

УДК 519.61

## МИНИМИЗАЦИОННЫЕ МЕТОДЫ АППРОКСИМАЦИИ ТЕНЗОРОВ И ИХ СРАВНЕНИЕ<sup>1)</sup>

© 2006 г. И. В. Оселедец, Д. В. Савостьянов

(119991 Москва, ул. Губкина, 8, Ин-т вычисл. матем. РАН)

e-mail: oseledets1983@yandex.ru; draug@bach.inm.ras.ru

Поступила в редакцию 30.01.2006 г.  
Переработанный вариант 14.02.2006 г.

Рассматривается приложение различных минимизационных методов к проблеме трилинейной аппроксимации тензоров, и дается их сравнение на основе численных экспериментов. Для метода Гаусса–Ньютона предлагается эффективная реализация и приводится оценка локальной скорости сходимости этого метода для случая полностью симметричного тензора. Библ. 12. Фиг. 2.

**Ключевые слова:** трилинейная аппроксимация, минимизационные методы, нелинейная аппроксимация.

### ВВЕДЕНИЕ

Рассматриваются различные стандартные минимизационные методы в применении к проблеме *трилинейной аппроксимации тензоров*, которая формулируется следующим образом. Пусть дан некий трехмерный массив (тензор)  $\mathcal{A} = [a_{ijk}]$  размером  $n_1 \times n_2 \times n_3$ . Требуется найти такие матрицы  $U = [u_{i\alpha}]$ ,  $V = [v_{j\alpha}]$ ,  $W = [w_{k\alpha}]$  размеров  $n_1 \times r$ ,  $n_2 \times r$ ,  $n_3 \times r$ , которые минимизируют функционал

$$\left\| a_{ijk} - \sum_{\alpha=1}^r u_{i\alpha} v_{j\alpha} w_{k\alpha} \right\|, \quad (1)$$

где

$$\|\Phi_{ijk}\| = \left( \sum_{ijk} \Phi_{ijk}^2 \right)^{1/2}.$$

В том случае. Когда минимум равен нулю, говорят о *каноническом разложении тензора*, которое в иностранной литературе называется моделью CANDECOMP–PARAFAC.

Изучение разложений трехмерных массивов впервые началось в факторном анализе в статистике [1]. Сейчас трилинейное разложение применяется в задачах обработки сигнала, химии, психологии и социологии. Читателю, интересующемуся этой областью приложения трехмерных массивов, мы рекомендуем обзор [2]. Еще одной областью применения тензорных разложений является эффективное решение интегральных и дифференциальных уравнений на тензорных сетках, т.е. сетках, являющихся прямым произведением одномерных [3].

Рассмотрим три метода решения задачи (1):

ALS – метод переменных направлений,

GN – метод Гаусса–Ньютона,

LM – метод Ньютона с подбором шага.

Проведем их сравнение на некоторых примерах. Для метода Гаусса–Ньютона мы предлагаем эффективную реализацию для рассматриваемого функционала (1), которая понижает вычислительную сложность по сравнению со стандартным методом Гаусса–Ньютона и приводим оценку локальной скорости сходимости для случая полностью симметричного тензора.

<sup>1)</sup> Работа выполнена при частичной финансовой поддержке РФФИ (код проектов 04-07-90336 и 05-01-00721) и программы фундаментальных исследований Отделения математических наук РАН “вычисл. и информационные проблемы решения больших задач” по проекту “Матричные методы в интегр. и дифференц. ур-ниях”.

Перейдем к изложению этих методов.

## 1. МЕТОД ПЕРЕМЕННЫХ НАПРАВЛЕНИЙ

Метод переменных направлений, который в английской литературе (см., например, [1], [4]) называется ALS, является самым простым способом решения задачи (1). Он состоит в следующем. Пусть даны некоторые приближения  $U, V, W$  к решению  $U^*, V^*, W^*$  проблемы (1). Тогда по  $V, W$  находится новое значение  $\hat{U}$  из решения задачи наименьших квадратов:

$$\hat{U} = \operatorname{argmin}_U \left( \left\| a_{ijk} - \sum_{\alpha=1}^r u_{i\alpha} v_{j\alpha} w_{k\alpha} \right\|^2 \right). \quad (2)$$

Задача (2) распадается на  $n_1$  независимых подзадач:  $i = 1, 2, \dots, n_1$ :

$$\sum_{j=1}^{n_2} \sum_{k=1}^{n_3} \left( a_{ijk} - \sum_{\alpha=1}^r \hat{u}_{i\alpha} v_{j\alpha} w_{k\alpha} \right)^2 \rightarrow \min.$$

Поэтому, для нахождения  $\hat{u}_{i\alpha}$  необходимо решить  $n_1$  линейных задач наименьших квадратов с одной и той же матрицей (обозначим ее через  $\Phi$ ) размеров  $(n_2 n_3) \times r$ , но с разными правыми частями. Прямое применение стандартных методов (например, QR-разложения) к матрице  $\Phi$  потребует  $O(n_2 n_3 r^2)$  арифметических операций. Однако для матрицы  $\Phi$  можно быстро вычислить матрицу Грама

$$\Gamma = \Phi^T \Phi = (V^T V) \circ (W^T W),$$

где через  $\circ$  обозначается адамарово (поэлементное) произведение матриц. Матрица  $\Gamma$  может быть получена за  $O((n_2 + n_3)r)$  арифметических операций. После этого для нахождения  $\hat{u}_{i\alpha}$  нужно решить  $n_1$  систем с матрицей  $\Phi$ . Стоимость такого вычисления составляет  $O(r^3 + n_1 r^2)$  операций. После того как новая матрица  $U$  найдена, аналогичным способом по матрицам  $U, W$  оценивается матрица  $V$ , и так далее, до тех пор, пока не будет достигнута сходимость.

Метод ALS несложно запрограммировать, стоимость одной итерации невелика, и на каждой итерации невязка не возрастает. Все эти достоинства делают рассматриваемый метод самым популярным методом вычисления тензорных аппроксимаций, см. [1], [4]–[6]. Однако ALS имеет несколько очень существенных недостатков. Во-первых, число итераций, необходимых для получения удовлетворительных результатов, может быть очень большим (тысячи или даже миллионы). Во-вторых (и это самое главное), метод может попасть в локальный минимум функционала (1). Однако если имеется хорошее начальное приближение к точке минимума, то применение метода может оказаться довольно полезным.

## 2. МЕТОД ГАУССА–НЬЮТОНА

### 2.1. Описание метода Гаусса–Ньютона и общие оценки сходимости

Опишем метод Гаусса–Ньютона для минимизации функционала (1). Напомним основную идею метода. Пусть нужно решить переопределенную систему нелинейных уравнений

$$f(x) = 0, \quad (3)$$

где  $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$ ,  $m > n$ . Решением назовем вектор  $x$ , который минимизирует функционал

$$\|f(x)\|^2 \rightarrow \min,$$

т.е. решение понимается в смысле наименьших квадратов. Линеаризуя (3) относительно некоторой точки  $x$  (как в обычном методе Ньютона), получаем

$$f(x) + Df(x)\Delta x = 0,$$

где  $Df$  – якобиан отображения. Находим нормальное псевдорешение получившейся переопреде-

ленной системы линейных уравнений

$$\Delta x = -(Df(x))^{-} f, \tag{4}$$

где  $(Df(x))^{-}$  – псевдообратная по Муру–Пенроузу матрица к  $Df(x)$ . Выражение (4) и есть формула одного шага метода Гаусса–Ньютона. Исследование локальных свойств сходимости метода Гаусса–Ньютона можно найти в [7], из которой мы отметим два факта:

1) если решение системы (3) существует, то сохраняется квадратичная скорость сходимости обычного метода Ньютона;

2) если равенства (3) выполняются лишь с некоторой точностью  $\epsilon$ , то скорость сходимости становится линейной, однако показатель сходимости ведет себя как

$$\lambda = O(\|Df^{\dagger}\|\epsilon).$$

### 2.2. Метод Гаусса–Ньютона в нашем случае

Рассмотрим теперь Гаусса–Ньютона для минимизации функционала (1). Подчеркнем, что соответствующая система уравнений будет переопределенной, если  $n_1 n_2 n_3 > r(n_1 + n_2 + n_3)$ . После линеаризации получаем

$$\left\| \mathcal{R}_{ijk} - \sum_{\alpha=1}^r (\Delta u_{i\alpha} v_{j\alpha} w_{k\alpha} + u_{i\alpha} \Delta v_{j\alpha} w_{k\alpha} + u_{i\alpha} v_{j\alpha} \Delta w_{k\alpha}) \right\|^2 \rightarrow \min, \tag{5}$$

где введена невязка

$$\mathcal{R}_{ijk} = a_{ijk} - \sum_{\alpha=1}^r u_{i\alpha} v_{j\alpha} w_{k\alpha}.$$

В отличие от метода ALS, система (5) уже не распадается на отдельные системы. Матрица система (обозначим ее опять через  $\Phi$ ) является структурированной и имеет блочный вид

$$\Phi = [\Phi_1, \Phi_2, \Phi_3],$$

где элементы матриц  $\Phi_1, \Phi_2, \Phi_3$  задаются в виде

$$(\Phi_1)_{(ijk)(i'\alpha)} = \delta_{i'i} v_{j\alpha} w_{k\alpha}, \quad (\Phi_2)_{(ijk)(j'\alpha)} = \delta_{j'j} u_{i\alpha} w_{k\alpha}, \quad (\Phi_3)_{(ijk)(k'\alpha)} = \delta_{k'k} u_{i\alpha} v_{j\alpha}.$$

В этих выражениях  $\Phi_1, \Phi_2, \Phi_3$  были пронумерованы тремя индексами  $(i, j, k)$  а столбцы – парами  $(i'\alpha), (j'\alpha), (k'\alpha)$  соответственно. Величина  $\delta_{i'i}$  равна единице при  $i = i'$  и нулю в противном случае.

Сложность прямого метода с использованием QR-разложения составляет

$$O(n_1 n_2 n_3 (n_1 + n_2 + n_3) r^2)$$

операций. Если же мы будем использовать метод нормальных уравнений, для которого необходимо вычислять матрицу Грама  $\Gamma = \Phi^T \Phi$ , то мы можем уменьшить вычислительную сложность. Система нормальных уравнений имеет вид

$$\begin{aligned} \sum_{\alpha} \Delta u_{i\alpha} (v_{\alpha}, v_{\beta})(w_{\alpha}, w_{\beta}) + \sum_{\alpha} u_{i\alpha} (\Delta v_{\alpha}, v_{\beta})(w_{\alpha}, w_{\beta}) &= \sum_{jk} \mathcal{R}_{ijk} v_{j\beta} w_{k\beta}, \\ \sum_{\alpha} \Delta v_{j\alpha} (w_{\alpha}, w_{\beta})(u_{\alpha}, u_{\beta}) + \sum_{\alpha} v_{j\alpha} (\Delta w_{\alpha}, w_{\beta})(v_{\alpha}, v_{\beta}) &= \sum_{ki} \mathcal{R}_{ijk} w_{k\beta} u_{i\beta}, \\ \sum_{\alpha} \Delta w_{k\alpha} (u_{\alpha}, u_{\beta})(v_{\alpha}, v_{\beta}) + \sum_{\alpha} w_{k\alpha} (\Delta u_{\alpha}, u_{\beta})(v_{\alpha}, v_{\beta}) &= \sum_{ij} \mathcal{R}_{ijk} u_{i\beta} v_{j\beta}. \end{aligned} \tag{6}$$

Введем “матрицы моментов”  $X, Y, Z$  и матрицы Грама  $\Gamma_u, \Gamma_v, \Gamma_w$

$$X = (\Delta u_{\alpha}, u_{\beta}), \quad Y = (\Delta v_{\alpha}, v_{\beta}), \quad Z = (\Delta w_{\alpha}, w_{\beta}),$$

$$\Gamma_u = U^T U = (u_{\alpha}, u_{\beta}), \quad \Gamma_v = V^T V = (v_{\alpha}, v_{\beta}), \quad \Gamma_w = W^T W = (w_{\alpha}, w_{\beta}).$$

Домножая равенства (6) на  $u_{i\alpha}, v_{j\alpha}, w_{k\alpha}$  и суммируя по  $i, j, k$  соответственно, получаем следующие уравнения для матриц моментов  $X, Y, Z$ :

$$\begin{aligned} (\Gamma_v \circ \Gamma_w)X + (\Gamma_w \circ Y^T + \Gamma_v \circ Z^T)\Gamma_u &= Q_1, \\ (\Gamma_w \circ \Gamma_u)Y + (\Gamma_u \circ Z^T + \Gamma_w \circ X^T)\Gamma_v &= Q_2, \\ (\Gamma_u \circ \Gamma_v)Z + (\Gamma_v \circ X^T + \Gamma_u \circ Y^T)\Gamma_w &= Q_3, \end{aligned} \tag{7}$$

где

$$Q_1 = \sum_{ijk} \mathcal{R}_{ijk} u_{i\alpha} v_{j\beta} w_{k\beta}, \quad Q_2 = \sum_{ijk} \mathcal{R}_{ijk} u_{i\beta} v_{j\alpha} w_{k\beta}, \quad Q_3 = \sum_{ijk} \mathcal{R}_{ijk} u_{i\beta} v_{j\beta} w_{k\alpha},$$

или в матрично-векторной форме

$$M(U, V, W)\Delta = q. \tag{8}$$

Число неизвестных теперь стало равным  $3r^2$ , поэтому для решения системы (7) прямым методом требуется  $O(r^6)$  арифметических операций. Суммарная стоимость вычисления матриц Грама и матриц  $Q_1, Q_2, Q_3$  составляет  $O(n_1 n_2 n_3 r^2)$  операций.

Важно отметить, что эта система структурированная, и ее матрица разрежена. Умножение матрицы системы на вектор требует  $O(r^3)$  операций, поэтому разумно использовать итерационный метод, например метод сопряженных градиентов. Однако для устойчивой сходимости нужно использовать предобуславливатель.

### 2.3. Сходимость

Анализ сходимости проведем в частном, но довольно часто встречающемся случае абсолютно симметричного тензора, т.е.  $n_1 = n_2 = n_3 = n$ , и

$$a_{ijk} = a_{P(i, j, k)},$$

где  $P(i, j, k)$  – любая перестановка индексов  $(i, j, k)$ . В этом случае факторы  $U, V, W$  в (1) совпадают. Для анализа сходимости мы должны определить, при выполнении каких условий матрица  $\Phi = Df(u)$  имеет полный столбцовый ранг, и оценить  $\|\Phi^{-1}\|$ . Ответ на эти вопросы дает следующая

**Теорема.** Пусть матрица  $U = [u_{i\alpha}] \in \mathbb{R}^{n \times r}$  имеет полный столбцовый ранг. Тогда якобиан  $Df(u)$  имеет полный столбцовый ранг  $nr$ , матрица  $M(U)$  из (8) невырождена и имеет место оценка

$$\sigma_{\min}(Df(u)) \geq \sigma_{\min}(U)^2.$$

**Доказательство.** Рассмотрим произведение матрица  $Df(u)$  на вектор  $\Delta u$ :

$$h = (Df(u)\Delta u)_{ijk} = \sum_{\alpha=1}^r \Delta u_{i\alpha} u_{j\alpha} u_{k\alpha} + u_{i\alpha} \Delta u_{j\alpha} u_{k\alpha} + u_{i\alpha} u_{j\alpha} \Delta u_{k\alpha}.$$

так как матрица  $U$  имеет полный столбцовый ранг, то существует такая невырожденная матрица  $\Psi \in \mathbb{R}^{n \times n}$ , что

$$\sum_{i'=1}^n u_{i'\alpha} \Psi_{i'i} = \delta_{i\alpha}, \quad i = 1, 2, \dots, n, \quad \alpha = 1, 2, \dots, r,$$

и при этом

$$\|\Psi\| \leq \frac{1}{\sigma_{\min}(U)}.$$

Определим вектор  $\hat{h}$ :

$$\hat{h} = \sum_{i'jk} h_{i'jk} \Psi_{i'i} \Psi_{j'j} \Psi_{k'k}.$$

Тогда имеем

$$\hat{h} = Df(I_r)\Delta\hat{u},$$

где через  $I_r$  обозначена матрица размера  $n$  на  $r$  с элементами  $\delta_{i\alpha}$ , а

$$\Delta\hat{u} = \Psi\Delta u.$$

Матрица  $Df(I_r)$  имеет полный столбцовый ранг. Ее сингулярные числа есть корни квадратные из собственных значений матрицы  $M(I_r) = Df(I_r)^T Df(I_r)$ . Легко видеть, что матрица  $M(I_r)$  является диагональной с элементами на диагонали, равными единице или тройке. Поэтому имеем

$$\|\hat{h}\| \geq \|\Delta\hat{u}\| \geq \|\Psi\|\|\Delta u\|.$$

Так как

$$\hat{h} = \sum_{i,j,k} h_{rjk} \Psi_{ir} \Psi_{jj} \Psi_{kk},$$

то

$$\|\hat{h}\| \leq \|\Psi\|^3 \|h\|,$$

поэтому

$$\|h\| = \|Df(u)\Delta u\| \geq \frac{\|\Delta u\|}{\|\Psi\|^2},$$

откуда следует, что  $Df(u)$  имеет полный столбцовый ранг, и справедлива оценка

$$\sigma_{\min}(Df(u)) \geq \sigma_{\min}(U)^2.$$

Из теоремы легко получить верхнюю оценку на показатель скорости сходимости метода

$$\lambda = O(\varepsilon \text{cond}(U)^2),$$

где  $\varepsilon$  – это значение минимума функционала отклонения. Надо отметить, что эта оценка является довольно грубой. В частности, матрица  $U$  может иметь линейно-зависимые столбцы, а якобиан все равно будет невырожденным.

### 3. МЕТОД НЬЮТОНА С ПОДБОРОМ ШАГА

#### 3.1. Описание метода

Мы также использовали для решения задачи (1) демпфированный метод Ньютона с подбором шага методом Левенберга–Марквардта. Применение этого метода можно рассматривать как обобщение подхода, описанного в [8], где для решения задачи 1) в случае  $r = 1$  применен обычный метод Ньютона. Демпфированный метод Ньютона является гибридом метода наискорейшего спуска для минимизации функционала  $F(x)$  шаги выбираются вдоль направления антиградиента (steepest descent)

$$h_{sd} = -F'(x),$$

а в методе Ньютона шаги определяются из решения уравнения

$$(F''(x))h_N = -F'(x).$$

Метод наискорейшего спуска обладает линейной сходимостью, метод Ньютона – квадратичной, но только в некоторой окрестности стационарной точки. Таким образом, метод наискорейшего спуска имеет смысл использовать на начальной, а метод Ньютона – на конечной стадии вычислений. Когда  $x$  близок к искомому  $x^*$ . Существуют различные методы “умного” переключения между этими двумя методами, одним из которых является демпфированный метод Ньютона. В нем шаг итерации определяется из решения уравнения

$$(F''(x) + \mu I)h = -F'(x),$$

и поведение алгоритма на текущей итерации определяется значением параметра  $\mu$ . Конкретнее,

при очень больших значениях  $\mu$  направление шага  $h$  практически совпадает с направлением  $h_{sd}$ :

$$h \approx -\frac{1}{\mu} F'(x),$$

при очень малых – с направлением  $h_N$ .

Поскольку параметр  $\mu$  определяет не только направление, но и длину шага, становится возможным предложить метод, не требующий отдельной процедуры выбора длины шага. Существенным, однако, является вопрос о выборе  $\mu$  в течение итерационного процесса. В методе Левенберга–Марквардта изменение величины  $\mu$  определяется тем, насколько хорошо поведение функции  $F(x)$  на текущем шаге может быть приближено с помощью квадратичной модели

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

Конкретнее, на каждом шаге вычисляется коэффициент выгода

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

Если  $\rho$  достаточно велик (близок к единице), это означает, что квадратичная модель  $L(h)$  достаточно близка к значению  $F(x+h)$ , а стало быть, в этой области хорошо применим метод Ньютона и значение  $\mu$  можно уменьшить, чтобы шаг следующей итерации был более близок к шагу  $h_N$ . Если значение  $\rho$  невелико или вообще отрицательно, то квадратичная модель плохо описывает поведение функции и необходимо увеличить значение  $\mu$ , с тем чтобы, во-первых, сделать шаг более близким к шагу метода наискорейшего спуска  $h_{sd}$ , а во-вторых, уменьшить его длину.

Изначально стратегия модернизации  $\mu$  была предложена Марквардтом, мы применяем более гладкий критерий, предложенный в [9]. Указанные стратегии обновления отображены графически на фиг. 1 (стратегия Марквардта обозначена штриховой линией, сглаженная – сплошной линией).

В окончательном виде метод Ньютона с выбором шага по критерию Левенберга–Марквардта выглядит так:

#### Алгоритм (Newton)

Установить  $k = 0$ ,  $x = x_0$ ,  $A = F''(x)$ ,  $g = F'(x)$ ;

done = .false.

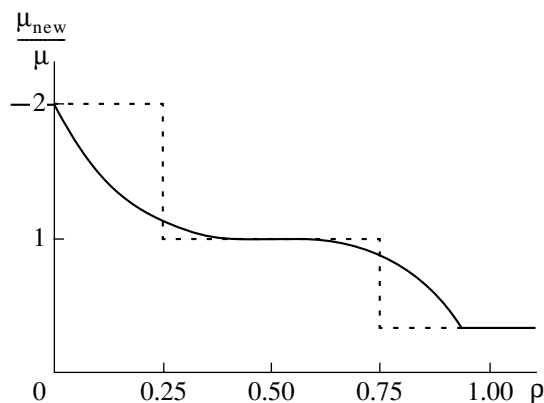
**пока** ( $k < k_{\max}$ ) и (.no.done)

установить  $k := k + 1$ ;

решить  $(A + \mu I)h = -g$  (разложением Холецкого);

**если** шаг  $h$  достаточно мал

done = .true.



Фиг. 1.

**иначе**

$$x_{\text{new}} = x + h, \rho := \frac{F(x) - F(x_{\text{new}})}{L(0) - L(h)};$$

**если**  $\rho > 0$

$$x := x_{\text{new}}, A := F''(x), g := F'(x);$$

**если**  $g$  достаточно мал

done = .true.

$$\mu := \mu \max\left(\frac{1}{3}, 1 - (2\rho - 1)^3\right);$$

**иначе**

$$\mu := 2\mu$$

Этот метод, как и метод наискорейшего спуска, является вполне робастным, и, в случае если  $F''(x^*)$  невырождена, его сходимость на финальном участке квадратичная. Сложность метода  $O(n^3)$  на одну итерацию (разложение Холецкого).

### 3.2. Функционал отклонения и его производные

Для решения задачи (1) проведем минимизацию функционала

$$F(u, v, w) = \frac{1}{2} \sum_{ijk} \left( \sum_{\alpha} u_{i\alpha} v_{j\alpha} w_{k\alpha} - a_{ijk} \right)^2.$$

Градиент  $F'$  выражается в виде

$$\frac{\partial F}{\partial u_{i\alpha}} = \sum_{jk} \mathcal{R}_{ijk} v_{j\alpha} w_{k\alpha}, \quad \frac{\partial F}{\partial v_{j\alpha}} = \sum_{ki} \mathcal{R}_{ijk} w_{k\alpha} u_{i\alpha}, \quad \frac{\partial F}{\partial w_{k\alpha}} = \sum_{ij} \mathcal{R}_{ijk} u_{i\alpha} v_{j\alpha},$$

где

$$\mathcal{R}_{ijk} = \sum_{\beta} u_{i\beta} v_{j\beta} w_{k\beta} - a_{ijk}.$$

Гессиан  $F''$  определяется формулами

$$\frac{\partial^2 F}{\partial u_{i\alpha} \partial u_{i\beta}} = \sum_{jk} v_{j\alpha} v_{j\beta} w_{k\alpha} w_{k\beta} = (v_{\alpha}, v_{\beta})(w_{\alpha}, w_{\beta}),$$

$$\frac{\partial^2 F}{\partial u_{i\alpha} \partial u_{r\beta}} = 0, \quad i \neq r,$$

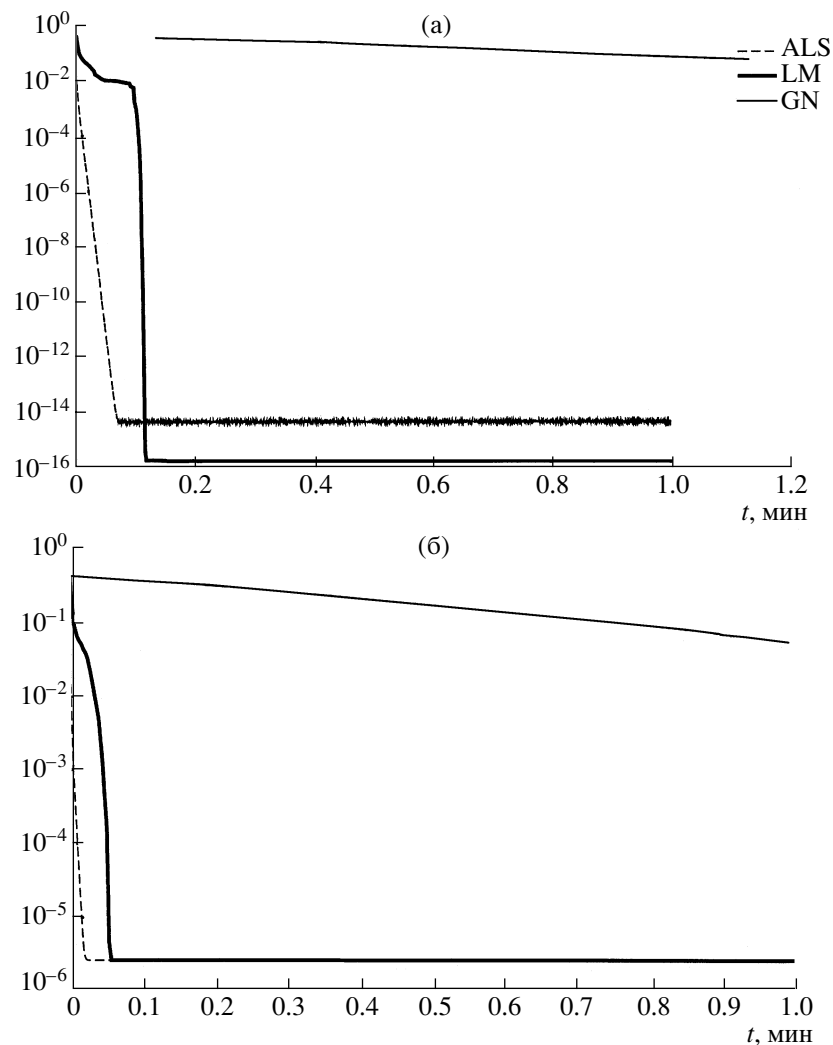
$$\frac{\partial^2 F}{\partial u_{i\alpha} \partial v_{j\beta}} = u_{i\beta} v_{j\alpha} (w_{\alpha}, w_{\beta}) + \delta_{\alpha\beta} \sum_k \mathcal{R}_{ijk} w_{k\alpha},$$

где  $\delta_{\alpha\beta} = 1$  при  $\alpha = \beta$  и  $\delta_{\alpha\beta} = 0$  при  $\alpha \neq \beta$ .

Прочие частные производные, необходимые для получения градиента  $F'(x)$  и гессиана  $F''(x)$ , определяются аналогично.

## 4. ЧИСЛЕННЫЕ ЭКСПЕРИМЕНТЫ

При сравнении различных минимизационных алгоритмов нас в первую очередь интересует скорость убывания целевой функции  $F(x)$ . Рассматриваемые нами методы обладают весьма разной асимптотикой сложности, поэтому использовать для их сравнения убывание  $F(x)$  после совершения одного и того же числа итераций было бы некорректно. Поэтому мы будем сравнивать уменьшение целевой функции (нормы отклонения), которое различные методы произвели за одинаковое время счета.



Фиг. 2.

Рассмотрев фиг. 2, демонстрирующие убывание невязки в ходе работы минимизационных методов на типовых для нас примерах массивов, можно дать следующие (не претендующие на глобальность) комментарии.

На фиг. 2а рассмотрен случайный массив размера  $16 \times 16 \times 16$  и точного ранга 16. Случайное начальное приближение.

Метод Ньютона дает более качественное разложение, чем метод переменных направлений, однако более медленное. Метод Гаусса–Ньютона не успевает заметно уменьшить невязку.

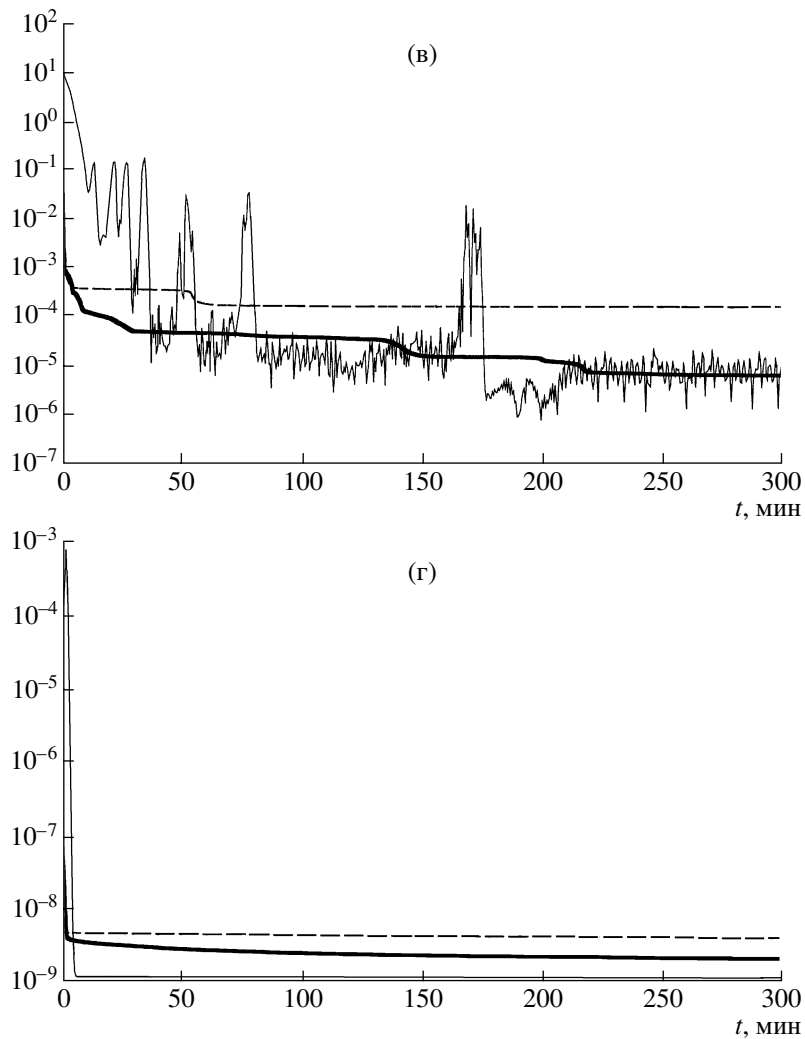
Случайный массив размера  $16 \times 16 \times 16$  и ранга 1 представлены на фиг. 2б. Наложены мультипликативный шум порядка  $10^{-5}$ . Случайное начальное приближение.

Метод Ньютона сходится к тому же решению, что и метод переменных направлений, но второй метод делает это быстрее. Метод Гаусса–Ньютона сходится слишком медленно.

На фиг. 2в рассматривается ядро разложения Таккера (см. [10]) размера  $20 \times 20 \times 20$ . Исходный массив имел элементы порядка  $1/r$ , где  $r \sim \sqrt{i^2 + j^2 + k^2}$ . Ищется трilinearное разложение ранга 20. Случайное начальное приближение.

Метод Ньютона находится более глубокий минимум функционала невязки, чем метод переменных направлений, однако это требует весьма значительного времени. Метод Гаусса–Ньютона дает еще больший выигрыш по точности, однако его сходимость столь же медленная, и кроме того, существенно немонотонная.





Фиг. 2. (Окончание).

Ядро разложения Таккера размера  $20 \times 20 \times 20$  рассматривается на фиг. 2. Исходный массив имел элементы порядка  $1/r$ . Ищется трилинейное разложение ранга 20. Начальное приближение – суперобобщенное разложение Шура (см. [11]).

Метод переменных направлений вновь оказывается самым быстрым, но наименее точным. Метод Ньютона работает медленнее, но обходит метод переменных направлений по точности. Метод Гаусса–Ньютона сначала ухудшает начальную невязку, но в итоге сходится к самому точному решению, хотя по скорости заметно проигрывает им обоим.

## ВЫВОДЫ

Итак, самый простой в реализации метод переменных направлений оказывается вполне хорошо применимым на практике. Для еще более хороших результатов можно использовать комбинации различных минимизационных техник, например, использовать метод переменных направлений на начальном шаге и найденное им решение подавать в метод Ньютона как начальное приближение. Решение задачи трилинейного разложения можно существенно ускорить также за счет использования для построения начального приближения специальных “матричных” алгоритмов, напрямую использующих специальный вид минимизируемого функционала (некоторый обзор таких методов предложен в [12]). Усовершенствование методов типа Ньютона возможно с применением методов доверительных областей.

Авторы выражают благодарность рецензенту за полезные замечания.

## СПИСОК ЛИТЕРАТУРЫ

1. *Harshman R.A.* Foundations of the Parafac procedure: Models and conditions for an explanatory multimodal factor analysis: UCLA Working Papers in Phonetics. 1970. V. 16. P. 1–84.
2. *Comon P.* Tensor decomposition: State of the art and applications // IMA Conf. Math. Signal Proc. Warwick, UK, Dec. 18–20, 2000. <http://www.i3s.fr/~comon/FichiersPs/ima2000.ps>
3. *Ibraghimov I.* Application of the three-way decomposition for matrix compression // Numer. Linear Algebra Appl. 2002. V. 9. № 6–7. P. 551–565.
4. *Caroll J.D., Chang J.J.* analysis of individual differences in multidimensional scaling via n-way generalization of Eckart-Young decomposition // Psychometrika. 1970. V. 35. P. 283–319.
5. *Bro R.* PARAFAC: Tutorial and applications // Chemom. Intel. Lab. Systems. 1997. V. 38. P. 149–171.
6. *Wang J.-H., Hopke P.K., Hanczewicz T.M., Zhang S.L.* Application of modified least squares regression to spectroscopic image analysis // *Analys Chim. Acta.* 2003. V. 476. P. 93–109.
7. *Dedieu J.P., Schub M.* Newton's method for overdetermined systems of equations // *Math. Comput.* 2000. V. 69(281). P. 1099–1115.
8. *Zhang T., Golub G.H.* Rank-one approximation to high-order tensors // *SIAM J. Matrix Analys Appl.* 2001. V. 23. P. 534–550.
9. *Nielsen H.B.* Damping parameter in Marquardt's method // *Inform. Math. Modellig, Denmark*, 1999. <http://www.imm.dtu.dk/~hbn/publ/TR9905.ps>
10. *Tucker L.R.* Some mathematical notes on three-mode factor analysis // *Psychometrika.* 1966. V. 31. P. 279–311.
11. *Оселедец И.В., Савостьянов Д.В.* Быстрый алгоритм для одновременного приведения матриц к треугольному виду и аппроксимации тензоров // *Матричные методы и технологии решения больших задач.* М.: ИВМ РАН, 2005. С. 101–116.
12. *Оселедец И.В., Савостьянов Д.В.* Методы разложения тензора // *Матричные методы и технологии решения больших задач.* М.: ИВМ РАН, 2005. С. 51–64.