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

Курсовая работа

Исследование методов восстановления малоранговых матриц и тензоров по части элементов.

выполнила студентка 503 группы
Ахиярова Э. А.

научный руководитель
д.ф.-м.н., профессор Оселедец И. В.

Москва, 2016

Содержание

1 Введение **2**

2 Восстановление матрицы **3**

2.1 Теоретические сведения 3

2.2 Численные эксперименты 6

3 Восстановление тензора **8**

3.1 Многомерное обобщение ядерной номы 8

3.2 Численные эксперименты 9

3.3 Альтернативный аналог 12

4 Заключение **14**

1 Введение

За последние несколько лет задача восстановления матрицы по части ее элементов привлекала большое внимание. Работы [1], [2], [3], [4], посвящены решению этой проблемы путем минимизации выпуклого функционала, представляющим собой ядерную норму матрицы. Одним из важных результатов этих работ является оценка на число элементов, необходимое для полного восстановления матрицы размера $n \times n$ - $O(nr\max\{\mu_0, \mu_1^2\}\log^2(n))$, где r - ранг матрицы, а $\mu_0,\ \mu_1$ - параметры когерентности. В свою очередь, Кандес и Тао доказали, что $O(nr\mu_0\log(n))$ известных элементов являются необходимым условием для того, чтобы минимизация ядерной нормы была близка к оптимальному решению исходной задачи. Представленный метод для восстановления матрицы по ее элементам теоретически обоснован и широко используется в приложениях.

В многомерном случае данный вопрос изучен менее подробно. Одной из причин этому является то, что задача нахождения ядерной нормы для тензора принадлежит NP классу ([5]), и поэтому минимизация соответствующего функционала является очень трудоемкой в плане компьютерных вычислений. В двумерном случае исходная проблема сводится к минимизации ядерной нормы матрицы, и как правило, тензорное обобщение данной нормы часто представляет собой линейные комбинации норм его разверток. Существующие нижние оценки на число известных элементов, необходимое для надежного восстановления тензора показывают неоптимальность выбранных моделей.

В данной работе рассмотрены теоретические сведения о методах восстановления матриц и тензоров по части их элементов, а также приведены соответствующие численные эксперименты.

2 Восстановление матрицы

2.1 Теоретические сведения

Во многих приложениях линейной алгебры возникает задача восстановления матрицы M размера $n \times n$ по известным m элементам. При этом обычно предполагается, что число известных m гораздо меньше общего числа элементов матрицы, n^2. Конечно, в общем случае решить такую задачу без дополнительных ограничений на матрицу невозможно. Очень часто, такая матрица является малоранговой, поэтому исходная проблема может быть сформулирована в следующем виде:

$$\begin{aligned} &\min_X rank(X), \\ &\quad X_{ij} = M_{ij}, \quad (i,j) \in \Omega, \quad |\Omega| = m \end{aligned} \tag{1}$$

К сожалению формулировка (1) почти не находит практического применения, так как принадлежит NP-классу.

В работе [1] было предложено заменить целевую функцию задачи (1) на ядерную норму:

Определение 1 **Ядерной нормой** *матрицы X называется сумма ее сингулярных значений:*

$$\|X\|_* = \sum_{k=1}^{r} \sigma_k(X) \tag{2}$$

Тогда (1) перепишется в виде:

$$\begin{aligned} &\min_X \|X\|_*, \\ &\quad X_{ij} = M_{ij}, \quad (i,j) \in \Omega, \quad |\Omega| = m \end{aligned} \tag{3}$$

Такая замена вполне естественна, так как любая матрица ранга r имеет r ненулевых сингулярных значений, и вместо того, чтобы минимизировать количество ненулевых сингулярных чисел, мы минимизируем их сумму. Более того, функционал (2) является выпуклым, и, как показанов работе [1], задача (3) может быть эффективно оптимизирована с помощью SDP.

Возникает вопрос, будет ли решение (3) соответствовать исходной матрице M? Ответ на него дает следующая теорема ([1]):

Теорема 1 *Пусть M - матрица размера $n \times n$ ранга r, m - число ее элементов, выбранных в произвольном порядке. Тогда, если найдутся такие положительные константы C и c, что*

$$m \geq Cn^{5/4} r \log n, \tag{4}$$

то, решение задачи (3) единственно и совпадает с матрицей M с вероятностью $\mathbf{P} \geq 1 - cn^{-3}$. Если же $r \leq n^{1/5}$, то точное восстановление матрицы с вероятностью $\mathbf{P} \geq 1 - cn^{-3}$ гарантируется при

$$m \geq Cn^{6/5} r \log n \tag{5}$$

То есть, для восстановления малоранговой матрицы достаточно знать довольно небольшое число ее элементов. Для малых r, например для $r = O(1)$ или $r = O(\log n)$ необходимо порядка $n^{6/5}$ элементов,что гораздо меньше n^2 - всего количества элементов квадратной матрицы. Таким образом, для большинства задач, минимизация ядерной нормы формально эквивалентна комбинаторной задаче минимизации ранга (1).

Теорема 1 является частным случаем другого утверждения, приведенного в работе [1]. Перед тем, как ознакомиться с ним, введем дополнительное определение параметра когерентности:

Определение 2 *Пусть U - подпространство R^n размерности n, а P_U - оператор ортогонального проектирования на U, тогда ко-*

герентность μ пространства U определяется следующим соотношением:

$$\mu(U) = \frac{n}{r} \max_{1 \leq i \leq n} \|P_U e_i\|^2, \tag{6}$$

где $e_i,\ i = \overline{1, n}$ - стандартный базис пространства R^n.

Пусть $U,\ V$ - левый и правый фактор сингулярного разложения матрицы M. Сделаем следующие предположения:

1. Существует μ_0, такое, что $\max(\mu(U), \mu(V)) \leq \mu_0$
2. $\|UV^*\|_C \leq \mu_1 \sqrt{r}/n$ для некоторого $\mu_1 > 0$

Основным результатом работы [1] является следующее утверждение:

Теорема 2 *Пусть для матрицы M размера $n \times n$ ранга r справедливы предположения 1 и 2, m - число известных элементов матрицы, выбранных в произвольном порядке. Если найдутся такие константы C и c, что:*

$$m \geq C \max(\mu_1^2, \mu_0^1 \mu_1, \mu_0 n^1) nr(\beta \log n) \tag{7}$$

для некоторого $\beta > 2$, то решение задачи (3) единственно и совпадает с матрицей M с вероятностью $\mathbf{P} \geq 1 - cn^{-\beta}$. Для $r \leq \mu_0^{-1} n^{1/5}$ оценка на m может быть улучшена с такой же вероятностью успешного восстановления:

$$m \geq C \mu_0 n^{6/5} r(\beta \log n) \tag{8}$$

Таким образом, чем меньше параметр когерентности матрицы, тем меньше параметров понадобится для ее восстановления. Например, если $\mu_0 = O(1)$, то для параметра m должно выполняться неравенство:

$$m \geq C n^{6/5} r \log n \tag{9}$$

2.2 Численные эксперименты

Для демонстрации практического примения минимизации ядерной нормы по части ее элементов было проведено несколько численных экспериментов для различных размеров матриц n, рангов r и чисел m. Для каждой тройки (n, m, r) следующая процедура повторялась 25 раз. Генерировалась произвольная матрица M ранга r. Выбирался произвольный набор элементов длины m и минимизация ядерной нормы

$$\begin{aligned} &\min_X \|X\|_*, \\ &\quad X_{ij} = M_{ij}, \quad (i,j) \in \Omega, \quad |\Omega| = m \end{aligned} \tag{10}$$

осуществлялась с помощью SDP решателя SCS пакета cvxpy. Если решение X_0, возвращаемое минимизирующей функцией удовлетворяла условию

$$\frac{\|X_0 - M\|_F}{\|M\|_F} \leq 10^{-2}, \tag{11}$$

то объявлялось, что матрица M восстановлена. Рис. 1 иллюстрирует результаты этих тестов при $n = 40$ и 50. По оси x откладывается отношение числа известных решателю элементов ко всему числу элементов матрицы. Ось y представляет собой отношение между размерностью множества матриц ранга r, $D = r(2n - r)$ и числом измерений m. Заметим, что оси x и y лежат в пределах отрезка $[0, 1]$. Если абсцисса больше 1, то это означает, что мы имеем дело с переопределенной системой, и решатель всегда будет выдавать успешный результат. Если же ордината больше 1, то это приводит к существованию бесконечного числа матриц ранга r с известными элементами. Цвет каждой клетки отражает степень восстановления матрицы за 25 запусков. Белый цвет обозначает успешное восстановление в каждом из 25 экспериментов, черный - абсолютный провал на всех запусках.

На Рис. 2 представлены процессы сходимости решателя для матрицы размера 40×40 для рангов $r = 1,\ 5,\ 10$ в зависимости от числа известных элементов m

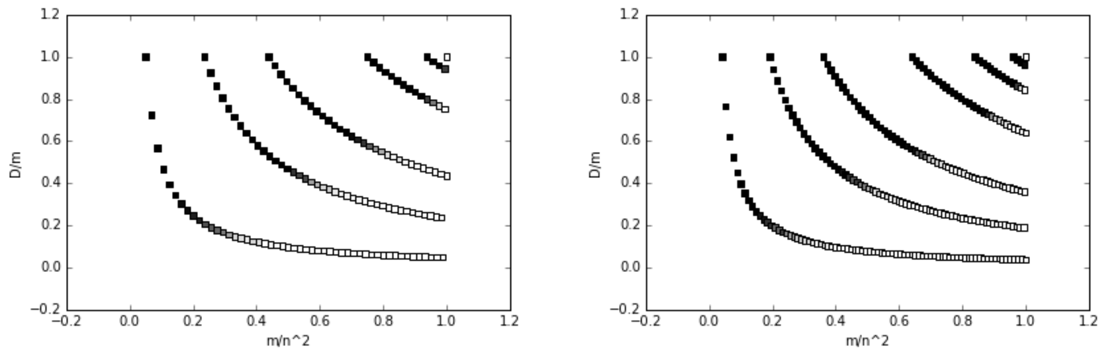

Рис. 1: $n = 40,\ 50$

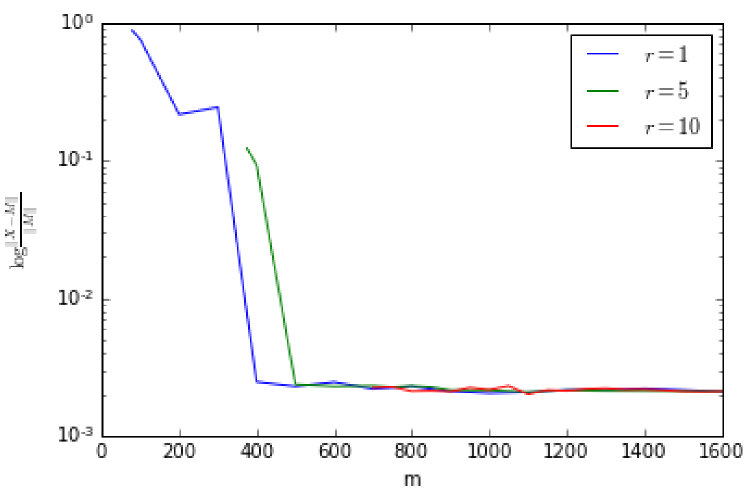

Рис. 2: $n = 40$

3 Восстановление тензора

3.1 Многомерное обобщение ядерной номы

Рассмотрим многомерный аналог задачи восстановления матрицы. Пусть известны m элементов тензора $\mathfrak{T} \in R^{n_1 \times \ldots, \times n_d}$. Как и в двумерном случае, мы предполагаем, что его ранг мал:

$$\begin{gathered} \min_{\mathfrak{X}} rank(\mathfrak{X}) \\ \mathfrak{X}_\Omega = \mathfrak{T}_\Omega \end{gathered} \tag{12}$$

Конечно, здесь не следует трактовать $rank(\mathfrak{T})$ в смысле канонического ранга из-за того, что в этом случае задача (12) становится NP-трудной. Используем понятие TT-ранга([8]):

$$\begin{gathered} \min_{\mathfrak{X}} rank(\mathfrak{X}) = (r_1, \ldots, r_{d-1}) \\ \mathfrak{X}_\Omega = \mathfrak{T}_\Omega, \end{gathered} \tag{13}$$

Здесь, числа $r_1, \ldots, r_{d-1}$ - ранги разверток тензора. Можно использовать следующее обобщение ядерной нормы:

$$\|\mathfrak{X}\|_* = \sum_{k=1}^{d-1} \alpha_k \|X_k\|_*, \tag{14}$$

где X_k $k = \overline{1, d-1}$ - развертки тензора ($X_k = reshape(\mathfrak{X}, n_1 \ldots n_k \times n_{k+1} \ldots n_d)$)

Таким образом, оптимизационная задача (13) примет вид:

$$\begin{gathered} \min_{\mathfrak{X}} \sum_{k=1}^{d-1} \alpha_k \|X_k\|_* \\ \mathfrak{X}_\Omega = \mathfrak{T}_\Omega \end{gathered} \tag{15}$$

Введем дополнительные неизвестные $M_k,\ k = \overline{1, d-1}$:

$$\min_{\mathfrak{X}} \sum_{k=1}^{d-1} \alpha_k \|X_k\|_* + \frac{\beta_k}{2}\|X_k - M_k\|_F^2 \qquad (16)$$
$$\mathfrak{X}_\Omega = \mathfrak{T}_\Omega$$

В [7] для решения (16) использовался метод переменных направлений. Представим численные эксперименты реализации этого метода, выполненного в рамках данной работы.

3.2 Численные эксперименты

На языке Python был реализован следующий алгоритм для решения (16).

• **Оптимизация по $\mathbf{M_k}$** , $k = \overline{1, d-1}$:

$$M_k = Shrink_{\frac{\alpha_k}{\beta_k}}(X_k) \qquad (17)$$

• **Оптимизация по $\mathfrak{X}$** :

$$\mathfrak{X}_{i_1 i_2 \ldots i_d} = \begin{cases} \dfrac{\sum\limits_{k=1}^{d-1} \beta_k Fold(M_k)}{\sum\limits_{k=1}^{d-1} \beta_k}, & i_1 i_2 \ldots i_d \notin \Omega \\ t_{i_1 i_2 \ldots i_d}, & i_1 i_2 \ldots i_d \in \Omega \end{cases} \qquad (18)$$

На Рис.4 Представлены процессы сходимости в зависимости от числа итераций для различных параметров α и β для задачи восстановления тензора $\mathfrak{T} \in R^{15\times15\times15\times15}$ с tt-рангами, равными 3. Здесь $err = \|\mathfrak{X} - \mathfrak{T}\|_F / \|\mathfrak{T}\|_F$: Видно, что сходимость метода переменных направлений довольна медленная. Однако для тензора размера $10 \times 10 \times 10$

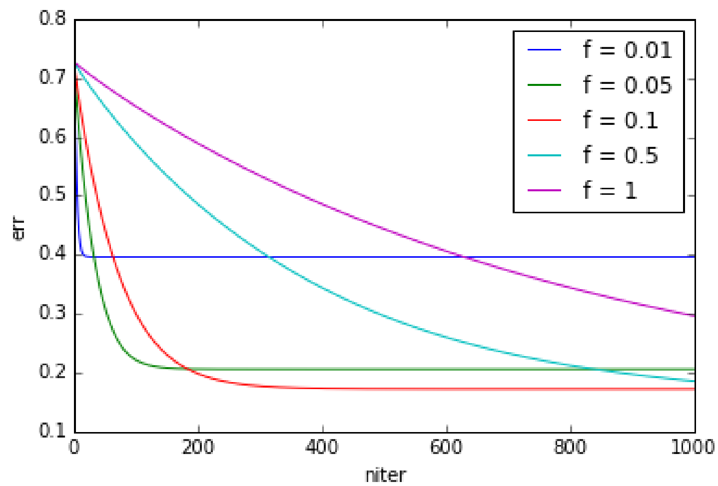

Рис. 3: $T \in R^{15\times15\times15\times15}$, $r_1 = r_2 = r_3 = 3$

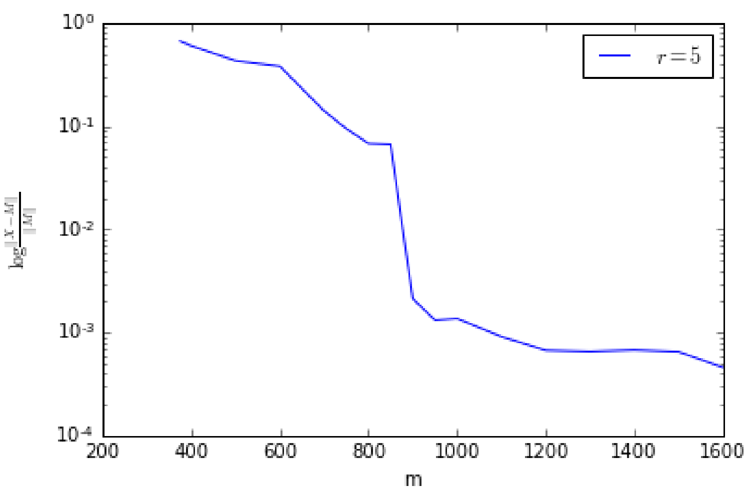

Рис. 4: $T \in R^{10\times10\times10}$, $r_1 = r_2 = 3$

удалось подобрать такие параметры α и β, что относительная погрешность решения вела себя гораздо лучше. (Рис. 4)

На Рис.5 изображена фазовая диаграмма, отображающая степень восстановления тензора той же размерности для $N = 15$ запусков алгоритма.

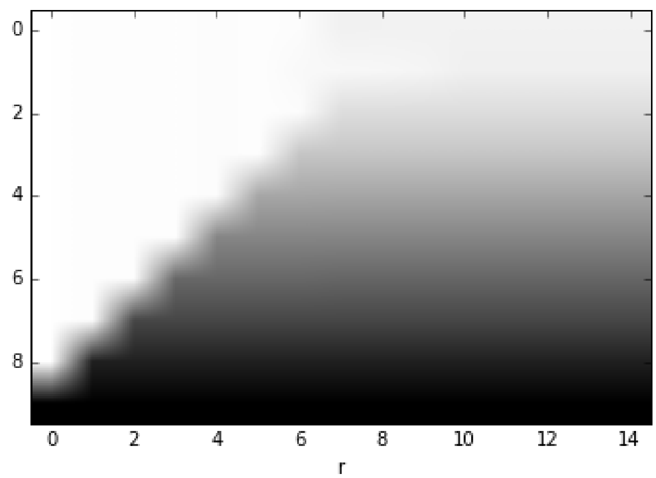

Рис. 5: $T \in R^{15\times15\times15\times15},\ r_1 = r_2 = r_3 = 3$

3.3 Альтернативный аналог

Рассмотрим другой эвристический подход для решения задачи восстановления малорангового тензора. Минимизируем следующий функционал:

$$\min_{r_1 \dots r_{d-1}} \quad n_1 r_1 + r_1 n_2 r_2 + r_2 n_3 r_3 + \dots + r_{d-2} n_{d-1} r_{d-1} + r_{d-1} n_d \tag{19}$$

Заменим $r_i, \ i = \overline{1, d-1}$ ядерной нормой $\| \cdot \|_*$ и получим:

$$\begin{aligned} &\min_{X_1, \dots, X_{d-1}} n_1 \|X_1\|_* + \sum_{k=1}^{d-1} n_{k+1} \|X_k\|_* \|X_{k+1}\|_* \\ &s.t. \quad \mathfrak{X}_\Omega = \mathfrak{T}_\Omega, \end{aligned} \tag{20}$$

Добавляя дополнительные неизвестные к проблеме (20) получим:

$$\begin{aligned} &\min_{\mathfrak{X}, M_k} n_1 \|M_1\|_* + \sum_{k=1}^{d-1} n_{k+1} \|M_k\|_* \|M_{k+1}\|_* + \frac{\beta_k}{2} \|X_k - M_k\|_F^2 \\ &s.t. \quad X_\Omega = T_\Omega \end{aligned} \tag{21}$$

Воспользуемся методом переменных направлений для решения (21):

- **Оптимизация по $\mathbf{M_k}$** , $k = \overline{1, d-1}$:

$$\min_{M_k} \left(n_k \|M_{k-1}\|_* + n_{k+1} \|M_{k+1}\|_*\right) \|M_k\|_* + \frac{\beta_k}{2} \|X_k - M_k\|_F^2 \tag{22}$$

воспользуемся заменой переменных и свойством TT-разложения:

$$\begin{aligned} &\alpha_k = n_k \|M_{k-1}\|_* + n_{k+1} \|M_{k+1}\|_* \\ &\|M_0\|_* = \|M_d\|_* = 1 \end{aligned} \tag{23}$$

подставим (23) в (22):

$$\min_{M_k} \ \alpha_k \|M_k\|_* + \frac{\beta_k}{2}\|X_k - M_k\|_F^2 \tag{24}$$

Решение (24):

$$\hat{M}_k = Shrink_{\frac{\alpha_k}{\beta_k}}(X_k) \tag{25}$$

• **Оптимизация по $\mathfrak{X}$**: Тензор $\mathfrak{X}$ может быть обновлен по следующему правилу:

$$X_{i_1 i_2 \dots i_d} = \begin{cases} \dfrac{\sum\limits_{k=1}^{d-1} \beta_k Fold(M_k)}{\sum\limits_{k=1}^{d-1} \beta_k}, & i_1 i_2 \dots i_d \notin \Omega \\ t_{i_1 i_2 \dots i_d}, & i_1 i_2 \dots i_d \in \Omega \end{cases} \tag{26}$$

4 Заключение

В работе были изучены методы восстановления малоранговых матриц и тензоров по части их элементов. Задача восстановления матрицы была решена с помощью пакета CVXPY, для тензоров реализован алгоритм, представленный в [7]. Вопрос о наилучшем способе обобщения ядерной нормы тензора для того, чтобы эффективно его восстанавливать до сих пор является открытым.

Список литературы

[1] *Emmanuel J Candes, Benjamin Recht* Exact matrix completion via convex optimization // Foundation of Computational mathematics, 2009, pp. 717-772

[2] *Emmanuel J Candes, Terence Tao* The power of convex relaxation: Near-optimal matrix completion// Information Theory, IEEE Transactions on, 2010, pp.2053-2080

[3] *David Gross* Recovering low-rank matrices from few coefficients in a new basis // Information Theory, IEEE Transactions on, 2011, pp. 1548-1566.

[4] *Benjamin Recht* A simpler approach to matrix completion // The Journal of Machine Learning Research, 2011, pp. 3413-3430.

[5] *J. Hastad* Tensor rank is NP-complete // J.of Algorithms 11, 1990, pp. 644–654.

[6] *Cun Mu, Bo Huang, John Wright, Donald Goldfarb* Square Deal: Lower Bounds and Improved Relaxations for Tensor Recovery // arXiv:1307.5870, Aug 2013

[7] *Ho N. Phien, Hoang D. Tuan, Johann A. Bengua and Minh N. Do* Efficient tensor completion - low rank tensor train // arXiv:1601.01083, Jan 2016

[8] *I.V. Oseledets* Tensor Train Decompsition // SIAM J. Sci. Comput., 2011, pp. 2295–2317.