} & \frac{\partial^2 F}{\partial X_1\cdot\partial X_2} & \cdots & \frac{\partial^2 F}{\partial X_1\cdot\partial X_n} \\ \frac{\partial^2 F}{\partial X_2\cdot\partial X_1} & \frac{\partial^2 F}{\partial X_2\cdot\partial X_2} & \cdots & \frac{\partial^2 F}{\partial X_2\cdot\partial X_n} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{\partial^2 F}{\partial X_n\cdot\partial X_1} & \frac{\partial^2 F}{\partial X_n\cdot\partial X_2} & \cdots & \frac{\partial^2 F}{\partial X_n\cdot\partial X_n} \end{bmatrix}. \quad (11.29)$$

В начале изучения квадратичных функций сделаем важное замечание, справедливое для любых дифференцируемых функций. Предположим, что дифференцируется функция $F(X)$, в заданном направлении S^k, и достигнут ее минимум в точке X^{k+1}. Градиент в точке должен быть ортогонален направлению поиска, т.е.

$$(S^k)^t\cdot\nabla F^{k+1}=(\nabla F^{k+1})^t\cdot S^k=0. \quad (11.30)$$

Предположим, что это неверно. В этом случае градиент должен иметь компоненту в направлении поиска, указывающую на то, что дальнейшее уменьшение функции еще возможно. Так как это противоречит предположению о том, что достигнут минимум, справедливость (11.30)

подтверждается. Эта ситуация подставлена на рисунке 11.2, где изображен фрагмент уровней трехмерной поверхности.

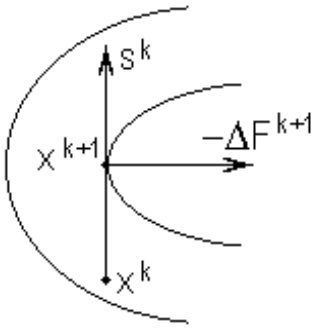


Рисунок 11.2 - Ортогональность векторов направления поиска s^k и градиента в точке минимума по этому направлению

Возвращаясь к квадратичной функции (11.28), предположим, что достигнута особая точка - экстремум $\hat{X}$, в которой градиент является нуль – вектором

$$\nabla F(\hat{X}) = 0. \qquad (11.31)$$

При этом (11.28) упрощается

$$F(\hat{X} + \Delta X) = F(\hat{X}) + 1/2 \cdot (\Delta X)^t \cdot G(\hat{X}) \cdot \Delta X. \qquad (11.32)$$

Произведение $(\Delta X)^t \cdot G(X) \cdot \Delta X$, называемое квадратичной формой, и определяет тип функции, с которой будем иметь дело.

Могут встретиться три возможных варианта значения этого произведения, определяемого второй производной $G(X)$:

1. Если квадратичная форма больше нуля для произвольно выбранного вектора ΔX, то эта точка является минимумом (возможно локальным).

2. Если квадратичная форма меньше нуля, для любых ΔX, то точка должна быть максимумом.

3. Если выбор ΔX может сделать квадратичную форму положительной, либо отрицательной, то точка $\hat{X}$ является седловой. Это название происходит от формы двумерной поверхности.

Эти три возможности определяются свойствами матрицы G, и для произвольной матрицы G размера $n * n$ вводятся следующее определение:

$$G = \begin{cases} положительно - определена, если : S^t \cdot G \cdot S > 0, \\ отрицательно - определена, если : S^t \cdot G \cdot S < 0, \\ неопределенная, если : S^t \cdot G \cdot S = 0. \end{cases} \qquad (11.33)$$

для всех ненулевых S.

Поскольку в дальнейшем будем рассматривать минимизацию функции, то положительная определенность матрицы G будет иметь первостепенное значение.

В качестве иллюстрации покажем, что матрица Гессе G, ранее рассматриваемой нами функции

$$F(X)=(x_2-x_1)^2+(1-x_1)^2,$$

положительно определена в точке минимума $X^*=[1 \quad 1]^t$. В соответствии с (11.29), матрица Гессе имеет вид

$$G=\begin{bmatrix} 4 & -2 \\ -2 & 2 \end{bmatrix},$$

и, в данном случае, не зависит от X. Выберем произвольное направление $S=[s_1 \quad s_2]$. Образуя квадратичную форму, видим, что

$$S^t \cdot G \cdot S=[s_1 \quad s_2]\cdot\begin{bmatrix} 4 & -2 \\ -2 & 2 \end{bmatrix}\cdot\begin{bmatrix} s_1 \\ s_2 \end{bmatrix}=2\cdot\left[(s_1-s_2)^2+s_1^2\right]>0,$$

для произвольного ненулевого вектора S.

Соотношение (11.33) устанавливает специальные свойства матрицы G, по отношению к вектору S.

Независимость векторов. Теперь определим основные свойства n векторов S^i, при $i=0,1,\ldots,n-1$, по отношению к положительно определенной матрице G, размера $n*n$. Если

$$(S^i)^t \cdot G \cdot S^j=\begin{cases} 0, при:\ i\neq j, \\ k_j>0, при:\ i=j, \end{cases} \tag{11.34}$$

то векторы S^i, S^j называют G- сопряженными и линейно независимыми. Линейная независимость может быть установлена по определению: n векторов линейно независимы, если уравнение

$$\sum_{j=0}^{n-1} a_j \cdot S^j=0 \tag{11.35}$$

удовлетворяется только при всех a_j, равных нулю. Для сопряженных векторов умножим уравнение (11.35) слева на $(S^i)^t \cdot G$. Как следует из (11.34), сумма должна быть равна $a_j \cdot k_j$.

Поскольку, по определению $k_j>0$, коэффициент a_j, должен быть равен нулю. Повторив это для всех $j=0,1,\ldots,n-1$, установим линейную независимость векторов.

Существует широкий класс G- сопряженных векторов, что демонстрируется в следующем простом примере.

Используя, матрицу Гессе из предыдущего примера

$$G = \begin{bmatrix} 4 & -2 \\ -2 & 2 \end{bmatrix},$$

найдем:

а) направление, сопряженное к $S^0 = [1 \quad 0]^t$;

b) направление, сопряженное к $S^0 = [1 \quad -1]^t$.

Для первого случая сформируем произведение

$$G \cdot S^0 = \begin{bmatrix} 4 & -2 \\ -2 & 2 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} -4 \\ 2 \end{bmatrix}.$$

Определим сопряженный вектор, как $S^1 = [a_1 \quad a_2]^t$ и потребуем, чтобы, $[a_1 \quad a_2] \cdot [-4 \quad 2]^t = 0$, откуда $a_2 = 2 \cdot a_1$. Выберем, например, вектор $S_1 = [1 \quad 2]^t$, который является решением.

Аналогично для второго случая сформируем произведение

$$G \cdot S^0 = \begin{bmatrix} 4 & -2 \\ -2 & 2 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} 6 \\ -4 \end{bmatrix}.$$

Определим вектор $S^1 = [b_1 \quad b_2]^t$ и потребуем, чтобы, $[b_1 \quad b_2] \cdot [6 \quad -4]^t = 6 \cdot b_1 - 4 \cdot b_2 = 0$, откуда $b_2 = 3/2 \cdot b_1$. Выбрав, например, $b_1 = 2$, получим, что сопряженное направление определяется вектором $S^1 = [2 \quad 3]^t$. Отсюда видно, что для любого ненулевого вектора можно найти G - сопряженный вектор.

Определим другие свойства, справедливые для произвольной квадратичной функции общего вида

$$F(X) = a + b^t \cdot X + 1/2 \cdot X^t \cdot G \cdot X, \qquad (11.36)$$

имеющей положительно определенную матрицу G. Градиент этой функции запишется

$$\nabla F + b + G \cdot X. \qquad (11.37)$$

Рассмотрим градиенты на двух соседних шагах итерации:

$$\nabla F^k = b + G \cdot X^k; \qquad \nabla F^{k+1} = b + G \cdot X^{k+1}.$$

Разность градиентов равна

$$\gamma^k = \nabla F^{k+1} - \nabla F^k = G \cdot (X^{k+1} - X^k). \qquad (11.38)$$

Возьмем аналогично разность двух соседних точек

$$X^{k+1} - X^k = \Delta X^k = d_k \cdot S^k, \qquad (11.39)$$

что соответствует выражению (11.15), где S^k - k - тое направление поиска. Подставив (11.39) в (11.38), получим

$$\gamma^k = \nabla F^{k+1} - \nabla F^k = G \cdot (X^{k+1} - X^k) = G \cdot \Delta X^k = d_k \cdot G \cdot S^k. \qquad (11.40)$$

В последующем рассмотрении будем предполагать, что длина шага d_k выбрана таким образом, что найден минимум в направлении S_k, т.е.

$d_k = d_{k\,min}$. Как следствие этого предположения, автоматическое выполнение условия ортогональности (11.30).

Преобразуем теперь (11.30) к следующему виду

$$\nabla F^{k+1} = \nabla F^k + d_k \cdot G \cdot S^k. \qquad (11.41)$$

Так как верхний индекс обозначает номер итерации, можно, повторно используя эту формулу, получить

$$\nabla F^{k+1} = \nabla F^{k-1} + d_{k-1} \cdot G \cdot S^{k-1} + d_k \cdot G \cdot S^k; \qquad (11.42)$$

$$\nabla F^{k+1} = \nabla F^{k-2} + d_{k-2} \cdot G \cdot S^{k-2} + d_{k-1} \cdot G \cdot S^{k-1} + d_k \cdot G \cdot S^k; \qquad (11.43)$$

и т.д.

Рассмотрим вектор - градиент (11.41), транспонируем его и помножим справа на вектор S^{k-1}, учитывая, что $G^t = G$

$$(\nabla F^{k+1})^t \cdot S^{k-1} = (\nabla F^k)^t \cdot S^{k-1} + d_k \cdot (S^k)^t \cdot G \cdot S^{k-1}.$$

Первый член в правой части равен нулю согласно (11.30). Если векторы S^k и S^{k-1} являются G- сопряженными, то второй член также равен нулю. Выполним аналогичные действия для (11.42), транспонируя и умножая на вектор S^{k-2}

$$(\nabla F^{k+1})^t \cdot S^{k-2} = (\nabla F^{k-1})^t \cdot S^{k-1} + d_{k-1} \cdot (S^{k-1})^t \cdot G \cdot S^{k-2} + d_k \cdot (S^k)^t \cdot G \cdot S^{k-2}$$

.

По тем же причинам заключаем, что правая часть этого равенства равна нулю.

Проводя далее аналогичные выкладки, получаем следующую общую формулу

$$(\nabla F^{k+1})^t \cdot S^j = 0, \qquad (11.44)$$

при $j = 0,1,\ldots,k$. Полагая, $k+1 = n$, можем записать

$$(\nabla F^n)^t \cdot S^j = 0, \qquad (11.45)$$

при $j = 0,1,\ldots,n-1$.

Так как векторы направлений поиска S^j, линейно независимы и уже все использованы, то градиент, на n- ом шаге, должен быть равен нулю $\nabla F^n = 0$.

Отсюда заключаем, что минимум, квадратичной положительно определенной функции, может быть достигнут, с помощью вышеописанного процесса, самое большое за n итераций. При этом предполагается, что на каждом направлении минимум находится за один шаг, а направления поиска линейно независимы.

Полученные результаты устанавливают некоторые общие правила, которых следует придерживаться при проведении процесса минимизации и при выборе направлений поиска. Выбор самих направлений пока не

излагался. Существует большое число различных вариантов выбора направлений поиска, которые излагается ниже.

11.4 Метода спуска при минимизации

Опишем некоторые хорошо известные методы, используемые при безусловной минимизации по мере их усложнения. Некоторые специальные детали приводятся без вывода и доказательства.

Метод наискорейшего спуска. Метод наискорейшего спуска самый ранний и наименее эффективный метод минимизации. При этом методе, в каждой точке рассчитывается вектор градиента, и направление поиска выбирается противоположно градиенту

$$S^k = -\nabla F^k. \qquad (11.46)$$

Метод имеет тенденции к колебаниям, если минимум представляет собой удлиненную изгибающуюся долину изображенную на рисунке 11.3.

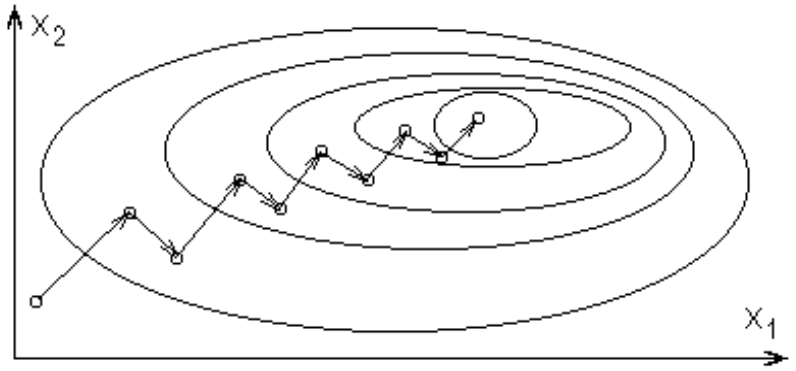


Рисунок 11.3 - Колебательный характер поиска в методе наискорейшего спуска

В результате сходимость метода наискорейшего спуска получается медленной. Позднее было предложено несколько модификаций этого метода, уменьшающих возможные колебания.

Метод сопряженного градиента. Этот метод использует информацию, полученную на предыдущих шагах, для определения нового направления. В предыдущем подразделе было показано, что для квадратичных функций можно найти ее минимум за n шагов, при условии, что в каждом направлении, минимум определяется точно.

Предполагается, что квадратичная функция также определяется выражением (11.28), однако информация о второй производной непосредственно не используется. Первое направление поиска выбирается так же, как и в методе наискорейшего спуска, а последующие направления являются линейными комбинациями вектора градиента и других выбранных предварительно направлений

$$
\begin{aligned}
&S^0 = -\nabla F^0;\\
&S^1 = -\nabla F^1 + k_{11}\cdot S^0;\\
&\ldots\ldots\ldots\ldots\ldots\ldots\\
&S^k = -\nabla F^k + \sum_{i=1}^{k} k_{ik}\cdot S^{i-1},
\end{aligned}
\qquad (11.47)
$$

где $1 \le k \le n-1$.

Предполагая законченную минимизацию в каждом направлении поиска, получим, как это следует из (11.30)

$$(S^i)^t \cdot \nabla F^{i+1} = 0, \qquad (11.48)$$

где $i = 0,1,\ldots,n-1$. Так как первое направление поиска определено, второе направление находится из условия G - сопряженности (11.34)

$$(S^1)^t \cdot G \cdot S^0 = (-\nabla F^1 + k_{11}\cdot S^0)^t \cdot G \cdot S^0 = 0. \qquad (11.49)$$

Коэффициент k_{11} можно рассчитать из этого выражения, если известна матрица G. Поскольку, в общем случае, она не известна, будем заменять матрицу G во всех выражениях значениями функции и или градиентов в соответствующих точках, предполагая квадратичность функции. В частности воспользуемся (11.40), переписав его в виде

$$G \cdot S^0 = (\nabla F^1 - \nabla F^0)/d_0. \qquad (11.50)$$

Подставляя (11.50) в (11.49) и, принимая во внимание (11.47), получаем

$$(-\nabla F^1 - k_{11}\cdot \nabla F^0)^t \cdot (\nabla F^1 - \nabla F^0)/d_0 = 0.$$

Сократив на не равный нулю множитель d_0, и перемножив скобки, получим

$$-(\nabla F^1)^t \cdot \nabla F^1 + (\nabla F^1)^t \cdot \nabla F^0 - k_{11}\cdot(\nabla F^0)^t \cdot \nabla F^1 + k_{11}\cdot(\nabla F^0)^t \cdot \nabla F^0 = 0.$$

Учитывая из (11.47), что, $-\nabla F^0 = S^0$, а также условие ортогональности направления поиска и сопряженного градиента в точке минимума (11.30) или (11.48) видим, что второе и третье слагаемые равны нулю. Это позволяет записать последнее соотношение в виде

$$k_{11}\cdot(\nabla F^0)^t \cdot \nabla F^0 = (\nabla F^1)^t \cdot \nabla F^1,$$

или

$$\alpha_1 = k_{11} = [(\nabla F^1)^t \cdot \nabla F^1]/[(\nabla F^0)^t \cdot \nabla F^0].$$

Выполнив аналогичные действия для следующих направлений, можно показать, что

$$\alpha_k = k_{kk} = [(\nabla F^k)^t \cdot \nabla F^k]/[(\nabla F^{k-1})^t \cdot \nabla F^{k-1}], \qquad (11.51)$$

а остальные коэффициенты k_{ik} равны нулю.

Итерационные шаги общего алгоритма, приведенные в предыдущем подразделе, остаются справедливыми, лишь третий шаг уточняется следующим образом: вычисляем α_k из (11.51) и, полагая

$$S^k = -\nabla F^k + \alpha_k \cdot S^{k-1}, \qquad (11.52)$$

производим его нормировку к единичной длине.

Если соотношения (11.51) и (11.52) применяются для не квадратичной функции, то наблюдается линейная сходимость до тех пор, пока направление поиска не начнет периодически повторятся. Это, как отмечалось, происходит после n шагов по направлениям $S^n = -\nabla F^n$.

Рассмотренный алгоритм прост для реализации и требует умеренный объем оперативной памяти, т. к. необходимо запоминать только предыдущее направление поиска и предыдущий градиент. Этот алгоритм часто используют для задач, имеющих большое число переменных. В литературе данный метод известен под названием метода сопряженного градиента Флетчера и Ривса.

Известны и другие версии метода сопряженного градиента с лучшей сходимостью, но требующие несколько большей памяти. При этом модификации подвергается формула определения направлений поиска (11.52) и изменяется последовательность шагов алгоритма.

Метод Ньютона. Итерационный метод, основанный на явном использовании вторых производных, известен под общим названием метод Ньютона. Пусть снова функция $F(X)$ разложена в ряд Тейлора и в нем удержано три члена. Результат разложения, представленный в выражении (11.28), перепишем в виде

$$F(X^k + \Delta X) - F(X^k) = (\Delta X)^t \cdot \nabla F^k + 1/2 \cdot (\Delta X)^t \cdot G^k \cdot \Delta X \qquad (11.53)$$

Пусть требуется минимизировать разность, стоящую в левой части. Это можно сделать дифференцированием (11.53), по ΔX и приравнивая результата к нулю

$$\partial [F(X^k + \Delta X) - F(X^k)] / \partial \Delta X = \nabla F^k + G^k \cdot \Delta X = 0,$$

откуда

$$G^k \cdot \Delta X = -\nabla F^k.$$

Это уравнение можно решить соответствующими методами относительно ΔX, например, с помощью LU- разложения. Формально решение можно записать

$$\Delta X = -(G^k)^{-1} \cdot \nabla F^k = -H^k \cdot \nabla F^k,$$

где $H = G^{-1}$. Направление поиска теперь полагаем совпадающим с вектором

$$S^k = \Delta X^k = -H^k \cdot \nabla F^k, \qquad (11.54)$$

и вновь повторяем общий алгоритм, изложенный в предыдущем подразделе.

При подходе к минимуму матрица Гессе G^k будет положительно определенной и можно использовать полный размер шага $d_k = 1$, т.е. не нужен поиск в направлении S^k. Однако вдали от минимума матрица Гессе

может и не быть положительно определенной. Более того, вычисление этой матрицы требует больших затрат, поэтому разработан целый класс других методов, называемых **методами с переменной метрикой или квазиньютоновскими**, которые лишены этого недостатка.

Методы с переменной метрикой. Эти методы были разработаны сравнительно давно, однако обобщены в последнее время. Они базируются на оценке градиентов и на аппроксимации матрицы Гессе или обратной от матрицы Гессе. Аппроксимация достигается преобразованием исходной положительно, определенной матрицы или тождественной ей матрицы таким образом, чтобы сохранить ее положительную определенность. Только при достижении минимума получаемая матрица аппроксимирует матрицу Гессе или обратную к ней.

Детальный вывод аппроксимирующих соотношений сложен и здесь не приводится, а используются лишь готовые выражения. Во всех методах направление поиска определяется, как и в методе Ньютона

$$S^k = -H^k \cdot \nabla F^k. \tag{11.55}$$

На каждой итерации по матрице H^k, согласно специальной формуле, получают матрицу H^{k+1}. Существует много формул для определения H^{k+1}. Приведем эту формулу в наиболее общей форме

$$H^{k+1} = f(\alpha^k, \phi^k, H^k, \Delta X^k, \gamma^k), \tag{11.56}$$

где

$$f(\alpha,\phi,H,\Delta X,\gamma) = \alpha \cdot H + \left[1 + \alpha \cdot \phi \cdot \frac{\gamma^t \cdot H \cdot \gamma}{(\Delta X)^t \cdot \gamma}\right] \cdot \frac{(\Delta X)\cdot(\Delta X)^t}{(\Delta X)^t \cdot \gamma} -$$
$$- \alpha \cdot \frac{1-\phi}{\gamma^k \cdot H \cdot \gamma} \cdot H \cdot \gamma \cdot \gamma^t \cdot H - \frac{\alpha \cdot \phi}{(\Delta X)^t \cdot \gamma} \cdot \left[(\Delta X)\cdot \gamma^t \cdot H + H \cdot \gamma \cdot (\Delta X)^t\right] \tag{11.57}$$

Эта прямая формула пригодна только в случае, когда $(\Delta X)^t \cdot \gamma \neq 0$ и $\gamma^t \cdot H \cdot \gamma \neq 0$. Здесь α и ϕ - скалярные величины, а $\gamma^k = \nabla F^{k+1} - \nabla F^k$. Другие хорошо известные формы, следуют из (11.57) при выборе соответствующих значений α и ϕ.

Например:

1. Если $\alpha^k \equiv 1,\ \phi^k \equiv 0$, то

$$f(1,0,H,\Delta X,\gamma) = H + \frac{(\Delta X)\cdot(\Delta X)^t}{(\Delta X)^t \cdot \gamma} - \frac{H \cdot \gamma \cdot \gamma^t \cdot H}{\gamma^t \cdot H \cdot \gamma}. \tag{12.58}$$

Эта формула получена Дэвидоном, Флетчером и Пауэлом и ее иногда называют ДФП - формулой.

2. Если $\alpha^k \equiv 1,\ \phi^k \equiv 1$, то

$$f(1,1,H,\Delta X,\gamma)=H+\left[1+\frac{\gamma^t\cdot H\cdot\gamma}{(\Delta X)^t\cdot\gamma}\right]\cdot\frac{(\Delta X)\cdot(\Delta X)^t}{(\Delta X)^t\cdot\gamma}-$$
$$-\frac{1}{(\Delta X)^t\cdot\gamma}\cdot\left[(\Delta X)\cdot\gamma^t\cdot H+H\cdot\gamma\cdot(\Delta X)^t\right]. \qquad (11.59)$$

Эта формула была получена независимо Бройденом, Флетчером, Гольдфарбом и Шенно и ее обычно называют БФГШ - формулой.

Существуют также и другие упрощения. Современные программы часто основываются на комбинации выражений (11.58) и (11.59). На каждой итерации правила решения заданы таким образом, что выбирается та или иная формула.

11.5 Минимизация при ограничениях

Основная задача минимизации была сформулирована в первом подразделе, где было введены понятие целевой функции и ограничения типа равенств и неравенств. Во втором подразделе обсуждалось решение задачи минимизации Лагранжа с ограничениями типа равенств. Здесь рассмотрим методы минимизации при наличии ограничений обоих видов, используя обозначения, введенные в первом подразделе.

Одна из возможностей решения задачи минимизации с ограничениями состоит в преобразовании задачи с ограничениями в задачу без ограничений, а затем в использовании любого из методов безусловной минимизации. Широко известный, но устаревший метод состоит в определении новой функции при введении штрафных коэффициентов. Упомянем два таких метода:

а) метод внутренней точки формирует новую функцию вида

$$P(X,r)=F(X)+r\cdot\sum_{j=1}^{k_1}1/g_j(X)+1/\sqrt{r}\cdot\sum_{i=1}^{k_1}[e_i(X)]^2;$$

б) метод внешней точки формирует функцию

$$P(X,r)=F(X)+1/r\cdot\sum_{j=1}^{k_1}[min(g_j(X),0)]^2+1/r\cdot\sum_{i=1}^{k_1}[e_i(X)]^2.$$

В обоих случаях коэффициент r вначале выбирается близким к единице и проводится минимизация, до тех пор, пока не будет, достигнут оптимума. На следующем шаге коэффициент r уменьшается, примерно на порядок, и минимизация проводится вновь. При повторном выполнении таких шагов и уменьшении коэффициента r комбинированная целевая функция увеличивается (штрафуется), если решение имеет тенденцию к нарушению ограничений. В конечном итоге, при очень малом r, функция минимизируется и ограничения выполняются.

Оба метода имеют тот недостаток, что при приближении к решению

задача становится плохо обусловленной. Чтобы этого избежать, современные методы используют подход, базирующийся на множителях Лагранжа. Принципы этих методов будут показаны на задачах с ограничениями типа равенств. Ограничения типа неравенств рассмотрим позже. При выводе соотношений сосредоточим основное внимание на существе методов, а не на деталях программирования.

Рассмотрим следующую задачу: найти минимум функции $F = F(X)$, при ограничениях $e_i(X) = 0$, где $i = 1,2,\ldots,k$. Следуя соотношениям, введенным во втором подразделе, введем функцию Лагранжа

$$L(X,\Lambda) = F(X) - \sum_i \lambda_i \cdot e_i(X) = F(X) - \Lambda^t \cdot E(X) = F(X) - E^t(X) \cdot \Lambda, \tag{11.60}$$

где $E = \begin{bmatrix} e_1 & e_2 & \cdots & e_{k_1} \end{bmatrix}^t$; $\Lambda = \begin{bmatrix} \lambda_1 & \lambda_2 & \cdots & \lambda_{k_1} \end{bmatrix}^t$. Выбор того или иного варианта соотношения (11.60) определяется удобством.

Дифференцируя (11.60) по X и Λ, и приравнивая производные нулю, находим систему уравнений определяющих минимум

$$\begin{aligned} \partial L / \partial X &= \nabla F - N(X) \cdot \Lambda = 0; \\ \partial L / \partial \Lambda &= -E(X) = 0, \end{aligned} \tag{11.61}$$

где

$$N = \begin{bmatrix} \nabla e_1 & \nabla e_2 & \cdots & \nabla e_{k_1} \end{bmatrix} = \begin{bmatrix} \partial e_1 / \partial x_1 & \partial e_2 / \partial x_1 & \cdot & \partial e_{k_1} / \partial x_1 \\ \partial e_1 / \partial x_2 & \partial e_2 / \partial x_2 & \cdot & \partial e_{k_1} / \partial x_2 \\ \cdot & \cdot & \cdot & \cdot \\ \partial e_1 / \partial x_n & \partial e_2 / \partial x_n & \cdot & \partial e_{k_1} / \partial x_n \end{bmatrix}. \tag{11.62}$$

Систему (11.61) можно решить, например, методом Ньютона – Рафсона. Якобиан при этом, согласно (11.62), имеет вид

$$M = \begin{bmatrix} M_1 & -N \\ -N^t & 0 \end{bmatrix}, \tag{11.63}$$

где элементы матрицы M_1 равны

$$m_{jl} = \frac{\partial^2 F}{\partial x_j \cdot \partial x_l} - \sum_{i=1}^{k_1} \lambda_i \cdot \frac{\partial^2 e_i}{\partial x_j \cdot \partial x_l}. \tag{11.64}$$

На k- ом шаге итерации система уравнений, решаемая методом Ньютона – Рафсона, запишется

$$M^K \cdot \begin{bmatrix} \Delta X^k \\ \Delta \Lambda^k \end{bmatrix} = -\begin{bmatrix} \nabla F^k - N^k \cdot \Lambda^k \\ -E^k \end{bmatrix}. \tag{11.65}$$

В первом уравнении системы (11.65) можно провести сокращения. Так, если переписать систему (11.65), в виде двух уравнений, раскрывая Якобиан M^k и заменяя $\Delta\Lambda^k = \Lambda^{k+1} - \Lambda^k$:

$$M_1^k \cdot \Delta X^k - N^k \cdot (\Lambda^{k+1} - \Lambda^k) = -\nabla F^k + N^k \cdot \Lambda^k;$$
$$-(N^k)^t \cdot \Delta X^k = E^k,$$

то, сократив в первом уравнении системы член $N^k \cdot \Lambda^k$, получим новую систему уравнений вида

$$\begin{bmatrix} M_1^k & -N^k \\ -(N^k)^t & 0 \end{bmatrix} \cdot \begin{bmatrix} \Delta X^k \\ \Lambda^{k=1} \end{bmatrix} = \begin{bmatrix} -\nabla F^k \\ E^k \end{bmatrix}, \tag{11.66}$$

в которой все изменения вектора Λ^{k+1} обусловлены влиянием коэффициентов λ_i^k, входящих в элементы матрицы M_1.

Можно показать, что точно такие же уравнения можно получить, если:

a) разложить функцию $F(X^k + \Delta X^k)$ в ряд Тейлора и удержать первые три члена;

b) разложить каждую из функций $e_i(X^k + \Delta X^k)$ в ряд Тейлора и удержать первые два члена.

Разложение функции $F(X^k + \Delta X^k)$ было получено в (11.28)

$$F(X^k + \Delta X^k) = F^k + (\Delta X^k)^t \cdot \nabla F^k + 1/2 \cdot (\Delta X^k)^t \cdot G^k \cdot \Delta X^k. \tag{11.67}$$

Разложение для ограничений типа равенств можно представить в виде

$$E(X^k + \Delta X^k) = E^k + (N^k)^t \cdot \Delta X^k, \tag{11.68}$$

и упрощенную квадратичную задачу с ограничениями можно представить следующим образом:

найти минимум функции

$$F(X^k) + (\Delta X^k)^t \cdot \nabla F^k + 1/2 \cdot (\Delta X^k)^t \cdot G^k \cdot \Delta X^k; \tag{11.69}$$

при ограничениях

$$E(X^k) + (N^k)^t \cdot \Delta X^k = 0. \tag{11.70}$$

Если теперь сформировать Лагранжиан по типу (11.60), используя (11.69) и (11.70), то получим

$$L(X^k, \Lambda^k) = F(X^k) + (\Delta X^k)^t \cdot \nabla F^k + 1/2 \cdot (\Delta X^k)^t \cdot G^k \cdot \Delta X^k - \\ - [E(X^k) + (N^k)^t \cdot \Delta X^k]^t \cdot \Lambda^k. \tag{11.71}$$

Дифференцируя (11.71), по ΔX^k, Λ^k и, приравнивая производные нулю, получим систему уравнений, определяющих минимум

$$\begin{bmatrix} G^k & -N^k \\ -(N^k)^t & 0 \end{bmatrix} \cdot \begin{vmatrix} \Delta X^k \\ \Lambda^k \end{vmatrix} = \begin{vmatrix} -\nabla F^k \\ E^k \end{vmatrix}, \tag{11.72}$$

и аналогичную системе (11.66). На каждой итерации общую формулировку задачи можно заменить упрощённой.

В общем случае, матрица G^k может и не быть положительно определённой. Для того чтобы обеспечить уменьшение целевой функции на каждом шаге, целесообразно, как и в случае безусловной минимизации,

заменить эту матрицу положительно определенной, рассчитанной по известному градиенту $F(X)$ и вектору $E(X)$. Для этой цели можно использовать формулы (11.58) и (11.59).

Поскольку при итерационном методе Ньютона - Рафсона могут возникать проблемы сходимости, если начальное приближение сильно отличается от решения, вместо исходной задачи целесообразно решать квадратичную, т.е. упрощенную задачу определяемую уравнениями (11.69) и (11.70). В этом случае вычисляются векторы ΔX^k, Λ^k и рассчитывается новая точка с помощью подстановки

$$X^{k+1} = X^k + \Delta X^k. \qquad (11.73)$$

После каждого шага вычисляется Лагранжиан, по формуле (11.71). Если он уменьшился, то начинается новая итерация, в противном случае предпринимается поиск в направлении ΔX^k, до нахождения точки минимума в этом направлении. Детали, связанные с квадратичной задачей и наилучшими методами поиска здесь не излагаются.

Используя соответствие между исходной задачей и квадратичной аппроксимацией, можно, аналогичным образом, рассмотреть проблему минимизации при ограничениях типа неравенств.

Рассмотрим следующую общую задачу: найти минимум функции $F(X)$, при ограничениях типа равенств $e_i(X)=0$ и неравенств $g_j(X)\geq 0$, где $i=1,2,\ldots k$; $j=1,2,\ldots k$. Сформируем Лагранжиан, используя два набора множителей Лагранжа

$$L(X,\Lambda,M) = F(X) - \lambda^t \cdot E(X) - \mu^t \cdot G(X), \qquad (11.74)$$

и исходную задачу, можно, на каждом итерационном шаге, аппроксимировать ее квадратичной задачей и найти минимум функции

$$F(X^k) + (\Delta X^k)^t \cdot \nabla F^k + 1/2 \cdot (\Delta X^k)^t \cdot G^k \cdot \Delta X^k, \qquad (11.75)$$

при ограничениях типа равенств и неравенств

$$E^k + (N_1^k)^t \cdot \Delta X^k = 0, \qquad (11.76)$$

$$G^k + (N_2^k)^t \cdot \Delta X^k \geq 0. \qquad (11.77)$$

На каждом шаге итерации коэффициенты μ_i должны сохраняться положительными, для того чтобы выполнялось условие $g_j(X)\geq 0$. Этот метод реализуется в некоторых известных программах оптимизации.

11.6 Алгоритмы оптимизации

Методы оптимизации можно разделить на два класса: методы, основанные на использовании только оценки целевой функции – прямые методы и методы, основанные на применении наряду с оценками целевой функции и информации о её градиенте – градиентные методы.

Преимуществами прямых методов являются: простота реализации, а также возможность использования их в случаях, когда градиент неизвестен

или не существует, например, для целевых функций имеющих разрывы. Основной недостаток прямых методов – снижение скорости сходимости вблизи минимума.

Преимуществом градиентных методов является высокая скорость сходимости вблизи минимума, а недостатками – сложность вычисления градиентов и сложность реализации самих методов.

Следует отметить, что методы оптимизации, в которых градиент не вычисляется в чистом виде, а используются различные оценки, например численное дифференцирование, строго говоря, относятся к прямым методам поиска оптимума.

Обычным приемом оценки эффективности метода оптимизации служит сравнение числа обращений к вычислению целевой функции и его градиента с теми же показателями для других методов. Кроме того, большинство описываемых в литературе методов, находят ближайший к начальной точке локальный минимум и для того, чтобы убедится в существовании других минимумов, в том числе и глобального, процесс поиска повторяют из разных начальных точек, используя для их задания опыт разработчика или датчик случайных чисел.

Опишем несколько методов поиска оптимума, описанных в литературе и получивших широкое распространение на практике.

Метод последовательного перебора параметров. По существу методы поиска простейшего типа заключаются в изменении каждый раз одной переменной при фиксации других, до тех пор, пока не будет, достигнут минимум в этом направлении. Затем это найденное значение переменной фиксируется и изменяется следующая переменная и т.д. Однако такой алгоритм работает плохо, если имеет место взаимодействие переменных, т.е. если в выражение для целевой функции входят члены содержащие произведение переменных, что чаще всего и наблюдается на практике. Таким образом, этот метод нельзя рекомендовать для практического применения.

Метод Хука - Дживса. Одним из наиболее простых и легко осваиваемых прямых методов поиска является метод Хука - Дживса. Хук и Дживс предложили логически простую стратегию поиска, использующую априорные сведения о характере топологии , целевой функции, постоянно обновляющиеся в процессе поиска. Суть метода заключается в попеременном применении двух типов поиска: исследующий поиск и поиск по образу. Исследующий поиск предназначен для выбора приемлемого направления поиска из текущей точки пространства параметров, движение по которому обеспечит минимизацию функции ошибки или, то же самое, целевой функции $F(X)$. Поиск по образцу используется для определения минимума целевой функции по направлению определяемому исследующим поиском.

Алгоритм Хука – Дживса, с использованием одномерной минимизации. Хук и Дживс предположили метод, не содержащий

одномерной минимизации, а использующий постоянные шаги по направлениям поиска. Этот вариант метода будет рассмотрен позднее. Здесь рассмотрим непрерывный вариант метода, использующий одномерную минимизацию вдоль координатных направлений $S^1, S^2, \ldots, S^n$ и направлений поиска по образцу.

Начальный этап. Выбрать число $\varepsilon > 0$ для остановки алгоритма. Выбрать исходные направления поиска $S^1, S^2, \ldots, S^n$, например, совпадающие со столбцами единичной матрицы. Выбрать начальную точку X^1, положить $Y^1 = X^1$, $j = k = 1$, где j – номер направления поиска, k – номер итерации и перейти к основному этапу.

Основной этап:

Шаг 1. Вычислить d_j – оптимальное решение задачи минимизации $F(Y^j + d_j \cdot S^j)$ в заданном направлении и положить $Y^{j+1} = Y^j + d_j \cdot S^j$. Если $j < n$, то $j = j + 1$ и повторить шаг 1. Если $j = n$, то положить $X^{k+1} = Y^{n+1}$ – последнее удачное решение. Если $\left\| X^{k+1} - X^k \right\| < \varepsilon$, то остановиться; в противном случае перейти к шагу 2.

Шаг 2. Положить $S^{k+1} = X^{k+1} - X^k$ и найти $\hat{\lambda}$ - оптимальное решение задачи минимизации $F(X^{k+1} + \lambda_{k+1} \cdot S^{k+1})$ в заданном направлении. Положить $Y^1 = X^{k+1} + \hat{\lambda} \cdot S^{k+1}$, $j = 1$, заменить $k = k + 1$ и перейти к шагу 1.

Анализ алгоритма оптимизации Хука - Дживса показывает, что минимизация целевой функции в заданном направлении используются, как на первом, так и на втором шаге основного этапа для нахождения следующей точки траектории поиска, поэтому ее целесообразно реализовать в виде отдельной процедуры.

Метод Хука - Дживса с дискретным шагом. Первоначально в методе Хука - Дживса не предполагалась одномерная минимизация. Одномерный поиск заменялся простой логической схемой, включающей вычисления целевой функции в процессе исследующего поиска и поиска по образцу. Приведем этот вариант алгоритма.

Начальный этап. Задать в качестве $S^1, S^2, \ldots, S^n$, координатные направления, например, совпадающие со столбцами единичной матрицы. Выбрать число $\varepsilon > 0$ для остановки алгоритма, начальный шаг $\Delta \geq \varepsilon$ и ускоряющий множитель $\alpha > 0$, обычно $\alpha \cong 2$. Выбрать начальную точку X^1. Положить $Y^1 = X^1$, $j = k = 1$, где j – номер направления поиска, k – номер итерации и перейти к основному этапу.

Основной этап:

Шаг 1. Если $F(Y^j + \Delta \cdot S^j) < F(Y^j)$, то шаг считается успешным; положить $Y^{j+1} = Y^j + \Delta \cdot S^j$ и перейти к шагу 2. Если $F(Y^j + \Delta \cdot S^j) \geq F(Y^j)$, то шаг считается неудачным. В этом случае, если $F(Y^j - \Delta \cdot S^j) < F(Y^j)$, то положить $Y^{j+1} = Y^j + \Delta \cdot S^j$ и перейти к шагу 2, если же $F(Y^j - \Delta \cdot S^j) \geq F(Y^j)$, то положить $Y^{j+1} = Y^j$ и перейти к шагу 2.

Шаг 2. Если $j < n$, то заменить $j = j + 1$ и вернуться к шагу 1. В противном случае, если $F(Y^{n+1}) < F(X^k)$, т.е. последний шаг был успешным, то перейти к шагу 3, а если $F(Y^{n+1}) \geq F(X^k)$, т. е. последний шаг был неуспешным, то перейти к шагу 4.

Шаг 3. Положить $X^{k+1} = Y^{n+1}$ – последнему удачному решению, $Y^1 = X^{k+1} + \alpha \cdot (X^{k+1} - X^k)$. Заменить $k = k + 1$, положить $j = 1$ и перейти к шагу 1.

Шаг 4. Если $\Delta < \varepsilon$, то остановиться; X^k решение. В противном случае, заменить Δ на $\Delta / 2$, положить $Y^1 = X^k$, $X^{k+1} = X^k$, т.е. вернуться в исходную точку. Заменить $k = k + 1$, положить $j = 1$ и вернуться к шагу 1.

Легко видеть, что описанные выше шаги 1 и 2 осуществляют исследующий поиск, а шаг 3 является ускоряющим шагом по направлению $X^{k+1} - X^k$. Заметим, что решение относительно того, делать ускоряющий шаг или нет, не применяется до тех пор, пока не будет выполнен исследующий поиск. На шаге 4 длина шага Δ сокращается. Алгоритм может быть легко модифицирован так, что по разным направлениям будут использоваться различные шаги Δ_i. Эта модификация иногда используется с целью масштабирования, что облегчает достижение минимума.

Анализ алгоритма оптимизации Хука - Дживса показывает, что исследующий поиск используются, как для определения направления поиска из базовой точки, так и в процессе поиска по образцу, поэтому его целесообразно реализовать в виде отдельной процедуры.

Основным недостатком данного метода является невозможность получения в некоторых случаях результата с помощью исследующего поиска, даже при значительном уменьшении шага, особенно вблизи минимума. Это происходит в непосредственной окрестности минимума, когда направление на него не совпадает ни с одной из осей пространства параметров. В таких точках процесс поиска может зайти в тупик даже для очень маленьких приращений Δ_j. При поиске в пространстве большой размерности метод Хука - Дживса не обеспечивает, сколько – ни будь значительного уменьшения целевой функции.

Необходимо заметить, что те же самые проблемы возникают и при использовании многих описанных в литературе методов, поэтому, когда

предполагают, что минимум достигнут, выбирают другую начальную точку и поиск минимума повторяют вновь. Только при совпадении минимумов целевой функции можно считать решение найденным.

Окончание поиска по методу Хука – Дживса осуществляется при длине шагов Δ_j меньше заданного значения. Если значение целевой функции в конечной точке не удовлетворяет проектировщика, то он должен или повторить процесс оптимизации, задавшись новыми начальными значениями параметров, или модифицировать принципиальную схему для получения желаемого результата.

Метод Розенброка. Метод Розенброка относится к прямым методам поиска, так как для выбора направлений поиска не использует непосредственно вычисления градиентов целевой функции, однако, как и метод сопряженных направлений использует набор взаимно ортогональных, а точнее ортонормированных направлений поиска. Кроме того, в процессе поиска производится адаптивный поворот системы координат, что делает стратегию поиска весьма гибкой. В качестве исходной ортонормированной матрицы направлений, обычно берется единичная матрица, столбцы которой выступают в качестве независимых направлений поиска. После исчерпания поиска по заданным направлениям, на основе линейной комбинации старых направлений и процедуры ортогонализации Грамма - Шмидта строится новый набор - матриц ортонормированных направлений и так до тех пор, пока не выполнятся условия окончания поиска оптимума.

В первоначальном варианте метода Розенброка не использовалась одномерная минимизация по направлению, а применялись дискретные шаги вдоль независимых направлений поиска. Здесь вначале изложим непрерывный вариант метода с применением одномерной минимизации, затем остановимся на дискретном варианте метода Розенброка.

Непрерывный вариант алгоритма Розенброка. В непрерывном варианте, на каждой итерации процедура осуществляет итеративный поиск вдоль n линейно независимых ортогональных направлений. После получения новой точки в конце итерации строятся новое множество ортогональных векторов.

Пусть $S^1, S^2, \ldots, S^n$ - линейно независимые и ортонормированные векторы, т.е. $S_i^t \cdot S_j = 0$, при $i \neq j$. Начиная из текущей точки X^k, целевая функция последовательно минимизируется вдоль каждого из направлений, в результате чего, получается точка X^{k+1}. Так, новое значение вектора параметров, определяется выражением

$$X^{k+1} = X^k + \sum_{j=1}^{n} d_j \cdot S^j,$$

где d_j – длина шага по направлению S^j. Новый набор направлений $\hat{S}^1, \hat{S}^2, \ldots, \hat{S}^n$ строится с помощью процедуры Грамма – Шмидта следующим образом:

$$A^j = \begin{cases} S^j, & \text{если } d_i = 0 \\ \sum_{i=j}^{n} d_i \cdot S^i, & \text{если } d_i \neq 0, \end{cases}$$

$$B^j = \begin{cases} A^j, & \text{при } j = 0, \\ A^j - \sum_{i=1}^{j-1} \left[(A^j)^t \cdot \hat{S}^i \right] \cdot \hat{S}^i, & \text{при } j \geq 2, \end{cases} \qquad (11.78)$$

$$\hat{S}^j = B^j / \| B^j \|,$$

где $\| B^j \| = [(B^j)^t * B^j]^{1/2}$ - норма вектора.

Пояснений требуют лишь формулы построения ортонормированных векторов B^j.

Так, в соответствии с процедурой ортогонализации Грамма – Шмидта, если известен набор векторов $A^1, A^2, \ldots, A^n$, то новая система ортонормальных векторов B^j строится как линейная комбинация этих векторов:

$$B^1 = A^1,$$
$$B^2 = A^2 + \lambda_{21} \cdot A^1,$$
$$\ldots\ldots\ldots\ldots$$
$$B^j = A^j + \sum_{i=1}^{j-1} \lambda_{ji} \cdot A^i.$$

Первый вектор новой системы берется, согласно (11.78), совпадающим с первым вектором старой системы $B^1 = A^1$.

Следующие вектора определены как линейная комбинация предыдущих векторов старой системы. Для определения коэффициента λ_{21} второго уравнения образуем произведение $(B^2)^t \cdot B^1$, в результате имеем

$$(B^2)^t \cdot B^1 = (A^2 + \lambda_{21} \cdot A^1)^t \cdot B^1 = 0.$$

Отсюда получаем

$$\lambda_{21} = -[(A^2)^t \cdot B^1] / [(A^1)^t \cdot B^1] = -[(A^2)^t \cdot B^1] / [(B^1)^t \cdot B^1].$$

В результате второй вектор новой системы запишется

$$B^2 = A^2 - [(A^2)^t \cdot B^1] / [(B^1)^t \cdot B^1] \cdot A^1,$$

или учитывая, что $A^1 = B^1$ и выражение для нормы вектора, можем записать

$$B^2 = A^2 - [(A^2)^t \cdot \hat{B}^1] \cdot \hat{B}^1.$$

Производя аналогичные операции с последующими уравнениями, получаем следующее обобщенное выражение

$$B^j = A^j - \sum_{i=1}^{j-1} [(A^j)^t * \hat{B}^i] * \hat{B}^i,$$

что полностью объясняет процедуру формирования новой системы векторов из (11.78). Нормировку, вновь образованных векторов, можно производить последовательно в процессе получения либо в конце процедуры.

Алгоритм Розенброка с линейным поиском по направлению. Приведем теперь алгоритм Розенброка, использующий линейный поиск по направлению для минимизации функций нескольких переменных.

Начальный этап. Пусть $\varepsilon > 0$ – скаляр, используемый в условии остановки. Выбираем в качестве $S^1, S^2, \ldots, S^n$ исходные направления поиска, обычно это столбцы единичной матрицы. Задаёмся начальным значением X^1, полагаем $Y^1 = X^1$, $j = k = 1$, где j – номер направления поиска, k – номер итерации и переходим к основному этапу.

Основной этап:

Шаг 1. Найти d_j - оптимальное решение задачи минимизации $F(Y^j + d_j \cdot S^j)$ в заданном направлении и положить $Y^{j+1} = Y^j + d_j \cdot S^j$. Если $j < n$, то $j = j + 1$ и вернуться к шагу 1, иначе перейти к шагу 2.

Шаг 2. Положить $X^{k+1} = Y^{n+1}$ - последний вектор, найденный на первом шаге. Если $\left\| X^{k+1} - X^k \right\| < \varepsilon$, то остановиться, в противном случае положить $Y^1 = X^{k+1}$, заменить $k = k + 1$, положить $j = 1$ и перейти к шагу 3.

Шаг 3. Построить новое множество линейно независимых и ортонормированных направлений в соответствии с (11.78). Обозначить новые направления через $S^1, S^2, \ldots, S^n$ и вернуться к шагу 3.

Как видно из приведенного алгоритма, при его реализации необходимо запоминать текущие направления и значения оптимальных шагов одномерных поисков, из которых в соответствии с первой частью выражений (11.78), строятся предварительные вектора новых направлений, ортогонализуемые затем с помощью процедуры Грамма - Шмидта.

Дискретный вариант алгоритма Розенброка. Теперь рассмотрим алгоритм Розенброка с дискретным шагом. Как уже отмечалось, предложенный Розенброком метод не использует одномерную минимизацию, вместо этого

по ортогональным направлениям делаются дискретные шаги, длина которых меняется в зависимости от знания целевой функции в текущей точке. Приведем алгоритм этого варианта метода.

Начальный этап. Выбрать число $\varepsilon > 0$, для остановки алгоритма, коэффициент растяжения $\alpha > 1$, обычно $\alpha \cong 3$ и коэффициент сжатия $\beta \subseteq (-1, 0)$, обычно $\beta \cong -0.5$. Взять в качестве $S^1, S^2, \ldots, S^n$ исходные направления поиска, обычно это столбцы единичной матрицы и выбрать $\hat{\Delta}_1, \hat{\Delta}_2, \ldots, \hat{\Delta}_n > 0$ - начальную длину шага вдоль каждого из направлений, например 5% от номинального значения. Выбрать начальную точку X^1, положить $Y^1 = X^1$, $j = k = 1$, где j - номер направления поиска, k - номер итерации. Положить $\Delta_j = \hat{\Delta}_j$ для всех j и перейти к основному этапу.

Основной этап:

Шаг 1. Если $F(Y^j + \Delta_j \cdot S^j) < F(Y^j)$, то шаг по j - му направлению считается успешным. Положить $Y^{j+1} = Y^j + \Delta_j \cdot S^j$ и заменить Δ_j на $\alpha \cdot \Delta_j$. Если же $F(Y^j + \Delta_j \cdot S^j) \geq F(Y^j)$, то шаг считается неудачным, при этом положить $Y^{j+1} = Y^j$ и Δ_j заменить на $\beta \cdot \Delta_j$. Если $j < n$, то $j = j + 1$ и повторить шаг 1, в противном случае, т.е. при $j = n$, перейти к шагу 2.

Шаг 2. Если $F(Y^{n+1}) < F(Y^1)$, т.е. если хотя бы один спуск по направлениям на шаге 1 оказался успешным, то положить $Y^1 = Y^{n+1}$ - последний вектор найденный на первом шаге, $j = 1$ и повторить шаг 1. Пусть $F(Y^{n+1}) = F(Y^1)$, т.е. если каждый из n последних спусков по направлениям на шаге 1 были неудачными. Если $F(Y^{n+1}) > F(Y^1)$, т.е. по крайней мере, один удачный спуск встретился в течение k - той итерации на шаге 1, то перейти к шагу 3. Если $F(Y^{n+1}) = F(X^k)$, т.е. не было ни одного удачного спуска по направлениям, то остановиться. При этом X^k является приближенным оптимальным решением, если $|\Delta_j| < \varepsilon$ для всех j. В противном случае положить $Y^1 = Y^{n+1}$ – последний вектор найденный на первом шаге, $j = 1$ и перейти к шагу 1.

Шаг 3. Положить $X^{k+1} = Y^{n+1}$ - последний вектор, найденный на первом шаге. Если $\left\| X^{k+1} - X^k \right\| < \varepsilon$, то остановиться; в противном случае, вычислить $d_1, d_2, \ldots, d_n$ из соотношения $X^{k+1} = X^k + \sum_{j=1}^{n} d_j \cdot S^j$, построить новые направления, в соответствии с (11.78), обозначить их

через $S^1, S^2, \ldots, S^n$, положить $\Delta_j = \hat{\Delta}_j$ для всех j, положить $Y^1 = X^{k+1}$, заменить $k = k+1$, положить $j = 1$ и перейти к шагу 1.

Заметим, что дискретные шаги выбираются вдоль n направлений поиска на шаге 1. Если движение вдоль S^j оказалось успешным, то Δ_j заменяется на $\alpha \cdot \Delta_j$, если же на этом направлении постигла неудача, то Δ_j заменяется на $\beta \cdot \Delta_j$. Так как $\beta < 0$, то неудача приводит к сдвигу в обратном направлении вдоль j - го вектора на следующей реализации шага 1. Отметим также, что шаг 1 повторяется до тех пор, пока неудача будет иметь место при спуске по каждому из направлений поиска. В случае, когда все направления поиска оказались неуспешными, строятся новые направления поиска, в соответствии с процедурой Грамма - Шмидта.

Прямые методы поиска, имеют меньшую скорость сходимости и реализуются в случае простых задач, однако на практике могут оказаться более удовлетворительными с точки зрения пользователя и решение задач с их помощью может обойтись дешевле.

Метод сопряженных градиентов. Как уже отмечалось в третьем подразделе настоящего раздела, что с понятием квадратичной формы тесно связано понятие сопряженных направлений, основанное на положительной определенности матрицы G. Два направления характеризуемые векторами S^i и S^j считаются сопряженными относительно положительно определенной матрицы G, если

$$(S^i)^t \cdot G \cdot S^j = 0, \qquad (11.79)$$

при $i \neq j$. Условие сопряжения напоминает условие ортогональности векторов S^i и S^j

$$(S^i)^t \cdot S^j = 0, \qquad (11.80)$$

где матрица G, подразумевается единичной. Это дает основание воспользоваться для выбора сопряженных направлений известной процедурой Грамма - Шмидта для ортогонализации пространства векторов, подразумевая вектор $G \cdot S^j$, ортогональным вектору S^i

$$(S^i)^t \cdot (G \cdot S^j) = 0. \qquad (11.81)$$

Первое направление поиска выбирается так же, как и ранее, противоположным направлению градиента в исходной точке, а последующие векторы градиента являются линейными комбинациями векторов предварительно выбранных направлений

$$
\begin{aligned}
&-\nabla F^0 = S^0\ ;\\
&-\nabla F^1 = S^1 + K_{10}\cdot S^0\ ;\\
&\ldots\quad \ldots\quad \ldots\quad \ldots\quad \ldots\quad \ldots\quad \ldots\\
&-\nabla F^k = S^k + \sum_{i=1}^{k} K_{i,k-1}\cdot S^{i-1},
\end{aligned}
\tag{11.82}
$$

где $1 < k < n-1$.

Так как первое направление поиска определено, второе направление находится из обобщённого условия ортогональности (11.81)

$$(-\nabla F^1)^t \cdot G \cdot S^0 = (S^1 + K_{10}\cdot S^0)^t \cdot G \cdot S^0 = 0.$$

Коэффициент K_{10} можно рассчитать из выражения

$$K_{10} = -[(\nabla F^1)^t \cdot G \cdot S^0] \,/\, [(S^0)^t \cdot G \cdot S^0],$$

если известна матрица G.

Подставив найденный коэффициент во второе уравнение системы (11.82), определяем первое направление за исходным направлением

$$S^1 = -\nabla F^1 + [(\nabla F^1)^t \cdot G \cdot S^0)] \,/\, [(S^0)^t \cdot G \cdot S^0]\cdot S^0.$$

Из последующих уравнений (11.82) определим коэффициенты K_{ij}, путём последовательного умножения на предыдущие выделенные направления и используя обобщённое условие ортогональности

$$K_{ij} = -[(\nabla F^i)^t \cdot G \cdot S^j] \,/\, [(S^j)^t \cdot G \cdot S^j], \tag{11.83}$$

где $j = 0,1,\ldots,i-1$.

Соответственно, новое сопряжённое направление определиться из выражения

$$
\begin{aligned}
S^i = &-\nabla F^i + [(\nabla F^i)^t \cdot G \cdot S^0)] \,/\, [(S^0)^t \cdot G \cdot S^0]\cdot S^0 + \cdots\\
&\cdots + [(\nabla F^i)^t \cdot G \cdot S^{i-1}] \,/\, [(S^{i-1})^t \cdot G \cdot S^{i-1}]\cdot S^{i-1}.
\end{aligned}
\tag{11.84}
$$

Таким образом, для вычисления нового сопряженного направления поиска необходимо уметь вычислять вектор градиента в текущей точке ∇F^i и матрицу Гессе G.

Поскольку в общем случае матрица G не известна, будем заменять матрицу G во всех выражениях значениями градиентов в соответствующих точках, предполагая квадратичность функции. В частности, воспользуемся (11.40), переписав его в виде

$$G \cdot S^i = (\nabla F^{i+1} - \nabla F^i)/d_i. \tag{11.85}$$

Подставляя (11.85) в (11.83), получаем

$$K_{ij} = [(-\nabla F^i)^t \cdot (\nabla F^{j+1} - \nabla F^j)] / [(S^j)^t \cdot (\nabla F^{j+1} - \nabla F^j)].$$

Учитывая условие ортогональности направления поиска и сопряженности градиента (11.48), а также тот факт, что в окрестности минимума градиенты в соседних точках минимума практически ортогональны, т.е. следующее

направление поиска совпадает с направлением градиента в предыдущей точке, видим, что практически отличными от нуля будут коэффициенты K_{ij}, при $j = i - 1$

$$K_{i,i-1} = [(\nabla F^i)^t \cdot \nabla F^i] / [(S^{i-1})^t \cdot \nabla F^{i-1}].$$

Теперь рассмотрим произведение $(S^k)^t \cdot \nabla F^k$. Согласно (11.82)

$$S^k = -\nabla F^k + \sum_{i=1}^{k} K_{i,k-1} \cdot S^{i-1}.$$

Образуя рассматриваемое произведение, получим

$$(S^k)^t \cdot \nabla F^k = [-\nabla F^k + \sum_{i=1}^{k} K_{i,k-1} \cdot S^{i-1}]^t \cdot \nabla F^k.$$

Учитывая условия сопряжения (11.79), получаем

$$(S^k)^t \cdot \nabla F^k = -(\nabla F^k)^t \cdot \nabla F^k. \qquad (11.86)$$

В результате, коэффициенты разложения по ортогональным направлениям, запишутся

$$\alpha_i = K_{i,i-1} = -[(\nabla F^i)^t \cdot \nabla F^i] / [(\nabla F^{i-1})^t \cdot \nabla F^{i-1}], \qquad (12.87)$$

что полностью соответствует полученному ранее соотношению (11.51), с учётом знаков коэффициентов в (11.47) и (11.82). Используя (11.82) и (11.87), перепишем соотношение (11.84) в виде

$$S^i = -\nabla F^i + \alpha_i \cdot S^{i-1}, \qquad (11.88)$$

которое для определения нового направления не требует вычисления матрицы Гессе.

Обобщим изложенный материал, представив общий алгоритм, метода сопряжённых градиентов.

1. Полагаем $k = 0$ и выбираем точку начального приближения X^0, полагая $S^0 = -\nabla F(X^0)$.

2. Вычислим $F(X^k)$ и $\nabla F(X^k)$ при $k = 0$.

3. Определяем направление поиска

$$S^k = -\nabla F(X^k) + [(\nabla F^k)^t \cdot \nabla F^k] / [(\nabla F^{k-1})^t \cdot \nabla F^{k-1}] \cdot S^{k-1}$$

и нормируем вектор направления поиска к единичной длине $S^k = S^k / \left\| S^k \right\|$.

4. В направлении поиска S^k найдем длину шага d_k такую, что или

$$F(X^k + d_k \cdot S^k) < F(X^k),$$

или функция $F(X^k + d_k \cdot S^k)$ минимальна в направлении S^k.

5. Вычислим:

$$\Delta X^k = d_k \cdot S^k;$$
$$X^{k+1} = X^k + \Delta X^k.$$

6. Если $\left|F(X^{k+1}-X^k)\right|<\varepsilon_1$ и $\left\|\Delta X^k\right\|<\varepsilon_2$, то процесс сошелся, иначе положим $k=k+1$ и перейдем к шагу 2.

Применение соотношений (11.87) и (11.88) для не квадратичной функции, обеспечивает линейную сходимость, до тех пор, пока направление поиска не начнет периодически повторятся.

Рассмотренный алгоритм прост для реализации и требует умеренный объем оперативной памяти, необходимо запоминать только предыдущее направление поиска и предыдущий градиент. Этот алгоритм часто используют для задач, имеющих большое число переменных. В литературе данный метод известен под названием метода сопряженного градиента Флетчера и Ривса. Известны и другие версии метода сопряжённого градиента с лучшей сходимостью, но требующие несколько большей памяти. При этом модификации подвергается формула определения направлений поиска (11.87) и изменяется последовательность шагов алгоритма.

Вариант алгоритма, использующий соотношения (11.83) и (11.84), практически не используется из-за необходимости вычисления матрицы Гессе G.

Таким образом, метод сопряженных направлений основан на использовании условия сопряжения, для выбора очередного направления поиска, причем, используя предположение о квадратичности целевой функции, из конечных выражений, исключается матрица Гессе и остается лишь информация о градиентах в предыдущей и текущей точках.

Метод Дэвидона-Флетчера-Пауэла. Как уже отмечалось, существует класс методов, называемых квазиньютоновскими или градиентными, с большим шагом, которые аппроксимируют матрицу Гессе или обратную к ней, используя информацию о первых производных.

Исходным моментом обоснования метода является разложение целевой функции $F(X)$ в ряд Тейлора с удержанием первых трех членов ряда и записи его в виде

$$F(X^k+\Delta X)-F(X^k)=(\Delta X^k)\cdot\nabla F^k+1/2\cdot(\Delta X^k)\cdot G^k\cdot\Delta X. \qquad (11.89)$$

Найдем минимум разности, стоящей в левой части, дифференцируя (11.89) по ΔX и приравнивая результат к нулю

$$\partial[F(X^k+\Delta X)-F(X^k)]/\partial\Delta X=\Delta F^k+G^k\cdot\Delta X=0,$$

откуда

$$G^k\cdot\Delta X=-\nabla F^k, \qquad (11.90)$$

Формально решение (11.90) можно записать

$$\Delta X=-(G^k)^{-1}\cdot\nabla F^k=-H^k\cdot\nabla F^k, \qquad (11.91)$$

где $H=G^{-1}$. Направление поиска теперь полагаем совпадающим с вектором

$$d_k\cdot S^k=\Delta X^k=-H^k\cdot\nabla F^k, \qquad (11.92)$$

откуда новое значение вектора X можно записать в виде

$$X^{k+1}=X^k+d_k\cdot S^k=X^k-d_k\cdot H(X^k)\cdot\nabla F(X^k). \qquad (11.93)$$

В данном методе матрица $H(X^k)$ является аппроксимацией обратной матрицы Гессе $(G^k)^{-1}$ и часто называется матрицей направлений. Для аппроксимации обратной матрицы Гессе используются различные соотношения, не требующие вычисления вторых частных производных $F(X)$ и обращения матрицы.

Вспомним соотношение (11.38), справедливое для квадратичных функций, и перепишем его в виде

$$\nabla F(X^{k+1}) - \nabla F(X^k) = G(X^k) \cdot (X^{k+1} - X^k). \qquad (11.94)$$

Формальное решение (11.94) запишется

$$(X^{k+1} - X^k) = H(X^k) \cdot [\nabla F(X^{k+1}) - \nabla F(X^k)], \qquad (11.95)$$

где $H(X^k) = G^{-1}(X^k)$. Вводя обозначения

$$\Delta X^k = X^{k+1} - X^k,$$

$$\Delta Q^k = \nabla F^{k+1} - \nabla F^k,$$

соотношение (11.95) перепишем в виде

$$\Delta X^k = H(X^k) \cdot \Delta Q^k. \qquad (11.96)$$

Отметим, что для квадратичной функции $F(X)$ матрицы G и H постоянны, поэтому (11.96) можно рассматривать как систему, используемую для оценки $H(X)$, при известных значениях $F(X)$, $\nabla F(X)$ и ΔX. Для решения этой системы могут быть использованы различные методы, каждый из которых приводит к различным методам переменной метрики.

В довольно большой группе методов $G^{-1}(X^{k+1})$ аппроксимируется с помощью векторов градиента и приращений, полученных на предыдущем шаге

$$G^{-1}(X^{k+1}) \cong \lambda \cdot H(X^{k+1}) = \lambda \cdot [H^k + \Delta H^k], \qquad (11.97)$$

где ΔH^k – искомая матрица, а λ - масштабный множитель, константа, обычно равная единице. Выбор ΔH^k, по существу, определяет метод переменной метрики. Для обеспечения сходимости $\lambda \cdot H^{k+1}$ должна быть положительно определённой и удовлетворять уравнению (11.96).

Пусть на $(k+1)$- ом шаге известны X^k, ∇F^k, ∇F^{k+1}, H^k и необходимо определить H^{k+1}, удовлетворяющую соотношению

$$H^{k+1} \cdot \Delta Q^k = 1/\lambda \cdot \Delta X^k. \qquad (11.98)$$

Так как $\Delta H^k = H^{k+1} - H^k$, то уравнение (11.98) перепишется в виде

$$\Delta H^k \cdot \Delta Q^k = 1/\lambda \cdot \Delta X^k - H^k \cdot \Delta Q^k. \qquad (11.99)$$

Прямой подстановкой результата, можно показать, что уравнение (11.99) имеет следующее решение

$$\Delta H^k = 1/\lambda \cdot \Delta X^k \cdot Y^t / (Y^t \cdot \Delta Q^k) - H^k \cdot \Delta Q^k \cdot Z^t / (Z^t \cdot \Delta Q^k), \qquad (11.100)$$

где Y и Z – произвольные векторы размерности n. Если при $\omega = 1 / \lambda = 1$ выбирается специальная линейная комбинация двух направлений ΔX^k и $H^k \cdot \Delta Q^k$, а именно

$$Y = Z = \Delta X^k - H^k \cdot \Delta Q^k, \tag{11.101}$$

то это соответствует, так называемому, алгоритму в модификации Бройдена. При этом (11.100) перепишется в виде

$$\Delta H^k = (\Delta X^k - H^k \cdot \Delta Q^k) \cdot (\Delta X^k - H^k \cdot \Delta Q^k)^t / [(\Delta X^k - H^k \cdot \Delta Q^k)^t \cdot \Delta Q^k]. \tag{11.102}$$

Если же комбинация удовлетворяет соотношениям

$$Y = \Delta X^k, \tag{11.103}$$

$$Z = H^k \cdot \Delta Q^k, \tag{11.104}$$

то H^{k+1}, с учётом (11.100), запишется в виде

$$H^{k+1} = H^k + \Delta X^k \cdot Y^t / (Y^t \cdot \Delta Q^k) - H^k \cdot \Delta Q^k \cdot Z^t / (Z^t \cdot \Delta Q^k), \tag{11.105}$$

что и соответствует описываемому алгоритму Дэвидсона-Флетчера-Пауэла.

Допустимы и другие подстановки векторов, на которых мы не будем останавливаться. Если шаги ΔX^k определяются последовательно путём минимизации $F(X)$ в направлении S^k, то все методы, с помощью которых вычисляют симметрическую матрицу H^{k+1}, удовлетворяющую (11.98), для квадратичной функции дают взаимно сопряжённые направления.

Метод Дэвидона-Флетчера-Пауэла попадает в общий класс квазиньютоновских процедур, в которых направления поиска задаются в виде $-H^k \cdot \nabla F^k$, т.е. совпадают с направлением градиента, отклонённым в результате умножения на $-H^k$. На следующем шаге, матрица направлений H^{k+1} представляется, в виде суммы двух симметрических матриц, ранга 1 каждая. В связи с этим данная модификация алгоритма называется схемой коррекции ранга 2.

Рассмотрим конкретно алгоритм Дэвидона-Флетчера-Пауэла минимизации дифференцируемой функции нескольких переменных. Как уже отмечалось, если функция квадратичная, то метод вырабатывает сопряженные направления и останавливается после выполнения одной итерации, т.е. после поиска вдоль каждого из сопряженных направлений.

Алгоритм включает в себя два этапа:

Начальный этап. Пусть $\varepsilon > 0$ – константа для остановки. Положить $j = k = 0$, где j – номер текущего направления, а k – номер итерации. Выбрать начальную точку X^0 и симметричную начальную положительно определённую матрицу H^0, например, равную единичной и принять $Y^0 = X^0$ и перейти к основному этапу.

Основной этап.

Шаг 1. Если $\left\|\nabla F(Y^j)\right\| \le \varepsilon$, то остановиться, в противном случае, положить $S^j = -H^j \cdot \nabla F(Y^j)$ и взять в качестве d_j оптимальное решение задачи минимизации $F(Y^j + d_j \cdot S^j)$, при $d \ge 0$. Если $j < n$, то перейти к шагу 2. Если $j = n-1$, то положить $Y^0 = X^{k+1} = Y^n -$ последний вектор, найденный на первом шаге, заменить k на $k+1$, положить $j = 1$ и повторить шаг 1.

Шаг 2. Построить H^{j+1}, следующим образом

$$H^{j+1} = H^j + Y^j \cdot (Y^j)^t / [(Y^j)^t \cdot \Delta Q^j] - Z^j \cdot (Z^j)^t / [(Z^j)^t \cdot \Delta Q^j],$$

где

$$Y^j = d_j \cdot S^j = \Delta X^j,$$

$$Z^j = H^j \cdot \Delta Q^j = H^j \cdot [\nabla F(Y^{j+1}) - \nabla F(Y^j)].$$

Заменить $j = j+1$ и перейти к шагу 1.

Таким образом, метод Дэвидона-Флетчера-Пауэла основан на аппроксимации обратной матрицы Гессе, входящей в выражение для нового направления поиска, исходя из предположения о квадратичности целевой функции.

Рассмотренный нами набор алгоритмов прямых и градиентных методов оптимизации позволяет достаточно эффективно решать практические задачи. Методы оптимизации могут быть использованы, как на этапе уточнения моделей сложных элементов электронных схем по измеренным характеристикам, так и для получения наиболее эффективных схемных решений РЭА.

ЗАКЛЮЧЕНИЕ

Рассмотренный нами круг вопросов дает представление о содержании курса "Основы АПР РЭУ". Конечно, рассмотренными вопросами содержание автоматизированного проектирования РЭУ не исчерпывается. В частности здесь не были рассмотрены вопросы проектирования импульсных устройств, цифровых логических устройств, устройств с распределенными параметрами и электродинамический расчет устройств СВЧ диапазона. Предполагается, что эти вопросы хотя бы частично будут рассмотрены в соответствующих курсах.

Здесь же были рассмотрены наиболее важные, т.е. основные вопросы, без которых немыслима постановка задачи автоматизированного проектирования РЭУ. Это, прежде всего методы формирования математических моделей цепи в виде систем алгебраических и дифференциальных уравнений, модели элементов электронных схем, численные методы решения систем алгебраических - линейных, нелинейных и дифференциальных уравнений, основные характеристики цепей в частотной и временной области и методы их расчета. В качестве основной характеристики связанной с производством и эксплуатацией рассмотрены вопросы расчета чувствительности характеристик к внешним и внутренним параметрам РЭА, а в качестве основного метода автоматизированного проектирования устройств РЭА с заданными характеристиками рассмотрены теория и методы оптимизации.

Все алгоритмы расчета и проектирования опираются на развитые аналитические и численные математические методы, которые излагаются попутно. Рассматриваются также вопросы, имеющие большое значение для реализации методов и алгоритмов на практике связанные с точностью и устойчивостью вычислений.

СПИСОК ЛИТЕРАТУРЫ

1. Влах И., Сингхал К. Машинные методы анализа и проектирования электронных схем. - М.: Радио и связь, 1888. - 560 с.
2. Фидлер Дж., Найтингейл К. Машинное проектирование электронных схем. - М.: Высшая школа, 1985. - 216 с.
3. Нерретер В. Расчет электрических цепей не персональной ЭВМ. М.: Энергоатомиздат, 1991. - 260 с.
4. Гупта К., Гардж Р., Чадха Р. Машинное проектирование СВЧ устройств. - М.: Радио и связь, 1987. - 432 с.
5. Ватанабе М., Асаки К., Кани К., Оцуки Т. Проектирование СБИС. - М.: 1988. - 304 с.
6. Автоматизация проектирования БМС. В 6-ти кн.: Практическое пособие. Кн. 3. В.В. Ермак, В.Н. Перминов, А.Г. Соколов. Рабочие станции в проектировании БИС. /Под ред. Г. Г. Казеннова. - М.: Высшая школа, 1990. - 272 с.
7. Баушев В.0., Бондарь В.А., Легостаев Н.С, Расчет и проектирование электронных схем. Учебное пособие. – Томск: изд-во ТГУ, 1990. - 265 с.
8. Автоматизация проектирования радиоэлектронных средств. /Под ред. О.В. Алексеева. - М.: Высшая школа, 2000. – 479 с.
9. Норенков И.П. Основы автоматизированного проектирования. - М.: Изд-во МГТУ, 2000. – 360 с.
10. Росадо Л. Физическая электроника и микроэлектроника.- М.: Высшая школа, 1991.- 351с.
11. Демидович Б.П., Марон И.А. Основы вычислительной математики. М.: Наука, 1966.-664 с.
12. Гилл Ф., Мюррей У., Райт М., Практическая оптимизация. - М.: Мир, 1985. - 509 с.