ЖУРНАЛ ВЫЧИСЛИТЕЛЬНОЙ МАТЕМАТИКИ И МАТЕМАТИЧЕСКОЙ ФИЗИКИ, 2024, том 64, № 9, с. 1578–1588

ОБЩИЕ ЧИСЛЕННЫЕ МЕТОДЫ

УДК 519.615.5

ОТЫСКАНИЕ КОМПЛЕКСНОЗНАЧНЫХ РЕШЕНИЙ УРАВНЕНИЙ БРЕНТА СВЕДЕНИЕМ К НЕЛИНЕЙНОЙ ЗАДАЧЕ НАИМЕНЬШИХ КВАДРАТОВ

© 2024 г. И. Е. Капорин[1,*]

[1] 119333 Москва, ул.Вавилова, 44, ФИЦ ИУ РАН, Россия
**e-mail: igorkaporin@mail.ru*

Поступила в редакцию 12.03.2024 г.
Переработанный вариант 12.03.2024 г.
Принята к публикации 31.05.2024 г.

Отыскание нетривиальных решений трилинейных уравнений Брента соответствует построению асимптотически быстрых алгоритмов перемножения матриц является важной, но в общем случае весьма сложной вычислительной задачей. Предлагаются способы параметризации уравнений Брента, основанные на использовании симметрий тензора матричного произведения, которые позволяют многократно уменьшить размерность задачи. Численное решение полученных трилинейных или кубических систем нелинейных уравнений осуществляется посредством сведения к нелинейной задаче наименьших квадратов и применения к ней специально разработанного итерационного метода, не требующего вычисления производных. Найденные решения параметризованных уравнений Брента, как правило, имеют ранг не больший (а иногда и меньший) по сравнению с известными результатами. Так, получен алгоритм перемножения двух матриц 4-го порядка за 48 активных умножений. Библ. 16. Табл. 1.

Ключевые слова: уравнения Брента; быстрое умножение матриц; алгоритм Штрассена; нелинейная задача наименьших квадратов.

DOI: 10.31857/S0044466924090015, **EDN**: WLIAJA

1. ВВЕДЕНИЕ

Уравнения, введенные Р. Брентом [1] в связи с задачей построения быстрых алгоритмов перемножения матриц, являются трилинейными относительно неизвестных. Формально, эти уравнения представляют собой каноническое разложение ранга r трехмерного $n_3n_1 \times n_1n_2 \times n_2n_3$-тензора специального вида, см. ниже (1). Точное решение уравнений Брента обеспечивает существование $O(N^\omega)$–алгоритма перемножения двух $N \times N$-матриц, где $\omega = 3\log r/\log(n_1n_2n_3)$. Соответствующую базовую конструкцию будем сокращенно обозначать как $(n_1, n_2, n_3; r)$-алгоритм. В некоторых случаях, частные решения уравнений Брента могут быть получены «на кончике пера», например в случаях, соответствующих алгоритмам Штрассена [2] ($n_1 = n_2 = n_3 = 2,\ r = 7$) или Ладермана [3], ($n_1 = n_2 = n_3 = 3,\ r = 23$) а также связанных с результатами Пана [4] по перемножению двух или трех независимых пар матриц. Однако для отыскания потенциально лучших схем матричного умножения приходится прибегать к попыткам численного решения уравнений Брента увеличенных размеров. Одним из последних достижений в этом направлении является (3,3,6;40)–алгоритм Смирнова [5]. Важным с практической точки зрения является преобразование полученного алгоритма, например с использованием технологии смены базиса [6] с целью понижения числа операций сложения и умножения на константу.

В настоящей работе рассматриваются некоторые известные и предлагаются новые способы специализации структуры решения уравнений Брента, предположительно не повышающие минимальный ранг, при котором решение существует. Частный случай предлагаемой параметризации (см. ниже (23)) был впервые представлен в [7] для вещественных матриц 3-го порядка для использования в качестве тестовой задачи.

С помощью предлагаемого итерационного метода решения нелинейной задачи наименьших квадратов корректность предлагаемых конструкций проверена экспериментально для значительного количества задач малого размера. При этом получены новые улучшения для малых размеров матриц, в частности, представлены алгоритмы типа (2,4,5;32) и (4,4,4;48).

2. ОБЩИЕ УРАВНЕНИЯ БРЕНТА И ИХ РЕДУЦИРОВАННАЯ ФОРМА

Ниже рассмотрим общий случай, когда размеры матриц n_1, n_2, n_3 принимают произвольные значения.

2.1. Стандартные уравнения Брента

В работе Р. Брента [1] была предложена следующая система трилинейных уравнений, где в качестве неизвестных фигурируют коэффициенты $n_1 \times n_3$-матриц X_t, $n_2 \times n_1$-матриц Y_t и $n_3 \times n_2$-матриц Z_t:

$$-\delta(i_2 - j_1)\delta(j_2 - k_1)\delta(k_2 - i_1) + \sum_{t=1}^{r}(X_t)_{i_2,i_1}(Y_t)_{j_2,j_1}(Z_t)_{k_2,k_1} = 0 \tag{1}$$

при всех

$$1 \leqslant i_2, j_1 \leqslant n_1, \quad 1 \leqslant j_2, k_1 \leqslant n_2, \quad 1 \leqslant k_2, i_1 \leqslant n_3.$$

Любому решению уравнения Брента при $r < n_1n_2n_3$ отвечает тройка нетривиальных билинейных алгоритмов типа $(n_1, n_2, n_3; r)$, позволяющих перемножать пары матриц размеров $n_1 \times n_2$ на $n_2 \times n_3$, или $n_2 \times n_3$ на $n_3 \times n_1$, или $n_3 \times n_1$ на $n_1 \times n_2$ с использованием r умножений в каждом из трех случаев. Чтобы в этом убедиться, достаточно домножить (1) на $(A)_{i_1,i_2}(B)_{j_1,j_2}$ и просуммировать по всем i_1, i_2, j_1, j_2, что приводит к

$$(AB)_{k_2,k_1} = \sum_{t=1}^{r}\left(\sum_{i_1,i_2}(X_t)_{i_2,i_1}(A)_{i_1,i_2}\right)\left(\sum_{j_1,j_2}(Y_t)_{j_2,j_1}(B)_{j_1,j_2}\right)(Z_t)_{k_2,k_1},$$

и аналогично в остальных двух случаях. Транспонирование дает еще три алгоритма, например, для умножения матриц $n_3 \times n_2$ на $n_2 \times n_1$ и т.д.

При $n \gg 1$ рекурсивное использование соотношений (1) путем подходящего разбиения матриц на блоки приводит к алгоритму перемножения двух $n \times n$-матриц за $O(n^{\omega})$ операций, где

$$\omega = \frac{3\log r}{\log(n_1n_2n_3)}. \tag{2}$$

Следует заметить, что отыскание решений уравнения Брента при значениях r, близких к минимально возможным, является крайне сложной задачей. Уравнение (1) появилось в работе [1] как инструмент построения алгоритмов матричного умножения, потенциально более эффективных, чем метод Штрассена [2]. Однако потребовалось более сорока лет, чтобы наконец в работе [5] появилось первое нетривиальное (в смысле выполнения условия $\omega = \log_{n_1n_2n_3} r^3 < \log_2 7$) решение уравнения Брента типа (3,3,6;40) для малых размеров задачи. Указанное решение обеспечивает построение $O(N^{2.7743})$-алгоритма перемножения двух матриц размеров $N \times N$. Позже в [6] была показана возможность сократить почти на порядок число операций типа сложения и умножения на константу в этом алгоритме.

2.2. Эквивалентное трилинейное тождество

Как заметил В.Я. Пан (см. [4] и цитированные там источники), решение уравнений Брента эквивалентно построению трилинейного тождества вида

$$\operatorname{tr}(ABC) = \sum_{t=1}^{r}\operatorname{tr}(X_tA)\operatorname{tr}(Y_tB)\operatorname{tr}(Z_tC), \tag{3}$$

где X_t, Y_t, Z_t те же, что и ранее искомые матричные коэффициенты, а A, B, C представляют собой свободные параметры в виде матриц с размерами $n_3 \times n_1$, $n_1 \times n_2$ и $n_2 \times n_3$ соответственно.

Действительно, (3) легко получаются из (1) домножением на $(A)_{i_1,i_2}(B)_{j_1,j_2}(C)_{k_1,k_2}$ и суммированием по всем $i_1, i_2, j_1, j_2, k_1, k_2$. В обратную сторону, чтобы получить уравнения Брента, достаточно подставить в (3)

$$A = e_{i_1}e_{i_2}^{\mathsf{T}}, \qquad B = e_{j_1}e_{j_2}^{\mathsf{T}}, \qquad C = e_{k_1}e_{k_2}^{\mathsf{T}}, \tag{4}$$

где e_i представляет собой i-й единичный столбцовый вектор соответствующей длины, и использовать соотношения типа $e_{i_2}^{\mathsf{T}}e_{j_1} = \delta(i_2 - j_1)$ и $\operatorname{tr}(X_te_{i_1}e_{i_2}^{\mathsf{T}}) = (X_t)_{i_2,i_1}$.

2.3. Параметризованные уравнения Брента

Трилинейная система (1) включает $M = (n_1n_2n_3)^2$ уравнений и $N = (n_3n_1 + n_1n_2 + n_2n_3)r$ неизвестных. Решение этой задачи либо не существует (при слишком малых r), либо неединственно, причем изолированные решения отсутствуют (т.е. каждая ветвь решений представляет собой непрерывное многообразие). Кроме того, попытки применения численных методов осложняются наличием “приближенных решений”, для которых ранг разложения обычно меньше, чем для точных решений уравнений Брента. При этом для таких “решений” стремление невязки уравнений Брента к нулю сопровождается неограниченным ростом некоторых коэффициентов разложения. Попытки применения к уравнениям Брента общих оптимизационных методов типа описанных в [8] оказываются недостаточно эффективными. Здесь можно упомянуть метод попеременной квадратичной минимизации, на основе которого получен результат [5], однако там были использованы дополнительные эвристики, направленные на отыскание решений с рациональными коэффициентами. Таким образом, по крайней мере для $\min(n_1, n_2, n_3) \geqslant 3$, отыскание частных решений задачи (1) является исключительно трудной вычислительной проблемой. Поэтому представляет интерес сведение уравнений Брента к задаче с меньшим числом уравнений и неизвестных. Подход, обсуждавшийся в [13] (в более узком контексте $n_1 = n_2 = n_3$), основан на частичном переносе свойств инвариантности левой части (3) на правую часть (3). Это достигается за счет специальной параметризации матричных коэффициентов X_t, Y_t, Z_t.

Определяющим свойством инвариантности функции

$$\tau(A, B, C) = \operatorname{tr}(ABC) \tag{5}$$

относительно преобразования аргументов является тождество

$$\tau(A, B, C) \equiv \tau\left(\mathrm{M}^{-1}A\mathrm{K},\ \mathrm{K}^{-1}B\Lambda,\ \Lambda^{-1}C\mathrm{M}\right), \tag{6}$$

где K, Λ, и M – произвольные невырожденные матрицы порядков n_1, n_2 и n_3 соответственно. Так, можно показать, что из (6) следует (5) с точностью до ненулевого скалярного множителя. Поэтому потребуем, чтобы свойство (6) соблюдалось также и для правой части (3), хотя бы и для ограниченного выбора матриц $\mathrm{K}, \Lambda, \mathrm{M}$:

$$\begin{aligned}\sum_{t=1}^{r} \operatorname{tr}(X_tA)\operatorname{tr}(Y_tB)\operatorname{tr}(Z_tC) &= \sum_{t=1}^{r} \operatorname{tr}(X_t\mathrm{M}^{-1}A\mathrm{K})\operatorname{tr}(Y_t\mathrm{K}^{-1}B\Lambda)\operatorname{tr}(Z_t\Lambda^{-1}C\mathrm{M}) = \\ &= \sum_{t=1}^{r} \operatorname{tr}(\mathrm{K}X_t\mathrm{M}^{-1}A)\operatorname{tr}(\Lambda Y_t\mathrm{K}^{-1}B)\operatorname{tr}(\mathrm{M}Z_t\Lambda^{-1}C).\end{aligned}$$

Естественным способом удовлетворить это равенство для некоторых троек матриц $\mathrm{K}, \Lambda, \mathrm{M}$, представляются условия

$$\mathrm{K}X_t\mathrm{M}^{-1} = X_{\sigma(t)}, \qquad \Lambda Y_t\mathrm{K}^{-1} = Y_{\sigma(t)}, \qquad \mathrm{M}Z_t\Lambda^{-1} = Z_{\sigma(t)}, \qquad t = 1, 2, \ldots, r, \tag{7}$$

где $\sigma(\cdot)$ является перестановкой чисел $(1, 2, \ldots, r)$. Чтобы согласованно замкнуть предлагаемые уравнения (7), положим

$$\sigma^p(t) = t, \qquad t = 1, 2, \ldots, r \tag{8}$$

(то есть $\sigma(\cdot)$ является корнем p-й степени из тождественной подстановки) и пусть

$$r = pq.$$

Тогда определим $\sigma(\cdot)$ как прямую сумму q циклических перестановок порядка p, то есть

$$\sigma(t) = t + 1 - p\,\delta(t(\mathrm{mod}p)), \qquad t = 1, \ldots, r.$$

Например, если $r = 15$, $p = 3$ и $q = 5$, то $\sigma = (2, 3, 1, 5, 6, 4, 8, 9, 7, 11, 12, 10, 14, 15, 13)$. Наконец, потребуем также

$$\mathrm{K}^p = I_{n_1}, \qquad \Lambda^p = I_{n_2}, \qquad \mathrm{M}^p = I_{n_3}, \tag{9}$$

то есть выберем матрицы $\mathrm{K}, \Lambda, \mathrm{M}$ как корни p-й степени из единичных матриц соответствующих размеров. Отсюда получаем формулы для матричных коэффициентов

$$X_t = \mathrm{K}^sX_{t-s}\mathrm{M}^{-s}, \qquad Y_t = \Lambda^sY_{t-s}\mathrm{K}^{-s}, \qquad Z_t = \mathrm{M}^sZ_{t-s}\Lambda^{-s},$$

где

$$s = (t-1)(\mathrm{mod}\ p), \qquad t = 1, \ldots, r.$$

Учитывая, что введенные выше индексы $t-s$ образуют неубывающую ступенчатую функцию вида $1,\dots,1,p+1,\dots,p+1,\dots,r-p+1,\dots,r-p+1$, соответствующая замена обозначений матричных коэффициентов X_t, Y_t, Z_t в виде $X_t^{\text{new}} = X_{(t-1)p+1}^{\text{old}}$ приводит к формуле

$$\operatorname{tr}(ABC) = \sum_{t=1}^{q}\sum_{s=0}^{p-1} \operatorname{tr}(\mathrm{K}^s X_t \mathrm{M}^{-s} A)\operatorname{tr}(\Lambda^s Y_t \mathrm{K}^{-s} B)\operatorname{tr}(\mathrm{M}^s Z_t \Lambda^{-s} C). \tag{10}$$

Полученное при условии (9) тождество (10) дает (при переходе к покомпонентной записи) *параметризованное уравнение Брента*. Уже на этом этапе видно, что в (10) число неизвестных X_t, Y_t, Z_t сократилось в p раз по сравнению с (3).

2.4. Редуцированные уравнения Брента

Если конкретизировать выбор матриц $\mathrm{K}, \Lambda, \mathrm{M}$ в виде

$$\mathrm{K} = \operatorname{Diag}_{1\leqslant m\leqslant n_1}(\Omega^{m-1}), \qquad \Lambda = \operatorname{Diag}_{1\leqslant m\leqslant n_2}(\Omega^{m-1}), \qquad \mathrm{M} = \operatorname{Diag}_{1\leqslant m\leqslant n_3}(\Omega^{m-1}), \tag{11}$$

где

$$\Omega = \exp(2\pi \mathbf{i}/p), \qquad \mathbf{i} = \sqrt{-1}, \tag{12}$$

то можно сократить примерно в p раз также и число уравнений. Для этого достаточно подставить указанные формулы для $\mathrm{K}, \Lambda, \mathrm{M}$ в (10) и перейти к покомпонентным соотношениям типа (1), подставляя (4):

$$\begin{aligned}
&\delta(i_2-j_1)\delta(j_2-k_1)\delta(k_2-i_1) = \sum_{t=1}^{q}\sum_{s=0}^{p-1}(e_{i_2}^T \mathrm{K}^s X_t \mathrm{M}^{-s} e_{i_1})(e_{j_2}^T \Lambda^s Y_t \mathrm{K}^{-s} e_{j_1})(e_{k_2}^T \mathrm{M}^s Z_t \Lambda^{-s} e_{k_1}) = \\
&= \sum_{t=1}^{q}\sum_{s=0}^{p-1}\left(\Omega^{(i_2-1)s}(X_t)_{i_2,i_1}\Omega^{(1-i_1)s}\right)\left(\Omega^{(j_2-1)s}(Y_t)_{j_2,j_1}\Omega^{(1-j_1)s}\right)\left(\Omega^{(k_2-1)s}(Z_t)_{k_2,k_1}\Omega^{(1-k_1)s}\right) = \\
&= \sum_{t=1}^{q}\left(\sum_{s=0}^{p-1}\Omega^{(i_2-i_1+j_2-j_1+k_2-k_1)s}\right)(X_t)_{i_2,i_1}(Y_t)_{j_2,j_1}(Z_t)_{k_2,k_1} = \\
&= p\,\delta((i_2-i_1+j_2-j_1+k_2-k_1)(\operatorname{mod} p))\sum_{t=1}^{q}(X_t)_{i_2,i_1}(Y_t)_{j_2,j_1}(Z_t)_{k_2,k_1}.
\end{aligned}$$

Учитывая, что в случае $\delta(i_2-j_1)\delta(j_2-k_1)\delta(k_2-i_1)=1$ также выполнено и равенство $\delta((i_2-i_1+j_2-j_1+k_2-k_1)(\operatorname{mod} p))=1$, получаем, умножая уравнения на p^{-1}, *редуцированные уравнения Брента*:

$$-\frac{1}{p}\delta(i_2-j_1)\delta(j_2-k_1)\delta(k_2-i_1) + \sum_{t=1}^{q}(X_t)_{i_2,i_1}(Y_t)_{j_2,j_1}(Z_t)_{k_2,k_1} = 0, \tag{13а}$$

$$1\leqslant i_2, j_1 \leqslant n_1, \quad 1\leqslant j_2, k_1 \leqslant n_2, \quad 1\leqslant k_2, i_1 \leqslant n_3, \tag{13б}$$

$$(i_2-i_1+j_2-j_1+k_2-k_1)\,(\operatorname{mod} p) = 0. \tag{13в}$$

Точному решению этих уравнений отвечает алгоритм перемножения двух $n\times n$-матриц за $O(n^{\omega})$ операций, где

$$\omega = \frac{3\log pq}{\log(n_1 n_2 n_3)}. \tag{14}$$

Замечание 1. Понятно, что свойство инвариантности правой части (3) в нашем случае ограничено использованием $p-1$ троек матриц вида

$$(\mathrm{K}^m, \Lambda^m, \mathrm{M}^m), \qquad m = 1,\dots,p-1.$$

Замечание 2. Построенная задача (13) не эквивалентна каноническому тензорному разложению, а является специальным случаем *частичного* разложения тензора в силу наличия дополнительного условия (13в).

Замечание 3. Можно рассмотреть несколько более общую конструкцию вида

$$\mathrm{K} = \operatorname{Diag}_{1\leqslant j\leqslant n_1}(\Omega^{\kappa(j)}), \qquad \Lambda = \operatorname{Diag}_{1\leqslant j\leqslant n_2}(\Omega^{\lambda(j)}), \qquad \mathrm{M} = \operatorname{Diag}_{1\leqslant j\leqslant n_3}(\Omega^{\mu(j)}),$$

где $\kappa(\cdot), \lambda(\cdot), \mu(\cdot)$ — произвольные целочисленные функции. В этом случае в редуцированных уравнениях Брента условие (13в) следует заменить на

$$(\kappa(i_2) - \mu(i_1) + \lambda(j_2) - \kappa(j_1) + \mu(k_2) - \lambda(k_1)) \pmod p = 0.$$

Неизвестно, можно ли на таком пути улучшить результаты, полученные при используемом в настоящей работе простейшем выборе $\kappa(j) = \lambda(j) = \mu(j) = j - 1$ для $j \geqslant 1$.

2.5. Обобщение редуцированных уравнений Брента

Сопоставление описанной выше конструкции (10) с результатами Штрассена (где $n_1 = n_2 = n_3 = 2$ и $r = 7$) и Ладермана (где $n_1 = n_2 = n_3 = 3$ и $r = 23$) приводит к необходимости ее обобщения на случай, не предполагающий наличия подходящего разложения r на множители.

Например, для $r = 7$ при $n_1 = n_2 = n_3 = 2$, выбор $p = 1$ тривиален, а при $p = 7$ – невозможен (детальное обсуждение нижних границ для q заслуживает отдельного рассмотрения). Однако в этом случае существует решение редуцированных уравнений Брента, отвечающих разложению $r = 8$ и $p = 2$. При этом полученное завышенное значение ранга можно понизить на единицу, если заметить, что одна из пар матричных коэффициентов, например X_1 и Y_1, образована диагональными матрицами. В этом случае $\mathrm{K}X_1\mathrm{M}^{-1} = X_1$ и $\Lambda Y_1\mathrm{K}^{-1} = Y_1$, так что первая сумма по s сворачивается в один член ранга 1:

$$\sum_{s=0}^{1} \operatorname{tr}(\mathrm{K}^s X_1 \mathrm{M}^{-s} A) \operatorname{tr}(\Lambda^s Y_1 \mathrm{K}^{-s} B) \operatorname{tr}(\mathrm{M}^s Z_1 \Lambda^{-s} C) = \operatorname{tr}(X_1 A) \operatorname{tr}(Y_1 B) \operatorname{tr}\Big(\Big(\sum_{s=0}^{1} \mathrm{M}^s Z_1 \Lambda^{-s}\Big) C\Big).$$

В общем случае для понижения ранга редуцированных уравнений Брента добавим к уравнениям (13) дополнительные $3q$ линейных матричных уравнений

$$\mathrm{K}^{d_t} X_t - X_t \mathrm{M}^{d_t} = 0, \qquad \Lambda^{d_t} Y_t - Y_t \mathrm{K}^{d_t} = 0, \qquad \mathrm{M}^{d_t} Z_t - Z_t \Lambda^{d_t} = 0, \qquad t = 1, \ldots, q, \tag{15}$$

где числа d_t образуют неубывающую последовательность,

$$1 \leqslant d_1 \leqslant d_2 \leqslant \ldots, \leqslant d_q = p,$$

и каждое из них является делителем p. Тогда, с учетом представления $s = u + d_t v$, где $0 \leqslant u \leqslant d_t - 1$ и $0 \leqslant v \leqslant \frac{p}{d_t} - 1$, (10) преобразуется к виду

$$\operatorname{tr}(ABC) = \sum_{t=1}^{q} \frac{p}{d_t} \sum_{u=0}^{d_t - 1} \operatorname{tr}(\mathrm{K}^u X_t \mathrm{M}^{-u} A) \operatorname{tr}(\Lambda^u Y_t \mathrm{K}^{-u} B) \operatorname{tr}(\mathrm{M}^u Z_t \Lambda^{-u} C). \tag{16}$$

Таким образом, ранг трилинейного разложения уменьшается до величины

$$r = \sum_{t=1}^{q} d_t. \tag{17}$$

Например, результату Штрассена соответствует разложение $7 = 1 + 2 + 2 + 2$ получаемое при $p = 2$ и $q = 4$, а результату Ладермана отвечает $23 = 1 + 2 + 4 + 4 + 4 + 4 + 4$ (при $p = 4$ и $q = 7$). Далее используется формула

$$r = (q - q_0)p + \sum_{t=1}^{q_0} d_t, \tag{18}$$

где обычно достаточно брать $q_0 \leqslant 2$.

Замечание 4. Одно из условий (15) является избыточным при используемом выборе диагональных матриц $\mathrm{K}, \Lambda, \mathrm{M}$, то есть выводится из остальных двух. Кроме того, можно сократить число неизвестных для каждого слагаемого при $d_t < p$, так как уравнения (15) фактически лишь определяют структуру разреженности матриц X_t, Y_t, Z_t согласно условиям

$$(X_t)_{i_2, i_1} = 0, \qquad i_2 - i_1 \neq 0 \pmod{p/d_t}, \quad t = 1, \ldots, q, \tag{19а}$$

$$(Y_t)_{j_2, j_1} = 0, \qquad j_2 - j_1 \neq 0 \pmod{p/d_t}, \quad t = 1, \ldots, q, \tag{19б}$$

$$(Z_t)_{k_2, k_1} = 0, \qquad k_2 - k_1 \neq 0 \pmod{p/d_t}, \quad t = 1, \ldots, q. \tag{19в}$$

Отметим также, что для слагаемых с $d_t = p$ условия (15) или (19) выполняются при любых X_t, Y_t, Z_t в силу (9).

Замечание 5. Наложение условий (15) отвечает использованию общего вида перестановки $\sigma(\cdot)$, удовлетворяющей условию $\sigma^p(t) = t$.

3. РЕДУЦИРОВАННЫЕ УРАВНЕНИЯ БРЕНТА ДЛЯ КВАДРАТНЫХ МАТРИЦ

В случае

$$n_1 = n_2 = n_3 = n \tag{20}$$

можно еще в три раза сократить число неизвестных за счет замены матриц Y_t и Z_t на матрицы X_s с подходящими индексами s.

3.1. Использование инвариантности относительно циклической перестановки

Потребуем, чтобы правая часть тождества (3) обладала тем же свойством, что и левая, а именно, была бы инвариантна относительно циклической перестановки матричных аргументов:

$$\operatorname{tr}(ABC) = \operatorname{tr}(BCA) = \operatorname{tr}(CAB). \tag{21}$$

Для этого произведем замену матриц вида $Y_t = X_{\alpha(t)}$ и $Z_t = X_{\alpha^2(t)}$:

$$\operatorname{tr}(ABC) = \sum_{t=1}^{r} \operatorname{tr}(X_k A)\, \operatorname{tr}(X_{\alpha(k)} B)\, \operatorname{tr}(X_{\alpha^2(k)} C), \tag{22}$$

где $\alpha(\cdot)$ является перестановкой $(1, 2, \ldots, r)$ такой, что ее для ее третьей степени справедливо $\alpha^3(t) = t$, то есть α представляет собой корень 3-й степени из тождественной перестановки длины r. Эта конструкция была представлена в [13], а ее обобщение – в [7], где впервые использована матричная параметризация уравнений Брента, обсуждаемая в настоящей работе.

При условии (20) упрощается выбор диагональных матриц в виде $\mathrm{K} = \Lambda = \mathrm{M}$. Тогда параметризованная форма трилинейного тождества приобретает вид

$$\operatorname{tr}(ABC) = \sum_{t=1}^{q} \sum_{s=0}^{p-1} \operatorname{tr}(\Lambda^s X_t \Lambda^{-s} A)\, \operatorname{tr}(\Lambda^s X_{\beta(t)} \Lambda^{-s} B)\, \operatorname{tr}(\Lambda^s X_{\beta^2(t)} \Lambda^{-s} C), \tag{23}$$

где

$$\Lambda = \operatorname{Diag}_{1 \leqslant j \leqslant n}(\Omega^{j-1}), \qquad \Omega = \exp(2\pi \mathbf{i}/p), \qquad \mathbf{i} = \sqrt{-1}. \tag{24}$$

и $\beta(\cdot)$ является перестановкой индексов $(1, 2, \ldots, q)$ такой, что $\beta^3(t) = t$, то есть $\beta(\cdot)$ представляет собой корень 3-й степени из тождественной перестановки длины q.

3.2. Редуцированные уравнения Брента для квадратных матриц

Переход к покомпонентной записи приводит к соответствующим версиям уравнений Брента:

$$\delta(i_2 - j_1)\delta(j_2 - k_1)\delta(k_2 - i_1) = \sum_{t=1}^{r} (X_t)_{i_2,i_1} (X_{\alpha(t)})_{j_2,j_1} (X_{\alpha^2(t)})_{k_2,k_1} \tag{25}$$

и, для параметризованной версии,

$$\delta(i_2 - j_1)\delta(j_2 - k_1)\delta(k_2 - i_1) = \sum_{t=1}^{q} \sum_{s=0}^{p-1} (\Lambda^s X_t \Lambda^{-s})_{i_2,i_1} (\Lambda^s X_{\beta(t)} \Lambda^{-s})_{j_2,j_1} (\Lambda^s X_{\beta^2(t)} \Lambda^{-s})_{k_2,k_1}, \tag{26}$$

где $1 \leqslant i_1, i_2, j_1, j_2, k_1, k_2 \leqslant n$. Подставляя явное выражение для Λ из (24), получаем редуцированные уравнения Брента для квадратных матриц:

$$-\frac{1}{p}\delta(i_2 - j_1)\delta(j_2 - k_1)\delta(k_2 - i_1) + \sum_{t=1}^{q} (X_t)_{i_2,i_1} (X_{\beta(t)})_{j_2,j_1} (X_{\beta^2(t)})_{k_2,k_1} = 0, \tag{27а}$$

$$1 \leqslant i_1, i_2, j_1, j_2, k_1, k_2 \leqslant n, \tag{27б}$$

$$(i_2 - i_1 + j_2 - j_1 + k_2 - k_1) \pmod p = 0. \tag{27в}$$

Используя представление $q = q_0 + 3q_1$, без ограничения общности можно положить

$$\begin{aligned} t &= (1\ 2\ 3\ \ldots\ q_0\ \ q_0+1\ \ q_0+2\ \ q_0+3\ \ \ldots\ \ q_0+3q_1), \\ \beta(t) &= (1\ 2\ 3\ \ldots\ q_0\ \ q_0+2\ \ q_0+3\ \ q_0+1\ \ \ldots\ \ q_0+3q_1-2), \\ \beta^2(t) &= (1\ 2\ 3\ \ldots\ q_0\ \ q_0+3\ \ q_0+1\ \ q_0+2\ \ \ldots\ \ q_0+3q_1-1). \end{aligned}$$

Так, для $q = 12$ при $q_0 = q_1 = 3$, такая конструкция имеет вид

$$\begin{aligned} t &= (1\ 2\ 3\ 4\ 5\ 6\ 7\ 8\ 9\ 10\ 11\ 12),\\ \beta(t) &= (1\ 2\ 3\ 5\ 6\ 4\ 8\ 9\ 7\ 11\ 12\ 10),\\ \beta^2(t) &= (1\ 2\ 3\ 6\ 4\ 5\ 9\ 7\ 8\ 12\ 10\ 11). \end{aligned}$$

Соответственно, параметризованная форма уравнений Брента для квадратных матриц принимает вид

$$\begin{aligned} &-\delta(i_2-j_1)\delta(j_2-k_1)\delta(k_2-i_1)+ \\ &+\sum_{t=1}^{q_0}\sum_{s=0}^{p-1}(\Lambda^s X_t\Lambda^{-s})_{i_2,i_1}(\Lambda^s X_t\Lambda^{-s})_{j_2,j_1}(\Lambda^s X_t\Lambda^{-s})_{k_2,k_1}+ \\ &+\sum_{t=1}^{q_1}\sum_{s=0}^{p-1}((\Lambda^s X_{q_0+3t-2}\Lambda^{-s})_{i_2,i_1}(\Lambda^s X_{q_0+3t-1}\Lambda^{-s})_{j_2,j_1}(\Lambda^s X_{q_0+3t}\Lambda^{-s})_{k_2,k_1}+ \\ &\quad+(\Lambda^s X_{q_0+3t-1}\Lambda^{-s})_{i_2,i_1}(\Lambda^s X_{q_0+3t}\Lambda^{-s})_{j_2,j_1}(\Lambda^s X_{q_0+3t-2}\Lambda^{-s})_{k_2,k_1}+ \\ &\quad+(\Lambda^s X_{q_0+3t}\Lambda^{-s})_{i_2,i_1}(\Lambda^s X_{q_0+3t-2}\Lambda^{-s})_{j_2,j_1}(\Lambda^s X_{q_0+3t-1}\Lambda^{-s})_{k_2,k_1})=0, \end{aligned} \tag{28}$$

и редуцированная форма перепишется в виде

$$\begin{aligned} &-\frac{1}{p}\delta(i_2-j_1)\delta(j_2-k_1)\delta(k_2-i_1)+ \\ &+\sum_{t=1}^{q_0}(X_t)_{i_2,i_1}(X_t)_{j_2,j_1}(X_t)_{k_2,k_1}+ \\ &+\sum_{t=1}^{q_1}((X_{q_0+3t-2})_{i_2,i_1}(X_{q_0+3t-1})_{j_2,j_1}(X_{q_0+3t})_{k_2,k_1}+ \\ &\quad+(X_{q_0+3t-1})_{i_2,i_1}(X_{q_0+3t})_{j_2,j_1}(X_{q_0+3t-2})_{k_2,k_1}+ \\ &\quad+(X_{q_0+3t})_{i_2,i_1}(X_{q_0+3t-2})_{j_2,j_1}(X_{q_0+3t-1})_{k_2,k_1})=0 \end{aligned} \tag{29}$$

для тех же $i_1,\dots,k_2$, что и в (27).

3.3. Обобщенная форма редуцированных уравнений Брента

Добавляя к уравнениям (29) дополнительные условия типа (15)

$$\Lambda^{d_t^{(0)}}X_t - X_t\Lambda^{d_t^{(0)}} = 0, \quad t = 1,\dots,q_0, \tag{30а}$$

$$\Lambda^{d_t^{(1)}}X_u - X_u\Lambda^{d_t^{(1)}} = 0, \quad u-q_0 \in \{3t-2,\ 3t-1,\ 3t\}, \quad t = 1,\dots,q_1, \tag{30б}$$

(здесь, как и ранее, предполагаем, что d_t^0 и d_t^1 являются делителями p), получаем обобщенную форму редуцированных уравнений Брента. Аналогично уравнениям (19), условия (30) можно заменить на требование разреженности вида

$$(X_t)_{i_2,i_1} = 0, \quad i_2 - i_1 \neq 0 \ (\mathrm{mod}\ \frac{p}{d_t^{(0)}}), \quad 1 \leqslant t \leqslant q_0, \tag{31а}$$

$$(X_u)_{i_2,i_1} = 0, \quad i_2 - i_1 \neq 0 \ (\mathrm{mod}\ \frac{p}{d_t^{(1)}}), \quad u-q_0 \in \{3t-2,\ 3t-1,\ 3t\}, \quad 1 \leqslant t \leqslant q_1. \tag{31б}$$

Соответствующие параметризованные уравнения Брента принимают вид

$$\begin{aligned} &-\delta(i_2-j_1)\delta(j_2-k_1)\delta(k_2-i_1) \quad + \\ &+\sum_{t=1}^{q_0}\frac{p}{d_t^{(0)}}\sum_{s=0}^{d_t^{(0)}-1}(\Lambda^s X_t\Lambda^{-s})_{i_2,i_1}(\Lambda^s X_t\Lambda^{-s})_{j_2,j_1}(\Lambda^s X_t\Lambda^{-s})_{k_2,k_1}+ \\ &+\sum_{t=1}^{q_1}\frac{p}{d_t^{(1)}}\sum_{s=0}^{d_t^{(1)}-1}((\Lambda^s X_{q_0+3t-2}\Lambda^{-s})_{i_2,i_1}(\Lambda^s X_{q_0+3t-1}\Lambda^{-s})_{j_2,j_1}(\Lambda^s X_{q_0+3t}\Lambda^{-s})_{k_2,k_1}+ \\ &\quad+(\Lambda^s X_{q_0+3t-1}\Lambda^{-s})_{i_2,i_1}(\Lambda^s X_{q_0+3t}\Lambda^{-s})_{j_2,j_1}(\Lambda^s X_{q_0+3t-2}\Lambda^{-s})_{k_2,k_1}+ \\ &\quad+(\Lambda^s X_{q_0+3t}\Lambda^{-s})_{i_2,i_1}(\Lambda^s X_{q_0+3t-2}\Lambda^{-s})_{j_2,j_1}(\Lambda^s X_{q_0+3t-1}\Lambda^{-s})_{k_2,k_1})=0, \end{aligned} \tag{32}$$

и ранг такого разложения равен

$$r = \sum_{t=1}^{q_0} d_t^{(0)} + 3\sum_{t=1}^{q_1} d_t^{(1)}. \tag{33}$$

4. НЕЛИНЕЙНЫЙ МЕТОД НАИМЕНЬШИХ КВАДРАТОВ

Ниже рассмотрим нелинейный метод наименьших квадратов, применимый в случае комплекснозначной функции невязки $f : \mathbf{C}^n \to \mathbf{C}^m$ и не требующий вычисления ее производных. Для достаточно гладкой функции невязки f рассмотрим нелинейную задачу наименьших квадратов

$$x = \arg\min_{x\in\mathbf{C}^n} \frac{1}{2}\|f(x)\|^2, \tag{34}$$

где $\|f(x)\|^2 = f^{\mathsf{H}}(x)f(x)$ и $f^{\mathsf{H}} = (\bar{f})^{\mathsf{T}}$ обозначает операцию транспонирования с сопряжением; также, будем обозначать $f = f(x)$. Предлагается, аналогично [11], использовать шаги минимизации нормы невязки вдоль направлений, выбираемых в подпространствах, образованных столбцами псевдослучайных прямоугольных $n \times d$-матриц U, где $d \leqslant n$. При этом предпочтительна (хотя бы приближенная) ортонормированность столбцов u_j и нормализованность строк каждой матрицы $U = [u_1|...|u_d]$. Если x является текущим приближением к решению, то следующее приближение строится в виде

$$x_+ = x + \alpha U s, \qquad 0 < \alpha < 2,$$

где параметр длины шага α получается из условия типа Армихо [12] (см. ниже (38)). В качестве вектора $s \in \mathbf{C}^d$ берется решение регуляризованной линейной задачи наименьших квадратов

$$s = \arg\min_{s\in\mathbf{C}^d} \left(\|f + Vs\|^2 + \xi\|s\|^2\right), \qquad 0 < \xi \ll \operatorname{tr}(V^{\mathsf{H}}V), \tag{35}$$

где ξ является параметром регуляризации, а $n \times d$-матрица V строится в виде

$$V = \left[\frac{f(x+\zeta u_1) - f(x)}{\zeta} \mid ... \mid \frac{f(x+\zeta u_d) - f(x)}{\zeta}\right] \in \mathbf{C}^{n\times d}. \tag{36}$$

Отсюда получаем

$$s = -(V^{\mathsf{H}}V + \xi I)^{-1}V^{\mathsf{H}}f.$$

Теперь, определяя величины

$$\theta^2 = \frac{-s^{\mathsf{H}}V^{\mathsf{H}}f}{f^{\mathsf{H}}f}, \qquad \gamma^2 = \frac{s^{\mathsf{H}}s}{f^{\mathsf{H}}f},$$

можно показать, что при выполнении *условия сходимости*

$$\left\|\frac{f(x+\alpha U s) - f}{\alpha} - Vs\right\| \leqslant \left(\left(1 - \frac{\alpha}{2}\right)\theta^2 + \frac{\alpha}{2}\xi\gamma^2\right)\|f\|, \tag{37}$$

справедлива следующая оценка убывания квадрата нормы невязки:

$$\frac{\|f(x+\alpha U s)\|^2}{\|f\|^2} \leqslant 1 - \left(\frac{\alpha(2-\alpha)\theta^2 + \alpha^2\xi\gamma^2}{2}\right)^2. \tag{38}$$

Таким образом, (37) представляет достаточное условие для убывания нормы невязки. Заметим, что левая часть (37) представляет собой норму разности двух различных аппроксимаций одной и той же величины $J(x)Us$, где $J(x) = \partial f/\partial x$ обозначает якобиан функции $f(x)$. Аналогичный подход был использован в [9, 10] и адаптирован к отказу от вычисления производных в [7].

Оценка (38) непосредственно используется для выбора подходящего значения параметра шага $\alpha \in \{1,\ 1/2,\ 1/4, \ldots\}$ аналогично [12]. Так, в случае липшицева якобиана J можно убедиться в справедливости (37) для всех достаточно малых α, что обеспечивает корректность предлагаемого способа выбора параметра шага. Естественным способом остановки итераций является (помимо задания предела числа итераций) появление предельно малых значений величины $\alpha(2-\alpha)\theta^2 + \alpha^2\xi\gamma^2$.

Замечание 6. В качестве псевдослучайной последовательности использовались точки на единичном круге в комплексной плоскости, получаемые по рекуррентной формуле

$$\eta_{k+1} = \eta_k/\bar{\eta}_k, \qquad k = 1, 2, \ldots,$$

с начальным условием, например, $\eta_0 = 0.6 + 0.8\,\mathbf{i}$. Элементы матрицы $U \in \mathbf{C}^{n\times d}$ задавались в виде $(U_k)_{i,j} = n^{-1/2}\eta_{i+(j-1)n+k}$, $1 \leqslant i \leqslant n$, $1 \leqslant j \leqslant d$, где через k обозначен номер итерации. Начальный вектор x также задавался псевдослучайным образом, например, $(x^{(m)})_j = 0.25\eta_{mj}$, $m = 1, 2, \ldots, m_{\max}$, где $m_{\max}$ обозначает число повторных применений итерационного метода для вычисления решения уравнения $f(x) = 0$.

Замечание 7. Для расчетов в двойной точности, в (35) и (36) выбирались значения $\xi = 10^{-13}$ и $\zeta = 10^{-8}$ соответственно. Максимальное число итераций и рестартов ограничивалось несколькими сотнями.

5. РЕЗУЛЬТАТЫ ЧИСЛЕННОГО ПОИСКА МАТРИЧНЫХ КОЭФФИЦИЕНТОВ

Расчеты были выполнены на настольном компьютере Pentium® Dual-Core CPU E6600@3.06 GHz (объем оперативной памяти 3.25 Gbytes). Использовался транслятор G95 Fortran 95 (http://www.g95.org).

Таблица 1. Найденные решения редуцированных уравнений Брента с наибольшими значениями p. В большинстве случаев удалось воспроизвести известные значения ранга, и даже улучшить два результата. Однако в трех случаях ранг увеличился.

n_1	n_2	n_3	r	ω	p	q	$\mathcal{D}$	$r(p,\mathcal{D})$	$\omega(p,\mathcal{D})$
2	2	2	7	2.807…	3	3	1	7	2.807…
2	2	3	11	2.894…	2	6	1	11	2.894…
2	2	4	14	2.855…	2	7	$\emptyset$	14	2.855…
2	2	5	18	2.894…	3	6	$\emptyset$	18	2.894…
2	2	6	21	2.873…	6	4	3	21	2.873…
2	3	3	15	2.810…	3	5	$\emptyset$	15	2.810…
2	3	4	20	2.827…	4	5	$\emptyset$	20	2.827…
2	3	5	25	2.839…	5	5	$\emptyset$	25	2.839…
2	3	6	30	2.847…	6	5	$\emptyset$	30	2.847…
2	4	4	26	2.820…	4	7	2	26	2.820…
2	4	5	33	2.843…	4	8	$\emptyset$	**32**	**2.818**…
2	4	6	39	2.839…	6	7	3	39	2.839…
2	5	5	40	2.828…	5	8	$\emptyset$	40	2.828…
2	5	6	48	2.836…	6	8	$\emptyset$	48	2.836…
2	6	6	57	2.836…	6	10	3	48	2.836…
3	3	3	23	2.854…	4	7	1;2	23	2.854…
3	3	4	**29**	**2.818**…	4	8	2	30	2.847…
3	3	5	36	2.824…	3	12	$\emptyset$	36	2.824…
3	3	6	**40**	**2.774**…	3	14	$\emptyset$	42	2.810…
3	4	5	**47**	**2.821**…	4	12	$\emptyset$	48	2.836…
2	2	2	7	2.807…	2	$4 = 1 + 3\cdot 1$	1	7	2.807…
3	3	3	23	2.854…	4	$7 = 4 + 3\cdot 1$	1;2	23	2.854…
4	4	4	49	2.807…	4	$12 = 3 + 3\cdot 3$	$\emptyset$	**48**	**2.792**…

Вектор x численного решения считался допустимым, если его невязка была близка к пределу машинной точности, а максимум модуля компонент был меньше единицы:

$$\|f(x)\| < 10^{-15}, \qquad \|x\|_\infty < 1. \tag{39}$$

Выполнение этих требований позволяет предполагать, что известные эквивалентные преобразования решения уравнений Брента (см., например, [14]) позволят получить явные выражения коэффициентов X_t, Y_t и Z_t через квадратичные и/или кубические иррациональности. Если получался вектор x с хорошей невязкой, но с нормой больше единицы, то он умножался на 0.9 и использовался как начальное приближение для рестарта, возможно, несколько раз. В итоге удавалось получить вектор x, удовлетворяющий (39).

Результаты расчетов приведены в таблице 1. В левой половине приводятся известные результаты, опубликованные в [2,3,5,15,16], а в правой – настройки редуцированных уравнений Брента и соответствующие значения ранга. Через $\mathcal{D}$ обозначается множество $\{d_1, d_2, \ldots\}$ нетривиальных делителей p, использованных для построения обобщенной формы редуцированных уравнений Брента (13),(19) или, для квадратных матриц, (29), (31). В нижней части таблицы представлены результаты для квадратных матриц с указанием параметров перестановок $\beta(\cdot)$, использованных в (27), в виде $q = q_0 + 3q_1$.

В случае несовпадения рангов, меньшее значение выделено жирным шрифтом. Из таблицы видно улучшение до $r = 32$, достигнутое для задачи перемножения прямоугольных матриц с $n_1 = 2$, $n_2 = 4$ и $n_3 = 5$, а также улучшение ранга до $r = 48$ для алгоритма перемножения двух квадратных матриц 4-го порядка.

Автор выражает благодарность Е. Е. Тыртышникову за полезные обсуждения тематики данной работы.

СПИСОК ЛИТЕРАТУРЫ

1. *Brent R. P.* Algorithms for matrix multiplication. (Report No. STAN-CS-70-157). Stanford Univ. CA Dept. of Computer Science, 1970, 58 p.

2. *Strassen V.* Gaussian elimination is not optimal // Numer. Math. **13**, 354–356 (1969).

3. *Laderman J. D.* A noncommutative algorithm for multiplying 3x3 matrices using 23 multiplications. Bull. Amer. Math. Soc. **82**, 126–128 (1976).

4. *Pan V.* How we can speed up matrix multiplication? // SIAM Review, **26**(3), 393–415 (1984).

5. *Smirnov A. V.* The bilinear complexity and practical algorithms for matrix multiplication // Comp. Math. Mathem. Phys. **53**, 1781–1785 (2013).

6. *Karstadt E., Schwartz O.* Matrix multiplication, a little faster // J. of the ACM (JACM). 2020, **67**(1), 1–31.

7. *Kaporin I.* A Derivative-Free Nonlinear Least Squares Solver. In: Olenev N.N., Evtushenko Y.G., Jacimovic M., Khachay M., Malkova V. (eds.) Optimization and Applications. OPTIMA 2021. Lecture Notes in Computer Science, V. 13078. P. 217–230. Springer, Cham. (2021). https://doi.org/10.1007/978-3-030-91059-4_16

8. *Oseledets I. V., Savostyanov D. V.* Minimization methods for approximating tensors and their comparison // Computational Mathematics and Mathematical Physics, 2006, **46**(10), 1641–1650

9. *Kaporin I. E., Axelsson O.* On a class of nonlinear equation solvers based on the residual norm reduction over a sequence of affine subspaces // SIAM J. Sci. Comput. **16**(1), 228–249 (1994)

10. *Kaporin I.* Preconditioned Subspace Descent Method for Nonlinear Systems of Equations // Open Computer Science, 2020, **10**(1), 71–81

11. *Kozak D., Molinari C., Rosasco L., Tenorio L., Villa S.* Zeroth order optimization with orthogonal random directions. // Mathematical Programming. 2022. V. 199. P. 1–41.

12. *Armijo L.* Minimization of functions having Lipschitz continuous first partial derivatives // Pacific Journal of mathematics **16**(1), 1–3 (1966).

13. *Ballard G., Ikenmeyer C., Landsberg J. M., Ryder N.* The geometry of rank decompositions of matrix multiplication II: 3x3 matrices. // Journal of Pure and Applied Algebra, 2019, **223**(8), 3205–3224.

14. *Berger G. O., Absil P. A., De Lathauwer L., Jungers R. M., Van Barel M.* Equivalent polyadic decompositions of matrix multiplication tensors // J. of Computational and Applied Mathematics, 2022, 406, 113941. https://doi.org/10.1016/j.cam.2021.113941.

15. *Hopcroft J. E., Kerr L. R.* On minimizing the number of multiplications necessary for matrix multiplication // SIAM Journal on Applied Mathematics, 1971, **20**(1):30–36.

16. *Fawzi A. et al.* Discovering faster matrix multiplication algorithms with reinforcement learning. Nature, 2022, **610**(7930): 47–53.

FINDING COMPLEX-VALUED SOLUTIONS TO THE BRENT EQUATIONS BY REDUCING THEM TO A NONLINEAR LEAST SQUARES PROBLEM

I. E. Kaporin*

119333 Moscow, Vavilov Str., 44, Federal Research Center Computer Science and Control of the Russian Academy of Sciences, Russia
**e-mail: igorkaporin@mail.ru*

Received: 12.03.2024
Revised: 12.03.2024
Accepted: 05.06.2024

Abstract. Finding nontrivial solutions to the trilinear Brent equations corresponds to the construction of asymptotically fast matrix multiplication algorithms is an important, but in general a very difficult computational task. Methods of parameterization of the Brent equations based on the use of symmetries of the matrix product tensor are proposed, which make it possible to repeatedly reduce the dimension of the problem. The numerical solution of the obtained trilinear or cubic systems of nonlinear equations is carried out by reducing to a nonlinear least squares problem and applying to it a specially developed iterative method that does not require calculation of derivatives. The found solutions of the parameterized Brent equations, as a rule, have a rank no higher (and sometimes even lower) than the known results. Thus, an algorithm for multiplying two 4th-order matrices in 48 active multiplications is obtained.

Keywords: Brent equations; fast matrix multiplication; Strassen algorithm; nonlinear least squares problem.