УДК 519.712.3+519.61

О БИЛИНЕЙНОЙ СЛОЖНОСТИ УМНОЖЕНИЯ МАТРИЦ РАЗМЕРОВ 5×2 И 2×2

В.Б. Алексеев

Аннотация

Статья посвящена изучению билинейной сложности (то есть наименьшего число умножений без учета коммутативности элементов) для задачи умножения матриц малых размеров. Показано, что билинейная сложность для задачи умножения матриц размеров 5×2 и 2×2 не может быть меньше 17 ни над каким полем.

Ключевые слова: умножение матриц, алгоритм, сложность, билинейная сложность.

Введение

Стандартный алгоритм («строка на столбец») для умножения матриц размера $n \times n$ использует порядка n^3 арифметических операций. Еще в 1969 г. Ф. Штрассен [1] построил первый асимптотически более быстрый алгоритм умножения матриц размера $n \times n$ (с числом арифметических операций $O\left(n^{\log_2 7}\right)$). В последующие 20 лет верхняя оценка сложности умножения двух матриц размера $n \times n$ была понижена до $O\left(n^{2.38}\right)$ [2], однако дальше (уже 25 лет) существенных продвижений в этой задаче нет. Результат Штрассена, а также быстрые алгоритмы для умножения чисел и полиномов дали толчок развитию важного нового направления – исследованию минимального числа алгебраических операций для вычислений в различных алгебрах (алгебраической сложности) [3]. Получаемые результаты позволяют лучше понять общие методы построения быстрых алгоритмов в алгебраических задачах. При этом отсутствие продвижений в общей задаче о сложности умножения матриц требует более глубокого изучения этой задачи. Одним из направлений является исследование сложности умножения матриц малого размера. Это связано, в частности, с тем, что алгоритм малой сложности для умножения матриц малого размера можно рекурсивно использовать для умножения матриц большого размера. Именно так получил свой результат Штрассен – он рекурсивно применял найденный им способ перемножения двух матриц размера 2×2 с использованием только 7 умножений вместо обычных 8. Для того чтобы можно было применять рекурсию, Штрассен построил для умножения матриц порядка 2 алгоритм специального вида, а именно так называемый билинейный алгоритм.

Определение 1. Пусть F – некоторое кольцо и пусть имеется 2 множества переменных $A = \{a_1, a_2, \ldots, a_r\}$ и $B = \{b_1, b_2, \ldots, b_s\}$. *Билинейными алгоритмами* над A и B и кольцом F называются алгоритмы, в которых сначала вычисляются произведения $\left(\sum_{i=1}^{r} \alpha_i^t a_i\right)\left(\sum_{j=1}^{s} \beta_j^t b_j\right)$ некоторых линейных форм (с коэффициентами из F) от первого множества переменных на некоторые линейные формы (с коэффициентами из F) от второго множества переменных, где $t = 1, 2, \ldots d$, а на втором этапе вычисляются некоторые линейные комбинации этих d произведений. При этом число умножений d называется *билинейной сложностью алгоритма.*

Определение 2. Будем говорить, что билинейный алгоритм над кольцом F вычисляет систему билинейных форм $C_k = \sum_{i=1}^{r}\sum_{j=1}^{s} c_{ij}^k a_i b_j,\ k = 1, \ldots, h$, где c_{ij}^k – произвольные константы из F, если каждая из этих билинейных форм оказывается вычисленной на втором этапе алгоритма. *Билинейной сложностью задачи* вычисления системы билинейных форм над кольцом F называется минимальная билинейная сложность алгоритмов над F, вычисляющих данную систему билинейных форм.

Обозначим через $\|a_{ij}\|_{m\times n}$ матрицу размера $m \times n$ над некоторым кольцом. Билинейная сложность задачи умножения матрицы $\|a_{ij}\|_{m\times n}$ на матрицу $\|b_{kl}\|_{n\times p}$ – это билинейная сложность вычисления системы из mp билинейных форм вида $\sum_{j=1}^{n} a_{ij}b_{jl}$.

Требование билинейности в алгоритмах при изучении сложности умножения матриц связано с тем, что при рекурсии вместо переменных подставляются матрицы, которые могут не коммутировать. Если отказаться от билинейности и считать, что элементы перемножаемых матриц коммутируют, то наименьшее число умножений в произвольных алгоритмах с операциями сложения, вычитания и умножения для вычисления произведения двух матриц заданных размеров называют *мультипликативной* сложностью задачи. Известно, что билинейная и мультипликативная сложности задачи могут не совпадать. Например, при наличии коммутативности можно построить алгоритм для умножения матрицы размера 3×2 на матрицу размера 2×2 с 10 умножениями [4], в то время как билинейная сложность этой задачи равна 11 [5].

Оказалось, что даже в задачах перемножения двух матриц достаточно малого размера не удается установить точное значение билинейной сложности. Например, для задачи перемножения двух матриц размера 3×3 к настоящему моменту известно только то, что билинейная сложность заключена между 19 и 23 [6, 7]. Для задачи перемножения двух матриц размера 4×4 верхняя оценка 49 на число умножений (вместо обычных 64) получается двукратным использованием алгоритма Штрассена, и эта оценка пока не понижена. Для задачи перемножения двух матриц размера 5×5 наилучшим остается алгоритм из [8] с числом умножений 100 вместо обычных 125. Из недавних результатов интересен результат А.В. Смирнова [9], который построил алгоритм для умножения матрицы размера 3×3 на матрицу размера 3×6 с 40 умножениями (вместо обычных 54).

Еще тяжелее обстоит дело с нижними оценками. Для задачи перемножения двух матриц размера 2×2 достаточно быстро было доказано, что оценка 7 на число умножений неулучшаема над произвольным полем [10]. Однако к данному моменту для билинейной сложности умножения матрицы размера $m \times n$ на матрицу размера $n \times p$ нижние оценки над произвольным полем, совпадающие с верхней оценкой, установлены только для нескольких значений параметров $m,\ n,\ p$.

Обозначим через $\langle m, n, p\rangle_F$ задачу умножения матрицы размера $m \times n$ на матрицу размера $n \times p$ над некоторым полем F. А через $rk_F\langle m, n, p\rangle$ обозначим билинейную сложность этой задачи. Теорема о двойственности [11] утверждает, что $rk_F\langle m, n, p\rangle$ не изменяется при любой перестановке чисел $m,\ n,\ p$.

Нетрудно показать, что $rk_F\langle m, 1, p\rangle = mp$. Из результата Штрассена легко получается, что $rk_F\langle m, 2, 2\rangle \le \lceil 7m/2\rceil$ для произвольного поля F. В работе [12] была получена такая же нижняя оценка, но только для поля из 2 элементов. Автором в работе [5] был рассмотрен случай $m = 3$ и доказано, что $rk_F\langle 3, 2, 2\rangle = 11$ для

произвольного поля F. В статье [13] доказано, что $rk_F\langle 4,2,2\rangle = 14$ для произвольного поля F. Пока только для этих параметров и двойственных к ним установлено точное значение для $rk_F\langle m,n,p\rangle$ над произвольным полем F. В настоящей работе рассматривается величина $rk_F\langle 5,2,2\rangle$, для которой получена нижняя оценка $rk_F\langle 5,2,2\rangle \geq 17$ над произвольным полем F. Отметим, что наилучшая известная верхняя оценка для этой задачи равна 18.

Кроме билинейных алгоритмов рассматривают также более общий класс так называемых приближенных билинейных алгоритмов и соответствующую приближенную билинейную сложность (определения см., например, в [13]). Для задачи умножения матрицы размера 5×2 на матрицу размера 2×2 приближенная билинейная сложность не превосходит 16 [9, 13]. Поэтому из полученного результата вытекает, что для задачи умножения матрицы размера 5×2 на матрицу размера 2×2 билинейная сложность и приближенная билинейная сложность различаются. Ранее различие билинейной сложности и приближенной билинейной сложности для задачи умножения матриц было установлено только для случаев $\langle 3,2,2\rangle$ [5, 14] и $\langle 4,2,2\rangle$ [13] (и двойственных к ним). Совпадение билинейной сложности и приближенной билинейной сложности доказано для задачи $\langle 2,2,2\rangle$ [15].

1. Нижняя оценка

В этом разделе мы покажем, что любой билинейный алгоритм (обычный, неприближенный) для задачи умножения матрицы размера 5×2 на матрицу размера 2×2 над любым полем имеет билинейную сложность не менее 17 (отметим, что в [12] для этой задачи получена нижняя оценка 18, но только для поля из 2 элементов). Основой доказательства является эквивалентность рассматриваемой задачи формулируемой ниже задаче P.

Пусть m, n, p – фиксированные натуральные числа и пусть P – произвольная матрица порядка mn. Тогда выражением $|P_1|P_2|\dots|P_m|$ будем обозначать матрицу P, разрезанную на m вертикальных блоков одинаковой ширины n. Для любых $1 \leq i \leq m$, $1 \leq l \leq p$ определим матрицу P_{il} размера $mn \times np$ следующим образом: $P_{il} = |0|\dots|0|P_i|0|\dots|0|$, где все p блоков имеют ширину n, l-й блок матрицы P_{il} равен P_i, то есть i-му блоку матрицы P, а остальные блоки матрицы P_{il} являются нулевыми матрицами. Рассмотрим следующую задачу.

Задача P: найти наименьшее число d матриц ранга 1, линейными комбинациями которых являются все указанные выше mp матриц P_{il} (при заданных фиксированных m, n, p).

В [13] доказано следующее утверждение (первая часть леммы 2).

Лемма 1. *Над любым полем F наименьшая сложность решения Задачи P при фиксированных m, n, p одинакова для всех невырожденных матриц P и равна билинейной сложности умножения матрицы $A = \|a_{ij}\|_{m\times n}$ на матрицу $B = \|b_{kl}\|_{n\times p}$.*

Эта лемма показывает, что для исследования билинейной сложности умножения двух матриц можно просто исследовать Задачу P с определенными параметрами. Следующие 3 леммы, доказанные в [13], позволяют упрощать решения Задачи P (лемма 2 – это вторая часть леммы 2 в [13]). В дальнейшем мы будем считать фиксированным некоторое (произвольное) поле F, над которым рассматриваются все матрицы, и не будем отмечать его в формулировках утверждений.

Лемма 2. *Если $D_1, D_2, \dots, D_d$ – решение Задачи P и C – невырожденная матрица, то $CD_1, CD_2, \dots, CD_d$ – решение Задачи CP.*

Лемма 2 позволяет из произвольного решения Задачи P получать решения более простого вида (возможно для другой матрицы).

Лемма 3. *Пусть $D_1, D_2, \ldots, D_d$ – решение Задачи P для некоторой невырожденной матрицы $P = |P_1|P_2|\ldots|P_m|$ порядка mn (блоки ширины n). Пусть каждый блок P_i заменяется на блок P'_i, являющийся линейной комбинацией блоков P_i. Тогда тот же набор матриц $D_1, D_2, \ldots, D_d$ является решением Задачи P' для матрицы $P' = |P'_1|P'_2|\ldots|P'_m|$.*

Замечание. При использовании леммы 3 надо будет следить за тем, чтобы матрица P' оставалась невырожденной.

Лемма 4. *Пусть $D_1, D_2, \ldots, D_d$ – решение Задачи P для некоторой невырожденной матрицы P. Пусть $D_t = |D_t^1|D_t^2|\ldots|D_t^p|$ – разбиение каждой матрицы D_t на блоки шириной n. Пусть одновременно во всех матрицах D_t делается одно и то же невырожденное преобразование с блоками: $\overline{D}_t^j = \sum_{i=1}^{p} l_{ij} D_t^i$, $j = 1, \ldots, p$. Пусть матрица этого преобразования $L = \|l_{ij}\|$ является невырожденной. Тогда полученные матрицы $\overline{D}_t$ также являются решением Задачи P для той же матрицы P. (Матрицы $\overline{D}_t$, очевидно, остаются матрицами первого ранга.)*

Основным результатом настоящей статьи является следующая

Теорема 1. *Любой билинейный алгоритм для умножения матрицы размера 5×2 на матрицу размера 2×2 над произвольным полем имеет билинейную сложность не менее 17. Тот же результат справедлив для умножения матрицы размера 2×5 на матрицу размера 5×2 и для умножения матрицы размера 2×2 на матрицу размера 2×5.*

Доказательство. Мы докажем от противного, что для задачи умножения матрицы размера 5×2 на матрицу размера 2×2 над произвольным полем не существует билинейного алгоритма с 16 умножениями. По лемме 1 это утверждение равносильно отсутствию решения у следующей задачи.

Пусть P – произвольная невырожденная матрица размера 10×10, разбитая на 5 подматриц размера 10×2: $P = |P_1|P_2|P_3|P_4|P_5|$. Пусть P_{i1} – матрица размера 10×4 вида $P_{i1} = |P_i|0|$, а P_{i2} – матрица размера 10×4 вида $P_{i2} = |0|P_i|$ (все блоки ширины 2). Спрашивается, существуют ли 16 матриц **первого ранга** $D_1, \ldots, D_{16}$ размера 10×4 такие, что все 10 матриц P_{i1}, $i = 1, \ldots, 5$, и P_{i2}, $i = 1, \ldots, 5$, являются линейными комбинациями матриц $D_1, \ldots, D_{16}$. Надо доказать, что эта задача не имеет решения ни для какой невырожденной матрицы P.

Допустим, что для некоторой невырожденной матрицы P размера 10×10 такое решение $D_1, \ldots, D_{16}$ из матриц ранга 1 существует. Это означает, что существуют такие коэффициенты $\alpha_{it}^1, \alpha_{it}^2$ (из рассматриваемого поля), что для $i = 1, \ldots, 5$ выполняются равенства

$$P_{i1} = \sum_{t=1}^{16} \alpha_{it}^1 D_t, \quad P_{i2} = \sum_{t=1}^{16} \alpha_{it}^2 D_t.$$

Разобьем все матрицы D_t на 2 блока размера 10×2: $D_t = |D_t^1|D_t^2|$. Тогда последние равенства равносильны равенствам

$$P_i = \sum_{t=1}^{16} \alpha_{it}^1 D_t^1, \quad \sum_{t=1}^{16} \alpha_{it}^1 D_t^2 = 0, \quad \sum_{t=1}^{16} \alpha_{it}^2 D_t^1 = 0, \quad P_i = \sum_{t=1}^{16} \alpha_{it}^2 D_t^2. \tag{1}$$

Так как матрица P невырожденная, то ее столбцы линейно независимы, поэтому, в частности, линейно независимы подматрицы $P_1,\ P_2,\ P_3,\ P_4,\ P_5$. Тогда из последнего равенства в (1) следует, что 5 векторов коэффициентов $(\alpha^2_{i1}, \alpha^2_{i2}, \ldots, \alpha^2_{i,16})$, $i = 1, \ldots, 5$ линейно независимы, а значит, из третьего равенства в (1) следует, что среди матриц $D^1_1, \ldots, D^1_{16}$ не более 11 линейно независимых. С другой стороны, в матрицах $P_1,\ P_2,\ P_3,\ P_4,\ P_5$ имеется 10 линейно независимых столбцов, которые согласно первому равенству в (1) должны быть линейными комбинациями столбцов из $D^1_1, \ldots, D^1_{16}$. Поскольку матрицы $D^1_1, \ldots, D^1_{16}$ имеют ранг 1, это означает, что среди $D^1_1, \ldots, D^1_{16}$ есть 10 ненулевых матриц первого ранга, построенных на 10 линейно независимых столбцах. Тогда эти 10 матриц линейно независимы. Получаем, что возможны только 2 случая: максимальное число линейно независимых матриц среди матриц $D^1_1, \ldots, D^1_{16}$ равно 10 или 11. Рассмотрим эти случаи отдельно.

Случай 1. Среди матриц $D^1_1, \ldots, D^1_{16}$ имеется ровно 10 линейно независимых матриц.

Пусть матрицы $D^1_1, \ldots, D^1_{10}$ линейно независимы, а матрицы $D^1_{11}, \ldots, D^1_{16}$ являются их линейными комбинациями. Это значит, что существуют такие коэффициенты b^p_t, k^p_i (из рассматриваемого поля), что выполняются равенства (см. (1))

$$D^1_t = \sum_{p=1}^{10} b^p_t D^1_p, \quad t = 11, \ldots, 16; \quad P_i = \sum_{t=1}^{10} k^t_i D^1_t, \quad i = 1, \ldots, 5. \tag{2}$$

Согласно (2) из столбцов матриц $D^1_1, \ldots, D^1_{10}$ линейными комбинациями должны получаться все 10 линейно независимых столбцов матрицы P. Так как все матрицы $D^1_1, \ldots, D^1_{10}$ имеют ранг 1, все матрицы $D^1_1, \ldots, D^1_{10}$ должны быть ненулевыми и построенными на 10 линейно независимых столбцах. Каждую матрицу D_t первого ранга можно представить в виде произведения $D_t = f_t \cdot g_t$, где f_t – некоторый вектор-столбец, а g_t – некоторая вектор-строка. Тогда $CD_t = (Cf_t) \cdot g_t$. Поскольку вектор-столбцы $f_1, f_2, \ldots, f_{10}$ линейно независимы, то можно найти такую невырожденную матрицу C, что $Cf_t = e_t$ при всех $t = 1, \ldots, 10$, где e_t – вектор-столбец с одной 1 на t-м месте. Тогда в матрице CD_t единственной ненулевой строкой будет строка с номером t. По лемме 2 матрицы $CD_t,\ t = 1, \ldots, 10$, будут опять давать решение Задачи P, но с заменой матрицы P на матрицу CP. При этом равенства (2) останутся справедливыми для матрицы CP и матриц CD_t. Чтобы избежать громоздких обозначений, можем считать, что решение $D_1, \ldots, D_{16}$ уже для исходной матрицы P имеет такой вид, то есть в матрице $D_t,\ t = 1, \ldots, 10$, единственной ненулевой строкой является строка с номером t.

Пусть

$$N_t = D_t - \sum_{p=1}^{10} b^p_t D_p, \quad t = 11, \ldots, 16. \tag{3}$$

Тогда из (2) получаем, что N_t имеет вид $N_t = |0|N^2_t|$, где N^2_t – подматрица размера 10×2. Так как среди матриц $D^1_1, \ldots, D^1_{16}$ ровно 10 линейно независимых матриц, то пространство линейных комбинаций $\sum\limits_{t=1}^{16} \mu_t D_t$ таких, что $\sum\limits_{t=1}^{16} \mu_t D^1_t = 0$, имеет размерность $16 - 10 = 6$. Но 6 линейных комбинаций $D_t - \sum\limits_{p=1}^{10} b^p_t D_p,\ t = 11, \ldots, 16$, выражающих N_t, входят в это пространство и линейно независимы, поскольку каждое слагаемое $D_{11}, \ldots, D_{16}$ входит ровно в одну из этих линейных комбинаций. Следовательно, каждая матрица, которая является линейной комбинацией

матриц $D_1,\dots,D_{16}$ и имеет вид $|0|H|$, является линейной комбинацией матриц $N_{11},\dots,N_{16}$. В частности, существуют такие константы e_i^t (из рассматриваемого поля), что выполняются равенства:

$$P_{i2}=|0|P_i|=\sum_{t=11}^{16}e_i^tN_t,\quad i=1,\dots,5;\quad P_i=\sum_{t=11}^{16}e_i^tN_t^2,\quad i=1,\dots,5. \tag{4}$$

Определение 3. Будем говорить, что матрица $N_t=|0|F_t|$ имеет тип F, если для нее в (3) все коэффициенты $b_t^p=0$. Будем говорить, что матрица $N_t=|0|G_t|$ имеет тип G, если для нее в (3) существует коэффициент $b_t^p\neq 0$.

Напомним, что каждая из матриц $D_1,\dots,D_{10}$ отлична от 0 только в одной строке (каждая в своей).

Лемма 5. *Если матрица $N_t=|0|F_t|$ имеет тип F, то ранг $rk(N_t)=1$. Если матрица $N_t=|0|G_t|$ имеет тип G, то она может быть отлична от 0 только в таких строках, в которых соответствующие строки из $D_1^1,\dots,D_{10}^1$ пропорциональны.*

Доказательство. В первом случае из (3) следует, что $N_t=D_t$ и ранг $\mathrm{rk}(N_t)=\mathrm{rk}(D_t)=1$. Во втором случае из (2) имеем $D_t^1=\sum\limits_{p=1}^{10}b_t^pD_p^1\neq 0$ и D_t^1 имеет ранг 1. Это значит, что все строки в ней пропорциональны. Поэтому в сумме $\sum\limits_{p=1}^{10}b_t^pD_p^1$ с ненулевыми коэффициентами b_t^p могут участвовать только те матрицы D_p^1, у которых единственные ненулевые строки пропорциональны. Но тогда матрица D_t^1 равна 0 во всех остальных строках, а поскольку $D_t^1\neq 0$ и матрица D_t имеет ранг 1, то и вся матрица D_t равна 0 во всех остальных строках. При этом и матрица $N_t=D_t-\sum\limits_{p=1}^{10}b_t^pD_p$ равна 0 во всех остальных строках. Лемма доказана. □

Так как в P_1, P_2, P_3, P_4, P_5 имеется 10 линейно независимых столбцов, из (4) и леммы 5 вытекает следующее утверждение.

Лемма 6. *В матрицах $N_{11}^2,\dots,N_{16}^2$ имеется 10 линейно независимых столбцов. В частности, среди $N_{11},\dots,N_{16}$ должно быть не менее 4 матриц типа G и, следовательно, не более 2 матриц типа F.*

Пусть для $j=1,2,\dots,10$ матрица D_j равна $|p_j|q_j|$ в j-й строке, где p_j, q_j – векторы размерности 2 (в остальных строках она равна 0). Для наглядности можно изображать набор матриц $D_1,\dots,D_{10}$ одной матрицей в виде

$$\left|\begin{array}{c|c|c|}
D_1: & p_1 & q_1 \\
D_2: & p_2 & q_2 \\
\dots & \dots & \dots \\
D_{10}: & p_{10} & q_{10}
\end{array}\right|$$

Лемма 7. *В каждой матрице P_i, $i=1,\dots,5$, в строке с номером j стоит вектор, пропорциональный вектору p_j.*

Это утверждение вытекает из второго равенства в (2).

Разобьем все 10 строк (или их номера) на классы эквивалентности $T_1,T_2,\dots,T_s$ так, что любые две строки с номерами r и l входят в один класс тогда и только тогда, когда векторы p_r и p_l пропорциональны (заметим, что все векторы $p_j\neq$ $\neq(0,0)$, поскольку матрицы $D_1^1,\dots,D_{10}^1$ линейно независимы).

Определение 4. Пусть зафиксирован некоторый класс эквивалентности T_m. Тогда для любой матрицы K с 10 строками будем через $\overline{K}$ обозначать ее подматрицу, образованную строками из T_m, а через $\overline{\overline{K}}$ – ее подматрицу, образованную остальными строками.

Лемма 8. *Пусть T_m – любой класс эквивалентности. Тогда* 5 *матриц* $\overline{\overline{P}}_1$, $\overline{\overline{P}}_2$, $\overline{\overline{P}}_3$, $\overline{\overline{P}}_4$, $\overline{\overline{P}}_5$ *линейно независимы.*

Доказательство. Допустим, что $\overline{\overline{P}}_1$, $\overline{\overline{P}}_2$, $\overline{\overline{P}}_3$, $\overline{\overline{P}}_4$, $\overline{\overline{P}}_5$ линейно зависимы. Тогда можно взять невырожденную линейную комбинацию матриц P_1, P_2, P_3, P_4, P_5 так, что получится матрица Q, в которой $\overline{\overline{Q}} \equiv 0$. При этом согласно лемме 7 в $\overline{Q}$ все строки (длины 2) будут пропорциональны, то есть матрица Q будет иметь ранг не более 1. Получаем, что в матрице P можно выполнить такое невырожденное линейное преобразование столбцов, что получится подматрица размера 10×2 ранга не более 1, то есть вся матрица будет вырожденной. Но это невозможно, поскольку P (по условию) – невырожденная матрица. Следовательно, утверждение леммы 8 верно (от противного). □

Лемма 9. *Пусть T_m – любой класс эквивалентности. Тогда все* 6 *матриц* $\overline{\overline{N}}^2_{11}, \ldots, \overline{\overline{N}}^2_{16}$ *линейно независимы.*

Доказательство. Согласно (4) все матрицы $\overline{\overline{P}}_1$, $\overline{\overline{P}}_2$, $\overline{\overline{P}}_3$, $\overline{\overline{P}}_4$, $\overline{\overline{P}}_5$ являются линейными комбинациями матриц $\overline{\overline{N}}^2_{11}, \ldots, \overline{\overline{N}}^2_{16}$, причем по лемме 8 матрицы $\overline{\overline{P}}_1$, $\overline{\overline{P}}_2$, $\overline{\overline{P}}_3$, $\overline{\overline{P}}_4$, $\overline{\overline{P}}_5$ линейно независимы. Следовательно, среди матриц $\overline{\overline{N}}^2_{11}, \ldots, \overline{\overline{N}}^2_{16}$ есть не менее 5 линейно независимых. Допустим, что среди них максимум 5 линейно независимых. Тогда линейные комбинации матриц $\overline{\overline{N}}^2_{11}, \ldots, \overline{\overline{N}}^2_{16}$ дают то же множество матриц, что и линейные комбинации матриц $\overline{\overline{P}}_1, \ldots, \overline{\overline{P}}_5$, и, следовательно, матрицы $\overline{\overline{N}}^2_{11}, \ldots, \overline{\overline{N}}^2_{16}$ являются линейными комбинациями матриц $\overline{\overline{P}}_1, \ldots, \overline{\overline{P}}_5$. Тогда по лемме 7 получаем, что для всех $j \notin T_m$ строки с номером j в матрицах $N^2_{11}, \ldots, N^2_{16}$ пропорциональны вектору p_j. Если $\overline{\overline{N}}^2_i \neq 0$ и N_i – матрица типа G, то согласно лемме 5 N_i может быть отлична от 0 только в строках некоторого одного класса эквивалентности T_s (отличного от T_m, поскольку $\overline{\overline{N}}^2_i \neq 0$). Причем все строки в N^2_i будут пропорциональны одному и тому же вектору p_j. Следовательно, если $\overline{\overline{N}}^2_i \neq 0$ и N_i – матрица типа G, то N_i имеет ранг 1. Поскольку все матрицы типа F имеют ранг 1 (по лемме 5), получаем, что не менее 5 матриц из N_i, $i = 11, \ldots, 16$ имеют ранг 1 (те, у которых $\overline{\overline{N}}^2_i \neq 0$). Тогда у всех матриц N_i, $i = 11, \ldots, 16$ не более 7 линейно независимых столбцов. Это противоречит лемме 6. От противного получаем утверждение леммы 9. □

Следствие. *Для любого класса эквивалентности T_m и любой матрицы N_i, $i = 11, \ldots, 16$, выполняется $\overline{\overline{N}}^2_i \neq 0$.*

По лемме 6 среди матриц N_i, $i = 11, \ldots, 16$ есть хотя бы одна матрица N типа G. По лемме 5 она может отличаться от 0 только в строках некоторого одного класса эквивалентности T_m. Это означает, что для подматрицы $\overline{\overline{N}}$, построенной для класса эквивалентности T_m, выполняется $\overline{\overline{N}} \equiv 0$. Если имеется более одного класса эквивалентности, то получаем противоречие со следствием. Если же имеется ровно один класс эквивалентности, то из леммы 7 следует, что в подматрице P_1 матрицы P все строки (длины 2) пропорциональны, и, следовательно, она имеет

ранг не более 1. Это противоречит невырожденности матрицы P. Получаем, что Случай 1 невозможен.

Случай 2. Среди матриц $D_1^1, \dots, D_{16}^1$ имеется 11 линейно независимых матриц.

Не ограничивая общности, будем считать, что матрицы $D_1^1, \dots, D_{11}^1$ линейно независимы, а матрицы $D_{12}^1, \dots, D_{16}^1$ являются их линейными комбинациями. Тогда из (1) вытекает, что существуют такие коэффициенты b_t^p, k_i^p (из рассматриваемого поля), что выполняются равенства

$$D_t^1 = \sum_{p=1}^{11} b_t^p D_p^1, \quad t = 12, \dots, 16; \quad P_i = \sum_{t=1}^{11} k_i^t D_t^1, \quad i = 1, \dots, 5. \tag{5}$$

Из второго равенства в (5) следует, что все столбцы матрицы P являются линейными комбинациями столбцов матриц $D_1^1, \dots, D_{11}^1$. Так как по условию матрица P невырожденная, то среди матриц первого ранга $D_1^1, \dots, D_{11}^1$ есть 10 ненулевых матриц, построенных на 10 линейно независимых столбцах. Не ограничивая общности, будем считать, что матрицы $D_1^1, \dots, D_{10}^1$ ненулевые и построены на 10 линейно независимых столбцах. Пусть в равенстве (5) среди k_1^{11}, k_2^{11}, k_3^{11}, k_4^{11}, k_5^{11} есть отличные от 0. Будем считать, что это k_1^{11}. Тогда, вычитая P_1 из P_2, P_3, P_4, P_5 с подходящими коэффициентами, можно для новых блоков получить

$$k_2^{11} = k_3^{11} = k_4^{11} = k_5^{11} = 0. \tag{6}$$

По лемме 3 те же матрицы $D_1, \dots, D_{16}$ будут давать решение для полученной после вычитания блоков матрицы P', причем легко видеть, что матрица P' останется невырожденной. Мы можем считать, что мы изначально рассматриваем такую матрицу, то есть, что равенства (6) выполняются (если изначально все k_1^{11}, k_2^{11}, k_3^{11}, k_4^{11}, k_5^{11} равны 0, то это тоже выполняется).

Положим

$$N_t = D_t - \sum_{p=1}^{11} b_t^p D_p, \quad t = 12, \dots, 16.$$

Тогда из первого равенства в (5) получаем, что N_t имеет вид $N_t = |0|N_t^2|$, где N_t^2 – подматрица размера 10×2.

Лемма 10. *Если матрица является линейной комбинацией матриц $D_1, \dots, D_{16}$ и имеет вид $|0|H|$, то она является линейной комбинацией матриц $N_{12}, \dots, N_{16}$.*

Доказательство. Рассмотрим всевозможные линейные комбинации $\sum_{t=1}^{16} \mu_t D_t$. Векторы коэффициентов этих линейных комбинаций образуют 16-мерное линейное пространство. По условию среди матриц $D_1^1, \dots, D_{16}^1$ ровно 11 линейно независимых матриц. Поэтому подпространство тех линейных комбинаций, для которых $\sum_{t=1}^{16} \mu_t D_t^1 = 0$, имеет размерность $16 - 11 = 5$. Но 5 линейных комбинаций $D_t - \sum_{p=1}^{11} b_t^p D_p$, $t = 12, \dots, 16$, выражающих N_t, входят в это подпространство и линейно независимы, поскольку каждое слагаемое $D_{12}, \dots, D_{16}$ входит ровно в одну из этих линейных комбинаций. Следовательно, они образуют базис в этом подпространстве. □

По определению решения задачи P все матрицы P_{i2} являются линейными комбинациями матриц $D_1, \dots, D_{16}$, где $P_{i2} = |0|P_i|$. Тогда по лемме 10 получаем, что 5 матриц P_{i2}, $i = 1, \dots, 5$, являются линейными комбинациями 5 матриц $N_{12}, \dots, N_{16}$. Но матрицы P_{i2}, $i = 1, \dots, 5$, линейно независимы, поскольку матрицы P_i, $i = 1, \dots, 5$, линейно независимы как блоки невырожденной матрицы P. Поэтому матрицы $N_{12}, \dots, N_{16}$, в свою очередь, выражаются как линейные комбинации матриц P_{12}, P_{22}, P_{32}, P_{42}, P_{52}, то есть существуют такие константы l_t^j, что

$$N_t = \sum_{j=1}^{5} l_t^j P_{j2}, \quad N_t^2 = \sum_{j=1}^{5} l_t^j P_j, \quad t = 12, \dots, 16. \tag{7}$$

По условию линейными комбинациями матриц $D_1, \dots, D_{16}$ являются также все матрицы $P_{i1} = |P_i|0|$, а значит, и матрицы $P_{i1} - \sum_{t=1}^{11} k_i^t D_t$. Но из (5) следует, что матрица $P_{i1} - \sum_{t=1}^{11} k_i^t D_t = |P_i|0| - \sum_{t=1}^{11} k_i^t |D_t^1|D_t^2|$ имеет вид $|0|H|$. По лемме 10 получаем, что все матрицы $P_{i1} - \sum_{t=1}^{11} k_i^t D_t$ являются линейными комбинациями 5 матриц $N_{12}, \dots, N_{16}$, а значит, согласно 7, и линейными комбинациями матриц P_{j2}. С учетом (6) получаем, что для некоторых констант r_i^j, $j = 1, \dots, 5$ для $i = 2, 3, 4, 5$ выполняется

$$P_{i1} = \sum_{t=1}^{10} k_i^t D_t + \sum_{j=1}^{5} r_i^j P_{j2}. \tag{8}$$

Если среди r_2^1, r_3^1, r_4^1, r_5^1 есть не равные 0, то пусть, например, $r_2^1 \neq 0$. Пусть тогда $v_3 = r_3^1/r_2^1$. Если среди r_2^1, r_3^1, r_4^1, r_5^1 все равны 0, то положим $v_3 = 0$. В любом случае $r_3^1 - v_3 r_2^1 = 0$. Рассмотрим матрицы $Q = P_3 - v_3 P_2$ и $Q_1 = |Q|0| = P_{31} - v_3 P_{21}$ (блоки ширины 2). Тогда из 8 получаем, что для некоторых констант $k_1, \dots, k_{10}$, w_2, w_3, w_4, w_5 выполняется равенство

$$Q_1 = \sum_{t=1}^{10} k_t D_t + w_2 P_{22} + w_3 P_{32} + w_4 P_{42} + w_5 P_{52}. \tag{9}$$

В 3-м и 4-м столбцах равенство (9) дает

$$0 = \sum_{t=1}^{10} k_t D_t^2 + w_2 P_2 + w_3 P_3 + w_4 P_4 + w_5 P_5. \tag{10}$$

Из равенств (5) и (6) следует, что все матрицы P_2, P_3, P_4, P_5 являются линейными комбинациями матриц $D_1^1, \dots, D_{10}^1$. Поэтому из (10) получаем, что для некоторых констант $h_1, \dots, h_{10}$ выполняется равенство

$$\sum_{t=1}^{10} k_t D_t^2 + \sum_{t=1}^{10} h_t D_t^1 = \sum_{t=1}^{10} (k_t D_t^2 + h_t D_t^1) \equiv 0. \tag{11}$$

Если для всех $t = 1, \dots, 10$ выполняется $k_t = 0$, то из (9) в первых двух столбцах получаем $Q \equiv 0$. Так как $Q = P_3 - v_3 P_2$, получаем, что блоки P_2 и P_3 в невырожденной матрице P линейно зависимы, что невозможно.

Значит, среди $k_1, \dots, k_{10}$ есть ненулевые. Не ограничивая общности, пусть $k_1 \neq \neq 0$. Напомним, что мы выбрали 10 матриц $D_1, \dots, D_{10}$ ранга 1 так, что они построены на 10 линейно независимых столбцах. Следовательно, матрицы $k_t D_t^2 + h_t D_t^1$ –

это матрицы не более чем первого ранга, построенные на 10 линейно независимых столбцах. Тогда равенство (11) возможно, только если $k_t D_t^2 + h_t D_t^1 \equiv 0$ для всех $t = 1, \ldots, 10$. В частности,

$$k_1 D_1^2 + h_1 D_1^1 \equiv 0. \tag{12}$$

Проделаем одновременно во всех матрицах D_t, $t = 1, \ldots, 16$, следующее преобразование над блоками

$$\begin{cases} H_t^1 = h_1 D_t^1 + k_1 D_t^2, \\ H_t^2 = D_t^1. \end{cases}$$

Определитель этого преобразования $\begin{vmatrix} h_1 & k_1 \\ 1 & 0 \end{vmatrix} = -k_1 \neq 0$. Поэтому это преобразование обратимое и по лемме 4 новые матрицы $H_t = |H_t^1|H_t^2|$, $t = 1, \ldots, 16$ дают решение Задачи P для той же невырожденной матрицы P. При этом из (12) следует, что $H_1^1 \equiv 0$.

Так как мы уже доказали, что Случай 1 невозможен, среди подматриц H_t^1, $t = 1, \ldots, 16$, должно быть ровно 11 линейно независимых. Поскольку $H_1^1 \equiv 0$, то H_1^1 не входит в их число. Тогда по H_1 будет строиться матрица типа N. В аналоге равенства (5) для H_1^1 (сменится нумерация) все коэффициенты b_k^p будут равны 0, и, следовательно, для соответствующей матрицы N_1 получим $N_1 = H_1$ и N_1 будет матрицей ранга 1. Тогда 5 матриц N_t будут иметь не более 9 линейно независимых столбцов. Но согласно лемме 10 их линейными комбинациями должны являться все матрицы $P_{i2} = |0|P_i|$, в которых 10 линейно независимых столбцов (ввиду невырожденности матрицы P).

Получаем, что Случай 2 также привел к противоречию. Поскольку полностью разобраны оба возможных случая, получаем, что для задачи умножения матрицы размера 5×2 на матрицу размера 2×2 над произвольным полем не существует билинейного алгоритма с 16 умножениями. Для других двух задач из теоремы 1 утверждение теоремы вытекает из равенства билинейной сложности для двойственных задач умножения матриц [11]. Теорема 1 полностью доказана. □

Выше отмечалось, что приближенная билинейная сложность задачи умножения матрицы размера 5×2 на матрицу размера 2×2 над произвольным полем не превышает 16 [9, 13]. С другой стороны, наилучший известный точный билинейный алгоритм для этой задачи имеет сложность 18 (надо дважды использовать алгоритм Штрассена), причем над полем из 2 элементов эту сложность понизить нельзя [12]. Если удастся показать, что точная билинейная сложность задачи умножения матрицы размера 5×2 на матрицу размера 2×2 над произвольным полем не меньше 18, это будет первым примером задачи об умножении матриц, в которой точная и приближенная билинейные сложности над произвольным полем различаются заведомо как минимум на 2.

Работа выполнена при финансовой поддержке Российского фонда фундаментальных исследований (проект № 12-01-91331-ННИО-а).

Summary

V.B. Alekseev. On Bilinear Complexity of Multiplication of 5×2 Matrix by 2×2 Matrix.

In this paper we study bilinear complexity (i.e. the minimum number of multiplications without using commutativity of the elements) for the problem of multiplication of matrices of small size. We show that the bilinear complexity for the problem of multiplication of a 5×2 matrix by a 2×2 matrix is at least 17 for any field.

Keywords: matrix multiplication, algorithm, complexity, bilinear complexity.

Литература

1. *Strassen V.* Gaussian elimination is not optimal // Numer. Math. – 1969. – V. 13. – P. 354–356. = *Штрассен В.* Алгоритм Гаусса не оптимален // Кибернетический сб. – М.: Мир, 1970. – Вып. 7. – С. 67–70.
2. *Coppersmith D., Winograd S.* Matrix Multiplication via Arithmetic Progressions // J. Symbolic Computation. – 1990. – V. 9, No 3. – P. 251–280.
3. *Bürgisser P., Clausen M., Shokrollahi M.A.* Algebraic complexity theory. – Berlin, Heidelberg: Springer, 1997. – 618 p.
4. *Waksman A.* On Winograd's algorithm for inner products // IEEE Trans. Comput. – 1970. – V. C-19. – P. 360–361.
5. *Alekseyev V.B.* On the complexity of some algorithms of matrix multiplication // J. Algorithms. – 1985. – V. 6, No 1. – P. 71–85.
6. *Laderman J.D.* A noncommutative algorithm for multiplying 3×3 matrices using 23 multiplications // Bull. Amer. Math. Soc. – 1976. – V. 82, No 1. – P. 126–128.
7. *Bläser M.* On the complexity of the multiplication of matrices of small formats // J. Complexity. – 2003. – V. 19. – P. 43–60.
8. *Макаров О.М.* Некоммутативный алгоритм умножения квадратных матриц пятого порядка, использующий сто умножений // Журн. вычисл. матем. и матем. физики. – 1987. – Т. 27, № 2. – С. 311–315.
9. *Смирнов А.В.* О билинейной сложности и практических алгоритмах умножения матриц // Журн. вычисл. матем. и матем. физики. – 2013. – Т. 53, № 12. – С. 1970–1984.
10. *Winograd S.* On multiplication of 2×2 matrices // Linear Algebra Appl. – 1971. – V. 4. – P. 381–388.
11. *Hopcroft J.E., Musinski J.* Duality applied to the complexity of matrix multiplication and other bilinear forms // SIAM J. Comput. – 1973. – V. 2, No 3. – P. 159–173.
12. *Hopcroft J.E., Kerr L.R.* On minimizing the number of multiplications necessary for matrix multiplication // SIAM J. Appl. Math. – 1971. – V. 20, No 1. – P. 127–148.
13. *Алексеев В.Б., Смирнов А.В.* О точной и приближенной билинейных сложностях умножения матриц размеров 4×2 и 2×2 // Современные проблемы математики. – 2013. – Вып. 17. – С. 135–152.
14. *Bini D., Capovani M., Lotti G., Romani F.* $O\left(n^{2.7799}\right)$ complexity for approximate matrix multiplication // Inform. Process. Lett. – 1979. – V. 8, No 5. – P. 234–235.
15. *Landsberg J.M.* The border rank of the multiplication of 2×2 matrices is seven // J. Amer. Math. Soc. – 2006. – V. 19, No 2. – P. 447–459.

Поступила в редакцию
15.08.14

Алексеев Валерий Борисович – доктор физико-математических наук, профессор, заведующий кафедрой математической кибернетики, Московский государственный университет имени М.В. Ломоносова, г. Москва, Россия.

E-mail: *vbalekseev@rambler.ru*