ДОКЛАДЫ РОССИЙСКОЙ АКАДЕМИИ НАУК. МАТЕМАТИКА, ИНФОРМАТИКА, ПРОЦЕССЫ УПРАВЛЕНИЯ, 2024, том 518, с. 29–34

МАТЕМАТИКА

УДК 519.615.5

ЧИСЛЕННО-АНАЛИТИЧЕСКОЕ РЕШЕНИЕ УРАВНЕНИЙ БРЕНТА

© 2024 г. И. Е. Капорин[1], *

Представлено академиком РАН Е.Е. Тыртышниковым
Поступило 13.03.2024 г.
После доработки 20.05.2024 г.
Принято к публикации 16.07.2024 г.

Предлагается параметризация канонических разложений тензоров матричного произведения с многократно меньшим (по сравнению со стандартными уравнениями Брента) числом переменных. Последние определяются численно с использованием итерационного метода решения задачи нелинейных наименьших квадратов. Получены более быстрые по сравнению с известными алгоритмы перемножения двух 4 × 4-матриц за 48 умножений и 2 × 4-матрицы на 4 × 5-матрицу за 32 умножения.

Ключевые слова: быстрое умножение матриц, алгоритм Штрассена, уравнения Брента

DOI: 10.31857/S2686954324040056, **EDN:** YZJHHB

1. ВВЕДЕНИЕ И ФОРМУЛИРОВКА РЕЗУЛЬТАТОВ

Трилинейные уравнения, введенные Р. Брентом [1] в связи с задачей построения быстрых алгоритмов перемножения матриц, представляют собой каноническое разложение ранга r трехмерного $n_3n_1 \times n_1n_2 \times n_2n_3$-тензора специального вида,

$$\delta(i_2 - j_1)\delta(j_2 - k_1)\delta(k_2 - i_1) = \\ = \sum_{l=1}^{r} (X_l)_{i_2,i_1} (Y_l)_{j_2,j_1} (Z_l)_{k_2,k_1} \quad (1)$$

где

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

Здесь в качестве неизвестных фигурируют элементы матриц $X_l \in \mathbf{C}^{n_1 \times n_3}$, $Y_l \in \mathbf{C}^{n_2 \times n_1}$ и $Z_l \in \mathbf{C}^{n_3 \times n_2}$, где $l = 1,\ldots,r$, и через $\delta(\cdot)$ обозначена дискретная дельта-функция, т.е., $\delta(0) = 1$ и $\delta(t) = 0$ при $t \ne 0$. Точное решение уравнений Брента непосредственно приводит, в частности, к методу умножения $n_1 \times n_2$-матрицы на $n_2 \times n_3$-матрицу с использованием r некоммутативных умножений или, сокращенно, к $(n_1, n_2, n_3; r)$-алгоритму. Разбиение матриц на блоки и рекурсия позволяют построить $O(N^{\omega})$-алгоритм перемножения двух $N \times N$-матриц, где

$$\omega = (3\log r) \,/\, \log(n_1n_2n_3). \quad (2)$$

Алгоритмы быстрого матричного умножения, возможно более эффективные, чем (2,2,2;7)—алгоритм Штрассена [2], могут быть найдены при помощи численного решения уравнений Брента для минимально возможных значений r. Одно из последних достижений в этом направлении представляет собой (3,3,6;40)-алгоритм Смирнова [3] с элементами матриц X_l, Y_l, Z_l, принимающими значения 0, 1/2, −1/2.

Важным с практической точки зрения является применение эквивалентных преобразований полученных матриц X_l, Y_l, Z_l с целью повышения их разреженности, что позволяет понизить число операций сложения и умножения на константы. При этом показатель степени (2) остается неизменным, но может быть существенно понижен коэффициент при главном члене в оценке $O(N^{\omega})$. Здесь можно отметить технологию смены базиса, описанную в [4]. Аналогично, рассматриваемые ниже схемы матричного умножения с комплексными коэффициентами могут быть более эффективны при перемножении комплекснозначных матриц по сравнению с их применением к вещественным матрицам. Более того, выявлены два комплекснозначных решения с рангом на единицу меньше, чем для известных вещественных решений, см. ниже табл. 1 и 2. С другой стороны, найденные значения компонент матриц X_l, Y_l, Z_l, как правило,

[1] *Федеральный исследовательский центр “Информатика и управление” Российской академии наук, Москва, Россия*
**E-mail: igorkaporin@mail.ru*

заметно меньше единицы по модулю, что улучшает оценки численной устойчивости получаемых рекурсивных алгоритмов в арифметике с плавающей запятой.

Трилинейная система (1) включает $(n_1 n_2 n_3)^2$ уравнений и $(n_3 n_1 + n_1 n_2 + n_2 n_3) r$ неизвестных. Решение задачи (1) либо не существует (при слишком малых r), либо неединственно, причем изолированные решения отсутствуют (т.е. каждая ветвь решений представляет собой непрерывное многообразие). Кроме того, существуют “приближенные решения”, 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 элементов разложения (1). Способы эффективного сведения уравнений Брента к легкорешаемым оптимизационным задачам в общем случае неизвестны. Здесь можно упомянуть метод попеременной квадратичной минимизации, на основе которого получен результат [3], однако там были использованы дополнительные эвристики. Суммируя результаты ряда публикаций за последние более чем 50 лет, можно утверждать, что, по крайней мере для $\min(n_1, n_2, n_3) \geq 3$, отыскание частных решений задачи (1) является исключительно трудной вычислительной проблемой.

Введение дополнительных естественных ограничений на неизвестные за счет использования симметрий, присущих левой части уравнения (1), может облегчить отыскание решений уравнений Брента, в частности, за счет сокращения размерностей задачи. Первый шаг в этом направлении был сделан в работе [5], где в случае $n_1 = n_2 = n_3$ к уравнениям (1) были добавлены равенства

$$Y_l = X_{\sigma(l)}, \qquad Z_l = X_{\sigma^2(l)}, \qquad 1 \leq l \leq r. \tag{3}$$

Здесь $\sigma(\cdot)$ представляет собой перестановку чисел $(1, 2, \ldots, r)$, третья степень которой равна тождественной перестановке. Таким образом, число неизвестных сокращается в 3 раза. Формально, при этом можно опасаться потери некоторых решений. Однако вычисления показывают (см. ниже табл. 2), что зачастую удается не только получать решения, достигающие известных значений ранга r, но иногда и улучшать существующие оценки ранга. В настоящей работе предлагаются новые способы параметризации уравнений Брента, позволяющие существенно уменьшить размерность задачи. Здесь и далее для остатка от деления целого k на натуральное p используется обозначение $k \bmod p$.

Теорема 1. *Пусть p и q — натуральные числа, при которых существует решение системы уравнений*

$$\begin{aligned} &\frac{1}{p}\delta(i_2 - j_1)\delta(j_2 - k_1)\delta(k_2 - i_1) = \\ &= \sum_{t=1}^{q} (\widehat{X}_t)_{i_2, i_1} (\widehat{Y}_t)_{j_2, j_1} (\widehat{Z}_t)_{k_2, k_1}, \end{aligned} \tag{4}$$

$$1 \leq i_2, j_1 \leq n_1, \quad 1 \leq j_2, k_1 \leq n_2, \quad 1 \leq k_2, i_1 \leq n_3, \tag{5}$$

$$(i_2 - i_1 + j_2 - j_1 + k_2 - k_1) \bmod p = 0. \tag{6}$$

Тогда решению этих уравнений отвечает $(n_1, n_2, n_3; pq)$*—алгоритм перемножения матриц, а решение соответствующих уравнений Брента* (1) *имеет вид*

$$\begin{aligned} (X_{t+sq})_{i_2, i_1} &= (\widehat{X}_t)_{i_2, i_1} \Omega_p^{(i_1 - i_2)s}, \\ (Y_{t+sq})_{j_2, j_1} &= (\widehat{Y}_t)_{j_2, j_1} \Omega_p^{(j_1 - j_2)s}, \\ (Z_{t+sq})_{k_2, k_1} &= (\widehat{Z}_t)_{k_2, k_1} \Omega_p^{(k_1 - k_2)s}, \end{aligned} \tag{7}$$

$$\begin{aligned} r = pq, \quad 0 \leq s \leq p-1, \quad 1 \leq t \leq q, \\ \Omega_p = \exp\left(\frac{2\pi i}{p}\right), \quad i = \sqrt{-1}. \end{aligned} \tag{8}$$

Замечание 1. *Задача* (4)–(6) *представляет собой* **частичное** *каноническое тензорное разложение в силу наличия условия* (6).

Подставляя формулы (7) в (1), получаем искомую параметризацию уравнений Брента:

$$\delta(i_2 - j_1)\delta(j_2 - k_1)\delta(k_2 - i_1) = \sum_{t=1}^{q}\sum_{s=0}^{p-1} (\mathrm{K}^s \widehat{X}_t \mathrm{M}^{-s})_{i_2, i_1} (\Lambda^s \widehat{Y}_t \mathrm{K}^{-s})_{j_2, j_1} (\mathrm{M}^s \widehat{Z}_t \Lambda^{-s})_{k_2, k_1}, \tag{9}$$

$$1 \leq i_2, j_1 \leq n_1, \quad 1 \leq j_2, k_1 \leq n_2, \quad 1 \leq k_2, i_1 \leq n_3, \tag{10}$$

где диагональные матрицы $\mathrm{K} \in C^{n_1 \times n_1}$, $\Lambda \in C^{n_2 \times n_2}$, $\mathrm{M} \in C^{n_3 \times n_3}$ определены как

$$\begin{aligned} \mathrm{K} &= \mathrm{D}iag_{1 \leq m \leq n_1}(\Omega_p^{m-1}), \\ \Lambda = \mathrm{D}iag_{1 \leq m \leq n_2}(\Omega_p^{m-1}), &\quad \mathrm{M} = \mathrm{D}iag_{1 \leq m \leq n_3}(\Omega_p^{m-1}). \end{aligned} \tag{11}$$

Заметим, что при этом выполнены равенства

$$\mathrm{K}^p = I_{n_1}, \qquad \Lambda^p = I_{n_2}, \qquad \mathrm{M}^p = I_{n_3}, \tag{12}$$

На первый взгляд, результат теоремы 1 обладает недостаточной общностью из-за требования, чтобы ранг r был составным числом. Так, многие известные решения уравнений Брента определены для значений r, равных простому числу, например $r = 7$ или $r = 23$ (см. [2] и [6], соответственно). Однако здесь можно применить следующее обобщение теоремы 1, связанное с требованием наличия специального шаблона разреженности для матриц $\widehat{X}_t, \widehat{Y}_t, \widehat{Z}_t$.

Теорема 2. *Пусть числа* d_t *образуют неубывающую последовательность*

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

и каждое из них является делителем p*, возможно включая* 1 *и* p*. Тогда при присоединении условий*

$$(\widehat{X}_t)_{i_2,i_1} = 0, \quad (i_2 - i_1) \mathrm{mod} \frac{p}{d_t} \ne 0, \quad t = 1,\ldots,q, \tag{13}$$

$$(\widehat{Y}_t)_{j_2,j_1} = 0, \quad (j_2 - j_1) \mathrm{mod} \frac{p}{d_t} \ne 0, \quad t = 1,\ldots,q, \tag{14}$$

$$(\widehat{Z}_t)_{k_2,k_1} = 0, \quad (k_2 - k_1) \mathrm{mod} \frac{p}{d_t} \ne 0, \quad t = 1,\ldots,q, \tag{15}$$

к условиям теоремы 1 ранг получаемого разложения сокращается до значения

$$r = r(p, D) = \sum_{t=1}^{q} d_t, \tag{16}$$

а соответствующее параметризованное разложение (9) *принимает вид*

$$\delta(i_2 - j_1)\delta(j_2 - k_1)\delta(k_2 - i_1) = \sum_{t=1}^{q} \frac{p}{d_t} \sum_{s=0}^{d_t - 1} (\mathrm{K}^s \widehat{X}_t \mathrm{M}^{-s})_{i_2,i_1} (\Lambda^s \widehat{Y}_t \mathrm{K}^{-s})_{j_2,j_1} (\mathrm{M}^s \widehat{Z}_t \Lambda^{-s})_{k_2,k_1}, \tag{17}$$

$$1 \le i_2, j_1 \le n_1, \quad 1 \le j_2, k_1 \le n_2, \quad 1 \le k_2, i_1 \le n_3. \tag{18}$$

Замечание 2. *В частном случае* $n_1 = n_2 = n_3$ *возникает возможность дополнительно сократить число неизвестных до трех раз, если использовать соотношения* (3) *с заменой l на t и r на q для преобразования суммы по* t *в* (9).

2. ДОКАЗАТЕЛЬСТВО РЕЗУЛЬТАТОВ

В этом разделе докажем теоремы 1 и 2.

Доказательство теоремы 1. Домножим левую и правую части равенства (4) соответственно на левую и правую части тождества

$$p\delta((i_2 - j_1 + j_2 - k_1 + k_2 - i_1) \mathrm{mod}\, p) = \sum_{s=0}^{p-1} \Omega_p^{(i_2 - i_1)s + (j_2 - j_1)s + (k_2 - k_1)s}.$$

При этом правая часть (4) с учетом (7) и (11) принимает требуемый вид правой части (9). Если условие (6) не выполнено, то $\delta((i_2 - j_1 + j_2 - k_1 + k_2 - i_1) \mathrm{mod}\, p) = 0$ и справедливость полученного соотношения очевидна. Если же (6) выполнено, то $\delta((i_2 - j_1 + j_2 - k_1 + k_2 - i_1) \mathrm{mod}\, p) = 1$, так что левая часть (4) остается неизменной с точностью до множителя p. Теорема доказана. □

Доказательство теоремы 2. Ниже для упрощения формул опустим индекс t. Требуемое утверждение легко следует, если установить справедливость равенств

$$(\mathrm{K}^s \widehat{X}_t \mathrm{M}^{-s})_{i_2,i_1} = (\widehat{X})_{i_2,i_1} \Omega_p^{(i_2 - i_1)s} = (\widehat{X})_{i_2,i_1} \Omega_p^{(i_2 - i_1)(s \bmod d)}$$

и аналогично для $\widehat{Y}$ и $\widehat{Z}$. В силу (8) и (13), для этого достаточно убедиться в том, что при

$$(i_2 - i_1) \bmod \frac{p}{d} = 0$$

нацело делится на p разность

$$(i_2 - i_1)s - (i_2 - i_1)(s \bmod d) = (i_2 - i_1)(s - s \bmod d) = (\frac{p}{d} b)(cd) = bcp,$$

где b и c — некоторые целые числа. Теорема доказана. □

3. ЧИСЛЕННЫЕ РАСЧЕТЫ

Для небольших размеров матриц (см. табл. 1 и 2) при помощи алгоритма решения задачи наименьших квадратов с нулевой невязкой $f(x) = 0$, описанного в [7], были получены числовые значения компонент вектора x, образованного элементами матриц $\widehat{X}_t$, $\widehat{Y}_t$ и $\widehat{Z}_t$. Уравнения, определяющие невязку $f(x)$, формировались исходя из (4)–(6), при $D \ne \varnothing$ дополненных уравнениями (13)–(15). При этом от решения требовалась строгая ограниченность согласно $\| x \|_\infty < 1$ и близкая к минимуму (при использовании двойной точности) норма невязки уравнений Брента, а именно, $\| f \|_2 < 10^{-15}$. Выполнение этих требований дает основания предполагать, что известные эквивалентные преобразования решения уравнений Брента [10] позволят получить явные выражения элементов матриц X_l, Y_l и Z_l через алгебраические числа. По крайней мере, для (4,4,4;48)-алгоритма с $p = 4$, представленного в [8], таковыми являются 144 компоненты из 192.

Таблица 1. Сравнение известных верхних границ ранга r для прямоугольных матриц с оценкой $r(p, D)$, где D — множество использованных делителей p, меньших p.

n_1	n_2	n_3	r	ω	ссылка	p	q	D	$r(p, D)$	$\omega(p, D)$	$\| x \|_\infty$	$\| f \|_2$
2	2	2	7	2.807...	[2]	2	4	1	7	2.807...	0.97	1.3E-16
						3	3	1	7	2.807...	0.90	1.2E-16
2	2	3	11	2.894...	[9]	2	6	1	11	2.894...	0.81	2.6E-16
2	2	4	14	2.855...	[9]	2	7	∅	14	2.855...	0.79	2.6E-16
						4	4	2	14	2.855...	0.64	1.8E-16
2	2	5	18	2.894...	[9]	2	9	∅	18	2.894...	0.79	3.4E-16
						3	6	∅	18	2.894...	0.79	2.0E-16
2	2	6	21	2.873...	[9]	2	11	1	21	2.873...	0.91	4.7E-16
						3	7	∅	21	2.873...	0.89	4.4E-16
						6	4	3	21	2.873...	0.86	1.1E-16
2	3	3	15	2.810...	[9]	2	8	1	15	2.810...	0.83	3.4E-16
						3	5	∅	15	2.810...	0.79	2.8E-16
2	3	4	20	2.827...	[9]	2	10	∅	20	2.827...	0.80	5.1E-16
						4	5	∅	20	2.827...	0.84	3.1E-16
2	3	5	25	2.839...	[9]	5	5	∅	25	2.839...	0.90	6.8E-16
2	3	6	30	2.847...	[9]	2	15	∅	30	2.847...	0.93	5.8E-16
						3	10	∅	30	2.847...	0.94	4.8E-16
						6	5	∅	30	2.847...	0.77	1.9E-16
2	4	4	26	2.820...	[9]	2	13	∅	26	2.820...	0.86	9.0E-16
						4	7	2	26	2.820...	0.91	2.2E-16
2	4	5	33	2.843...	[9]	4	8	∅	**32**	**2.818...**	0.92	7.1E-16
						4	9	1	33	2.843...	0.79	4.9E-16
2	4	6	39	2.839...	[9]	6	7	3	39	2.839...	0.79	6.3E-16
2	4	8	52	2.850...	[9]	8	7	4	52	2.850...	0.95	1.4E-15
2	5	5	40	2.828...	[9]	5	8	∅	40	2.828...	0.82	3.7E-16
2	5	6	48	2.836...	[9]	6	8	∅	48	2.836...	0.79	1.1E-15
2	5	7	56	2.842...	[9]	7	8	∅	56	2.842...	0.80	6.7E-16
2	6	6	57	2.836...	[9]	6	10	3	57	2.836...	0.87	4.7E-16
3	3	3	23	2.854...	[6]	2	12	1	23	2.854...	0.90	5.2E-16
						4	7	1;2	23	2.854...	0.80	3.5E-16
						2	12	∅	24	2.892...	0.92	4.2E-16
						3	8	∅	24	2.892...	0.92	2.5E-16
						4	6	∅	24	2.892...	0.84	2.1E-16
3	3	4	**29**	**2.818...**	[3]	3	10	∅	30	2.847...	0.86	8.7E-16
						4	8	2	30	2.847...	0.95	2.7E-16
3	3	5	36	2.824...	[3]	3	12	∅	36	2.824...	0.89	1.9E-15
3	3	6	**40**	**2.774...**	[3]	3	14	∅	42	2.810...	0.97	6.9E-16
3	4	4	**38**	**2.818...**	[3]	3	13	∅	39	2.839...	0.87	9.6E-16
3	4	5	**47**	**2.821...**	[11]	4	12	∅	48	2.836...	0.99	2.1E-15

Таблица 2. Сравнение известных верхних границ ранга r для квадратных матриц с $r(p, D)$, (где D — множество делителей p, меньших p) при использовании (3); выражения $q = q' + 3q''$ характеризуют перестановку $\sigma(\cdot)$, отвечающую блочно-диагональной матрице перестановок с ведущим блоком $I_{q'}$ и остальными q'' блоками в виде циклических перестановок порядка 3.

n_1	n_2	n_3	r	ω	ссылка	p	q	D	$r(p, D)$	$\omega(p, D)$	$\| x \|_\infty$	$\| f \|_2$
2	2	2	7	2.807...	[2]	2	$4 = 1 + 3 \cdot 1$	1	7	2.807...	0.81	6.6E-17
3	3	3	23	2.854...	[6]	4	$7 = 4 + 3 \cdot 1$	1;2	23	2.854...	0.79	1.6E-14
4	4	4	49	2.807...	[2]	2	$24 = 6 + 3 \cdot 6$	∅	**48**	**2.792...**	0.77	1.6E-15
						4	$12 = 3 + 3 \cdot 3$	∅	**48**	**2.792...**	0.75	8.1E-16

Результаты численного решения сведены в таблицы 1 и 2. В случае несовпадения рангов меньшее значение выделено жирным шрифтом. Из таблицы 1 видно улучшение до $r = 32$, достигнутое для задачи перемножения прямоугольных матриц с $n_1 = 2$, $n_2 = 4$ и $n_3 = 5$, а таблица 2 представляет два алгоритма для перемножения двух квадратных матриц 4-го порядка за 48 активных умножений. Отсутствие этих результатов в известных источниках (где рассматривались вещественные числа) позволяет предположить, что ранг тензора матричного умножения может понижаться при переходе к комплексным числам. Отметим, что в [12] утверждается существование $(4,4,4;r)$-алгоритма с $r < 49$ над полем комплексных чисел.

4. ЗАКЛЮЧЕНИЕ

Рассмотренные выше представления решений уравнений Брента позволяют, с одной стороны, облегчить численное решение последних за счет кратного сокращения числа уравнений и неизвестных, и, с другой стороны, не слишком сильно сузить многообразие решений, чтобы сдержать возможное завышение оценок ранга. Результат численного решения полученных трилинейных или кубических систем нелинейных уравнений считался допустимым, если не только невязка уравнений Брента была близка к пределу машинной точности, но и максимум модуля компонент решения был меньше единицы. Это позволило надежно отсеять так называемые “приближенные алгоритмы”, зависящие от малого параметра. Последние гораздо менее пригодны для вычислений в арифметике с плавающей точкой ввиду неограниченного роста величин некоторых компонент решения уравнения Брента при устремлении нормы невязки к нулю, см. [13] и цитированные там источники. Можно предполагать, что известные эквивалентные преобразования решения уравнений Брента позволят получить явные выражения коэффициентов через константы, являющиеся алгебраическими числами.

Найденные решения параметризованных уравнений Брента для большинства рассчитанных вариантов имеют ранг не больший (а иногда и меньший) по сравнению с известными результатами. В частности, получен алгоритм перемножения двух матриц 4-го порядка за 48 активных умножений, а также (2,4,5;32)-алгоритм для прямоугольных матриц.

БЛАГОДАРНОСТИ

Автор выражает благодарность Е.Е. Тыртышникову за полезные обсуждения данной работы, а также анонимным рецензентам за ценные замечания, улучшившие ее изложение.

СПИСОК ЛИТЕРАТУРЫ

1. *Brent R.P.* Algorithms for matrix multiplication. (Report No. STAN-CS-70-157). Stanford Univ. CA Dept. of Computer Science, 1970, 58p.
2. *Strassen V.* Gaussian elimination is not optimal // Numer. Math. 1969. V. 13. № 4. P. 354–356.
3. *Смирнов А.В.* О билинейной сложности и практических алгоритмах умножения матриц // ЖВМ и МФ. 2013. Т. 53. № 12. С. 1970–1984.
4. *Beniamini G. et al.*, Sparsifying the operators of fast matrix multiplication algorithms. arXiv preprint arXiv:2008.03759 (2020) https://arxiv.org/pdf/2008.03759.pdf
5. *Ballard G., Ikenmeyer C., Landsberg J.M., Ryder N.* The geometry of rank decompositions of matrix multiplication II: 3×3 matrices // J. of Pure and Applied Algebra. 2019. V. 223. № 8. P. 3205–3224.
6. *Laderman J.D.* A noncommutative algorithm for multiplying 3x3 matrices using 23 multiplications // Bull. Amer. Math. Soc. 1976. V. 82. № 1. P. 126–128.
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. *Kaporin I.* Verifying the correctness of the (4,4,4;48) matrix multiplication scheme with complex coefficients exact up to the floating point tolerance // 2024. URL: https://cloud.mail.ru/public/Yfij/ErDxopqBh
9. *Hopcroft J.E., Kerr L.R.* On minimizing the number of multiplications necessary for matrix multiplication // SIAM Journal on Appl. Math. 1971. V. 20. № 1 P. 30–36.
10. *Berger G.O., Absil P.A., De Lathauwer L., Jungers R.M., Van Barel M.* Equivalent polyadic decompositions of matrix multiplication tensors // J. of Comput. and Appl. Math. 2022. V. 406. P. 113941. https://doi.org/10.1016/j.cam.2021.113941
11. *Fawzi A. et al.* Discovering faster matrix multiplication algorithms with reinforcement learning // Nature. 2022. V. 610. № 7930. P. 47–53.
12. *Li X., Zhang L., Ke Y.* Deflation conjecture and local dimensions of Brent equations // arXiv preprint arXiv:2310.11686. 2023 Oct 18.
13. *Ballard G., Weissenberger J., Zhang L.* Accelerating neural network training using arbitrary precision approximating matrix multiplication algorithms // 50th International Conference on Parallel Processing Workshop 2021 Aug 9. P. 1–8. https://doi.org/10.1145/3458744.3474050

SEMI-ANALYTICAL SOLUTION OF BRENT EQUATIONS

I. E. Kaporin[a]

Presented by Academician of the RAS E. E. Tyrtyshnikov

[a]*FRC CSC RAS, Moscow, Russia*

A parametrization of Brent equations is proposed which admit for a several times reduction of the number of unknowns and equations. The arising equations are solved numerically, and for the resulting fast matrix multiplication algorithms many known values of rank are reproduced and even improved, in particular, the designs $(4,4,4;48)$ and $(2,4,5;32)$ are found.

Keywords: fast matrix multiplication, Brent equations, Strassen algorithm