Московский государственный университет имени М.В. Ломоносова
Факультет вычислительной математики и кибернетики
Кафедра вычислительных технологий и моделирования

Адимов Арсений Владимирович

Метод оптимизации приближенных тензорных разложений

ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА

Научный руководитель:
д.ф.-м.н., профессор, чл.-корр. РАН, зав. кафедрой ВТМ
Е.Е. Тыртышников

Москва, 2016

Содержание

Аннотация

В данной работе предложен новый метод для задач канонического разложения и приближения 3-тензоров.

1 Введение

Многомерные массивы возникают во многих приложениях. Среди различных применений можно упомянуть анализ данных и вычислительную сложность алгоритмов (см. секцию 2.1). В большинстве приложений тензоры обладают какой-то скрытой структурой, для решения задач важно выявить эту структуру и использовать ее для малопараметрического представления данных. Часто представление тензора в виде суммы тензоров ранга 1 позволяет решить эту задачу. Тензоры могут не точно соответсвовать описываемой модели, в таком случае возникает задача приближения тензора, вместо нахождения точного разложения. Также приближать тензоры необходимо после различных операций с ними.

Целью данной работы является построение алгоритма для разложения и приближения 3-тензоров с вещественными элементами. Известно, что ранг тензора зависит от поля, над которым он рассматривается. В данной работе рассматриваются 3-тензоры над полем рациональных функций. К заданному тензору добавляется полиномиальное возмущение, и для полученного тензора исследуются методы минимизации общего вида для непосредственного вычисления элементов канонических факторов в виде полиномов Лорана ограниченной степени такие как метод наименьших квадратов с переменными направлениями и метод Ньютона (см. секцию 3).

1.1 Определения и обозначения

Тензор $A = (a_{ijk}) \in \mathbb{R}^{N_1 \times N_2 \times N_3}$ может быть представлен в виде канонического разложения:

$$A = \sum_{t=1}^{T} F_t, \quad F_t = \alpha^t \otimes \beta^t \otimes \gamma^t = (\alpha_i^t \beta_j^t \gamma_k^t)$$

Минимальное число слагаемых в таком представлении называется каноническим тензорным рангом или просто тензорным рангом тензора A, обозначается trank(A) или r(A).

Минимальное число t, такое что выполнено:

$$\forall \varepsilon > 0 \; \exists E = (e_{ijk}) : |e_{ijk}| < \varepsilon : \; r(A+E) = t < r(A) \tag{1}$$

называется граничным рангом тензора A и обозначается br(A) [2].

Допустим, что разложение (1) тензора A, может быть полученно с поправкой E, элементы которой являются многочленами переменной ε. Другими словами, мы предполагаем, что существует $P(\varepsilon) = (p_{ijk}(\varepsilon)$, $p_{ijk}(\varepsilon)$ – многочлены, такие что $p_{ijk}(0) = 0$ и

$$A + P(\varepsilon) = \sum_{t=1}^{T_0} F_t(\varepsilon), \; F_t(\varepsilon) = (\alpha_i^t(\varepsilon)\beta_j^t(\varepsilon)\gamma_k^t(\varepsilon)) \tag{2}$$

является каноническим разложением тензора $A + P(\varepsilon)$ ранга T_0, в котором коэффициенты $\alpha_i^t(\varepsilon), \beta_j^t(\varepsilon), \gamma_k^t(\varepsilon)$ принадлежат полю рациональных функций над $\mathbb{R}$.

Если $br_0(A)$ минимальное число слагаемых в представлении (2), то $br(A) \leqslant br_0(A)$

Так как $\lim\limits_{\varepsilon \to 0}(A + P(\varepsilon)) = A$, то разложение (2) является приближенным каноническим разложением тензора A.

Будем искать разложение (2) с коэффициентами в виде полиномов Лорана ограниченной степени.

1.2 Постановка задачи

Дан 3-тензор $\sigma(i,j,k) \in \mathbb{R}^{N_1 \times N_2 \times N_3}$ и натуральные числа T и D.

Требуется получить приближенное каноническое разложение с рангом T: $\sigma_{ijk} = \lim\limits_{\varepsilon \to 0} \sum\limits_{t=1}^{T} \alpha_i^t(\varepsilon)\beta_j^t(\varepsilon)\gamma_k^t(\varepsilon)$, где коэффиценты являются многочленами Лорана ограниченной степени D.

Зададим случайное полиномиальное возмущение $p_{ijk}(\varepsilon)$, в котором $p_{ijk}(0) = 0$ и степени полиномов не выше $d \leqslant 3 \cdot D$.

$$\sigma_{ijk} + p_{ijk}(\varepsilon) = \sum_{t=1}^{T} \alpha_i^t(\varepsilon)\beta_j^t(\varepsilon)\gamma_k^t(\varepsilon)$$

Приравнивая коэффициенты при одинаковых степенях, приходим к системе уравнений:

$$\sigma_{ijk}^s = \sum_{t=1}^{T} \sum_{s_i+s_j+s_k=s} \alpha_i^{t,s_i}\beta_j^{t,s_j}\gamma_k^{t,s_k} \tag{3}$$

$$1 \leqslant i \leqslant N_1,\ 1 \leqslant j \leqslant N_2,\ 1 \leqslant k \leqslant N_3, -D \leqslant s_i, s_j, s_k \leqslant D,\ -3 \cdot D \leqslant s \leqslant 3 \cdot D$$

Нужно исследовать данную систему и предложить метод ее решения.

Требуется для заданного числа ε найти тензор $L(i,j,k) \in \mathbb{R}^{N_1 \times N_2 \times N_3}$ ранга T, являющимся ε-аппроксимацией тензора σ: $\|\sigma - L\|_2^2 \leqslant \varepsilon$

2 Обзор литературы

2.1 Быстрое умножение матриц

Пусть даны матрицы $A(i_1, i_2) \in \mathbb{R}^{m \times n}$ и $B(j_1, j_2) \in \mathbb{R}^{n \times p}$. По определению, задача умножения матрицы A на матрицу B состоит в вычислении mp билинейных форм вида $\sum\limits_{k=1}^{n} a_{i_1 k} b_{k j_2}$. Если напрямую вычислять по этой формуле, то потребуется mnp умножений, если $m = n = p$, то будет n^3 умножений. Быстрыми называются алгоритмы, в которых требуется меньше операций умножения, чем при обычном алгоритме. В 1969 году Штрассен [14] предложил алгоритм, в котором требуемое число операций $O(n^{\log_2 7})$. Усилиями многих авторов эта оценка была уменьшена примерно до $O(n^{2.37})$. Последняя оценка носит теоретический характер, так как она справедлива для матриц больших размеров. На практике для перемножения матриц ограниченного размера(например $n \leqslant 1000000$) применяются алгоритмы со сложностью $O(n^{2.77})$. Краткое резюме исследований задачи быстрого умножения матриц содержится в работе В. Пана [10].

Алгоритм Штрассена для перемножения матриц порядка n со сложностью $O(n^{\log_2 7})$ основан на билинейном алгоритме, который был найден им для перемножения матриц порядка 2, требующего 7 умножений вместо 8, как в обычном алгоритме[14]. Затем этот алгоритм был применен рекурсивно для перемножения матриц порядка 2^k с числом умножений 7^k.

Пусть $A = a_1, \ldots, a_r$ и $B = b_1, \ldots, b_s$, билинейными алгоритмами называются алгоритмы следующего вида. Сначала вычисляются произведения

$$q^t = \left(\sum_{i=1}^{r} \alpha_i^t a_i \right) \left(\sum_{j=1}^{s} \beta_j^t b_j \right) \text{ для } \; t = 1, \ldots, T. \tag{4}$$

Затем вычисляются линейные комбинации этих произведений:

$$C_k = \sum_{t=1}^{T} \gamma_k^t q^t. \tag{5}$$

Для анализа сложности задачи умножения матриц билинейность необходима, потому что при рекурсии числа заменяются матрицами, а они могут не коммутировать.

Подставляя (4) в (5), используя определение матричного умножения получим:

$$\sum_{t=1}^{T} \gamma^t_{k_1 k_2} \left(\sum_{i}^{r} \alpha^t_i a_i \right) \left(\sum_{j=1}^{s} \beta^t_j b_j \right) = \sum_{k=1}^{n} a_{k_1 k} b_{k k_2}$$

приравнивая коэффициенты, приходим к системе уравнений Брента [11]:

$$\sum_{t=1}^{T} \alpha^t_i \beta^t_j \gamma^t_k = \sigma_{ijk}, \quad \sigma_{ijk} = \delta(k_1, i_1) \cdot \delta(i_2, j_1) \cdot \delta(j_2, k_2), \tag{6}$$

где $\delta(i,j) = \begin{cases} 1, \ i = j \\ 0, \ i \neq j \end{cases}$ – символ Кронекера.

Минимальное число $T = R = R(m,n,p)$, при котором система (6) разрешима, называется рангом задачи умножения матриц или билинейной сложностью. Если для некоторого числа Т, получено решение системы уравнений (6), то Т называется длиной этого решения, соответствующий набор коэффициентов $\alpha^t_i, \beta^t_j, \gamma^t_k$ является билинейным алгоритмом для перемножения матриц, требующим Т умножений. Можно показать, что такой алгоритм дает асимптотическую оценку сложности умножения матриц $O(N^p)$, где $p = 3 \log_{mnp}(T)$.

2.2 Приближенные алгоритмы

Набор 3T матриц с элементами, являющимися полиномами Лорана

$\alpha^t_{i_1,i_2}, \quad i_1 = 1, 2, \ldots, n_1, \quad i_2 = 1, 2, \ldots, n_2, \quad t = 1, 2, \ldots, T$

$\beta^t_{j_1,j_2} \quad j_1 = 1, 2, \ldots, n_2, \quad j_2 = 1, 2, \ldots, n_3, \quad t = 1, 2, \ldots, T$

$\gamma^t_{k_1,k_2} \quad k_1 = 1, 2, \ldots, n_1, \quad k_2 = 1, 2, \ldots, n_3, \quad t = 1, 2, \ldots, T$

является приближенным билинейным алгоритом длины T для перемножения матриц $A(i_1, i_2) \in \mathbb{R}^{n_1 \times n_2}$ и $B(j_1, j_2) \in \mathbb{R}^{n_2 \times n_3}$ если:

$$\sum_{t=1}^{T} \gamma^t_{k_1 k_2} \left(\sum_{i_1,i_2=1}^{n_1,n_2} \alpha^t_{i_1 i_2} a_{i_1 i_2} \right) \left(\sum_{j_1,j_2=1}^{n_2,n_3} \beta^t_{j_1 j_2} b_{j_1 j_2} \right) = \sum_{k=1}^{n_2} a_{k_1 k} b_{k k_2} + O(x)$$

Соотношение между алгоритмами приближающими результат с произвольной точностью и точными алгоритмами было описано Д.Бини [2]. Если получено приближенное разложение (2) с рангом T и возмущением степени не выше d, то можно получить каноническое разложение ранга (1+d)T.

$$A + P(\varepsilon) = \sum_{t=1}^{T} F_t(\varepsilon)$$

Если взять d+1 различную точку $\varepsilon_1, \varepsilon_2, \ldots \varepsilon_{d+1}$, то система уравнений

$$\begin{bmatrix} 1 & 1 & \cdots & 1 \\ \varepsilon_1 & \varepsilon_2 & \cdots & \varepsilon_{d+1} \\ \vdots & \vdots & \ddots & \vdots \\ \varepsilon_1^d & \varepsilon_2^d & \cdots & \varepsilon_{d+1}^d \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_{d+1} \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

будет иметь единственное решение.

$$\sum_{i=1}^{d+1} \alpha_i A + \sum_{i=1}^{d+1} \alpha_i P(\varepsilon_i) = \sum_{i=1}^{d+1} \alpha_i \sum_{t=1}^{T} F_t(\varepsilon_i) \;\Rightarrow\; A = \sum_{i=1}^{d+1} \alpha_i \sum_{t=1}^{T} F_t(\varepsilon_i)$$

Так как $\otimes^k (A + P(\varepsilon)) = \otimes^k A + \widetilde{P}(\varepsilon) = \sum_{t=1}^{T^k} \widetilde{F}_t(\varepsilon)$, $\widetilde{F}_t(\varepsilon) = F_{t_1}(\varepsilon) \otimes F_{t_2}(\varepsilon) \otimes x \ldots \otimes F_{t_k}(\varepsilon)$ и $\widetilde{P}(\varepsilon)$ полиномиальное возмущение степени dk, такое что $\widetilde{e}_{ijk}(0) = 0$, то существует каноническое разложение тензора $\otimes^k A$ ранга $(1 + dk)(T)^k$

2.3 Обзор существующих решений

Ранг тензора может быть вычислен за конечное число шагов. Тем не менее эффетивных алгоритмов для его вычисления и вычисления канонического разложения пока не найдено.

Используются различные методы минимизации общего вида. Однако, задача оптимизации ошибки может оказаться некорректно поставленной [4]: можно построить такой тензор, и такую последовательность канонических факторов, что ошибка аппроксимации будет стремится к нулю, тогда как сами элементы разложения будут расходится к неопределенности $\infty - \infty$.

Одним из наиболее популярных методов минимизации для задачи малоранговой аппроксимации с фиксированным рангом является метод наименьших квадратов с переменными направлениями. В [6] этот метод применялся для метода главных компонент 3-тензора и в [1] для тензоров, представленных в виде канонического разложения. Также для задачи применялись методы Ньютона: метод Ньютона с регуляризацией [5] и метод Ньютона в доверительной области [7]. Сравнение различных методов приведено в [3, 9].

Недавно А. Смирнов в [13] предложил численный, частично эвристический метод для получения решений системы уравнений Брента(6), которые имеют рациональные компоненты и содержат большое число нулей. С помощью него он получил новые оценки граничных рангов тензора матричного умножения [12, 13]

3 Алгоритмы решения задачи

3.1 Метод наименьших квадратов

Введем целевую функцию

$$F = \left\| \sigma - \sum_{t=1}^{T} \alpha^t \otimes \beta^t \otimes \gamma^t \right\|_2^2 + \lambda(\|z - \widetilde{z}\|_2^2 \quad z = (\alpha, \beta, \gamma) \tag{7}$$

$$f_1 = \sum_{i,j,k,s} \left(\sum_{t=1}^{T} \sum_{s_i+s_j+s_k=s} \alpha_i^{t,s_i} \beta_j^{t,s_j} \gamma_k^{t,s_k} - \sigma_{ijk}^s \right)^2$$

$$f_2 = \lambda \sum_{t=1,s=-D}^{T,D} \left(\sum_{i=1}^{N_1} (\alpha_i^{t,s} - \widetilde{\alpha}_i^{t,s})^2 + \sum_{j=1}^{N_2} (\beta_j^{t,s} - \widetilde{\beta}_j^{t,s})^2 + \sum_{k=1}^{N_3} (\gamma_k^{t,s} - \widetilde{\gamma}_k^{t,s})^2 \right)$$

$F = f_1 + f_2$, $\widetilde{z} = (\widetilde{\alpha}, \widetilde{\beta}, \widetilde{\gamma})$ – фиксированные коэффициенты, удоавлетворяющие следующим ограничениям:

$$\begin{cases} \widetilde{\alpha}_i^{t,s} = \alpha_i^{t,s}, \text{ если } |\alpha_i^{t,s}| \leqslant B; \ \widetilde{\alpha}_i^{t,s} = -B, \text{ если } \alpha_i^{t,s} < -B; \ \widetilde{\alpha}_i^{t,s} = B, \text{ если } \alpha_i^{t,s} > B; \\ \widetilde{\beta}_j^{t,s} = \beta_j^{t,s}, \text{ если } |\beta_j^{t,s}| \leqslant B; \ \widetilde{\beta}_j^{t,s} = -B, \text{ если } \beta_j^{t,s} < -B; \ \widetilde{\beta}_j^{t,s} = B, \text{ если } \beta_j^{t,s} > B; \\ \widetilde{\gamma}_k^{t,s} = \gamma_k^{t,s}, \text{ если } |\gamma_k^{t,s}| \leqslant B; \ \widetilde{\gamma}_k^{t,s} = -B, \text{ если } \gamma_k^{t,s} < -B; \ \widetilde{\gamma}_k^{t,s} = B, \text{ если } \gamma_k^{t,s} > B; \end{cases} \tag{8}$$

Здесь $\alpha_i^{t,s}, \beta_j^{t,s}$ и $\gamma_k^{t,s}$ значения неизвестных, полученных на предыдущем шаге вычислений.

Легко заметить, что система (3) может быть разделена на 3 линейных подсистемы, каждая для своей группы неизвестных(α, β, γ). При текущих значений α_i^{t,s_i} и β_j^{t,s_j}, обозначая $x_{p,t}$ для $\gamma_k^{t,p}$ при фиксированном k, мы хотим найти решение в смысле метода наименьших квадратов системы $Ax = b$, где матрица A имеет вид:

$$A = \left(\sum_{s_i+s_j=s-p} \alpha_i^{t,s_i} \beta_j^{t,s_j} \right)_{(s,i,j),(p,t)} \in \mathbb{R}^{(6D+1)N_1N_2 \times (2D+1)\cdot T}, \ -3{\cdot}D \leqslant s \leqslant 3{\cdot}D, \ -D \leqslant p \leqslant D.$$

Решение выражается как $x = (A^T A + \lambda \mathrm{I})^{-1}(A^T b + \lambda \widetilde{\gamma}_k^{t,s})$

$$A^T A = \left(\sum_{s,i,j} \left(\sum_{s_i+s_j=s-p_i} \alpha_i^{t,s_i} \beta_j^{t,s_j} \right) \left(\sum_{s_i+s_j=s-p_j} \alpha_i^{u,s_i} \beta_j^{u,s_j} \right) \right)_{(p_i,t),(p_j,u)}$$

$$A^T b = \left(\sum_{s,i,j} \left(\sum_{s_i+s_j=s-p} \alpha_i^{t,s_i} \beta_j^{t,s_j} \right) \cdot \sigma_{ijk}^s \right)_{(p,t)}$$

Алгоритм:

Сгенерировать возмущение и начальное приближение. Взять маленькое λ и число B. На каждом шаге итерации решить $N_1 + N_2 + N_3$ систем уравнений для каждой группы коэффициентов α, β или γ, после решения кадой системы, установлия значения коэффициентов $\widetilde{\alpha}_i^{t,s}, \widetilde{\beta}_j^{t,s}$, или $\widetilde{\gamma}_k^{t,s}$ в соответсвии с (8). Постепенно уменьшая λ, продолжать минимизировать целевую функцию f, пока она не станет достаточно малой.

Метод описанный выше прост в реализации и вычислительная сложность каждого шага мала. Но число итераций для достижения требуемой точности может быть очень большим, скорость сходимости замедляется при малой целевой функции, также метод может не сходится, или сходится к локальному минимуму. Тем не менее различные эвристики и комбинации этого метода с другими могут разрешить эти проблемы.

3.2 Метод Ньютона

Для решения (3) рассмотрим такую же целевую функцию $F(7)$, обозначим вектор неизвестных $\overrightarrow{x} = (\overrightarrow{\alpha}, \overrightarrow{\beta}, \overrightarrow{\gamma})$

$$F = ||\overrightarrow{f}(\overrightarrow{x})|| = \sum_{v \in ijks} f_v(\overrightarrow{x})^2, \quad f_{ijk}^s = \sum_{t=1}^{T} \sum_{s_i+s_j+s_k=s} \alpha_i^{t,s_i} \beta_j^{t,s_j} \gamma_k^{t,s_k} - \sigma_{ijk}^s$$

Направление поиска $\overrightarrow{h_k}$ в методе Ньютона определяется из системы:

$$(H(\overrightarrow{x_k}) + \lambda_k \mathrm{I})\overrightarrow{h_k} = -\nabla F,$$

где $H(\overrightarrow{x})$ - матрица Гессе функции $F(\overrightarrow{x})$.

$$\overrightarrow{x}_{k+1} = \overrightarrow{x}_k + \overrightarrow{h_k}$$

В методе Левенберга-Марквардта параметр λ зависит от того, насколько хорошо функция F на текущем шаге может быть аппроксимирована со вторым порядком.

$$L(h) = F(x) + F'(x)h + \frac{1}{2} h^T F''(x)h$$

На каждом шаге вычисляется коэффициент

$$\rho = \frac{F(x) - F(x+h)}{L(0) - L(h)}$$

Если $\rho > 0$, то вычисляется новое приближение x_{k+1}, и параметры $\lambda_{k+1} = \lambda_k max(\frac{1}{3}, 1-(2p-1)^3)$ и $\nu = 2$, иначе x_k не изменяется, вычисляется $\lambda_{k+1} = \nu\lambda_k$ и $\nu = 2\nu$. Впервые стратегия выбора параметра λ была предложена Марквардтом. Используется метод, описанный в [8] и использованный в [3] и [9].

$$F_k^{t,s_k} = \frac{\partial F}{2 \cdot \partial \gamma_k^{t,s_k}} = \sum_{i,j,s} G_{ijk}^s \cdot W_{ij}^{t,s,s_k} + \lambda \gamma_k^{t,s_k}$$

$$G_{ijk}^s = \sum_{t=1}^{T} \sum_{s_i+s_j+s_k=s} \alpha_i^{t,s_i} \beta_j^{t,s_j} \gamma_k^{t,s_k} - \sigma_{ijk}^s, \quad W_{ij}^{t,s,s_k} = \sum_{s_i+s_j=s-s_k} \alpha_i^{t,s_i} \beta_j^{t,s_j}$$

$$\frac{\partial F_k^{t,s_k}}{\partial \gamma_{k'}^{p,s'_k}} = \begin{cases} 0, k \neq k' \\ \sum\limits_{s,i,j} W_{ij}^{p,s,s'_k} \cdot W_{ij}^{t,s,s_k} + \lambda \delta(p,t) \delta(s_k, s'_k) \end{cases}$$

$$\frac{\partial F_k^{t,s_k}}{\partial \alpha_i^{p,s_i}} = \sum_{s,j} (\beta_j^{p,s-s_i-s_k} \gamma_k^{p,s_k} \cdot W_{ij}^{t,s,s_k} + \delta(t,p) G_{ijk}^s \cdot \beta_j^{p,s-s_i-s_k})$$

Остальные производные считаются аналогично.

Алгоритм:

$k := 0,\ x := x_0,\ A := H(x),\ g := -\nabla F(x),\ done := false$

Пока (($k < k_{max}$) и (done $\neq$ true))

$k := k + 1$

$h = (A + \lambda \mathrm{I})^{-1} g$

Если h достаточно мало

$done := true$

Иначе

$x_{new} := x + h,\ p := \frac{F(x) - F(x_{new})}{L(0) - L(h_k)}$

Если ($p > 0$)

$x := x_{new},\ A := H(x),\ g := -\nabla F(x)$

Если g достаточно мало

done := true

$\lambda := \lambda \cdot max(\frac{1}{3},\ 1 - (2p-1)^3),\ \ \nu := 2$

Иначе

$\lambda := \nu\lambda,\ \ \nu := 2\nu$

4 Вычислительные эксперименты

4.1 Исходные данные и условия эксперимента

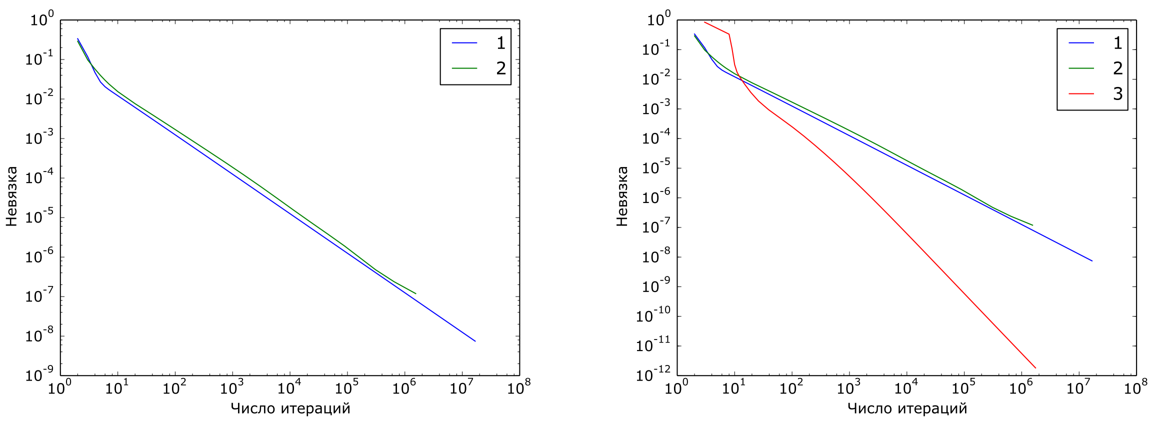
Пусть $x_1, y_1 \in \mathbb{R}_1^n, x_2, y_2 \in \mathbb{R}_2^n, x_3, y_3 \in \mathbb{R}_3^n$ пары линейно независимых векторов. Тогда тензор ранга 3: $A = x_1 \otimes x_2 \otimes y_3 + x_1 \otimes y_2 \otimes x_3 + y_1 \otimes x_2 \otimes x_3$ может быть приближен последовательностью тензоров ранга 2:

$$A_n = n(x_1 + \frac{1}{n}y_1) \otimes (x_2 + \frac{1}{n}y_2) \otimes (x_3 + \frac{1}{n}y_3) - nx_1 \otimes x_2 \otimes x_3, \quad \lim_{n\to\infty} A_n = A$$

Для первого эксперимента рассмотрим тензор $A = u \otimes u \otimes v + u \otimes v \otimes u + v \otimes u \otimes u$ u $= \begin{pmatrix} 1 \\ 0 \end{pmatrix}$, v $= \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ и будем искать приближение ранга 2.

4.2 Результаты экспериментов

Эксперимент 1:



а) б)

Рис. 1:

1 – обычный МНК,

2 – МНК для тензора с возмущением $|p^s_{ijk}| \leqslant 10^{-6}$,

3 – МНК для тензора с возмущнением $|p^s_{ijk}| \leqslant 1$.

4.3 Обсуждение и выводы

После того как численное решение найдено, погрешность решения может содержаться при отрицательных степенях коэффициентов. Таким образом мы не можем стремить $\varepsilon \to 0$, чтобы получить последовательность тензоров, стремящихся к исходному.

Применяя метод описанный в секции 2.2, мы можем получить каноническое разложение исходного тензора ранга (1+2D)T, для этого достаточно взять 2D+1 различных точек. Также для этого полиномиальное возмущение можно делать с отрицательными степенями.

Если получено численное решение L(i,j,k,s) системы (3) с полиномиальным возмущением P и с точностью ε_1:

$L^s_{ijk} = \sum\limits_{t=1}^{T} \sum\limits_{s_i+s_j+s_k=s} \alpha_i^{t,s_i} \beta_j^{t,s_j} \gamma_k^{t,s_k}, \quad \left\| L^s_{ijk} - (\sigma_{ijk} + P^s_{ijk}) \right\|_2^2 \leqslant \varepsilon_1$, то, рассматривая в некоторой точке полиномы Лорана $L_{ijk}(x) = \sum\limits_{s=-3D}^{3D} L^s_{ijk} x^s$, мы можем приблизить исходный тензор и оценить ошибку аппроксимации $\|L_{ijk}(x) - \sigma_{ijk}\|_2^2$:

$$\|L(x) - \sigma\|_2^2 \leqslant (\|L(x) - (\sigma + P(x))\|_2 + \|P(x)\|_2)^2 \leqslant 2 \cdot (\|L(x) - (\sigma + P(x))\|_2^2 + \|P(x)\|_2^2) \tag{9}$$

Рассматривая в точке x = 1:

$$\|L(1) - (\sigma + P(1))\|_2^2 = \sum_{ijk} \left(\sum_s (L^s_{ijk} - \sigma^s_{ijk}) \right)^2 \leqslant (6D+1) \sum_{ijks} (L^s_{ijk} - \sigma^s_{ijk})^2 = (6D+1)\varepsilon_1$$

Если ограничить коэффициенты возмущения $|P_{ijk}| \leqslant M$, то

$$\|P(1)\|_2^2 \leqslant M^2 (3D)^2 N_1 N_2 N_3 \leqslant C\varepsilon_1, \quad \text{при } M \leqslant \frac{\sqrt{C\varepsilon_1}}{3D\sqrt{N_1 N_2 N_3}}$$

Таким образом $\|L(1) - \sigma\|_2^2 \leqslant 2(6D + 1 + C)\varepsilon_1$

5 Заключение

Предложен новый подход к задаче малоранговой канонической аппроксимации с фиксированным рангом для 3-тензора. Проведены вычислительные эксперименты, которые показывают быструю сходимость предложенного алгоритма в случае хороших начальных приближений и возмущений.

Список литературы

1. *Beylkin, G.* Algorithms for numerical analysis in high dimensions / G. Beylkin, M. J. Mohlenkamp // *SIAM J. Sci. Comput.* "— 2005. "— Vol. 26, no. 6. "— P. 2133–2159.

2. *Bini, D.* Relations between exact and approximate bilinear algorithms: Applications / D. Bini // *Calcolo.* "— 1980. "— Vol. 17. "— P. 87–97.

3. *Comon, P.* Tensor decompositions, alternating least squares and other tales / P. Comon, X. Luciani, A. L. D. Almeida // *Journal of Chemometrics.* "— 2009. "— no. 23 (7-8). "— P. 393–405.

4. *de Silva, V.* Tensor rank and the ill-posedness of the best low-rank approximation problem / V. de Silva, L.-H. Lim // *SIAM J. Matrix Anal. Appl.* "— 2008. "— Vol. 30, no. 3. "— P. 1084–1127.

5. *Espig, M.* A regularized newton method for the efficient approximation of tensors represented in the canonical tensor format / M. Espig, W. Hackbusch // *Numer. Math.* "— 2012. "— Vol. 122, no. 3. "— P. 489–525.

6. *Harshman, R.* Foundations of the parafac procedure / R. Harshman // *UCLA Working Papers in Phonetics 16.* "— 1970. "— P. 1–84.

7. *Kazeev, V.* Structure of the hessian matrix and an economical implementation of newton's method in the problem of canonical approximation of tensors / V. Kazeev, E. Tyrtyshnikov // *Comput. Math. and Math. Phys.* "— 2010. "— Vol. 50, no. 6. "— P. 979–998.

8. *Madsen, K.* Methods for non-linear least squares problems / K. Madsen, H. B. Nielsen, O. Tingleff // *Informatics and mathematical Modelling. Technical University of Denmark.* "— 2004.

9. *Oseledets, I. V.* Minimization methods for approximating tensors and their comparison / I. V. Oseledets, D. V. Savost'yanov // *Comput. Math. and Math. Phys.* "— 2006. "— no. 46 (10). "— P. 1641–1650.

10. *Pan, V.* Matrix multiplication, trilinear decompositions, apa algorithms, and summation. "— Faculty Works. Paper 48. "— 2016.

11. *R.P.Brent,.* Algorithms for matrix multiplications. "— Comput.Sci.Dept.Report CS 157 Stanford Univ. "— 1970.

12. *Smirnov, A. V.* The bilinear complexity and practical algorithms for matrix multiplication / A. V. Smirnov // *Comput. Math. and Math. Phys.* "— 2013. "— no. 53 (12). "— P. 1781–1795.

13. *Smirnov, A. V.* A bilinear algorithm of length 22 for approximate multiplication of 2×7 and 7×2 matrices / A. V. Smirnov // *Comput. Math. and Math. Phys.* "— 2015. "— no. 55 (4). "— Pp. 541–545.

14. *Strassen, V.* Gaussian elimination is not optimal / V. Strassen // *Numer. Math.* "— 1969. "— no. 13. "— P. 354–356.