

О минимизации числа арифметических операций при одном преобразовании матриц

В. В. Ключев, Н. И. Коковкин-Щербак

Рассмотрим множество невырожденных квадратных матриц n -го порядка A, B, C, \dots с элементами из числового поля P .

Определение 1. Всякое преобразование матрицы A в матрицу B , осуществляемое линейным комбинированием строк и столбцов матрицы A , называется прямым методом преобразования A в B .

Определение 2. Пусть a и b — элементы числового поля P . Будем считать, что произведение $a \cdot b$ или частное $a : b$ дают одну арифметическую операцию второго порядка (операцию \times), если: 1) в произведении $a \cdot b$ сомножители a и b не равны 0 и ± 1 ; 2) в частном $a : b$ компоненты $a \neq 0$ и $b \neq 0, \pm 1$.

При желании часть этих ограничений может быть снята, что отразится лишь на подсчете числа операций.

Рассмотрим множество всех прямых методов $M = \{\alpha, \beta, \dots\}$ преобразования матрицы A в матрицу B , имеющую k фиксированных элементов, $0 < k < n^2$ (чаще всего фиксированными элементами матрицы B будут в дальнейшем нули или нули и единицы).

Обозначим через $f_\alpha(A \rightarrow B, P)$ число операций \times , необходимое для преобразования фиксированной матрицы A в матрицу B данным прямым методом α . Величина $f_\alpha(A \rightarrow B, P)$ зависит от прямого метода α и от $2n^2 - k$ произвольных элементов матриц A и B . Здесь предполагается, что число операций сложения и вычитания не ограничивается, а операция \times к ним не сводится.

Не исключено, что при некоторых прямых методах $f_\alpha(A \rightarrow B, P) = +\infty$. Но для дальнейшего это не существенно. Существенно лишь то, что всегда можно найти такой прямой метод β преобразования A в B , при котором величина $f_\alpha(A \rightarrow B, P)$ конечна. Определим функцию

$$f(n, k) = \inf_{\alpha \in M} \sup_P f_\alpha(A \rightarrow B, P), \quad (1)$$

где верхняя грань берется по всевозможным подмножествам поля P из нефиксированных элементов матриц A и B , а нижняя грань — по всевозможным прямым методам $\alpha \in M$ преобразования A в B .

Теорема 1. Функция $f(n, k)$, определяемая соотношением (1), существует и существует, по крайней мере, один прямой метод $\alpha_0 \in M$ такой, что $\sup_P f_{\alpha_0}(A \rightarrow B, P) = f(n, k)$.

Доказательство основано на том, что все конечные значения ограниченных сверху функций $f_{\alpha}(A \rightarrow B, P)$ по определению есть натуральные числа.

Определение 3. Функцию $f(n, k)$, определяемую условием (1), назовем минимизирующей функцией числа операций \times преобразования A в B , а реализующий эту функцию прямой метод α_0 — минимизирующим методом того же преобразования.

В статье рассмотрены некоторые конструктивные методы построения функции $f(n, k)$ и соответствующих ей минимизирующих методов.

Всякое линейное комбинирование рядов матрицы A можно свести к последовательности двух элементарных преобразований (умножению элементов ряда на число и сложению соответствующих элементов двух рядов). Следовательно, всякий прямой метод α преобразования A в B можно разложить в некоторую последовательность элементарных преобразований

$$E^{(0)}, E^{(1)}, E^{(2)}, \dots, E^{(p_{\alpha}-1)}, \quad (2)$$

дающую последовательность матриц

$$A = A_0 \rightarrow A_1 \rightarrow A_2 \rightarrow \dots \rightarrow A_{p_{\alpha}} = B, \quad (3)$$

где переход от матрицы A , к матрице A_{r+1} реализуется элементарным преобразованием $E^{(r)}$.

Определение 4. Подпоследовательность

$$A_q \rightarrow A_{q+1} \rightarrow \dots \rightarrow A_{q+t} \quad (4)$$

последовательности (3) называется циклом, если $A_{q+t} = A_q$. Эту же подпоследовательность (4) будем называть квазициклом для элементов $a_{ik_1}, a_{ik_2}, \dots, a_{ik_e}, 1 \leq e \leq n$ (для элементов $a_{i_1k}, a_{i_2k}, \dots, a_{i_ek}, 1 \leq e \leq n$) матрицы B , полученных в матрице A_q , в следующих двух случаях: 1) если матрицы A_{q+t} и A_q отличаются лишь элементами $a_{ik}, k \neq k_1, k_2, \dots, k_e (i \neq i_1, i_2, \dots, i_e)$, причем эти элементы a_{ik} не являются элементами матрицы B и i -ая строка (k -ый столбец) матрицы A_{q+t} не используется последующими элементарными преобразованиями для получения элементов матрицы B без предварительного умножения элементов i -ой строки (k -го столбца) на специально выбранное число; 2) если матрица A_{q+t} содержит лишь часть из элементов $a_{ik_1}, a_{ik_2}, \dots, a_{ik_e} (a_{i_1k}, a_{i_2k}, \dots, a_{i_ek})$ или совсем их не содержит. Подпоследовательность (4) будем называть просто квазициклом, если матрицы A_{q+t} и A_q , отличающиеся лишь элементами i -ой строки (k -го столбца), не содержат в i -ой строке (k -ом столбце) элементов матрицы B и i -ая строка (k -ый столбец) матрицы A_{q+t} не используется ни одним из последующих элементарных преобразований для получения элементов матрицы B в другом ряду без предварительного умножения элементов i -ой строки (k -го столбца) на специально выбранное число, отличное от 0 и ± 1 .

Теорема 2. Если M — множество всех прямых методов преобразования матрицы A в B , а M' — подмножество всех тех прямых методов множества M , для которых соответствующие им последовательности элементарных преобразований (3) не содержат циклов и квазициклов, то минимизирующие функции преобразования A в B во множествах M и M' одинаковы.

Доказательство. Пусть прямому методу α из M соответствует последовательность (3). Если в (3) есть цикл (4), то $A_{q+t} = A_q$ и поэтому последовательность (3) можно заменить последовательностью

$$A = A_0 \rightarrow A_1 \rightarrow A_2 \rightarrow \dots \rightarrow A_q \rightarrow A_{q+t+1} \rightarrow \dots \rightarrow A_{p_{\alpha}} = B. \quad (5)$$

Прямой метод, соответствующий последовательности (5), требует, очевидно для своей реализации операций \times не больше, чем α .

Пусть поэтому последовательность (3) не содержит циклов и (4) есть ее квазицикл. Тогда $A_{q+t} \neq A_q$ и (3) нельзя в общем случае заменить последовательностью (5). Рассмотрим сперва случай, когда квазицикл (4) не содержит в себе других квазициклов, кроме самого себя. Назовем такой квазицикл минимальным и для определенности будем считать, что A_{q+t} и A_q отличаются лишь элементами i -ой строки. Квазицикл (4) получается в результате применения подпоследовательности

$$E^{(q)}, \dots, E^{(q+t-1)}. \quad (6)$$

Не трудно видеть, что подпоследовательность (6) не содержит элементарных преобразований столбцов. Но тогда последовательность (2) можно заменить последовательностью

$$E^{(0)}, E^{(1)}, \dots, E^{(q-1)}, E^{(q+t)}, \dots, E^{(p\alpha-1)}. \quad (7)$$

Прямой метод α' , описываемый последовательностью (7), даст последовательность матриц

$$A = A_0 \rightarrow A_1 \rightarrow \dots \rightarrow A_q \rightarrow A_{q+t+1} \rightarrow \dots \rightarrow A_{p\alpha} = B,$$

в которой элементы матриц A_{q+t+s} , $s \geq 1$, в общем случае отличаются от элементов матриц A_{q+t+s} из (3). Но с точки зрения подсчета числа операций \times различие в значениях элементов a_{ik} матриц A_q и A_{q+t} , являющихся в общем случае произвольными элементами поля P , не существенно. Существенна лишь последовательность применения элементарных преобразований, заготавливающих нужные числа и обеспечивающих тем самым получение элементов матрицы B . Поэтому, опуская элементарные преобразования, дающие минимальные квазициклы, последовательно исключим все квазициклы, причем прямой метод α' , не дающий квазициклов, будет затрачивать операций \times не больше, чем α . Отсюда и следует утверждение теоремы.

Опираясь на теорему 2 при построении функции $f(n, k)$ ограничимся в дальнейшем подмножеством прямых методов преобразования A в B , не содержащих циклов и квазициклов.

2°. Обозначим через $F_k = F(q_1, t_1; q_2, t_2; \dots; q_k, t_k)$ квадратную матрицу n -го порядка, у которой в первых k столбцах, считая слева направо, и в первых k строках, считая сверху вниз, $1 \leq k \leq n-1$, соответственно q_1, q_2, \dots, q_k и t_1, t_2, \dots, t_k элементов ниже и правее главной диагонали равны нулю, элементы $a_{pp} = 1$, $p = 1, 2, \dots, k$, а остальные элементы в этих k рядах и остальных рядах имеют произвольные значения.

Л е м м а 1. Минимизирующая функция числа операций \times преобразования $A \rightarrow F_1 \equiv F(q, t)$ во множестве всех прямых методов равна

$$f(n, q+t) = (n-1)(t+q+1) - tq. \quad (8)$$

Д о к а з а т е л ь с т в о. Пусть прямой метод α , реализующий преобразование A в F_1 , дает нули матрицы F_1 в последовательности: сначала получаются s_1 нулей в первом столбце, затем p_1 нулей в первой строке, потом опять s_2 нулей в первом столбце, после чего p_2 нулей в первой строке и так далее по очереди, то в столбце, то в строке, т. е. нули получаются в последовательности $s_1, p_1, s_2, p_2, \dots, s_r, p_r$; $s_1 + \dots + s_r = q < n-1$; $p_1 + \dots + p_r = t < n-1$; $s_i \geq 0, p_i \geq 0$.

1) Если прямой метод таков, что единица на месте (1, 1) получается перед получением всех нулей, то на ее получение будет затрачено в общем случае не меньше $n-1$ операций \times . Обозначим через $f(s_i)$ и $\varphi(p_i)$, $i = 1, 2, \dots, r$, число операций \times , необходимое для получения соответственно s_i нулей в первом столбце и p_i нулей в первой строке. В силу произвольности

элементов матрицы A получим следующие оценки снизу:

$$\begin{aligned} f(s_1) &\geq (n-1) \cdot s_1, & \varphi(p_1) &\geq (n-1-s_1) p_1, \\ \dots & & \dots & \\ f(s_r) &\geq (n-1-p_1-\dots-p_{r-1}) s_r, & \varphi(p_r) &\geq (n-1-s_1-\dots-s_r) \cdot p_r. \end{aligned}$$

Суммируя эти неравенства, получим

$$\begin{aligned} \sum_{m=1}^r f(s_m) + \sum_{m=1}^r \varphi(p_m) &\geq \sum_{m=1}^r \left(n-1 - \sum_{m=1}^{r-1} p_m \right) \cdot s_m + \\ + \sum_{m=1}^r \left(n-1 - \sum_{m=1}^r s_m \right) \cdot p_m &= (n-1)(q+t+1) - qt. \end{aligned}$$

Отсюда и следует (8) для данного случая. Заметим, что полученная оценка совершенно не зависит от последовательности получения $q+t$ нулей.

2) Пусть теперь прямой метод α такой, что преобразование элемента a_{11} в единицу будет выполняться лишь после того, как получено m_1 нулей в первом столбце и m_2 нулей в первой строке, $0 \leq m_1 + m_2 \leq q+t$. Рассуждая как и в первом случае, получим, с учетом произвольности a_{11} , что на преобразование $m_1 + m_2$ элементов в нули будет затрачено операций \times не меньше, чем $n(m_1 + m_2) - m_1 m_2$. Пусть для определенности $m_1 \leq m_2$. Тогда на преобразование a_{11} в единицу пойдет не менее $(n-1-m_2)$ операций \times . Продолжая рассуждать как и в первом случае, находим, что на получение оставшихся $q-m_1$ нулей в столбце и $t-m_2$ нулей в строке будет затрачено операций \times не меньше, чем $(n-1-m_2)(q-m_1) + (n-1-m_1)(t-m_2) - (q-m_1)(t-m_2)$. А всего тем самым на преобразование A в F_1 будет затрачено операций \times не меньше, чем (8). Из рассмотренных случаев следует, что функция (8) минимизирующая.

Лемма 2. Минимизирующая функция числа операций \times во множестве всех прямых методов преобразования $A \rightarrow F_2 \equiv F(q_1, t_1; q_2, t_2)$ равна

$$f(n, q_1 + \dots + t_2) = \sum_{r=1}^2 [(n-r)(q_r + t_r + 1) - q_r t_r]. \quad (9)$$

Доказательство. 1 случай. Пусть нули матрицы F_2 получают в последовательности q_1, t_1, q_2, t_2 , т. е. q_1 нулей в первом столбце, t_1 нулей в первой строке, затем q_2 нулей во втором столбце и t_2 нулей во второй строке. По лемме 1 на получение $q_1 + t_1$ нулей в первых рядах матрицы F_2 требуется операций \times не меньше, чем $(n-1)(q_1 + t_1 + 1) - q_1 t_1$, причем элемент a_{11} будет преобразован в единицу. Оценим теперь снизу число операций \times , необходимое для получения $q_2 + t_2$ нулей.

а) Пусть среди $q_1 + t_1$ нулей есть нули на местах (1, 2) и (2, 1). Тогда для получения нулей во вторых рядах достаточно воспользоваться только линейной комбинацией нужных строк (столбцов) лишь со второй строкой (столбцом). На это по лемме 1 потребуется операций \times не меньше, чем $(n-2)(q_2 + t_2 + 1) - q_2 t_2$. А всего на преобразование A в F_2 потребуется в этом случае операций \times не меньше, чем (9).

Этот результат можно усилить. Пусть даны четыре произвольных последовательности целых чисел $s'_1, \dots, s'_k; p'_1, \dots, p'_k; s''_1, \dots, s''_k; p''_1, \dots, p''_k$, удовлетворяющих условию $s'_1 + \dots + s'_k = q_1; p'_1 + p'_2 + \dots + p'_k = t_1; s''_1 + \dots + s''_k = q_2; p''_1 + p''_2 + \dots + p''_k = t_2$. образуем из них последовательность $s'_1, p'_1, s''_1, p''_1, s'_2, p'_2, \dots, p''_k$ и будем считать, что матрица A преобразуется в матрицу F_2 так, что нули матрицы F_2 получают в этой последовательности. Пусть среди s'_1 и p'_1 нулей есть нули на местах (2, 1) и (1, 2). Тогда к первым и вторым рядам отдельно можно применить лем-

му 1, ибо получение нулей в этих рядах можно производить независимо друг от друга. И мы опять получаем (9).

б) Пусть теперь среди $q_1 + t_1$ нет нулей на местах (1, 2) и (2, 1). Тогда $q_2 \leq q_1 - 1$ и $t_2 \leq t_1 - 1$, причем все q_2 и t_2 нулей находятся в тех же рядах, в которых имеются нули из числа q_1 и t_1 нулей. Если допустить противное, то прямой метод рассматриваемого преобразования будет содержать квазицикл. Тогда для получения любого из q_2 или t_2 нулей следует воспользоваться лишь комбинацией тех строк (столбцов), в которых есть нули из числа $q_1(t_1)$ нулей. Так как в этом случае комбинировать со второй строкой (столбцом) нельзя, чтобы не получить квазицикл, то преобразование a_{22} в единицу можно сделать после того, как получено $\max(q_2, t_2)$ нулей в одном из вторых рядов. В таком случае будет затрачено операций \times не меньше, чем

$$(n-1)(q_2 + t_2 - 1) - \max(q_2, t_2) > (n-2)(q_2 + t_2 + 1) - q_2 t_2.$$

в) Пусть, наконец, на одном из мест (2, 1) или (1, 2) перед получением $q_2 + t_2$ нулей уже получен нуль, а на другом нет. Изучая обе возможности аналогично предыдущему случаю, получим оценку

$$\begin{aligned} (n-2)(q_2 + t_2) - q_2 t_2 + q_2 + t_2 + \max(n-2 - q_2; n-1 - t_2) &\geq \\ &\geq (n-2)(q_2 + t_2 + 1) - q_2 t_2. \end{aligned}$$

Очевидно этот результат может быть усилен построением некоторой последовательности, в которой будут получаться $q_1 + t_1 + q_2 + t_2$ нулей.

Рассмотренные подслучаи, с учетом теоремы 2, доказывают лемму.

З а м е ч а н и я. 1. Для подсчета числа операций \times преобразование A в F_2 всегда можно начинать с получения части нулей в первых рядах матрицы A , ибо в начале преобразования все ряды равноправны с точки зрения этого подсчета.

2. Хотя доказательство леммы строилось так, что нули матрицы F_2 ниже (выше) главной диагонали получались только линейным комбинированием строк (столбцов), но это ограничение не увеличивает минимизирующую функцию преобразования A в F_2 . В самом деле, пусть на некотором шаге нуль получается на одном из мест (2, k), $3 \leq k \leq n$, с помощью линейного комбинирования двух строк. Для того чтобы не получился квазицикл нужно, чтобы элемент на месте (2, 1) был отличен от нуля. Однако при дальнейшем преобразовании, т. е. после получения нуля на месте (2, k) должен появиться квазицикл либо для элемента на месте (1, k), либо для элемента на месте (2, 1). Остается опять сослаться на теорему 2.

3. Если рассматривать преобразование $A \rightarrow F_k \equiv F(q_1, t_1; \dots; q_k, t_k)$, то всегда можно считать, что после получения некоторого числа нулей в первом столбце и в первой строке, получаются нули во втором столбце и во второй строке, ибо с точки зрения подсчета числа операций \times все остальные ряды матрицы, начиная со второго, равнозначны.

Теорема 3. Минимизирующая функция числа операций \times во множестве прямых методов преобразования A в $F_k \equiv F(q_1, t_1; q_2, t_2; \dots; q_k, t_k)$ равна

$$f(n, q_1 + \dots + t_k) = \sum_{r=1}^k [(n-r)(q_r + t_r + 1) - q_r t_r]. \quad (10)$$

Минимизирующий метод, реализующий эту функцию, укладывается в следующую схему: прежде чем будет получен хотя бы один нуль из q_r и t_r нулей в r -ом ряду соответственно ниже и правее главной диагонали матрицы F_k , нужно получить нули на месте всех элементов, кроме диагональных, преобразуемых в единицу, в матричной квадратной клетке $(r-1)$ -го порядка, стоящей в левом верхнем углу матрицы A . Остальные нули в первых $r-1$ рядах могут быть получены в произвольной последовательности.

Доказательство методом математической индукции по r , $1 \leq r \leq k$. Пусть теорема верна для некоторого r . Это значит, что элементы на местах (p, q) , $1 \leq q \leq r-1$, $1 \leq p \leq r-1$, $p \neq q$ преобразованы в нули, а на местах (p, p) , $1 \leq p \leq r-1$ в единицу и минимизирующая функция числа операций \times преобразования A в F_r имеет вид

$$f(n, q_1 + \dots + t_r) = \sum_{i=1}^r [(n-i)(q_i + t_i + 1) - q_i t_i],$$

где q_i и t_i — число нулей, полученных в i -ом столбце и i -ой строке ниже и правее главной диагонали. Рассмотрим квадратную матричную клетку $(n+1-r)$ -го порядка, стоящую в правом нижнем углу частично преобразованной матрицы A . Тогда q_r и t_r нулей будут нулями в первой строке и первом столбце этой клетки. Пусть теперь среди первых $r-1$ элементов $r+1$ строки (столбца) m_1 (m_2) элементов отличны от нуля. В рассматриваемой матричной клетке элементы r -го и $(r+1)$ -го рядов основной матрицы соответственно ниже и правее главной диагонали являются элементами первого и второго рядов. Если положить $m_1 = m_2 = 0$, то к этой матричной клетке $(n+1-r)$ -го порядка можно применить лемму 2.

Так как на получение q_r и t_r нулей по предположению потребовалось операций \times не меньше, чем $(n-r)(q_r + t_r + 1) - q_r \cdot t_r$, то по лемме 2 на получение q_{r+1} и t_{r+1} нулей потребуется не менее, чем $(n-r-1)(q_{r+1} + t_{r+1} + 1) - q_{r+1} \cdot t_{r+1}$ операций, причем элементы на месте $(r+1, r)$ и $(r, r+1)$ будут преобразованы в нули. В этом случае теорема доказана. Если же $m_1 \neq 0$ и $m_2 \neq 0$, то на получение q_{r+1} и t_{r+1} нулей пойдет операций \times не меньше, чем $(n-r-1)(q_{r+1} + t_{r+1} + 1) - q_{r+1} \cdot t_{r+1} + m_1 q_{r+1} + m_2 \cdot t_{r+1}$, если комбинировать все ряды только с $(r+1)$ -ми рядами. Отсюда видно, что эта величина минимизируется при $m_1 = m_2 = 0$.

Итак, переход от r к $r+1$ переводит минимизирующую функцию $f(n, q_1 + \dots + t_r)$ в минимизирующую функцию $f(n, q_1 + \dots + t_{r+1})$, что и доказывает теорему, если учесть, что при $r=1$ теорема верна по лемме 1 и что достаточно рассмотреть множество прямых методов, не содержащих циклов и квазициклов.

Следствие 1. Минимизирующая функция числа операций \times во множестве прямых методов преобразования матрицы A в единичную матрицу E равна

$$f(n, n^2 - n) = \frac{1}{3} n(n^2 - 1),$$

причем одним из минимизирующих методов является известная схема исключения Гаусса.

Действительно, положив $k \leq n-1$, $q_r = t_r = n-r$, $1 \leq r \leq n-1$, получим

$$f(n, n^2 - n) = \sum_{r=1}^{n-1} [(n-r)(n-r+n-r+1) - (n-r)^2] = \frac{1}{3} n(n^2 - 1).$$

Следствие 2. Минимизирующая функция числа операций \times , необходимых для вычисления определителя n -го порядка приведением его матрицы к треугольному виду во множестве всех прямых методов равна

$$f(n) = \frac{1}{3} (n-1)(n^2 + n + 3).$$

Действительно, по теореме 3 на преобразование матрицы определителя к треугольному виду будет затрачено операций \times не меньше, чем $\frac{1}{3} n(n^2 - 1)$.

Добавив $n - 1$ операцию \times , идущую на вычисление произведения диагональных элементов, получим $\frac{1}{3}(n - 1)(n^2 + n + 3)$.

Наконец, теорему 3 можно использовать для оценки числа операций \times при решении линейных алгебраических систем уравнений прямыми методами. В самом деле, пусть $AX = B$ линейная система, причем A — невырожденная квадратная матрица n -го порядка и D — матрица-столбец решения системы. Тогда процедуру получения решения D можно рассматривать как преобразование матрицы A_1 в D_1 , где

$$A_1 = \begin{bmatrix} A & B \\ 0 & 1 \end{bmatrix}, \quad D_1 = \begin{bmatrix} E & D \\ 0 & 1 \end{bmatrix}.$$

($0 = (0, \dots, 0)$ — матрица-строка из нулевых элементов).

Следствие 3. Минимизирующая функция числа операций \times преобразования A_1 в D_1 во множестве всех прямых методов равна

$$f(n) = \frac{1}{3} n(n^2 + 3n - 1),$$

причем в качестве минимизирующего метода можно взять известную схему единственного деления Гаусса.

ЛИТЕРАТУРА

1. Д. К. Фадеев, В. Н. Фадеева, Вычислительные методы линейной алгебры, Физматгиз, 1960.
2. Г. Белага, О вычислении значений многочлена от одного переменного с предварительной обработкой коэффициентов, Проблемы киберн., № 5, 1961.
3. А. М. Островски, On two problems in abstract algebra connected with Horner's rule, Studies presented to R. von Mises, Acad. Pr., New York, 1954.
4. Т. С. Мотзкин, Evaluation of polynomials and Evaluation of rational functions, Bull. Amer. Math. Soc., 1955.

Поступила 14.VII 1964 г.

Черкассы