Записки научных
семинаров ПОМИ
Том 499, 2021 г.

В. Н. Малозёмов, Г. Ш. Тамасян

ФАКТОРИЗАЦИЯ МАТРИЦЫ ОРТОГОНАЛЬНОГО ПРОЕКТИРОВАНИЯ НА ПОДПРОСТРАНСТВО

Факторизация матриц является одним из основных методов построения быстрых алгоритмов. В данной заметке получена глубокая факторизация матрицы ортогонального проектирования на подпространство. Такие матрицы играют важную роль в методах условной оптимизации. В качестве промежуточного используется LQ-разложение, где L – нижнетреугольная матрица с положительными диагональными элементами и Q – ортогональная матрица. Для построения матрицы Q привлекается метод последовательного понижения ранга матрицы.

§1

Пусть m, n – натуральные числа, $M = 1 : m$, $N = 1 : n$ – индексные множества и $A = A[M, N]$ – матрица с линейно независимыми строками. Строки матрицы A будем обозначать так:

$$A[i, N] \quad \text{или} \quad a_i^T,\ i \in M.$$

Рассмотрим в пространстве $\mathbb{R}^n$ линейное подпространство

$$\mathscr{L} = \{x \in \mathbb{R}^n \mid Ax = \mathbf{0}\} = \big\{x \in \mathbb{R}^n \mid \langle a_i, x\rangle = 0,\ i \in M\big\}.$$

Ортогональной проекцией точки $c \in \mathbb{R}^n$ на подпространство $\mathscr{L}$ называется решение экстремальной задачи:

$$\textit{минимизировать } \|x - c\| \textit{ по всем } x \in \mathscr{L}.$$

Как известно (см., например, [1, с. 52–55]), решение x_* этой задачи существует, единственно и допускает представление $x_* = Pc$, где

$$P = E - A^T(AA^T)^{-1}A. \tag{1}$$

Ключевые слова: ортогональное проектирование, факторизация матрицы, QR-разложение, LQ-разложение.

Результаты раздела 7 получены в Институте проблем машиноведения РАН при финансовой поддержке Российского научного фонда (проект No. 20-71-10032).

Здесь E – единичная матрица порядка n. Матрица P называется *матрицей ортогонального проектирования на подпространство* $\mathscr{L}$. Отметим, что линейная независимость строк матрицы A существенна для обратимости матрицы AA^T.

При $m = 1$ подпространство $\mathscr{L}$ является гиперплоскостью, определяемой соотношением $\langle a, x\rangle = 0$. Линейная независимость строк матрицы A в данном случае соответствует условию $a \neq \mathbf{0}$. Обозначим $\widehat{a} = a/\|a\|$. Тогда матрица P примет вид

$$P = E - \widehat{a}\,\widehat{a}^T. \tag{2}$$

Теорема 1. *Если строки* $a_1^T, a_2^T, \ldots, a_m^T$ *матрицы* A *ортонормированы, то матрицу ортогонального проектирования* P *можно разложить на множители* (*факторизовать*):

$$P = \prod_{i=1}^{m}(E - a_i a_i^T). \tag{3}$$

Произведение не зависит от порядка сомножителей.

Доказательство. В условиях теоремы произведение AA^T есть единичная матрица порядка m. Формула (1) упрощается:

$$P = E - A^T A. \tag{4}$$

При $m = 1$ равенство (3) следует как из (4), так и из (2).

Сделаем индукционный переход от m к $m+1$. Воспользуемся свойством произведения двух матриц, которое с помощью индексной техники можно записать так:

$$B[N, M] \times C[M, N] = \sum_{i\in M} B[N, i] \times C[i, N]. \tag{5}$$

Это значит, что произведение двух матриц равно сумме матриц, каждая из которых есть произведение вектора-столбца на вектор-строку. Согласно (4) и (5) имеем

$$P = E - a_1 a_1^T - a_2 a_2^T - \ldots - a_m a_m^T. \tag{6}$$

Пусть a_{m+1}^T – нормированная вектор-строка, ортогональная всем строкам матрицы A. Нужно проверить справедливость следующего равенства

$$P - a_{m+1}a_{m+1}^T = \Big(\prod_{i=1}^{m}(E - a_i a_i^T)\Big)\left(E - a_{m+1}a_{m+1}^T\right),$$

которое в силу индукционного предположения (3) можно переписать в виде

$$P a_{m+1} a_{m+1}^T = a_{m+1} a_{m+1}^T. \tag{7}$$

Подробная запись левой части этого равенства выглядит так:

$$(E - a_1 a_1^T)(E - a_2 a_2^T) \dots (E - a_m a_m^T) a_{m+1} a_{m+1}^T.$$

Производя последовательные умножения справа налево и учитывая ортогональность вектора a_{m+1} всем векторам $a_m, a_{m-1}, \dots, a_1$, приходим к равенству (7). Формула (3) установлена.

То, что сомножители в произведении из правой части формулы (3) можно переставлять, следует из формулы (6).

Теорема доказана. □

Следствие. *В случае ортонормированности строк матрицы A проекцию x_* точки c на подпространство $\mathscr{L}$ можно получить, последовательно проектируя (в любом порядке) эту точку на гиперплоскости*

$$\mathscr{L}_i = \big\{ x \in \mathbb{R}^n \mid \langle a_i, x \rangle = 0 \big\}, \quad i \in 1 : m.$$

§2

Предположим теперь, что строки матрицы A лишь линейно независимы. Покажем, что и в этом случае матрица ортогонального проектирования P допускает разложение на простые множители.

Нам потребуется следующее утверждение.

Теорема 2. *Матрицу $A[M, N]$ с линейно независимыми строками можно представить в виде*

$$A[M, N] = L[M, M] \times Q[M, N], \tag{8}$$

где L – нижнетреугольная матрица с положительными диагональными элементами и Q – матрица с ортонормированными строками.

Формула (8) непосредственно связана с QR-разложением матрицы [2–4]. В §7 будет дано прямое доказательство теоремы 2 на основе метода последовательного понижения ранга матрицы [1, с. 24–30].

Ясно, что L – обратимая матрица и что QQ^T – единичная матрица порядка m. Эти свойства вместе с формулой (8) позволяют доказать, что

$$P = E - Q^T Q. \tag{9}$$

Действительно, согласно (1), имеем

$$
\begin{aligned}
P &= E - Q^T L^T \big(L \big(Q Q^T \big) L^T \big)^{-1} L Q \\
&= E - Q^T L^T \big(L^T \big)^{-1} L^{-1} L Q = E - Q^T Q.
\end{aligned}
$$

Обозначим строки матрицы Q через x_1^T, $x_2^T, \ldots, x_m^T$. На основании теоремы 1 получаем

$$
P = \prod_{i=1}^{m} (E - x_i x_i^T).
$$

Произведение не зависит от порядка сомножителей.

§3

Приводимая ниже лемма играет решающую роль при доказательстве теоремы 2.

Лемма. *Пусть $B = B[N, M]$ – прямоугольная матрица ранга $r \geqslant 1$. Введём матрицу*

$$
F = B - x y^T,
$$

где $x \in \mathbb{R}^n$, $y \in \mathbb{R}^m$. Ранг матрицы F равен $r - 1$, если векторы x, y допускают представление

$$
x = B u_0, \qquad y = B^T v_0, \tag{10}
$$

причём

$$
\langle B u_0, v_0 \rangle = 1. \tag{11}
$$

Доказательство. Согласно (10) имеем

$$
F = B - B u_0 v_0^T B.
$$

Отсюда и из (11), в частности, следует, что $F u_0 = \mathbf{0}$. Кроме u_0, системе $F u = \mathbf{0}$ удовлетворяет любое решение системы $B u = \mathbf{0}$. Так как $\operatorname{rank} B = r$ и $u \in \mathbb{R}^m$, то система $B u = \mathbf{0}$ имеет $m - r$ линейно независимых решений u_1, $u_2, \ldots, u_{m-r}$. Получили, что

$$
F u_i = \mathbf{0}, \quad i \in 0 : m - r.
$$

В силу (11), $B u_0 \neq \mathbf{0}$. Это гарантирует линейную независимость системы векторов u_0, $u_1, \ldots, u_{m-r}$.

Как известно, максимальное число линейно независимых решений системы $F u = \mathbf{0}$ равно $m - \operatorname{rank} F$. Значит,

$$
m - \operatorname{rank} F \geqslant m - r + 1.
$$

Приходим к неравенству $\operatorname{rank} F \leqslant r-1$.

Вместе с тем,

$$r = \operatorname{rank} B = \operatorname{rank}\left(F + xy^T\right) \leqslant \operatorname{rank} F + \operatorname{rank}\left(xy^T\right).$$

Равенство $\langle Bu_0, v_0\rangle = \langle u_0, B^T v_0\rangle$ позволяет переписать условие (11) так:

$$\langle x, v_0\rangle = \langle u_0, y\rangle = 1.$$

Отсюда следует, что векторы x и y ненулевые, так что $\operatorname{rank}\left(xy^T\right) = 1$. Приходим к обратному неравенству $\operatorname{rank} F \geqslant r-1$. В результате получаем равенство $\operatorname{rank} F = r-1$.

Лемма доказана. □

Замечание. Условия леммы являются не только достаточными, но и необходимыми для того, чтобы ранг матрицы F равнялся $r-1$.

§4

Опишем конкретный способ понижения ранга матрицы B.

Возьмём любой ненулевой столбец $b_j = B[N, j]$. Положим

$$x = \frac{1}{\|b_j\|} b_j, \qquad y = B^T x.$$

Это соответствует тому, что $u_0 = \frac{1}{\|b_j\|} e_j$, где e_j – j-й орт, и $v_0 = x$. Так как

$$\langle Bu_0, v_0\rangle = \langle x, x\rangle = 1,$$

то по лемме ранг матрицы

$$F := B - xy^T = \left(E - xx^T\right) B \tag{12}$$

равен $r-1$.

На основании (12) и (2) можно утверждать, что столбцы матрицы F являются проекциями соответствующих столбцов матрицы B на гиперплоскость $\{v \in \mathbb{R}^n \mid \langle x, v\rangle = 0\}$. При этом ненулевой столбец $B[N, j]$ переходит в нулевой столбец $F[N, j]$. Действительно,

$$F[N, j] = \left(E - xx^T\right) b_j = \left(E - \frac{1}{\|b_j\|^2} b_j b_j^T\right) b_j = \mathbf{0}.$$

§5

Теперь будем считать, что ранг матрицы $B = B[N, M]$ равен m, то есть что столбцы матрицы B линейно независимы. Опишем метод последовательного понижения ранга матрицы B.

Начальный шаг. Положим $F_0 = B$.

k-й шаг. Имеем матрицу F_{k-1}. Берём k-й столбец $f_k = F_{k-1}[N, k]$. Вычисляем

$$x_k = \frac{1}{\|f_k\|} f_k, \qquad y_k = F_{k-1}^T x_k,$$
$$F_k = F_{k-1} - x_k y_k^T = \big(E - x_k x_k^T\big) F_{k-1}. \tag{13}$$

Эти преобразования проведём при $k = 1, 2, \ldots, m$. В результате получим

$$F_m = F_0 - \sum_{k=1}^{m} x_k y_k^T. \tag{14}$$

Сделаем общие замечания по поводу описанного алгоритма:

- нулевые столбцы матрицы F_{k-1} переходят в нулевые столбцы матрицы F_k;
- нулевые столбцы матрицы F_{k-1} порождают нулевые компоненты вектора y_k;
- справедливо равенство $y_k[k] = \langle f_k, x_k \rangle = \|f_k\|$.

§6

Проследим подробнее за первыми двумя шагами алгоритма.

В силу условия $\operatorname{rank} B = m$ все столбцы матрицы F_0 ненулевые. При $k = 1$ возьмём первый столбец $f_1 = F_0[N, 1]$ и вычислим

$$x_1 = \frac{1}{\|f_1\|} f_1, \quad y_1 = F_0^T x_1, \quad F_1 = F_0 - x_1 y_1^T.$$

По теории $\operatorname{rank} F_1 = m - 1$ и первый столбец матрицы F_1 нулевой. В частности, все столбцы матрицы F_1, начиная со второго, ненулевые. Отметим также, что $y_1[1] = \|f_1\| > 0$.

При $k = 2$ берём ненулевой столбец $f_2 = F_1[N, 2]$ и вычисляем

$$x_2 = \frac{1}{\|f_2\|} f_2, \quad y_2 = F_1^T x_2, \quad F_2 = F_1 - x_2 y_2^T.$$

По теории у матрицы F_2 первые два столбца нулевые и $\operatorname{rank} F_2 = m-2$. В частности, все столбцы матрицы F_2, начиная с третьего, ненулевые. Отметим также, что $y_2[1] = 0$, $y_2[2] = \|f_2\| > 0$.

Теперь обратим внимание на следующее свойство:

$$F_1^T x_1 = \left(F_0^T - y_1 x_1^T\right) x_1 = \mathbf{0}\,. \tag{15}$$

Вычислим вторую компоненту вектора $F_1^T x_1$. Так как $F_1[N,2] = f_2 = \|f_2\| x_2$, то $F_1^T[2,N] \times x_1[N] = \|f_2\| x_2^T x_1$. Согласно (15) получаем $\langle x_2, x_1\rangle = 0$, так что нормированные векторы x_1, x_2 ортогональны.

Пусть уже имеется матрица F_{k-1} ранга $m-k+1$ с $k-1$ нулевыми первыми столбцами и ненулевыми остальными столбцами. При этом построены векторы $x_1, \ldots, x_{k-1}$, нормированные и попарно ортогональные. Возьмём ненулевой столбец $f_k = F_{k-1}[N,k]$ и по формулам (13) вычислим x_k, y_k, F_k. По теории справедливы следующие утверждения:

- ранг матрицы F_k равен $m-k$;
- первые k столбцов матрицы F_k нулевые, а остальные – ненулевые;
- первые $k-1$ компонент вектора y_k равны нулю, а k-я компонента положительна ($y_k[k] = \|f_k\| > 0$).

Дополнительно покажем, что вектор x_k ортогонален всем векторам $x_1, \ldots, x_{k-1}$. При $j \in 1 : k-1$ в силу попарной ортогональности векторов $x_1, \ldots, x_{k-1}$ имеем

$$F_{k-1}^T x_j = \Big(F_{j-1}^T - \sum_{i=j}^{k-1} y_i x_i^T\Big) x_j = y_j - y_j \|x_j\|^2 = \mathbf{0}\,. \tag{16}$$

Вычислим k-ю компоненту вектора $F_{k-1}^T x_j$. Так как $F_{k-1}[N,k] = f_k = \|f_k\| x_k$, то $F_{k-1}^T[k,N] \times x_j[N] = \|f_k\| x_k^T x_j$. Согласно (16) получаем $\langle x_k, x_j\rangle = 0$ при всех $j \in 1 : k-1$, так что нормированные векторы $x_1, \ldots, x_{k-1}, x_k$ попарно ортогональны.

У матрицы F_{m-1} ненулевым будет только последний столбец. Матрица F_m – нулевая.

Вернёмся к формуле (14). Вспоминая, что $F_0 = B$, приходим к представлению

$$B = \sum_{k=1}^{m} x_k y_k^T. \tag{17}$$

Обозначим через X матрицу со столбцами $x_1, x_2, \ldots, x_m$ и через Y –матрицу со столбцами $y_1, y_2, \ldots, y_m$. Тогда формулу (17) можно переписать в виде

$$B = XY^T. \tag{18}$$

По построению матрица Y является нижнетреугольной с положительными диагональными элементами, поэтому Y^T – верхнетреугольная матрица с положительными диагональными элементами. Матрица X состоит из ортонормированных столбцов. Формула (18) называется QR-разложением прямоугольной матрицы B с линейно независимыми столбцами.

§7

Переходим к доказательству теоремы 2. Пусть $A = A[M, N]$ – прямоугольная матрица с линейно независимыми строками. Перепишем алгоритм (13) в терминах транспонированных матриц.

Начальный шаг. Положим $F_0^T = A$.

k-й шаг. Имеем матрицу F_{k-1}^T. Берём k-ю строку $f_k^T = F_{k-1}^T[k, N]$ и вычисляем

$$x_k^T = \frac{1}{\|f_k\|} f_k^T, \qquad y_k = F_{k-1}^T x_k,$$
$$F_k^T = F_{k-1}^T - y_k x_k^T.$$

Выполнив указанное преобразование при $k = 1, 2, \ldots, m$, получим

$$F_m^T = F_0^T - \sum_{k=1}^{m} y_k x_k^T.$$

Учитывая, что F_m^T – нулевая матрица и $F_0^T = A$, приходим к разложению

$$A = \sum_{k=1}^{m} y_k x_k^T =: L \cdot Q. \tag{19}$$

Здесь L – нижнетреугольная $(m \times m)$-матрица с положительными диагональными элементами и Q – $(m \times n)$-матрица с ортонормированными строками. Формула (19) называется LQ-разложением прямоугольной матрицы A с линейно независимыми строками.

Обозначим $A_k = F_k^T$. Тогда алгоритму LQ-разложения можно придать другую – рабочую форму.

Начальный шаг. Положим $A_0 = A$.

k-й шаг. Имеем матрицу A_{k-1}. Берём k-ю строку $a_k^T = A_{k-1}[k, N]$ и вычисляем

$$\begin{gathered} x_k^T = \frac{1}{\|a_k\|} a_k^T, \qquad y_k = A_{k-1} x_k, \\ A_k = A_{k-1} - y_k x_k^T. \end{gathered} \tag{20}$$

Выполнив указанные преобразования при $k = 1, 2, \ldots, m$, получим LQ-разложение (19).

§8

Рассмотрим простой пример на LQ-разложение. Возьмём (3×4)-матрицу A с линейно независимыми строками,

$$A = \begin{pmatrix} 1 & -1 & -1 & -1 \\ 1 & 2 & 2 & -1 \\ 1 & 0 & 1 & 0 \end{pmatrix}.$$

Выполним преобразования по формулам (20).

Начальный шаг. Полагаем $A_0 = A$.

Первый шаг. Берём первую строку матрицы A_0,

$$a_1^T = \begin{pmatrix} 1 & -1 & -1 & -1 \end{pmatrix},$$

и вычисляем

$$x_1^T = \frac{1}{\|a_1\|} a_1^T = \begin{pmatrix} \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} \end{pmatrix}, \qquad y_1 = A_0 x_1 = \begin{pmatrix} 2 \\ -1 \\ 0 \end{pmatrix},$$

$$A_1 = A_0 - y_1 x_1^T = \begin{pmatrix} 0 & 0 & 0 & 0 \\ \frac{3}{2} & \frac{3}{2} & \frac{3}{2} & -\frac{3}{2} \\ 1 & 0 & 1 & 0 \end{pmatrix}.$$

Второй шаг. Берём вторую строку матрицы A_1,

$$a_2^T = \begin{pmatrix} \frac{3}{2} & \frac{3}{2} & \frac{3}{2} & -\frac{3}{2} \end{pmatrix},$$

и вычисляем

$$x_2^T = \frac{1}{\|a_2\|} a_2^T = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & -\frac{1}{2} \end{pmatrix}, \qquad y_2 = A_1 x_2 = \begin{pmatrix} 0 \\ 3 \\ 1 \end{pmatrix},$$

$$A_2 = A_1 - y_2 x_2^T = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \frac{1}{2} & -\frac{1}{2} & \frac{1}{2} & \frac{1}{2} \end{pmatrix}.$$

Третий шаг. Берём третью строку матрицы A_2,

$$a_3^T = \begin{pmatrix} \frac{1}{2} & -\frac{1}{2} & \frac{1}{2} & \frac{1}{2} \end{pmatrix},$$

и вычисляем

$$x_3^T = \frac{1}{\|a_3\|} a_3^T = \begin{pmatrix} \frac{1}{2} & -\frac{1}{2} & \frac{1}{2} & \frac{1}{2} \end{pmatrix}, \qquad y_3 = A_2 x_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix},$$

A_3 – нулевая матрица.

Приходим к LQ-разложению, $A = L \cdot Q$, где

$$L = \begin{pmatrix} 2 & 0 & 0 \\ -1 & 3 & 0 \\ 0 & 1 & 1 \end{pmatrix}, \qquad Q = \frac{1}{2} \begin{pmatrix} 1 & -1 & -1 & -1 \\ 1 & 1 & 1 & -1 \\ 1 & -1 & 1 & 1 \end{pmatrix}.$$

Список литературы

1. В. Н. Малозёмов, *Линейная алгебра без определителей. Квадратичная функция.* СПб.: Изд-во СПбГУ, 1997.
2. Р. Хорн, Ч. Джонсон, *Матричный анализ.* М.: Мир, 1989.
3. G. H. Golub, C. F. Van Loan, *Matrix Computations.* (4th ed.) — The Johns Hopkins University Press, Baltimore (2013).
4. Линеал. Базовая электронная энциклопедия по линейной алгебре. [http://lineal.guru.ru/].

Malozemov V. N., Tamasyan G. Sh. Factorization of the projection matrix onto a subspace.

A deep factorization of theorthogonal projection matrix onto a subspace is obtained. LQ decomposition isused. For construction of the orthogonal matrix Q the method of successive rankreduction is applied.

С.-Петербургский государственный университет
E-mail: v.malozemov@spbu.ru

Поступило 4 октября 2020 г.

С.-Петербургский государственный университет,
Институт проблем машиноведения РАН
E-mail: g.tamasyan@spbu.ru