Метод наименьших квадратов и сингулярное разложение

Н. Н. Осипов

e-mail: nnosipov@rambler.ru

Аннотация

Статья представляет собой более подробную версию открытой лекции, прочитанной автором в мае 2016 года студентам-первокурсникам специальности «Компьютерная безопасность».

I. Рассмотрим, вообще говоря, *переопределённую* систему линейных уравнений:

$$\sum_{j=1}^{n} a_{ij}x_j = b_i, \quad i = 1, \ldots, m, \tag{0.1}$$

где $m \geqslant n$. Эту СЛУ можно записать в виде уравнения

$$\mathcal{A}(x) = b, \tag{0.2}$$

где $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$, $b = (b_1, \ldots, b_m) \in \mathbb{R}^m$ и

$$\mathcal{A} : \mathbb{R}^n \to \mathbb{R}^m$$

— *линейное отображение*, заданное матрицей $A = (a_{ij}) \in M_{m\times n}(\mathbb{R})$ в стандартных базисах арифметических векторных пространств $\mathbb{R}^n$ и $\mathbb{R}^m$.[1)] Пусть

$$a_j = \mathcal{A}(e_j) \in \mathbb{R}^m, \quad j = 1, \ldots, n,$$

— векторы, столбцы координат которых совпадают со столбцами матрицы A. Тогда уравнение (0.2) примет вид

$$x_1a_1 + \ldots + x_na_n = b.$$

Для простоты далее будем считать, что отображение $\mathcal{A}$ невырождено:

$$\operatorname{Ker}\mathcal{A} = \{0\} \tag{0.3}$$

или, в других терминах, матрица A имеет полный ранг (система векторов $(a_1, \ldots, a_n)$ линейно независима). При $m > n$ (и, тем более, при $m \gg n$, как часто бывает на практике) уравнение (0.2) обычно неразрешимо, ибо почти всегда вектор b не принадлежит подпространству

$$L = \operatorname{Im}\mathcal{A} = \langle a_1, \ldots, a_n\rangle.$$

Задача состоит в том, чтобы найти некоторое *псевдорешение*

$$x^0 = (x_1^0, \ldots, x_n^0) \in \mathbb{R}^n$$

[1)]Напомним, что стандартным базисом арифметического n-мерного векторного пространства $\mathbb{R}^n$ называется базис $(e_1, \ldots, e_n)$ из так называемых *единичных векторов* (все компоненты вектора e_j равны нулю, кроме j-й, которая равна 1).

уравнения (0.2) и тем самым как-то «решить» несовместную СЛУ (0.1).

Будем считать пространства $\mathbb{R}^n$ и $\mathbb{R}^m$ евклидовыми.[2)] В *методе наименьших квадратов* псевдорешение $x^0 \in \mathbb{R}^n$ находят, исходя из следующего условия:

$$\min_{x \in \mathbb{R}^n} |b - \mathcal{A}(x)|^2 = |b - \mathcal{A}(x^0)|^2.$$

Это условие объясняет название самого метода, так как величина $|b - \mathcal{A}(x)|^2$, которую требуется минимизировать, есть сумма квадратов

$$\sum_{i=1}^{m} \left(b_i - \sum_{j=1}^{n} a_{ij} x_j \right)^2 .$$

Метод наименьших квадратов широко применяется как в прикладной математике, так и в статистике.[3)]

Имеем $b = b' + b''$, где $b' \in L$ — *ортогональная проекция* вектора b на подпространство L, а $b'' \in L^\perp$. Пусть $a = \mathcal{A}(x) = x_1 a_1 + \ldots + x_n a_n \in L$. Тогда по *теореме Пифагора*

$$|b - \mathcal{A}(x)|^2 = |b - a|^2 = |(b - b') + (b' - a)| = |b - b'|^2 + |b' - a|^2 \geqslant |b - b'|^2,$$

причём равенство возможно только если $a = b'$. Итак, псевдорешение x^0 должно удовлетворять уравнению

$$x_1 a_1 + \ldots + x_n a_n = b'. \tag{0.4}$$

Из условия (0.3) следует, что это уравнение разрешимо и имеет единственное решение.

Стандартный теоретический способ решить уравнение (0.4) состоит в том, чтобы решить так называемую *нормальную* СЛУ

$$\sum_{j=1}^{n} (a_i, a_j) x_j = (a_i, b), \quad i = 1, \ldots, n, \tag{0.5}$$

которая получится, если для каждого $i = 1, \ldots, n$ обе части уравнения (0.4) скалярно домножить на вектор a_i и учесть, что $(a_i, b') = (a_i, b)$. Матрица системы (0.5) есть *матрица Грама*

$$G = ((a_i, a_j)) = A^t A \in M_{n \times n}(\mathbb{R})$$

системы векторов $(a_1, \ldots, a_n)$. Поскольку последняя является базисом подпространства L, имеем $\det G > 0$.[4)] Таким образом, псевдорешение x^0 можно найти по формуле

$$\begin{bmatrix} x_1^0 \\ \vdots \\ x_n^0 \end{bmatrix} = G^{-1} A^t \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix} . \tag{0.6}$$

[2)]Термин «евклидово пространство $\mathbb{R}^n$» в данном случае означает, что в пространстве $\mathbb{R}^n$ задано стандартное скалярное произведение:

$$(x, y) = x_1 y_1 + \ldots + x_n y_n,$$

где $x = (x_1, \ldots, x_n)$, $y = (y_1, \ldots, y_n)$.

[3)]Подробнее см. по ссылке [3].

[4)]Действительно, пусть $(w_1, \ldots, w_n)$ — какой-нибудь ортонормированный базис L. Тогда $G = T^t T$, где T — *матрица перехода* от $(w_1, \ldots, w_n)$ к $(a_1, \ldots, a_n)$. Следовательно, $\det G = (\det T)^2 > 0$.

II. Насколько хорош описанный выше способ отыскания псевдорешения с практической точки зрения?

Дело в том, что как матрица A, так и вектор b обычно заданы с какой-то погрешностью:

$$A \approx \hat{A}, \quad b \approx \hat{b}.$$

Возникает вопрос, насколько сильно может отличаться фактически вычисленное псевдорешение $\hat{x}^0$, соответствующее матрице $\hat{A}$ и вектору $\hat{b}$, от истинного псевдорешения x^0.[5)]

В том, что применение формулы (0.6) на практике может оказаться небезобидным, показывает уже тривиальный случай $m = n$. В этом случае

$$G^{-1}A^t = (A^tA)^{-1}A^t = A^{-1},$$

и формула (0.6) упрощается до традиционной:

$$\begin{bmatrix} x_1^0 \\ \vdots \\ x_n^0 \end{bmatrix} = A^{-1} \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix}.$$

Проводить вычисления по этой формуле надёжнее, чем по формуле (0.6), поскольку лишние умножения матриц, вообще говоря, способствуют росту погрешности.

III. Опишем другой способ отыскания псевдорешения x^0, основанный на так называемом *сингулярном разложении* матрицы A, которое имеет вид

$$A = U\Sigma V^t. \tag{0.7}$$

Здесь $U \in M_{m\times m}(\mathbb{R})$, $V \in M_{n\times n}(\mathbb{R})$ — *ортогональные матрицы*, $\Sigma \in M_{m\times n}(\mathbb{R})$ — *диагональная прямоугольная матрица*.[6)]

Объясним, как можно прийти к равенству (0.7). Рассмотрим линейный оператор

$$\mathcal{G} = \mathcal{A}^*\mathcal{A} : \mathbb{R}^n \to \mathbb{R}^n,$$

заданный матрицей G в стандартном базисе. Оператор $\mathcal{G}$ является *самосопряжённым*, поскольку $G^t = G$. Следовательно, существует ортонормированный базис $(v_1, \ldots, v_n)$ пространства $\mathbb{R}^n$, состоящий из собственных векторов этого оператора:

$$\mathcal{G}(v_j) = \lambda_j v_j, \quad j = 1, \ldots, n.$$

Этот базис называется *первым сингулярным базисом* матрицы A.[7)] Кроме того, имеем

$$\lambda_j = (v_j, \mathcal{G}(v_j)) = (v_j, \mathcal{A}^*(\mathcal{A}(v_j))) = (\mathcal{A}(v_j), \mathcal{A}(v_j)) = |\mathcal{A}(v_j)|^2 > 0$$

5) Отметим, что и сам процесс вычисления псевдорешения $\hat{x}^0$ происходит с какой-то погрешностью (например, имеют место ошибки округления), которую также необходимо учитывать.

6) Напомним, что квадратная вещественная матрица C называется ортогональной, если $C^{-1} = C^t$. Подробнее о Singular Value Decomposition (SVD) и современных методах его нахождения см. по ссылке [4].

7) Или линейного отображения $\mathcal{A}$. Все результаты о сингулярном разложении можно излагать как на языке матриц, так и на языке линейных отображений (операторов).

для любого $j = 1, \ldots, n$. Числа

$$\sigma_j = \sqrt{\lambda_j} = |\mathcal{A}(v_j)|, \quad j = 1, \ldots, n,$$

называются *сингулярными числами* матрицы A. Далее положим

$$u_j = \frac{\mathcal{A}(v_j)}{\sigma_j}, \quad j = 1, \ldots, n.$$

Система векторов $(u_1, \ldots, u_n)$ является ортонормированным базисом подпространства L:

$$(u_i, u_j) = \frac{1}{\sigma_i \sigma_j}\,(\mathcal{A}(v_j), \mathcal{A}(v_j)) = \frac{1}{\sigma_i \sigma_j}\,(v_i, \mathcal{G}(v_j)) = \frac{\lambda_j}{\sigma_i \sigma_j}\,(v_i, v_j) = \begin{cases} 1, & \text{если } i = j, \\ 0, & \text{если } i \neq j. \end{cases}$$

Дополним её до ортонормированного базиса $(u_1, \ldots, u_m)$ всего пространства $\mathbb{R}^m$. Он называется *вторым сингулярным* базисом матрицы A.

Теперь составим ортогональные матрицы U и V из столбцов координат векторов второго и первого сингулярных базисов соответственно, а диагональную матрицу Σ заполним сингулярными числами. Тогда

$$U\Sigma = AV$$

по построению, откуда, домножив справа на $V^{-1} = V^t$, получим (0.7).

Зная сингулярное разложение (0.7) матрицы A, псевдорешение x^0 СЛУ (0.1) можно найти по формуле

$$x^0 = \sum_{j=1}^{n} \frac{(u_j, b)}{\sigma_j}\, v_j. \tag{0.8}$$

Для доказательства формулы (0.8) достаточно проверить, что она даёт решение уравнения (0.4). В самом деле, имеем

$$x_1^0 a_1 + \ldots + x_n^0 a_n = \mathcal{A}(x^0) = \sum_{j=1}^{n} \frac{(u_j, b)}{\sigma_j} \mathcal{A}(v_j) = \sum_{j=1}^{n} (u_j, b) u_j = b'.$$

IV. Вернёмся к вопросу о погрешности при вычислении псевдорешения x^0, которая может возникнуть из-за того, что матрица A и вектора b сами заданы с некоторой погрешностью.

Для простоты будем считать, что матрица A задана точно ($\hat{A} = A$), а возмущению подвергается только вектор b. Упорядочим сингулярные числа матрицы A по невозрастанию:

$$\sigma_1 \geqslant \ldots \geqslant \sigma_n > 0.$$

Пусть теперь $\hat{b} = b + \varepsilon u_n$, где коэффициент ε — малое число. Из формулы (0.8) находим

$$\hat{x}^0 = x^0 + \frac{\varepsilon}{\sigma_n}\, v_n$$

при этом коэффициент ε/σ_n может оказаться существенным, если сингулярное число σ_n мало, а коэффициент ε соизмерим с σ_n. Таким образом, при практическом вычислении следует учитывать, что псевдорешение x^0 будет особенно чувствительно к возмущениям вектора b вдоль тех векторов второго сингулярного базиса матрицы A, которые соответствуют малым сингулярным числам.

Впрочем, практический интерес представляет оценка не абсолютной, а относительной погрешности вычисления x^0, а именно, сравнение последней с той относительной погрешностью, с которой задан вектор b. Пусть

$$\Delta = \frac{|\hat{x}^0 - x^0|}{|x^0|}, \quad \delta = \frac{|\hat{b} - b|}{|b|}.$$

С помощью формулы (0.8) нетрудно получить оценку

$$\frac{\Delta}{\delta} \leqslant \frac{|b|}{|b'|} \cdot \frac{|u'|}{|u|} \cdot \frac{\sigma_1}{\sigma_n},$$

где $u = \hat{b} - b = u' + u''$ ($u' \in L$, $u'' \in L^\perp$). При $b \in L$ и $u \in L$ эта оценка упрощается до

$$\frac{\Delta}{\delta} \leqslant \frac{\sigma_1}{\sigma_n}.$$

Отношение σ_1/σ_n называется *числом обусловленности* матрицы A и обозначается $\mu(A)$.[8)] Ясно, что чем больше $\mu(A)$, тем более неустойчивым будет вычисление псевдорешения x^0.

Отметим, что сингулярное разложение матрицы G можно получить из сингулярного разложения (0.7) матрицы A. Оно имеет вид

$$G = V\Lambda V^t,$$

при этом на диагонали матрицы

$$\Lambda = \Sigma^t\Sigma \in M_{n\times n}(\mathbb{R})$$

находятся собственные числа $\lambda_j = \sigma_j^2$ матрицы G. Поэтому $\mu(G) = \mu(A)^2 \gg \mu(A)$, если $\mu(A)$ велико. Это значит, что в этом случае вычисление псевдорешения x^0 по формуле (0.6) будет гораздо более неустойчивым, чем по формуле (0.8).

Замечание. 1. Подробное изложение различных практических способов поиска псевдорешений читатель может найти, например, в учебнике [1], глава IV. Там же (см. § 4 главы IV) можно узнать о применении метода наименьших квадратов в статистике (*линейная регрессия*).

2. При $m = n$ из сингулярного разложения (0.7) матрицы A можно получить её *полярное разложение* (см. учебник [1], §§ 1 — 3 главы I) в виде

$$A = SR, \tag{0.9}$$

где $S = U\Sigma U^t$, $R = UV^t$. Последние формулы также показывают, как от полярного разложения (0.9) матрицы A прийти к её сингулярному разложению.[9)]

[8)] Число $\mu(A)$ естественным образом возникает при оценке относительной погрешности решения невырожденной системы линейных уравнений (см., например, [2], стр. 14).

[9)] Такая тесная связь между полярным и сингулярным разложениями матрицы неудивительна, поскольку конструкция обоих разложений основана на сингулярных базисах.

Список литературы

[1] *Беклемишев Д.В.* Дополнительные главы линейной алгебры. М.: Наука, 1983.

[2] *Годунов С.К.* Решение систем линейных уравнений. Новосибирск: Наука, 1980.

[3] https://en.wikipedia.org/wiki/Linear_least_squares_(mathematics)

[4] https://en.wikipedia.org/wiki/Singular_value_decomposition