

Курс «Вычислительные алгоритмы теории автоматического управления».

Лекция 3. Вычислительные методы решения проблемы собственных значений. Численные методы анализа и моделирования детерминированных гладких нелинейных систем.

Несимметричная проблема собственных чисел. Вычисление характеристического многочлена матрицы. Квадратичные матричные уравнения и неравенства. Алгебраическое уравнение Риккати. Вычислительные задачи анализа и моделирования детерминированных гладких нелинейных систем. Понятие нормированного пространства. Методы численного решения систем нелинейных алгебраических уравнений. Численные методы вычисления матрицы Якоби. Численные методы интегрирования ОДУ.

Несимметричная проблема собственных чисел.

Для вычисления собственных значений и собственных векторов в пакете MatLab используется соответствующая команда *eig*. Используя вышеуказанные функции и команды для нахождения собственных значений и собственных векторов матрицы, тем самым применяют, так называемые, QR – алгоритмы.

Собственные вектора и значения. Характеристическое уравнение матрицы.

Число $\lambda \in \mathbb{C}$ и ненулевой вектор $x \in \mathbb{C}^n$, где \mathbb{C} поле комплексных чисел, называются собственным значением и собственным вектором матрицы $A \in \mathbb{R}^{n \times n}$, если $Ax = \lambda x$.

Говорят, что «собственное число λ отвечает собственному вектору x », а (λ, x) – называют собственной парой матрицы $A \in \mathbb{R}^{n \times n}$. Собственные значения матрицы $A \in \mathbb{R}^{n \times n}$ суть корни ее характеристического уравнения $\det(\lambda E - A) = 0$. Левая часть этого уравнения представляет собой многочлен от λ степени n – так называемый характеристический многочлен матрицы A . Поэтому всякая матрица порядка n имеет с учетом кратностей ровно n собственных значений. Кратность собственного значения λ , как корня характеристического многочлена, называется его алгебраической кратностью. Собственные векторы, отвечающие числу λ , могут быть найдены как ненулевые решения однородной линейной системы $\det(\lambda E - A) = 0$. Размерность подпространства решений этой системы называется геометрической кратностью собственного значения λ . Напомним, что геометрическая кратность никогда не превосходит алгебраическую.

Под полной проблемой собственных значений в численной линейной алгебре понимают задачу вычисления всех собственных значений или собственных векторов матрицы. Если же нужно одно или небольшая группа собственных значений (векторов), то говорят о частичной проблеме. В случае симметричной или эрмитовой матрицы употребляют термин симметричная проблема собственных значений, в противном случае – несимметричная проблема.

Решение задачи анализа асимптотической устойчивости линейной однородной системы $\dot{x} = Ax$, очевидно, сводится к полной несимметричной проблеме */Икрамов Нес проблема с61/*.

Вычисление характеристического многочлена матрицы.

Определитель характеристической матрицы $\Delta(\lambda) = \det(\lambda E - A)$ представляет собой скалярный многочлен относительно λ и называется характеристическим многочленом матрицы A . Матрицу $D(\lambda) = [d_{ik}(\lambda)]$, $i, k = 1; n$, где $d_{ik}(\lambda)$ – алгебраическое дополнение элемента $\lambda \delta_{ik} - a_{ki}$ в определителе $\Delta(\lambda)$, называют присоединенной матрицей для матрицы A .

Пример. Пусть $A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$ будем иметь $\lambda E - A = \begin{pmatrix} \lambda - a_{11} & -a_{12} \\ -a_{21} & \lambda - a_{22} \end{pmatrix}$,

$$\Delta(\lambda) = \lambda^2 - (a_{11} + a_{22})\lambda + a_{11}a_{22} - a_{12}a_{21}, \quad D(\lambda) = \begin{pmatrix} \lambda - a_{22} & a_{12} \\ a_{21} & \lambda - a_{11} \end{pmatrix}.$$

Из приведенных определений можно записать следующие тождества:

$$(\lambda E - A)D(\lambda) = \Delta(\lambda)E \text{ и } D(\lambda)(\lambda E - A) = \Delta(\lambda)E.$$

Теорема (Гамильтона-Кэли). Всякая квадратная матрица A удовлетворяет своему характеристическому уравнению, то есть $\Delta(A) = 0$.

Пример. Пусть $A = \begin{pmatrix} 2 & 1 \\ -1 & 3 \end{pmatrix}$. Тогда $\Delta(\lambda) = \lambda^2 - 5\lambda + 7$. Отсюда $\Delta(A) = A^2 - 5A + 7E = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$

Теорема /Гантмахер с94/. Если $\lambda_1, \lambda_2, \dots, \lambda_n$ все характеристические числа (с учетом кратностей) матрицы A , а $g(\mu)$ - некоторый скалярный многочлен, то $g(\lambda_1), g(\lambda_2), \dots, g(\lambda_n)$ - все характеристические числа матрицы $g(A)$.

Пусть характеристический многочлен имеет $\Delta(\lambda) = \lambda^n + a_{n-1}\lambda^{n-1} + \dots + a_0$. Тогда присоединенную матрицу можно записать в следующем виде:

$$D(\lambda) = D_{n-1}\lambda^{n-1} + D_{n-2}\lambda^{n-2} + \dots + D_0,$$

где

$$D_{n-1} = E,$$

$$D_{n-2} = A + a_{n-1}E,$$

$$D_{n-3} = A^2 + a_{n-1}A + a_{n-2}E,$$

....,

$$D_{n-k} = A^{k-1} + a_{n-1}A^{k-2} + \dots + a_{n-k+1}E, \quad k = 1, 2, \dots, n-2$$

При этом, если $k = n-1$, то можно записать следующее тождество $A \cdot D_0 + a_0E = 0$.

Последнее тождество следует из следующего утверждения.

Теорема (обобщенная теорема Безу). При правом (левом) делении матричного многочлена $F(\lambda)$ на бином $\lambda E - A$ остаток от деления равен $F(A)$ (соответственно $\hat{F}(A)$).

Из данной теоремы следует, что многочлен $F(\lambda)$ делится без остатка справа (слева) на бином $\lambda E - A$ тогда и только тогда, когда $F(A) = 0$ (соответственно $\hat{F}(A) = 0$).

Метод Леверье-Фадеева вычисления характеристического многочлена и присоединенной матрицы.

Пусть характеристический многочлен матрицы A имеет вид $\Delta(\lambda) = \lambda^n + a_{n-1}\lambda^{n-1} + \dots + a_0$, а присоединенная матрица задается выражением $D(\lambda) = D_{n-1}\lambda^{n-1} + D_{n-2}\lambda^{n-2} + \dots + D_0$ где $D_{n-1} = E$.

Используя свойства присоединенной матрицы можно записать соотношение

$$(\lambda E - A)^{-1} = \frac{D(\lambda)}{\Delta(\lambda)}.$$

Таким образом, если вычислить коэффициенты характеристического

многочлена a_{n-1}, \dots, a_0 и матрицы $D_{n-1}, D_{n-2}, \dots, D_0$, то можно определить и обратную матрицу $(\lambda E - A)^{-1}$. Решение данной задачи дается следующим алгоритмом Леверье-Фадеева:

1. $D_{n-1} = E, a_n = 1$
2. $F_{n-1} = A; a_{n-1} = -\text{tr}(F_{n-1}); D_{n-2} = F_{n-1} + a_{n-1}E$
3. $F_{n-2} = AD_{n-2}; a_{n-2} = -\text{tr}(F_{n-2}) / 2; D_{n-3} = F_{n-2} + a_{n-2}E$
-
- n-2. $F_1 = AD_1; a_1 = -\text{tr}(F_1) / (n-1); D_0 = F_1 + a_1E$

Соответственно, (n-1) – ый шаг алгоритма можно использовать для проверки вычислений. Действительно, вычисленные параметры $F_0 = AD_0; a_0 = -\text{tr}(F_0) / n$ должны удовлетворять

равенству $AD_0 + a_0E = 0$. Отсюда в частности легко получить, что $A^{-1} = \frac{D_0}{a_0}$. Здесь при

вычислениях используется обозначение $\text{tr}(A) = \sum_{i=1}^n a_{ii}$. Таким образом, метод Леверье-Фадеева

относится к прямым методам, позволяющим получить результат за конечное, заранее определенное, число шагов.

Эквивалентность задачи на собственные значения и задачи определения корней многочленного уравнения.

Определение собственных значений матрицы A эквивалентно определению корней многочленного уравнения $\det(\lambda E - A) = 0$. В принципе, верно и обратное: если имеется способ нахождения корней произвольного многочлена, то можно найти и собственные значения любой матрицы. Пусть некоторый многочлен имеет нормированную форму вида

$P(\lambda) = \lambda^n + a_{n-1}\lambda^{n-1} + \dots + a_0$. Построим для данного многочлена сопровождающую матрицу

$A_p \in \mathbb{R}^{n \times n}$ следующего вида $A_p = \begin{pmatrix} 0 & 1 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \dots & 1 \\ -a_0 & -a_1 & \dots & -a_{n-1} \end{pmatrix}$. Тогда имеет место следующее

утверждение.

Теорема. Пусть $P(\lambda) = \lambda^n + a_{n-1}\lambda^{n-1} + \dots + a_0$ - нормированный многочлен степени n и A_p его сопровождающая матрица. Тогда характеристический многочлен матрицы A_p есть многочлен $P(\lambda)$. Таким образом, корни уравнения $P(\lambda) = 0$ являются собственными значениями матрицы A_p .

Из приведенной теоремы следует, задача вычисления корней характеристического многочлена математически эквивалентна задаче вычисления собственных значений матрицы, как это было показано выше. Однако, вычислительная практика показала, что корни многих многочленов очень чувствительны к малым изменениям коэффициентов. Это относится и к многочленам с вещественными и хорошо разделенными корнями. Приведем хорошо известный пример. Многочлен $P(s) = (s-1)(s-2)\dots(s-20) = s^{20} - 210s^{19} + \dots$ имеет простые корни 1, 2, ..., 19, 20. Если к значению -210 коэффициента добавить 2^{-23} , что соответствует ошибке представления числа с плавающей точкой, то корни значительно изменятся. Наличие и распространенность подобных примеров вызвали в 50 годах прошлого столетия переоценку применявшихся тогда методов численного решения спектральных задач. Развиваемые с тех пор методы предусматривают вычисление собственных значений, минуя построение

характеристического многочлена. Это позволило добиться того, что соответствующие возмущения решения стали подчиняться закону: $\frac{\|\delta x\|}{\|x\|} \leq \frac{\|\delta A\|}{\|A\|}$.

Однако, для небольших размерностей $n < 10$ и хорошо обусловленных сопровождающих матриц, методы определения собственных значений с помощью нахождения корней характеристического многочлена дают достаточно хорошие оценки. Для решения задачи вычисления (отделения) корней характеристического многочлена используются разнообразные итерационные алгоритмы.

Итерационный метод вычисления собственных значений и векторов

Итерационные методы, в отличие от прямых методов, дают последовательность приближений, которые сходятся, в общем случае за бесконечное число шагов к истинному решению задачи. На практике бесконечное выполнение алгоритма невозможно, и, поэтому, его останавливают, когда получают хорошее, в смысле некоторого критерия, приближение, которое можно принять за решение. В частности, теорема Абеля показывает, что не существует прямых методов решения общей задачи на собственные значения. Так как существование конечной, заранее прописанной процедуры, приведет к существованию (возможно очень сложной) формулы для решения любого многочленного уравнения.

Примечание. Нильс Абель смог доказать, что не существует общей формулы (то есть содержащей сложение, вычитание, умножение, деление и извлечение корней) для выражения корней полинома степени $n > 4$.

Итерационные методы порождают последовательности векторов, которые сходятся к некоторому предельному значению – вектору. Запись $q_j \rightarrow v$ означает, что каждая компонента вектора q_j сходится к соответствующей компоненте вектора v , когда $j \rightarrow \infty$. Нетрудно проверить, что такая сходимость имеет место тогда и только тогда, когда $\|q_j - v\|_\infty \rightarrow 0$, или $\|q_j - v\|_1 \rightarrow 0$, или $\|q_j - v\|_2 \rightarrow 0$, когда $j \rightarrow \infty$. Кроме наличия сходимости, не менее важным фактором оценки итерационного метода является скорость сходимости. Сходящийся итерационный метод не имеет практического значения, если его временная сложность превышает некоторый приемлемый порог.

Теорема Шура. QR – метод.

При вычислении собственных значений и собственных векторов матрицы A не слишком высокого порядка и, уместающейся в виде квадратно $(n \times n)$ массива в оперативной памяти ЭВМ, обычно применяют так называемый итерационный QR – метод. Основой данного метода является теорема Шура об унитарной триангуляции.

Теорема Шура об унитарной триангуляции /Хорн с.101/. Пусть задана матрица $A \in \mathbb{C}^{n \times n}$ и зафиксирован какой-то порядок ее собственных значений $\lambda_1, \lambda_2, \dots, \lambda_n$. Тогда существует унитарная матрица $U \in \mathbb{C}^{n \times n}$, такая, что $U^* \cdot A \cdot U = T = \{t_{ij}\}$ – верхняя треугольная матрица с диагональными элементами $t_{ii} = \lambda_i; i = 1, 2, \dots, n$. Таким образом, любая квадратная матрица A унитарно эквивалентна треугольной матрице, в которой диагональные элементы представляют собой собственные значения матрицы A , записанные в заранее заданном порядке. Кроме того, если $A \in \mathbb{R}^{n \times n}$ и все ее собственные значения вещественны, то U можно выбрать вещественной и ортогональной.

Как частный случай теоремы Шура, существует также ее чисто вещественный вариант.

Теорема (вещественный вариант теоремы Шура). Для любой матрицы $A \in \mathbb{R}^{n \times n}$ существует

вещественная ортогональная матрица $Q \in \mathbb{R}^{n \times n}$, такая, что $Q^T \cdot A \cdot Q = \begin{pmatrix} A_1 & \cdot & \dots & X \\ \cdot & A_2 & \dots & \cdot \\ \cdot & \cdot & \dots & \cdot \\ 0 & \cdot & \dots & A_k \end{pmatrix}$,

где для каждого i матрица A_i имеет размер 1×1 или 2×2 , и отвечает соответственно вещественному значению или не вещественной паре комплексно-сопряженных собственных значений матрицы A . Блоки A_i можно расположить в любом заданном порядке.

Таким образом, если матрица A имеет не вещественные собственные значения, то привести ее к верхней треугольной форме с помощью ортогонального преобразования нельзя, однако в любом случае можно получить за конечное число шагов преобразований верхнюю хессенбергову форму вида:

$$\begin{pmatrix} X & X & X & X & X \\ X & X & X & X & X \\ & X & X & X & X \\ & & X & X & X \\ & & & X & X \end{pmatrix}$$

В случае вещественных собственных значений матрицу A можно представить в следующем виде $A = Q \cdot R$, где Q - ортогональная, а R - верхняя треугольная матрица. QR - разложение, можно применять точно также, как LU - разложение, для решения линейной системы $Ax = b$. Однако в основном QR разложение используют для оценки собственных значений матрицы A .

В первом приближении под QR – алгоритмом понимают итерационный процесс, протекающий при $k=1,2,\dots$ по формулам: $A_k \rightarrow Q_k \cdot R_k$, $R_k \cdot Q_k \rightarrow A_{k+1}$ /Икрамов Нес проблема с111/. Первая формула означает, что очередная матрица A_k подвергается ортогонально-треугольному (в комплексном случае унитарно-треугольному) разложению или коротко QR – разложению.

Вторая формула показывает, что сомножители QR – разложения перемножаются в обратном порядке, порождая следующую матрицу A_{k+1} . В качестве начальной матрицы A_1 берут матрицу A .

В надлежащий момент можно прекратить QR – итерации, приняв диагональные элементы последней матрицы A_e за приближения к собственным значениям. При необходимости из A_e и промежуточных матриц Q_1, Q_2, \dots можно определить собственные векторы матрицы A .

Следует отметить, что основной QR – алгоритм сходится очень медленно. Кроме того, каждый шаг алгоритма очень дорог (каждая QR итерация требует $O(n^3)$ операций): нужно выполнить ортогонально-треугольное разложение, а затем перемножить две матрицы порядка n . Поэтому в существующих реализациях алгоритма используют такие приемы, как предварительно приведенное матрицы A к удобной для алгоритма форме (хессенберговой форме), введение промежуточных итераций (так называемых сдвигов) между основными шагами.

Практический QR - алгоритм содержит следующие основные этапы.

1. Масштабирование матрицы: $A \rightarrow B = D^{-1} \cdot A \cdot D$
2. Приведение матрицы к форме Хессенберга: $B \rightarrow H = P^{-1} \cdot B \cdot P$.
3. Итерационный QR процесс: $H_1 \rightarrow H$, $H_k \rightarrow Q_k \cdot R_k$, $R_k \cdot Q_k \rightarrow H_{k+1}$, $k=1,2,\dots$ (процесс QR -разложения на каждой итерации осуществляется путем совокупности ряда элементарных операции (вращений и сдвигов) над соответствующей матрицей H_k).

В пакетах MatLab и MatCad имеются соответствующие функции, реализующие QR -разложение.

Квадратичные матричные уравнения. Алгебраическое уравнение Риккати.

В общем случае задача синтеза оптимального линейного регулятора линейной стационарной системы $\dot{x} = Ax + Bu$, $x \in \mathbb{R}^n$, $u \in \mathbb{R}^m$ с критерием качества в виде

$$J = \int_0^{\infty} \{u^T(\tau)Gu(\tau) + x^T(\tau)Qx(\tau)\}d\tau$$

сводится, как известно, к нахождению положительно

определенного решения V уравнения Риккати: $A^T \cdot V + V \cdot A - V \cdot B \cdot G^{-1} \cdot B^T \cdot V = -Q$. Для справедливости такого утверждения достаточно, чтобы пара $\{A, B\}$ была стабилизируемой.

Квадратичные матричные уравнения.

Уравнение Риккати есть частный случай, более общего вида, квадратичного матричного уравнения вида $A \cdot X + X \cdot B - X \cdot F \cdot X + C = 0$ /Икрамов МатрУр с134/. Здесь принимается, что матрица A – квадратная матрица порядка n , B – квадратная матрица порядка m , F – матрица $(m \times n)$, а C – матрица $(n \times m)$.

Рассмотрим общий случай квадратичного матричного уравнения, когда его коэффициенты и решение являются комплексными матрицами. Это упрощает изложение его метода решения. В дальнейшем приведем дополнительные условия, которые позволяют перейти к случаю нахождения вещественных решений.

Построим следующую блочную матрицу: $M = \begin{pmatrix} -B & -F \\ C & A \end{pmatrix}$. Порядок матрицы M равен $(n + m)$, а ее диагональные блоки $-B$ и A являются квадратными матрицами.

Рассмотрим матрицу M , как линейный оператор, действующий в пространстве \mathbb{C}^{n+m} . У такого оператора имеются инвариантные подпространства всех размерностей, то есть совокупности L^k , $k=1,2,\dots,m+n$ векторов $x \in \mathbb{C}^k \rightarrow Mx \in \mathbb{C}^k$. Пусть U некоторая базисная матрица подпространства L^m . Тогда существует матрица M_L размерности $(m \times m)$ такая, что будет выполняться соотношение: $MU = UM_L$.

Представим матрицу U в блочном виде: $U = \begin{pmatrix} U_1 \\ U_2 \end{pmatrix}$, $U_1 \in \mathbb{C}^{m \times m}$. Предположим, что матрица U_1 невырождена. Тогда, можно показать, $X_0 = U_2 \cdot (U_1)^{-1}$ является решением исходного квадратичного матричного уравнения. При этом, если используется другая базисная матрица $\bar{U} = \begin{pmatrix} \bar{U}_1 \\ \bar{U}_2 \end{pmatrix}$, то будет справедливо соотношение $X_0 = U_2 \cdot (U_1)^{-1} = \bar{U}_2 \cdot (\bar{U}_1)^{-1}$.

Способ построения базиса подпространства L^m и соответственно матрицы U базируется на теореме Шура, а практическим средством его реализации является QR-алгоритм.

Алгоритм решения квадратичных матричных уравнений.

Рассмотрим случай, когда достаточно построить какое-либо решение исходного матричного уравнения.

1. Составляется матрица M по коэффициентам исходного уравнения.

2. Осуществляется приведение матрицы M к форме Шура с помощью QR-алгоритма.

Пусть U унитарная матрица, трансформирующая матрицу M к верхней форме Шура S , то есть:

$$U^* \cdot M \cdot U = S. \text{ Осуществляется представление матрицы } U \text{ в блочном виде: } U = \begin{pmatrix} U_{11} & U_{12} \\ U_{21} & U_{22} \end{pmatrix},$$

где $U_{11} \in \mathbb{C}^{m \times m}$.

3. Вычисляется решение X_0 по следующей формуле: $X_0 = U_{21} \cdot (U_{11})^{-1}$.

Если матрица M вещественная, то весь процесс вычисления вещественного решения X_0 можно провести в вещественной арифметике. В этом случае QR-алгоритм используется, как метод построения вещественной формы Шура, и матрица U будет вещественной и ортогональной.

Решение уравнения Риккати.

Для уравнения Риккати матрица M определяется следующим образом:

$$M = \begin{pmatrix} A & -BG^{-1}B^T \\ -Q & -A^T \end{pmatrix}, \text{ где матрицы } A, B \text{ определяют уравнение системы, а матрицы } Q, G$$

соответствующий функционал качества. Так как порядок матрицы A равен n , то матрица M имеет порядок $2n$.

Матрица M является гамильтоновой, то есть удовлетворяет следующему уравнению

$$J^T M^T J = -M, \text{ где } J = \begin{pmatrix} 0 & E_n \\ -E_n & 0 \end{pmatrix}. \text{ Спектр гамильтоновой матрицы имеет следующее свойство:}$$

наряду с каждым собственным значением λ он содержит, причем с той же кратностью, и значение $-\lambda$. При этом, условия, наложенные на матрицы A, B и Q, G , обеспечивают, что матрица M не имеет чисто мнимых собственных значений. Следовательно, ровно половина спектра матрицы M принадлежит полуплоскости $\operatorname{Re} \lambda < 0$. Это позволяет упростить задачу вычисления положительно определенного решения уравнения Риккати.

Алгоритм нахождения такого решения состоит из следующих этапов.

1. Составление матрицы M по исходным матрицам задачи.

2. Приведение матрицы M к форме Шура S .

3. Переупорядочение формы Шура.

4. Вычисление решения V .

Смысл этапа 3 заключается в переходе от первоначально вычисленной формы S к новой форме \bar{S} , содержащей в первых n позициях главной диагонали устойчивые собственные значения.

Весь процесс вычисления положительно определенного решения 3 можно провести в вещественной арифметике. На этапе 2 матрица M , посредством QR-алгоритма, приводится к некоторой вещественной форме Шура. Затем на этапе 3 эта форма преобразуется в другую, также вещественную, форму Шура \bar{S} . Верхняя угловая подматрица размерности $(n \times n)$ матрицы \bar{S} соответствует устойчивым собственным значениям. Таким образом, матрица U , используемая при построении решения V , является вещественной.

Следует отметить, что при построении матрицы M требуется дополнительная вычислительная работа, связанная с обращением матрицы G и вычислением произведения $-B \cdot G^{-1} \cdot B^T$. Если матрица G плохо обусловлена, то может потребоваться соответствующая модификация алгоритма, связанная с необходимостью вычисления обратной матрицы.

В ряде случаев, требуется определить не саму матрицу V , а некоторую совокупность параметров (настроек регулятора $r \in R^q$), от которых зависят элементы решения – матрицы V . В этом случае, в ряде случаев, может быть целесообразным искать сразу вектор r с использованием соответствующих прямых методов решения систем нелинейных уравнений.

Вычислительные задачи анализа и моделирования детерминированных гладких нелинейных систем.

Условно все задачи анализа нелинейных систем можно разделить на два основных раздела: исследование стационарных решений (стационарный анализ) и анализ динамического поведения.

В стационарном анализе можно выделить следующие основные задачи: получение стационарных решений и исследования их зависимостей от параметра, получение линеаризованных моделей первого приближения в окрестности стационарных решений, оценка устойчивости стационарных решений, анализ точек ветвления и определение бифуркаций.

В анализе динамического поведения нелинейных систем следует выделить, прежде всего, задачу получения численных решений (переходных процессов), исследование геометрической структуры решений (фазовых портретов), оценка инвариантных и притягивающих множеств, в том числе периодических решений и странных аттракторов, исследование устойчивости решений и определение областей управляемости.

Линейная аппроксимация нелинейных математических моделей.

Рассмотрим нелинейное уравнение вида: $\dot{x}(t) = F(x(t), u(t))$, где функция F определяет отображение $F: \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}^n$. Точка $\bar{x} \in \mathbb{R}^n$ называется изолированной или гиперболической точкой равновесия, если существует такой постоянный входной сигнал $\bar{u} \in \mathbb{R}^m$, что $F(\bar{x}, \bar{u}) = 0$, и в ближайшей окрестности точки (\bar{x}, \bar{u}) нет других точек равновесия, и, кроме того, $\det\left(\frac{\partial F}{\partial x}\bigg|_{x=\bar{x}, u=\bar{u}}\right) \neq 0$.

Если, начальное состояние системы равно $x(t_0) = \bar{x}$ при входном воздействии \bar{u} , то, очевидно, что $x(t) \equiv \bar{x}$ для $\forall t \geq t_0$.

Определим вариации переменных относительно точки равновесия: $\delta x(t) = x(t) - \bar{x}$, $\delta u(t) = u(t) - \bar{u}$, где $x(t)$ решение уравнения $\dot{x}(t) = F(x(t), u(t))$. Тогда можно записать линейную стационарную систему первого приближения $\delta \dot{x}(t) = A \delta x(t) + B \delta u$, где

$A = \frac{\partial F}{\partial x}\bigg|_{x=\bar{x}, u=\bar{u}} \in \mathbb{R}^{n \times n}$ и $B = \frac{\partial F}{\partial u}\bigg|_{x=\bar{x}, u=\bar{u}} \in \mathbb{R}^{n \times m}$ являются постоянными, вещественными матрицами (матрицами Якоби).

Однако, довольно часто, при линеаризации, возникают более сложные случаи. Предположим, что исследуемая система (объект) следует вдоль некой программной траектории $\hat{x}(t), \hat{u}(t)$, где $\dot{\hat{x}}(t) = F(\hat{x}(t), \hat{u}(t))$. Определим, как уже указывалось выше, $\delta x(t) = x(t) - \hat{x}(t)$, $\delta u(t) = u(t) - \hat{u}(t)$. Тогда можно записать уравнения:

$$\dot{\hat{x}}(t) + \delta \dot{x}(t) \approx F(\hat{x}(t), \hat{u}(t)) + \frac{\partial F}{\partial x}\bigg|_{x=\hat{x}(t), u=\hat{u}(t)} \cdot \delta x(t) + \frac{\partial F}{\partial u}\bigg|_{x=\hat{x}(t), u=\hat{u}(t)} \cdot \delta u.$$

Так как $\dot{\hat{x}}(t) = F(\hat{x}(t), \hat{u}(t))$, то получим:

$$\delta \dot{x}(t) \approx \frac{\partial F}{\partial x}\bigg|_{x=\hat{x}, u=\hat{u}} \cdot \delta x(t) + \frac{\partial F}{\partial u}\bigg|_{x=\hat{x}, u=\hat{u}} \cdot \delta u.$$

В этом случае матрицы Якоби $A(t) = \frac{\partial F}{\partial x} \Big|_{\substack{x=\hat{x}(t) \\ u=\hat{u}(t)}} \in \mathbb{R}^{n \times n}$ и $B(t) = \frac{\partial F}{\partial u} \Big|_{\substack{x=\hat{x}(t) \\ u=\hat{u}(t)}} \in \mathbb{R}^{n \times m}$ зависят от параметра t (или от времени). То есть линеаризованная система первого приближения является нестационарной моделью.

В общем случае, точек равновесия системы может быть некоторое множество, которое будет определяться решением уравнения $F(\bar{x}, \bar{u}) = 0$ в зависимости от значений параметра \bar{u} .

Очевидно, что каждому такому регулярному решению (то есть, когда матрица Якоби

$A = \frac{\partial F}{\partial x} \Big|_{\substack{x=\bar{x} \\ u=\bar{u}}} \in \mathbb{R}^{n \times n}$ является невырожденной) будет соответствовать в его окрестности своя

линеаризованная система первого приближения.

Обозначим множество всех решений такой системы как:

$$S(F) = \{(x, u) \in \mathbb{R}^n \times \mathbb{R}^m, F(x, u) = 0\}.$$

Если $m = 1$, то множество $S(F)$ обычно представляет собой объединение нескольких кривых в пространстве \mathbb{R}^{n+1} , хотя может включать в себя и отдельные изолированные точки. При $n = 1$, множество $S(F)$ удобно изображать на двумерной плоскости. Такое представление обычно называют диаграммой стационарных решений.

Следует отметить, что в случае наличия периодического движения в системе или странного аттрактора, указанный подход не позволит оценить инвариантное множество траекторий и его область притяжения. Решение подобной задачи требует других, более сложных методов, опирающихся на более детальное изучение структуры траекторий движения системы.

Понятие нормированного пространства.

В некоторых линейных пространствах удается ввести метрику, задавая норму элемента. Понятие нормы эквивалентно понятию длины вектора в конечномерном пространстве. В пространствах с нормой метрика $\rho(x, y)$ будет инвариантна относительно сдвига, то есть:

$$\rho(x, y) = \rho(x + z, y + z) \forall x, y, z \in X,$$

а само пространство X будет линейным метрическим пространством.

Определение. Линейное пространство X называется нормированным пространством, если любому элементу $x \in X$ поставлено в соответствие норма $\|x\|$ такая, что для $\|x\|$ выполнены следующие аксиомы:

- 1) $\|x\| > 0$; $\|x\| = 0 \Leftrightarrow x = 0$;
- 2) $\|\lambda x\| = |\lambda| \cdot \|x\|$;
- 3) $\|x + y\| \leq \|x\| + \|y\|$.

Замечание. В нормированном пространстве можно ввести понятие расстояния между элементами следующим образом: $\rho(x, y) = \|x - y\|$ и, значит, метрическое пространство можно считать обобщением нормированного.

Примеры нормированных пространств:

- 1) Пространство \mathbb{R}^n . В нем можно ввести норму следующим образом:

$$a) \|x\|_1 = \sum_{i=1}^n |x_i|;$$

$$б) \|x\|_2 = \left(\sum_{i=1}^n x_i^2 \right)^{1/2};$$

$$в) \|x\|_\infty = \max_i |x_i|.$$

2) Пространство ограниченных числовых последовательностей m или l_∞ . Это множество бесконечных последовательностей $x = (x_1, x_2, \dots, x_n, \dots)$, для которых $\sup_i |x_i| < \infty$. Норму в этом пространстве можно ввести как: $\|x\| = \sup_i |x_i|$.

3) Пространство l_p - это множество всех последовательностей $x = (x_1, x_2, \dots, x_n, \dots)$, таких, что ряд $\sum_{i=1}^{\infty} |x_i|^p$ - сходится. Норму в пространстве l_p вводится следующим образом:

$$\|x\|_p = \left(\sum_{i=1}^{\infty} |x_i|^p \right)^{1/p}.$$

4) Пространство непрерывных функций $C[a, b]$. Норму в этом пространстве вводится следующим образом: $\|f\| = \max_{t \in [a, b]} |f(t)|$.

5) Пространство k раз непрерывно дифференцируемых функций $C^k[a, b]$. Норму в этом пространстве вводится следующим образом: $\|f\| = \sum_{i=0}^k \max_{t \in [a, b]} |f^{(i)}(t)|$.

6) Пространство $L^2[a, b]$ - пространство непрерывных функций с нормой:

$$\|f\| = \left(\int_a^b f^2(t) dt \right)^{1/2}.$$

Определение. Пусть X - линейное пространство и в X введены две нормы: $\|x\|_1, \|x\|_2$. Нормы $\|x\|_1, \|x\|_2$ называются эквивалентными, если $\exists \alpha, \beta > 0: \forall x \in X \alpha \|x\|_1 \leq \|x\|_2 \leq \beta \|x\|_1$.

Определение. Две нормы $\|x\|_1, \|x\|_2 \in X$. Норма $\|x\|_1$ называется подчиненной норме $\|x\|_2$, если $\exists \beta > 0$, то $\forall x \in X \|x\|_1 \leq \beta \|x\|_2$.

Замечание. Если каждая из норм подчинена другой, то нормы являются эквивалентными.

Определение. Вещественное пространство называется евклидовым, если $\forall x, y \in X$ можно поставить в соответствие вещественное число (x, y) , которое называется скалярным произведением, для которого выполняются аксиомы:

$$1) (x, x) \geq 0; (x, x) = 0 \Leftrightarrow x = 0;$$

$$2) (x, y) = (y, x);$$

$$3) (\mu x, y) = \mu(x, y);$$

$$4) (x + y, z) = (x, z) + (y, z).$$

Замечание: Евклидово пространство можно превратить в нормированное, если определить норму:

$$\|x\| = \sqrt{(x, x)}.$$

Свойства скалярного произведения:

1) Непрерывность скалярного произведения

Пусть $x_n \rightarrow x, y_n \rightarrow y, n \rightarrow \infty$. Тогда $(x_n, y_n) \rightarrow (x, y)$.

Доказательство:

$$\text{Рассмотрим } (x_n, y_n) - (x, y) = (x_n - x, y_n - y) + (x, y - y_n)$$

По неравенству Коши – Буняковского:

$$|(x_n, y_n) - (x, y)| \leq |(x_n - x, y_n - y) + (x, y - y_n)| \leq \|x_n - x\| \cdot \|y_n - y\| + \|x\| \cdot \|y - y_n\| \rightarrow 0 \text{ при } n \rightarrow \infty. \text{ Значит, скалярное произведение будет непрерывной функцией.}$$

2) Равенство параллелограмма:

Во всяком пространстве со скалярным произведением справедливо равенство:

$$\|x + y\|^2 + \|x - y\|^2 = 2(\|x\|^2 + \|y\|^2).$$

Доказательство:

$$\|x + y\|^2 + \|x - y\|^2 = (x + y, x + y) + (x - y, x - y) = 2(\|x\|^2 + \|y\|^2).$$

Банаховы пространства.

Определение. Полное нормированное пространство называется банаховым.

Примеры банаховых пространств: 1) Пространство \mathfrak{R}^m с нормой $\|x\|_2 = \left(\sum_{i=1}^n x_i^2 \right)^{1/2}$;

2) Пространство $C[a, b]$ с нормой $\|f\| = \max_{t \in [a, b]} |f(t)|$.

Гильбертовы пространства.

Определение. Гильбертово пространство – это пространство со скалярным произведением, полное в норме, порожденной данным скалярным произведением.

Примеры гильбертовых пространств:

1) Евклидово пространство \mathfrak{R}^m полно в евклидовой норме $\|x\|_2 = \left(\sum_{i=1}^m x_i^2 \right)^{1/2}$, порожденной

скалярным произведением $(x, y) = \sum_{i=1}^m x_i y_i$.

2) Пространство l_2 бесконечных последовательностей $x = (x_1, x_2, \dots, x_n, \dots)$, таких, что ряд $\sum_{i=1}^{\infty} |x_i|^2$ сходится, с нормой $\|x\|_2 = \left(\sum_{i=1}^{\infty} |x_i|^2 \right)^{1/2}$, порожденной скалярным произведением $(x, y) = \sum_{i=1}^{\infty} x_i y_i$.

Линейные операторы. Непрерывность и ограниченность.

Пусть X, Y – некоторые линейные пространства и множество $D \subseteq X$.

Определение. Если каждому элементу $x \in D$ можно поставить в соответствие определенный элемент $y \in Y$, то говорят, что задан оператор $y = F(x)$. При этом, множество D – область определения оператора F .

Определение. Множество $R(F) = \{y : y = F(x)\}$ называется областью значений оператора F , элемент y называется образом элемента x , элемент x называется прообразом.

Определение. Элемент $y_0 \in Y$ называется пределом оператора $F(x)$ при $x \rightarrow x_0$, если $\forall \varepsilon > 0 \exists \delta(\varepsilon) > 0 \|x - x_0\|_X < \delta \rightarrow \|F(x) - y_0\|_Y < \varepsilon$.

Определение. Оператор $F : X \rightarrow Y$, определенный в окрестности точки x_0 , называется непрерывным в точке x_0 , если $F(x) \rightarrow F(x_0)$, $x \rightarrow x_0$.

Определение. Пусть $F(x)$ – оператор с $D(F) \subset X$, $R(F) \subset Y$, X, Y – нормированные пространства. Оператор F называется ограниченным, если он переводит всякое ограниченное множество из множества $D(F)$ во множество, ограниченное в пространстве Y .

Определение. Линейный оператор, определенный на множестве $D(A) \subset X = B$, $R(B) \subset Y$ называется ограниченным, если он ограничен на единичном шаре $\bar{B}_1(0)$, то есть, если ограничена величина $\{\|Ax\|_Y, \|x\| < 1\}$

С другой стороны, если оператор A ограничен, то $\exists c > 0 : \forall x \in \bar{B}_1(0), \|x\| < 1$ для которой выполняется: $\|Ax\|_Y \leq c$.

Теорема: Оператор A ограничен тогда и только тогда, когда выполняется оценка $\|Ax\|_Y \leq c \|x\|_X, x \in X, c = \text{const}$.

Доказательство.

При $x = 0$ справедливость утверждения очевидна. Пусть $x \neq 0$. Рассмотрим элемент

$\tilde{x} = \frac{x}{\|x\|}, \|\tilde{x}\|_X = 1$. Поэтому получим, что $\|A\tilde{x}\|_Y \leq c$, значит, $\left\| A \left(\frac{x}{\|x\|} \right) \right\|_Y \leq c$. Так как норма

аддитивна, а оператор A линеен, то получаем: $\|Ax\|_Y \leq c \|x\|_X$. С другой стороны, если верно $\|Ax\|_Y \leq c \|x\|_X$, то тогда в замкнутом шаре $\bar{B}_1(0): \|x\|_X \leq 1$. Отсюда следует, что оператор A ограничен.

Следствие: Если оператор A - ограничен, то он ограничен на любом шаре.

Теорема (Об эквивалентности понятий ограниченности и непрерывности линейного оператора).

Пусть $A: X \rightarrow Y$, X, Y - банаховы пространства. Пусть область определения оператора $A: D(A) = X$. Для того, чтобы A был непрерывен, необходимо и достаточно, чтобы он был ограничен.

Доказательство.

Пусть A непрерывен. Предположим, что A - неограничен. Значит, значение $A(\bar{B}_1(0))$ будет неограниченно: $\forall n, \exists x_n \in X$ с нормой $\|x_n\| \leq 1$, такой что $\|Ax_n\| \geq n$. Введем элемент $\tilde{x}_n = \frac{x_n}{n}$, $\|\tilde{x}_n\| \leq \frac{1}{n} \rightarrow 0, n \rightarrow \infty$. При этом $\|A\tilde{x}_n\| = \frac{1}{n} \|Ax_n\| \geq 1$. А из непрерывности следует, что $\|A\tilde{x}_n\| \leq \varepsilon \rightarrow 0, n \rightarrow \infty$. Противоречие доказывает теорему.

Примеры линейных операторов:

1) В пространстве \mathbb{R}^n равенство $y = A \cdot x$, где A - матрица $(n \times n)$ задает некоторый линейный оператор.

2) Интегральное выражение $y(x) = \int_a^b K(x,s)x(s)ds$, где $K(x,s) \in C([a,b] \times [a,b])$ может определять различные интегральные операторы $A: X \rightarrow Y$, где, например, пространство $X = C[a,b]$, $Y = C[a,b]$ или пространство $X = L_2(a,b)$, $Y = L_2(a,b)$.

Пространство линейных операторов

В линейном пространстве операторов определено понятие нормы: $\|A\| = \sup_{\|x\|=1} \|Ax\|$. По

свойствам ограниченного оператора и из определения нормы получается: $\|Ax\| \leq \|A\| \cdot \|x\|$.

Для введенного определения нормы выполняются все аксиомы норм. Поэтому $L(X,Y)$ - линейное нормированное пространство.

Определение. Пусть последовательность операторов $\{A_n\} \in L(X,Y)$ последовательность операторов сходится в пространстве операторов $A_n \rightarrow A, n \rightarrow \infty$ равномерно, если $\|A_n - A\| \rightarrow 0, n \rightarrow \infty$.

Теорема: Чтобы последовательность операторов сходилась равномерно $A_n \rightarrow A$ необходимо и достаточно, чтобы $A_n(x) \rightarrow A(x), n \rightarrow \infty$ равномерно по x в шаре $\bar{B}_1(0), \|x\| \leq 1$.

Доказательство.

Можно записать следующее неравенство:

$$\|A_n(x) - Ax\| = \|(A_n - A)x\| \leq \|A_n - A\| \|x\| \leq \|A_n - A\|.$$

Если $\|A_n - A\| \rightarrow 0$, то тогда получим: $\forall \varepsilon > 0 \exists N : \forall n > N \|A_n - A\| < \varepsilon$. Отсюда найдем, что $\|A_n x - Ax\| < \varepsilon \quad \forall x \in \overline{B}_1(0)$. Это означает равномерную сходимость $A_n x \rightarrow Ax$.

С другой стороны, пусть $\forall \varepsilon > 0 \exists N : \forall n > N \|A_n x - Ax\| < \frac{\varepsilon}{2}$ для $\forall x \in \overline{B}_1(0)$. Тогда

можно записать, что $\sup_{\|x\| \leq 1} \|A_n x - Ax\| \leq \frac{\varepsilon}{2} < \varepsilon$. Значит, $\|A_n - A\| < \varepsilon$ для $\forall x \in \overline{B}_1(0)$, то есть

$A_n x \rightarrow Ax$ равномерно.

Теорема (Принцип равномерной ограниченности): Если последовательность $\{A_n(x)\}$ ограничена при каждом фиксированном x , то последовательность $\{\|A_n\|\}$ тоже ограничена.

Методы численного решения систем нелинейных уравнений.

Как уже указывалось, задача определения изолированных точек равновесия нелинейной автономной системы вида $\dot{x} = F(x)$, где $x \in \mathbb{R}^n; F: \mathbb{R}^n \rightarrow \mathbb{R}^n$, сводится к решению системы уравнений $F(x) = 0$.

Решение скалярных уравнений.

Рассмотрим задачу нахождения корней уравнения $f(x) = 0; f, x \in \mathbb{R}$. Для корректного решения поставленной задачи примем, что на отрезке $[a, b]$ имеется только один корень. Таким образом, если необходимо найти все корни уравнения, то решение задачи будет проводиться последовательно, находя корни поочередно. При этом после нахождения простого корня x^* необходимо исходную функцию разделить на множитель $(x - x^*)$ и, далее искать решение модифицированной задачи.

Метод деления отрезка пополам.

Пусть: $a_0 = a; b_0 = b; c_1 = \frac{a_0 + b_0}{2}$. Если $f(c_1) = 0$, то решение задачи найдено. Если $f(c_1) > 0$, то полагаем $a_1 = a_0, b_1 = c_1$. Если же $f(c_1) < 0$, то полагаем $a_1 = c_1, b_1 = b_0$. И повторяем действия алгоритма далее. В результате на k -ом шаге получим отрезок, содержащий решение, длиной:

$$b_k - a_k = \frac{b - a}{2^k}; k = 0, 1, \dots$$

Если в качестве приближения к решению x^* взять $x_k = \frac{a_k + b_k}{2}$, то получим:

$$|x_k - x^*| \leq \frac{1}{2} \cdot (b - a) \cdot 2^{-k}.$$

Для нахождения решения с точностью ε требуется выполнить следующее число итераций:

$$k = \left\lceil \frac{\ln(\frac{b-a}{2\varepsilon})}{\ln(2)} \right\rceil + 1,$$

где $[x]$ - целая часть числа x .

Методы типа простой итерации.

Уравнение: $f(x) = 0; f, x \in \mathbb{R}$ заменим на соответствующее уравнение: $x = \phi(x); \phi, x \in \mathbb{R}$, эквивалентное исходному. Как правило:

$$\phi(x) = x + \tau(x) \cdot f(x),$$

Где $\tau(x)$ - знакопостоянная на отрезке $[a, b]$ функция. Построим алгоритм в виде итерационного процесса:

$$x_{k+1} = \phi(x_k); k = 0, 1, \dots$$

Теорема. О сходимости метода скалярной простой итерации / *Галанин* /.

Пусть функция $\phi(x)$ непрерывная, ограниченная по Липшицу, функция с постоянной $q \in (0, 1)$ на отрезке $[x - \delta, x + \delta] = \Delta_\delta(x)$, то есть для любых значений $x_1, x_2 \in \Delta_\delta(x)$ выполняются неравенства: $|\phi(x_1) - \phi(x_2)| \leq q \cdot |x_1 - x_2|$ и $|\phi(x) - x| \leq (1 - q) \cdot \delta$. Тогда уравнение $x = \phi(x)$ имеет единственное решение x^* на отрезке $\Delta_\delta(x)$, которое может быть найдено в результате итерационного процесса $x_{k+1} = \phi(x_k); k = 0, 1, \dots$.

При этом погрешность найденного решения имеет следующую оценку

$$|x_k - x^*| \leq q \cdot |x_0 - x^*| \text{ или } |x_k - x^*| \leq \frac{q^k}{1 - q} \cdot |\phi(x_0) - x_0|; k = 0, 1, \dots$$

Следствие. Если верно не условие непрерывности по Липшицу, а следующее неравенство $|\phi'(x)| \leq q < 1$ для $\forall x \in \Delta_\delta(x)$, то все условия теоремы выполняются, и ее выводы справедливы.

Если погрешность метода удовлетворяет оценке $|x_k - x^*| \leq L \cdot q^k \cdot |x_0 - x^*|$, где $L > 0$ - некоторая постоянная величина, то говорят, что имеет место линейная сходимость алгоритма со скоростью геометрической прогрессии с показателем q .

Рассмотрим различные реализации метода простой итерации.

Если $\phi(x) = x + \tau \cdot f(x)$, то такая реализация называется методом релаксации.

Итерационный процесс тогда имеет вид:

$$\frac{x_{k+1} - x_k}{\tau} = f(x_k).$$

При этом будет выполняться соотношение:

$$\phi'(x) = 1 + \tau \cdot f'_x(x)$$

Очевидно, что метод будет сходиться, если выполняется неравенство $|1 + \tau \cdot f'_x(x)| < 1$ или $-2 < \tau \cdot f'_x(x) < 0$.

Если $\phi(x) = x - \frac{f(x)}{f'(x)}$, то такая реализация называется методом Ньютона или методом касательных.

Производная функции $\phi'_x(x)$, как несложно вычислить, определяется равенством:

$$\phi'_x(x) = \frac{f(x) \cdot f''_{xx}(x)}{(f'_x(x))^2}.$$

Пусть на отрезке $[a, b]$ выполняются следующие неравенства:

$$|f'_x(x)| \geq m > 0, \quad |f''_{xx}(x)| \leq M.$$

Тогда существует ε - окрестность корня x^* такая, что если $x_0 \in \Delta_\varepsilon(x^*)$, то итерационный процесс будет сходиться к решению x^* . Это следует из того, что всюду на отрезке $[a, b]$ имеет место неравенство:

$$|\phi'_x(x)| = \left| \frac{f(x) \cdot f''_{xx}(x)}{(f'_x(x))^2} \right| \leq \frac{|f(x)|}{m^2} \cdot M.$$

С другой стороны из непрерывности функции $f(x)$ следует, что существует некоторая ε - окрестность корня x^* , где выполняется неравенство:

$$|f(x)| \leq q \cdot \frac{m^2}{M}; 0 < q < 1$$

Таким образом, в этой окрестности будут справедливы условия указанной выше теоремы и ее выводы. Однако о сходимости метода Ньютона можно говорить только в малом, то есть при попадании начального приближения x_0 в окрестность корня x^* .

Используя разложение функции $f(x^*)$ до второго порядка можно записать следующее равенство:

$$f(x^*) = f(x_k) + f'_x(x_k) \cdot (x^* - x_k) + \frac{1}{2} \cdot f''_{xx}(\xi) \cdot (x^* - x_k)^2 = 0,$$

где $\xi \in (x_k, x^*)$ или $\xi \in (x^*, x_k)$. Отсюда получим следующее соотношение:

$$x_{k+1} = x_k - \frac{f(x_k)}{f'_x(x_k)} = x_k - \frac{f(x_k) - f(x^*)}{f'_x(x_k)} = x^* + \frac{1}{2} \cdot \frac{f''_{xx}(\xi)}{f'_x(x_k)} \cdot (x^* - x_k)^2.$$

Следовательно:

$$|x_{k+1} - x^*| \leq \frac{M}{2m} \cdot |x_k - x^*|^2 \leq \left(\frac{M}{2m}\right)^{2^{k+1}-1} \cdot |x_0 - x^*|^{2^{k+1}}$$

Таким образом, метод Ньютона будет сходиться с квадратичной скоростью.

В общем случае, если итерационный алгоритм сходится к решению с оценкой:

$$|x_{k+1} - x^*| \leq L \cdot |x_k - x^*|^p,$$

где $L > 0$ - некоторая постоянная, то говорят, что метод сходится со скоростью порядка p .

Рассмотрим теперь метод хорд или метод секущих.

В этом случае имеем:

$$\phi(x) = x - f(x) \cdot \frac{b - x}{f(b) - f(x)},$$

где принимается $\phi(b) = b - f(x) \cdot \frac{f(b)}{f'_x(b)}$.

Примем, что функция $f(x)$ является дважды дифференцируемой и, что выполняются неравенства $f'_x(x) > 0; f''_{xx}(x) \geq 0$ на отрезке $[a, b]$. Пусть также, что на k -ой итерации выполняется неравенство $x_k < x^* < b$.

Тогда с учетом приведенных условий и формулы Лагранжа можно записать соотношение:

$$x_{k+1} - x_k = -f(x_k) \cdot \frac{b - x_k}{f(b) - f(x_k)} = -f'_x(\xi) \cdot (x_k - x^*) \cdot \frac{b - x_k}{f'_x(\eta) \cdot (b - x^*) + f'_x(x) \cdot (x^* - x_k)},$$

где $\xi \in (x_k, x^*); \eta \in (x^*, b); \xi < \eta$.

Отсюда получим неравенство:

$$0 < x_{k+1} - x_k \leq x^* - x_k.$$

То есть последовательность итерационных приближений $\{x_k\}$ монотонно возрастает и ограничена сверху значением корня. Следовательно, по теореме Вейерштрасса она сходится. Аналогично можно рассмотреть ситуацию, когда $f'_x(x) < 0; f''_{xx}(x) \leq 0$.

Для разных знаков первой и второй производной, метод хорд надо использовать с заменой концевых точек отрезка, то с заменой b на a .

Методы решения нелинейного векторного уравнения.

Пусть необходимо решить уравнение $F(x) = 0; F, x \in \mathfrak{R}^n$. Рассмотрим простые итерационные методы решения такого уравнения вида:

$$A_{k+1} \cdot \frac{x^{(k+1)} - x^{(k)}}{\tau_{k+1}} + F(x^{(k)}) = 0; k = 0, 1, 2, \dots$$

где $x^{(k)}$ - приближение решения на k -ой итерации; A_{k+1} - невырожденный линейный оператор (матрица); τ_{k+1} - весовые коэффициенты алгоритма.

Для нахождения приближения $x^{(k+1)}$ надо решить уравнение:

$$A_{k+1} \cdot x^{(k+1)} = A_{k+1} \cdot x^{(k)} - \tau_{k+1} \cdot F(x^{(k)})$$

Рассмотрим стационарный случай, когда $A_{k+1} = A; \tau_{k+1} = \tau$. Тогда получим следующее соотношение:

$$x^{(k+1)} = x^{(k)} - \tau \cdot A^{-1} \cdot F(x^{(k)}) = \Phi(x^{(k)})$$

Тогда метод простых итераций будет определяться следующим формальным рекуррентным равенством $x^{(k+1)} = \Phi(x^{(k)})$. При определенных условиях последовательность $\{x^{(k)}\}$ будет приближаться к вектору x^* - неподвижной точке изображения $\Phi(x)$, то есть $x^* = \Phi(x^*)$.

Теорема. Пусть функция $\Phi(x)$ и замкнутое множество $M \subseteq \mathbb{R}^n$ таковы, что: $\Phi(x) \in M$ для $\forall x \in M$; существует, такое значение $0 < q < 1$, что $\|\Phi(x) - \Phi(\tilde{x})\| \leq q \|x - \tilde{x}\|$ для $\forall x, \tilde{x} \in M$. Тогда $\Phi(x)$ имеет в множестве $M \subseteq \mathbb{R}^n$ единственную неподвижную точку x^* , а последовательность $\{x^{(k)}\}$, определяемая соотношением $x^{(k+1)} = \Phi(x^{(k)})$, сходится при любом $x^{(0)} \in M$ к точке x^* . При этом справедливы следующие оценки

$$\|x^* - x^{(k)}\| \leq \frac{q}{1-q} \|x^{(k)} - x^{(k-1)}\| \leq \frac{q^k}{1-q} \|x^{(1)} - x^{(0)}\|, \forall k \in \mathbb{N}.$$

В случае, когда выбрано хорошее начальное приближение $x^{(0)}$ и удалось построить оператор $\Phi: \mathbb{R}^n \rightarrow \mathbb{R}^n$ с хорошими сжимающими свойствами, можно сформулировать следующее утверждение.

Теорема. Пусть функция $\Phi(x)$ дифференцируема в замкнутом шаре $S(x^{(0)}, r) \subseteq M \subset \mathbb{R}^n$, причем $\exists q \in (0,1)$ такое, что $\sup_{x \in S} \|\dot{\Phi}(x)\| \leq q$. Тогда, если центр $x^{(0)}$ и радиус шара S таковы, что выполняется неравенство $\|x^{(0)} - \Phi(x^{(0)})\| \leq r(1-q)$, то последовательность $\{x^{(k)}\}$, определяемая соотношением $x^{(k+1)} = \Phi(x^{(k)})$, сходится к единственной неподвижной точке x^* при $M = S$.

Метод Ньютона для векторного уравнения.

Если известно достаточно хорошее начальное приближение к решению системы уравнений $F(x) = 0$, где $F = (f_1, f_2, \dots, f_n); x \in \mathbb{R}^n$, то эффективным методом повышения точности является метод Ньютона. Основная идея метода Ньютона заключается в замене исходной задачи на некоторую вспомогательную линейную задачу в окрестности приближения $x^{(k)}$.

Пусть $F(x)$ - некоторый оператор, отображающий полное линейное нормированное пространство X на полное линейное нормированное пространство Y . Линейный оператор P , действующий из пространства X в пространство Y , назовем производной Фреше оператора $F(x)$ в точке x , если:

$$F(x + \eta) - F(x) = P \cdot \eta + r_0(x, \eta).$$

Причем для остаточного члена верно соотношение $\frac{\|r_0(x, \eta)\|_Y}{\|\eta\|_X} \xrightarrow{\|\eta\|_X \rightarrow 0} 0$ или

$$\|F(x + \eta) - F(x) - P\eta\|_Y = o(\|\eta\|_X).$$

Если функции f_i непрерывно дифференцируемы в окрестности данной точки x , то можно записать следующее соотношение:

$$f_i(x_1 + \eta_1, \dots, x_n + \eta_n) = f_i(x_1, \dots, x_n) + \sum_{j=1}^n \frac{\partial f_i}{\partial x_j} \eta_j + o(\|\eta\|_X).$$

Тогда, в качестве оператора P примем, оператор умножения слева на матрицу Якоби $F'_x(x) = \left(\frac{\partial F}{\partial x} \right)$. Пусть x^* решение уравнения $F(x^*) = 0$, а $x^{(k)}$ некоторое приближение к точке x^* .

В предположении существования производной $F'_x(x) = \left(\frac{\partial F}{\partial x} \right)$ можно записать

соотношение:

$$\| F(x^*) - F(x^{(k)}) - F'_x(x^{(k)})(x^* - x^{(k)}) \|_Y = o(\| x^* - x^{(k)} \|_X).$$

Если выбрано хорошее приближение к решению, то величина $\| x^* - x^{(k)} \|_X$ мала. Тогда можно записать соотношение $F(x^*) \approx F(x^{(k)}) + F'_x(x^{(k)})(x^* - x^{(k)})$. Так как $F(x^*) = 0$, то можно записать следующее итерационное соотношение $x^{(k+1)} = x^{(k)} - (F'_x(x^{(k)}))^{-1} F(x^{(k)})$. Такой итерационный процесс называют методом Ньютона. Пусть $\Omega_a = \{x : \| x - x^* \|_X < a\}$ и, при некоторых a, a_1, a_2 , таких, что $a > 0$ $0 \leq a_1, a_2 < \infty$, выполнены условия $\| (F'_x(x))^{-1} \|_Y \leq a_1$ при $\forall x \in \Omega_a$ и $\| F(u_1) - F(u_2) - F'_u(u_2)(u_1 - u_2) \|_Y \leq a_2 \| x^* - x^{(k)} \|_X^2$ при $\forall u_1, u_2 \in \Omega_a$. Тогда можно показать, что итерационный процесс Ньютона сходится с оценкой погрешности /Бахвалов с333/ :

$$\| x^{(k)} - x^* \|_X \leq C^{-1} (b \| x^{(0)} - x^* \|_X)^{2^n},$$

где $C = a_1 \cdot a_2$, $b = \min\{a, C^{-1}\}$.

Рассмотренные итерационные методы являются линейными. Однако широко используются и нелинейные итерационные методы.

Одним из популярных нелинейных методов является метод Зейделя, который является модификацией метода простых итераций. В алгоритме, после задания начального приближения $x^{(0)}$, вместо параллельного итерирования производится последовательное итерирование, причем на каждой итерации в каждое последующее уравнение подставляются значения неизвестных, полученных из предыдущих уравнений. То есть векторное уравнение $x = \Phi(x)$ записывается в виде системы n скалярных уравнений вида $x_i = \phi_i(x_1, x_2, \dots, x_n); i = 1, 2, \dots, n$ и реализуется следующий итерационный процесс:

$$x_i^{(k+1)} = \phi_i(x_1^{(k+1)}, x_2^{(k+1)}, \dots, x_{i-1}^{(k+1)}, x_i^{(k)}, \dots, x_n^{(k)}); i = 1, 2, \dots, n; k = 0, 1, 2, \dots$$

Условием прекращения процесса является выполнение неравенства:

$$\max_{i=1,2,\dots,n} \{ | x_i^{(k+1)} - x_i^{(k)} | \} < \varepsilon,$$

где ε - заданная точность решения; при этом полагается, что $x_i^* \approx x_i^{(k+1)}; i = 1, 2, \dots, n$.

Отличие метода Зейделя от метода простых итераций заключается в следующем:

- при вычислении $x_2^{(k+1)}$ вместо значения $x_1^{(k)}$ используется только что вычисленное значение $x_1^{(k+1)}$;
- значение $x_3^{(k+1)}$ вычисляется уже с использованием вычисленных в текущей итерации значений $x_1^{(k+1)}$, $x_2^{(k+1)}$ и так далее.

Такой прием позволяет увеличить скорость сходимости итераций. Преимущество метода Зейделя в скорости тем больше, чем больше порядок рассматриваемой системы.

Решатели нелинейных алгебраических уравнений в среде Matlab.

Для нахождения нуля функции $f(x)$ одной переменной x в окрестности заданной точки можно использовать функцию $fzero(\cdot)$, вызов которой описывается как:

$$z = fzero(hfun, x0, tol),$$

где:

аргумент $hfun$ - указатель на функцию вычисления значений $f(x)$;

аргумент $x0$ - значение начального приближения;

аргумент tol - относительная погрешность.

Функция возвращает результат решения уравнения $f(x) = 0$ с относительной погрешностью tol .

Для решения системы нелинейных алгебраических уравнений $F(x) = 0$ используется функция $fsolve(\cdot)$ пакета Optimization Toolbox Matlab. Метод решения базируется на минимизации суммы квадратов компонент функции $F(x)$ на основе алгоритмов Гаусса – Ньютона и Левенберга – Маркварда. Простейший вызов функции имеет вид:

$$x = fsolve(hfun, x0, options),$$

где:

аргумент $hfun$ - указатель на функцию вычисления значений $F(x)$;

аргумент $x0$ - значение начального приближения;

аргумент $options$ - обеспечивает учет специфических деталей функции $F(x)$ в виде вектора параметров.

Численные методы вычисления матрицы Якоби.

Для вычисления матрицы Якоби функции $F(x)$; $F, x \in \mathbb{R}^n$ в заданной точке $\bar{x} \in \mathbb{R}^n$, где $F(\bar{x}) = 0$, необходимо найти частные производные всех функций системы по всем переменным. Для

вычисления производной $\left. \frac{\partial F(x)}{\partial x} \right|_{x=\bar{x}}$ применяются методы численные методы вычисления первой производной.

Формула для отдельного элемента матрицы Якоби при использовании правой разностной производной имеет вид:

$$J_{ij} = \frac{F_i(x_1, x_2, \dots, x_j + h_{ij}, x_{j+1}, \dots, x_n) - F_i(x_1, x_2, \dots, x_j, \dots, x_n)}{h_{ij}}$$

Формула для элемента якобиана при использовании центральной разностной производной:

$$J_{ij} = \frac{F_i(x_1, x_2, \dots, x_j + h_{ij}, x_{j+1}, \dots, x_n) - F_i(x_1, x_2, \dots, x_j - h_{ij}, x_{j+1}, \dots, x_n)}{2h_{ij}}.$$

Вычисление якобиана с использованием правой разностной производной требует вычислять значения функций в $n \cdot (n + 1)$ точках. Если использовать центральную производную, то нужно находить значения функций $\bar{x} = (\bar{x}_1, \bar{x}_2, \dots, \bar{x}_n)^T$ в $2 \cdot n^2$ точках. С другой, стороны погрешность правой производной имеет порядок $O(h)$, а в центральной производной - $O(h^2)$. В большинстве

случаев вычисление значения функции в точке $\bar{x} = (\bar{x}_1, \bar{x}_2, \dots, \bar{x}_n)^T$ - это затратная по времени операция, поэтому используется правая разностная производная.

Основная проблема при вычислении каждого элемента матрицы Якоби - как правильно выбрать шаг метода h_{ij} . Шаг выбирается независимо для каждого элемента матрицы.

Проанализируем зависимость погрешности метода от величины шага в случае использования правой разностной производной. Для сокращения записи введём следующее обозначение $F(\xi_j) = F(x_1, x_2, \dots, \xi_j, x_{j+1}, \dots, x_n)$. Остаточный член в соотношении:

$$F'(x) \approx \frac{F(x+h) - F(x)}{h}$$

имеет вид: $r_1 = -\frac{F''_{xx}(\xi) \cdot h}{2}$, где $\xi \in (x, x+h)$.

Если $|F''_{xx}(\xi)| \leq M_2$, то $|r_1| \leq \frac{M_2 \cdot h}{2}$. Соответственно, если значения $F(x+h)$ и $F(x)$

заданы с погрешностями $\varepsilon_1, \varepsilon_2 \leq e$, то погрешность будет содержать ещё одно слагаемое:

$$r_2 \leq -\frac{\varepsilon_1 - \varepsilon_2}{h} \leq \frac{2 \cdot e}{h}.$$

Таким образом, оценка суммарной погрешности имеет вид:

$$|r| \leq |r_1| + |r_2| \leq \frac{M_2 \cdot h}{2} + \frac{2 \cdot e}{h}.$$

Очевидно, что эта оценка достигает минимума при значении $h = h_0 = 2 \sqrt{\frac{e}{M_2}}$. Такая оценка

погрешности имеет один глобальный минимум. Поэтому выбор очень маленького шага не будет приводить к росту точности. При величине шага, близкой к значению $h \approx h_0$, погрешность имеет порядок $O(\sqrt{e})$.

Численные методы интегрирования обыкновенных дифференциальных уравнений.

Необходимость численного решения систем обыкновенных дифференциальных уравнений (ОДУ) возникает в случае, когда аналитическое решение найти либо невозможно, либо нерационально, а приближенное решение (в виде набора интерполирующих функций) не дает требуемой точности. Численное решение системы ОДУ позволяет построить только таблицу значений неизвестных функций, удовлетворяющих начальным условиям. По этим значениям можно построить графики данных функций, что обычно и требуется в физических или инженерных задачах.

Будем называть дифференциальным уравнением n -го порядка соотношение вида $\Phi(t, x, \dot{x}, \ddot{x}, \dots, x^{(n)}) = 0$. Если функцию Φ можно разрешить относительно старшей производной $x^{(n)} = F(t, \dot{x}, \dots, x^{(n-1)})$, то такое уравнение будем называть разрешенным относительно старшей производной. Введем следующие обозначения $x = y_1; \dot{x} = y_2; \dots; x^{(n-1)} = y_n$, то уравнение можно записать в виде следующей системы дифференциальных уравнений первого порядка: $\dot{y}_1 = y_2; \dots; \dot{y}_{n-1} = y_n; \dot{y}_n = F(y_1, \dots, y_n)$. Задачу нахождения решения такой системы, удовлетворяющей начальным условиям $y_1(t_0) = y_{10}, \dots, y_{n-1}(t_0) = y_{(n-1)0}$, называется задачей Коши

. Будем предполагать, что выполняются известные требования, обеспечивающие существование и единственность решения $y = (y_1, y_2, \dots, y_n)^T$ системы в некоторой заданной области переменных $y \in D \subseteq \mathbb{R}^n$. В дальнейшем будем также предполагать, что решение системы всегда обладает той или иной степенью гладкости, необходимой для построения и применения того или иного расчетного метода.

Методы Эйлера.

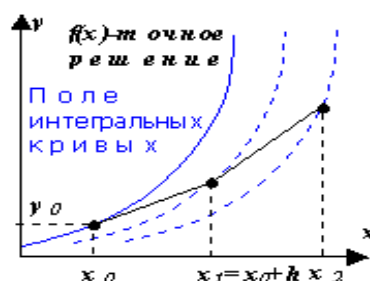
Рассмотрим задачу Коши для уравнения первого порядка:

$\dot{y}(t) = f(t, y(t)), t_0 \leq t \leq t_{\max}; y(t_0) = y_0$. В окрестности точки y_0 функцию $y(t)$ разложим в ряд Тейлора:

$$y(t) = y(t_0) + \dot{y}(t_0)(t - t_0) + \frac{1}{2} \ddot{y}(t_0)(t - t_0)^2 + \dots$$

Идея этого и последующих методов основывается на том, что если функция $f(t, y)$ имеет q непрерывных производных, то в разложении нужно учитывать члены вплоть до $O(h^{q+1})$, при этом стоящие в правой части производные можно найти, дифференцируя уравнение $\dot{y}(t) = f(t, y(t))$ требуемое число раз. В случае метода Эйлера ограничимся только двумя членами разложения. Пусть h малое приращение аргумента. Тогда можно записать соотношение

$y(t_0 + h) = y(t_0) + h \cdot \dot{y}(t_0) + O(h^2)$ или приближенно $y(t_0 + h) \approx y(t_0) + h \cdot \dot{y}(t_0)$. Таким образом, расчет можно организовать по следующей итерационной формуле: $y_{k+1} = y_k + h \cdot f(t_k, y_k)$, где $y_0 = y(t_0)$, а все y_k - приближенные значения искомой функции.



В методе Эйлера происходит движение не по интегральной кривой, а по касательной к ней. На каждом шаге касательная находится уже для новой интегральной кривой (что и дало название методу – метод ломаных), таким образом, ошибка будет возрастать с отдалением t от t_0 .

При значении $h \rightarrow 0$, приближенное решение сходится к точному решению равномерно с первым порядком точности. То есть, метод дает весьма низкую точность вычислений: погрешность на

элементарном шаге h составляет $\frac{1}{2} h^2 \ddot{y}[\frac{1}{2}(t_k + t_{k+1})]$, а для всей интегральной кривой порядка h .

Метод Эйлера легко обобщается для систем ОДУ. Общая схема процесса может быть записана так: $y_n^{(k+1)} = y_n^{(k)} + h \cdot F(t^{(k)}, y_1^{(k)}, \dots, y_n^{(k)})$. Рассмотрим следующую модификацию метода Эйлера. Запишем для интервала $t_0 \leq t \leq T$ следующее интегральное уравнение:

$$y(T) = y(t_0) + \int_{t_0}^T f(t, y(t)) dt.$$

Применяя к интегралу простейшую формулу трапеций можно записать следующее

итерационное равенство: $y^{(k+1)} = y^{(k)} + \frac{h}{2}[(f(t^{(k)}, y^{(k)}) + f(t^{(k+1)}, y^{(k+1)}))], k = 0:n$. Данный метод является неявным, так как требует решения на каждом шаге соответствующих уравнений относительно $y^{(k+1)}$. Представляет интерес совместное применение явного метода Эйлера и неявного метода трапеций:

$$y^{(k+1)} = y^{(k)} + \frac{h}{2}[f(t^{(k)}, y^{(k)}) + f(t^{(k)} + h, y^{(k)} + h \cdot f(t^{(k)}, y^{(k)}))], k = 0:n$$

Метод расчета, реализуемый с помощью указанной итерационной формулы, называется методом Хойна /Вержбицкий с222/. Если сделать для уточнения значения $y^{(k+1)}$ несколько внутренних шагов (итераций) по формуле Эйлера, то можно еще больше повысить точность расчетов. Такой вариант совместного применения прямого метода Эйлера и метода интегрирования трапеций называют модифицированным методом Эйлера-Коши. Этот метод относят к двух-шаговым процессам численного интегрирования ОДУ.

Прямой алгоритм Эйлера (метод ломанных).

Обозначения: N - количество точек интегрирования; $h = \frac{(T - t_0)}{N}$ – шаг интегрирования;

$y_0 = y(t_0)$ – начальное значение; $Z(k, 0)$ – столбец изменения аргумента решения t ; $Z(k, 1)$ – столбец изменения решения, $k = 0, 1, \dots, N$.

```

Z(0,0) = t0; Z(0,1) = y0;
for i = 1:N
    Z(i,0) = t0 + h · i;
    Z(i,1) = Z((i-1),1) + h · (f(Z((i-1),0), Z((i-1),1)));
end for i

```

Одношаговый метод Эйлера-Коши.

Обозначения: N - количество точек интегрирования; $h = \frac{(T - t_0)}{N}$ – шаг интегрирования;

$y_0 = y(t_0)$ – начальное значение; $Z(k, 0)$ – столбец изменения аргумента решения t ; $Z(k, 1)$ – столбец изменения решения, $k = 0, 1, \dots, N$.

```

Z(0,0) = t0; Z(0,1) = y0;
for i = 1:N
    Z(i,0) = t0 + h · i;
    Z1 = Z((i-1),1) + h · f(Z((i-1),0), Z((i-1),1));
    Z(i,1) = Z((i-1),1) + (h / 2) · (f(Z((i-1),0), Z((i-1),1)) + f(Z(i,0), Z1));
end for i

```

Одним из основных недостатков метода Эйлера является также необходимость вычисления на каждом шаге частных производных при увеличении количества членов разложения функции $f(t, y)$ в ряд Тейлора.

Семейство прямых методов Рунге-Кутты.

Приведенных выше недостатков в определенной степени лишены методы Рунге-Кутты. Идея такого метода заключается в получении приближений к значениям $y^{(k+1)}$ по так называемому

«правилу средней точки». Первый шаг длины h при таком подходе должен иметь вид

$$y^{(1)} = y(t_0 + h) \approx y^{(0)} + hf(t^{(0)} + \frac{h}{2}, y(t^{(0)} + \frac{h}{2})).$$

Для вычисления значения $y(t^{(0)} + \frac{h}{2})$ в правой

части соотношения, используется один малый шаг метода Эйлера длины $\frac{h}{2}$. Тогда соотношение

для первого шага можно записать в виде следующих соотношений:

$$\begin{cases} k_1 = f(t^{(0)}, y^{(0)}) \\ k_2 = f(t^{(0)} + \frac{h}{2}, y^{(0)} + \frac{h}{2}k_1) \\ y^{(1)} = y^{(0)} + hk_2 \end{cases}$$

результате такого построения процесса вычислений влияние погрешности становится менее существенным, чем в методе Эйлера. Вычислим для функции $y^{(1)}$ разложение Тейлора по степеням величины h :

$$y^{(1)} = y^{(0)} + hf(t^{(0)}, y^{(0)}) + \frac{h^2}{2}(f_t + f_y f)|_{(t^{(0)}, y^{(0)})} + \frac{h^3}{8}(f_{tt} + 2f_{ty}f + f_{yy}f^2)|_{(t^{(0)}, y^{(0)})} + \dots$$

Полученное разложение сравним с рядом Тейлора для точного решения, который получается из уравнения $\dot{y}(t) = f(t, y(t))$, $y(t_0) = y_0$ повторным дифференцированием с заменой \dot{y} на f каждый раз, когда оно появляется:

$$y(t_0 + h) = y^{(0)} + hf(t^{(0)}, y^{(0)}) + \frac{h^2}{2}(f_t + f_y f)|_{(t^{(0)}, y^{(0)})} + \frac{h^3}{6}(f_{tt} + 2f_{ty}f + f_{yy}f^2 + f_y f_t + f_y^2 f)|_{(t^{(0)}, y^{(0)})} + \dots$$

Вычитая из последнего соотношения предыдущее, получим для погрешности первого шага выражение:

$$y(t_0 + h) - y^{(1)} = \frac{h^3}{24}[f_{tt} + 2f_{ty}f + f_{yy}f^2 + 4(f_y f_t + f_y^2 f)]|_{(t^{(0)}, y^{(0)})} + \dots$$

Таким образом, если все частные производные f второго порядка ограничены, то $\|y(t^{(0)} + h) - y^{(1)}\| \leq Kh^3$.

В 1901 году Кутта сформировал общую численную схему интегрирования дифференциальных уравнений, которая теперь называется методом Рунге-Кутты.

Определение. Пусть s - целое положительное число (число этапов схемы) и

$a_{21}, a_{31}, a_{32}, \dots, a_{s1}, a_{s2}, \dots, a_{s,s-1}, b_1, \dots, b_s, c_2, \dots, c_s$ - вещественные коэффициенты. Тогда схема

$$\begin{cases} k_1 = f(t^{(0)}, y^{(0)}) \\ k_2 = f(t^{(0)} + c_2 h, y^{(0)} + h a_{21} k_1) \\ k_3 = f(t^{(0)} + c_3 h, y^{(0)} + h(a_{31} k_1 + a_{32} k_2)) \\ \dots \\ k_s = f(t^{(0)} + c_s h, y^{(0)} + h(a_{s1} k_1 + \dots + a_{s,s-1} k_{s-1})) \\ y_1 = y_0 + h(b_1 k_1 + \dots + b_s k_s) \end{cases}$$

называется s - стадийным (этапным) явным методом Рунге-Кутты.

Обычно коэффициенты c_i удовлетворяют условиям $c_2 = a_{21}$, $c_3 = a_{31} + a_{32}, \dots$, $c_s = a_{s1} + \dots + a_{s,s-1}$ или, короче $c_i = \sum_j a_{ij}$. Смысл этих условий состоит в том, что все точки, в

которых вычисляется f , являются приближениями первого порядка к решению /Хайер т1 с141/.

Выбор параметров $a_{21}, a_{31}, a_{32}, \dots, a_{s1}, a_{s2}, \dots, a_{s,s-1}, b_1, \dots, b_s, c_2, \dots, c_s$ осуществляется таким образом, чтобы функция $h(b_1 k_1 + \dots + b_s k_s)$ приближала отрезок ряда Тейлора функции $f(t, y)$ до некоторого p -го порядка и не содержала частных производных.

Определение. Метод Рунге-Кутты имеет порядок p , если для достаточно гладких задач

$\dot{y}(t) = f(t, y(t)), y(t_0) = y_0$ выполняется соотношение $\|y(t^{(0)} + h) - y^{(1)}\| \leq Kh^{p+1}$, то есть ряды Тейлора для точного решения $y(t^{(0)} + h)$ и для схемы Рунге-Кутты $y^{(1)}$ совпадают до члена h^p включительно.

Символически метод Рунге-Кутты представляют с помощью следующей формы, называемой таблицей Бутчера.

Методы Рунге—Кутты низших порядков				
0	Рунге, порядок 2		Рунге, порядок 3	
c_2	a_{21}		0	
c_3	a_{31} a_{32}		$\frac{1}{2}$ $\frac{1}{2}$	
\vdots	\vdots \vdots		1 0 1	
c_s	a_{s1} a_{s2} \dots $a_{s,s-1}$		1 0 0 1	
	b_1 b_2 \dots b_{s-1} b_s	$\frac{1}{2}$ $\frac{1}{2}$	$\frac{1}{6}$ $\frac{2}{3}$ 0 $\frac{1}{6}$	0 $\frac{1}{3}$
		0 1		$\frac{2}{3}$ 0 $\frac{2}{3}$
				$\frac{1}{4}$ 0 $\frac{3}{4}$

В семействе методов Рунге-Кутты наиболее употребительным является метод четвертого порядка. Чтобы получить оценки параметров схемы интегрирования, надо вычислить производные порядков 1, 2, 3 и 4 от функции $y^{(1)} = y^{(1)}(h)$ при $h = 0$ и сравнить их с производными точного решения. Отсюда получим следующие условия $\sum_i b_i = b_1 + b_2 + b_3 + b_4 = 1$;

$$\sum_i b_i c_i = b_2 c_2 + b_3 c_3 + b_4 c_4 = \frac{1}{2};$$

$$\begin{aligned} \sum_i b_i c_i^2 &= b_2 c_2^2 + b_3 c_3^2 + b_4 c_4^2 = \frac{1}{3}, & \sum_{i,j} b_i c_i a_{ij} c_j &= b_3 c_3 a_{32} c_2 + b_4 c_4 (a_{42} c_2 + a_{43} c_3) = \frac{1}{8}, \\ \sum_{i,j} b_i a_{ij} c_j &= b_3 a_{32} c_2 + b_4 (a_{42} c_2 + a_{43} c_3) = \frac{1}{6}, & \sum_{i,j} b_i a_{ij} c_i^2 &= b_3 a_{32} c_2^2 + b_4 (a_{42} c_2^2 + a_{43} c_3^2) = \frac{1}{12}, \\ \sum_i b_i c_i^3 &= b_2 c_2^3 + b_3 c_3^3 + b_4 c_4^3 = \frac{1}{4}, & \sum_{i,j,k} b_i a_{ij} a_{jk} c_k &= b_4 a_{43} a_{32} c_2 = \frac{1}{24}. \end{aligned}$$

Решая эти уравнения относительно искомых параметров при различных допущениях, можно получать различные схемы метода Рунге-Кутты 4-го порядка. Наиболее популярными стали следующие два варианта. Первый из методов более популярен, а второй более точен.

«Классический» метод Рунге— Кутты					Правило 3/8				
0					0				
$\frac{1}{2}$	$\frac{1}{2}$				$\frac{1}{3}$	$\frac{1}{3}$			
$\frac{1}{2}$	0	$\frac{1}{2}$			$\frac{2}{3}$	$-\frac{1}{3}$	1		
1	0	0	1		1	1	-1	1	
	$\frac{1}{6}$	$\frac{2}{6}$	$\frac{2}{6}$	$\frac{1}{6}$		$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

Метод Рунге – Кутты - Мерсона.

Этот метод отличается от метода Рунге – Кутты четвертого порядка возможностью оценивать погрешность на каждом шаге и в зависимости от этого принимать решение об изменении шага. Один из вариантов формул этого метода имеет вид:

$$\left\{ \begin{array}{l} y^{(k+1)} = y^{(k)} + \frac{1}{2}(m_4^{(k)} + m_5^{(k)}) + O(h^5) \\ m_1^{(k)} = h_3 f(t^{(k)}, y^{(k)}); h_3 = \frac{h}{3} \\ m_2^{(k)} = h_3 f(t^{(k)} + h_3, y^{(k)} + m_1^{(k)}) \\ m_3^{(k)} = hf[t^{(k)} + h_3, y^{(k)} + \frac{1}{2}(m_1^{(k)} + m_2^{(k)})] \\ m_4^{(k)} = m_1^{(k)} + 4h_3 f[t^{(k)} + \frac{2}{3}h, y^{(k)} + 0.375(m_1^{(k)} + m_3^{(k)})] \\ m_5^{(k)} = h_3 f[t^{(0)} + h, y^{(0)} + \frac{3}{2}(m_4^{(k)} - m_3^{(k)})] \end{array} \right.$$

Здесь $R_k = 0.2m_4 - 0.3m_3 - 0.1m_5$ - погрешность на каждом шаге. Пусть задана максимальная погрешность ξ .

Тогда, если $|R_{n+1}| > \xi$, $h = h/2$, то $(k+1)$ цикл расчета повторяется (с точки $y^{(k)}$) с новым шагом равным $h = h/2$. Если же $|R_{n+1}| < \frac{\xi}{32}$, то цикл расчета повторяется с точки $y^{(k)}$ с новым шагом $h = h/2$. Автоматический выбор шага позволяет значительно сократить время решения ОДУ.

Линейные многошаговые разностные методы.

Рассмотрим задачу Коши для дифференциального уравнения:

$$\dot{y} = f(t, y); \quad y(0) = y_0; t > 0; \quad y, f \in \mathbb{R}^n.$$

Пусть решение уравнения заданно на сетке ω_τ с постоянным шагом $\tau > 0$. Обозначим точки решения в моменты t_k , как $y^{(k)} = y(t_k)$ и, соответственно, значения правых частей $f^{(k)} = f(t_k, y^{(k)})$.

Тогда линейные m - шаговые разностные методы решения обыкновенных дифференциальных уравнений описываются с помощью систем разностных уравнений вида:

$$\frac{a_0 \cdot y^{(k)} + a_1 \cdot y^{(k-1)} + \dots + a_m \cdot y^{(k-m)}}{\tau} = b_0 \cdot f^{(k)} + b_1 \cdot f^{(k-1)} + \dots + b_m \cdot f^{(k-m)}; k = m, m+1, \dots$$

Решение системы разностных уравнений начинается со значения $k = m$. Следовательно, для обеспечения возможности расчета необходимо задать значения: $y^{(0)}, y^{(1)}, \dots, y^{(m-1)}$. Обычно эти значения заранее неизвестны (за исключением $y^{(0)} = y(0)$), но их можно вычислить с помощью каких – либо иных методов, например Рунге – Кутты. В рассматриваемом выражении все коэффициенты a_i, b_i - заранее заданные коэффициенты и $a_0 \neq 0$. Так как метод является определенным, когда коэффициенты заданы с точностью до постоянного сомножителя, то обычно задается условие нормировки в виде:

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

Если $b_0 = 0$, то метод называется явным, если $b_0 \neq 0$, то неявным.

Устойчивость и сходимость разностных методов.

Обычно методы высокого порядка аппроксимации редко используются, в связи с трудностями обеспечения их устойчивости.

Рассмотрим соответствующее однородное разностное уравнение с постоянными коэффициентами:

$$\sum_{j=0}^m a_j y_{k-j} = 0,$$

Данному уравнению будет соответствовать следующее характеристическое уравнение:

$$\sum_{j=0}^m a_j z^{m-j} = 0.$$

Очевидно, что многошаговый процесс разностный процесс будет устойчивым, если все корни характеристического уравнения будут лежать внутри единичной окружности на комплексной плоскости. Однако, данное условие очень общее и, никак не учитывает структуру правой части обыкновенного дифференциального уравнения, а, следовательно, характерных особенностей решения.

Пример.

Рассмотрим, следующую задачу Коши: $\dot{y} = -\alpha^2 \cdot y$; $y(0) = y_0$

Очевидно, что $y(t) = y_0 \cdot \exp(-\alpha^2 \cdot t)$. То есть решение монотонно убывает.

Рассмотрим метод Эйлера: $\frac{y^{(k+1)} - y^{(k)}}{\tau} = -\alpha^2 \cdot y^{(k)}$.

Отсюда найдем $y^{(k+1)} = y^{(k)} \cdot (1 - \alpha^2 \cdot \tau)$. Тогда метод Эйлера будет устойчив, при условии: $0 \leq \tau \leq \frac{2}{\alpha^2}$.

Пример.

Рассмотрим задачу Коши: $\begin{cases} \dot{y}_1 = -y_1 \\ \dot{y}_2 = -\varepsilon^2 \cdot y_2 \end{cases}$. Тогда получим: $\begin{cases} y_1 = y_{1,0} \cdot \exp(-t) \\ y_2 = y_{2,0} \cdot \exp(-\varepsilon^2 \cdot t) \end{cases}$

Пусть задача решается на участке $(0, T)$, причем $\varepsilon^2 T \gg 1$. Если система решается явным методом Эйлера,

то должно быть выполнено условие: $\tau \leq \min\{2, \frac{2}{\varepsilon^2}\}$.

Очевидно, выполнение этого ограничения становится бессмысленным с некоторого момента времени, так как функция $y_1(t)$ к моменту времени $t = 5$ становится пренебрежимо малой, а именно с ней связывается

количество временных шагов, равных $\frac{T}{\tau} = \frac{5 \cdot \varepsilon^2}{2}$.

Метод Адамса.

Линейный метод с коэффициентами $a_0 = -a_1 = 1$; $a_k = 0$; $k = 2, 3, \dots, m$ называется методом Адамса. Метод может быть описан соотношением:

$$\frac{y^{(k+1)} - y^{(k)}}{h} = \sum_{i=0}^m b_i f^{(k-i)}.$$

Будем считать, что уже найдено несколько приближенных значений $y_j \approx y(t_j)$, $j = 1, 2, \dots, m$, решения задачи $\dot{y} = f(t, y)$; $y(t_0) = y_0$. Построим правила для вычисления значения $y^{(k+1)} \approx y(t^{(k+1)})$. Для этого будем использовать интегрально интерполяционный подход. В

уравнение $y^{(L+1)} = y^{(L)} + \int_{L_t}^{(L+1)t} f(t, y(t)) dt$, под знак интеграла, вместо функции $f(t, y)$, подставим ее интерполяционный многочлен Ньютона-Лейбница:

$$P_L(t) = P_L(t^{(k)} + qh) = f_{(k)} + q\Delta f_{(k-1)} + \frac{q(q+1)}{2!} \Delta^2 f_{(k-2)} + \dots$$

Проведя вычисления, можно получить следующую конечно-разностную формулу, определяющую экстраполяционный многочлен Адамса - Башфорта:

$$y^{(k+1)} = y^{(k)} + h(f_{(k)} + \frac{1}{2}\Delta f_{(k-1)} + \frac{5}{12}\Delta^2 f_{(k-2)} + \frac{3}{8}\Delta^3 f_{(k-3)} + \dots)$$

Рассмотрим конкретные формулы для метода Адамса второго и третьего ($k=2,3$) порядков:

$$\left\{ \begin{array}{l} y^{(k+1)} = y^{(k)} + \frac{h}{12}[23f(t^{(k)}, y^{(k)}) - 16f(t^{(k-1)}, y^{(k-1)}) + 5f(t^{(k-2)}, y^{(k-2)})]; k = 2 \\ y^{(k+1)} = y^{(k)} + \frac{h}{24}[55f(t^{(k)}, y^{(k)}) - 59f(t^{(k-1)}, y^{(k-1)}) + 37f(t^{(k-2)}, y^{(k-2)}) - 9f(t^{(k-3)}, y^{(k-3)})]; k = 3 \end{array} \right.$$

Можно показать, что метод Адамса, порождаемый интерполированием с помощью многочлена k -ой степени, является методом $(k+1)$ -го порядка точности $O(h^{k+1})$.

Методы прогноза и коррекции (предиктор-корректорные методы Адамса).

Под названием «метод прогноза и коррекции» понимается совместное применение явных и неявных методов одинакового или смежных порядков. С помощью явной экстраполяционной формуле значение $y^{(k+1)}$ прогнозируется, то есть определяется его достаточно грубое приближение, а с помощью неявной формулы, в правую часть которой подставляется спрогнозированное значение, полученное прогнозное значение уточняется.

Ниже приведены формулы метода прогноза и коррекции разного порядка.

$$\begin{aligned}
 & \left\{ \begin{array}{l} \text{1-ый_порядок} \\ y_p^{(k+1)} = y^{(k)} + hf(t^{(k)}, y^{(k)}) \\ y^{(k+1)} = y^{(k)} + hf(t^{(k+1)}, y_p^{(k+1)}) \end{array} \right\}, \quad \left\{ \begin{array}{l} \text{2-ый_порядок} \\ y_p^{(k+1)} = y^{(k)} + \frac{h}{2}[3f(t^{(k)}, y^{(k)}) - f(t^{(k-1)}, y^{(k-1)})] \\ y^{(k+1)} = y^{(k)} + \frac{h}{2}[f(t^{(k+1)}, y_p^{(k+1)}) + f(t^{(k)}, y^{(k)})] \end{array} \right\} \\
 & \qquad \qquad \qquad \text{4-ый_порядок} \\
 & \left\{ \begin{array}{l} y_p^{(k+1)} = y^{(k)} + \frac{h}{24}[55f(t^{(k)}, y^{(k)}) - 59f(t^{(k-1)}, y^{(k-1)}) + 37f(t^{(k-2)}, y^{(k-2)}) - 9f(t^{(k-3)}, y^{(k-3)})] \\ y^{(k+1)} = y^{(k)} + \frac{h}{24}[9f(t^{(k+1)}, y_p^{(k+1)}) + 19f(t^{(k)}, y^{(k)}) - 5f(t^{(k-1)}, y^{(k-1)}) + f(t^{(k-2)}, y^{(k-2)})] \end{array} \right\}.
 \end{aligned}$$

Все S-шаговые методы Адамса можно описать одной формулой:

$$y^{(k+1)} = y^{(k)} + h \sum_{j=0}^S \beta_j f(t^{(k+1-j)}, y^{(k+1-j)}),$$

где S должно быть фиксированным натуральным числом, а K может принимать значения $S-1, S, S+1, \dots$. При $\beta_0 = 0$ приведенная формула определяет явные, а при $\beta_0 \neq 0$ неявные методы.

Методы Адамса высокого порядка выгодно применять при построении решения на длительном интервале. При этом возникают проблемы с вычислением первых $S-1$ разгонных значений $y^{(1)}, y^{(2)}, \dots, y^{(S-1)}$. Для их получения можно использовать одношаговые методы Рунге-Кутты.

В пакете MatLab реализованы следующие методы решения обыкновенных дифференциальных уравнений: ode45 – одношаговые явные методы Рунге-Кутты 4-го и 5-го порядка; ode23 – одношаговые явные методы Рунге-Кутты 2-го и 4-го порядка; ode113 – многошаговый метод Адамса – Башворта - Мултона переменного порядка и другие.

Численные методы интегрирования жестких системы.

Явление жесткости дифференциального уравнения состоит в том, что решение, которое нужно вычислить меняется медленно. Однако существуют быстро затухающие возмущения. Наличие таких возмущений затрудняет получение медленно меняющегося решения численным способом.

Например, рассмотрим задачу: $\dot{y} = -100y + 100; y(0) = 2$. Точным решением уравнения является функция $y = 1 + e^{-100t}$. Тогда выбор очень мелкого шага приведет к удовлетворительной оценке начального изменения, но потребует больших вычислительных затрат на получение всего решения. Кроме того, наличие большого количества шагов может привести к накоплению ошибок и потере устойчивости численного метода. А выбор относительно большого шага не позволит оценить все решение, включая начальный участок.

Рассмотрим предварительно следующую линейную систему $\dot{x} = Ax$. Система с постоянной $(n \times n)$ матрицей A называется жесткой, если собственные числа λ_k матрицы A удовлетворяют следующим условиям:

$$1. \operatorname{Re}(\lambda_k) < 0; k = 1:n$$

$$2. \text{Число жесткости (обусловленности)} \quad g(A) = \frac{\max \|\operatorname{Re} \lambda_k\|}{\min \|\operatorname{Re} \lambda_k\|} \text{ велико.}$$

Вместо числа жесткости можно использовать значение числа обусловленности матрицы

$k(A) = \|A\| \cdot \|A^{-1}\|$. При определении жесткости нелинейной системы $\dot{x} = F(t, x)$ в основу определения чисел обусловленности или жесткости кладут свойства матрицы Якоби $\left\| \frac{\partial F}{\partial x} \right\|$.

Важным является вопрос изучения сходимости различных алгоритмов, используемых для решения жестких уравнений. Для этого используют, так называемую, модельную задачу вида $\dot{z} = \Lambda z$, где Λ - диагональная матрица с различными значениями собственных чисел.

Определение. Численный метод называется A -устойчивым, если его область устойчивости включает всю полуплоскость $\text{Re}(\lambda_k) < 0; k = 1:n$.

Поэтому в качестве основы для выбора алгоритмов для решения жестких систем обычно кладут анализ областей абсолютной устойчивости при решении модельной задачи.

Рассмотрим линейные многошаговые методы вида:

$$\frac{1}{h} \cdot \sum_{i=0}^m a_i \cdot y_{k-i} = \sum_{i=0}^m b_i \cdot f_{k-i}.$$

Применительно к уравнению $\dot{y} = \Lambda y$, где Λ - диагональная матрица с одинаковыми значениями λ собственных чисел. В этом случае разностный метод сводится к итерационному решению следующей системы алгебраических уравнений:

$$\sum_{i=0}^m (a_i - \lambda \cdot h \cdot b_i) \cdot y_{k-i} = 0.$$

Соответствующее характеристическое уравнение разностного процесса имеет вид:

$$\sum_{i=0}^m (a_i - \lambda \cdot h \cdot b_i) \cdot z^{m-i} = 0.$$

Будем называть областью устойчивости m -шагового разностного алгоритма множество точек $\lambda \cdot h$ комплексной плоскости, для которых данный метод применительно к уравнению $\dot{y} = \Lambda y$ будет устойчив.

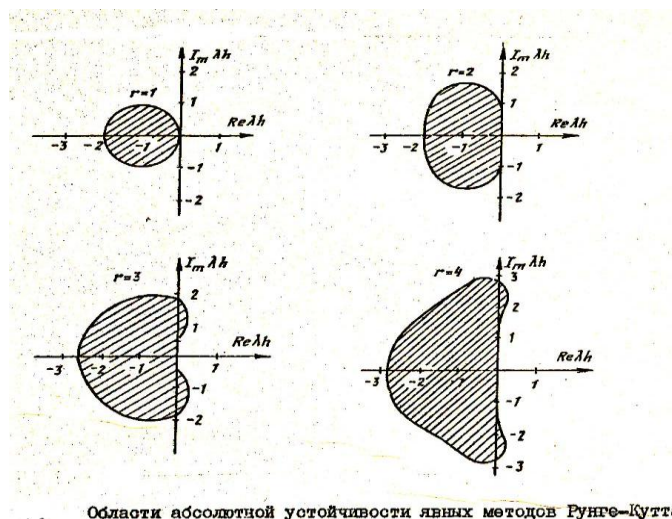
Пример.

Явный метод Эйлера для скалярного уравнения $\dot{y} = \lambda \cdot y$ можно записать в следующем виде:

$y_{k+1} = y_k \cdot (1 + \lambda \cdot h)$. Отсюда получим условие устойчивости: $|1 + \lambda \cdot h| \leq 1$. Таким образом, область устойчивости на комплексной плоскости представляет собой круг единичного радиуса с центром в точке $(-1, 0)$.

Соответственно неявный метод Эйлера имеет вид $y_{k+1} = y_k \cdot (1 - \lambda \cdot h)^{-1}$. Тогда условие устойчивости имеет вид: $|(1 - \lambda \cdot h)^{-1}| \leq 1$, что соответствует внешности единичного круга с центром в точке $(1, 0)$.

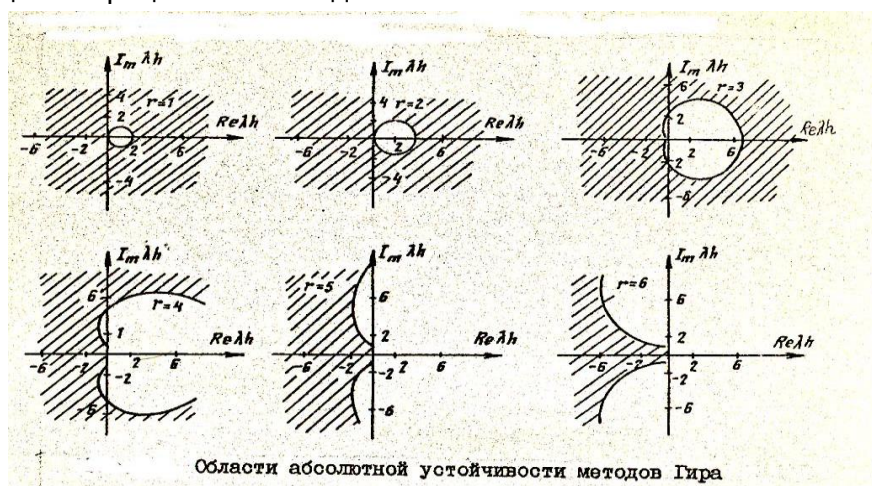
В частности, анализ областей устойчивости явных методов Рунге-Кутты показал, что, за исключением методов Адамса-Мултона первого и второго порядков, они не удовлетворяют требованиям для численного решения дифференциальных уравнений, так как не являются A -устойчивыми /Кудинов с23/.



Одним из методов, хорошо предназначенных для решения жестких систем, является неявный метод Гира. Ниже приведены параметры этого класса методов для $r = 1, 2, 3, 4$ /Кудинов с21/:

$$\begin{cases} y^{(k+1)} = y^{(k)} + hf(t^{(k+1)}, y^{(k+1)}); r = 1 \\ y^{(k+1)} = \frac{4}{3}y^{(k)} - \frac{1}{3}y^{(k-1)} + \frac{2}{3}hf(t^{(k+1)}, y^{(k+1)}); r = 2 \\ y^{(k+1)} = \frac{18}{11}y^{(k)} - \frac{9}{11}y^{(k-1)} + \frac{2}{11}y^{(k-2)} + \frac{6}{11}hf(t^{(k+1)}, y^{(k+1)}); r = 3 \\ y^{(k+1)} = \frac{48}{25}y^{(k)} - \frac{36}{25}y^{(k-1)} + \frac{16}{25}y^{(k-2)} - \frac{3}{25}y^{(k-3)} + \frac{12}{25}hf(t^{(k+1)}, y^{(k+1)}); r = 4 \end{cases}$$

Методы Гира представляют собой семейство r -шаговых неявных методов, требующих для вычисления $y^{(k+1)}$ значений в r предыдущих точках, а именно $y^{(k)}, \dots, y^{(k-r+1)}$. Для реализации алгоритма необходимо на каждом шаге искать решение, в общем случае, системы нелинейных уравнений с помощью итерационных методов.



Построение вычислительного процесса численного решения методом Гира порядка r , как и всяким многшаговым методом, требует предварительного вычисления значений искомой функции в предыдущих точках $y^{(1)}, \dots, y^{(r-1)}$. Это можно осуществить подходящим одношаговым методом, например Рунге-Кутты.

Основные численные решатели ОДУ в среде Matlab.

В этом разделе обобщенное название solver (решатель) означает один из возможных численных методов решения ОДУ: например, ode45, ode23, ode113, ode15s, ode23s, ode23t , ode23tb, bvp4c и прочие.

Решатели реализуют основные численные методы решения систем дифференциальных уравнений. Причем для решения жестких систем уравнений рекомендуется использовать только специальные решатели.

ode45 — одношаговые явные методы Рунге-Кутты 4-го и 5-го порядка. Это классический метод, рекомендуемый для начальной пробы решения. Во многих случаях он дает хорошие результаты;

ode23 — одношаговые явные методы Рунге-Кутты 2-го и 4-го порядка. При умеренной жесткости системы ОДУ и низких требованиях к точности этот метод может дать выигрыш в скорости решения;

ode113 — многошаговый метод Адамса – Башворта - Мултона переменного порядка. Это адаптивный метод, который может обеспечить высокую точность решения

ode23tb — неявный метод Рунге-Кутты в начале решения и метод, использующий формулы обратного дифференцирования 2-го порядка в последующих шагах. Несмотря на сравнительно низкую точность, этот метод может оказаться более эффективным, чем ode15s;

ode15s — многошаговый метод переменного порядка (от 1 до 5, по умолчанию 5), использующий формулы численного дифференцирования. Это адаптивный метод, и его стоит применять, если решатель ode45 не обеспечивает решения;

ode23s — одношаговый метод, использующий модифицированную формулу Розенброка 2-го порядка. Может обеспечить высокую скорость вычислений при низкой точности решения жесткой системы дифференциальных уравнений;

ode23t — метод трапеций с интерполяцией. Этот метод дает хорошие результаты при решении задач, описывающих колебательные системы с почти гармоническим выходным сигналом;

bvp4c - служит для проблемы граничных значений систем дифференциальных уравнений (краевой задачи).