Метод максимизации столбцов матрицы*

В. Г. ЛЕЖНЁВ
Кубанский государственный университет

УДК 519.61

Ключевые слова: матрица, сингулярное разложение, спектральная задача.

Аннотация

Рассматривается итерационный метод сингулярного разложения вещественных матриц, использующий односторонние вращения. Алгоритм применён также для решения спектральной проблемы симметричных матриц.

Abstract

V. G. Lezhnev, A method for maximizing matrix columns, Fundamentalnaya i prikladnaya matematika, vol. 14 (2008), no. 2, pp. 113—119.

We describe a singular-value decomposition method, where one-side rotation is used. The algorithm is also applied for a symmetrical spectral problem for matrices.

1. Максимизация нормы столбца

Пусть A — вещественная квадратная матрица порядка n с элементами a_{ij}. Будем обозначать $a_k = (a_{1k}, \ldots, a_{nk})^{\mathrm{T}}$ её k-й столбец (T — знак транспонирования), а через $|a_k|$ — евклидову норму вектор-столбца a_k. Пусть

$$(a_k, a_p) = \sum_{i=1}^{n} a_{ik} a_{ip}.$$

Будем далее для простоты полагать, что норма первого столбца наибольшая, $|a_1| \geqslant |a_k|$, $k = 2, \ldots, n$.

Обозначим через $U_p(\varphi) = (u_{ij})$ ортогональную матрицу простого поворота с элементами $u_{11} = u_{pp} = c$, $-u_{1p} = u_{p1} = s$, где $c = \cos\varphi$, $s = \sin\varphi$, $p > 1$, остальные диагональные элементы равны единице, внедиагональные — нулю.

Рассмотрим матрицу $B = AU_p(\varphi)$. Для столбцов $b_k = (b_{1k}, \ldots, b_{nk})^{\mathrm{T}}$ матрицы $B = (b_{ij})$ имеем

$$b_1 = ca_1 + sa_p, \quad b_p = -sa_1 + ca_p, \quad b_k = a_k, \quad k \neq 1, p, \tag{1}$$

столбцы b_1 и b_p являются линейными комбинациями столбцов a_1 и a_p, остальные столбцы матриц A и B совпадают.

*Работа выполнена в рамках проектов РФФИ № 07-01-00269, 06-01-96645.

Фундаментальная и прикладная математика, 2008, том 14, № 2, с. 113—119.

Свойство 1. При любых углах φ выполняется равенство

$$|b_1|^2 + |b_p|^2 = |a_1|^2 + |a_p|^2.$$

Угол φ будем выбирать так, чтобы норма первого столбца $|b_1|$ была максимальной [3, 5]. Рассмотрим функцию

$$f(\varphi) := \cos^2\varphi|a_1|^2 + 2\cos\varphi\sin\varphi(a_1, a_p) + \sin^2\varphi|a_p|^2.$$

Имеем

$$f'(\varphi) = 2\cos 2\varphi(a_1, a_p) - \sin 2\varphi[|a_1|^2 - |a_p|^2],$$

функция $f(\varphi)$ периодическая с периодом $T = \pi$.

Для определения максимизирующего угла $\varphi = \varphi_p$, $\|b_1\|^2 = f(\varphi_p)$, получаем (приравнивая производную к нулю) следующее последовательное правило а)—в):

а) если $(a_1, a_p) = 0$, то $\varphi_p = 0$;

б) если $|a_1|^2 - |a_p|^2 = 0$, то

$$2\varphi_p = \frac{\pi}{2}\operatorname{sgn}(a_1, a_p); \tag{2}$$

в) если $|a_1|^2 - |a_p|^2 \neq 0$, то угол φ_p определяется равенством

$$\operatorname{tg} 2\varphi_p = \frac{2(a_1, a_p)}{|a_1|^2 - |a_p|^2}, \quad \varphi_p \in \left(-\frac{\pi}{4}, \frac{\pi}{4}\right). \tag{3}$$

Отсюда следует, что $f'(\varphi_p) = 0$. Покажем, что угол φ_p является максимизирующим.

Если $(a_1, a_p) = 0$ и $\varphi_p = 0$, то

$$f(\varphi) = \cos^2\varphi|a_1|^2 + \sin^2\varphi|a_2|^2 \leqslant |a_1|^2 = f(\varphi_p).$$

Если $|a_1|^2 - |a_p|^2 = 0$, $\varphi_p = \pm\pi/4$, то $\sin 2\varphi_p(a_1, a_p) = |(a_1, a_p)|$, т. е.

$$f(\varphi) = |a_1|^2 + \sin 2\varphi(a_1, a_p) \leqslant |a_1|^2 + |(a_1, a_p)| = f(\varphi_p) = |b_1|^2,$$

и $|a_1|^2 < |b_1|^2$ при $(a_1, a_p) \neq 0$.

Если $|a_1|^2 - |a_p|^2 > 0$, то, поскольку $\varphi_p \in (-\pi/4, \pi/4)$, $f'(\varphi_p) = 0$, $f'(\pm\pi/4) = \mp[|a_1|^2 - |a_p|^2]$, функция $f(\varphi)$ строго выпукла вверх на $(-\pi/4, \pi/4)$, точка φ_p — точка максимума,

$$|b_1|^2 = f(\varphi_p) \geqslant f(0) = |a_1|^2,$$

неравенство строгое при $\varphi_p \neq 0$, т. е. при $(a_1, a_p) \neq 0$. Доказано следующее утверждение.

Свойство 2. Если $|a_1| \geqslant |a_p|$ и угол φ_p определяется правилом а)—в), то

$$|b_1| \geqslant |a_1| \geqslant |a_p| \geqslant |b_p|$$

и при $(a_1, a_p) \neq 0$ первое неравенство является строгим.

Равенство $|b_1| = |a_1|$ достигается только при $\varphi_p = 0$, т. е. когда $a_1 \perp a_p$, при этом $U_p(\varphi_p) = E$. Норма первого столбца не убывает при рассматриваемом преобразовании $A \to B$.

Свойство 3. Если $|a_1| \geqslant |a_p|$, то столбцы b_1 и b_p в матрице B ортогональны, $b_1 \perp b_p$.

Действительно, в силу выбора угла вращения φ_p имеем

$$\begin{aligned}(b_1, b_p) &= -cs(a_1, a_1) - s^2(a_p, a_1) + c^2(a_1, a_p) + cs(a_p, a_p) = \\ &= (c^2 - s^2)(a_1, a_p) - cs[|a_1|^2 - |a_p|^2] = \\ &= \cos 2\varphi_p(a_1, a_p) - \sin 2\varphi_p \frac{|a_1|^2 - |a_p|}{2} = 0,\end{aligned}$$

последнее равенство следует из а)—в).

2. Последовательности матриц

Пусть в вещественной матрице $A = (a_{ij})$ порядка n максимальную норму имеет m-й столбец. Обозначим через $U_1 = (u_{ij})$ ортогональную матрицу перестановки первого и m-го столбцов, именно $u_{1m} = u_{m1} = 1$, $u_{11} = u_{mm} = 0$, остальные диагональные элементы равны единице, внедиагональные — нулю. Первый столбец матрицы $A_1 = AU_1$ имеет наибольшую норму.

Обозначим $A_0 = A$ и рассмотрим последовательность матриц $A_k = (a_{ij}^k)$, где

$$\begin{gathered} A_1 = A_0 U_1, \quad A_2 = A_1 U_2(\varphi_2), \ldots, \quad A_{k+1} = A_k U_{p(k)}\big(\varphi_{p(k)}\big), \ldots, \\ p(k) = k \bmod (n-1) + 2, \quad k = 0, 1, 2, \ldots. \end{gathered} \tag{4}$$

Целочисленный индекс $p = p(k)$ изменяется циклически от 2 до n при возрастании k, в частности $p(n-2) = n$, $p(n-1) = 2$. Номер матрицы U_p обозначает номер столбца, который преобразуется вместе с первым столбцом, угол $\varphi_{p(k)}$ определяется матрицей A_k, $\varphi_{p(k)} = \varphi(p, A_k)$.

Итерационная последовательность A_k формируется умножением справа на матрицы простого поворота $U_p(\varphi_p)$, и угол φ_p определяется правилом а)—в). Столбцы матрицы A_{k+1} определяются равенствами (1), где $a_m^{k+1} = b_m$. Следовательно, выполняются свойства 1—3. В частности, для любой матрицы A_k, $k = 1, 2, \ldots$, норма первого столбца не меньше чем норма остальных её столбцов и монотонно не убывает при росте k.

Лемма 1. *Последовательность норм $|a_1^k|$ первых столбцов сходится при $k \to \infty$.*

Доказательство следует из ограниченности и монотонности последовательности норм.

Лемма 2. *Последовательность матриц A_k при $k \to \infty$ сходится поэлементно.*

Доказательство. Достаточно показать, что последовательность ортогональных матриц $U_p(\varphi_p)$, $p = p(k)$, сходится к единичной матрице, т. е. что угол $\varphi_{p(k)}$ стремится к нулю при $k \to \infty$.

Действительно, из (1) имеем

$$a_1^{k+1} = ca_1^k + sa_p^k, \quad a_p^{k+1} = -sa_1^k + ca_p^k, \quad a_m^{k+1} = a_m^k, \quad m \neq 1, p. \tag{5}$$

Если $\varphi(k) \to 0$ при $k \to \infty$, то для любого элемента a_{ij}^{k+1} матрицы A_{k+1} получаем, что

$$a_{ij}^{k+1} = \cos\varphi_{p(k)} a_{ij}^k + o(1),$$

и следовательно, ограниченная при $k \to \infty$ последовательность a_{ij}^{k+1} может иметь лишь единственную точку сгущения, т. е. сходится, если $\cos\varphi_{p(k)} \to 1$.

Покажем, что угол φ_p стремится к нулю. Используя формулы (5), получаем для $A_{k+1} = A_k U_{p(k)}\big(\varphi_{p(k)}\big)$, что

$$\begin{aligned}|a_1^{k+1}|^2 - \big|a_{p(k)}^{k+1}\big|^2 &= (c^2 - s^2)\big[|a_1^k|^2 - \big|a_{p(k)}^k\big|^2\big] + 4cs\big(a_1^k, a_{p(k)}^k\big) = \\ &= \cos 2\varphi_{p(k)}\big[|a_1^k|^2 - \big|a_{p(k)}^k\big|^2\big] + 2\sin 2\varphi_{p(k)}\big(a_1^k, a_{p(k)}^k\big) = \\ &= \cos 2\varphi_{p(k)}\big[|a_1^k|^2 - \big|a_{p(k)}^k\big|^2\big] + \big[|a_1^k|^2 - \big|a_{p(k)}^k\big|^2\big]\frac{\sin^2 2\varphi_{p(k)}}{\cos 2\varphi_{p(k)}},\end{aligned}$$

где последнее слагаемое получено согласно (3). Следовательно,

$$|a_1^{k+1}|^2 - \big|a_{p(k)}^{k+1}\big|^2 = \big[|a_1^k|^2 - \big|a_{p(k)}^k\big|^2\big]\frac{1}{\cos 2\varphi_{p(k)}}. \tag{6}$$

Покажем, что $\varphi_{p(k)} \to 0$ при $k \to \infty$. Последовательность $\varphi_{p(k)} \equiv \varphi(k)$ распадается на $n-1$ последовательность соответственно значениям индекса $p = p(k)$, $p = 2, 3, \ldots, n$.

Предположим противное, тогда одна из этих частичных последовательностей не стремится к нулю, например соответствующая столбцу с индексом q. Пусть номера её не стремящейся к нулю подпоследовательности есть k_i, $i = 1, 2, \ldots$. Тогда $|\varphi(k_i)| > \delta > 0$ при $k_i > k_0$,

$$\cos\varphi(k_i) < \cos\delta < 1, \quad k_i > k_0.$$

Из (6) следует неравенство

$$|a_1^{k+1}|^2 - |a_q^{k+1}|^2 > (|a_1^k|^2 - |a_q^k|^2)\frac{1}{\cos\delta}, \quad k = k_i > k_0.$$

Так как нормы $|a_1^m|$ не убывают, а $|a_q^m|$ не возрастают при росте m и так как $k_i > k_{i-1} + 1$, имеем

$$|a_1^{k_i}|^2 - |a_q^{k_i}|^2 \geqslant |a_1^{k_{i-1}+1}|^2 - |a_q^{k_{i-1}+1}|^2.$$

Используя предыдущее неравенство, имеем

$$|a_1^{k_i+1}|^2 - |a_q^{k_i+1}|^2 > (|a_1^{k_{i-1}+1}|^2 - |a_q^{k_{i-1}+1}|^2)\frac{1}{\cos\delta}.$$

Продолжая рекуррентно по i, получим

$$|a_1^{k_i+1}|^2 - |a_q^{k_i+1}|^2 > (|a_1^{k_1}|^2 - |a_q^{k_1}|^2)\frac{1}{(\cos\delta)^m},$$

где m — количество номеров k_j рассматриваемой подпоследовательности, когда номер k последовательности A_k изменяется от k_1 до k_i. Так как $m \to \infty$ при

$k_i \to \infty$, то правая часть неравенства неограниченно растёт вместе с m, а левая часть остаётся ограниченной. Полученное противоречие завершает доказательство леммы. $\square$

Итак, матричная последовательность A_k сходится поэлементно. Обозначим предельную матрицу A_∞,

$$\lim_{k\to\infty} A_k = A_\infty.$$

Следствие. *Первый столбец a_1^∞ предельной матрицы A_∞ является наибольшим по норме и уже не может быть увеличен предложенной итерационной процедурой. Отсюда следует, что этот столбец a_1^∞ ортогонален всем остальным, иначе можно было бы построить матрицу U_p, которая увеличила бы норму $|a_1^\infty|$.*

Обозначим

$$V_k = U_1 \prod_{m=0}^{k} U_{p(m)},$$

тогда $A_{k+1} = A_k U_{p(k)} = AV_k$. Из леммы 2 следует также сходимость ортогональных матриц V_k, обозначим их предельную ортогональную матрицу через V_∞,

$$A_\infty = AV_\infty.$$

3. Сингулярное разложение

Для вещественной матрицы $A = (a_{ij})$ рассмотрим алгоритм получения сингулярного разложения, в основе которого лежит построение последовательности (4), образованной по правилу а)—в).

Процедурой (2)—(5) будем называть построение последовательности (4) по формулам (5), максимизирующее первый столбец матрицы, где угол $\varphi_{p(k)}$ для матриц $U_{p(k)}$ определяется по правилу а)—в).

Эта процедура для A приводит к матрице A_∞, первый столбец которой ортогонален остальным; обозначим здесь эту матрицу через $A_\infty^{(1)}$ с первым столбцом a_∞^1, обозначим также $V_\infty^{(1)} := V_\infty$.

Для $A_\infty^{(1)}$ с исключённым первым столбцом может быть осуществлена процедура (2)—(5), т. е. максимизация второго столбца (мы предварительно поставим на его место максимальный по норме из последующих номеров). Полученную в пределе матрицу с прежним первым столбцом a_∞^1 обозначим $A_\infty^{(2)}$; её первые два столбца ортогональны остальным. Предельную ортогональную матрицу в этой второй процедуре будем обозначать $V_\infty^{(2)}$, т. е.

$$A_\infty^{(2)} = A_\infty^{(1)} V_\infty^{(2)} = AV_\infty^{(1)} V_\infty^{(2)}.$$

Для $A_\infty^{(2)}$ проведём максимизацию нормы третьего столбца, предельные матрицы обозначим через $A_\infty^{(3)}$ (её три первых столбца ортогональны всем остальным), ортогональную матрицу преобразования — через $V_\infty^{(3)}$.

Продолжая таким образом, получим после применения $n-1$ раза к матрице $A_\infty^{(1)}$ этой процедуры матрицу $A_\infty^{(n-1)} = \big(a_{ij}^{(n-1)}\big)$, все столбцы которой взаимно ортогональны, для которой выполняется равенство

$$A_\infty^{(n-1)} = AV, \tag{7}$$

где $V = V_\infty^{(1)} \dots V_\infty^{(n-1)}$ — ортогональная матрица.

Пусть $S = \operatorname{diag}(s_1, \dots, s_n)$, где s_k — норма k-го столбца матрицы $A_\infty^{(n-1)}$. Мы имеем $A_\infty^{(n-1)} = WS$, W — матрица со взаимно ортогональными нормированными столбцами, и для A получаем сингулярное разложение

$$A = WSV^{\mathrm{T}}. \tag{8}$$

4. Спектральная задача для симметричной матрицы

Пусть теперь A — симметричная матрица. Применяя $n-1$ раз процедуру (2)—(5), получим матрицу $A_\infty^{(n-1)}$ и равенство (7). Обозначим

$$D := \big(A_\infty^{(n-1)}\big)^{\mathrm{T}} A_\infty^{(n-1)},$$

т. е.

$$A^2 = VDV^{\mathrm{T}}. \tag{9}$$

Матрица D диагональная вследствие взаимной ортогональности столбцов матрицы $A_\infty^{(n-1)}$,

$$D = \operatorname{diag}(b_1, \dots, b_n), \quad b_k = s_k^2,$$

b_k — собственные числа симметричной матрицы A^2, столбцы матрицы V — собственные векторы w матрицы A^2.

Собственными числами λ_k матрицы A могут быть только числа $\pm\sqrt{b_k}$. Будем для краткости обозначать далее через μ и w собственные числа и векторы матрицы A^2, через λ и v — собственные числа и векторы A. Укажем правило получения собственных чисел и векторов матрицы A, используя равенства (8) и (9), которые определяют μ и w.

Возможны два случая.

1. μ — простое собственное число матрицы A^2. Тогда собственным числом матрицы A будет либо $\lambda = \sqrt{\mu}$, либо $\lambda = -\sqrt{\mu}$, и это собственное число простое, его собственная функция v является и собственной функцией для A^2, т. е. $v = w$. Знак собственного числа λ определяется равенством

$$Aw = (\operatorname{sgn}\lambda)\sqrt{\mu}w.$$

2. μ — кратное собственное число. Пусть $\lambda^+ = \sqrt{\mu}$, $\lambda^- = -\sqrt{\mu}$ — собственные числа A с собственными функциями $v_1^+, \dots, v_m^+$ и $v_1^-, \dots, v_p^-$ соответственно, при этом μ является $(m+p)$-кратным.

Любой собственный вектор w матрицы A^2 для собственного числа μ имеет вид $w = \alpha v^+ + \beta v^-$, где v^+ и v^- — линейные комбинации собственных векторов при λ^+ и λ^-. Для определения собственных векторов v^+ и v^- матрицы A рассмотрим равенства

$$Aw = \sqrt{\mu}(\alpha v^+ - \beta v^-), \quad w = \alpha v^+ + \beta v^-.$$

Отсюда

$$2\sqrt{\mu}\alpha v^+ = \sqrt{\mu}w + Aw, \quad 2\sqrt{\mu}\beta v^- = \sqrt{\mu}w - Aw.$$

Непосредственно проверяется, что векторы v^+ и v^- ортогональны.

Рассматривая $m + p$ раз ортогональные собственные векторы $w = w_k$ матрицы A^2 с собственным числом μ, получим указанным образом взаимно ортогональные векторы v_k^+ и v_k^- матрицы A и их собственные числа λ^+ и λ^-. Если одно из них не является собственным числом, то α или β равно нулю.

5. Численная реализация

На матрицах порядка $10 \div 30$, взятых из [4], данный алгоритм даёт все приведённые (десять—пятнадцать) десятичные знаки сингулярных или собственных чисел, кроме одного или двух последних.

Метод может быть применён для решения частичной проблемы SVD, для вычисления только наибольшего или наименьшего сингулярного числа матрицы, для определения вершины многогранника, в задаче сжатия цифровых изображений.

Литература

[1] Голуб Дж., Ван Лоун Ч. Матричные вычисления. — М.: Мир, 1999.

[2] Деммель Дж. Вычислительная линейная алгебра. Теория и приложения. — М.: Мир, 2001.

[3] Лежнёв В. Г. Метод решения алгебраической спектральной задачи // Численные методы анализа. — М.: Изд-во Моск. ун-та, 1995. — С. 16—22.

[4] Фаддеева В. Н., Колотилина Л. Ю. Вычислительные методы линейной алгебры. Набор матриц для тестирования. — Вып. I—III. — Ленинград: ЛОМИ АН СССР, 1982.

[5] Lezhnev V. G., Nesterenko A. G. Some estimate for generalized algebraic spectral problem // Spectral and Evaluational Problems; Proc. of the 7th Crimean Autumn Mathematical School-Symphosium. Vol. 7. — Simferopol, 1997. — P. 61—62.