
**ОБЩИЕ
ЧИСЛЕННЫЕ МЕТОДЫ**

УДК 519.6

**ПРИБЛИЖЕННЫЕ АЛГОРИТМЫ МАЛОРАНГОВОЙ
АППРОКСИМАЦИИ В ЗАДАЧЕ ВОСПОЛНЕНИЯ МАТРИЦЫ
НА СЛУЧАЙНОМ ШАБЛОНЕ¹⁾**

© 2021 г. О. С. Лебедева¹, А. И. Осинский^{2,**}, С. В. Петров^{1,*}¹ 119333 Москва, ул. Губкина, 8, ИВМ РАН им. Г.И. Марчука, Россия² 121205 Москва, Большой бульвар, 30, стр. 1, Сколтех, Россия

*e-mail: spetrov.msk@gmail.com

**e-mail: sasha_o@list.ru

Поступила в редакцию 24.11.2020 г.

Переработанный вариант 24.11.2020 г.

Принята к публикации 14.01.2021 г.

Изучается возможность ускорения алгоритма проектирования на старшие сингулярные пространства в задаче “восполнения” матрицы малого ранга по небольшому числу ее элементов. Идея работы состоит в замене процедуры поиска наилучшего приближения во фробениусовой норме на быстрые приближенные алгоритмы. Рассматриваются два метода вычисления таких приближений: (а) проектирование на случайные подпространства; (б) метод крестовой аппроксимации. Доказаны теоремы о геометрической сходимости алгоритмов с приближенными проекциями. Проведены численные эксперименты, показывающие эффективность обоих вариантов по сравнению с точной проекцией. Библиография: 18. Фиг. 4.

Ключевые слова: матрицы малого ранга, восполнение матриц, Singular Value Projection, метод крестовой аппроксимации, случайные подпространства.

DOI: 10.31857/S0044466921050136

1. ВВЕДЕНИЕ

Пусть $X \in \mathbb{R}^{m \times n}$ – неизвестная матрица ранга k , причем $k \ll \min(m, n)$, а Ω – случайный набор пар индексов (i, j) , где $i \in 1, 2, \dots, m$ и $j \in 1, 2, \dots, n$. Задача восстановления всей матрицы X лишь по элементам X_{ij} для $(i, j) \in \Omega$ называется задачей матричного восполнения (Matrix completion problem).

Современные приложения задачи матричного восполнения включают рекомендательные системы (см. [1]), обработку коррелирующих сигналов (см. [2]–[4]), машинное обучение (см. [5], [6]) и многое другое.

Одним из эффективных методов, применяемых для решения задачи восполнения, является алгоритм *проекции на главные сингулярные пространства* (Singular Value Projection method), предложенный в [7], [8] (далее для краткости мы будем называть этот метод SVP).

По своей структуре алгоритм SVP относится к классу методов проективного градиента. Каждая его итерация состоит из двух шагов: шага градиентного спуска и последующей проекции полученного приближения на многообразие матриц ранга k . Сложность одной итерации SVP практически полностью определяется сложностью построения наилучшего приближения с помощью сингулярного разложения (SVD), применяемого к $m \times n$ -матрицам общего вида.

В работе рассматриваются способы ускорения метода SVP за счет замены медленного алгоритма SVD наилучшего приближения матрицами предписанного ранга k на известные быстрые методы построения малоранговых приближений. При этом не предполагается высокой точности получаемых приближений. Напротив, достаточно такой точности, при которой алгоритм восполнения сохраняет свойство глобальной сходимости к искомой матрице X , а возможное увели-

¹⁾Работа поддержана Отделением Московского центра фундаментальной и прикладной математики в ИВМ РАН (Соглашение № 075-15-2019-1624 с Минобрнауки РФ).

чение общего числа итераций компенсируется скоростью построения приближениями малого ранга на каждой итерации. Как будет показано, применение такого подхода во многих случаях приводит к существенному сокращению времени работы всего алгоритма. Семейство алгоритмов, использующих алгоритмы приближенного проектирования будем называть методом ASVP (Approximate Singular Value Projection).

В работе рассматриваются два быстрых алгоритма малоранговых приближений:

- метод проектирования на случайные пространства (см. [9], [10]);
- метод псевдо-крестовой аппроксимации, основанной на принципе большого проективного объема (см. [11]–[13]).

Нашей целью является построение алгоритма ASVP, допускающего теоретическое обоснование, а также получение значимых оценок на его сложность.

Как было отмечено в [4], алгоритм SVP может проявлять нерегулярное поведение. По ходу работы алгоритма гипотезы, используемые в доказательстве его сходимости могут начать нарушаться, даже если они верны для искомой матрицы и исходного приближения; в таких случаях на практике алгоритм расходится. Для восстановления сходимости метода в [4] предложено ограничивать длину шага в градиентном спуске. Последнее замедляет сходимость и увеличивает сложность SVP. В случае метода ASVP нерегулярное поведение имеет даже большее значение. Более того, на практике нерегулярное поведение является типичным! В работе дается анализ причин нерегулярного поведения алгоритмов SVP и ASVP и предлагаются изменения, которые сохраняют сходимость алгоритма без уменьшения скорости сходимости.

2. ТЕОРИЯ СХОДИМОСТИ АЛГОРИТМА SVP

Следуя [7], мы рассматриваем теорию алгоритма SVP в несколько более общей постановке задачи восполнения.

А именно, пусть на множестве $m \times n$ -матриц задан аффинный оператор $\mathcal{A} : \mathbb{R}^{m \times n} \rightarrow \mathbb{R}^M$, где M обозначает число аффинных соотношений на элементах $m \times n$ -матриц. Тогда для произвольного вектора $\mathbf{B} \in \mathbb{R}^M$ задачу малорангового восполнения можно поставить как задачу оптимизации

$$\psi_{\mathcal{A}}(X) = \frac{1}{2} \|\mathcal{A}(X) - \mathbf{B}\|_F^2 \rightarrow \inf, \quad \text{rank}(X) \leq k, \quad (1)$$

с квадратичным функционалом $\psi_{\mathcal{A}}(X)$. Следуя [7], заметим, что градиент $\nabla \psi_{\mathcal{A}}(Y)$ функционала (1) в произвольной точке $Y \in \mathbb{R}^{m \times n}$ может быть записан в виде

$$\nabla \psi_{\mathcal{A}}(Y) = \mathcal{A}^T(\mathcal{A}(Y) - \mathbf{B}),$$

где оператор $\mathcal{A}^T : \mathbb{R}^M \rightarrow \mathbb{R}^{m \times n}$ является транспонированным к \mathcal{A} (везде предполагается, что в пространствах $\mathbb{R}^{m \times n}$ и \mathbb{R}^M задано стандартное скалярное произведение).

Алгоритм SVP состоит из двух шагов: шага градиентного спуска и проекции нового приближения на многообразие матриц ранга не выше k . Другими словами, если X_t – приближение, полученное на шаге t алгоритма, то следующее приближение X_{t+1} получается в виде

$$X_{t+1} = \mathcal{P}_k(X_t - \tau \nabla \psi_{\mathcal{A}}(X_t)) = \mathcal{P}_k(X_t - \tau \mathcal{A}^T(\mathcal{A}(X_t) - \mathbf{B})), \quad (2)$$

где $\tau \in \mathbb{R}$ – некоторый шаг градиентного спуска, который мы определим позже, а \mathcal{P}_k – наилучший во фробениусовой норме проектор на множество матриц ранга не выше k . Для сходимости SVP алгоритма в [7] доказана следующая теорема.

Теорема 1 (см. [7]). Пусть X_* – решение (1), для которого $\psi_{\mathcal{A}}(X_*) = 0$, а оператор \mathcal{A} удовлетворяет условию ограниченной изометрии вида

$$(1 - \sigma) \|X\|_F^2 \leq \|\mathcal{A}(X)\|_2^2 \leq (1 + \sigma) \|X\|_F^2 \quad (3)$$

с параметром $0 < \sigma < 1$ для всех матриц X ранга не выше $2k$. Если дополнительно

$$\frac{2\sigma}{1 - \sigma} < 1,$$

то алгоритм SVP с постоянным шагом $\tau = 1/(1 + \sigma)$ сходится к решению X_* , а скорость сходимости определяется соотношением

$$\psi_{\mathcal{A}}(X_{t+1}) \leq \frac{2\sigma}{1 - \sigma} \psi_{\mathcal{A}}(X_t).$$

Из (2) следует, что алгоритмическая сложность SVP определяется сложностью вычисления наилучшей проекции \mathcal{P}_k на многообразие матриц ранга не выше k . Если $m \approx n$, то сложность такого вычисления, основанного на сингулярном разложении, будет порядка $\mathcal{O}(n^3)$. Цель настоящей работы состоит в исследовании возможности ускорения SVP путем замены оптимального проектора \mathcal{P}_k на быстро вычисляемый приближенный $\hat{\mathcal{P}}_k$.

3. ТЕОРИЯ СХОДИМОСТИ АЛГОРИТМА SVP С ПРИБЛИЖЕННЫМ ВЫЧИСЛЕНИЕМ ПРОЕКЦИЙ (ASVP)

Формализуем понятие приближенного проектирования на многообразии матриц ранга не выше k . Пусть оператор $\hat{\mathcal{P}}_k : \mathbb{R}^{m \times n} \rightarrow \mathbb{R}^{m \times n}$, заданный произвольным образом, всегда возвращает матрицу ранга не выше k . Будем называть $\hat{\mathcal{P}}_k$ оператором приближенного проектирования, если для любой матрицы Y выполнено

$$\|\hat{\mathcal{P}}_k(Y) - Y\|_F^2 \leq (1 + \varepsilon) \|\mathcal{P}_k(Y) - Y\|_F^2 \tag{4}$$

с некоторой константой $\varepsilon > 0$.

По аналогии с алгоритмом SVP определим итерацию приближенного алгоритма ASVP равенством

$$X_{t+1} = \hat{\mathcal{P}}_k(X_t - \tau \mathcal{A}^T(\mathcal{A}(X) - \mathbf{B})).$$

Покажем, что для такого алгоритма ASVP справедлива теорема, аналогичная теореме 1.

Теорема 2. Пусть на матрице X_* ранга не выше k достигается минимум функционала

$$\psi_{\mathcal{A}}(X) = \frac{1}{2} \|\mathcal{A}(X) - \mathbf{B}\|_F^2 \rightarrow \inf, \quad \text{rank}(X) \leq k, \tag{5}$$

с оператором \mathcal{A} , для которого условие ограниченной изометрии

$$(1 - \sigma) \|X\|_F^2 \leq \|\mathcal{A}(X)\|_2^2 \leq (1 + \sigma) \|X\|_F^2 \tag{6}$$

выполняется на всех матрицах ранга не выше $2k$.

В этом случае для алгоритма ASVP с постоянным шагом $1/(1 + \sigma)$ и оператором $\hat{\mathcal{P}}_k$ приближенного проектирования, удовлетворяющим условию (4) с константой ε , справедливо неравенство

$$\begin{aligned} \psi_{\mathcal{A}}(X_{t+1}) &\leq (1 + \varepsilon) \psi_{\mathcal{A}}(X_*) - \varepsilon \psi_{\mathcal{A}}(X_t) + \\ &+ (1 + \varepsilon) \frac{\sigma}{1 - \sigma} \|\mathcal{A}(X_* - X_t)\|_F^2 + \frac{\varepsilon \|\mathcal{A}^T\|^2}{2(1 + \sigma)} \|\mathcal{A}(X_t) - \mathbf{B}\|_F^2. \end{aligned}$$

Доказательство. Используя арифметическое тождество

$$(b - a)^2 - (c - a)^2 = (b - c)^2 + 2(c - a)(b - c)$$

и символьную подстановку

$$a \leftrightarrow \mathbf{B}, \quad b \leftrightarrow \mathcal{A}(X), \quad c \leftrightarrow \mathcal{A}(X_t),$$

для произвольной матрицы X можно записать

$$\begin{aligned} \psi_{\mathcal{A}}(X) - \psi_{\mathcal{A}}(X_t) &= \frac{1}{2} \|\mathcal{A}(X) - \mathbf{B}\|_F^2 - \frac{1}{2} \|\mathcal{A}(X_t) - \mathbf{B}\|_F^2 = \\ &= (\mathcal{A}^T(\mathcal{A}(X_t) - \mathbf{B}), X - X_t) + \frac{1}{2} \|\mathcal{A}(X - X_t)\|_F^2, \end{aligned} \tag{7}$$

где скалярное произведение (U, V) для матриц $U, V \in \mathbb{R}^{m \times n}$ определяется соотношением $(U, V) = \text{tr}(V^T U)$. Для произвольной матрицы X ранга не выше k воспользуемся во втором слагаемом (7) свойством ограниченной изометрии оператора \mathcal{A} и установим верхнюю оценку для разности:

$$\begin{aligned} \psi_{\mathcal{A}}(X) - \psi_{\mathcal{A}}(X_t) &= (\mathcal{A}^T(\mathcal{A}(X_t) - \mathbf{B}), X - X_t) + \frac{1}{2} \|\mathcal{A}(X - X_t)\|_F^2 \leq \\ &\leq (\mathcal{A}^T(\mathcal{A}(X_t) - \mathbf{B}), X - X_t) + \frac{1+\sigma}{2} \|X - X_t\|_F^2. \end{aligned} \quad (8)$$

Правую часть (8) обозначим через $f_t(X)$. Таким образом,

$$f_t(X) = (\mathcal{A}^T(\mathcal{A}(X_t) - \mathbf{B}), X - X_t) + \frac{(1+\sigma)}{2} \|X - X_t\|_F^2 \quad (9)$$

и

$$\psi_{\mathcal{A}}(X) - \psi_{\mathcal{A}}(X_t) \leq f_t(X) \quad (10)$$

для всех матриц X , ранг которых не превосходит k .

Значения $f_t(X)$ дают оценку на падение функционала невязки на всем множестве матриц ранга не выше k , а вид $f_t(X)$ позволяет найти в явном виде матрицу ранга не выше k , на которой падение невязки является значительным.

Действительно, используем арифметическое тождество

$$\frac{1+\sigma}{2} (2pq + p^2) = \frac{1+\sigma}{2} [(p+q)^2 - q^2]$$

вместе с соответствием

$$p \leftrightarrow X - X_t, \quad q \leftrightarrow \frac{1}{1+\sigma} \mathcal{A}^T(\mathcal{A}(X_t) - \mathbf{B}), \quad f_t(X) \leftrightarrow \frac{1+\sigma}{2} (2pq + p^2).$$

После замен $f_t(X)$ принимает вид

$$f_t(X) = \frac{1+\sigma}{2} \left\| X - \left(X_t - \frac{1}{1+\sigma} \mathcal{A}^T(\mathcal{A}(X_t) - \mathbf{B}) \right) \right\|_F^2 - \frac{1}{2(1+\sigma)} \left\| \mathcal{A}^T(\mathcal{A}(X_t) - \mathbf{B}) \right\|_F^2.$$

Определяя Y_{t+1} равенством

$$Y_{t+1} = X_t - \frac{1}{1+\sigma} \mathcal{A}^T(\mathcal{A}(X_t) - \mathbf{B}),$$

представим $f_t(X)$ в виде суммы:

$$f_t(X) = \frac{1+\sigma}{2} \|X - Y_{t+1}\|_F^2 - \frac{1}{2(1+\sigma)} \left\| \mathcal{A}^T(\mathcal{A}(X_t) - \mathbf{B}) \right\|_F^2. \quad (11)$$

Так как второе слагаемое в (11) не зависит от X , то минимум $f_t(X)$ на множестве всех матриц ранга не выше k достигается на матрице $Z_{t+1} = \mathcal{P}_k(Y_{t+1})$. При этом сама матрица Y_{t+1} формально совпадает с матрицей, полученной градиентным спуском из точки X_t с величиной шага $\tau = 1/(1+\sigma)$, не зависящей от его номера t .

Видно, что Z_{t+1} совпадает с приближением на шаге $t+1$ в алгоритме SVP.

Так как Z_{t+1} дает минимум $f_t(X)$, то значение $f_t(Z_{t+1})$ можно оценить сверху значением $f_t(X_*)$:

$$f_t(Z_{t+1}) \leq f_t(X_*) = (\mathcal{A}^T(\mathcal{A}(X_t) - \mathbf{B}), X_* - X_t) + \frac{1+\sigma}{2} \|X_* - X_t\|_F^2.$$

Используя свойство ограниченной изометрии оператора \mathcal{A} на матрице $(X_* - X_t)$, ранг которой очевидным образом не превосходит $2k$, преобразуем последнее выражение к виду

$$f_t(Z_{t+1}) \leq (\mathcal{A}^T(\mathcal{A}(X_t) - \mathbf{B}), X_* - X_t) + \frac{1+\sigma}{2(1-\sigma)} \left\| \mathcal{A}(X_* - X_t) \right\|_F^2.$$

Теперь воспользуемся формулой (7) для первого слагаемого, чтобы получить

$$f_i(Z_{t+1}) \leq \psi_{\mathcal{A}}(X_*) - \psi_{\mathcal{A}}(X_t) + \frac{\sigma}{1 - \sigma} \|\mathcal{A}(X_* - X_t)\|_F^2. \quad (12)$$

Положим $X_{t+1} = \hat{\mathcal{P}}_k(Y_{t+1})$. В силу определения $\hat{\mathcal{P}}_k$ выполняется неравенство

$$\|X_{t+1} - Y_{t+1}\|_F^2 \leq (1 + \varepsilon) \|Z_{t+1} - Y_{t+1}\|_F^2. \quad (13)$$

Рассмотрим разность $f_i(X_{t+1}) - f_i(Z_{t+1})$. Подставляя для $f_i(X_{t+1})$ и $f_i(Z_{t+1})$ их представление из (11), сокращая постоянное слагаемое и учитывая (13), получаем

$$f_i(X_{t+1}) - f_i(Z_{t+1}) = \frac{1 + \sigma}{2} (\|X_{t+1} - Y_{t+1}\|_F^2 - \|Z_{t+1} - Y_{t+1}\|_F^2) \leq \varepsilon \frac{1 + \sigma}{2} \|Z_{t+1} - Y_{t+1}\|_F^2. \quad (14)$$

Еще раз применим (11) теперь к правой части (14). Тогда

$$\begin{aligned} f_i(X_{t+1}) - f_i(Z_{t+1}) &\leq \varepsilon f_i(Z_{t+1}) + \varepsilon \frac{1}{2(1 + \sigma)} \|\mathcal{A}^T(\mathcal{A}(X_t) - \mathbf{B})\|_F^2 \leq \\ &\leq \varepsilon f_i(Z_{t+1}) + \varepsilon \frac{\|\mathcal{A}^T\|^2}{2(1 + \sigma)} \|\mathcal{A}(X_t) - \mathbf{B}\|_F^2, \end{aligned}$$

где $\|\mathcal{A}^T\|$ — операторная норма \mathcal{A}^T , подчиненная второй норме в пространстве \mathbb{R}^M и фробениусовой норме в пространстве $\mathbb{R}^{m \times n}$. Теперь мы готовы оценить разность $\psi(X_{t+1}) - \psi(X_t)$. Действительно, как следует из (10),

$$\begin{aligned} \psi_{\mathcal{A}}(X_{t+1}) - \psi_{\mathcal{A}}(X_t) &\leq f_t(X_{t+1}) = f_t(X_{t+1}) - (f_t(Z_{t+1}) - f_t(Z_{t+1})) = \\ &= (f_t(X_{t+1}) - f_t(Z_{t+1})) + f_t(Z_{t+1}) \leq (1 + \varepsilon) f_t(Z_{t+1}) + \varepsilon \frac{\|\mathcal{A}^T\|^2}{2(1 + \sigma)} \|\mathcal{A}(X_t) - \mathbf{B}\|_F^2. \end{aligned}$$

Заменяя $f_t(Z_{t+1})$ выражением (12), преобразуем последнее соотношение к окончательному виду

$$\begin{aligned} \psi_{\mathcal{A}}(X_{t+1}) - \psi_{\mathcal{A}}(X_t) &\leq (1 + \varepsilon) (\psi_{\mathcal{A}}(X_*) - \psi_{\mathcal{A}}(X_t)) + \\ &+ (1 + \varepsilon) \frac{\sigma}{1 - \sigma} \|\mathcal{A}(X_* - X_t)\|_F^2 + \varepsilon \frac{\|\mathcal{A}^T\|^2}{2(1 + \sigma)} \|\mathcal{A}(X_t) - \mathbf{B}\|_F^2. \end{aligned}$$

Откуда после сокращения слагаемого $\psi(X_t)$ в левой и правой частях неравенства заканчиваем доказательство.

Следствие 1. Пусть существует матрица X_* ранга не выше k , для которой $\psi_{\mathcal{A}}(X_*) = 0$ и выполняется условие ограниченной изометрии (3). Алгоритм ASVP с постоянным шагом $\tau = \frac{1}{1 + \sigma}$ сходится к решению X_* , а скорость сходимости определяется соотношением

$$\psi_{\mathcal{A}}(X_{t+1}) \leq \psi_{\mathcal{A}}(X_t) \left(\frac{2\sigma}{1 - \sigma} + \varepsilon \frac{\|\mathcal{A}^T\|^2}{1 + \sigma} \right). \quad (15)$$

Доказательство. Из предыдущей теоремы и в силу соотношений

$$\begin{aligned} \psi_{\mathcal{A}}(X_*) &= 0, \\ \mathcal{A}(X_*) &= \mathbf{B}, \\ \psi_{\mathcal{A}}(X_t) &= \frac{1}{2} \|\mathcal{A}(X_t - X_*)\|_F^2 = \frac{1}{2} \|\mathcal{A}(X_t) - \mathbf{B}\|_F^2 \end{aligned}$$

имеем

$$\begin{aligned} \psi_{\mathcal{A}}(X_{t+1}) &\leq \left(\frac{2\sigma(1+\varepsilon)}{1-\sigma} + \varepsilon \frac{\|\mathcal{A}^T\|^2}{1+\sigma} - \varepsilon \right) \psi_{\mathcal{A}}(X_t) = \\ &= \left[\frac{2\sigma}{1-\sigma} + \varepsilon \left(\frac{2\sigma}{1-\sigma} + \frac{\|\mathcal{A}^T\|^2}{1+\sigma} - 1 \right) \right] \psi_{\mathcal{A}}(X_t). \end{aligned} \quad (16)$$

В предположении $\sigma < 1/3$, аналогичном тому, что используется для доказательства сходимости SVP в [7], видим, что ε можно выбрать достаточно малой константой, чтобы из неравенства (16) следовала геометрическая сходимость ASVP.

Как следует из формулы (15), использование приближенного проектора $\hat{\mathcal{P}}_k$ в алгоритме ASVP ухудшает коэффициент линейной сходимости, что, конечно, не является неожиданным. Однако, как будет показано далее, выигрыш, получаемый за счет упрощения \mathcal{P}_k , превосходит проигрыш, связанный с увеличением числа итераций в ASVP.

В следствии 1 предполагается существование решения X_* , для которого $\psi_{\mathcal{A}}(X_*) = 0$. Однако на практике для заданного вектора $\mathbf{B} \in \mathbb{R}^M$ минимум функционала может быть не равен нулю, например, в случае $\mathbf{B} = \mathcal{A}(X) + \Delta$, где X – матрица малого ранга, а Δ имеет малую норму. Оценки для такого случая представлены в следствии 2.

Следствие 2. Пусть на матрице X_* достигается минимум функционала $\psi_{\mathcal{A}}(X)$ среди всех матриц ранга не выше k , а $\Delta = \mathbf{B} - \mathcal{A}(X_*)$ имеет норму $\delta = \|\Delta\|_2$.

Если для приближения X_t выполняется неравенство $\psi_{\mathcal{A}}(X_t) \geq C^2 \frac{\delta^2}{2}$ с константой $C > 0$, то для следующего приближения, получаемого в методе ASVP, справедливо соотношение

$$\psi_{\mathcal{A}}(X_{t+1}) < \kappa \psi_{\mathcal{A}}(X_t)$$

с константой κ , удовлетворяющей неравенству

$$\kappa \leq D_0 + \varepsilon(D_0 + D_1), \quad (17)$$

где

$$D_0 = \frac{1}{C^2} + \frac{2\sigma}{1-\sigma} \left(1 + \frac{1}{C}\right)^2,$$

$$D_1 = \frac{\|\mathcal{A}^T\|^2}{1+\sigma} - 1.$$

Доказательство. Из теоремы 2

$$\psi_{\mathcal{A}}(X_{t+1}) \leq (1+\varepsilon) \frac{\delta^2}{2} + (1+\varepsilon) \frac{\sigma}{1-\sigma} \|\mathbf{B} - \mathcal{A}(X_t) - \Delta\|_F^2 + \varepsilon \left(\frac{\|\mathcal{A}^T\|^2}{1+\sigma} - 1 \right) \psi_{\mathcal{A}}(X_t).$$

Учитывая, что $\delta^2 \leq \frac{2}{C^2} \psi_{\mathcal{A}}(X_t)$, и раскрывая второе слагаемое, получаем

$$\begin{aligned} \psi_{\mathcal{A}}(X_{t+1}) &\leq \frac{(1+\varepsilon)}{C^2} \psi_{\mathcal{A}}(X_t) + (1+\varepsilon) \frac{2\sigma}{1-\sigma} \left(\psi_{\mathcal{A}}(X_t) + \frac{2}{C} \psi_{\mathcal{A}}(X_t) + \frac{1}{C^2} \psi_{\mathcal{A}}(X_t) \right) + \\ &+ \varepsilon \left(\frac{\|\mathcal{A}^T\|^2}{1+\sigma} - 1 \right) \psi_{\mathcal{A}}(X_t) \leq \left[\frac{1}{C^2} + \frac{2\sigma}{1-\sigma} \left(1 + \frac{1}{C}\right)^2 \right] \psi_{\mathcal{A}}(X_t) + \\ &+ \varepsilon \left[\frac{1}{C^2} + \frac{2\sigma}{1-\sigma} \left(1 + \frac{1}{C}\right)^2 + \frac{\|\mathcal{A}^T\|^2}{1+\sigma} - 1 \right] \psi_{\mathcal{A}}(X_t). \end{aligned}$$

Полагая

$$D_0 = \frac{1}{C^2} + \frac{2\sigma}{1-\sigma} \left(1 + \frac{1}{C}\right)^2 \quad \text{и} \quad D_1 = \frac{\|\mathcal{A}^\Gamma\|^2}{1+\sigma} - 1,$$

окончательно получаем

$$\Psi_{\mathcal{A}}(X_{t+1}) \leq [D_0 + \varepsilon(D_0 + D_1)]\Psi_{\mathcal{A}}(X_t). \tag{18}$$

4. ПРИМЕНЕНИЕ АЛГОРИТМОВ SVP И ASVP К ЗАДАЧЕ ВОСПОЛНЕНИЯ МАТРИЦ МАЛОГО РАНГА

Рассмотрим применение алгоритма ASVP к задаче восполнения матриц малого ранга по случайному набору элементов (на случайном шаблоне). С точки зрения анализа сходимости алгоритмов SVP и ASVP задача восполнения сводится к специальному способу выбора аффинного преобразования \mathcal{A} .

Пусть $X_* \in \mathbb{R}^{m \times n}$ – неизвестная матрица ранга k , причем $k \ll \min(m, n)$, а Ω – случайный набор пар индексов (i, j) , где $i \in 1, 2, \dots, m$ и $j \in 1, 2, \dots, n$, причем любая пара (i, j) может быть выбрана равновероятно среди всех mn пар с вероятностью q , которую мы определим позже. Обозначим мощность множества Ω как $M = |\Omega|$, будем считать, что пары $(i, j) \in \Omega$ линейно упорядочены; будем также для простоты считать, что выполнено в точности $M = qmn$, а q будем также называть плотностью известных элементов матрицы.

Определим линейный оператор $\mathcal{S} : \mathbb{R}^{m \times n} \rightarrow \mathbb{R}^M$ по правилу

$$(\mathcal{S}(X))_k = \frac{1}{\sqrt{q}} X_{ij}, \quad k = 1, 2, \dots, M, \tag{19}$$

где $(\mathcal{S}(X))_k$ – компонента вектора $\mathcal{S}(X)$, соответствующая паре (i, j) в выбранном линейном упорядочении. Определим B формулой $B = \mathcal{S}(X_*) \in \mathbb{R}^M$ и зададим функционал $\Psi_{\mathcal{S}}(X) = \frac{1}{2} \|\mathcal{S}(X) - B\|_2^2$. Задача восполнения матрицы X_* по части ее элементов записывается в виде

$$\Psi_{\mathcal{S}}(X) = \frac{1}{2} \|\mathcal{S}(X) - B\|_2^2 \rightarrow \inf, \quad \text{rank}(X) \leq k. \tag{20}$$

Для упрощения записи в дальнейшем индекс \mathcal{S} в обозначении функционала Ψ опускается.

В качестве оператора $\hat{\mathcal{P}}_k$ при этом можно использовать произвольный алгоритм построения малоранговой аппроксимации $\tilde{A} = \hat{\mathcal{P}}_k A$, достаточно близкой к проекции с использованием сингулярного разложения

$$\|A - \tilde{A}\|_F \leq (1 + \varepsilon) \|A - A_k\|_F.$$

Нижний индекс k у матрицы здесь и далее обозначает наилучшее приближение ранга k по норме Фробениуса, которое можно получить с помощью сингулярного разложения.

При $\sigma \approx 0$ из следствия 2 подстановкой $\mathcal{A} = \mathcal{S}$, используя тривиальную оценку $\|\mathcal{S}^\Gamma\|^2 \leq \frac{1}{q}$, мы получаем условие

$$\varepsilon \|\mathcal{S}^\Gamma\|^2 < 1, \quad \frac{\varepsilon}{q} < 1,$$

а потому для сходимости метода восполнения матриц достаточно $\varepsilon = O(q)$.

Отметим, что для задачи восполнения матрицы малого ранга по значениям ее элементов на разреженном шаблоне Ω напрямую воспользоваться результатами разд. 2 и 3 нельзя.

Так, например, в случае произвольного шаблона Ω оператор $\mathcal{S}(X)$, определенный ранее, не удовлетворяет условию ограниченной изометрии, которое является ключевым при доказательстве сходимости метода. Тем не менее, как показано в [7], если Ω имеет случайный характер, то условие ограниченной изометрии для оператора \mathcal{S} выполняется с вероятностью, неотличимой от 1, на почти всем множестве матриц ранга, не превосходящего $2k$. Чтобы сформулировать соответствующее утверждение, нам потребуется понятие μ -некогерентных матриц.

Определение 1. Пусть для матрицы $X \in \mathbb{R}^{m \times n}$ задано ее сингулярное разложение $X = U\Sigma V^T$. Матрица X называется μ -некогерентной, если для матриц сингулярных векторов U и V справедливы неравенства

$$\max_{i,j} |U_{ij}| \leq \frac{\sqrt{\mu}}{\sqrt{m}}, \quad \max_{i,j} |V_{ij}| \leq \frac{\sqrt{\mu}}{\sqrt{n}}.$$

Используя определение 1, сформулируем следующий результат о выполнении свойства ограниченной изометрии для оператора \mathcal{S} .

Теорема 3 [Теорема 4.2 из [7]]. Существует константа $C \geq 0$ такая, что для любого $0 < \sigma < 1$, любых $\mu \geq 1$ и $n \geq t \geq 3$, и для любого шаблона разреженности Ω , пары индексов которого выбираются случайно в соответствии с моделью Бернулли с параметром

$$q \geq C \frac{\mu^2 k^2 \log(n)}{\sigma^2 t},$$

определяющим вероятность для пары индексов (i, j) принадлежать маске Ω , оператор \mathcal{S} удовлетворяет на множестве всех μ -некогерентных матриц свойству ограниченной изометрии вида

$$(1 - \sigma) \|X\|_F^2 \leq \|\mathcal{S}(X)\|_F^2 \leq (1 + \sigma) \|X\|_F^2$$

с вероятностью не меньше $1 - \exp(-n \log(n))$.

Из теорем 2 и 3 следует сходимость алгоритмов SVP и ASVP в случае, если на всех итерациях (т.е. для всех t) матрицы $X_{t+1} - X_t$ и $X_t - X_*$ обладают ограниченной константой некогерентности μ . Отметим, что для разности $X_{t+1} - X_t$ формальная проверка μ -некогерентности является возможной и вычислительно недорогой процедурой. В то же время для разности $X_t - X_*$ на практике оценить константу некогерентности невозможно.

Из сказанного выше следует, что для задачи восполнения матриц малого ранга, заданных на случайном шаблоне, обоснование алгоритмов SVP и ASVP не является безусловным. Формальные реализации этих методов могут и проявляют нерегулярное поведение, т.е. расходятся при выборе теоретически обоснованного шага градиентного спуска τ (см., например, [4]) или в том случае, когда сингулярные числа решения X_* быстро убывают.

На практике получение (быстро) сходящихся реализаций алгоритмов возможно только с использованием дополнительных приемов, среди которых мы выделяем два: последовательный набор ранга приближения и выбор шага τ в градиентном методе.

4.1. Последовательный набор ранга приближения

Этот прием необходим в ситуациях, когда сингулярные числа решения X_* задачи восполнения быстро убывают, или когда ранг X_* заранее неизвестен. Причиной нерегулярного поведения алгоритмов SVP и ASVP алгоритмов в этом случае является существование решений задачи восполнения с большим рангом и обладающих большой константой некогерентности.

Рассмотрим искусственный, но проясняющий ситуацию пример. Пусть $E \in \mathbb{R}^{m \times n}$ — матрица ранга 1, каждый элемент которой равен единице. Такая матрица является μ -некогерентной с параметром некогерентности $\mu = 1$. Зададим $M \ll \min(m, n)$ пар индексов (i, j) и поставим задачу восполнения матрицы E среди матриц ранга M . Среди возможных решений этой задачи существует разреженное решение ранга M , все исходно неизвестные компоненты которого равны нулю. Однако такое решение имеет константу некогерентности порядка $\max(\sqrt{m}, \sqrt{n})$.

Чтобы избежать нерегулярного поведения ASVP в рассматриваемых ситуациях, предлагается использовать последовательный набор ранга приближений. Другими словами, вместо того, чтобы решать задачу восполнения сразу для предписанного ранга k , задачу восполнения предлагается последовательно решать для рангов $1, 2, \dots, k$, и для каждого последующего ранга решение, соответствующее предыдущему рангу, использовать в качестве начального приближения.

Наиболее примитивный подход к набору ранга решения состоит в том, чтобы заранее фиксировать время работы алгоритма и отвести константные отрезки времени на работу алгоритма с

каждым рангом приближения k . При этом можно, например, отвести первую половину всего времени работы алгоритма на стадию набора ранга, а вторую – на итерации с необходимым конечным рангом приближения.

Другой более обоснованный способ набора ранга может быть основан на проверке выполнения условия μ -некогерентности для разности $X_t - X_{t-1}$. Анализ численных экспериментов показал, что при нерегулярном поведении алгоритмов SVP и ASVP параметр μ -некогерентности последовательности матриц $X_{t+1} - X_t$ увеличивается от итерации к итерации и стремится к максимально возможному значению $\max(\sqrt{m}, \sqrt{n})$. Таким образом, на каждой итерации можно вычислять текущее значение некогерентности $X_{t+1} - X_t$ и проводить увеличение ранга приближения на единицу только в случае, если на протяжении нескольких итераций эта величина оказывается достаточно малой; для определения допустимой величины некогерентности полезно иметь априорную оценку на эту величину для искомой матрицы.

Сложность вычисления μ для матрицы $X_t - X_{t-1}$ определяется сложностью вычисления сингулярного разложения объединенных левых и объединенных правых факторов матриц X_t, X_{t-1} , что в случае $m = n$ требует $O(nk^2) \ll n^3$ операций.

При этом точно указать необходимое число последовательных шагов алгоритма, при которых величина некогерентности $X_t - X_{t-1}$ остается ниже допустимой границы, затруднительно; на основе проведенных экспериментов установлено, что в случае матриц порядка 1000 и случайных факторов достаточно около пяти шагов. Это число последовательных шагов можно оценивать по числу предшествующих шагов, при которых величины некогерентности были, наоборот, большими, например, используя следующую процедуру.

Алгоритм 1

Входные данные: граничное значение μ_{crit} , по которому определяется, что некогерентность на текущей итерации достаточно мала. Константа минимального числа последовательных итераций с низкой некогерентностью s_{crit} . Вводятся переменные, сохраняющие число предшествующих итераций SVP (ASVP) с большими и малыми величинами μ : $s_{\text{high}} = 0, s_{\text{low}} = 0$.

Цель: ранг k увеличивается, если на протяжении нескольких итераций $\mu < \mu_{\text{crit}}$.

1: если $\mu(X_t - X_{t-1}) > \mu_{\text{crit}}$, то

2: $s_{\text{high}} := s_{\text{high}} + 1$

3: $s_{\text{crit}} := \max(s_{\text{high}}, s_{\text{crit}})$

4: иначе

5: $s_{\text{low}} := s_{\text{low}} + 1$

6: если $s_{\text{low}} > s_{\text{crit}}$, то

7: $k := k + 1$

8: $s_{\text{low}} := 0$

9: $s_{\text{high}} := 0$

Такой способ набора ранга оказался особенно эффективен для варианта ASVP, основанного на псевдоскелетном методе с использованием подматриц большого объема (см. далее в п. 5.2).

4.2. Выбор параметра шага градиентного метода

Численные эксперименты показывают, что нерегулярного поведения алгоритмов SVP, ASVP можно избежать путем понижения параметра шага градиентного метода τ . Так, теорема 2 гарантирует, что в условиях теоремы метод SVP сходится геометрически при $\tau = 1/(1 + \sigma) \approx 3/4$ с учетом того, что, согласно определению оператора задачи восполнения матриц \mathcal{S} , все элементы на маске Ω скалируются на константу $1/\sqrt{q}$, использование такого шага интуитивно неустойчиво при критически малых значениях плотности известных элементов q . На практике оказывается, что при таком шаге в случае малых q и быстром падении сингулярных чисел искомой матрицы алгоритмы SVP, ASVP могут расходиться, но при тех же условиях и малом шаге порядка q расхо-

димости не наблюдается. Экспериментально установлено, что при шаге τ в пределах от $[q, 2q]$ при достаточно малых q сходимость SVP, ASVP наблюдается всегда.

Случаи расходимости SVP метода с большим шагом были замечены уже в [4]. Для повышения устойчивости алгоритма ее авторы также предложили выбирать τ из интервала $[q, 2q]$. Не проводя формального доказательства сходимости алгоритма SVP с шагом $\tau = q$, приведем некоторые соображения в пользу его устойчивости.

Действительно, если X_t и X_{t+1} — приближения с номерами t и $t + 1$ и $B = S(X_*)$, то шаг градиентного метода $\tau = q$ соответствует присваиванию каждому элементу текущего приближения, лежащему на маске известных элементов, соответствующего известного элемента. Это значит, что

$$\|X_t - X_*\|_{F,\Omega} = \|X_t - Y_{t+1}\|_F \geq \|Y_{t+1} - X_{t+1}\|_F \geq \|X_{t+1} - X_*\|_{F,\Omega}, \quad (21)$$

где $\|X\|_{F,\Omega} = \|\mathcal{P}(X)\|_F$. Неравенство (21) можно записать в эквивалентной форме:

$$\psi(X_{t+1}) \leq \psi(X_t),$$

откуда имеем, что невязка на итерациях SVP алгоритма с шагом $\tau = q$ монотонно не возрастает.

Тем не менее эксперименты показывают, что в случаях, когда при использовании шага $\tau \approx 3/4$ на практике не нарушаются гипотезы о некогерентности, сходимость алгоритмов SVP, ASVP с шагом $\tau \approx 3/4$ значительно быстрее, чем с шагом $\tau = q$, т.е. на практике использование большого шага является предпочтительным. Чтобы на практике использовать преимущество большого шага в скорости сходимости, но не допускать нерегулярного поведения алгоритмов, предлагается использовать адаптивные процедуры изменения шага τ по ходу алгоритма. Для этого предлагается рассматривать изменение невязки $\|\mathcal{P}(X_t - X_*)\|_F$ между итерациями SVP, ASVP. Если отношение невязок на итерациях $t + 1$ и t

$$\alpha_{t+1,t} := \frac{\|\mathcal{P}(X_{t+1} - X_*)\|_2}{\|\mathcal{P}(X_t - X_*)\|_2} > \alpha_{\text{crit}} > 1,$$

то делается вывод, что алгоритм начал вести себя нерегулярно, в связи с чем шаг градиентного метода понижается, а последняя итерация пересчитывается с уменьшенным шагом. Формально после выполнения итерации $t + 1$ алгоритмов SVP, ASVP выполняется следующий

Алгоритм 2

Входные данные: пограничная величина $\alpha_{\text{crit}} > 1$; стартовое значение шага $\tau = 3/4$; константы увеличения и уменьшения шага $\beta_{\text{inc}} < 1$, $\beta_{\text{dec}} < 1$.

Цель: шаг τ изменяется в зависимости от роста или падения погрешности.

1: если $(\alpha_{t+1,t} > \alpha_{\text{crit}})$, то

2: $\tau := \tau + \beta_{\text{dec}}(q - \tau)$

3: Итерация $t + 1$ проводится заново с уменьшенным шагом.

4: иначе

5: $\tau := \tau + \beta_{\text{inc}}\left(\frac{3}{4} - \tau\right)$.

Этот адаптивный алгоритм выбора шага оказался эффективен на практике для ASVP с обоими вариантами метода приближенного проектирования, которые будут рассмотрены далее. Однако вычисление невязки может быть асимптотически сложной процедурой, так как требует не менее $O(mnqk)$ операций, поэтому возможно использовать эту процедуру обновления величины шага не после каждой итерации, а например, после каждой десятой итерации ASVP.

5. ВОЗМОЖНЫЕ МЕТОДЫ ПРИБЛИЖЕННОГО ПРОЕКТИРОВАНИЯ ASVP

5.1. Проектирование на случайные подпространства

Рассмотрим один из возможных методов случайного проектирования, необходимых для построения алгоритма ASVP. Будем использовать способ приближенного вычисления частичного

сингулярного разложения на основе проектирования на случайные подпространства. Такой способ введен и исследован в [9]. Будем использовать вспомогательную матрицу

$$J \in \mathbb{R}^{n \times l}, \quad l = k + p, \quad p > 0,$$

все элементы которой выбираются независимо и случайно по стандартному нормальному распределению. Тогда в качестве проектора $\hat{\mathcal{P}}$ предлагается использовать

$$\hat{\mathcal{P}}_k(Y) = Q(Q^T Y)_k, \quad Y \in \mathbb{R}^{m \times n},$$

где матрица Q определяется как ортогональный базис столбцов YJ : $YJ = QR$, а $(Q^T Y)_k$ – наилучшее приближение матрицы $(Q^T Y)$ матрицей ранга k , вычисляемое с помощью сингулярного разложения вытянутой матрицы. Из леммы 6.1 в [18], которая обобщает результаты из [9], следует, что для любой матрицы $Y \in \mathbb{R}^{m \times n}$ справедлива

Лемма 1 (см. [18]):

$$\mathbb{E}_J \|Y - Q(Q^T Y)_k\|_F^2 \leq \left(1 + \frac{k}{p-1}\right) \|Y - Y_k\|_F^2. \quad (22)$$

В [9], [18] приведены оценки на отклонение от матожидания, которые опустим в силу их громоздкости. Таким образом, в [9] предложен следующий алгоритм вычисления приближенного частичного сингулярного разложения матрицы Y .

1. Создается случайная матрица $J \in \mathbb{R}^{n \times l}$.
2. Вычисляется произведение YJ .
3. Для произведения YJ вычисляется QR -разложение: $YJ = QR$.
4. Вычисляется произведение $Q^T Y$.
5. Вычисляется сокращенное сингулярное разложение для вытянутой матрицы

$$(Q^T Y)_k = U \Sigma V^T, \quad \text{где } U \in \mathbb{R}^{l \times k}, \quad V^T \in \mathbb{R}^{k \times n}.$$

6. Так как

$$(QU) \Sigma V^T = Q(Q^T Y)_k,$$

то матрицы QU и V^T являются искомым приближением сингулярных векторов матрицы Y .

Согласно оценке (22), введенный таким образом оператор $\hat{\mathcal{P}}$ удовлетворяет определению оператора приближенного проектирования (4) с точностью $\epsilon = k/(p-1)$ в среднем. Оценим сложность одной итерации полученного приближенного алгоритма ASVP.

- Умножение на случайную матрицу: так как приближенное сингулярное разложение вычисляется для суммы матрицы малого ранга (матрица предыдущей итерации) и разреженной матрицы (градиент), имеем $O((n+m)kl + mnlq)$ операций.

- QR -разложение YJ : $O(ml^2)$ операций.

- Умножение на матрицу Q^* слева: аналогично первому пункту, $O((n+m)kl + mnlq)$ операций.

- Сингулярное разложение вытянутой матрицы: $O(nl^2)$ операций.

- Восстановление левого сингулярного базиса: $O(ml^2)$ операций.

Таким образом, имеем суммарную сложность одной итерации приближенного алгоритма SVP $O((m+n)l^2 + mnlq)$. С учетом оценки $\epsilon = O\left(\frac{k}{p}\right)$ можно считать, что при малых ϵ выполнено $p \gg k, l \approx p$, а вычислительная сложность одной итерации соответственно оценивается величиной

$$O((m+n)p^2 + mnpq).$$

5.2. Псевдоскелетный метод на матрицах большого проективного объема

Опишем метод ASVP с использованием в качестве проектора $\hat{\mathcal{P}}_k$ аппроксимации на основе крестового (псевдоскелетного) CGR-приближения, в котором используются строки $R \in \mathbb{R}^{p \times n}$ и столбцы $C \in \mathbb{R}^{m \times p}$ приближаемой матрицы

$$Y = X_t - \tau \mathcal{P}^T(\mathcal{G}(X_t) - \mathbf{B}).$$

Один из самых быстрых способов построения аппроксимации – построение Fast CGR из [17]. Построение аппроксимации состоит из следующих этапов.

1. Использование алгоритма `maxvol` (см. [16]) для поиска подматрицы $\hat{Y} \in \mathbb{R}^{2k \times 2k}$.

Увеличенный в 2 раза ранг позволяет позже использовать усеченное сингулярное разложение для увеличения точности аппроксимации (по сравнению с прямым поиском аппроксимации ранга k), а также позволяет алгоритму не “зацикливаться” на одной и той же подматрице.

Эксперименты в [17] показывают, что в случае прямого построения аппроксимации $\tilde{Y} = C\hat{Y}^{-1}R$, коэффициент ϵ растет линейно с ростом r , поэтому нельзя гарантировать произвольную точность приближения ϵ , ограничиваясь лишь алгоритмом `maxvol`.

При фиксированном числе шагов алгоритм требует $O((m+n)k^2)$ операций и знания $O(k)$ строк и столбцов матрицы.

2. Увеличение числа строк и столбцов до p .

Новые строки и столбцы выбираются с помощью ускоренного варианта алгоритма `maxvol2` (см. [15]). Расширение требует лишь знания уже найденных с помощью `maxvol` строк и столбцов и занимает $O((m+n)kp)$ операций.

Согласно гипотезе из [17], для аппроксимации ранга $2k$ с p столбцами справедлива оценка коэффициента

$$1 + \epsilon \leq 1 + \frac{k}{p - k + 1}, \quad (23)$$

а потому достаточно набрать $p = O(k/\epsilon)$ строк и столбцов. При этом в качестве ядра CGR-аппроксимации выбирается матрица $G = \hat{Y}_{2k}^+$, для чего потребуется сингулярное разложение матрицы \hat{A} , занимающее $O(p^3)$ операций.

3. Сингулярное разложение аппроксимации.

Найденная на предыдущем шаге CGR-аппроксимация ранга $2k$ подвергается сингулярному разложению с сохранением лишь k максимальных сингулярных чисел.

Сингулярное разложение произведения $C\hat{Y}_{2k}^+R$ требует $O((m+n)kp)$ операций и выполняется в следующем порядке.

(а) Факторы $\hat{U} \in \mathbb{R}^{p \times 2k}$ и $\hat{V}^T \in \mathbb{R}^{2k \times p}$ сингулярного разложения матрицы \hat{Y}_{2k}^+ умножаются на C и R соответственно.

(б) Для матрицы $C\hat{U}$ вычисляется представление в виде произведения матрицы с ортогональными столбцами и верхней треугольной матрицы; для матрицы $\hat{V}^T R$ вычисляется представление в виде произведения нижней треугольной матрицы и матрицы с ортогональными строками.

(в) Выполняется сингулярное разложение произведения двух полученных треугольных матриц, с помощью которого находится оптимальное приближение этого произведения матрицей ранга k .

Полученная таким образом аппроксимация ранга k найденной на предыдущем шаге матрицы CGR считается искомым приближением для Y_k .

Таким образом, полная сложность алгоритма аппроксимации составляет $O((m+n)kp + p^3) = O((m+n)p^2q + p^3)$. Данная аппроксимация может быть использована в качестве оператора приближенного проектирования $\hat{\mathcal{P}}_k$ для алгоритма ASVP.

Полученное CGR -разложение записывается в виде $CGR = X_{t+1} = U_{t+1}V_{t+1}^T$ с $U_{t+1} \in \mathbb{R}^{m \times k}$, $V_{t+1}^T \in \mathbb{R}^{k \times n}$.

Напомним, что приближенное проектирование (аппроксимация) проводится для матрицы вида

$$Y = X_t - \tau \mathcal{S}^T(\mathcal{S}(X_t) - \mathbf{B}).$$

Всего нам требуется $O(p)$ ее строк и столбцов, которые легко найти на основе матрицы $X_t = U_t V_t^T$ за $O((m+n)kp)$ операций, что не увеличивает асимптотическую сложность алгоритма.

Заметим, что на практике число шагов алгоритма `maxvol` можно считать константой, так как по ходу итераций алгоритма ASVP приближение каждой итерации меняется медленно, откуда имеем, что и объем подматриц приближения меняется слабо. Алгоритм `maxvol` основан на максимизации объема, а потому будет слабо реагировать на небольшие изменения входных данных.

Более того, на основе сингулярных чисел, которые отбрасываются на стадии сингулярного разложения аппроксимации CGR , можно судить о скорости сходимости алгоритма и иногда вообще не изменять выбранные ранее строки и столбцы.

В численных экспериментах использовались

$$p = 2k + \left\lceil 0.7 \frac{k}{q} \right\rceil$$

строк и столбцов. Это число намеренно ниже оценки (23): численные эксперименты из [17] показали, что реальные значения погрешности гораздо ниже верхних оценок, хотя те и предсказывают верную асимптотическую зависимость.

6. СРАВНИТЕЛЬНЫЙ АНАЛИЗ ВЫЧИСЛИТЕЛЬНОЙ СЛОЖНОСТИ АЛГОРИТМОВ ASVP И SVP

Рассмотрим соотношения сложностей алгоритма SVP с использованием полного сингулярного разложения матрицы и алгоритма ASVP с использованием приближенного проектирования матрицы. Для анализа отметим следующие особенности алгоритмов.

1. Доказательство сходимости алгоритмов SVP и ASVP базируется на использовании свойства ограниченной изометрии (3) рассматриваемого оператора A .

2. В случае, если A – оператор задачи матричного восполнения, существует асимптотическая оценка на минимальное количество известных элементов матрицы, которого достаточно для полного восполнения. Эта оценка определяется размерами неизвестной матрицы, рангом неизвестной матрицы, и “некогерентностью” неизвестной матрицы, численной величиной, характеризующей разреженность сингулярных факторов матрицы.

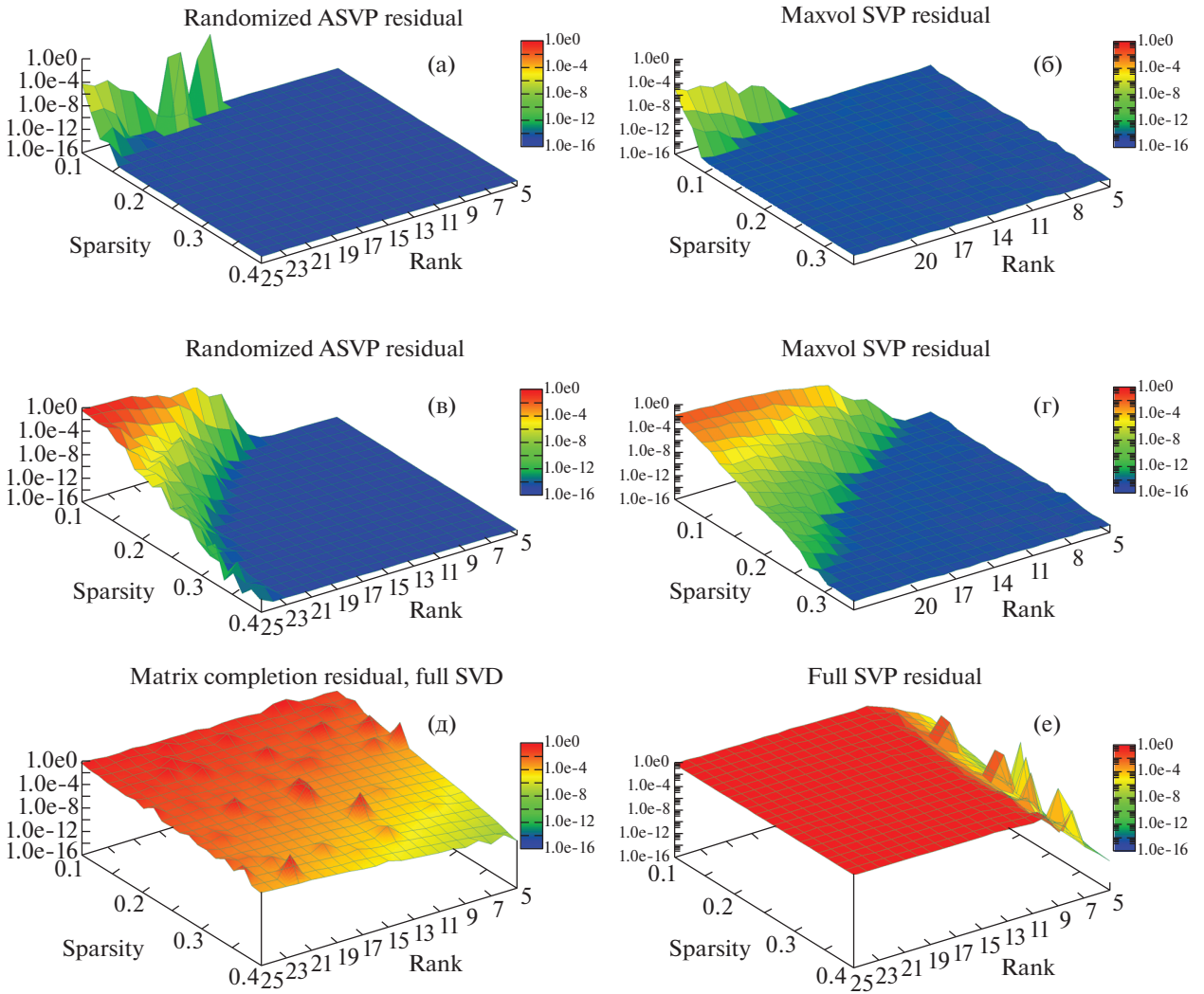
3. Не обязательно использовать именно такой порядок числа известных элементов матрицы, допустимо использовать и большие порядки.

4. В случае алгоритма ASVP гарантируется геометрическая сходимость алгоритма с порядком не менее $O\left(\frac{\epsilon}{q}\right)$, где q – плотность известных элементов матрицы, а ϵ – относительная ошибка приближенного проектирования. Таким образом, при уменьшении числа известных элементов матрицы требуется большая точность приближенного проектирования.

5. И в случае ASVP с использованием проектирования на случайные подпространства, и в случае ASVP с использованием псевдоскелетного метода, вводится дополнительный параметр p , определяющий некую вспомогательную размерность метода, причем с увеличением p увеличивается и точность, и вычислительная сложность приближенного проектирования, и наоборот.

6. Таким образом, с уменьшением плотности известных элементов матрицы требуется большая точность приближения, что приводит к увеличению вспомогательных размерностей рассматриваемых методов случайного проектирования, что может привести к увеличению сложности алгоритма ASVP.

Проведем формальное сравнение алгоритмов. Сложность SVP и ASVP удобно выразить через параметры m, n, k, q, p . Кроме того, $q \geq q_{\min}(m, n, k, \mu)$; также и для псевдоскелетного метода, и



Фиг. 1. Графики итоговой относительной невязки $\|\mathcal{F}(X_t - X_*)\|_2 / \|\mathcal{F}(X_*)\|_2$ восполнения матриц с сингулярными числами вида $\sigma_i = 1$: (а), (в) – ASVP со случайным проектированием, ограничение по времени в 50 и в 10 итераций SVP соответственно; (б), (г) – псевдоскелетный ASVP, ограничение по времени в 50 и в 10 итераций SVP соответственно; (д), (е) – полный SVP (50 итераций) с параметрами шага $\tau = q$ и $\tau = 3/4$ соответственно.

для случайного проектирования выполнено $\epsilon = O\left(\frac{k}{p}\right)$; считая множитель геометрической сходимости константой, чтобы обеспечить константный порядок числа итераций алгоритма, имеем $\epsilon = O(q)$, откуда $p = O\left(\frac{k}{q}\right)$.

Алгоритм SVP состоит из двух основных операций, оценим их сложность.

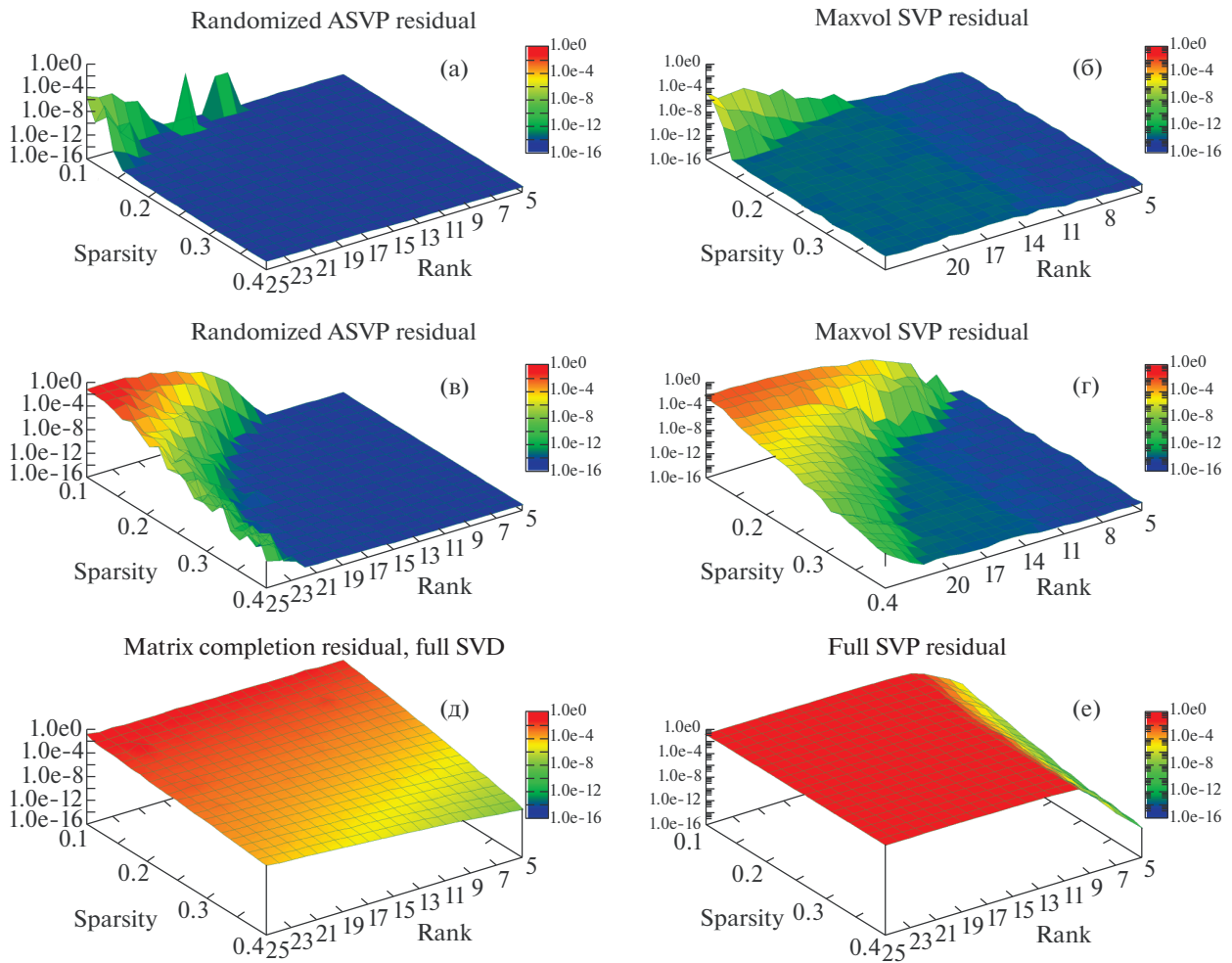
1. Вычисление невязки на маске известных элементов покомпонентно: $O(k)$ операций на каждый элемент, $O(mnkq)$ операций всего.

2. Вычисление сингулярного разложения для получения приближения следующей итерации: $O(mn \min(m, n))$.

Так как $q < 1$, $k < \min(m, n)$, имеем сложность одной итерации алгоритма $O(mn \min(m, n))$, кубическую в случае квадратной матрицы.

Рассмотрим полученные выше оценки на сложность итерации алгоритма ASVP.

- $O((m + n)p^2) + O(mnpq)$ для случайного проектирования,



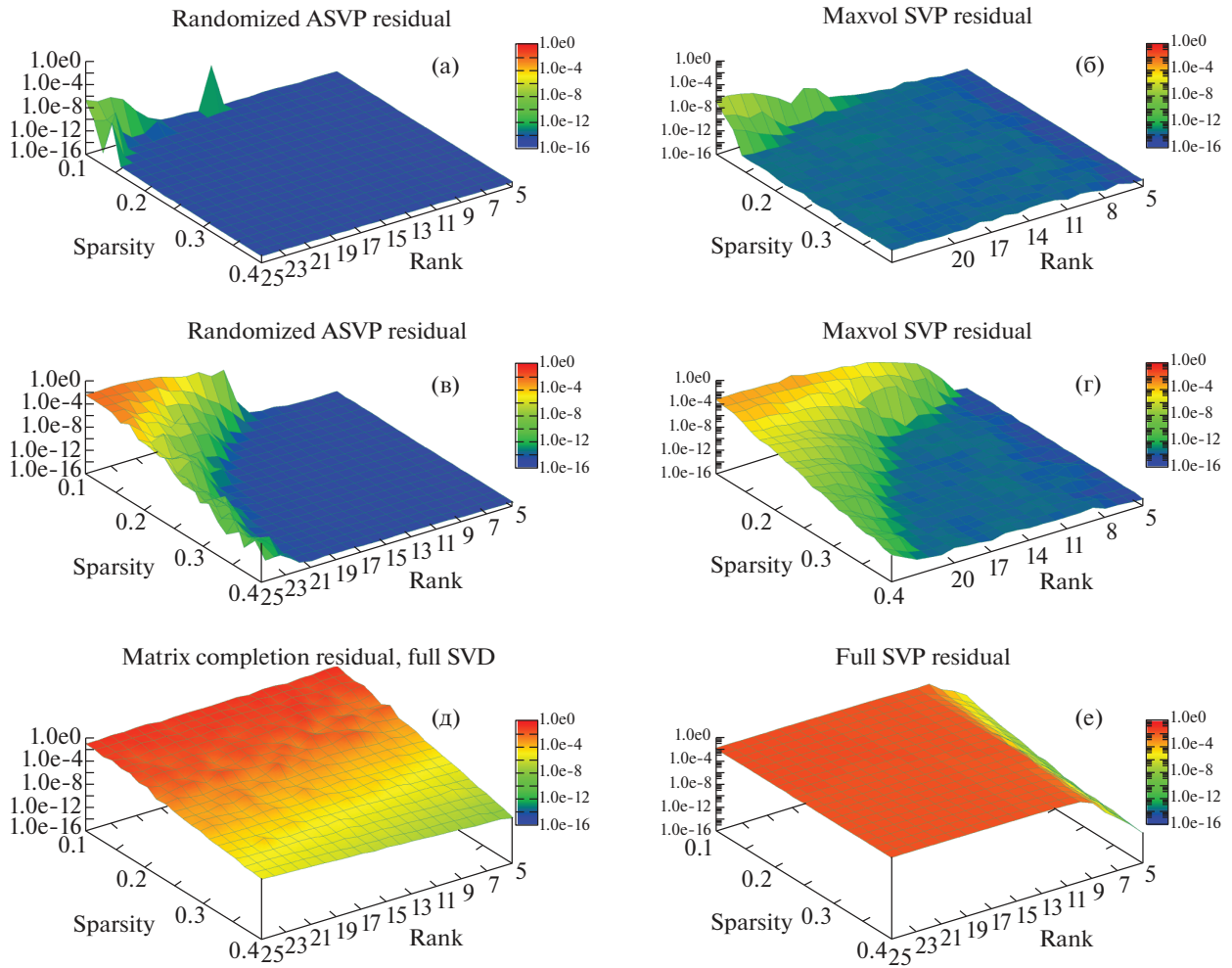
Фиг. 2. Графики итоговой относительной невязки $\|\mathcal{F}(X_t - X_*)\|_2 / \|\mathcal{F}(X_*)\|_2$ восполнения матриц с сингулярными числами вида $\sigma_i = \frac{1}{i}$: (а), (в) – ASVP со случайным проектированием, ограничение по времени в 50 и в 10 итераций SVP соответственно; (б), (г) – псевдоскелетный ASVP, ограничение по времени в 50 и в 10 итераций SVP соответственно; (д), (е) – полный SVP (50 итераций) с параметрами шага $\tau = q$ и $\tau = 3/4$ соответственно.

- $O((m + n)kp) + O(p^3)$ для псевдоскелетного метода.

С учетом $q = O(\epsilon) = O\left(\frac{k}{p}\right)$ для метода случайного проектирования имеем, что асимптотика сложности одной итерации псевдоскелетного метода в $O(\min(m, n)/p)$ раз меньше сложности одной итерации метода случайного проектирования.

Теперь рассмотрим сложности ASVP в зависимости от различных q . Пусть $n \geq m$. В случае минимально возможного $q = q_{\min} = O\left(\frac{k^2 \mu^2 \log n}{m}\right)$ имеем сложность порядка

- $O\left(\frac{(m + n)m^2}{k^2 \mu^4 \log^2 n}\right) + O(mnk)$ для случайного проектирования;
- $O\left(\frac{m^3}{\mu^6 k^3 \log^3 n}\right) + O\left(\frac{(m + n)m}{\mu^2 \log n}\right)$ для псевдоскелетного метода.



Фиг. 3. Графики итоговой относительной невязки $\|\mathcal{F}(X_t - X_*)\|_2 / \|\mathcal{F}(X_*)\|_2$ выполнения матриц с сингулярными числами вида $\sigma_i = \frac{1}{i^2}$: (а), (в) – ASVP со случайным проектированием, ограничение по времени в 50 и в 10 итераций SVP соответственно; (б), (г) – псевдоскелетный ASVP, ограничение по времени в 50 и в 10 итераций SVP соответственно; (д), (е) – полный SVP (50 итераций) с параметрами шага $\tau = q$ и $\tau = 3/4$ соответственно.

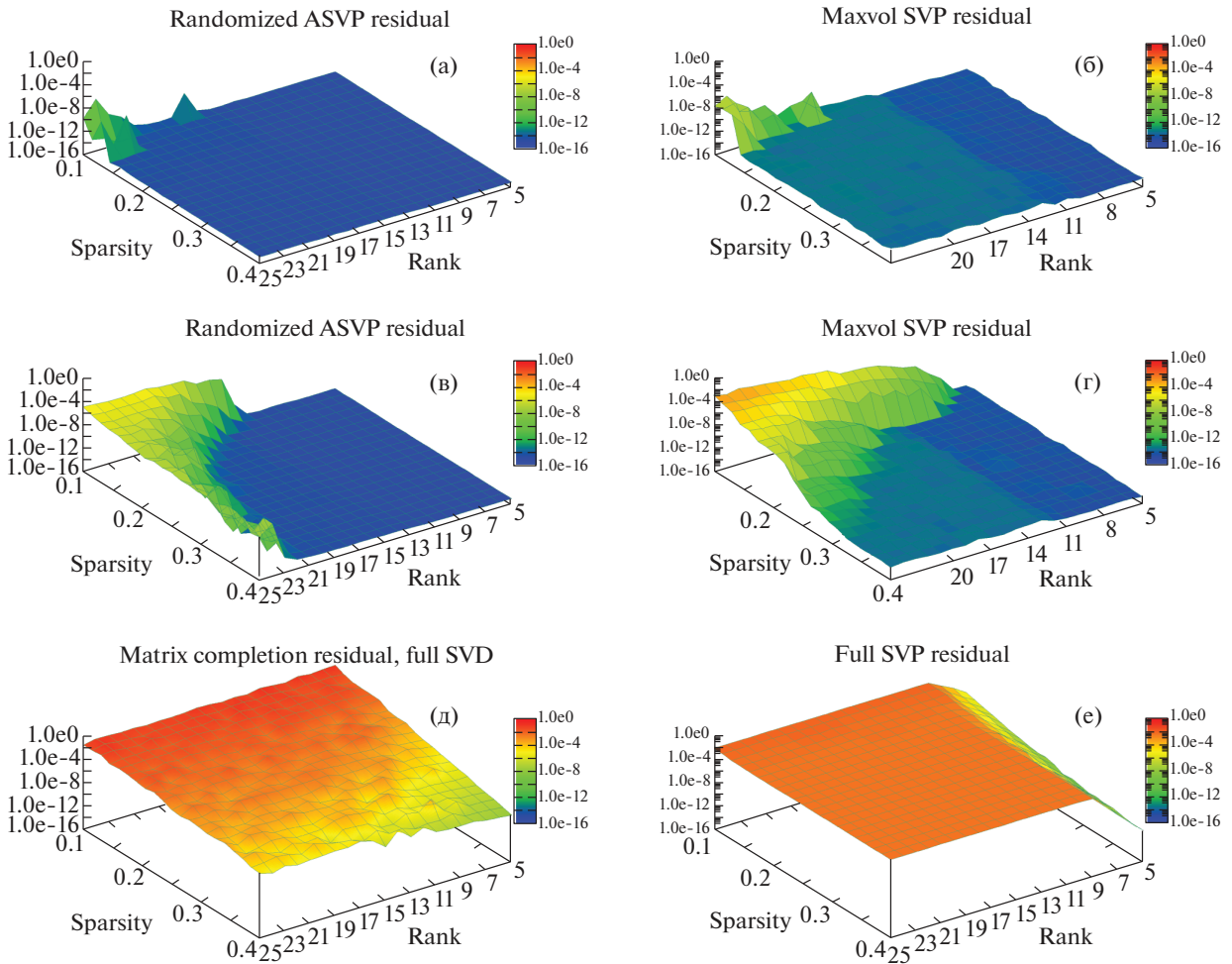
Покажем, что при больших порядках q можно достичь и более низкой сложности ASVP. Пусть $q = m^{-1/2} \gg \frac{k^2 \mu^2 \log n}{m}$ (при малых порядках μ , близких к константе или к логарифму от размеров матрицы); тогда подстановкой имеем сложность

- $O((m+n)mk^2)$ для случайного проектирования;
- $O((m+n)m^{1/2}k^2) + O(m^{3/2}k^3)$ для псевдоскелетного метода.

7. ЧИСЛЕННЫЕ ЭКСПЕРИМЕНТЫ

Проведем экспериментальное сравнение алгоритмов SVP и ASVP с двумя рассмотренными вариантами методов приближенного проектирования.

- Рассматриваются квадратные матрицы размера $m = n = 1000$. Рассматриваются ранги матриц k от 5 до 25 и плотности распределения известных элементов q от 0.1 до 0.4.



Фиг. 4. Графики итоговой относительной невязки $\|\mathcal{F}(X_t - X_*)\|_2 / \|\mathcal{F}(X_*)\|_2$ восполнения матриц с сингулярными числами вида $\sigma_i = \frac{1}{2^i}$: (а), (в) – ASVP со случайным проектированием, ограничение по времени в 50 и в 10 итераций SVP соответственно; (б), (г) – псевдоскелетный ASVP, ограничение по времени в 50 и в 10 итераций SVP соответственно; (д), (е) – полный SVP (50 итераций) с параметрами шага $\tau = q$ и $\tau = 3/4$ соответственно.

- Матрицы формируются случайно следующим образом.

- Формируются случайные матрицы $\hat{U}_k \in \mathbb{R}^{m \times k}$ и $\hat{V}_k \in \mathbb{R}^{n \times k}$, каждый элемент которых распределен случайно по стандартному нормальному распределению.

- Вычисляются ортогональные базисы U_k и V_k в пространствах столбцов \hat{U}_k и \hat{V}_k соответственно, например, с помощью QR-разложения. Экспериментально установлено, что некогерентность факторов U_k, V_k , выбранных таким образом, растет логарифмически от m, n .

- Искомая матрица X_* задается в виде $X_* = U_k \Sigma_k V_k^T$ с матрицей Σ_k сингулярных значений $\sigma_1 \geq \sigma_2 \geq \sigma_3 \geq \dots$. Экспериментально установлено, что сходимость методов SVP, ASVP зависит от характера убывания сингулярных чисел σ_j ; в экспериментах рассмотрены следующие законы σ_j : $\sigma_i = 1$ (фиг. 1), $\sigma_i = 1/i$ (фиг. 2), $\sigma_i = 1/i^2$ (фиг. 3), $\sigma_i = 1/2^i$ (фиг. 4).

- Сравнение проводится с ограничением по времени практической работы алгоритмов. Ограничение задается константой, которая определяется временем выполнения десяти или пятидесяти итераций алгоритма SVP с полным SVD; предполагается, что за это время приближенный алгоритм ASVP выполнит большее число итераций.

Ниже приведены графики зависимостей невязки от q и k при фиксированном размере матрицы в случае алгоритма SVP и двух рассмотренных вариантов алгоритма ASVP при различных зонах падения сингулярных чисел неизвестной матрицы. Значения полных относительных ошибок по всей матрице на практике во всех экспериментах отличаются от относительных невязок на маске известных элементов не более, чем в два раза.

Графики сходимости в случае полного SVP приведены в двух вариантах: с шагом q (фиг. 1д–4д) и с шагом $3/4$ (фиг. 1е–4е). Так как рассматривалось ограничение в 50 итераций SVP, вариант SVP с большим шагом на большей части графиков не достиг сходимости, так как такого числа итераций оказалось недостаточно для завершения последовательного набора ранга.

8. ЗАКЛЮЧЕНИЕ

В данной статье рассмотрено применение методов приближенного вычисления частичного сингулярного разложения к существующему итерационному алгоритму восполнения матриц малого ранга, сходящемуся геометрически, основной операцией которого является сингулярное разложение матрицы с последующим отсечением всех неглавных компонент (SVD-проекция) на каждой итерации. Для этого алгоритма получен теоретический результат, при определенных предположениях гарантирующий сохранение геометрической сходимости даже в случае замены точного вычисления SVD-проекции на приближенное. Сформулировано условие приближения, выполнения которого достаточно для геометрической сходимости при произвольном методе получения самой приближенной проекции. Рассмотрены два возможных варианта таких методов, удовлетворяющих этому условию в среднем или с высокой вероятностью. Для обоих этих методов проведены численные эксперименты и получены похожие результаты, показывающие большую вычислительную эффективность использования приближенных SVD-проекций.

СПИСОК ЛИТЕРАТУРЫ

1. Kang Z., Peng C., Cheng Q. Top-N Recommender System via Matrix Completion // arXiv preprint arXiv:1601.04800v1, 01/2016.
2. Ahmed A., Romberg J. Compressive multiplexing of correlated signals // IEEE Trans. Inform. Theory. 2015. V. 61. № 1. P. 479–498.
3. Davies M., Eldar Y. Rank awareness in joint sparse recovery // IEEE Trans. Inform. Theory. 2012. V. 58. № 2. P. 1135–1146.
4. Hu R., Tong J., Xi J., Guo Q., Yu Y. Low-Complexity and Basis-Free Channel Estimation for Switch-Based mmWave MIMO Systems via Matrix Completion. arXiv preprint arXiv:1609.05693v2[cs.IT], 11/2016.
5. Argyriou A., Evgeniou T., Pontil M. Convex multi-tasks feature learning // Machine Learn. 2008. V. 73. № 3. P. 243–272.
6. Blei D. Probabilistic topics models // Comm. ACM. 2012. V. 55. № 4. P. 77–84.
7. Inderjit S. Dhillon Paghu Meka, Prateek Jain. Guaranteed rank minimization via singular value projection // arXiv preprint arXiv:0909.5457, 09/2009.
8. Tanner J., Wei K. Normalized iterative hard thresholding for matrix completion // SIAM J. Sci. Comput. 2013. V. 35. № 5. P. 104–125.
9. Tropp J.A., Halko N., Martinsson P.G. Finding structure with randomness: probabilistic algorithms for constructing approximate matrix decompositions // SIAM Rev. 2011. V. 53. № 2. P. 217–288.
10. Drienas P., Mahoney M.W. Lectures on Randomized Numerical Linear Algebra // arXiv preprint arXiv:1712.08880v1 [cs.DS], 12/2017.
11. Goreinov S.A., Tyrtshnikov E.E., Zamarashkin N.L. A theory of pseudoskeleton approximations // Linear Algebra and its Appl. 1996. V. 261. № 1–3. P. 1–21.
12. Osinsky A.I., Zamarashkin N.L. Pseudo-skeleton approximations with better accuracy estimates // Linear Algebra and its Appl. 2018. V. 537. P. 221–249.
13. Osinsky A.I., Zamarashkin N.L. New accuracy estimates for pseudoskeleton approximations of matrices // Doklady Math. 2016. V. 94. № 3. P. 643–645.
14. Simon H.D., Zha H. Low-rank matrix approximation using the Lanczos Bidiagonalization process with applications // SIAM J. Sci. Comput. 2000. V. 21. № 6. P. 2257–2274.
15. Mikhalev A.Y., Oseledets I.V. Rectangular submatrices of maximum volume and their computation // Doklady Math. 2015. V. 91. № 3. P. 267–268.
16. Goreinov S.A., Oseledets I.V., Savostyanov D.V. et al. How to find a good submatrix / Matrix Methods: Theory, Algorithms, Applications. Ed. by V. Olshevsky, E. Tyrtshnikov. World Sci. Publ., 2010. P. 247–256.
17. Osinsky A.I. Rectangular maximum volume and projective volume search algorithms // arXiv 1809.02334 (Submitted on 7 Sep 2018)
18. Boutsidis C., Drineas P., Magdon-Ismail M. Near-optimal column-based matrix reconstruction // SIAM J. Comput. 2014. V. 43. № 2. P. 687–717.