Современная вычислительная математика содержит в себе три главных части:

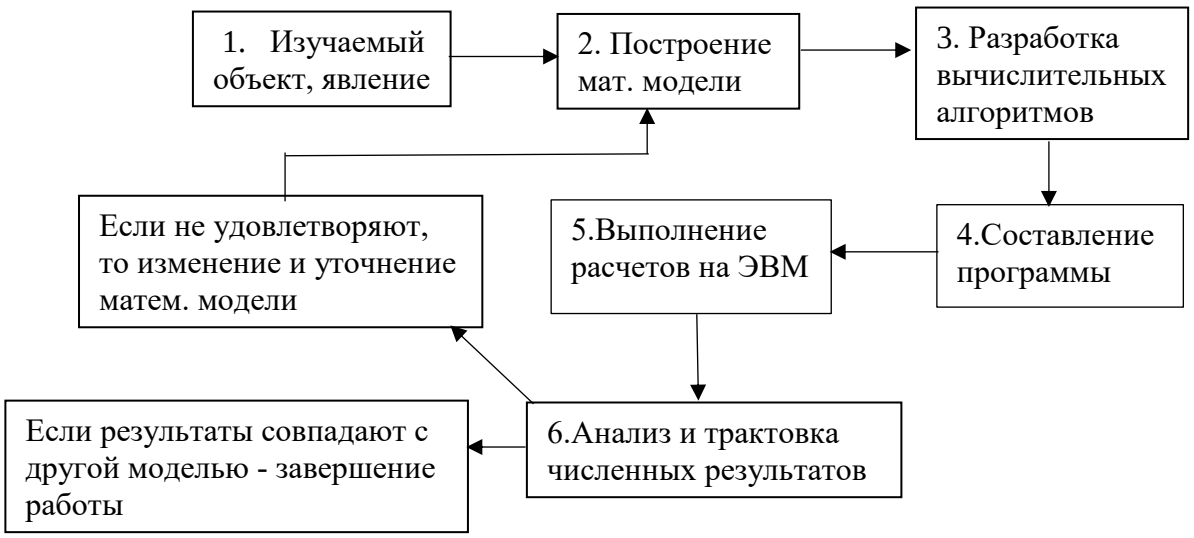
1. Теорию вычислительных методов.
2. Приборы, позволяющие автоматизировать вычисления.
3. Вспомогательные средства, обеспечивающие управление работой вычислительной машины, к ним относятся алгебраические языки, стандартные программы, содержащие наиболее часто употребляемые вычислительные процессы и т.д.

Две крупные проблемы значительно подтолкнут развитие вычислительных методов: овладение ядерной энергией и освоение космоса.

Вычислительные методы, предназначенные для быстродействующих ЭВМ, должны удовлетворять многообразным и зачастую противоречивым требованиям:

1. Метод должен быть целенаправленным.
2. Внутренние свойства метода должны совпадать с внешними.
3. Должны быть установлены границы применения алгоритмов.
4. Результаты должны быть надежными и достоверными.
5. Алгоритмы должны быть универсальными.
6. Методы устойчивые и сходящиеся.
7. Алгоритмы экономичные и точные.

В настоящее время выработалась технология исследования сложных проблем, основанная на построении и анализе с помощью ЭВМ математических моделей изучаемого объекта. Такой метод исследования называется вычислительным экспериментом.



Метод Гаусса Цель В.Э.: за возможно меньшее время получить наиболее достоверные результаты.

Численные методы линейной алгебры

К численным методам линейной алгебры традиционно относят методы решения СЛАУ, обращения матриц, вычисления определителей, нахождения собственных значений и собственных векторов матриц и нулей многочленов.

При формальном подходе решение таких задач не вызывает затруднений: решение системы можно найти, раскрыв определители в формуле Крамера, для нахождения собственных значений матрицы достаточно выписать характеристическое уравнение и найти его корни. Однако эти манипуляции встречают возражения со многих сторон. Так, при непосредственном раскрытии определителей решения, решение СЛАУ с m неизвестными требует $m!\,m$ арифметических операций, уже при $m = 30$ такое число операций уже недоступно для современных ЭВМ. Другой причиной, по которой эти классические способы неприменимы даже при малых m, является сильное влияние на окончательный результат округлений при вычислениях. Уже при $m = 20$ при расчетах на современных ЭВМ типичная аварийная остановка из-за переполнения порядка чисел.

Методы решения алгебраических задач разделяются на точные, итерационные и вероятностные. В настоящее время точные методы

обычно применяются для решения систем до порядка 10^4, итерационные до порядка 10^7, свыше вероятностные.

Классы задач, для решения которых обычно применяются методы этих групп можно условно назвать классами задач соответственно с малым, средним и большим числом неизвестных.

Введение

К численным методам относят решения СЛАУ, обращение матриц, задача на собственные вектора и собственные значение, поиск нулей многочлена. При формальном похоже это несложно. К примеру, систему можно решить методом Крамера, однако это потребует $m!\,m$ арифметических операций, где n – число неизвестных, m – число уравнений. Другой причиной, по которой классические методы неприемлемы, является влияние погрешностей. Так, при использовании методов для $n = 20$, погрешности округления приводят к тому, что получаемый результат далек от реальности. Поэтому важно выбрать правильный метод.

Вычислительные схемы метода Гаусса.

Метод Гаусса является примером точного метода.

Пусть задана система линейных алгебраических уравнений:

$$\begin{aligned} &Ax = f, \\ &A = (a_{ij}),\ i,j = 1..n; \\ &f = (f_i). \end{aligned} \qquad \textbf{(1)}$$

Будем рассматривать систему (1) в виде:

$$\begin{aligned} &a_{11}x_1 + a_{12}x_2 + a_{12}x_3 + \ldots + a_{1n}x_n = f_1 \\ &\ldots \\ &\ldots \\ &\ldots \\ &a_{n1}x_1 + a_{n2}x_2 + a_{n2}x_3 + \ldots + a_{nn}x_n = f_n \end{aligned} \qquad (\mathbf{2})$$

Схема единственного деления.

Выберем из системы какое-либо уравнение, а в нем какою-нибудь неизвестную x_i, коэффициент которой не равен 0. Не уменьшая

общности, можно считать, что это первое уравнение и неизвестное x_1, т.е. предполагаем $a_{11} \neq 0$. Разделим первое уравнение на a_{11}, которое будем называть ведущим:

$$x_1 + b_{12}x_2 + b_{12}x_3 + \ldots + b_{1n}x_n = g_1 \quad (3)$$
$$b_{1j} = \frac{a_{ij}}{a_{11}}, g_1 = \frac{f_1}{a_{11}}$$

Умножим (3) последовательно на $a_{21}, a_{31}, \ldots, a_{n1}$ и вычтем последовательно из второго, третьего и т.д. уравнений системы (2) , тем самым мы исключим неизвестную x_1 из второго и т.д. уравнений. Преобразованные уравнения будут иметь вид:

$$a_{22}{}^1 x_2 + a_{23}{}^1 x_3 + \ldots + a_{2n}{}^1 x_n = f_2{}^1$$
$$\ldots \quad (4)$$
$$a_{n2}{}^1 x_2 + a_{n2}{}^1 x_3 + \ldots + a_{nn}{}^1 x_n = f^1{}_n,$$

где $a_{ij}^1 = a_{ij} - a_{i1}b_{1j}$, $f_i{}^1 = a_{i1}g_1$.

Систему (4) можно рассматривать как систему с $n-1$ уравнением с $n-1$ неизвестными $x_2, x_3, \ldots, x_n$ и с ней поступим аналогичным образом как с системой (2). Продолжая этот процесс, предположим, что он возможен до m – го шага. Тогда на m - том шаге получим

$$x_m + b_{mn-1}x_{m+1} + \ldots + b_{mn}x_n = g_m \quad (5)$$
$$b_{m+1m+1}{}^m x_{m+1} + b_{m+1n}{}^m x_n = f_{m+1}{}^m$$
$$\ldots$$
$$b_{nm+1}{}^m x_{m+1} + \ldots + b_{nn}{}^m x_n = f_n{}^m, \quad (6)$$

Предположим, что m-ый шаг – это последний возможный шаг преобразования.

1. Если $m = n$, то соединив все первые уравнения до n-го шага включительно, получим систему :

$$x_1 + b_{12}x_2 + \cdots + b_{1n}x_n = g_1$$
$$\ldots \quad (7)$$
$$x_{n-1} + b_{n-1n}x_n = g_{n-1}$$
$$x_n = g_n$$

Из последнего уравнения найдем x:

$$x_n = g_n$$
$$x_{n-1} = -b_{n-1n}x_n + g_{n-1}$$

$$\dots \qquad (8)$$
$$x_1 = g_1 - b_{12}x_2 - \dots - b_{1n}x_n$$

Процесс нахождения x_n $(n = 1..j)$ по (8) называется обратным ходом метода Гаусса. Процесс приведения к (7) – прямым ходом метода Гаусса. Если $m = n$, то получаем единственное решение (2).

2. Пусть $m < n$ и m – последний возможный шаг преобразования, т.е. на следующем шаге мы не можем найти ведущего элемента отличного от 0

$$\begin{aligned} &x_m + b_{mm+1}x_{n+1} + \dots + b_{mn}x_n = g_m \\ &0 = f_{m+1}^m \\ &\qquad \dots \\ &0 = f_n^m \end{aligned} \qquad (9)$$

Если все $f_j = 0$, j $= m + 1$...n, то на n-том шаге исходная система может быть записана в виде:

$$\begin{aligned} &x_1 + b_{12}x_2 + \cdots + b_{1n}x_n = g_1 \\ &x_2 + \cdots + b_{2n}x_n = g_2 \\ &\qquad \dots \\ &x_m + b_{mn}x_n = g_m \end{aligned} \qquad (10)$$

Это означает, что неизвестные $x_1, x_2, \dots, x_m$ можно было выразить через $x_{m+1}, x_{m+2}, \dots, x_n$, а тогда (2) имеет множество решений.

Если $m < n$ и хотя бы один из коэффициентов $f^m{}_j \neq 0$, $j = m + 1, \dots n$, то (2) не имеет решений.

На практике часто осуществляется контроль вычислений.

Метод Гаусса относится к точным методам решений СЛУ, в котором четко закреплены действия.

Число всех умножений и делений равно: $S_n = \frac{n}{3}(n^2 + 6n - 1)$.

Часто может оказаться, что на первом шаге коэффициент $a_{11} \neq 0$, но ≈ 0, в этом случае

1. Схема с выбором максимального элемента по строке.

Находят $|a_{1j_0}| = max|a_{ij}|, j = 1 \dots n$. Объявляют a_{1j_0}- ведущей и исключают неизвестные x_{j_0}из второго, ..., n-го уравнения. Аналогично поступают и с другими неизвестными.

2. Схема с выбором максимального элемента по столбцу.

$a_{i_0 1}$ - ведущий, но $|a_{i_0 1}| = max|a_{ij}|, i = 1 \dots n$ и исключают неизвестную x_1 из 1-го, ..., $i_0 - 1, i_0 + 1$,..., n-го уравнений.

3. Схема с выбором максимального элемента по всей матрице.

$a_{i_0 j_0}$- ведущий, но $|a_{i_0 j_0}| = max|a_{ij}|, i, j = 1 \dots n.$ В этом случае неизвестную x_{j_0} исключают из 1-го, ..., $i_0 - 1, i_0 + 1$,..., n уравнений системы.

4. Схема Жордана или схема оптимального исключения:

1.

$$\begin{bmatrix} 1 & * & \dots & * \\ 0 & * & \dots & * \\ \vdots & \vdots & \vdots & \vdots \\ 0 & * & \dots & * \end{bmatrix} \qquad \begin{bmatrix} 1 & * & \dots & * \\ * & * & \dots & * \\ \vdots & \vdots & \vdots & \vdots \\ * & * & \dots & * \end{bmatrix}$$

2.

$$\begin{bmatrix} 1 & * & \dots & * \\ 0 & 1 & \dots & * \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & * \end{bmatrix} \qquad \begin{bmatrix} 1 & 0 & * \dots & * \\ 0 & 1 & * \dots & * \\ \vdots & \vdots & \vdots & \vdots \\ * & * & * \dots & * \end{bmatrix}$$

.....

n.

$$\begin{bmatrix} 1 & * & * \dots & * \\ 0 & 1 & * \dots & * \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 \dots & * \end{bmatrix} \qquad \begin{bmatrix} 1 & * & \dots & 0 \\ * & 1 & \dots & * \\ \vdots & \vdots & \vdots & \vdots \\ 0 & * & \dots & 1 \end{bmatrix}$$

Таким образом, в методе Гаусса исходная матрица приводится к треугольному виду, а матрица Жордана – к диагональному. Число операций приблизительно в 2 раза меньше числа операций в матрице Жордана.

К точным методам также можно отнести:

1. Метод ортогонализации.
2. Метод окаймления.
3. Метод квадратного корня и другие.

Итерационные методы. Некоторые сведения о нормах векторов матриц. Метод простой итерации.

Нормой вектора x называется сопоставляемое этому вектору действительное число:

1. $\|x\| > 0, x \neq 0$
   $\|0\| = 0;$
2. $\|\alpha\, x\| = |\alpha|\, \|x\|;$
3. $\|x + y\| \leq \|x\| + \|y\|.$

Вводить норму можно различными способами:

1. Кубическая $x \in R^n, x = (x_1, x_2, \dots, x_n);$
   $$||x||_I = max_{1 \leq i \leq n} |x_i|;$$
2. Октаэдрическая $\|x\|_{II} = \sum_{i=1}^{n} |x_i|;$
3. Сферическая или евклидова норма вектора $\|x\|_{III} = (\sum |x_i|^2)^{1/2}.$

Справедливы 2 равнозначных определения сходимости последовательности вектора:

1. $x_k \to x^*$, если $\|x^k - x^*\| \to 0, k \to \infty$. (сходимость по норме);
2. $x_k \to x^*$, если $\lim\limits_{k \to \infty} x_k^{(i)} = x^{(i)}, i = \overline{1, n}$ (покоординатная сходимость).

Нормой квадратной матрицы A назовем сопоставляемое этой матрице A число $\|A\|$, удовлетворяющее условиям:

1. $\|A\| > 0, A \neq 0, \|0\| = 0;$
2. $\|\alpha\, \mathrm{A}\| = |\alpha|\, \|A\|;$
3. $\|A + B\| \leq \|A\| + \|B\|;$
4. $\|A\, B\| \leq \|A\|\, \|B\|.$

Будем говорить, что $\|A\|$ согласованна с данной нормой вектора, если для любой квадратной матрицы A, размерность которой равна порядку матрицы, выполняется$\|A\, x\| \leq \|A\|\, \|x\|$.

Среди всех норм матриц, согласованных с данной нормой, выберем наименьшую. Для этой цели за норму матрицы A выберем Ax в

предположении, что вектор x пробегает множество всех векторов, нормы которых равны 1, т.е. $\|A\| = max\|Ax\|, \|x\| = 1.$

Для каждой матрицы A, в силу непрерывности норм матриц, этот максимум достигается, т. е. всегда найдется вектор x^*, такой что $\|A\| = \|Ax^*\|, \|x^*\| = 1.$

Введенную таким образом норму будем называть подчиненной данной норме вектора.

Примеры подчиненных норм:

1. $\|x\|_I: \|A\|_I = max \sum_{j=1}^{n} |a_{ij}|, 1 \le j \le n;$
2. $\|x\|_{II}: \ \|A\|_{II} = max \sum_{i=1}^{n} |a_{ij}|, 1 \le i \le n;$
3. $\|x\|_{III}: \ \|A\|_{III} = \sqrt{\lambda_1(T)} \quad \mathrm{T} = A^*A > 0$

$\lambda_1 > \lambda_2 > \lambda_3 > \ldots > \lambda_n > 0 -$ собственные значения,

Если A - симметрическая положительно определенная, то $A = A^T > 0$, то $\|A\|_{III} = max\, \lambda_j\,(A), 1 \le j \le n.$

Лемма

Модуль каждого собственного значения матрицы не превосходит любую из ее норм.

Доказательство: обозначим $\lambda_i(A)$ - собственные значения соответствующие собственному вектору x_i матрицы A, тогда для любых λ_i справедливо равенство

$$Ax_i = \lambda_i x_i$$

$$\|Ax_i\| = \|\lambda_i x_i\| = |\lambda_i|\|x_i\|$$

т. к. $\|Ax_i\| \le \|A\|\|x_i\|,$

то $|\lambda_i|\|x_i\| \le \|A\|\|x_i\|$

$\Rightarrow x_i -$ собственный вектор, $|\lambda_i| \le \|A\|.$

Лемма доказана.

Рассмотрим СЛАУ

$$\mathrm{A}x = f \qquad (\mathbf{1})$$

при условии, что определитель не равен нулю и матрица A – положительно определена. Это значит, что квадратичная форма матрицы больше нуля. Запишем (1) в каноническом виде:

$$x = \phi(x) \qquad \mathbf{(2)}$$

Решить (1) значит найти неподвижную точку отображения (2), т.е.

$$x^* = \varphi(x^*) \quad (3)$$
$$x^* = A^{-1}f$$

Для этого (1) можно записать : $\frac{x-x}{\tau} = -Ax + f$

$$x = (E - \tau A)x + \tau f \quad (4)$$
$$H(\tau) = E - \tau A$$
$$x = H(\tau)x + \tau f \quad (5)$$

Для нахождения решения (1) строим итерационный процесс по правилу:

$$x_{k+1} = H(\tau)x_k + \tau\delta, x_0 - \textit{начальное приближение} \quad (6)$$

Метод (6) – метод простой итерации для решения СЛАУ;

x_k - итерационная последовательность;

$H(\tau)$ - матрица перехода от x_k до x_{k+1};

τ - стационарный параметр.

В нашем случае метод (5) – одношаговый стационарный метод. Если бы $H(\tau) = H(\tau, k)$, то в этом случае метод назывался бы нестационарным.

Достоинства: простота, самоисправляемость.

Установим поведение $\varepsilon_к$:

$$\varepsilon_к = x_к - x^* \quad (7)$$

$\boldsymbol{\varepsilon_к}$ - погрешность итерационного метода (6).

Нам нужно установить эту погрешность при $\|\varepsilon_к\| \to 0, к \to \infty$.

$$x_{к+1} = H(\tau)x_k + \tau f$$
$$-$$
$$x^* = H(\tau)x^* + \tau f$$
$$\boldsymbol{\varepsilon_{k+1}} = \boldsymbol{H(\tau)\varepsilon_k} \quad (8)$$

Из (8) получаем:

$$\boldsymbol{\varepsilon_1} = \boldsymbol{H(\tau)\varepsilon_0}$$
$$\boldsymbol{\varepsilon_2} = \boldsymbol{H(\tau)}\,\boldsymbol{\varepsilon_1} = \boldsymbol{(H(\tau))^2\varepsilon_0}$$
$$\ldots$$
$$\boldsymbol{\varepsilon_k} = \boldsymbol{(H(\tau))^k\varepsilon_0} \quad (9)$$

Т.к. x_0- произвольный начальный вектор, то можно считать, что ε_0- произвольный вектор, тогда можно положить $\boldsymbol{\varepsilon_0} = z_i, z_i$ - собственный вектор матрицы А:

$$Az_i = \lambda_i z_i,$$

тогда из (8) $\Rightarrow \quad \varepsilon_к = (H(\tau))^k z_i = (1 - \lambda_i \tau)^k z_i$ **(10)**

Теорема

Пусть $\{x_k\}_{k=0}^{\infty}$ - последовательность, построенная по итерационному методу (6), τ - стационарный параметр, x_0 - произвольное начальное приближение. Тогда для сходимости последовательности $x_к \to x^*, к \to \infty,\ x^*$- решение (1), достаточным является условие:

$$\rho = \rho(\tau) = \|E - \tau A\| < 1 \qquad \textbf{(11)}$$

При этом скорость сходимости линейная (геометрическая прогрессия). Если $A = A^T$- симметричная и сходимость имеет место для любых x_0, то условие (11) будет также достаточным.

Общий неявный метод простой итерации.

В приложении возникает необходимость решения систем высокого порядка, матрицы которых обладают рядом специфических свойств: симметричность, положительно определена.

Такие системы в частности возникают при использовании метода сеток при решении задач мат. физики. Наиболее интересные результаты решения таких задач получил Самарский.

Пусть задана система:

$$\text{Ax} = f,\ A = \left(a_{ij}\right), i,j = \overline{1,n}\ , \qquad \textbf{(1)}$$

Матрица A симметрическая и положительно определенная, в том смысле, что $A = A^T, A > 0, (Ax, x) \geq \mu(x, x), \mu > 0, x \neq 0.$

Для системы (1) запишем итерационный метод:

$$B\frac{x_{k+1}-x_k}{\tau} + Ax_k = f,\ k = 0,1,2 \ldots, \tau - \text{параметр.} \qquad \textbf{(2)}$$

Обычно матрица B с определителем не равным нулю выбирается таким образом, чтобы легко вычислялась матрица B^{-1}(треугольная диагональная).

При этом B такова, что система $Bz = g$ имеет легко вычисляемые решения.

Перепишем метод (2) в виде:

$$x_{k+1} = (E - \tau B^{-1}A)x_k + \tau B^{-1}f, k = 0,1,2,\ldots \qquad \textbf{(3)}$$

(3) - общий неявный метод простой итерации для СЛАУ.

В (3) $H(\tau) = E - \tau B^{-1}A$

$$x_{k+1} = H(\tau)x_k + \tau B^{-1}f_i \qquad \textbf{(4)}$$

$H(\tau)$ играет роль матрицы перехода от k итерации к $k+1$. Если $H(\tau)$ не зависит от k, то процесс, называется стационарным итерационным методом. Иначе - нестационарным. Мы будем рассматривать стационарные методы.

Условия, в которых итерационная последовательность $x_к \to x*$, $к \to \infty$, определённая (3) при любом начальном приближении сходятся к $x^* = A^{-1}f$ в (1).

Теорема Самарского

Если выполнены условия:

1. $A = A^T, A > 0$;

2. $B > 0$, то для сходимости последовательности построенной по (3) для любого начального приближения $x_0 = R^n, n = 0,1,2,..$ достаточно выполнения неравенств:

$$2B > \tau A, \qquad (\alpha)$$

$$\tau A > 0, \qquad (\beta)$$

которые называются условиями Самарского.

Если кроме того выполняются:

3. $B = B^T$;

4. $AB = BA$

то (α) и (β) (условия Самарского) являются также необходимыми для сходимости последовательности $x_к \to x^*$ (к решению системы (1) при любых x_0).

(Без доказательства)

Теорема Самарского устанавливает факт сходимости последовательности $\{x_к\}_{к=0}^{\infty}$ к $x^* = A^{-1}f$, но ничего не говорит о скорости такой сходимости.

Справедливы следующие теоремы:

Теорема 2

Пусть $A = A^T > 0, B = B^T > 0$ и выполняются условия (α) и (β), сформулированные в теореме Самарского, тогда

$$(B\varepsilon_{к+1}, \varepsilon_{к+1}) \le (B\varepsilon_к, \varepsilon_к),$$

где $\varepsilon_k = x_k - x^*$ - погрешность k-го приближения.

Теорема 3

Пусть выполнены условия :

1. $A = A^T > 0, B = B^T > 0,$

2. Выполняются условия (α) и (β) теоремы Самарского,

3. A, B, τ и некоторое число ρ согласованы в том смысле, что выполнено неравенство

$$\frac{1-\rho}{\tau}B \le A \le \frac{1+\rho}{\tau}B, \qquad (5)$$

тогда для погрешности $\varepsilon_к = x_к - x^*$справедливы оценки

$$\|\varepsilon_к\|_B \le \rho^к \|\varepsilon_0\|_B, \ 0 < \rho < 1, \ \text{где } \|\varepsilon_к\|_B^2 = (B\varepsilon_к, \varepsilon_к) \qquad (6)$$

Из теоремы 3 следует, что если $\|\varepsilon_к\|_B \to 0,_{k\to\infty} \Rightarrow$ и$\|\varepsilon_к\| \to 0,_{k\to\infty}$ также, как $\rho^к \to 0,_{k\to\infty}$. Т.е. тем быстрее, чем меньше ρ.

Рассмотрим частный случай общего неявного метода простой итерации: матрицу B нужно выбирать таким образом, чтобы она конкретно отображала свойства и была связана с матрицей A.

1. Метод простой итерации

$B = diag(a_{11}, a_{22}, \dots, a_{nn})$, a_{ii} −элемент главной диагонали матрицы A.

$\tau = 1$, тогда формула (3) примет вид:

$$x_{k+1} = (E - B^{-1}A)x_k + B^{-1}f,\ k = 0,1,\dots \quad \textbf{(7)}$$

Метод будет сходится, если выполняется одно из условий:

$$q = \|E - B^{-1}A\|_I = max \sum_{j=1}^{n} \left|\frac{a_{ij}}{a_{ii}}\right| \left(1 - \delta_{ij}\right) < 1,$$

$$q = \|E - B^{-1}A\|_{II} = max \sum_{i=1}^{n} \left|\frac{a_{ij}}{a_{ii}}\right| (1 - \delta_{ij}) < 1,$$

$\delta_{ij} = \begin{cases} 1, i = j \\ 0, i \neq j \end{cases}$- символ Кронекера.

Метод сходящийся, если в матрице A имеется диагональное преобладание по строкам, либо столбцам. Т.е. на диагонали стоят элементы по модулю больше суммы модулей остальных элементов.
$(q < 1)$

Если диагонального преобладания нет, то его пытаются достичь путём алгебраических преобразований над уравнениями системы.

Отметим, что для сходимости (7) достаточно выполнение условия $2B > A$, $A = A^T > 0$, это следует из теоремы Самарского.

2. Метод Зейделя

Пусть $A = A^T > 0$, тогда она может быть представлена в виде:

$$A = L + D + L^T,$$

$$L = \begin{pmatrix} 0 & 0 & \dots & 0 \\ a_{21} & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ a_{n1} & \dots & a_{nn-1} & 0 \end{pmatrix},\ \mathrm{D} = \begin{pmatrix} a_{11} & \dots & 0 \\ \dots & \dots & \dots \\ 0 & \dots & a_{\mathrm{nn}} \end{pmatrix}$$

$\mathrm{B} = D + L$, $\tau = 1$.

Тогда общий неявный метод простой итерации в данном случае примет вид:

$$(D + L)x_{k+1} = (D + L - A)x_k + f, k = 0,1,.. \quad \textbf{(8)}$$

(8) - метод Зейделя для системы $Ax = f$.

Метод Зейделя сходится для любой СЛАУ, матрица которой симметрическая положительно определённая.

3. Метод релаксации.

$B = D + \omega L, \omega = \tau, \omega$ −параметр.

Учитывая это, общий неявный метод простой итерации в данном случае примет вид:

$$(D + \omega L)\frac{x_{k+1} - x_k}{\omega} + Ax_k = f \quad \textbf{(9)}$$

(9) – метод релаксации для системы $Ax = f$. При этом если

$0 < \omega < 1$ - то метод нижней релаксации,

$\omega = 1$ - полной релаксации,

$1 < \omega < 2$- метод верхней релаксации.

Для сходимости метода релаксации (9) достаточно выполнение условий:

1. $A = A^T > 0$,

2. $0 < \omega < 2$.

Метод скорейшего спуска.

Этот метод предназначен для решения СЛАУ

$$Ax = f \quad \textbf{(1)}$$

с вещественной, симметрической, положительно определенной матрицей A.

В методе скорейшего спуска, а также в методе сопряженных градиентов отыскание решения системы (1) связано с задачей минимизации следующего функционала

$$F(x) = (Ax, x) - 2(f, x) \quad \textbf{(2)}$$

который является квадратичной функцией от переменных $x_1, x_2, \dots, x_n$. Это объясняется тем, что решение системы (1) $x^* = A^{-1}f$ достигает минимум функционала (2) на множество векторов из вещественного векторного пространства.

Доказательство:

$$F(x) - F(x^*) = (Ax, x) - 2(f, x) - (Ax^*, x^*) + 2(f, x^*) =$$

$$= (Ax, x) - 2(Ax^*, x) - (Ax^*, x^*) + 2(Ax^*, x^*) = (A(x - x^*), x - x^*) \geq 0 \quad \textbf{(3)}$$

при этом равенство в (3) возможно, когда $x - x^* = 0$, $x = x^*$.

Таким образом, задача нахождения решения системы (1) сводится к задаче отыскания вектора x, доставляющего минимум функционалов $F(x)$.

Метод имеет следующую вычислительную структуру. Исходя из x_0 к решению x^* системы $Ax = f$ вычисляется вектор $r_0 = f - Ax_0$,

число $\alpha_0 = \frac{(r_0, r_0)}{(r_0, Ar_0)}$.

Следующее приближение x_1 определяется по формуле:

$$x_1 = x_0 + \alpha_0 r_0.$$

Вектор x_2 вычисляем из условия минимума функции $F(x_1 + \alpha r_1)$, где $r_1 = f - Ax_1 = r_0 - \alpha_0 Ar_0$.

Исходя из условия минимума функции $F(x_1 + \alpha r_1)$ мы найдем α_1 по формуле $\alpha_1 = \frac{(r_1, r_1)}{(r_1, Ar_1)}$ и $x_2 = x_1 + \alpha_1 r_1$.

Далее процесс построения последовательных приближений осуществляется рекуррентно, по формулам:

$$r_k = f - Ax_k = r_{k-1} - \alpha_{k-1} Ar_{k-1}, \quad \textbf{(4)}$$

$$x_{k+1} = x_k + \alpha_k r_k, \quad \textbf{(5)}$$

$$\alpha_k = \frac{(r_k, r_k)}{(r_k, Ar_k)}. \quad \textbf{(6)}$$

Отметим, что r_k особенно при большом порядке матрицы A удобно вычислять по формуле:

$$r_k = r_{k-1} - \alpha_{k-1} A r_{k-1}.$$

Чтобы из-за ошибок округления таким образом вычисляемые r_k через несколько шагов не стали сильно отличаться от истинных невязок $f - Ax_k$, их надо время от времени вычислять по формуле $r_k = f - Ax_k$.

В методе скорейшего спуска ортогонализация векторов невязок системы r_k не производится.

О погрешности приближенного решения систем линейных алгебраических уравнений, об обусловленности систем и матриц.

В некоторых методах численного решения СЛАУ о точности полученного приближения решения чаще всего судят по векторам невязок системы.

Однако если для одного класса матриц малость вектора невязок системы в некоторой метрике означает и малость компонент вектора погрешностей, то для другого класса такой связи может и не быть.

Рассмотрим

$$Ax = f, \quad \textbf{(1)}$$

x^*- точное решение, y - некоторое приближенное решение. Рассмотрим вектора:

$$\varepsilon = x^* - y \text{ - вектор погрешности,} \quad \textbf{(2)}$$

$$r = f - Ay \text{ - вектор невязки.} \quad \textbf{(3)}$$

Пусть матрица A системы (1) имеет хотя бы одно очень малое по модулю собственное значение λ. И пусть z - собственный вектор, который соответствует λ.

Тогда

$$A(x^* + z) = Ax^* + Az = f + \lambda z \quad \textbf{(4)}$$

Компоненты вектора $x^* + z$ могут отличаться весьма сильно от компонент вектора x^* хотя бы в силу малости λ. Компоненты вектора $f + \lambda z$ будут мало отличаться от компонент вектора f. В связи с этим необходимо ввести такие соотношения между векторами ε и r, которые позволяли бы по величине r судить более точно о величине вектора ε.

В практике вычислений большое значение имеют не нормы векторов ε, r, а отношения $\frac{\|\varepsilon\|}{\|x^*\|}, \frac{\|r\|}{\|f\|}$, которые являются в некотором смысле относительными погрешностями. Для количественной характеристики

этих соотношений, а также векторов ε и r, введено понятие обусловленности систем и матриц. Введем в рассмотрение величину:

$$\mu = \sup_r \left(\frac{\|\varepsilon\|}{\|x^*\|}, \frac{\|r\|}{\|f\|}\right). \quad \textbf{(5)}$$

Если μ мало, то из **(5)** $\Rightarrow$

$$\|\varepsilon\| \leq \mu \cdot \frac{\|x^*\|}{\|f\|} \cdot \|r\| \quad \textbf{(6)}$$

и малость нормы вектора невязок r означает малость нормы погрешности ε. В этом случае говорят, что **(1)** хорошо обусловлена. Если μ велико, то малость нормы вектора невязок r не означает малости нормы ε. В этом случае говорят, что **(1)** плохо обусловлена.

Число μ называется мерой обусловленности системы **(1)**. По аналогии введем понятие обусловленности матрицы. Из определения вектора невязок и ε, а также из определения нормы матрицы имеем

$$\sup_r \frac{\|\varepsilon\|}{\|r\|} = \sup_r \frac{\|x^* - y\|}{\|r\|} = \sup_r \frac{\|A^{-1}r\|}{\|r\|} = \|A^{-1}\|. \quad \textbf{(7)}$$

Учитывая (3) и (4) получаем:

$$\mu = \frac{\|f\|}{\|x^*\|} \cdot \|A^{-1}\|. \quad \textbf{(8)}$$

Будем рассматривать систему **(1)** при всевозможных значениях f. Тогда решением этой совокупности будет некоторое множество X векторов x^*, отвечающих соответствующим значениям f при одной и той же матрице A.

$$v = \sup_{x^* \in X} \frac{\|Ax^*\| \cdot \|A^{-1}\|}{\|x^*\|} = \|A\| \cdot \|A^{-1}\|, \quad \textbf{(9)}$$

v - число обусловленности матрицы A.

Из **(9)** следует, что если A близка к особенной, то v будет для такой матрицы велико. В этом случае говорят, что матрица A плохо обусловлена.

Если v мало, то соответствующую матрицу A называют хорошо обусловленной.

Как правило система с хорошо (плохо) обусловленной матрицей будет хорошо (плохо) обусловлена. Значения v зависят от того, каким образом мы определим норму матрицы A, однако неравенство $v \geq 1$ справедливо при любом выборе норм матриц.

Оценка погрешности в сильной степени зависит от того, как изменяется решение **(1)** при малых изменениях ее коэффициентов и

свободных членов, а это означает, что оценка ε зависит от меры и числа обусловленности матрицы системы.

Проблема собственных значений.

Пусть задана $A = (a_{ij}), i, j = \overline{1, n}$. Собственным значением матрицы A называется число λ, такое что система линейных алгебраических уравнений:

$$Ax = \lambda x \quad \textbf{(1)}$$

имеет ненулевое число решений $x \neq 0$. Вектор x, отвечающий числу λ, - собственный вектор.

Система **(1)** имеет нулевое решение $\Leftrightarrow det|A - \lambda E| = 0$ **(2)**

(2) представляет собой уравнение n-ной степени относительно λ, со старшим коэффициентом $(-1)^n$, который можно записать в виде:

$$(-1)^n(\lambda^n - p_1\lambda^{n-1} - \ldots - p_n) = 0$$

$\lambda^n - p_1\lambda^{n-1} - \ldots - p_n$ - характеристический многочлен матрицы A. Его корни – собственные значения матрицы A.

При решении различных задач возникают разные требования о собственных значениях и собственных векторах матрицы и это порождает многообразие проблем и методов решения этой задачи:

1. Для решения ряда задач механики, физики, химии требуется вычисление всех λ_i , а иногда и всех собственных векторов. Это задача – полная проблема собственных значений.
2. В ряде случаев требуется найти лишь минимальное или максимальное по модулю собственное значение. Возникает в физике. Здесь приходится решать задачи, эквивалентные задаче отыскания собственных значений матриц размерности порядка 10^3 – 10^6. В таких задачах при малых размерностях матриц используются итерационные методы, при больших – вероятностные.
3. При исследовании колебательных процессов иногда требуются определить два максимальные по модулю собственные значения матрицы. Причем меньшее из них обычно достаточно определить с меньшей точностью.

Задачи 2,3 являются частным случаем общей проблемы СЗ и достаточно ограничиться набором методов для решения полной проблемы собственных значений.

Однако такой подход приведет к неоправданно большому объему вычислений.

Рассмотрим типичную задачу нахождения максимальных по модулю СЗ.

Предположим, что A обладает e_i - полная система собственных векторов.

$$Ae_i = \lambda_i e_i, i = \overline{1,n}.$$

$$(e_i, e_j) = \delta_{ij}.$$

$$|\lambda_1| > |\lambda_2| \geq |\lambda_3| \geq \ldots \geq |\lambda_n|.$$

Выберем произвольный вектор x_0 и будем строить последовательность векторов $x_{m+1} = Ax_m, m = 0,1,2,\ldots$

Представим x_0 в виде: $x_0 = \sum_{i=1}^{n} C_i\, e_i$.

Тогда $x_m = \sum_{i=1}^{n} C_i\, A^m e_i = \sum_{i=1}^{n} C_i\, \lambda_i^m e_i$.

Так как у нас λ_1 равняется максимальному по модулю значению, то можно записать $x_m = C_1\lambda_1^m e_1 + o(|\lambda_2|^m)$. Рассмотрим скалярное произведение

$$(x_m, x_m) = (C_1\lambda_1^m e_1 + o(|\lambda_2|^m), C_1\lambda_1^m e_1 + o(|\lambda_2|^m) = C_1^2\lambda_1^{2m}\left(1 + o\left(\left|\frac{\lambda_2}{\lambda_1}\right|^m\right)\right)$$

$$(x_{m+1}, x_m) = \lambda_1 C_1^2\lambda_1^{2m+1}\left(1 + o\left(\left|\frac{\lambda_2}{\lambda_1}\right|^m\right)\right).$$

Положим:

$$\lambda_1^m = \frac{(x_{m+1}, x_m)}{(x_m, x_m)} = \frac{\lambda_1 C_1^2\lambda_1^{2m+1}\left(1+o_1\left(\left|\frac{\lambda_2}{\lambda_1}\right|^m\right)\right)}{C_1^2\lambda_1^{2m}\left(1+o\left(\left|\frac{\lambda_2}{\lambda_1}\right|^m\right)\right)} = \lambda_1 + o_2\left(\left|\frac{\lambda_2}{\lambda_1}\right|^m\right).$$

Таким образом для достаточно больших m мы можем получить с заданной точностью максимальное по модулю λ_1^m и соответствующий ему вектор e_1^m.

Обзор методов численного решения проблемы СЗ

Метод Крылова.

Для иллюстрации основной идеи метода Крылов вводит в рассмотрение каноническую систему обыкновенных дифференциальных уравнений первого порядка с постоянными коэффициентами:

$$\begin{aligned} y_1{}' &= a_{11}y_1 + a_{12}y_2 + \ldots + a_{1n}y_n \\ y_2{}' &= a_{21}y_1 + a_{22}y_2 + \ldots + a_{2n}y_n \\ &\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ y_n' &= a_{n1}y_1 + a_{n2}y_2 + \ldots + a_{nn}y_n, \quad a_{ij} \in A \end{aligned} \qquad \textbf{(3)}$$

Характеристическое уравнение этой системы совпадает с характеристическим уравнением матрицы A. Корни этого уравнения являются собственными значениями матрицы A.

Если заданную систему обыкновенных дифференциальных уравнений первого порядка удается свести к одному дифференциальному уравнению порядка n с постоянными коэффициентами:

$$y^{(n)} = p_1 y^{(n-1)} + p_2 y^{(n-2)} + \ldots + p_{n-1}y' + p_n y,$$

по виду этого уравнения можно выписать его характеристическое уравнение:

$$\lambda^n - p_1\lambda^{n-1} - \ldots - p_n = 0.$$

Крылов указал также на возможность алгебраической интерпретации этой идеи. Рассмотрим любой вектор $c_0 \neq 0$, согласованный по размерности с размерностью матрицы A. По этому вектору будем строить последовательность векторов

$$\begin{aligned} c_1 &= Ac_0; \\ c_2 &= Ac_1 = A^2c_0; \\ &\ldots \\ c_m &= A^m c_0 \end{aligned}$$

до тех пор пока не встретим первый вектор, например, c_m, который будет являться линейной комбинацией предыдущих линейно независимых векторов $c_0, c_1, \ldots, c_{m-1}$, то есть пока не будет выполняться равенство

$$c_m = q_1 c_{m-1} + q_2 c_{m-2} + \ldots + q_m c_0.$$

Если $m = n$, то q_i будут коэффициентами многочлена степени m, тем самым мы построим характеристический многочлен.

Если $m < n$, то q_i будут коэффициентами многочлена

$$P_m(\lambda) = \lambda^m - q_1\lambda^{m-1} - \ldots - q_m = 0$$

который является делителем многочлена $P_n(\lambda)$, т.е. решив $P_m(\lambda) = 0$ мы найдем m собственных значений матрицы A.

Метод Данилевского.

Этот метод построен на известном из линейной алгебры факте о том, что преобразование подобия $S^{-1}AS$ не изменяет характеристического значения матрицы A.

$$A = \begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1n-1} & a_{1n} \\ . & . & . & . & . \\ a_{n1} & a_{n2} & \ldots & a_{nn-1} & a_{nn} \end{pmatrix}$$

$$|S^{-1}AS - \lambda E| = |S'AS - \lambda S^{-1}S| = |S'||A - \lambda E||S| = |A - \lambda E|$$

Поэтому удачно подобрав преобразование подобия можно надеяться получить матрицу, собственный многочлен которой легко выписывается по виду этой матрицы. Данилевский предложил преобразованием подобия приводить исходную матрицу к виду Фробениуса:

$$\Phi = \begin{pmatrix} p_1 & p_2 & \ldots & p_n \\ 1 & 0 & \ldots & 0 \\ . & . & . & . \\ 0 & \ldots & 1 & 0 \end{pmatrix}$$

Тогда характеристический многочлен матрицы Φ совпадает с характеристическим многочленом матрицы A, а именно имеет вид

$$\lambda^n - p_1\lambda^{n-1} - \ldots - p_n = 0.$$

Преобразование подобия S осуществляем по шагам

$$S = M_{n-1}M_{n-2}\ldots M_1,$$

где каждая последующая матрица начиная с M_{n-1} приводит к виду Фробениуса только одну строку, начиная с последней. Сначала приводим к виду Фробениуса последнюю строку матрицы A, то есть в последней строке должны быть все нули и 1 под главной диагональю. Для этой цели A умножаем на ${M_{n-1}}^{-1}$ и M_{n-1}, где

$$M_{n-1} = \begin{pmatrix} 1 & 0 & \dots & 0 & 0 \\ . & . & . & . & . \\ . & . & . & . & . \\ -\frac{a_{n1}}{a_{nn-1}} & -\frac{a_{n2}}{a_{nn-1}} & \dots & \frac{1}{a_{nn-1}} & -\frac{a_{nn}}{a_{nn-1}} \\ 0 & 0 & \dots & 0 & 1 \end{pmatrix}.$$

Вид Фробениуса это последняя строка $(0 \quad 0 \quad \dots \quad 1 \quad 0)$.

$${M^{-1}}_{n-1} = \begin{pmatrix} 1 & 0 & \dots & 0 & 0 \\ . & . & . & . & . \\ . & . & . & . & . \\ a_{n1} & a_{n2} & \dots & a_{nn-1} & a_{nn} \\ 0 & 0 & \dots & 0 & 1 \end{pmatrix}$$

Таким образом преобразование ${M_{n-1}}^{-1} A M_{n-1} = A^{(1)}$.

$$A^{(1)} = \begin{pmatrix} {a^{(1)}}_{1\,1} & {a^{(1)}}_{1\,2} & \dots & {a^{(1)}}_{1n-1} & {a^{(1)}}_{1\,n} \\ . & . & . & . & . \\ . & . & . & . & . \\ {a^{(1)}}_{n-1\,1} & {a^{(1)}}_{n-2\,2} & \dots & {a^{(1)}}_{n-1\,n-1} & {a^{(1)}}_{n-1\,n} \\ 0 & 0 & \dots & 1 & 0 \end{pmatrix}$$

Далее используем преобразование подобия для матрицы $A^{(1)}$, т.е. ${M_{n-2}}^{-1} A^{(1)} M_{n-2}$ и до тех пор пока полностью не будет приведена к виду Фробениуса вся матрица. После нахождения характеристического многочлена и вычисления его корней метод Данилевского позволяет уменьшить объем вычислений при определении собственного вектора. Пусть мы нашли собственное значение λ. Тогда собственный вектор матрицы Фробениуса имеет вид $y = \begin{pmatrix} \lambda^{n-1} \\ \lambda^{n-2} \\ \dots \\ \lambda \\ 1 \end{pmatrix}$, а собственный вектор матрицы A вычисляется по правилу $x = Sy = M_{n-1} M_{n-2} \dots M_1 y$ (сначала $M_1 y$ и дальше последовательно домножаем слева).

Т.е. в методе Данилевского можно вычислить все СЗ и все СВ, отвечающие данным СЗ.

Треугольный степенной метод.

Разрабатывался Бауэром и предназначается для вычисления итерационным путем всех собственных значений матрицы A. При этом предполагается, что для собственных значений матрицы A справедливо

распределение $|\lambda_1| > |\lambda_2| > \ldots > |\lambda_n|$. Пусть $c_0 = (c_{ij}^0), i.j = \overline{1,n}$ - некоторая матрица задаваемая вычислителем. В основе вычислительной схемы метода лежит последовательное вычисление матриц

$$C_k = \begin{pmatrix} 1 & 0 & 0 & \ldots & 0 \\ C_{21}^{(k)} & 1 & \ldots & \ldots & 0 \\ C_{31}^{(k)} & C_{32}^{(k)} & 1 & \ldots & 0 \\ . & . & . & . & . \\ C_{n1}^{(k)} & C_{n2}^{(k)} & \ldots & C_{n\,n-1}^{(k)} & 1 \end{pmatrix}$$

и вычисление матриц

$$R_k = \begin{pmatrix} r_{11}^{(k)} & r_{12}^{(k)} & \ldots & \ldots & r_{1n}^{(k)} \\ 0 & r_{22}^{(k)} & \ldots & \ldots & r_{2n}^{(k)} \\ . & . & . & . & . \\ . & . & . & . & . \\ 0 & 0 & \ldots & \ldots & r_{nn}^{(k)} \end{pmatrix}.$$

Последовательность вычислений матриц осуществляется по правилу:

1. $AC_0 = C_1R_1$, где A заданная матрица, C_0 - некоторая невырожденная матрица, заданная вычислителем, C_1, R_1- искомые матрицы. Характер вычислений при определении коэффициентов матриц C_1, R_1 аналогичен обратному ходу метода Гаусса, то есть не вызывает затруднений.

2. $AC_1 = C_2R_2$ Из этого соотношения определяем C_2, R_2.

Процесс продолжается до тех пор, пока не окажется, что с заданной степенью точности диагональные элементы матриц R_k и R_{k+1} не будут совпадать между собой.

k. $AC_{k-1} = C_kR_k$

Оказывается, что $\lim\limits_{k\to\infty} r_{ii}^{(k)} = \lambda_i$.

Поэтому при достаточно больших k можно предположить

$$\lambda_i = r_{ii}^{(k)}, i = \overline{1,n}.$$

Тем самым можем вычислить все собственные значения матрицы A.

Метод вращений.

Пусть A эрмитова матрица ($A = A^*$). Основой для построения метода служит известная теорема из алгебры, утверждающая, что если A эрмитова матрица, то существует унитарная матрица V, для которой

$V^{-1} = V^*$, а преобразование подобия с этой матрицей приводит A к диагональному виду

$$V^{-1}AV = \Lambda \qquad \textbf{(4)}$$

Λ - диагональная матрица, на диагонали которой стоят собственные значения матрицы A. Так как для унитарной матрицы справедливо $V^{-1} = V^*$, то последнее равенство можно записать в виде

$$V^*AV = \Lambda. \qquad \textbf{(4')}$$

Эта формула не может быть использована для прямого вычисления элементов матриц $\boldsymbol{V}$и $\boldsymbol{\Lambda}$, так как она представляет систему $\boldsymbol{n^2}$ уравнений с $\boldsymbol{n^2 + n}$ неизвестными. Здесь $\boldsymbol{n^2}$ элементов матрицы Vи $\boldsymbol{n}$ элементов матрицы $\boldsymbol{\Lambda}$. Однако можно трактовать задачу приведения матрицы A к диагональному виду, как приближенную задачу.

Предположим, что некоторым преобразованием вида (5) матрица $\boldsymbol{A}$ приведется к некоторой матрице

$$\widetilde{\boldsymbol{\Lambda}} = \begin{pmatrix} \tilde{\boldsymbol{\lambda}}_{\mathbf{1}} & \tilde{\boldsymbol{\lambda}}_{\mathbf{12}} & \dots & \tilde{\boldsymbol{\lambda}}_{\mathbf{1}\boldsymbol{n}} \\ \tilde{\boldsymbol{\lambda}}_{\mathbf{21}} & \tilde{\boldsymbol{\lambda}}_{\mathbf{2}} & \dots & \tilde{\boldsymbol{\lambda}}_{\mathbf{2}\boldsymbol{n}} \\ . & . & . & . \\ \tilde{\boldsymbol{\lambda}}_{\boldsymbol{n}\mathbf{1}} & \tilde{\boldsymbol{\lambda}}_{\boldsymbol{n}\mathbf{2}} & \dots & \tilde{\boldsymbol{\lambda}}_{\boldsymbol{n}} \end{pmatrix}$$

Предположим также, что внедиагональные элементы матрицы $\tilde{\Lambda}$ таковы, что ими можно пренебречь в условиях заданной точности. Тогда на диагонали стоят элементы, совпадающие с заданной точностью с собственными значениями матрицы A.

Пусть матрица $A = A^T$, $a_{ij} \in R$. Для такой матрицы метод вращений заключается в построении последовательности матриц $A^{(0)} = A; A^{(1)}; A^{(2)}; \ldots$ каждая последующая получается из предыдущей при помощи элементарного шага, состоящего из преобразования подобия путем умножения предыдущей матрицы на некоторую матрицу, называемую матрицей вращений, имеющей вид

$$V_{ij}(\phi) = \begin{pmatrix} 1 & 0 & & & & & 0 \\ & . & & & & & \\ & & \cos\phi & & & -\sin\phi & \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ & & \sin\phi & & & \cos\phi & \\ & & & & & & \\ 0 & & & & & & 1 \end{pmatrix} \begin{matrix} \\ \\ i \\ \\ j \\ \\ \\ \end{matrix}$$

$$\qquad\qquad i \qquad\qquad\qquad j$$

где

$$\cos\varphi = \sqrt{\frac{1}{2}\left(1 + \frac{1}{\sqrt{1+\mu^2}}\right)},$$

$$\sin\varphi = sign\,\mu\sqrt{\frac{1}{2}\left(1 - \frac{1}{\sqrt{1+\mu^2}}\right)},$$

$$\mu = \frac{2a_{ij}}{a_{ii} - a_{ij}}, \qquad a_{ii} - a_{ij} \neq 0.$$

Здесь a_{ij} – максимальный по модулю из внедиагональных элементов предыдущей матрицы.

QR-алгоритм

В настоящее время, это основной метод, который используется при нахождении полной проблемы собственных значений.

Основная идея QR-алгоритма состоит в разложении исходной матрицы в произведение ортогональной и верхнетреугольной матрицы.

Предположим, что:

$$\lambda_1, \dots, \lambda_n;\ |\lambda_1| > |\lambda_2| > \dots > |\lambda_n|.$$

Порядок построения $\{A_i\}_{i=1}^{\infty}$.

Предполагаем, что $A_0 = A$.

Вычисляем разложение A_0 в виде произведения ортогональной матрицы Q_0 и верхнетреугольной R_0:

$$A_0 = Q_0 R_0.$$

Далее образуем $A_1 = Q_0 R_0$ и ищем представление $A_1 = Q_1 R_1$, где Q_1 - ортогональная, R_1 - верхтреугольная.

$$A_2 = Q_1 R_1,$$

…

Аналогичным образом продолжаем процесс.

QR-алгоритм легко осуществляется с помощью преобразования Хаусхолдера.

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix}$$

$$W_1^T = \mu(a_{11} - S, a_{21}, \dots, a_{n1}),$$

где $S = \left(\sum_{j=1}^{n} a_{j1}^2\right)^{\frac{1}{2}}, \mu = \left(2S(S - a_{11})\right)^{-\frac{1}{2}}$.

Поскольку $A_0 = A$, то:

$$A_0^{(1)} = (E - 2W_1 W_1^T)A_0 = \begin{pmatrix} * & * & \dots & * \\ 0 & * & \dots & * \\ \dots & \dots & \dots & \dots \\ 0 & * & \dots & * \end{pmatrix}$$

Вычичсляем:

$$W_2^T = \hat{\mu}\left(0, \hat{a}_{22} - \hat{S}, \hat{a}_{32}, \dots, \hat{a}_{n2}\right), \text{где}$$

$$\hat{S} = \left(\sum_{j=2}^{n} a_{j2}^2 \right)^{\frac{1}{2}}, \hat{\mu} = \left(2\hat{S}(\hat{S} - \hat{a}_{22}) \right)^{-\frac{1}{2}}$$

Где $\hat{}$ означает, что рассматриваются элементы $A_0^{(1)}$.

Тогда $A_0^{(2)} = (E - 2W_2W_2^T)A_0^{(1)}$- матрица, у которой в первых двух столбцах элементы под главной диагональю $= 0$.

Продолжая этот процесс с помощью векторов W_i через $n-1$ шаг получим $A_0^{(n-1)} = (E - 2W_{n-1}W_{n-1}^T) \dots (E - 2W_2W_2^T)(E - 2W_1W_1^T)A_0$, которая будет являться верхнетреугольной. Обозначим ее $A_0^{(n-1)} = R_0$.

$Q_0 = (E - 2W_1W_1^T)(E - 2W_2W_2^T) \dots (E - 2W_{n-1}W_{n-1}^T)$.

Тогда $Q_0^T A_0 = R_0$.

Т.к. каждая матрица $(E - 2W_iW_i^T)$ является ортогональной, то произведение их с Q_0 также ортогонально.

Значит: $Q_0^{-1} = Q_0^T$.

Т.о. $A_0 = Q_0R_0$.

Эти все преобразования - 1-й шаг QR-алгоритма.

Далее определяется $A_1 = Q_0R_0$ и для нее снова ищется разложение вида $A_1 = Q_1R_1$ т.е. по матрице A_1 строим $A_1^{(n-1)}$ и т.д.. Процесс продолжаем до тех пор, пока диагональные элементы матрицы $A^{(n)}$ не будут совпадать в пределах заданной точности с диагональными элементами матрицы $A^{(n-1)}$. Тогда собственные значения матрицы А это

$$\lambda_i(A) = \lim_{k \to \infty} \lambda_i(A_k^{(n-1)}) = \lim_{k \to \infty} a_{ii}^{(k)}; i = 1, \dots, n.$$

Решение систем нелинейных уравнений. Постановка задачи. Метод итерации.

Задача решения уравнений в достаточно общем виде может быть сформулирована так:

Пусть заданы X и Y, элементы, которых будем обозначать x, y. Природа этих элементов может быть любой. Это могут быть числа, функции, линии, поверхности и так далее.

Говорят, что в некотором X задан оператор A со значениями во множестве Y, если для каждого элемента из X существует некоторый $y = Ax \in Y$. Этот оператор можно рассматривать как преобразование Xв Y. x - оригинал, y - изображение.

Допустим, что в X задан оператор A со значениями в Y. Допустим также, что во множестве Y взят некоторый элемент y_0 Нужно найти такие $x \in X$, которые являются для y_0 оригиналами.

Записать такую задачу можно в виде уравнения

$$Ax = y_0 \quad \textbf{(1)}$$

Особое значение для нас будут иметь уравнения, в которых неизвестными величинами будут числа. Такие уравнения являются частным случаем операторных уравнений, когда X и Y - численные пространства конечных размерностей. В этом случае

$$f(x) = 0, \quad f: R^n \to R^n \quad \textbf{(2)}$$

$$\begin{cases} f_1(x_1, \dots, x_n) = 0 \\ f_2(x_1, \dots, x_n) = 0 \\ \dots \\ f_n(x_1, \dots, x_n) = 0 \end{cases} \quad \textbf{(3)}$$

системы вида (3) имеют очень важное значение в науке и ее приложениях. Много задач сводятся к системам вида (3). Одним из наиболее распространенных методов решения системы (3) является метод итерации или метод повторных подстановок, который является общим методом решения нелинейных уравнений и применим к весьма

широкому классу операторных уравнений. Общность метода, его простота и удобная реализация сделали метод итерации одним из основных методов вычислительной математики.

Приведем (2) к виду

$$x = \varphi(x) - \textit{каноническое уравнение.} \qquad \textbf{(4)}$$

Множество значений, которое может принимать x обозначим X. В приложениях X чаще всего конечный или бесконечный отрезок числовой прямой. Множество значений, которое может принимать $\varphi(x)$, когда $x \in X$, обозначим Y. Функцию $\varphi(x)$ можно рассматривать как оператор, преобразующий X в Y. При использовании метода итераций мы должны обозначить свои промежутки принадлежности корней. Для каждого промежутка расчеты нужно производить заново. Уравнение (4) означает, что нужно найти такие $x \in X$, которые при преобразовании множества X оператором φ переходят сами в себя, иначе говоря, во множестве X нужно найти точки, остающиеся неподвижными при преобразовании множества X оператором φ. В методе итерации последовательные приближения вычисляются по правилу:

$$x_{k+1} = \varphi(x_k), \qquad k = 0,1,2, \ldots \qquad \textbf{(5)}$$

x_0 - начальное приближение.

Для начала построений последовательных приближений нужно уравнение в виде (2) привести к каноническому виду (4). Одним из способов такого приведения является преобразование

$$x = x + A(x)f(x),$$

где $A(x) = (a_{ij}(x)), ij = \overline{1,n}$ выбирается так, чтобы она не обращалась в ноль для $x \in X$. При этом (4) должно быть выбрано т.о., чтобы φ было преобразованием сжатия, т.е.

$$\|\varphi(x) - \varphi(y)\| \le q\|xy\|,$$
$$0 < q < 1, \qquad q << 1,$$

в области, где ищется решение (2) и задано начальное приближение $\boldsymbol{x_0}$. Кроме того, $\boldsymbol{x_0}$ и $\boldsymbol{q}$ и размер области должны быть согласованы между собой.

При решении уравнений методом итераций речь всегда идет только о вычислении 1 корня, т.к. для нахождения другого надо находить другие $\boldsymbol{\varphi}$, начальное приближение и область.

Теорема 1 (о сходимости метода итерации)

Пусть x_0- начальное приближение к корню x^* уравнения $f(x)=0$. $\{x_k\}_{k=0}^{\infty}$ итерационная последовательность, построенная по правилу:

$x_{k+1}=\phi(x_k);\quad \varphi(x)=x+A(x)f(x)$

Пусть выполняются условия:

1. $\varphi_i(x),\quad i=\overline{1,n}$ определены и непрерывны по всем $x_j, j=\overline{1,n}$ в области $\overline{S}=\overline{S}(x_0,r)=\{x\in R^n | \|x-x_0\|\le r, r>0\}$.

2. $\varphi(x)$ являются отображением сжатия в $\overline{S}$, то есть для $\forall x,z\in\overline{S}$ выполняется $\|\varphi(x)-\varphi(z)\|\le q\|x-z\|,\quad 0<q<1$

3. Функция $\varphi(x)$, числа q,r и x_0 согласованы так, что выполняется неравенство: $\frac{\|\varphi(x_0)-x_0\|}{1-q}\le r.$

Тогда

$\alpha)$ Все $x_k\in\overline{S}(x_0,r)$.

$\beta)$ $\lim\limits_{k\to\infty}x_k=x^*,\ \ x^*=\varphi(x^*);\qquad x^*\in\overline{S}(x_0,r).$

$\gamma)$ Справедлива оценка скорости сходимости

$$\|x_k-x^*\|\le\frac{\|\varphi(x_0)-x_0\|}{1-q}q^k.$$

Доказательство:

Докажем $\alpha)$ методом математической индукции.

$x_0\in\overline{S}(x_0,r)$ - очевидно.

Предположим, что $x_1,\dots,x_k\in\overline{S}(x_0,r)$.

Покажем, что $x_{k+1}\in\overline{S}(x_0,r)$.

$$\|x_{k+1}-x_0\|\le\|x_{k+1}-x_k\|+\|x_k-x_{k-1}\|+\dots+\|x_1-x_0\|$$

$$\|x_{k+1}-x_k\|=\|\varphi(x_k)-\varphi(x_{k-1})\|\le q\|x_k-x_{k-1}\|\le\dots\le q^k\|x_1-x_0\|\ \ \textbf{(*)}$$

$$\|x_1-x_0\|=\|\varphi(x_0)-x_0\|=m_0.$$

Применяя оценку (*) ко всем звеньям в неравенстве

$$\|x_{k+1}-x_k\|+\|x_k-x_{k-1}\|+\dots+\|x_1-x_0\|\le q^km_0+q^{k-1}m_0+\dots+m_0\le$$
$$\le\frac{1}{1-q}m_0\le r\ \ \Rightarrow\ \ x_{k+1}\in\overline{S}(x_0,r).$$

$\beta)$ Для доказательства сходимости $\{x_k\}_{k=0}^{\infty}$ применим признак Больцано-Коши.

$$\|x_{k+p}-x_k\|\le\|x_{k+p}-x_{k+p-1}\|+\dots+\|x_{k+1}-x_k\|\le$$

$$\leq q^{k+p-1}m_0+\ldots+q^k m_0 \leq \frac{q^k}{1-q}m_0 \quad \textbf{(6)}$$

Правая часть (6) не зависит от p и так как $0 < q < 1$, то $\forall \varepsilon > 0$ начиная с некоторого значения k правая часть станет меньше ε при любом p. Поэтому существует $\lim\limits_{k\to\infty}\{x_k\}_{k=0}^{\infty}$, т. е. $\lim\limits_{k\to\infty} x_k = x^*$. И так как все $x_k \in \overline{S}(x_0, r)$, то x^* также будет принадлежать $\overline{S}(x_0, r)$.

Чтобы показать, что x^* является корнем уравнения $x = \varphi(x)$ надо в левой и правой части равенства $x_{k+1} = \varphi(x_k)$ перейти к пределу. Так как $x_{k+1} \to x^*$ и $\varphi(x_k) \to \varphi(x^*)$, то $x^* = \varphi(x^*)$.

γ) Мы получили (6) $\left\|x_{k+p} - x_k\right\| \leq \frac{q^k}{1-q}m_0$. Если устремить $p \to \infty$, то так как $x_{k+p} \to x*$, в пределе будем иметь $\|x_k - x^*\| \leq \frac{q^k}{1-q}m_0$.

Теорема 2

Во всяком множестве точек, где для $\varphi(x)$, $\forall x, y$ справедливо неравенство:

$$\|\varphi(x) - \varphi(y)\| < \|x - y\|.$$

Уравнение $x = \varphi(x)$ имеет не более одного решения.

Доказательство:

Допустим противоположное. Будем считать, что это уравнение имеет два различных решения x, y, $x \neq y$.

$\|x - y\| = \|\varphi(x) - \varphi(y)\| < \|x - y\|$ в силу условия теоремы. Получили противоречие.

Замечание

Итерационный процесс сходится со скоростью геометрической прогрессии и, чтобы скорость была удовлетворительной, $\phi(x)$ должна быть выбрана так, чтобы $0 < q < 1$.

Этого можно добиться за счёт удачного выбора начального приближения, области $S(x_0, r)$, а также $\varphi(x)$.

Есть наборы на правильное нахождение $\varphi(x)$.

Об ускорении сходимости простой одношаговой итерации

Проблема ускорения сходимости является одной из общих задач вычислительной математики и может возникать всюду, где для разыскиваемого элемента строится последовательность вычисляемых элементов, неограниченно, в той или иной мере, приближающихся к нему.

Рассмотрим метод простой итерации. Эйткен построил правила улучшения сходимости для сходящейся последовательности $\boldsymbol{S_n}$,

$n = 0,1,2 \dots, S_n \to S$, у которых погрешность $\varepsilon_n = S_n - S$ изменяется по закону, близкому к геометрической прогрессии:

$$\varepsilon_{n+1} \approx q\varepsilon_n, \qquad 0 < q < 1.$$

Указанный закон изменения ε_n весьма прост и по нему легко определяется представление любого члена последовательности S_n, для которой он выполняется. Так как в приближенном равенстве $\varepsilon_{n+1} \approx q\varepsilon_n$ погрешность должна быть малой величиной порядка выше, чем $q\varepsilon_n$, то эта погрешность должна иметь представление $q\varepsilon_n\eta_n, \eta_n \to 0, n \to \infty$. Тогда можно записать точное равенство $\varepsilon_{n+1} = q\varepsilon_n(1 + \eta_n)$. Если применить это равенство несколько раз начиная с ε_n, то получится последовательность равенств, последний член которой даст нужное представление ε_n через ε_0:

$$\varepsilon_n = q\varepsilon_{n-1}(1 + \eta_{n-1}) = q^2\varepsilon_{n-2}(1 + \eta_{n-1})(1 + \eta_{n-2}) = \cdots$$
$$\dots = q^n\varepsilon_0(1 + \eta_{n-1})(1 + \eta_{n-2}) \dots (1 + \eta_0).$$

Тогда для последовательности S_n получим:

$$S_n = S + Aq^n(1 + \eta_0) \dots (1 + \eta_{n-1}), \qquad (1)$$
$$A = \varepsilon_0, \qquad \eta_n \to 0, n \to \infty.$$

За каноническую принято считать последовательность, для которой все $\varepsilon_n = 0$.

$$S_n = Aq^n. \qquad (2)$$

Эта последовательность зависит от трех величин S, A, q. Они могут быть найдены по трем любым значениям S_n. Нас интересует предельное значение S. Чтобы определить его, рассмотрим равенство (2) при трех последовательных значениях n.

$$S_{n-1} - S = Aq^{n-1},$$
$$S_n - S = Aq^n,$$
$$S_{n+1} - S = Aq^{n+1}.$$

Разделим второе равенство на первое и третье на второе. Получим:

$$\frac{S_n - S}{S_{n-1} - S} = q,$$
$$\frac{S_{n+1} - S}{S_n - S} = q.$$

Приравняем две части и выразим S:

$$S = \frac{S_{n+1}S_{n-1} - S_n^2}{S_{n+1} - 2S_n + S_{n-1}}. \quad (3)$$

Равенство (3) можно применить не к канонической последовательности, а к любой последовательности, близкой к канонической. Но тогда мы получим не предельное значение S, а получится значение тем ближе к S, чем больше n. Это значение обозначим через:

$$\sigma_n = \frac{S_{n+1}S_{n-1} - S_n^2}{S_{n+1} - 2S_n + S_{n-1}}, S_{n+1} - 2S_n + S_{n-1} \neq 0 \quad (4)$$

Равенство (4) называется пребразованием Эйткена, заданным последовательностью S_n в новую последовательность σ_n. Оно применимо ко всякой последовательности S_n, для которой знаменатель

$$S_{n+1} - 2S_n + S_{n-1} \neq 0.$$

Изложенное выше позволяет утверждать, что последовательность σ_n будет сходиться к S быстрее, чем последовательность S_n.

Возвратимся к методу итераций:

$$x_{n+1} = \varphi(x_n), n = 0,1,2, \ldots \quad (5)$$
$$x^* + \varepsilon_{n+1} = \varphi(x^* + \varepsilon_{n+1}) = \varphi(x^*) + \varepsilon_{n+1}\varphi'(x^*) + O(\varepsilon_n^2)\,, \quad (6)$$
$$\varepsilon_{n+1} \approx \varphi'(x^*)\varepsilon_n. \quad (7)$$

Для метода итераций, при выполнения условия $\mathrm{q} = \left|\varphi'^{(x^*)}\right| < 1$, закон изменения погрешности ε_n имеет как раз такую же форму, которая используется при построении правила Эйткена, но в данном случае, как было предложено Стеффенсоном, этот итерационный процесс изменяют таким образом, чтобы он стал одношаговым, то есть найденное улучшенное приближение сразу используется при вычислении.

Обозначим $x_n' = \varphi(x_n), x_n'' = \varphi(x_n') = \varphi\big(\varphi(x_n)\big)$,

$$x_{n+1} = \frac{x_n''x_n - (x_n')^2}{x_n'' - 2x_n' + x_n} = \frac{\varphi(\varphi(x_n))x_n - (\varphi(x_n))^2}{\varphi(\varphi(x_n)) - 2\varphi(x_n) + x_n}\,. \quad (8)$$

Равенство (8) называется итерационной формулой Стеффенсона.

Это правило является одношаговым и требует дополнительно вычисления двух значений φ на каждом шаге.

Доказано, что при использовании правила (8) для погрешности

$$\varepsilon_{n+1} = B\varepsilon_n^2,$$

где B – константа, не завивсящая от ε_n, то есть при использовании правила Эйткена, погрешность изменяетя по квадратичному закону в отличие от линейного закона, который получается при использовании метода итераций.

Метод Ньютона для операторных уравнений

Этот метод, так же, как и метод итераций, является общим методом и может применяться для решения нелинейных уравнений многих видов.

Его значение: он позволяет решение нелинейных задач привести к решению последовательности линейных задач, каждая из которых более доступна в решении, чем нелинейные.

X и Y - полные линейные нормированный пространства, $x \in X$, $y \in Y$.

В области D пространства X определен нелинейный оператор $y = f(x), f(x) \in Y$.

$f(x)$ предполагается дважды дифференцируемым в смысле Фреше.

Определение

Оператор f дифференцируем в смысле Фреше на x, если существует оператор $H: X \to Y$, что

$$\|f(x + \triangle x) - f(x) - H(\triangle x)\| \leq \|\triangle x\| E(\|\triangle x\|),$$

где $E(\delta) \underset{\delta \to 0}{\to} 0$.

Оператор H называют производной f на элементах x.

$$\dot{H} = f'(x)$$

Линейное преобразование $H(\triangle x)$ имеет смысл дифференциала оператора.

Рассмотрим нелинейное уравнение

$$f(x) = 0 \qquad (1)$$

В правой части $0 \in Y$ (нулевой элемент).

Будем считать, что знаем нулевое приближение x_0 к решению (1).

Правило нахождения следующего приближения по предыдущему:

$$f(x) = f(x_0) + (f(x) - f(x_0))$$

Вместо разности возьмем дифференциал оператора на ${x_0}'$. $(f'(x_0)(x - x_0))$

Заменим (1) приближенным равенством:

$$f(x_0) + f'(x_0)(x - x_0) \approx 0 \qquad (2)$$

В предположении, что существует оператор $(f'(x_0))^{-1}$, переводящий Y в X, для x_1 из (2) можно получить следующее выражение:

$$x_1 = x_0 - \left(f'(x_0)\right)^{-1} f(x_0)$$

Повторяя этот процесс, получим в общем виде правило нахождения следующего приближения по предыдущему:

$$x_{n+1} = x_n - \left(f'(x_n)\right)^{-1} f(x_n), n = 0, 1, 2, \ldots \quad (4)$$

(4) возможно, если:

1. $x_n;\ n = 0,1,2, \ldots : x_n \in D \Longrightarrow$ все x_n принадлежат D

2. Существует $\left(f'(x_n)\right)^{-1}, n = 0, 1, 2, \ldots$

Справедлива теорема о сходимости метода Ньютона.

Теорема

Пусть выполняются:

1) $f(x)$ определен на замкнутом шаре и дважды дифференцируем там:

$$\|x - x_0\| \leq \delta,$$

и $D^2(f)$ по норме ограничена константой k:

$$\|f''(x)\| \leq k.$$

2) $f(x)$ имеет $f'^{-1}(x)$ в точке x_0: $\Gamma_0 = \left(f'(x_0)\right)^{-1}$ и известна оценка его нормы $\|\Gamma_0\| \leq b$.

На начальном приближении x_0 выполняются:

3) $\|\Gamma_0 f(x_0)\| \leq \eta$

4) k, b, η связаны соотношением:

$$h = k\, b\, \eta \leq \frac{1}{2}$$

5) для δ выполнено:

$$\left(\frac{1 - \sqrt{1 - 2h}}{h}\right)\eta \leq \delta$$

Тогда:

1) $f(x) = 0$ имеет в области (5) решение.
2) Последовательные приближения $x_n, n = 0,1,2, \ldots$ могут быть построены по правилу (4), которое называется методом Ньютона для операторных уравнений и сходящимся к решению (1):

$$\lim_{n\to\infty} x_n = x^*$$
$$f(x^*) = 0.$$

3) Скорость сходимости итерационного процесса:

$$\|x^* - x_n\| \leq t^* - t_n$$

где t_n - последовательность Ньютоновских приближений:

$$t_{n+1} = t_n - \frac{P(t_n)}{P_1(t_n)}$$

а t^* - меньший корень уравнения $P(t) = 0$

$$P(t) = \frac{1}{2}\, k\, t^2 - \frac{1}{b}\, t + \frac{\eta}{b}$$

Замечание

Основными требованиями, которые должны выполняться для сходимости итерационного процесса являются условия (4) и (5), где k, b, η - просто вычисляемые константы. А условия (4), (5) будут выполняться, если начальное приближение x_0 выбрано достаточно близки к корню уравнения $\Longrightarrow \eta \to 0$.

При решении некоторых уравнений проверка условий теоремы часто оказывается более затруднительным, чем нахождение решения. В этом случае корни находятся методом вычислительного эксперимента: выбрать x_0 и построить итерационный процесс. Если он сходящийся, то строят решение. В противном случае стараются лучше подобрать x_0.

Метод Ньютона для систем уравнений.

Пусть дано

$$f(x) = 0, \quad \textbf{(1)}$$

где x – числовая переменная, $f(x)$ - достаточно гладкая функция. Обозначим через x^* точное решение уравнения $f(x) = 0$.

Предположим, что каким-либо образом указано для x^* начальное приближение x_0. Последовательные приближения в методе Ньютона строятся по правилу:

$$x_{n+1} = x_n - \frac{f(x_n)}{f\prime(x_n)}, \quad \textbf{(2)}$$
$$x_0,\ n = 0,\ 1,\ 2,\ \ldots$$
$$f'(x_n) \neq 0$$

В метода Ньютона для операторного уравнения $(f'(x_n))^{-1}$ – матрица. Здесь это число.

Условиями возможности построения итерационного процесса (2) являются следующие требования:

1) Все x_nпринадлежат области определения функции $f(x)$;

2) $f'(x_n) \neq 0,\ n = 0,\ 1,\ 2,\ \ldots$;

(2) имеет простой геометрический смысл.

Корень – абсцисса точки пересечения графика функции с Ох. Отметим т.$M_n(x_n, f(x_n))$ на графике и проведём касательную в этой точке к $y = f(x)$. Уравнение касательной

$$y - f(x_n) = f'(x_n)(x - x_n)$$

Положив $y = 0$ в том равенстве и приняв за следующее приближение абсциссу точки пересечения касательной с Ох, из заданного равенства получим $x_{n+1} = x_n - \frac{f(x_n)}{f\prime(x_n)}$, что совпадает со значением полученным по (2).

Поэтому метод Ньютона часто называют методом касательной.

Выясним, как ведут себя последовательные приближения x_n вблизи точного решения x^*. Обозначим $\varepsilon_n = x^* - x_n$. Из равенства $x^* = x^*$ вычтем почленно уравнение (2), получим:

$$\varepsilon_{n+1} = \varepsilon_n + \frac{f(x_n)}{f'(x_n)} = \frac{\varepsilon_n f'(x_n) + f(x_n)}{f'(x_n)} \quad \textbf{(3)}$$

Воспользовавшись рядом Тейлора заметим, что

$$0 = f(x^*) = f(x_n + \varepsilon_n) = f(x_n) + f'(x_n)\varepsilon_n + \frac{1}{2}\varepsilon_n^2 f''(x_n + \theta\varepsilon_n).$$

Заменим $f(x_n) + f'(x_n)\varepsilon_n$ на $-\frac{1}{2}\varepsilon_n^2 f''(x_n + \theta\varepsilon_n)$, тогда получим:

$$\varepsilon_{n+1} = \frac{-\frac{1}{2}\varepsilon_n^2 f''(x_n+\theta\varepsilon_n)}{f'(x_n)} = \frac{-\frac{1}{2}f''(x^*)}{f'(x^*)}\varepsilon_n^2 + o(\varepsilon_n^2) \qquad \textbf{(4)}$$

Погрешность в методе Ньютона на каждом шаге изменяется по квадратичному закону в отличии от закона геометрической прогрессии, который имеет для метода итераций.

Теорема о сходимости метода Ньютона для одного уравнения такая же, что и для операторных, но здесь числовая функция.

Теорема 1

Пусть выполнены следующие условия:

1) функция $f(x)$ определена и дважды непрерывно-дифференцируема на отрезке $\|x - x_0\| < \delta$, (α), при этом $f''(x) \le k$, $\forall x \in \alpha$;

2) $f'(x_0) \neq 0$, $\frac{1}{|f'(x_0)|} \le B$;

3) $\left|\frac{f(x_0)}{f'(x_0)}\right| \le \eta$;

4) $h = kB\eta \le \frac{1}{2}$;

5) $\frac{1-\sqrt{1-2h}}{h}\eta \le \delta$;

Тогда итерационный процесс может быть построен по (2). Он сходится к корню $f(x^*) = 0$ на отрезке $|x - x_0| \le \delta$.

Верна оценка для погрешности $|x^* - x_n| \le \frac{1}{2^{n-1}}(2n)^{2^{n-1}}\eta$

Следует отметить, что условия теоремы будут выполняться, если начальное приближение x_0 расположено вблизи корня x^*

Замечание

Метод Ньютона – метод линеаризации, позволяющий сводить решение нелинейных задач к последовательному решению ряда линейных задач. Он является одним из старейших вычислительных методов решения нелинейных уравнений.

Рассмотрим несколько модификаций этого метода. Все его изменения связаны либо с увеличением скорости сходимости, либо с

упрощением и уменьшением объёма вычислений, так как метод Ньютона - частный случай общего метода простой итерации.

$x_{n+1} = \varphi(x_n),\ \varphi(x) = x - \frac{f(x)}{f'(x)}$, то для него справедливо правило увеличения скорости сходимости Эйткена. Так как основной объём вычислений в методе Ньютона связан с нахождением и обращением $f'(x_n)$, то модификации связанные с уменьшением объёма вычислений основаны на упрощении вычисления $f'(x_n)$.

Метод секущих.

В методе секущих $f'(x_n)$ заменяется на $f'(x_n) = \frac{f(x_n) - f'(x_{n-1})}{x_n - x_{n-1}}$, тогда

$$x_{n+1} = x_n - \frac{(x_n - x_{n-1})f(x_n)}{f(x_n) - f(x_{n-1})} = \frac{x_{n-1}f(x_n) - x_n f(x_{n-1})}{f(x_n) - f(x_{n-1})} \quad \textbf{(6)}.$$

Итерационный процесс (6) может быть осуществлён, если:

1. Все x_n принадлежат области определения функции $f(x)$
2. На каждой итерации выполняется неравенство:

$$f(x_n) - f(x_{n-1}) \neq 0,\ n = 1, 2, \ldots$$

В отличии от метода Ньютона, этот метод является 2-х шаговым и для начала вычислительного процесса нужно знать начальное приближения x_0, x_1 к решению (x^*) уравнения (1).

Погрешность в (6) изменяется по правилу: $\varepsilon_{n+1} \approx -\frac{f''(x^*)}{2f'(x^*)}\varepsilon_n\varepsilon_{n-1}$.

т.е. по закону близкому к квадратичному закону в основном методе Ньютона.

Метод Ньютона с постоянной касательной.

В этой модификации метод Ньютона освобождается от вычисления производных $f'(x_n)$ на каждом шаге и пользуется только лишь производной $f'(x_0)$. Итерационный процесс строится по правилу:

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_0)},\ n = 1, 2, \ldots \quad \textbf{(7)}$$

Объём вычислений в (7) значительно меньше, чем в методе Ньютона, но погрешность в нём изменяется не по квадратичному закону, а по

закону геометрической прогрессии со знаменателем $q = 1 - \frac{f'(x^*)}{f'(x_0)}$, и если точка x_0 близка к решению x^*, то q близок к нулю, то есть скорость сходимости достаточно высока. В этом случае метод (7) является быстросходящимся.

Демпфированный метод Ньютона.

Рассмотрим этот метод на примере решения систем нелинейных уравнений $\Phi(x) = 0$, $\Phi : R^n \to R^n$. Итерационный процесс строится по правилу:

$$x_{n+1} = x_n - \lambda_n \left(\frac{\partial \Phi}{\partial x}(x_n)\right)^{-1} \Phi(x_n), \qquad \textbf{(8)}$$

$0 \le \lambda_n \le 1$, λ_n выбираем из условия, чтобы функция

$$\bar{f}(x) = \sum_{i=1}^{n} \alpha_i^2 \, \Phi_i^2(x), \; \alpha_i \neq 0,$$

λ_n находится из $\bar{f}(x_{n+1}) < \bar{f}(x_n)$. Однако процесс его нахождения сложен.

Преимущество этого метода заключается в том, что он сходится к решению с любого начального приближения, выбранного на промежутке, где существует и при том единственный корень системы $\Phi(x) = 0$.

Ещё раз о целях модификации:

1. упрощение вычислений;
2. уменьшение объёма вычислений.

Также метод Ньютона будет сходится при удачном выборе начального приближения.

Выбор начального приближения может осуществляться градиентными методами и методами оптимизации.

Метод Ньютона для систем.

Рассмотрим систему нелинейных уравнений

$$f(x) = 0, \qquad \textbf{(9)}$$

$f : R^n \to R^n$. Запишем её подробнее:

$$\begin{cases} f_1(x_1, \dots, x_n) = 0 \\ \dots\dots\dots\dots\dots \\ f_n(x_1, \dots, x_n) = 0 \end{cases} \quad \textbf{(10)}$$

Метод Ньютона для системы (10) является обобщённым случаем для одного уравнения.

Пусть нам известно x_0 - начальное приближение к решению системы (10). Итерационный процесс строится следующим образом: каждое последовательное приближение к решению системы (10) находится из системы линейных алгебраических уравнений, составленных по предыдущему приближению, а именно, решения находятся как решения системы

$$\sum_{j=1}^{n} \frac{\partial f_i(x^{(k)})}{\partial x_j} \left(x_j^{(k+1)} - x_j^{(k)}\right) = f_i\left(x^{(k)}\right),\ i = \overline{1,n} \quad \textbf{(11)}$$

Формулировка теоремы о сходимости метода зависит от нормы матрицы, в которой будем рассматривать систему. Сразу рассмотрим случай кубической, октаэдрической, сферической нормы.

Теорема.

a – кубическая,

b – октаэдрическая

c - сферическая

Пусть выполнены условия:

1. $f_i(x) = f_i(x_1, \dots, x_n), \quad i = \overline{1,n}$ определены и дважды непрерывно дифференцируемы в области

(a): $\left|x_i - x_i^{(0)}\right| \le \delta,\ i = \overline{1,n}$

(b): $\sum_{i=1}^{n} \left|x_i - x_i^{(0)}\right| \le \delta$

(c): $\sum_{i=1}^{n} \left(x_i - x_i^{(0)}\right)^2 \le \delta^2$

При этом для вторых производных в области справедливы неравенства:

a) $\sum_{j,k=1}^{n} \left|\frac{\partial^2 f_i(x)}{\partial x_j \partial x_k}\right| \le K,\ i = \overline{1,n};$

b) $\left|\frac{\partial^2 f_i(x)}{\partial x_j \partial x_k}\right| \le L_i,\ L_1 + L_2 + \dots + L_n = K,\ j,k = \overline{1,n};$

c) $\sum_{i,j,k=1}^{n}\left(\frac{\partial^2 f_i(x)}{\partial x_j \partial x_k}\right)^2 \le K^2$;

2. Значения $x_1^{(0)},\dots,x_n^{(0)}$ являются начальным приближением к решению системы (10) и для справедливы неравенства:

a) $\left|f_i\left(x^{(0)}\right)\right| \le \eta,\ i=\overline{1,n}$;

b) $\sum_{i=1}^{n}\left|f_i\left(x^{(0)}\right)\right| \le \eta$;

c) $\sum_{i=1}^{n}\left(f_i\left(x^{(0)}\right)\right)^2 \le \eta^2$;

3. Матрица Якоби $f'(x)$ имеет в точке $x^{(0)}$ определитель $D\left(f'\left(x^{(0)}\right)\right)$ отличный от нуля и если D_{jk} есть алгебраическое дополнение к элементу $\frac{\partial f_j(x^{(0)})}{\partial x_n}$, то:

a) $\frac{1}{D}\sum_{j=1}^{n}\left|D_{jk}\right| \le B,\ k=\overline{1,n}$;

b) $\frac{1}{D}\sum_{k=1}^{n}\left|D_{jk}\right| \le B,\ j=\overline{1,n}$;

c) $\frac{1}{D}\left(\sum_{j,k=1}^{n}\left(D_{jk}\right)^2\right)^{\frac{1}{2}} \le B$;

4. Для чисел K, B, η выполняется условие $h = KB^2\eta \le \frac{1}{2}$;

5. Для числа δ выполняется условие $\frac{1-\sqrt{1-2h}}{h}B\eta \le \delta$,

Тогда:

1. Система $f(x)=0$ имеет в области (α) решение $x^* = (x_1^*,\dots,x_n^*)^T$;

2. Последовательность Ньютоновских приближений (11) может быть построена, принадлежит области (α) и сходится к решению x^*;

3. Скорость сходимости оценивается неравенством:

a) $\left|x_i^* - x_i^{(k)}\right| \le t^* - t^{(k)},\ i=\overline{1,n}$;

b) $\sum_{i=1}^{n}\left|x_i^* - x_i^{(k)}\right| \le t^* - t^{(k)}$;

c) $\left(\sum_{i=1}^{n}\left(x_i^* - x_i^{(k)}\right)^2\right)^{\frac{1}{2}} \le t^* - t^{(k)}$;

где $t^* = \frac{1-\sqrt{1-2h}}{h}B\eta$ - меньший корень уравнения $\frac{1}{2}Kt^2 - \frac{1}{B}t + \eta = 0$ и $t^{(k)}$ последовательность Ньютоновских приближений к корню t^*, построенная при начальном приближении $t^{(0)} = 0$.

Доказательство - сходимость метода Ньютона для операторных уравнений.

Решение задачи Коши для обыкновенных дифференциальных уравнений. Метод Эйлера.

Будем рассматривать задачу Коши для систем обыкновенных уравнений

$$\begin{cases} y' = f(x, y),\ x_0 < x \le X \\ y(x_0) = y_0 \end{cases} \quad \textbf{(1)}$$

Систему (1) запишем покоординатно:

$$\begin{cases} \frac{dy_i}{dx} = f_i(x, y_1, .., y_n), i = \overline{1, n} \\ y_i(x_0) = y_i^{(0)}, i = \overline{1, n} \end{cases} \quad \textbf{(2)}$$

Из теории дифференциальных уравнений известны условия, гарантирующие существование и единственность задачи Коши.

Предположим, что функции $f_i(x, y_1, .., y_n), i = \overline{1, n}$ непрерывны по всем аргументам в замкнутой области

$$D = \{x_0 \le x \le X, \left|y_i - y_i^{(0)}\right| \le b, i = \overline{1, n}\}$$

Из непрерывности функций f_i в замкнутой области следует их ограниченность, т.е. существует константа $M > 0$ такая, что всюду в области D выполняется неравенство $|f_i(x, y_1, .., y_n)| \le M, i = \overline{1, n}.$

Предположим, кроме того, что в области D функции f_i удовлетворяет условию Липшица по аргументам $y_1, .., y_n$,

$$|f_i(x, y_1', \dots, y_n') - f_i(x, y_1'', \dots, y_n'')| \le L(|y_1' - y_1''| + \cdots + |y_n' - y_n''|)$$

для $\forall$ точек $(x, y_1', \dots, y_n')$ и $(x, y_1'', \dots, y_n'')$.

Если выполнены сформулированные выше требования, то в области D существует единственное решение задачи Коши **(1) :**

$$y_1 = y_1(x), \dots, y_n = y_n(x).$$

В дальнейшем будем предполагать, что решение задачи **(1)** существует, единственно и обладает необходимыми свойствами гладкости.

Заменим непрерывную область $[x_0, X]$ разностной (или дискретной) областью.

Рассмотрим сетку

$$w_k = \{x_k, x_{k+1} = x_k + h, k = \overline{0, N-1}, x_N \le X \le x_{N+1}\}$$

w_k – сетка, x_k– узел сетки, h – шаг сетки.

В данном случае сетка равномерная.

Если $x_{k+1} - x_k = h_k$ и $h_k \neq h_i$ хотя бы для одной пары чисел k и i, то сетка называется неравномерной.

Непрерывную модель $y' = f(x, y)$ заменим дискретной моделью, которая строится следующим образом: решение задачи рассматривается в узлах сетки, а именно, вычисляются значения $y(x_0), y(x_1), \dots, y(x_N)$.

Внутри дискретной модели будем работать только со значениями сеточной задачи, а именно, будем рассматривать величины $z_0, z_1, \dots, z_N$ – решения дискретной модели.

Задача состоит в следующем: построить сеточную модель таким образом, чтобы выполнялось условие

$$z_k = y(x_k).$$

$$y_i(x_{k+1}) = y_i(x_k + h) = y_i(x_k) + h\, y_i'(x_k) + \frac{h^2}{2} y_i''(x_k + \theta\, h) =$$

$$= y_i(x_k) + h\, f_i\big(x_k, y(x_k)\big) + r_k^{(i)}(h) \qquad \textbf{(3)}$$

где $r_k^{(i)}(h) = \frac{h^2}{2} y_i''(x_k + \theta h), 0 \le \theta \le 1$,

при $h \to 0, y_i(x_k) \to z_k^i$.

Из **(3)** получим

$$\mathrm{z}_{k+1}^i = z_k^i + h\, f_i(x_k, z_k), i = \overline{1, n} \qquad \textbf{(4)}$$

или в векторном виде:

$$z_{k+1} = z_k + hf(x_k, z_k), k = \overline{0, N-1}, z_0 = y_0. \qquad \textbf{(5)}$$

(4) – метод Эйлера решения задачи Коши для обыкновенных дифференциальных уравнений.

Определим погрешность

$$\varepsilon_k = z_k - y(x_k), k = \overline{0, N}, \qquad \textbf{(6)}$$

нам нужно оценить $\|\varepsilon_k\|$ и установить, стремится ли $\|\varepsilon_k\| \underset{h \to 0}{\to} 0$.

Из **(4)** имеем

$$z_{k+1} = z_k + hf(x_k, z_k) + \delta_k \qquad \textbf{(7)}$$

здесь δ_k– погрешность округления.

Будем предполагать, что $\|\delta_k\| \le \delta, k = \overline{0, N-1}$.

Из **(3)** следует

$$y(x_{k+1}) = y(x_k) + hf(x_k, y(x_k)) + r_k(h), \qquad \textbf{(8)}$$

где $r_k(h)$ – погрешность аппроксимации метода Эйлера.

$\|r_k(h)\| \le Mh^2$**,** где M– константа, не зависящая от h.

Из **(7)** и **(8)** получим

$$\varepsilon_{k+1} = \varepsilon_k + h(f\big(x_k, y(x_k)\big) - f(x_k, z_k)) + (r_k(h) - \delta_k), \quad \textbf{(9)}$$
$$\varepsilon_0 = 0$$

Предположим, что функции $f(x,y)$ – удовлетворяют условию Липшица с постоянной B, а именно:

$$\|f(x, z_k) - f(x_k, y(x_k)\| \le B\|z_k - y(x_k)\| = B\|\varepsilon_k\|.$$

Из **(9)** получим

$$\|\varepsilon_{k+1}\| \le \|\varepsilon_k\| + hB\|\varepsilon_k\| + \mu \quad \textbf{(10)}$$

где $\mu = \delta + Mh^2$.

Из **(10)**, при $k = 1$ получим

$$\|\varepsilon_1\| \le \mu$$
$$\|\varepsilon_2\| \le (1 + hB)\mu + \mu \le (1 + \ell^{hB})\mu.$$

При $k = k$,

$$\big||\varepsilon_{k+1}|\big| \le (1 + e^{hB} + e^{2hB} + \cdots + e^{khB})\mu \quad \textbf{(11)}$$
$$\|\varepsilon_{k+1}\| \le (k+1)\ell^{khB}\mu \quad \textbf{(12)}$$

Тогда, т.к. правая часть (12) не зависит от k, получим, что

$$\varepsilon(h) = (X - x_0)e^{(X-x_0)B}(Mh + \frac{\delta}{h}) \quad \textbf{(13)}$$

Т.о. получили формулу для оценки погрешности метода Эйлера.

Следует отметить, что если отрезок $[\boldsymbol{x_0}, \boldsymbol{X}]$ – большой и константа $B >> 1$, то чтобы $\|\varepsilon_k\| \to 0$ нужно брать очень маленьким шаг h, но в этом случае из-за большого числа шагов, увеличивается погрешность округления. Но, выбирая маленьким h, получаем увеличенные погрешности вычисления $\varphi(h) = Mh + \frac{\delta}{h}$, если δ– фиксированная величина, то оптимальный шаг для минимизации этой величины нужно брать по формуле $h_{\text{опт}} = \sqrt{\frac{\delta}{M}}$. В этом случае $\varphi_{опт} = 2\sqrt{M\delta}$.

Для увеличения точности вычислений используются модификации МЭ. Одна из них – метод Эйлера-Мултона (МЭМ), где последовательные приближения

$$\begin{cases} z_{k+1} = z_k + \dfrac{h}{2}(f(x_k, z_k) + f(x_{k+1}, z_{k+1})) \\ z_0 = y_0 \end{cases}$$

МЭ погрешность: $O(h^2)$

МЭМ погрешность: $O(h^3)$

МЭМ – неявный метод. В правой части формулы стоит величина, которую надо найти. Для её нахождения на каждом шаге используется итерационный процесс.

$$z_{k+1}^{(i+1)} = z_k + \frac{h}{2}(f(x_k, z_k) + f(x_{k+1}, z_{k+1}^{(i)})),$$

$$z_0 = y_0.$$

Методы типа Рунге –Кутта.

Рассмотрим дифференциальное уравнение

$$y' = f(x, y), x_0 \le x \le X \quad \textbf{(1)}$$

и начальное условие $y(x_0) = y_0$.

Здесь f– непрерывно дифференцируемая функция

$$f: [x_0, X] \times R^n \to R^n$$
$$y: R^n \to R^n.$$

Предположим, что решение задачи **(1),(2)** существует и единственно, и обладает необходимыми условиями гладкости.

Общая характеристика методов Рунге – Кутта:

1. Методы не чувствительны к сетке. Сетка может быть равноотстоящей и не равноотстоящей.
2. Точность методов может быть любой.
3. Методы универсальные, гибкие, адаптируемые к любому классу задач.
4. Методы одношаговые.

Будем искать значение приближения решения задачи **(1),(2)** лишь в фиксированных точках $x_i, i = \overline{0, N}$ этого отрезка.

Выберем равноотстоящую сетку: $x_i = x_0 + ih, i = \overline{0, N}, N = [\frac{X - x_0}{h}]$.

Связь между двумя соседними значениями функции $y(x)$ дает следующее очевидное равенство:

$$y(x_{n+1}) = y(x_n) + \int_{x_n}^{x_n+1} y'(t)dt. \quad \textbf{(2)}$$

Поскольку при построении одинаковых методов используется информация о решении лишь на отрезке длинной в один шаг, то в формуле **(2)** можно опустить индекс, означающий номер шага и переписать формулу **(2)** в соответствующем виде:

$$y(x + h) = y(x) + \int_x^{x+h} y'(t)dt. \quad \textbf{(3)}$$

Обозначим $y(x + h) - y(x) = \Delta y$ и, учитывая **(1)**, используя замену $t = x + \alpha h$, последнее равенство можно переписать в виде:

$$\Delta y = \int_x^{x+h} y'(t)dt = h \int_0^1 f(x + \alpha h, y(x + \alpha h))d\alpha. \quad \textbf{(4)}$$

Если значение $y(x)$ известно, то чтобы найти $y(x+h)$, нам нужно найти поправку Δy. Для построения интересующей нас формулы введем три набора параметров:

$$\alpha_2, \alpha_3, \dots, \alpha_r \qquad (\boldsymbol{\alpha})$$

$$\beta_{21}$$

$$\beta_{31}, \beta_{32}$$

$$\dots\dots\dots\dots \qquad (\boldsymbol{\beta})$$

$$\beta_{r1}, \beta_{r2}, \dots, \beta_{rr-1}$$

$$p_1, p_2, \dots, p_r \qquad (\boldsymbol{p})$$

С помощью параметров (α) и (β) составим величины:

$$k_1 = hf(x, y)$$
$$k_2 = hf(x + \alpha_2 h, y + \beta_{21} k_1)$$
$$\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots$$
$$k_r = hf(x + \alpha_r h, y + \beta_{r1} k_1 + \dots + \beta_{rr-1} k_{r-1}).$$

Если параметры (α) и (β) выбраны, то значения $k_1, k_2, .., k_r$ вычисляются последовательно.

Заметим, что величины $k_i = hf(x + \alpha_i h, y + \beta_{i1} k_1 + \dots + \beta_{ii-1} k_{i-1})$ вообще говоря не равны величинам $hf(x + \alpha_i h, y(x + \alpha_i h)), i = \overline{1, r}$. Однако при удачном выборе (β) их можно трактовать как умноженные на h приближенные значения интегрируемой функции

$$f(x + \alpha h, y(x + \alpha h)).$$

Поэтому можно надеяться, что с помощью параметров (p) нам удастся создать такую комбинацию величин, которая будет являться квадратурной суммой и позволит найти приближенное значение интеграла **(4)**:

$$\Delta y \approx \sum_{i=1}^{r} p_i k_i. \qquad \textbf{(5)}$$

Отсюда становится понятен смысл введенных $(\alpha), (\beta), (p)$.

Величины $\varphi_r(h) = \Delta y - \sum_{i=1}^{r} p_i\, k_i$ представляют собой погрешность приближенного равенства **(5)**.

Если правая часть уравнения **(1)** является достаточно гладкой функцией, то $\varphi_r(h)$ будет иметь достаточно большое число производных.

Запишем разложение $\varphi_r(h)$ в ряд Маклорена:

$$\varphi_r(h) = \sum_{j=0}^{s} \frac{h^j}{j!} \varphi^{(j)}(0) + \frac{h^{s+1}}{(s+1)!} \varphi^{(s+1)}(\theta h). \qquad \textbf{(6)}$$

Если удастся выбрать $(\alpha),(\beta),(p)$ так, что $\varphi_r^{(j)}(0)=0, j=\overline{0,s}$, а $\varphi_r^{(s+1)}(\theta h)\neq 0$, то погрешность в формуле **(5)** будет величиной порядка h^{s+1}.

$$\varphi_r(h)=\frac{h^{s+1}}{(s+1)!}\varphi_r^{(s+1)}(\theta h), \qquad \textbf{(7)}$$

Тогда число s −порядок (или степень точности) данного метода типа Рунге-Кутты. Такое определение точности связано с тем, что если на каждом шаге погрешность расчетной формулы данного метода типа Рунге-Кутты имеет порядок h^{s+1}, а функция $\varphi(x,y)$ является гладкой в окрестности решения, то погрешность данного метода (часть погрешности приближенного решения, определяемая неточностью расчетной формулы) на конечном отрезке будет величиной порядка h.

Для построения по методу Рунге-Кутты при данном r одношаговых правил возможно более высокого порядка точности s выражают величины $f(x,y)$ по h в заданной степени.

Требуют, чтобы для любой гладкой $f(x,y)$ обращалось в нуль возможно большее число этих величин. Иными словами, $(\alpha),(\beta),(p)$ выбираются исходя из требования, чтобы разложение

$$\Delta y=y(x+h)-y(x)=\frac{h}{1!}y'(x)+\frac{h^2}{2!}y''(x)+\frac{h^3}{3!}y'''(x)+\cdots$$

и разложение линейной комбинации $\sum_{i=1}^{r}p_ik_i$ по степеням h совпадали для $\forall f(x,y)$ до членов с возможно более высокими степенями h.

Записать в общем виде систему уравнений для определения $(\alpha),(\beta),(p)$ крайне затруднительно. Поэтому рассмотрим лишь несколько примеров построения одношаговых правил по методу Рунге-Кутты.

Метод первого порядка точности.

(5) принимает в случае $r=1$ вид: $\Delta y\approx hf(x,y)$.

Т. о. построенный одношаговый метод Рунге-Кутта в точности совпадает с методом Эйлера.

Метод второго порядка точности.

$$\Delta y=\frac{1}{2}(k_1+k_2)+o(h^3); k_1=hf(x,y), k_2=hf(x+h,y+k_1).$$

Эта формула – аналог квадратурной формулы трапеции.

Положив

$$\Delta y = k_2 + o(h^3);$$
$$k_1 = hf(x, y),$$
$$k_2 = hf(x + \frac{1}{2}h, y + \frac{1}{2}k_1)$$

Аналог квадратурной формулы средних прямоугольников.

Методы третьего порядка точности.

$\Delta y = p_1 k_1 + p_2 k_2 + p_3 k_3$ и одна из формул имеет вид:

$$\Delta y = \frac{1}{6}(k_1 + 4k_2 + k_3) + o(h^4),$$
$$k_1 = hf(x, y),$$
$$k_2 = hf(x + \frac{1}{2}h, y + \frac{1}{2}k_1),$$
$$k_3 = hf(x + h, y - k_1 + 2k_2).$$

Методы четвертого порядка точности.

Наиболее употребительным методом решения задачи Коши является следующее правило:

$$\Delta y = \frac{1}{6}(k_1 + 2k_2 + 2k_3 + k_4) + o(h^5),$$
$$k_1 = hf(x, y),$$
$$k_2 = hf(x + \frac{h}{2}, y + \frac{1}{2}k_1),$$
$$k_3 = hf(x + \frac{h}{2}, y + \frac{1}{2}k_2),$$
$$k_4 = hf(x + h, y + k_3).$$

Это наиболее часто используемые методы для решения задачи Коши. В литературе они упоминаются без ссылок на порядок и точность.

Сходимость и оценка погрешности одношаговых методов.

Можно считать, что малость ошибки при одном шаге вычислений вообще говоря не гарантирует малость ошибки при счете на большом числе шагов. При переходе от шага к шагу, ошибки, допущенные на каждом шаге вычислений, накапливаются и могут быстро расти. Поведение накопившейся погрешности при этом зависит и от особенностей решаемой задачи, и от свойств избранного вычислительного метода, а также от погрешности задания начальных данных и неточностей выполнения вычислений. Оценим поведение этой погрешности в случае одношагового метода.

Рассмотрим задачу Коши:

$$y' = f(x, y), x_0 \le x \le X,$$
$$y(x_0) = y_0. \quad \textbf{(1)}$$

Для приближенного решения в точках $x_0, x_1, .., x_n$ решение $y(x)$ этой задачи при использовании одношагового метода можно определить по формуле:

$$y_{n+1} = y_n + h\Phi_f(h, x_n, y_n). \quad \textbf{(2)}$$

Будем предполагать в плоскости (x, y) существование выпуклой в направлении оси y области D, содержащей внутри себя $grad$ точного и приближенного решений исходной задачи и такой, что функция $f(x, y)$ имеет в D непрерывную производную по y такую, что $\frac{\partial f(x,y)}{\partial y} \le F$.

При решении реальной задачи вычисления по формуле **(2)** выполняются, как правило, не точно (из-за ошибок округления) и величины $\tilde{y}_1, \tilde{y}_2, .., \tilde{y}_n$ не будут удовлетворять разностному уравнению (2), а будут связаны соотношением

$$\tilde{y}_{n+1} = \tilde{y}_n + \Phi_f(h, x_n, \tilde{y}_n) - \alpha_n \quad \textbf{(3)}$$

Здесь $\tilde{y}_0 = y(x_0)$.

Величина α_n, определяемая формулой (3) – погрешность округления на n–ом шаге.

Обозначим через $\tilde{y}_n(x) -$ решение дифференциального уравнения, удовлетворяющего начальному условию $\tilde{y}_n(x_n) = \tilde{y}_n$, а через r_n - погрешность формулы **(2)** на $n -$ом шаге. Величина r_n - характеризует

локальную погрешность избранного метода или ошибку приближенного решения на одном шаге без учета погрешности округления и может быть введена посредством равенства:

$$\tilde{y}_n(x_{n+1}) = \tilde{y}_n(x_n) + h\Phi_f(n, x_n, \tilde{y}_n(x_n)) + r_n. \quad \textbf{(4)}$$

Так как $\tilde{y}_n(x_n) = \tilde{y}_n$, то из последнего равенства с учетом (3) и (4) получим:

$$\tilde{y}_n(x_{n+1}) = \tilde{y}_n + h\Phi_f(h, x_n, \tilde{y}_n(x_n)) + r_n = \tilde{y}_{n+1} + \alpha_n + r_n. \quad \textbf{(5)}$$

Поэтому, локальная ошибка приближенного решения с учетом погрешности округления может быть задана по формуле:

$$\tilde{y}_n(x_{n+1}) - \tilde{y}_{n+1} = \alpha_n + r_n. \quad \textbf{(6)}$$

Оценив сумму, стоящую в правой части равенства (6), мы можем сделать заключение о малости ошибки на одном шаге вычислений. Чтобы иметь возможность судить о величине погрешности при счете на большое число шагов, нужно получить удобное представление для разности $\varepsilon_n = y(x_n) - \tilde{y}_n$. Эта разность – погрешность приближенного решения. Она, очевидно, должна выражаться через погрешность формулы, погрешность округления и погрешность задания начального условия – $\varepsilon_0 = y(x_0) - \tilde{y}_0$.

Для величины

$$\varepsilon_n = y(x_n) - \tilde{y}_n = y(x_n) - \tilde{y}_0(x_n) + \sum_{i=1}^{n}(\tilde{y}_{i-1}(x_n) - \tilde{y}_i\,(x_n)), \quad \textbf{(7)}$$

оценив каждое слагаемое в равенстве (7), можно получить оценку:

$$|\varepsilon_n| \le (\varepsilon + \frac{r+\alpha}{h}(X - x_0))e^{L(X-x_0)} \quad \textbf{(8)}$$

где $L = \max\{F; 0\}$

На основании последней оценки можно утверждать, что, если выполняются условия:

1) $\varepsilon \underset{h\to 0}{\to} 0$,
2) $\frac{r}{h} \underset{h\to 0}{\to} 0$,
3) $\frac{\alpha}{n} \underset{n\to 0}{\to} 0$.

Таким образом, приближенное решение задачи **(1),** полученное одношаговым методом вида **(2)** равномерно на отрезке $[x_0, X]$ сходится к точному решению этой задачи.

При реальных расчётах величина α обычно фиксируется, поэтому отношение $\frac{\alpha}{h}$ возрастает с уменьшением h.

Правда, погрешности могут компенсироваться в случае разных знаков, но, если не согласовывать перед началом вычислений возможности машины и точность округлений, то результат может быть далёк от точного.