левых элементов. Обычно эти матрицы имеют так называемую *ленточную структуру*. Более точно, матрицу A называют (2q + 1)-диагональной или имеющей *ленточную структуру*, если $a_{ij} = 0$ при |i - j| > q. Число 2q+1 называют *шириной ленты*. Оказывается, что при решении системы уравнений с ленточной матрицей методом Гаусса число арифметических операций и требуемый объем памяти ЭВМ могут быть существенно сокращены.

Задача 1. Исследовать характеристики метода Гаусса и метода решения системы с помощью разложения ленточной матрицы A на произведение левой и правой треугольных матриц. Показать, что для нахождения решения требуется O($mq2$) арифметических операций (при $m, q \to \infty$). Найти главный член числа операций при условии $1 \ll q \ll m$.

Задача 2. Оценить объем загружаемой памяти **ЭВМ** в методе Гаусса для ленточных матриц.

При вычислениях без помощи ЭВМ велика вероятность случайных погрешностей. Для устранения таких погрешностей иногда вводят *контрольный столбец системы* $a_{m+2} = (a_1,\ _{m+2}, ..., a_m,\ _{m+2})^T$, состоящий из контрольных эле- ментов уравнений системы

$$a_{i,m+2} = \sum_{j=1}^{k+1} a_{ij}.$$

При преобразовании уравнений над контрольными элементами производятся те же операции, что и над свободными членами уравнений. В результате этого контрольный элемент каждого нового уравнения должен равняться сумме коэффициентов этого уравнения. Большое расхождение между ними указывает на погрешности в вычислениях или на неустойчивость алгоритма вычислений по отношению к вычислительной погрешности.

К примеру, в случае приведения системы уравнений $A\mathbf{x} = \mathbf{b}$ к виду $D\mathbf{x} = \mathbf{d}$ с помощью формул (4) контрольный элемент $d_{i,\ m+2}$ каждого из уравнений системы $D\mathbf{x} = \mathbf{d}$ вычисляется по тем же формулам (4). После вычисления всех элементов d_{ij} при фиксированном i контроль осуществляется проверкой равенства

$$\sum_{j=i}^{m+1} d_{ij} = d_{i,\ m+2}.$$

Обратный ход метода Гаусса также сопровождается вычислением контрольных элементов строк системы.

Чтобы избежать катастрофического влияния вычислительной погрешности, применяют метод Гаусса с выбором главного элемента. Его отличие от описанной выше схемы метода Гаусса состоит в следующем. Пусть по ходу исключения неизвестных получена система уравнений

$$x_i + \sum_{j=i+1}^{m} a_{ij}^i x_j = a_{i,\ m+1}^i, \quad i = 1, ..., k,$$

$$\sum_{j=k+1}^{m} a_{ij}^k x_j = a_i^k,\ _{m+1},\quad i = k+1, ..., m,$$

Найдем l такое, что $a_{k+1,l}^k = \max a_{k+1,j}^k$ и переобозначим $x_{k+1} = x_l$ и $x_l = x_{k+1}$; далее произведем исключение неизвестной x_{k+1} из всех урав- нений, начиная с $(k+2)$-го. Такое переобозначение приводит к изменению порядка исключения неизвестных и во многих случаях существенно уменьшает чувствительность решения к погрешностям округления при вычислениях. Часто требуется решить несколько систем уравнений $A\mathbf{x} = \mathbf{b}_q, q = 1, ..., p$, с одной и той же матрицей A. Удобно поступить следующим образом: введя обозначения

$$\mathbf{b}_q = (a_{1,m+q}\ , ..., a_{m,m+q}\ )^T,\ q = 1, ..., p.$$

произведем вычисления по формулам (4), причем элементы dik вычислим при i < k m + p. В результате будут получены p систем уравнений с треугольной матрицей, соответствующих исходной задаче

$$D\mathbf{x} = \mathbf{d}_q, \mathbf{d}_q = (d1_{,m+q}\ , ..., d_{m,m+q}\ )^T, q = 1, ..., p.$$

Решаем эти системы каждую в отдельности. Оказывается, что общее чис- ло арифметических действий при решении p систем уравнений таким способом
$N \sim 2m^3/3 + 2pm^2$

Описанный выше прием иногда используется для того, чтобы без существен- ных дополнительных затрат получить суждение о погрешности решения, яв- ляющейся следствием погрешностей округления при вычислениях. Задаются вектором z с компонентами, имеющими по возможности тот же порядок и знак, что и компоненты искомого решения; часто из-за отсутствия достаточной информации берут $z = (1, ..., 1)^T$. Вычисляется вектор $\mathbf{c} = A\mathbf{z}$, и наряду с исходной системой уравнений решается система $\mathbf{A}z = \mathbf{c}$.
Пусть $\mathbf{x}'$ и **z'** — реально получаемые решения этих систем. Суждение о погрешности $\mathbf{x}' - \mathbf{x}$ искомого решения можно получить, основываясь на гипотезе: относительные погрешности при решении методом исключения систем с одной и той же матрицей и различными правыми частями, которыми являются соответственно величины $\|\mathbf{x} - \mathbf{x}'\| / \|\mathbf{x}'\|$ и $\|\mathbf{z} - \mathbf{z}'\| / \|\mathbf{z}'\|$, отличаются не в очень большое число раз.

Другой прием для получения суждения о реальной величине погрешности, возникающей за счет округлений при вычислениях, состоит в *изменении масштабов*, меняющем картину накопления вычислительной погрешности. Наряду с исходной системой тем же методом решается система

$$(\alpha A)\mathbf{x}' = \beta\mathbf{b}, \text{ где } \alpha \text{ и } \beta - \text{ числа.}$$

При α и β, не являющихся целыми степенями двойки, сравнение векторов $\mathbf{x}$ и $\alpha\beta^{-1}\mathbf{x}'$ дает представление о величине вычислительной погрешности. Например,можно взять $\alpha = \sqrt{2}$, $\beta = \sqrt{3}$.

Изучение многих задач приводит к необходимости решения систем ли- нейных уравнений с симметричной положительно определенной матрицей. Такие системы

возникают, например, при решении дифференциальных уравнений методом конечных элементов или же конечно-разностными ме- тодами. В этих случаях матрица системы имеет также и ленточную структуру.

Для решения таких систем, а также систем уравнений более обще- го вида с эрмитовой не обязательно положительно определенной матрицей применяется *метод квадратного корня (метод Холецкого)*. Матрица A представляется в виде

$$A = S^* DS,$$

где S—правая треугольная матрица, S^* - сопряженная с ней, т.е.

$$S = \begin{pmatrix} s_{11} & s_{12} & ... \\ 0 & s_{22} & ... \\ ... & ... & ... \end{pmatrix},$$

причем все $s_{ii} > 0$, D — диагональная матрица с элементами d_{ii}, равными $+1$ или 1. Матричное равенство (6) образует систему уравнений

$$a_{ij} = \sum_{k=1}^{i} \bar{s_{ki}} s_{kj} d_{kk} = \bar{s_{1i}} s_{1j} d_{11} + ... + \bar{s_{ii}} s_{ij} d_{ii} \quad \text{при} \quad i \le j.$$

Аналогичные уравнения при $i > j$ отброшены, так как уравнения, соот- ветствующие парам (i, j) и (j, i), эквивалентны. Отсюда получаем рекуррентные формулы для определения элементов d_{ii} и s_{ij} :

$$d_{ii} = sign\left(a_{ii} \sum_{k=1}^{i-1} s_k i^2 d_{kk}\right), \quad s_{ii} = \sqrt{a_{ii} - \sum_{k=1}^{i-1} s_k i^2 d_{kk}},$$

$$s_{ij} = \frac{a_{ij} - \sum\limits_{k=1}^{i-1} \bar{s_{ki}} s_{kj} d_{kk}}{s_{ii} d_{ii}} \quad \text{при } i < j.$$

Матрица S является правой треугольной, и, таким образом, после по- лучения представления (6) решение исходной системы также сводится к последовательному решению двух систем с треугольными матрицами. Заметим, что в случае $A > 0$ все $d_{ii} = 1$ и $A = S^* S$.

Задача 3. Оценить число арифметических операций и загрузку памяти ЭВМ (при условии $a_{ij} = a_{ji}$ объем памяти, требуемый для запоминания матрицы A, уменьшается) при решении системы с вещественной положи- тельно определеннной матрицей A методом квадратного корня.

Многие пакеты прикладных программ для решения краевых задач ма- тематической физики методом конечных элементов организованы по сле- дующей схеме. После формирования матрицы системы A путем переста- новки строк и столбцов (одновременно переставляются i-я и j-я строки и i-й и j-й столбцы) система преобразуется к виду с наименьшей шириной ленты. Далее применяется метод квадратного корня. При этом с целью уменьшения объема вычислений при решении системы $A\mathbf{x} = \mathbf{b}$ с другими правыми частями матрица S запоминается.

Замечание. Часто этот метод уступает по эффективности итерационным методам.

Задача 4. Оценить число арифметических операций и объем требуемой памяти метода квадратного корня в случае матриц ленточной структуры.

Если есть подозрение, что реально полученное решение x^1 сильно искажено вычислительной погрешностью, то можно поступить следующим образом. Определим вектор $\mathbf{b}^1 = \mathbf{b} - A\mathbf{x}^1$. Погрешность $\mathbf{r}^1 = \mathbf{x} - \mathbf{x}^1$ удо- влетворяет системе уравнений

$$A\mathbf{r}^1 = A\mathbf{x} - A\mathbf{x}^1 = \mathbf{b}^1.$$

Решая эту систему в условиях реальных округлений, получаем прибли- жение $\mathbf{r}^{(1)}$ к $\mathbf{r}^1$. Полагаем $\mathbf{x}^2 = \mathbf{x}^1 + \mathbf{r}^{(1)}$. Если точность нового прибли- жения представляется неудовлетворительной, то повторяем эту операцию. При решении системы (7) над компонентами правой части производятся те же линейные операции, что и над компонентами правой части при решении системы (1). Поэтому при вычислениях на ЭВМ с плавающей запятой естественно ожидать, что относительные погрешности решений этих систем будут одного порядка. Поскольку погрешности округлений обычно малы, то $\|\mathbf{b}^1\| \ll \|\mathbf{b}\|$ тогда $\|\mathbf{r}^1\| \ll \|\mathbf{x}^1\|$, и, как правило, решение (7) определится с существенно меньшей абсолютной погрешностью, чем решение системы (1). Таким образом, применение описанного прие- ма приводит к повышению точности приближенного решения.

Особенно удобно применять этот прием, когда по ходу вычислений в памяти ЭВМ сохраняются матрицы B и D. Тогда для каждого уточнения требуется найти вектор $\mathbf{b}^k = \mathbf{b} - A\mathbf{x}^k$ и решить две системы с треугольными матрицами. Это потребует всего $N_1 \sim 4m^2$арифметических операций, что составит малую долю от числа операций $N_0 \sim 2m^3/3$, требующихся для представления матрицы A в виде $A = BD$.

Идея описанного приема последовательного уточнения приближений к реше- нию часто реализуется в такой форме. Пусть матрица B близка в каком-то смысле к матрице A, но решение системы $B\mathbf{x} = \mathbf{c}$ требует существенно мень- шего объема вычислений по сравнению с решением системы $A\mathbf{x} = \mathbf{b}$. Решение системы $B\mathbf{x} = \mathbf{b}$ принимаем в качестве первого приближения $\mathbf{x}^1$ к решению. Разность x x1 удовлетворяет системе уравнений

$$A(\mathbf{x} - \mathbf{x}') = \mathbf{b} - A\mathbf{x}^1.$$

Вместо решения этой системы находим решение системы

$$B\mathbf{r}^1 = \mathbf{b}A\mathbf{x}'$$

и полагаем $\mathbf{x} = \mathbf{x}^1 + \mathbf{r}^1$. Таким образом, каждое приближение находится из предыдущего по формуле

$$\mathbf{x}^{n+1} = \mathbf{x}^n + B^{-1}(\mathbf{b} - A\mathbf{x}^n) = (E - B^{-1}A)\mathbf{x}^n + B^{-1}\mathbf{b}.$$

Если матрицы A и B достаточно близки, то матрица $E - B^{-1}A$имеет малую норму и такой итерационный процесс быстро сходится (см. также §10).

Значительно более редкой, чем задача решения системы уравнений, является задача обращения матриц. Для обратной матрицы$X = A^{-1}$ имеем равенство $AX = BDX = E$. Таким образом, для нахождения матрицы X достаточно последовательно решить две матричные систе- мы $BY = E$, $DX = Y$. Нетрудно подсчитать, что при нахождении на таком пути матрицы A^{-1} общий объем вычислений составит $N_2 \sim 2m^3$ арифметических операций. В случае необходимости уточнения приближения к обратной мат- рице могут производиться при помощи итерационного процесса $X_k = X_{k-1}(2E - AX_{k-1})$. Для исследования сходимости итерационного процесса рассмотрим матрицы $G_k = E - AX_k$. Имеем равенство

$$G_k = E - AX_k = E - AX_{k-1}(2E - AX_{k-1}) = (E - AX_{k-1})^2 = G_{k-1}^2.$$

Отсюда получаем цепочку равенств

$$G_k = G_{k-1}^2 = G_{k-2}^4 = ... = G_0^{2^k}.$$

Поскольку

$$A^{-1} - X_k = A^{-1}(E - AX_k) = A^{-1}G_k = A^{-1}G_0^{2^k},$$

то имеем оценку

$$\|A^{-1} - X_k\| \le \|A^{-1}\| \cdot \|G_0\|^{2^k}$$

Таким образом, при достаточно хорошем начальном приближении, т.е. если $\|E - AX_0\| \le 1$, этот итерационный процесс сходится со скоростью более быстрой, чем геометрическая прогрессия.

§2. Метод отражений

В настоящее время разработано так много точных методов численного решения систем линейных алгебраических уравнений, что даже простое перечисление их затруднительно. Большинство этих методов, как и метод исключения Гаусса, основано на переходе от заданной системы $A\mathbf{x} = \mathbf{b}$ к новой системе $CA\mathbf{x} = C\mathbf{b}$ такой, что система $Bx = \mathbf{d}$, где $B = CA$ и $\mathbf{d} = C\mathbf{b}$, решается проще, чем исходная. При выборе подходящей матрицы C нужно учитывать по крайней мере следующие два фактора. Во-первых, ее вычисление не должно быть чересчур сложным и трудоемким. Во- вторых, умножение на матрицу C не должно в каком-то смысле портить матрицу A (мера обусловленности матрицы не должна меняться сильно (см. §11)).

Этим условиям в определенной степени удовлетворяет описываемый ниже *метод отражений*. Среди методов, требующих для своей реализа- ции $N \sim 4m^3/3$ операций, этот метод в настоящее время рассматривается как один из наиболее устойчивых к вычислительной погрешности. Среди методов, требующих для своей реализации $N \sim 4m^3/3$ операций, как наи- более устойчивый к вычислительной погрешности рассматривается *метод вращений*.

Рассмотрим случай вещественной матрицы A. Если $\mathbf{w}$ — некоторый вектор-столбец единичной длины, $(\mathbf{w}, \mathbf{w}) = 1$, то матрицу

$$U = E - 2\mathbf{w}\mathbf{w}^T$$

называют *матрицей отражений*. Под $\mathbf{w}\mathbf{w}^T$ здесь понимается матрица, являющаяся произведением вектора-столбца $\mathbf{w}$ на вектор-строку $\mathbf{w}^T$, т.е. $\mathbf{w}\mathbf{w}^T = (w_{ij})$, где

$w_{ij} = w_i w_j$. Из определения следует, что $\mathbf{w}\mathbf{w}^T$ — сим- метричная матрица.

Непосредственной проверкой убеждаемся, что $U = U^T$ и
$UU^T = (E - 2\mathbf{w}\mathbf{w}^T)^T(E - 2\mathbf{w}\mathbf{w}^T)^T = E - 2\mathbf{w}\mathbf{w}^T - 2\mathbf{w}\mathbf{w}^T + 4\mathbf{w}\mathbf{w}^T\mathbf{w}\mathbf{w}^T = E$;
здесь мы воспользовались тем, что

$$\mathbf{w}^T\mathbf{w} = (\mathbf{w}, \mathbf{w}) = 1.$$

Таким образом, матрица U —симметричная и ортогональная.

> Напомним один факт из алгебры. Пусть U и B — две матрицы порядка m, B — многочлен от U, $B = P_l(U)$. Тогда можно переупорядочить их собственные BU значения так, что $\lambda_j^B = P_l(\lambda_j^U)$ при $j = 1, ..., m$.

Поскольку U симметрична и $U^2 = UU^T = E$, а все собственные числа E равны 1, то все собственные числа матрицы U удовлетворяют условию $\lambda)U^2 = 1$, т.е. равны или $+1$ или -1.

Собственному значению 1 отвечает собственный вектор w. В самом деле,

$$U\mathbf{w} = \mathbf{w} - 2\mathbf{w}\mathbf{w}^T\mathbf{w} = \mathbf{w} - 2\mathbf{w} = -\mathbf{w}.$$

Все векторы, ортогональные вектору w, являются собственными. Им соответствует собственное значение, равное $+1$. Действительно, пусть $(v, w) = 0$. Тогда имеем

$$U\mathbf{v} = \mathbf{v} - 2\mathbf{w}\mathbf{w}^T\mathbf{v} = \mathbf{v} - 2\mathbf{w}(\mathbf{w}, \mathbf{v}) = \mathbf{v}.$$

Представим произвольный вектор y в виде $\mathbf{y} = \mathbf{z}+\mathbf{v}$, где $\mathbf{z} = \gamma\mathbf{w}$, $(\mathbf{v}, \mathbf{w}) = 0$. Для этого следует взять в качестве $\mathbf{z}$ проекцию вектора $\mathbf{y}$ на вектор $\mathbf{w}$, т.е. $\mathbf{z} = (\mathbf{y}, \mathbf{w})\mathbf{w}$, и $\mathbf{v} = \mathbf{y}(\mathbf{y}, \mathbf{w})\mathbf{w}$. Вследствие (2) и (3) имеем $U_\mathbf{y} = \mathbf{z} + \mathbf{v}$. Таким образом, $U_\mathbf{y}$ есть зеркальное отражение вектора $\mathbf{y}$ относительно гиперплоскости, ортогональной вектору $\mathbf{w}$. Используя геометрические свойства матрицы отражений, нетрудно решить следующую задачу: подобрать вектор $\mathbf{w}$ в матрице отражений так, чтобы заданный вектор $y \neq 0$ имел в результате преобразования $U_\mathbf{y}$ матрицей отражения $U = E2ww^T$направление заданного единичного вектора $\mathbf{e}$.

Так как U — ортогональная матрица, а при ортогональных преобразо- ваниях длины векторов сохраняются, то мы должны иметь $U_\mathbf{y} = \alpha\mathbf{e}$ или $U_\mathbf{y} = \alpha - \mathbf{e}$, где $\alpha = (y, y)$. Поэтому направление, перпендикулярное плоскости отражения, будет определяться либо вектором $\mathbf{y}\alpha\mathbf{e}$, либо вектором $\mathbf{y} + \alpha\mathbf{e}$ (см. рис. 6.2.1).

Таким образом, векторы $\mathbf{w}_1 = \pm\rho_1^{-1}(\mathbf{y} - \alpha\mathbf{e})$ или $\mathbf{w}_2 = \pm\rho_2^{-1}(\mathbf{y} + \alpha\mathbf{e})$ где $\rho_1 = \sqrt{(\mathbf{y} - \alpha\mathbf{e}, y - \alpha\mathbf{e})}$, $\rho_2 = \sqrt{(\mathbf{y} + \alpha\mathbf{e}, y + \alpha\mathbf{e})}$, будут искомыми. Ясно, что данный процесс всегда осуществим. Если векторы $\mathbf{y}$ и $\mathbf{e}$ коллинеарны, а в этом случае либо ρ_1, либо ρ_2 будет равно нулю, то никаких отражений делать не надо

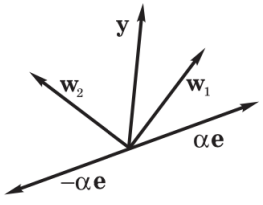

Рис. 6.2.1

Матрицы отражения нашли широкое применение при численном решении раз- личных задач линейной алгебры (в частности, в рассматриваемой нами задаче приведения матрицы системы уравнений к треугольному виду).

Лемма. *Произвольная квадратная матрица может быть представлена в виде произведения ортогональной и верхней треугольной матриц.*

Доказательство. Пусть дана квадратная матрица порядка m. Будем приводить ее к правой треугольной матрице путем последовательного умножения слева на ортогональные матрицы. На первом шаге приведения рассмотрим в качестве вектора $\mathbf{y}$ из предыдущего рассуждения первый столбец матрицы A:

$$\mathbf{y}_l = (a_{11}, ..., a_{m1})^T.$$

Если $a_{21} = a_{31} = ... = a_{m1} = 0$, то переходим к следующему шагу, положив $A^{(1)} = A, U_1 = E$ и введя обозначения $a_{ij}^{(1)} = a_{ij}$. В противном случае умножаем матрицу A слева на матрицу отражения $U_1 = E_m - 2\mathbf{w}_1\mathbf{w}_1^T$, где w_1 подбирается так, чтобы вектор $U_1\mathbf{y}_1$ был коллинеарен вектору $\mathbf{e}_1 = (1, 0, ..., 0)^T$. Здесь и далее $\mathbf{E}_q$ — единичная матрица размерности q.

На этом первый шаг закончен, и на следующем шаге будем рассмат- ривать матрицу $A^{(1)}$ с элементами $a_{ij}^{(1)}$, которая либо равна A, если имеет место первый случай, либо $A^{(1)} = U_1A$, если имеет место второй случай. Пусть мы уже осуществили $l1 > 0$шагов и пришли к матрице $A^{(l1)}$ с элементами $a_{ij}^{(l1)}$такими, что $a_{ij}^{(l1)} = 0$ при $i > j$, $j = 1, 2, ..., l1$. В пространстве R_{ml+1}векторов размерности $ml + 1$ рассмотрим вектор

$$\mathbf{y}_l = \left(a_{l,l}^{(l-1)}, a_{l+1,l}^{(l-1)}, ..., a_{m,l}^{(l-1)}\right)^T.$$

Если$a_{l+1,l}^{(l-1)} = a_{l+2,l}^{(l-1)} = ... = 0$, то переходим к следующему шагу, полагая $A^{(l)} = A^{(l1)}, U_l = E$. В противном случае строим матрицу отражения $V_l = E_{m-l+1} - 2\mathbf{w}_l\mathbf{w}_l^T$(размеры матрицы V_l и вектора $\mathbf{w}_l$ равны $m - l + 1$), переводящую вектор $\mathbf{y}_l$ в вектор, коллинеарный $\mathbf{e}_l = (1, 0, ..., 0)^T \in R_{m-l+1}$, и переходим к матрице

$$A^l = U_lA^{(l-1)};$$

здесь $U_l = \begin{pmatrix} E_{l-1} & 0 \\ 0 & V_l \end{pmatrix}$. Ясно, что процесс всегда осуществим, и после $(m-1)$-го шага мы приходим к матрице

$$A^{(m-1)} = U_{m-1}U_{m-2}...U_1A,$$

имеющей правую треугольную форму.

Если обозначить $U_{m-1}U_{m-2}...U_1 = U$, то из последнего равенства сле- дует, что $A = U^TA^{(m1)}$, где U^T — ортогональная, а $A^{(m-1)}$ — правая тре- угольная матрицы. Лемма доказана.

Вернемся к решению системы $A\mathbf{x} = \mathbf{b}$. С помощью указанных пре- образований отражения последовательно приводим ее к эквивалентному

$$A^{(m-1)}\mathbf{x} = U\mathbf{b},$$

где $A^{(m-1)}$ — правая треугольная матрица. Если все диагональные эле- менты $A^{(m-1)}$ отличны от нуля, то последовательно находим $x_x, ..., x_1$. Если же хотя бы один из диагональных элементов равен нулю, то по- следняя система вырождена и в силу эквивалентности вырождена и ис- ходная система.

Задача 1. Получить асимптотику числа операций метода отражений при $m \to \infty$

Рассмотрим случай системы уравнений $A\mathbf{x} = \mathbf{b}$ с комплексными A и $\mathbf{b}$. Пусть

$$A = A_1 + iA_2, \mathbf{b} = \mathbf{b}_1 + i\mathbf{b}_1, \mathbf{x} = \mathbf{x}_1 + i\mathbf{x}_2.$$

Исходная система уравнений равносильна системе

$$C\mathbf{y} = \mathbf{d}$$

с вещественными C и $\mathbf{d}$:

$$C = \begin{pmatrix} A_1 & -A_2 \\ A_2 & A_1 \end{pmatrix}, \quad \mathbf{d} = \begin{pmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \end{pmatrix}, \quad \mathbf{y} = \begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{pmatrix}.$$

Поэтому вместо непосредственного решения исходной задачи можно пе- рейти к решению задачи (4) и применить для решения последней метод отражений.

Однако возможен и другой путь—непосредственное применение мето- да отражений к исходной системе $A\mathbf{x} = \mathbf{b}$. Здесь матрица отражения $U = E - 2\mathbf{w}\mathbf{w}^*, w^* = (\bar{w}_1, ..., \bar{w}_m)^T$ будет унитарной с собственными зна- чениями вида $\lambda_U = e^{i\phi}$. (Через z обозначено комплексное число, сопря- женное с z).

Задача 2. Перенести метод отражений на случай комплексных матриц.

Задача 3. Исследовать метод отражений в случае его применения для реше- ния систем уравнений с ленточной матрицей.

§3 Метод простой итерации

Простейшим итерационным методом решения систем линейных уравнений яв- ляется *метод простой итерации.* Система уравнений

$$A\mathbf{x} = \mathbf{b} \tag{1}$$

преобразуется к виду

$$\mathbf{x} = B\mathbf{x} + c, \tag{2}$$

и ее решение находится как предел последовательности

$$\mathbf{x}^{n+1} = B\mathbf{x}^n + \mathbf{c}. \tag{3}$$

Всякая система

$$\mathbf{x} = \mathbf{x} + D(A\mathbf{x} - \mathbf{b}) \tag{4}$$

имеет вид (2) и при $\det D \neq 0$ эквивалентна системе (1). В то же время вся- кая система (2), эквивалентная (1), записывается в виде (4) с матри- цей $D = (E - B)A^{-1}$.

Теорема (о достаточном условии сходимости метода простой итерации). Если

$\|B\| < 1$, *то система уравнений* (2) *имеет единственное решение и итерационный процесс* (3) *сходится к решению со скоростью геометриче- ской прогрессии. Доказательство.* Для всякого решения системы (2) имеет место $\|\mathbf{x}\| \leqslant \|B\|\|\mathbf{x}\| + \|\mathbf{c}\|$, поэтому справедливо неравенство $\|\mathbf{x}\|(1-\|B\|) \leqslant \|\mathbf{c}\|$ или $\|\mathbf{x}\| \leqslant (1-\|B\|)^{-1}\|\mathbf{c}\|$. Отсюда следует существование и единственность решения однородной системы $\mathbf{x} = B\mathbf{x}$, а следовательно, и системы (2). Пусть $\mathbf{X}$—решение системы (2). Из (2) и (3) получаем уравнение отно- сительно погрешности $\mathbf{r}^n = \mathbf{x}^n - \mathbf{X}$:

$$\mathbf{r}^{n+1} = B\mathbf{r}^n. \tag{5}$$

Из (5) получаем равенство

$$\mathbf{r}^n = B^n\mathbf{r}^0. \tag{6}$$

Отсюда следует, что $\|\mathbf{r}^n\| \leqslant \|B\|^n\|\mathbf{r}^0\| \to 0$. Теорема доказана.

Качество итерационного процесса удобно характеризовать скоростью убывания отношения погрешности после n итераций к начальной погреш- ности:

$$s_n = \sup_{\mathbf{x}^0 \neq \mathbf{X}} \frac{\|\mathbf{r}^n\|}{\|\mathbf{r}^0\|} = \sup_{\mathbf{r}^0 \neq 0} \frac{\|B\mathbf{r}^0\|}{\|\mathbf{r}^0\|} = \|B^n\|.$$

Можно гарантировать, что величина $s_n \leqslant \varepsilon$, если $\|B\|^n \leqslant \varepsilon$, т. е. при

$$n \geqslant n_\varepsilon = \ln(\varepsilon^{-1})/\ln(\|B\|^{-1}). \tag{7}$$

Если существуют постоянные $\gamma_{\alpha\beta}$, $\gamma_{\beta\alpha}$ такие, что при $\mathbf{x} \neq 0$

$$\|\mathbf{x}\|_\beta/\|\mathbf{x}\|_\alpha \leqslant \gamma_{\alpha\beta}, \|\mathbf{x}\|_\alpha/\|\mathbf{x}\|_\beta \leqslant \gamma_{\beta\alpha},$$

то нормы $\|\mathbf{x}\|_\alpha$ и $\|\mathbf{x}\|_\beta$называются *эквивалентными.* Имеем

$$\|\mathbf{r}^n_\beta\| \leqslant \gamma_{\alpha\beta}\|\mathbf{r}^n\|_\alpha \leqslant \gamma_{\alpha\beta}\|B\|^n_\alpha\|\mathbf{r}^0\| \leqslant \gamma_{\alpha\beta}\gamma_{\beta\alpha}\|B\|^n_\alpha\|\mathbf{r}^0\|_\beta.$$

Таким образом, если условие доказанной теоремы выполнено для нормы $\|\cdot\|_\alpha$, то утверждение справедливо относительно любой эквивалентной ей нормы.

Любые две нормы в конечномерном пространстве являются эквива- лентными. В частности, нормы $\|\mathbf{x}_1\|$, $\|\mathbf{x}_2\|$, $\|\mathbf{x}_3\|$, вычисляемые соответ- ственно по формулам (2), (3), (4), приведенным во введении к настоя- щей главе, эквивалентны между собой вследствие справедливости цепоч- ки неравенств

$$\|\mathbf{x}_\infty\| \leqslant \|\mathbf{x}\|_2 \leqslant \|\mathbf{x}\|_1 \leqslant m\|\mathbf{x}\|_\infty.$$

Лемма. *Пусть все собственные значения λ_i матрицы B лежат в круге $|\lambda| \leqslant q$, причем собственным значениям, по модулю равным q, соответ- ствуют жордановы клетки размерности 1. Тогда существует матрица $\Lambda = D^{-1}BD$ с нормой $\|\Lambda\|_\infty \leqslant q$.*

Доказательство. Положим $\eta = q - \max\limits_{|\lambda_i|<q} \mid \lambda_i \mid$. Собственными значениями матрицы $\eta^{-1}B$ будут $\eta^{-1}\lambda_i$. Преобразуем матрицу $\eta^{-1}B$ к жордановой форме

$$D^{-1}(\eta^{-1}B)D = \begin{pmatrix} \eta^{-1}\lambda_1 & \alpha_{12} & 0 & ... \\ 0 & \eta^{-1}\lambda_2 & \alpha_2 & ... \\ . & . & . & ... \end{pmatrix},$$

где $\alpha_{i,i+1}$ принимают значения 0 или 1. После умножения на η получим

$$\Lambda = D^{-1}BD = \begin{pmatrix} \lambda_1 & \alpha_{12}\eta & 0 & ... \\ 0 & \lambda_2 & \alpha_{23}\eta & ... \\ . & . & . & ... \end{pmatrix},$$

Если $\mid \lambda_i \mid = q$, , то согласно условиям леммы, $\alpha_{i,i+1} = 0$. Отсюда следует, что $\mid \lambda_i \mid + \mid \alpha_{i,i+1}\eta \mid = q$. Если $\mid \lambda_i < q \mid$, то

$$\mid \lambda_i \mid + \mid \alpha_{i,i+1}\eta \mid \leqslant \max_{|\lambda_i|<q} \mid \lambda_i \mid + \eta = q.$$

Таким образом, $\|\Lambda\|_\infty = \max\limits_i(\mid \lambda_i \mid + \mid \alpha_{i,i+1}\eta \mid) \leqslant q$.

Теорема (о необходимом и достаточном условии сходимости метода про- стой итерации). *Пусть система* (2) *имеет единственное решение. Итерационный процесс* (3) *сходится к решению системы* (2) *при любом начальном приближении тогда и только тогда, когда все собственные значения матрицы B по модулю меньше 1.*

Доказательство. Достаточность. Возьмем произвольное q в пределах $\max\limits_i \mid \lambda_i \mid < q < 1$. Условие леммы выполнено по отношению к этому q, поэтому существует матрица D такая, что $\|\Lambda\|_\infty \leqslant q$ при $\Lambda = D^{-1}BD$. Поскольку $B = D\Lambda D^{-1}$, то

$$B^n = D\Lambda D^{-1}D...D^{-1}D\Lambda D^{-1} = D\Lambda^n D^{-1}.$$

Поэтому

$$\|B^n\|_\infty \leqslant \|D\|_\infty \|D^{-1}\|_\infty q^n \to 0$$

и

$$\|\mathbf{x}^n - \mathbf{X}\|_\infty \leqslant \|D\|_\infty \|D^{-1}\|_\infty q^n \|\mathbf{x}^0 - \mathbf{X}\|_\infty \to 0 \tag{8}$$

при $n \to \infty$. Следовательно, и $\|\mathbf{x}^n - \mathbf{X}\|_1, \|\mathbf{x}^n - \mathbf{X}\|_2 \to 0$.

Если χ_i — координатные орты, $\mathbf{x} = (x_1, ..., x_m)^T$, то $\mathbf{x} = \sum\limits_i x_i\chi_i$ некоторая норма, тогда

$$\|\mathbf{x}\| \leqslant \sum_i \mid x_i \mid \|\chi_i\| \leqslant \|\mathbf{x}\|_\infty \sum_i \|\chi_i\|.$$

Поэтому при любой норме $\|\cdot\|$ имеем

$$\|\mathbf{x}^n - \mathbf{X}\| \leqslant \left(\sum_i \|\chi_i\|\right) \|D\|_\infty \|D^{-1}\|_\infty q^n \|\mathbf{x}^0 - \mathbf{X}\|_\infty \to 0 \tag{9}$$

Соотношения (8), (9) означают также, что любые нормы погрешности убывают быстрее любой геометрической прогрессии со знаменателем, большим $\max\limits_i \mid \lambda_i \mid$.

Необходимость. Пусть $\mid \lambda_i \mid \geqslant 1$ и $\mathbf{e}_1$ — соответствующий собственный вектор матрицы B. Тогда при начальном приближении $\mathbf{x}^0 = \mathbf{X} + c\mathbf{e}^1, c \neq 0$ имеем

$$\mathbf{r}^0 = c\mathbf{e}_1 \quad \text{и} \quad \mathbf{r}^n = \lambda_l^n c\mathbf{e}_1 \nrightarrow 0 \quad \text{при} \quad n \to \infty.$$

Задача 1. Пусть все собственные значения матрицы B, за исключением простого $\lambda_1 = 1$, лежат внутри единичного круга и система (2) имеет решение X. Решением системы будут также все $\mathbf{x} = \mathbf{X} + c\mathbf{e}_1$. Доказать, что итерационный процесс (3) сходится к одному из таких решений.

§4. Особенности реализации метода простой итерации на ЭВМ

Если все собственные значения матрицы B лежат внутри единичного круга, то может показаться, что не возникает никаких проблем отно- сительно поведения метода в реальных условиях ограниченности поряд- ков чисел в ЭВМ и присутствия округлений. В обоснование этого ино- гда приводят следующий довод: возмущения приближений в результате округлений равносильны возмущениям начальных условий итерационно- го процесса. Поскольку процесс сходящийся, «самоисправляющийся», эти возмущения в конце концов затухнут, и будет получено хорошее прибли- жение к решению исходной задачи.

Однако при решении некоторых систем возникала следующая ситуа- ция. Все собственные значения матрицы B лежали в круге $\mid \lambda \mid \leqslant 1/2$, а итерационный процесс останавливался после некоторого числа итера- ций из-за переполнения порядков чисел в ЭВМ. В других случаях та- кого переполнения не происходило, но векторы $\mathbf{x}^n$, получаемые при вы- числениях, не сходились к решению. Последний случай особенно опасен по следующей причине. Можно необоснованно решить, что при условии $\mid \lambda_i \mid \leqslant 1/2$ какое-то определенное число итераций, например 100, заве- домо достаточно для получения решения с требуемой точностью. Затем производим эти 100 итераций и рассматриваем полученный результат как требуемый. Поэтому наличие подобных явлений послужило толчком к бо- лее детальному исследованию итерационных процессов и формированию новых понятий в теории операторов.

Чтобы понять сущность явления, полезно построить пример, где это явление прослеживается в явном виде. В качестве модели выберем ите- рационный процесс, соответствующий двухдиагональной матрице

$$B_0 = \begin{pmatrix} \alpha & \beta & 0 & \ldots & 0 \\ 0 & \alpha & \beta & \ldots & 0 \\ . & . & . & \ldots & . \\ 0 & 0 & 0 & \ldots & \alpha \end{pmatrix}.$$

При возведении матрицы $B0$ в степень n, получается треугольная мат- рица

$$B_0^n = (b_{ij}^{(n)} = \begin{pmatrix} \alpha^n & C_n^1\alpha^{n-1}\beta & C_n^2\alpha^{n-2}\beta^2 & \ldots \\ 0 & \alpha^n & C_n^1\alpha^{n-1}\beta & \ldots \\ . & . & . & \ldots \end{pmatrix}$$

с элементами $b_{ij}^{(n)} = C_n^{j-i}\alpha^{n-(j-i)}\beta^{j-i}$. Если $\mathbf{r}^0 = (0,...,0,1)^T$, то

$$\mathbf{r}^n = B_0^n\mathbf{r}^0 = (b_{1m}^{(n)},...,b_{mm}^{(n)})^T, \quad \|\mathbf{r}^n\|_1 = \sum_{i=1}^{m} \mid b_{im}^{(n)} \mid .$$

При $n < m$ последнее выражение упрощается:

$$\|\mathbf{r}^n\|_1 = \sum_{i=1}^{m} C_n^{m-i} \mid \alpha \mid^{n-(m-i)} \mid \beta \mid^{m-i} =$$
$$= \sum_{k=0}^{m-1} C_n^k \mid \alpha \mid^{n-k)} \mid \beta \mid^k = \sum_{k=0}^{n} C_n^k \mid \alpha \mid^{n-k)} \mid \beta \mid^k = (|\alpha| + \mid \beta \mid)^n .$$

Рассмотрим случай $|\alpha| < 1, |\alpha|+|\beta| > 1, |\beta/(1-\alpha)| < 1$. Пусть $\mathbf{c} = \mathbf{c}^0 = (0,...,0,1)^T$. Непосредственно проверяется, что при таком с решением рассматриваемой системы будет

$$\mathbf{X}^0 = \left(\frac{1}{1-\alpha}\right)\left(\frac{\beta}{1-\alpha}\right)^{m-1}, ..., \frac{1}{1-\alpha})^T.$$

Справедлива оценка

$$\|\mathbf{X}^0\|_1 \leqslant \omega,$$

где

$$\omega = \frac{1}{|1-\alpha|}\sum_{k=0}^{\infty}\left|\frac{\beta}{1-\alpha}\right|^k = \frac{1}{|1-\alpha|\left(1-\left|\frac{\beta}{1-\alpha}\right|\right)}.$$

При начальном приближении $\mathbf{x}^0 = \mathbf{X}^0 + \mathbf{c}^0$ имеем $\mathbf{r}^0 = \mathbf{c}^0$ и, согласно проводившимся выше построениям,

$$\|\mathbf{r}^n\|_1 = (|\alpha| + |\beta|)^n \quad \text{для} \quad n < m.$$

Выберем m таким, чтобы число $\sigma = [(|\alpha|+|\beta|)^{m-1} - \omega]/m$ превосходило пределы, допустимые в ЭВМ. Из полученных ранее соотношений следует, что

$$\|\mathbf{x}^{m-1}\|_\infty \geqslant \|\mathbf{x}^{m-1}\|_{1/m} \geqslant \left(\|\mathbf{r}^{m-1}\|_1 - \|\mathbf{x}^0\|_1\right)/m \geqslant \sigma.$$

Поэтому построенный пример обладает следующими свойствами: норма начального приближения невелика, итерационный процесс сходится при отсутствии округлений и ограничения на порядки чисел в ЭВМ, но оста- навливается не позднее чем при $n = m1$ из-за недопустимо больших значений компонент приближений.

Обратимся к реальной ситуации, когда на каждом шаге вычислений происходят округления. Рассмотрим подробнее случай, когда переполне- ние не происходит. Вместо $\mathbf{x}^n$ получаются векторы $\mathbf{x}^{*n}$, связанные соот- ношениями

$$\mathbf{x}^{*n+1} = B\mathbf{x}^{*n} + \mathbf{c} + \rho^n,$$

где ρ^nсуммарное округление на шаге итерации.

Отсюда и из (3.2) получается уравнение относительно погрешности $\mathbf{r}^n = \mathbf{x}^n - \mathbf{X}$:

$$\mathbf{r}^n = \rho^n + B\mathbf{r}^n. \tag{1}$$

Выражая каждое $\mathbf{r}^{*n}$ через предыдущее, получаем

$$\mathbf{r}^{*n} = \rho^n + B\mathbf{r}^{n-1} = \rho^n + B(\rho^{n-2} + B\mathbf{r}^{*n-2} =$$

$$= \rho^{n-1} + B\rho^{n-2} + ... + B^{n-1}\rho^0 + B^n\mathbf{r}^0. \qquad (2)$$

Как мы видели, норма $\|B_0^n$ при $|\alpha|+|\beta| > 1$ имеет следующий характер поведения: при малых n она имеет тенденцию к возрастанию, при больших n стремится к нулю. (Можно показать, что максимальное значение $\varphi(B_0) = \max\limits_n \|B_0^n\|$ достигается при значении $n = n_0$ порядка m.)
При таком характере поведения норм B^n может возникнуть следующая ситуация. Величина $\max\limits_n \|\mathbf{x}^{*n}\|$ не настолько велика, чтобы происходило
переполнение и остановка ЭВМ; в то же время $\varphi(B)2^{-t} \gg R$, где R — максимально допустимая погрешность решения. Поэтому, как правило, при $n > n_0$ среди слагаемых в правой части (2) присутствует слагаемое $B^{n_0}\rho^{n-1-n_0}$ с нормой, много большей, чем R. В результате установление приближений $\mathbf{x}^n$ с приемлемой точностью не происходит.

Подведем некоторый итог проведенных построений. Матрицы высокой размерности обладают свойствами, существенно отличными от свойств матриц малой размерности. Кроме собственных значений у таких матриц есть почти собственные значения, т.е. λ такие, что $\|A\mathbf{x} - \lambda\mathbf{x}\| \leqslant \varepsilon\|\mathbf{x}\|$ при $\|\mathbf{x}\| \neq 0$
и очень малом ε.

Например, в случае матрицы B_0 при любом λ_0, лежащем в круге $|\alpha - \lambda| < |\beta|$, можно построить вектор $\mathbf{x}_\lambda$ такой, что $\|B_0\mathbf{x}_\lambda - \lambda\mathbf{x}_\lambda\|_\infty \leqslant \varepsilon_\lambda\|\mathbf{x}_\lambda\|_\infty$, где $\varepsilon_\lambda = |\beta|\,|(\lambda - \alpha)/\beta|^m$. Поведение степеней матрицы B^n n при порядка m определяется во многом такими «почти собственными векторами» $\mathbf{x}\lambda$ и «почти собственными значени- ями» λ.

Задача 2. Построить «почти собственный вектор» $\mathbf{x}_\lambda$, соответствующий значению ε_λ, приведенному выше.

Суммарная вычислительная погрешность $\rho_n = \sum\limits_{j=0}^{n-1} B^{n-1-j}\rho^j$ может оказаться большой не только из-за большой величины отдельных слагаемых, но и из-за того, что их много. Пусть B — симметричная матрица и$\|B\|_2 = \max\limits_i \left|\lambda_B^i\right| = \lambda_B^1 < 1, \mathbf{e}^1$ соответствующий λ_B^1 нормированный собственный вектор. Предположим, что на каждом $j-$м шаге происходит округление $\rho^j = \rho\mathbf{e}^1$, где ρ порядка 2^{2-t}. Имеем равенство

$$\rho_n = \rho\sum_{j=0}^{n-1}(\lambda_B^1)^j\mathbf{e}^1$$

Поскольку число итераций берется таким, что $\|B^n\| \gg 1$,а $\|B^n\| = (\lambda_B^1)^n$ то можно считать, что $\|\rho_n\| \approx \rho/(1 - \lambda_B^1)$. Таким образом, если λ_B^1 близко к 1, то суммарное влияние округлений на шагах интегрирования может оказаться довольно большим. Покажем, что вычислительная погрешность такого порядка является неизбежной. Предположим, что вместо системы (3.2) решается система $\mathbf{X} = B\mathbf{X} + \mathbf{c} + \rho\mathbf{e}_1$. Разность $\mathbf{X} - \mathbf{x}$ решений этих систем удовлетворяет соотношению $(\mathbf{X} - \mathbf{x}) = B(\mathbf{X} - \mathbf{x}) + \rho\mathbf{e}_1$, отсюда $\mathbf{X} - \mathbf{x} = (E - B)^{-1}\rho\mathbf{e}_1$ Поэтому погрешность порядка $(1 - \lambda_B^1)^{-1}\rho$ является неустранимой; возмущение прибли- жений, создаваемое в ходе итераций, сравнимо с неустранимой погрешностью.

§5. δ^2-процесс практической оценки погрешности и ускорения сходимости

Рассмотрим вопрос об оценке погрешности приближенного решения си- стемы уравнений. Если $\mathbf{X}^*$ — приближенное решение системы $A\mathbf{X} = \mathbf{b}$, а $\mathbf{X}$—точное решение этой системы, то можно написать равенство

$$\|\mathbf{X}^* - \mathbf{X}\| = \|A^{-1}(A\mathbf{X}^* - \mathbf{b})\| \leqslant \|A^{-1}\|\|A\mathbf{X}^* - \mathbf{b}\|,$$

которое редко применяется из-за сложности оценки $\|A^{-1}\|$. Поэтому при практическом анализе погрешности приближений, получаемых итераци- онными методами, обычно вместо этой оценки используется рассматри- ваемая далее нестрогая, но более простая оценка погрешности, которая строится на основании дополнительной информации, получаемой в про- цессе вычислений.

Примем следующий *критерий разумности практической оценки погреш- ности:* $\mathbf{v}^n$ принимается за практическую погрешность приближения $\mathbf{x}^n$, стремящегося к $\mathbf{X}$ при $\to \infty$, если

$$\|\mathbf{v}^n - (\mathbf{x}^n - \mathbf{X})\|/\|\mathbf{x}^n - \mathbf{X}\| \to 0 \quad \text{при} \quad n \to \infty \tag{1}$$

Ясно, что тогда $\|\mathbf{v}^n\| \sim \|\mathbf{x}^n - \mathbf{X}\|$.

Рассмотрим метод простой итерации $\mathbf{x}^{n+1} = B\mathbf{x}^n + \mathbf{c}$ Для краткости изложения ограничимся случаем, когда матрица B простой структуры (т.е. ее жорданова форма диагональна и поэтому она обладает полной системой собственных векторов).

Пусть $\lambda_i, i = 1, ..., m$,—собственные значения матрицы B, занумеро- ванные в порядке убывания $|\lambda_i|$, причем $1 > |\lambda_1| > |\lambda_2| \geqslant |\lambda_3| \geqslant ... \geqslant |\lambda_m|$, а $\mathbf{e}_i, \|\mathbf{e}_i\| = 1$, — соответствующие собственные векторы, образующие полную систему. Разложим вектор $\mathbf{r}^0$ по базису $\mathbf{e}_i$: $\mathbf{r}^0 = \sum c_i\mathbf{e}_i$. Тогда

$$\mathbf{r}^0 = \mathbf{x}^n - \mathbf{X} = B^n\mathbf{r}^0 = \sum c_i\lambda_i^n\mathbf{e}_i = c_1\lambda_i^n\mathbf{e}_1 + O(|\lambda_2|^n). \tag{2}$$

Здесь и далее выражение $\mathbf{x}^n = \mathbf{y}^n + O(\varepsilon_n)$ имеет следующий смысл:

$$\|\mathbf{x}^n - \mathbf{y}^n\| = O(\varepsilon_n) \quad \text{при} \quad n \to \infty.$$

Далее в этом параграфе $\|\mathbf{x}\|$ — это $\|\mathbf{x}_2\|$.

Укажем способ построения приближения к вектору $\mathbf{w}^n = c_1\lambda_1^n\mathbf{e}_1$ на основании информации, получающейся в ходе вычислений. Согласно (2) имеем

$$\begin{aligned}
\mathbf{x}^{n-2} - \mathbf{X} &= \mathbf{w}^n\lambda_1^{-2} + O(|\lambda_2|^n),\\
\mathbf{x}^{n-1} - \mathbf{X} &= \mathbf{w}^n\lambda_1^{-1} + O(|\lambda_2|^n),\\
\mathbf{x}^n - \mathbf{X} &= \mathbf{w}^n + O(|\lambda_2|^n).
\end{aligned}$$

Вычитая друг из друга соседние соотношения, получим

$$\begin{aligned}
\mathbf{x}^{n-1} - \mathbf{x}^{n-2} &= \mathbf{w}^n(1 - \lambda_1^{-1})\lambda_1^{-1} + O(|\lambda_2|^n),\\
\mathbf{x}^n - \mathbf{x}^{n-1} &= \mathbf{w}^n(1 - \lambda_1^{-1})\lambda_1^{-1} + O(|\lambda_2|^n).
\end{aligned} \tag{3}$$

Положим

$$\lambda_1^{(n)} = \frac{(\mathbf{x}^n - \mathbf{x}^{n-1}, \mathbf{x}^n - \mathbf{x}^{n-1})}{(\mathbf{x}^{n-1} - \mathbf{x}^{n-2}, \mathbf{x}^n - \mathbf{x}^{n-1})}.$$

Воспользуемся соотношениями (4) и в предположении $c_1 \neq 0$ поделим числитель и знаменатель выражении для $\lambda_1^{(n)}$ на $\|\mathbf{w}^n\|^2 \left|1 - \lambda_1^{-1}\right|^2 \lambda_1^{-1}$; в результате получим

$$\lambda_1^{(n)} = \frac{\lambda_1 + O\left(\frac{|\lambda_2|^n}{\|\mathbf{w}^n\|}\right)}{1 + O\left(\frac{|\lambda_2|^n}{\|\mathbf{w}^n\|}\right)}.$$

Поскольку

$$\|\mathbf{w}^n\| = |c_1|\ \lambda_1|^n \tag{5}$$

то

$$\lambda_1^{(n)} = \lambda_1 + O\left(|\lambda_2/\lambda_1|^n\right). \tag{6}$$

Поделив второе из соотношений (3) на $1 - \left(\lambda_1^{(n)}\right)^{-1}$, получим

$$\frac{\mathbf{x}^n - \mathbf{x}^{n-1}}{1 - \left(\lambda_1^{(n)}\right)^{-1}} = \mathbf{w}^n \frac{1 - \lambda_1^{-1}}{1 - \left(\lambda_1^{(n)}\right)^{-1}} + O\left(|\lambda_2|^n\right) = \mathbf{w}^n + \mathbf{w}^n \frac{\lambda_1 - \lambda_1^{(n)}}{\lambda_1(\lambda_1^{(n)} - 1)} + O\left(|\lambda_2|^n\right).$$

Из (5), (6) следует $\|\mathbf{w}^n(\lambda_1 - \lambda_1^{(n)})\| = O\left(|\lambda_2|^n\right)$; поэтому

$$\frac{\mathbf{x}^n - \mathbf{x}^{n-1}}{1 - \left(\lambda_1^{(n)}\right)^{-1}} = \mathbf{w}^n + O\left(|\lambda_2|^n\right).$$

Отсюда и из (2) получаем

$$\mathbf{x}^n - \mathbf{X} = \mathbf{v}^n + O(|\lambda_2|^n),$$

где $\mathbf{v}_n = \left(\mathbf{x}^n - \mathbf{x}^{n-1}\right)/(1 - \lambda_1^n)^{-1})$. Заметим, что согласно (3), (6) $\|\mathbf{v}^n\| = |c_1|\,|\lambda_1|^n + O(|\lambda_2|^n)$. Из этих равенств вытекает, что $\mathbf{v}^n$ удовле- творяет критерию (1), и поэтому его можно принять за практическую погрешность приближения $\mathbf{x}^n$.

В случае $c_1 = ... = c_l = 0, c_{l+1} \neq 0$ проведенные рассуждения останутся в силе, если $|\lambda_{l+1}| > |\lambda_{l+2}|$ Во всех соотношениях следует заменить лишь $\lambda_i, c_i, \mathbf{e}_i$ при $i = 1, 2$ на $\lambda_{l+i}, c_{l+i}, \mathbf{e}_{l+i}$. Описанный способ получения оценки приближенного решения называется δ^2- *процессом.*

Если положить $\mathbf{y}^n = \mathbf{x}^n - \mathbf{v}^n$, то $\mathbf{y}^n - \mathbf{X} = O(|\lambda_2|^n)$, и поэтому $\mathbf{y}^n$, вообще говоря, является лучшим начальным условием для последующих итераций по сравнению с $\mathbf{x}^n$. Производя время от времени такие уточнения, иногда удается существенно уменьшить общее число итераций.

Для справедливости приближенного равенства

$$\mathbf{x}^n - \mathbf{X} \approx \mathbf{v}^n$$

необходимо, чтобы в правой части равенства

$$\mathbf{x}^n - \mathbf{X} = \sum_i c_i \lambda_i^n \mathbf{e}_i$$

$\mathbf{x}^n - \mathbf{x}^{n-1}, \mathbf{x}^{n-1} - \mathbf{x}^{n-2}$ приблизительно пропорциональны и

$$\mu_n = \frac{\left|(\mathbf{x}^{n-1} - \mathbf{x}^{n-2}, \mathbf{x}^{n-1} - \mathbf{x}^n)\right|}{\|\mathbf{x}^{n-1} - \mathbf{x}^{n-2}\|\|\mathbf{x}^{n-1} - \mathbf{x}^n\|} \approx 1.$$

Таким образом, условие $\mu_1 \approx 1$ является необходимым для того, чтобы проводившиеся ранее построения были справедливы. Поэтому его можно принять за условие практической применимости (7).

Например, возможна следующая схема метода простой итерации с примене- нием δ^2-процесса ускорения сходимости. Задаются некоторым η' в пределах $1 > \eta > 0$ и малым $\eta > 0$.Если по ходу итераций оказалось,что $\mu_n \geqslant 1 - \eta$, то вычисляется $\mathbf{v}^n$ и вектор $\mathbf{y}^n$ принимается за начальное приближение для последующих итераций. Итерационный процесс прекращается, если μ_n и $\|\mathbf{v}^n\| \leqslant \varepsilon$, где ε требуемая точность.

Если η очень мало, то условие $\eta_n \geqslant 1 - \eta$ будет выполняться только после большого числа итераций, ускорение сходимости не будет иметь места. При большом η соотношения, положенные в основу наших построений, выполня- ются грубо, поэтому не исключено, что применение δ^2-процесса сходимости замедлит итерационный процесс. Картина итераций также осложняется нали- чием погрешности округлений, так что описанная выше схема требует практи- ческой отработки на большом числе примеров с целью выбора оптимальных η', η и указания нижней границы значений ε, при которых алгоритм применим. Если однородный итерационный процесс подвергается перестройке (в нашем случае при переходе от $\mathbf{x}^n$ к $\mathbf{y}^n$), то иногда полезно проверить, не ведет ли эта перестройка к ухудшению. В качестве критерия целесообразности перестройки можно взять некоторое соотношение, связывающее нормы невязок для $\mathbf{x}^n, \mathbf{y}^n$, например неравенство вида

$$\|(E - B)\mathbf{y}^n - \mathbf{c}\| \leqslant q\|(E - B)\mathbf{x}^n - \mathbf{c}\|.$$

Замечание о необходимости указания нижней грани значений ε вызывается следующим обстоятельством. Пусть для определенности $\lambda_1 > 0$. Уже при вычислении $\mathbf{x}^n$ по заданному $\mathbf{x}^{n-1}$ погрешности округления могут возмутить ре- n зультат на величину $\delta\mathbf{x}^n$ нормой порядка ρ. Следствием этого может явиться возмущение $\delta\mathbf{v}^n$, имеющее норму порядка $(1\lambda_1)^1\rho$. Отсюда следует, что в случае $\varepsilon < (1\lambda_1)^{-1}\rho$ итерационный процесс может никогда не закончиться. Проведенные построения показывают, что при реализации метода возника- ет много таких моментов, разбор которых требует серьезной математической подготовки и проведения большой серии численных экспериментов. Поэтому, несмотря на «простоту» метода простой итерации, будет вполне оправданным создание стандартной программы этого метода.

§6. Оптимизация скорости сходимости итерационных процессов

Рассмотрим простейший итерационный способ решения системы уравне- ний

$A\mathbf{x} = \mathbf{b}$:

$$x^{n-1} = \mathbf{x}^n - \alpha(A\mathbf{x}^n - \mathbf{b}).$$

Мы видели, что скорость сходимости такого итерационного процес- са существенно зависит от максимального модуля собственных значений матрицы $B = E$. Если $\lambda_1, ..., \lambda_n$ — собственные значения матрицы A, то $\max\limits_i |\lambda_1 B| = \max\limits_i |1 - \alpha\lambda_1|$ Из рис. 6.6.1 видно, что при действительных собственных значениях различных знаков этот максимум больше 1 и итерационный процесс расходится.

Обратимся к часто встречающемуся случаю, когда все $\lambda_i > 0$. Значения λ_i бывают известны крайне редко, однако довольно типичен случай, когда известна оценка для этих чисел вида $0 < \mu \leqslant \lambda_i \leqslant M < \infty$ при всех i. Скорость сходимости итерационного процесса можно характеризовать величиной

$$\rho(\alpha) = \max_{\mu \leqslant \lambda \leqslant M} |1 - \alpha\lambda| .$$

Рассмотрим задачу минимизации $\rho(\alpha)$ за счет выбора α.

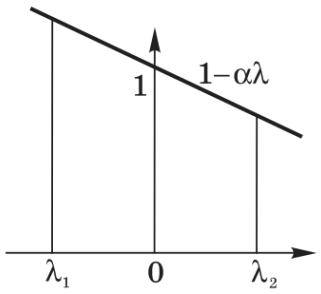

Рис. 6.6.1

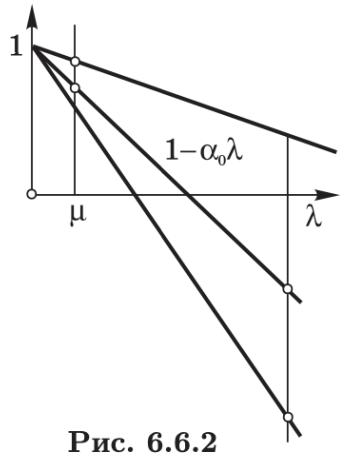

Рис. 6.6.2

Для нахождения $\min\limits_\alpha \rho(\alpha)$ удобно обратиться к геометрической картине α (рис. 6.6.2). Ясно, что $\rho(\alpha) \geqslant 1$ при $\alpha \leqslant 0$. При $0 < \alpha \leqslant M^1$ функция $1\alpha\lambda$ неотрицательна и монотонно убывает на отрезке $[\mu, M]$, поэтому $\rho(\alpha) = 1\alpha\mu$. При $M^1 < \alpha$ величина $1\alpha M$ отрицательна и модуль ее растет с ростом α. При некотором $\alpha = \alpha_0$ наступит момент, когда

$$1 - \alpha_0\mu = -(1 - \alpha_0 M), \tag{1}$$

и тогда $\rho(\alpha_0) = |1 - \alpha_0\mu|$. Если $\alpha < \alpha_0$, то $\rho(\alpha) = 1 - \alpha\mu > 1 - \alpha_0\mu = \rho(\alpha_0)$; если $\alpha_0 < \alpha$, то $\rho(\alpha) \geqslant |1 - \alpha M| = M\alpha - 1 \geqslant M\alpha_0 - 1 = \rho(\alpha_0)$.

Таким образом, значение $\alpha = \alpha_0$ является искомым. Решая уравнение (1) относительно α_0, получим $\alpha_0 = 2/(M + \mu)$. Отсюда

$$\rho(\alpha_0) = (M - \mu)/(M + \mu).$$

Задача 1. Доказать сходимость итерационного процесса при $\alpha = \|A\|^{-1}$.

На примере систем с матрицей $A > 0$ (здесь и далее неравенство $A > 0$ означает, что A — симметричная положительно определенная мат- рица) рассмотрим

более формализованные постановки проблем оптимиза- ции скорости сходимости итерационных процессов.

Если число ненулевых элементов матрицы много больше ее размерно- сти, то операция умножения матрицы на вектор более трудоемка, чем умножение числа на вектор или сложение векторов. Поэтому при оценке трудоемкости итерационных процессов и оптимизации этих процессов да- лее за меру трудоемкости мы неявно принимаем число умножений мат- рицы A на вектор.

Всякая система $A\mathbf{x} = \mathbf{b}$ с $detA \neq 0$, вообще говоря, может быть приве- дена (как говорят, *симметризована*) умножением обеих частей уравнения на матрицу A^T к системе с симметричной положительно определенной матрицей. В самом деле, система $A^TA\mathbf{x} = A^T\mathbf{b}$ эквивалентна исходной, матрица A^TA симметричная, так как $A^TA = (A^TA)^T$, и положительно определена, так как $(A^TA\mathbf{x}, \mathbf{x}) = \|A\mathbf{x}\|^2 > 0$ при $\mathbf{x} \neq 0$. По возможности стараются избегать симметризации, поскольку, как мы увидем далее, она часто приводит к ухудшению сходимости итерационных процессов.

Рассмотрим несколько более общий итерационный метод, чем метод простой итерации. А именно, в методе простой итерации

$$\mathbf{x}^{k+1} = \mathbf{x}^k - \tau(A\mathbf{x}^k - \mathbf{b})$$

будем считать, что итерационный параметр τ может изменяться от шага к шагу. Тогда метод примет вид

$$\mathbf{x}^{k+1} = \mathbf{x}^k - \tau_{k+1}(A\mathbf{x}^k - \mathbf{b}), \quad k = 0, 1, ..., \tag{2}$$

где $\mathbf{x}_0$—некоторое начальное приближение.

Зададимся некоторым целым $n > 0$ и произведем n итераций по фор- муле (2). Согласно (2) погрешность $\mathbf{r}^k = \mathbf{x}^k - \mathbf{X}$ удовлетворяет соотно- шению

$$\mathbf{r}^{k+1} = \mathbf{r}^k - \tau_{k+1}A\mathbf{r}^k = (E - \tau_{k+1}A)\mathbf{r}^k. \tag{3}$$

Тогда через n шагов итерационного метода (2) погрешность $\mathbf{r}^n$ будет выражать- ся через погрешность начального приближения $\mathbf{r}^0$ следующим образом:

$$\mathbf{r}^n = (E - \tau_nA)\mathbf{r}^{n-1} = ... = (E - \tau_nA)...(E - \tau_1A)\mathbf{r}^0, \tag{4}$$

где $\mathbf{r}_0 = \mathbf{x}_0 - \mathbf{X}$−погрешность начального приближения.