Отчет по численным методам: итерационный метод Чебышева

Сотникова Виктория
307 группа

1 Постановка задачи

Требуется явным итерационным методом Чебышева с оптимальным набором итерационных параметров приближенно решить систему линейных алгебраических уравнений

$$x + Ax = (I + A)x = F$$

с симметричной положительно определенной матрицей $A \in R^{n\times n}$.

2 Описание метода решения задачи

2.1 Оценка спектра матрицы [3]

Теорема (Гершгорина): каждое характеристическое число λ матрицы $A = \|a_{ik}\|_1^n$ всегда расположено в одном из кругов

$$|a_{ii} - \lambda| \le \sum_{j=1, j\ne i}^{n} |a_{ij}| \quad (i = 1, \ldots, n)$$

Объединение всех точек кругов Гершгорина дает некоторую область локализации характеристических чисел матрицы A, т. е. область, в которой заведомо лежат все характеристические числа матрицы A. Исходя из условий Адамара для столбцов, получаем область локализации в виде объединения n кругов

$$|\lambda - a_{ii}| \le \sum_{j=1, j\ne i}^{n} |a_{ij}| \quad (i = 1, \ldots, n)$$

Тогда границы спектра матрицы можно оценить следующими значениями сверху и снизу соответственно:

$$\max_i \left(a_{ii} + \sum_{j\ne i} |a_{ij}| \right), \quad \min_i \left(a_{ii} - \sum_{j\ne i} |a_{ij}| \right)$$

где a_{ij} — элементы матрицы A, a_{ii} — модуль диагонального элемента, $\sum_{j\ne i} |a_{ij}|$ — сумма модулей остальных элементов строки.

2.2 Метод Чебышева [1]

8.3 Нестационарные итерационные методы

Другой путь увеличения скорости сходимости итерационного метода (8.2) состоит в том, чтобы вместо одного итерационного параметра τ использовать несколько — свой итерационный параметр на каждой итерации. Итерационные методы такого типа называются нестационарными и имеют вид

$$B\,\frac{x^{k+1}-x^k}{\tau_{k+1}} + Ax^k = b, \quad k = 0,1,\dots \tag{8.15}$$

или с учетом (8.7), (8.10)

$$\frac{y^{k+1}-y^k}{\tau_{k+1}} + Cy^k = f, \quad k = 0,1,\dots. \tag{8.16}$$

Укажем один из возможных способов выбора итерационных параметров τ_k. Введем обозначение

$$z^k = y^k - y,$$

где

$$Cy = f, \tag{8.17}$$

и вычтем (8.17) из (8.16). В результате получим задачу для z^k:

$$\frac{z^{k+1}-z^k}{\tau_{k+1}} + Cz^k = 0, \quad k = 0,1,\dots, \quad z^0 = y^0 - y. \tag{8.18}$$

Отсюда

$$z^{k+1} = (I - \tau_{k+1}C)z^k,$$

и, следовательно,

$$z^k = \prod_{j=1}^{k}(I - \tau_j C)z^0, \tag{8.19}$$

т.е.

$$z^k = P_k(C)z^0,$$

где

$$P_k(t) = \prod_{j=1}^{k}(1 - \tau_j t) = 1 + a_1^{(k)}t + \dots + a_k^{(k)}t^k. \tag{8.20}$$

Из (8.19) находим, что

$$\|z^k\| = \|y^k - y\| = \left\|\prod_{j=1}^{k}(I - \tau_j C)z^0\right\| \leqslant \left\|\prod_{j=1}^{k}(I - \tau_j C)\right\| \|z^0\|. \tag{8.21}$$

Рис. 1: Андреев В.Б. *Численные методы.* 2013.

Но

$$\left\| \prod_{j=1}^{k} (I - \tau_j C) \right\| = \max_l \left| \lambda_l \big(P_k(C) \big) \right| = \max_l \left| P_k \big(\lambda_l(C) \big) \right|. \tag{8.22}$$

Поскольку мы хотим, чтобы итерации сходились как можно быстрее, то можно поставить задачу о минимизации $\|P_k(C)\|$ в зависимости от итерационных параметров τ_j, $j = \overline{1,k}$. В силу (8.20), (8.22) эта задача эквивалентна задаче построения многочлена $P_k(t)$ степени k с единичным свободным членом, который в точках спектра матрицы C наиболее близок к нулю. Но поставленная задача практически не разрешима. Однако вместо нее можно поставить близкую задачу о построении $P_k(t)$, наименее отклоняющегося от нуля не на спектре, а на отрезке $[\lambda_1, \lambda_n]$, где этот спектр расположен. Эта задача много проще, и решение ее известно. Найдем это решение.

Итак, среди многочленов степени k таких, что $Q_k(0) = 1$ (см. (8.20)), требуется найти многочлен $P_k(t)$, максимум модуля которого на $[\lambda_1, \lambda_n]$ минимален.

Линейной заменой переменной $t = ax + b$ переведем отрезок $[\lambda_1, \lambda_n]$ в отрезок $[1, -1]$. Имеем

$$\left. \begin{array}{rcrcl} \lambda_1 & = & a & + & b \\ \lambda_n & = & -a & + & b \end{array} \right\}, \qquad b = \frac{\lambda_1 + \lambda_n}{2}, \qquad a = \frac{\lambda_1 - \lambda_n}{2},$$

т.е.

$$t = \frac{\lambda_n + \lambda_1}{2} - \frac{\lambda_n - \lambda_1}{2} x = \frac{\lambda_1 + \lambda_n}{2} \left[1 - \frac{\lambda_n - \lambda_1}{\lambda_n + \lambda_1} x \right] = \frac{1}{\tau_0} [1 - \rho_0 x], \tag{8.23}$$

где

$$\tau_0 = \frac{2}{\lambda_1 + \lambda_n}, \quad \rho_0 = \frac{\lambda_n - \lambda_1}{\lambda_n + \lambda_1} < 1. \tag{8.24}$$

Заметим, что τ_0 совпадает с τ из (8.5), (8.11) для стационарного итерационного процесса, а ρ_0 совпадает с q из (8.6), (8.12) и характеризует скорость сходимости этого процесса.

Пусть

$$P_k(t) = \widehat{P}_k(x). \tag{8.25}$$

Поскольку $t = 0$ отвечает точка $x = \rho_0^{-1}$ (см. (8.23)), то должно быть

$$P_k(0) = \widehat{P}_k \left(\frac{1}{\rho_0} \right) = 1. \tag{8.26}$$

Наша задача свелась к отысканию многочлена $\widehat{P}_k(x)$, наименее отклоняющегося от нуля на $[-1, 1]$ и удовлетворяющего условию (8.26). С похожей задачей мы уже сталкивались на прошлой лекции. Мы знаем, что среди многочленов степени k вида $x^k + \ldots$ наименее отклоняется от нуля на $[-1, 1]$ многочлен

$$\overline{T}_k(x) = \frac{1}{2^{k-1}} T_k(x),$$

Рис. 2: Андреев В.Б. *Численные методы.* 2013.

где $T_k(x)$ — многочлен Чебышева первого рода. Разумеется, сам многочлен Чебышева $T_k(x)$ является наименее отклоняющимся от нуля на $[-1,1]$ среди многочленов вида $P_k(x) = 2^{k-1}x^k + \dots$.

Пусть $T_k\left(\frac{1}{\rho_0}\right) = \frac{1}{q_k}$. Тогда очевидно, что искомое решение дает многочлен

$$\widehat{P}_k(x) = q_k T_k(x). \tag{8.27}$$

Из свойства 6° многочленов T_k

$$q_k = \left[\operatorname{ch} k \operatorname{Arch} \frac{1}{\rho_0}\right]^{-1} < 1. \tag{8.28}$$

Получим еще одно представление для q_k. Из (8.28)

$$\operatorname{ch} k \operatorname{Arch} \frac{1}{\rho_0} = \frac{1}{q_k}.$$

Отсюда

$$k \operatorname{Arch} \frac{1}{\rho_0} \equiv k \, \ln \frac{1+\sqrt{1-\rho_0^2}}{\rho_0} = \operatorname{Arch} \frac{1}{q_k} \equiv \ln \frac{1+\sqrt{1-q_k^2}}{q_k}.$$

Пусть

$$\begin{aligned}\rho_1 &= \frac{\rho_0}{1+\sqrt{1-\rho_0^2}} = \frac{\frac{\lambda_n-\lambda_1}{\lambda_n+\lambda_1}}{1+\sqrt{1-\left(\frac{\lambda_n-\lambda_1}{\lambda_n+\lambda_1}\right)^2}} = \frac{\lambda_n-\lambda_1}{\lambda_n+\lambda_1+2\sqrt{\lambda_1\lambda_n}} = \\ &= \frac{\sqrt{\lambda_n}-\sqrt{\lambda_1}}{\sqrt{\lambda_n}+\sqrt{\lambda_1}} = \frac{1-\sqrt{\lambda_1/\lambda_n}}{1+\sqrt{\lambda_1/\lambda_n}}.\end{aligned} \tag{8.29}$$

Тогда

$$\ln \frac{1+\sqrt{1-q_k^2}}{q_k} = k \, \ln \frac{1}{\rho_1} = \ln \rho_1^{-k}$$

и

$$\frac{1+\sqrt{1-q_k^2}}{q_k} = \rho_1^{-k}.$$

Отсюда, переходя к квадратному уравнению относительно q_k^2, находим, что

$$q_k = \frac{2\rho_1^{-k}}{1+\rho_1^{-2k}} = \frac{2\rho_1^k}{1+\rho_1^{2k}}. \tag{8.30}$$

Итак, мы нашли такой многочлен

$$\widehat{P}_k(x) = q_k T_k(x),$$

Рис. 3: Андреев В.Б. *Численные методы.* 2013.

что

$$\max_{[-1,1]} \left| \widehat{P}_k(x) \right| = q_k$$

и, следовательно, в силу (8.21), (8.7) имеем оценку сходимости

$$\|z^k\| = \|x - x^k\|_B \leqslant q_k \, \|x - x^0\|_B. \tag{8.31}$$

Получим формулы для итерационных параметров τ_j. Из (8.25), (8.27) и (8.23) следует, что нули полиномов $P_k(t)$ и $T_k\left(\frac{1-\tau_0 t}{\rho_0}\right)$ совпадают. Так как полином $P_k(t)$ имеет нули в точках $t = 1/\tau_j$, $j = \overline{1,k}$, а нулями полинома Чебышева $T_k(x)$ являются числа (7.21)

$$x_j = -\cos\frac{(2j-1)\pi}{2k}, \quad j = \overline{1,k},$$

то с учетом (8.23) находим, что

$$\tau_j = \frac{\tau_0}{1 + \rho_0 \mu_j}, \quad j = \overline{1,k}, \tag{8.32}$$

где

$$\mu_j \in \mathfrak{M}_k = \left\{ -\cos\frac{2i-1}{2k}\pi, \quad i = \overline{1,k} \right\}. \tag{8.33}$$

Рис. 4: Андреев В.Б. *Численные методы.* 2013.

2.3 Построение оптимальной последовательности итерационных параметров [2]

Порядок использования итерационных параметров τ_k в чебышевском методе существенно влияет на сходимость метода. Поэтому возникает задача построения наилучшей последовательности итерационных параметров, обеспечивающей минимальное влияние вычислительной погрешности метода. Так как последовательность параметров определяется упорядочением множества $\mathfrak{M}_n$, то необходимо построить оптимальное упорядочение множества $\mathfrak{M}_n$.

Пусть сначала число итераций есть степень 2: $n = 2^p$. Обозначим через θ_m множество, состоящее из m целых чисел:

$$\theta_m = \left\{ \theta_1^{(m)}, \theta_2^{(m)}, \ldots, \theta_m^{(m)} \right\}$$

Исходя из множества $\theta_1 = \{1\}$, построим множество $\theta_2 p$ по следующему правилу. Пусть множество θ_m построено. Тогда множество θ_{2m} определим по формулам

$$\theta_{2m} = \left\{ \theta_{2i}^{(2m)} = 4m - \theta_i^{(m)}, \theta_{2i-1}^{(2m)} = \theta_i^{(m)}, i = 1, 2, \ldots, m \right\}, \\ m = 1, 2, 4, \ldots, 2^{p-1}$$

Нетрудно убедиться, что множество θ_{2^k} состоит из нечетных чисел от 1 до $2^{k+1} - 1$. Используя построенное множество $\theta_2 p$, упорядочим множество $\mathfrak{M}_2 p$ следующим образом:

$$\mathfrak{M}_n^* = \left\{ -\cos \beta i, \beta_i = \frac{\pi}{2n} \theta_i^{(n)}, i = 1, 2, \ldots, n \right\}, n = 2^p.$$

Это и есть искомое упорядочение множества $\mathfrak{M}_n$ в случае, когда $n = 2^p$. Для соответствующей этому упорядочению последовательности итерационных параметров доказана оценка

$$\sum_{i=1}^{n} \tau_j \left\| T_{n,j} \right\| \frac{1 - q_n}{\gamma_1}.$$

Построенное упорядочение множества $\mathfrak{N}_n^*$ обеспечивает минимальное влияние вычислительной погрешности на сходимость чебышевского метода.

3 Листинг программы, реализующей метод Чебышева

Основная функция, реализующая метод Чебышева:

```
Matrix Chebyshev(const Matrix& R, int k, const Vector& f, double
   lambdaMin, double lambdaMax) {
    int n = R.size(), m, parityCounter = 0;;
    Vector xPrev(n, 0.0), mi(k, 0.0), teta0(k, 0.0), teta1(k, 0.0), Rx(
       n, 0.0);
    Matrix solutions(k, Vector(n, 0.0));

    double tau_k, tau0 = 2.0 / (lambdaMax + lambdaMin);
    double ro0 = (lambdaMax - lambdaMin) / (lambdaMax + lambdaMin);
    teta0[0] = 1;

    for (m = 1; m < k; m *= 2) {
        Vector& currentTeta = (parityCounter % 2 == 0) ? teta1 : teta0;
        Vector& previousTeta = (parityCounter % 2 == 0) ? teta0 : teta1
           ;

        for (int i = 1; i <= m; ++i) {
            currentTeta[2 * i - 2] = previousTeta[i - 1];
            currentTeta[2 * i - 1] = 4 * m - previousTeta[i - 1];
        }

        parityCounter++;
    }

    Vector& currentTeta = (parityCounter % 2 == 1) ? teta1 : teta0;

    for (int i = 1; i <= k; ++i) {
        mi[i - 1] = cos(M_PI / (2 * k) * currentTeta[i - 1]);
    }

    for (int i = 0; i < k; ++i) {
        Rx = multiplyR(R, xPrev);
        tau_k = tau0 / (1 - ro0 * mi[i]);

        for (int j = 0; j < n; ++j) {
            solutions[i][j] = (f[j] - Rx[j]) * tau_k + xPrev[j];
            xPrev[j] = solutions[i][j];
        }
    }

    return solutions;
}
```

Исходный код приложен к отчету.

Команда для компиляции:

```
g++ 1.cpp -o 1
```

Команда для запуска:

```
./1
```

Замечание. Как правило, методу требуется от 512 до 2048 итераций. Исходя из этого расчеты ограничены соответствующим количеством итераций. Но бывают редкие случаи, когда нормы погрешностей очень близки и тогда необходимо повысить количество итераций, изменив const int exponent в начале 1.cpp на большую степень двойки.

4 Результаты

Ниже представлены результаты вычислений.

Границы спектра матрицы I + A: [1;154.4]

Таблица 1: Результаты вычислений

Итерации	Норма погрешности	Относительная погрешность	Показатель степени
1	3.732035e+01	6.738653e+01	0
2	2.059595e+01	3.718856e+01	1
4	1.043445e+00	1.884070e+00	2
8	7.560321e-01	1.365110e+00	3
16	2.816696e-01	5.085895e-01	4
32	4.941594e-02	8.922662e-02	5
64	3.378608e-03	6.100497e-03	6
128	5.163887e-06	9.324040e-06	7
256	9.549431e-13	1.724269e-12	8
512	7.384630e-17	1.333387e-16	9
1024	7.382687e-17	1.333036e-16	10
2048	7.382472e-17	1.332997e-16	11
4096	7.382855e-17	1.333066e-16	12
8192	7.383801e-17	1.333237e-16	13
16384	7.383049e-17	1.333101e-16	14
32768	7.382024e-17	1.332916e-16	15

Наименьший показатель степени двойки, при котором погрешность решения не превосходит погрешность прямого метода: 11.

То есть потребовалось 2048 итераций для достижения необходимой точности.
Относительная погрешность: 1.332997e-16

Таблица 2: Среднеквадратическая норма погрешности

Метод	Значение
Прямой метод	7.38254×10^{-17}
Метод Чебышева	7.382472×10^{-17}

График среднеквадратической нормы погрешности решения:

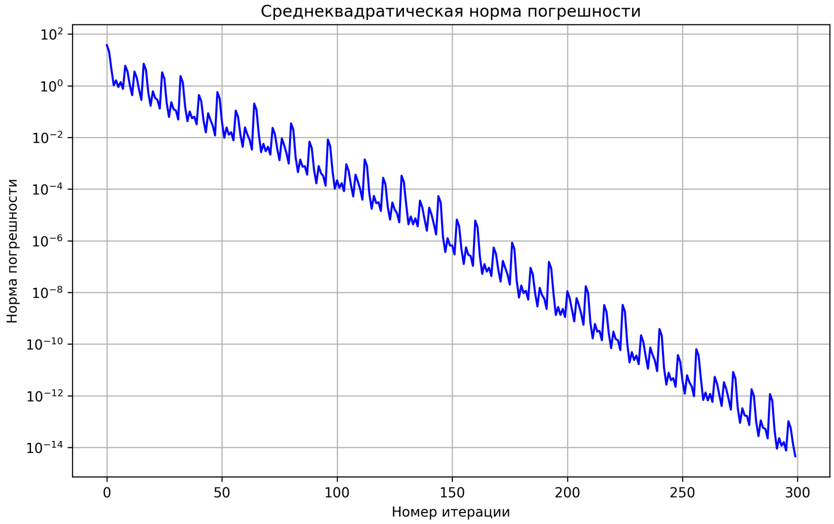

Рис. 5: График (вариант 1)

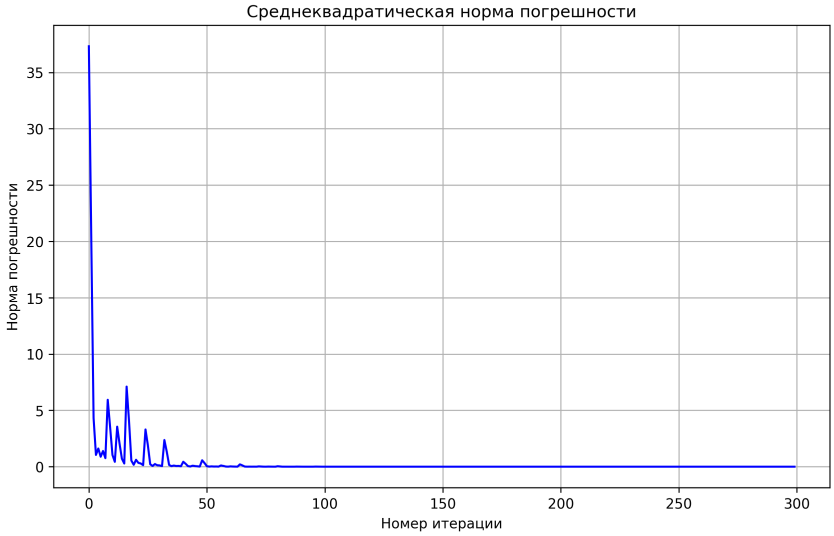

Рис. 6: График (вариант 2)

Список литературы

1. Андреев В.Б. *Численные методы*. 2013.
2. Самарский А.А., Николаев Е.С. *Методы решения сеточных уравнений*. 1978.
3. Гантмахер Ф.Р. *Теория матриц*. 1966.