ЖУРНАЛ ВЫЧИСЛИТЕЛЬНОЙ МАТЕМАТИКИ И МАТЕМАТИЧЕСКОЙ ФИЗИКИ, 2021, том 61, № 11, с. 1814–1824

ОПТИМАЛЬНОЕ УПРАВЛЕНИЕ

УДК 519.853.6

МЕТОД ПРОЕКЦИИ ГРАДИЕНТА С ШАГОМ АРМИХО НА МНОГООБРАЗИЯХ

© 2021 г. **М. В. Балашов[1,*], Р. А. Камалов[1,**]**

[1] 117997 Москва, ул. Профсоюзная, 65, Институт проблем управления РАН им. В.А. Трапезникова, Россия

*e-mail: balashov73@mail.ru

**e-mail: kamalov.ra@phystech.edu

Поступила в редакцию 23.10.2020 г.
Переработанный вариант 23.10.2020 г.
Принята к публикации 09.07.2021 г.

Рассматривается задача минимизации функции с непрерывным по Липшицу градиентом на проксимально гладком подмножестве, которое является гладким многообразием без края. Обсуждается метод проекции градиента с шагом Армихо и доказывается его линейная сходимость. Для различных матричных множеств и многообразий получена точная константа проксимальной гладкости. Библ. 21.

Ключевые слова: проксимальная гладкость, метод проекции градиента, невыпуклая экстремальная задача, шаг Армихо, матричные многообразия.

DOI: 10.31857/S004446692111003X

1. ВВЕДЕНИЕ И ОСНОВНЫЕ ОБОЗНАЧЕНИЯ

1.1. Введение

Рассмотрим в $\mathbb{R}^n$ задачу

$$\min_{x\in S} f(x). \tag{1}$$

Для решения задачи (1) в выпуклом случае давно используется метод проекции градиента (далее МПГ), который появился в [1], [2]. Напомним, что если функция в (1) сильно выпуклая с непрерывным по Липшицу градиентом и множество S выпукло и замкнуто, то МПГ сходится со скоростью геометрической прогрессии (или линейной скоростью).

Мы предполагаем, что S – гладкое многообразие без края, а f – функция с непрерывным по Липшицу градиентом, которая необязательно выпуклая. Пусть T_x – касательное подпространство к S в точке $x \in S$ и $T_x^{\perp}$ – его ортогональное дополнение. Мы рассматриваем МПГ вида $x_0 \in S$, $k = 0$,

$$t_k > 0, \quad z_k = P_{T_{x_k}}(x_k - t_k f'(x_k)), \quad x_{k+1} = R_S z_k, \quad k = k+1. \tag{2}$$

Здесь P_A – оператор метрического проектирования на замкнутое множество $A \subset \mathbb{R}^n$, а $R_S z_k$ – некоторая ретракция (см. подробности в [3]) точки z_k на множество S, $R_S z_k \in S$. Часто используется $R_S = P_S$. Другой вид ретракции мы обсудим ниже.

Примером актуальной задачи (1) является задача минимизации гладкой функции f на некотором матричном многообразии S без края (см. [4], [5]).

Традиционно в задаче (1) используются варианты метода проекции градиента с шагами, связанными с локальными геодезическими на многообразии (см. [5]–[9]), а также метод Ньютона (см. [4], [7]).

В последнее время появилось много работ, где используется идеология (2). Основными трудностями являются выбор шага t_k и доказательство сходимости алгоритма при разумных предположениях.

В [10] рассмотрен МПГ в задаче (1) с шагом t_k Армихо (см. определение А1). Однако по сути доказана лишь асимптотика последовательности $\{x_k\}$, но не явные оценки сходимости (см. [10, Theorem 2.3]). Аналогичный результат при некотором фиксированном $t_k = t > 0$ для всех k получен и в [11, Corollary 4.2] при условии, что кривизны многообразия S ограничены. В обоих работах для линейной сходимости принципиально условие Лежанского–Поляка–Лоясевича (условие ЛПЛ) функции f на поверхности S (см. ниже определение 1). Близкий к алгоритму из [10] алгоритм с шагом Армихо рассмотрен в [12, Algorithm 2.1]. Однако в [12] исследовался только факт сходимости алгоритма без оценки скорости сходимости. Кроме того, множество S предполагалось выпуклым, а функция f невыпуклой.

Обозначим через $B_R(x)$ замкнутый шар с центром x радиуса R. В [13, Theorems 2, 3] получен следующий результат.

Теорема А. *Пусть многообразие S гладкое и проксимально гладкое с константой $\frac{\pi}{2}R$ многообразие без края, $x_0 \in S$. Предположим, что $f : \mathbb{R}^n \to \mathbb{R}$ функция обладает следующими свойствами*:

1) *f липшицево дифференцируема с константой L_1*,

2) $L = \sup\limits_{x : \varrho(x,S)<R} \|f'(x)\| < +\infty$, *где* $\varrho(x,S) = \inf_{s\in S}\|x - s\|$,

3) *выполнено условие ЛПЛ для функции f на $S \cap \mathcal{L}_f(f(x_0))$, $\mathcal{L}_f(f(x_0)) = \{x \in \mathbb{R}^n \,|\, f(x) \le f(x_0)\}$, с константой $\mu > 0$, т.е. для всех $x \in S \cap \mathcal{L}_f(f(x_0))$*

$$\mu(f(x) - \inf_{x\in S} f) \le \|P_{T_x} f'(x)\|^2. \tag{3}$$

Пусть $t_0 = \dfrac{1}{\frac{2L}{R} + L_1}$ *и* $t \in (0, t_0]$. *Положим* $q(t) = t - t^2\frac{L}{R} - t^2\frac{L_1}{2}$. *Тогда итерации* $x_0 \in S$,

$$1)\quad z_k = x_k - tP_{T_{x_k}} f'(x_k), \qquad x_{k+1} = S \cap (z_k + T_{x_k}^{\perp}) \cap B_R(x_k),$$

или

$$2)\quad z_k = x_k - tP_{T_{x_k}} f'(x_k), \qquad x_{k+1} = P_S z_k,$$

сходятся с линейной скоростью по функции

$$f(x_{k+1}) - f_* \le (1 - \mu q(t))(f(x_0) - f_*), \qquad f_* = \inf_{x\in S} f(x).$$

При этом

$$f(x_{k+1}) - f(x_k) \le -\|P_{T_{x_k}} f'(x_k)\|^2 q(t). \tag{4}$$

Кроме того, для случая 1) *имеет место следующая линейная скорость сходимости по точке*

$$\|x_{k+1} - x_k\|^2 \le \frac{t^2 + t^4\frac{L^2}{R^2}}{q(t)}(1 - \mu q(t))^k (f(x_0) - f_*), \qquad f_* = \inf_{x\in S} f(x). \tag{5}$$

Отметим, что $1 - \mu q(t) \in (0, 1)$.

В случае итераций 1) пересечение $S \cap (z_k + T_{x_k}^{\perp}) \cap B_R(x_k)$ одноточечно для всякого k (см. [13, Lemma 5]).

Покажем, что условие (5) действительно означает линейную сходимость к решению (1). Положим

$$\theta = \sqrt{1 - \mu q(t)} \in (0, 1),\; C = \frac{t^2 + t^4\frac{L^2}{R^2}}{q(t)}(f(x_0) - f_*)$$

Для $M \geq N$

$$\|x_M - x_N\| \leq \sum_{k=N}^{M-1} \|x_{k+1} - x_k\| \leq \frac{C\theta^N}{1-\theta} \to 0, \quad N \to \infty.$$

Значит, $x_N \to x_* \in S$. Очевидно, что $\|x_N - x_*\| \leq \frac{C\theta^N}{1-\theta}$, т.е. сходимость $\{x_k\}$ линейная. В силу [13, Theorem 2] $f(x_k) - f(x_{k+1}) \geq q(t)\left\|P_{T_{x_k}} f'(x_k)\right\|^2$ для всех k и в силу условия ЛПЛ

$$f(x_k) - f_* \leq \left\|P_{T_{x_k}} f'(x_k)\right\|^2 \Big/ \mu \leq \frac{f(x_k) - f(x_{k+1})}{q(t)\mu}.$$

Отсюда $f(x_*) = f_*$.

Условие ЛПЛ с показателем 2 является одним из наиболее общих условий на гладкую функцию на многообразии, которое обеспечивает линейную сходимость градиентных методов (см. [10]). В частности, сильно выпуклая функция на многообразии при определенном соотношении константы сильной выпуклости и других параметров удовлетворяет условию ЛПЛ, а также некоторые функции с условием квадратичного роста на множестве (см. [13, Theorem 1]). Также условию ЛПЛ удовлетворяет квадратичная форма на сфере или, более общо, квадратичная форма на многообразии Штифеля (см. [14]).

Теорема A решает вопрос об оценке скорости сходимости алгоритма и выборе шага $t_k = t$ через константы Липшица L, L_1 и константу проксимальной гладкости R многообразия. Константа μ из условия ЛПЛ возникает в оценке (5), она не нужна для выбора шага t. Ключевое отличие в доказательстве теоремы A от приведенных выше результатов состоит в использовании проксимальной гладкости множества S, что позволяет получить оценку сходимости (5) в явном виде.

Тем не менее на практике константы L, L_1 и R могут быть неизвестны. В этой ситуации становится актуальным выбор шага t_k в (2) по некоторому правилу, которое не требует знания упомянутых констант. Правило Армихо как раз и является одним из таких способов выбора шага t_k.

В работе рассмотрены два способа выбора шага t_k по правилу Армихо. Это правило A1, заимствованное в [10], а также правило A2, сформулированное нами.

В обоих случаях мы получаем явную оценку скорости сходимости метода (2) для правил выбора шага A1 и A2.

При правиле A1 нам не нужно знать никаких констант.

В случае, когда константа проксимальной гладкости известна и известна оценка сверху для константы L, можно применять правило A2. Правило A2 имеет техническое преимущество: в отличие от правила A1 при подсчете шага t_k ретракцию точки на множество нужно вычислять 1 раз.

В Приложении мы приводим точные константы проксимальной гладкости R для основных матричных многообразий.

1.2. Основные обозначения

Через $\mathbb{R}^n$ будем обозначать вещественное евклидово пространство n измерений со скалярным произведением (x, y) $\forall x, y \in \mathbb{R}^n$. Обозначим через $B_R(x)$ замкнутый шар с центром x радиуса R. Далее по тексту для задачи (1) T_x – касательное подпространство к многообразию S в точке $x \in S$, $f'(x)$ – градиент Фреше функции f в точке x. Для точки $x_k \in S$ для краткости обозначим $\xi_k = P_{T_{x_k}} f'(x_k)$. Отметим, что $(f'(x_k), \xi_k) = \|\xi_k\|^2$.

Если f' – липшицева функция с константой L_1, то для любых $x, y \in \mathbb{R}^n$ верны оценки (см. [15, Лемма 1.2.3])

$$f(x) + (f'(x), y - x) - \frac{L_1}{2}\|y - x\|^2 \leq f(y) \leq f(x) + (f'(x), y - x) + \frac{L_1}{2}\|y - x\|^2.$$

Лебеговым множеством функции f для $\beta \in \mathbb{R}$ называют множество $\mathcal{L}_f(\beta) = \{x \in R^n \mid f(x) \leq \beta\}$.

Определение 1 (см. [10], [16, Definition 1]). Пусть S является C^1 многообразием, функция $f : \mathbb{R}^n \to \mathbb{R}$ дифференцируема. Пусть $\mu > 0$, $\beta \in \mathbb{R}$, $f_* = \inf_{x \in S} f(x)$. Будем говорить, что функция f удовлетворяет условию ЛПЛ на множестве $S \cap \mathcal{L}_f(\beta)$, если $\|P_{T_x} f'(x)\|^2 \geq \mu(f(x) - f_*)$ $\forall x \in S \cap \mathcal{L}_f(\beta)$.

Определим $O(n)$ – ортогональные матрицы размера $n \times n$. Для матриц $X \in \mathbb{R}^{n \times k}$ и $Y \in \mathbb{R}^{n \times k}$ определим скалярное произведение $(X, Y) = \operatorname{tr} X^{\mathrm{T}Y}$ и норму Фробениуса $\|X\| = \sqrt{\operatorname{tr} X^{\mathrm{T}} X}$. Напомним (см. [17, Theorem 7.3.2]), что для каждой матрицы $X \in \mathbb{R}^{n \times k}$ ранга r существует сингулярное разложение вида $X = U \Sigma V^{\mathrm{T}}$, где

$$U \in O(n), V \in O(k), \Sigma = \begin{pmatrix} \sigma_1 & \dots & \dots & \dots & \dots & \dots & 0 \\ 0 & \sigma_2 & \dots & \dots & \dots & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & \sigma_r & \dots & \dots & 0 \\ 0 & 0 & \dots & \dots & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & \dots & \dots & \dots & 0 \end{pmatrix} \in \mathbb{R}^{n \times k}.$$

Числа $\sigma_1 \geq \ldots \geq \sigma_r > 0$ называются сингулярными числами матрицы X. Для случая $k \leq n$ договоримся обозначать оставшиеся $k - r$ нулей на главной диагонали матрицы Σ через $\sigma_{r+1}, \ldots, \sigma_k$.

Мы будем пользоваться следующей формулой (см. [17, Corollary 7.3.5]) для $X, Y \in \mathbb{R}^{n \times k}$:

$$\sum_{i=1}^{k} (\sigma_i(X) - \sigma_i(Y))^2 \leq \|X - Y\|^2. \tag{6}$$

Через I_k обозначим единичную матрицу размера $k \times k$.

Множество $A \subset \mathbb{R}^n$ называется проксимально гладким (прокс-регулярным, слабо выпуклым) с константой $R > 0$ (см. [18], [19]), если функция расстояния $\varrho(x, A) = \inf_{a \in A} \|x - a\|$ непрерывно дифференцируема на множестве $U_A(R) = \{x \in \mathbb{R}^n \,|\, 0 < \varrho(x, A) < R\}$. Эквивалентным условием проксимальной гладкости множества $A \subset \mathbb{R}^n$ является условие, что для всякой точки $x \in U_A(R)$ метрическая проекция $P_A(x)$ является одноточечным множеством.

Напомним определения основных матричных многообразий и множеств (далее $k \leq n$):

(i) многообразие Штифеля $\mathcal{S}_{n,k} = \{X \in \mathbb{R}^{n \times k} \,|\, X^{\mathrm{T}} X = I_k\}$;

(ii) многообразие Грассмана $\mathcal{G}_{n,k}$ – множество всех k-мерных подпространств в $\mathbb{R}^n$. Мы рассматриваем реализацию $\mathcal{G}_{n,k}$ вида $\mathcal{G}_{n,k} = \{XX^{\mathrm{T}} \,|\, X \in \mathcal{S}_{n,k}\}$ (см. [20]), т.е. подмножество во множестве симметричных матриц $n \times n$;

(iii) многообразие $\mathfrak{M}_r = \{X \in \mathbb{R}^{n \times k} \,|\, \operatorname{rank} X = r, \sigma_i(X) \geq \sigma_0 > 0 \ \ \forall i = \overline{1, r}\}$ – матрицы ранга r с сингулярными числами $\geq \sigma_0 > 0$;

(iv) множество $\mathfrak{L}_r = \{X \in \mathbb{R}^{n \times k} \,|\, 0 < \operatorname{rank} X \leq r, \sigma_i(X) \geq \sigma_0 > 0 \ \ \forall i = \overline{1, \operatorname{rank}(X)}\}$ – матрицы ранга $> 0, \leq r$ с сингулярными числами $\geq \sigma_0 > 0$.

В [21] было показано, что $\mathcal{S}_{n,k}$ – проксимально гладкое множество с $R = 1$, а $\mathcal{G}_{n,k}$ (в указанной реализации) – проксимально гладкое множество с $R = \frac{1}{\sqrt{2}}$. Там же приводятся формулы для метрической проекции матрицы на многообразие Штифеля или Грассмана. Точные значения констант проксимальной гладкости для множеств $\mathfrak{M}_r$, $\mathfrak{L}_r$ и метрические проекции на них найдены нами в Приложении.

2. РЕТРАКЦИИ И ПРАВИЛО АРМИХО

Мы будем рассматривать две возможности для ретракции.

Во-первых, в качестве ретракции выступает оператор метрического проектирования P_S. В Приложении мы покажем, что, например, на большинство матричных многообразий метрическая проекция может быть найдена с помощью сингулярного разложения матрицы.

Во-вторых, мы рассмотрим упомянутый в теореме А выбор $x_{k+1} = (z_k + T_{x_k}^{\perp}) \cap S \cap B_R(x_k)$. В [13, Theorem 2] доказано, что при условии $t_k \|\xi_k\| \le \frac{\sqrt{3}}{2}R$ точка x_{k+1} определена корректно и однозначно. Далее будем обозначать указанную ретракцию через R_S.

Пусть $d > 0$; $\alpha, \beta \in (0, 1)$. Мы рассмотрим два способа выбора шага в алгоритме (2) по правилу Армихо.

Определение A1 (см. [10]). Определим t_k на k-м шаге по правилу

$$z_k = x_k - t_k\xi_k, \quad x_{k+1} \in P_S z_k, \quad \text{где}$$

$$t_k = \max_{m\in\mathbb{N}\cup\{0\}} \{d\beta^m \,|\, f(x_{k+1}) \le f(x_k) - \alpha d\beta^m \|\xi_k\|^2\}.$$

Отметим, что несмотря на проксимальную гладкость S, мы не можем гарантировать попадание точки z_k в проксимальный слой множества S, поэтому множество $P_S z_k$ может быть неодноточечно и x_{k+1} выбирается из него произвольно.

Определение A2. Пусть $\alpha_1 \in (0, 1)$, $\alpha_1 < \alpha$, множество S проксимально гладкое с константой $\frac{\pi}{2}R$ и функция f липшицева с константой $\le L$. Определим t_k на k-м шаге по правилу

$$d < \alpha_1 \frac{\sqrt{3}R}{2L}, \quad z_k = x_k - t_k\xi_k, \quad \text{где}$$

$$t_k = \max_{m\in\mathbb{N}\cup\{0\}} \{d\beta^m \,|\, f(z_k) \le f(x_k) - \alpha d\beta^m \|\xi_k\|^2\}.$$

При этом $x_{k+1} = P_S z_k$ или $x_{k+1} = R_S z_k$.

Покажем, что обе ретракции в случае определения $A2$ корректно определены. Действительно, когда шаг t_k найден, вычисляется $x_{k+1} = P_S z_k$. В силу $t_k \le d < \alpha_1 \frac{\sqrt{3}R}{2L}$ точка $z_k = x_k - t_k\xi_k$ находится в проксимальном слое: $\varrho(z_k, S) \le \|z_k - x_k\| = t_k \|\xi_k\| < \alpha_1 \frac{R}{L} L = \alpha_1 R < R$. Поэтому $P_S z_k$ — одноточечное множество.

В случае $x_{k+1} = R_S z_k$ имеем $t_k \|\xi_k\| \le t_k L \le \frac{\sqrt{3}}{2}R$, что, как отмечалось выше, гарантирует существование и единственность $R_S z_k$.

Обсудим очевидные свойства правил A1 и A2. Правило A1 не требует знания констант Липшица L_1 для f', L для f и константы проксимальной гладкости S для выбора t_k, в отличие от теоремы А. В силу теоремы А (см. (4)) при $t_k < t_0$ имеем

$$f(x_{k+1}) - f(x_k) \le -\|\xi_k\|^2 \left[t_k - \left(\frac{L}{R} + \frac{L_1}{2}\right)t_k^2\right],$$

правая часть последнего неравенства очевидно меньше $-\alpha t_k \|\xi_k\|^2$ при достаточно малых t_k. Следовательно, максимум в определении t_k по правилу Армихо A1 достигается. Очевидным недостатком является необходимость находить проекцию $P_S z_k$ точки z_k на множество S при переборе $m = 0, 1, \dots$.

Правило A2 требует точное знание (либо оценку снизу) константы R и оценки сверху для константы Липшица функции f. Преимуществом является необходимость всего одного проектирования вектора $f'(x_k)$ на подпространство T_{x_k}. Далее t_k ищется перебором $m = 0, 1, \dots$ аналогично правилу Армихо в безусловной минимизации. Когда шаг t_k найден, вычисляется x_{k+1}: $x_{k+1} = P_S z_k$ или $x_{k+1} = R_S z_k$.

3. СХОДИМОСТЬ МПГ В ЗАДАЧЕ (1) С ШАГОМ АРМИХО

3.1. Шаг Армихо A1

Теорема 1. *Предположим, что в задаче* (1) *функция f липшицева с константой L, функция f' липшицева с константой L_1, S — проксимально гладкое множество с константой $\frac{\pi}{2}R$. Числа L, L_1, R неизвестны. Пусть выполнено условие ЛПЛ для функции f на $S \cap \mathcal{L}_f(f(x_0))$.*

Определим $B = \min\left\{d; \frac{\beta(1-\alpha)}{C}\right\}$, где $C = \frac{L_1}{2} + \frac{L}{R}$. Тогда алгоритм (2) *с выбором шага A*1 *сходится с линейной скоростью по функции* (8) *и по точке* (9).

Доказательство. Рассмотрим алгоритм с шагом A1 и случаи *а* и *б*:

(*а*) Если $\alpha t_k > t_k - Ct_k^2$, т.е. $t_k > \frac{1-\alpha}{C}$, то на k-м шаге имеем

$$f(x_{k+1}) - f(x_k) \le -\frac{\alpha(1-\alpha)}{C}\|\xi_k\|^2.$$

(*б*) Пусть $t_k - Ct_k^2 \ge \alpha t_k$, т.е. $t_k \le \frac{1-\alpha}{C} < \frac{1}{C}$.

Если $d > \frac{1-\alpha}{C}$, то $t_k \ge \frac{(1-\alpha)\beta}{C}$ в силу определения шага Армихо A1.

Если $d \le \frac{1-\alpha}{C}$, то $t_k = d$ (и $m = 0$). Покажем это. По теореме А (4) при $t_k < \frac{1}{C}$ (оценка (4) верна для всех t_k)

$$f(x_{k+1}) - f(x_k) \le -\|\xi_k\|^2 (t_k - Ct_k^2) \le -\alpha t_k \|\xi_k\|^2. \tag{7}$$

Максимальное t_k, удовлетворяющее предыдущей формуле, равно d.

Таким образом, в случаях (а) и (б), которые исчерпывают весь диапазон $t_k > 0$,

$$t_k \ge B := \min\left\{d; \frac{\beta(1-\alpha)}{C}\right\}.$$

В силу неравенства ЛПЛ

$$\|\xi_k\|^2 \ge \mu(f(x_k) - f_*), \quad \text{где} \quad f_* = \inf_{x\in S} f(x),$$

получаем, что

$$f(x_{k+1}) - f(x_k) \le -\alpha\mu B(f(x_k) - f_*).$$

Для функции $\varphi(x) = f(x) - f_*$ имеем

$$\varphi(x_{k+1}) \le (1 - \alpha\mu B)\,\varphi(x_k), \quad B = \min\left\{d; \frac{\beta(1-\alpha)}{C}\right\}. \tag{8}$$

Для сходимости по точке с учетом формулы

$$\alpha B\|\xi_k\|^2 \le f(x_k) - f(x_{k+1}) \le \varphi(x_k)$$

и неравенства $\|\xi_k\| \le L$ имеем

$$\begin{gathered} \|x_{k+1} - x_k\| \le 2\|z_k - x_k\| = 2t_k\|\xi_k\| \le 2d\|\xi_k\|, \\ \|x_{k+1} - x_k\|^2 = 4d^2\|\xi_k\|^2 \le \frac{4d^2}{\alpha B}\varphi(x_k). \end{gathered} \tag{9}$$

Теорема доказана.

3.2. Шаг Армихо А2

Теорема 2. *Пусть в задаче* (1) *известна константа* $\frac{\pi}{2}R$ *проксимальной гладкости множества* S (*или ее оценка снизу*), *а также известна оценка сверху* L *константы Липшица* f. *Пусть* f' *— липшицева функция с неизвестной константой* L_1. *Предположим, что выполнено условие ЛПЛ для функции* f *на* $S \cap \mathcal{L}_f(f(x_0))$.

Пусть $\alpha_1 \in (0, 1)$, $\alpha > \alpha_1$ *и* $d \le \alpha_1 \frac{\sqrt{3}R}{2L}$. *Тогда алгоритм* (2) *с выбором шага А2 сходится с линейной скоростью по функции* (14) *и по точке* (15) (*для* R_S) *или* (16) (*для* P_S).

Доказательство. Рассмотрим ретракцию R_S. Для $z_k = x_k - t_k\xi_k$ имеем

$$f(z_k) - f(x_k) \le (f'(x_k), z_k - x_k) + \frac{L_1}{2}\|z_k - x_k\|^2 =$$
$$= -t_k\|\xi_k\|^2 + \frac{L_1}{2}t_k^2\|\xi_k\|^2 = -\|\xi_k\|^2\left(t_k - \frac{L_1}{2}t_k^2\right).$$

Рассмотрим альтернативы (а) и (б):

(а) $\alpha t_k > t_k - \frac{L_1}{2}t_k^2$, т.е. $t_k > \frac{2(1-\alpha)}{L_1}$,

(б) $\alpha t_k \le t_k - \frac{L_1}{2}t_k^2$, т.е. $t_k \le \frac{2(1-\alpha)}{L_1}$.

Аналогично п. (б) теоремы 1, $t_k \ge \frac{2\beta(1-\alpha)}{L_1}$ в случае (*б*) при условии $d \ge \frac{2(1-\alpha)}{L_1}$. Если $d < \frac{2(1-\alpha)}{L_1}$, то, опять же аналогично доказательству теоремы 1, $t_k = d$ и $m = 0$.

Итак, в любом случае $t_k \ge E := \min\left\{d; \frac{2\beta(1-\alpha)}{L_1}\right\}$.

Имеем

$$f(x_{k+1}) - f(x_k) = f(x_{k+1}) - f(z_k) + f(z_k) - f(x_k),$$
$$f(z_k) - f(x_k) \le -\alpha t_k\|\xi_k\|^2. \tag{10}$$

По теореме Пифагора $\|x_{k+1} - x_k\|^2 = \|x_k - z_k\|^2 + \|z_k - x_{k+1}\|^2$. При этом при условии $t_k\|\xi_k\| < \frac{\sqrt{3}}{2}R$ (см. [13, Theorem 2 (20)])

$$\|z_k - x_{k+1}\| \le \frac{\|x_k - z_k\|^2}{R}. \tag{11}$$

В силу (11)

$$f(x_{k+1}) - f(z_k) \le L\|x_{k+1} - z_k\| \le \frac{L}{R}\|x_k - z_k\|^2 \le \frac{L}{R}t_k^2\|\xi_k\|^2. \tag{12}$$

Из (10) и (12) получаем

$$f(x_{k+1}) - f(x_k) \le -\alpha t_k\|\xi_k\|^2 + \frac{L}{R}t_k^2\|\xi_k\|^2 = -t_k\|\xi_k\|^2\left(\alpha - \frac{L}{R}t_k\right).$$

В силу условий $d \le \alpha_1\frac{\sqrt{3}R}{2L}$, $\alpha > \alpha_1$ имеем

$$t_k \le d < \alpha_1\frac{R}{L}, \quad \alpha - \frac{L}{R}t_k > \alpha - \frac{L}{R}\alpha_1\frac{R}{L} = \alpha - \alpha_1 > 0,$$
$$-t_k\|\xi_k\|^2\left(\alpha - \frac{L}{R}t_k\right) \le -t_k\|\xi_k\|^2(\alpha - \alpha_1).$$

Вспоминая, что $t_k \geq E$, окончательно получаем

$$f(x_{k+1}) - f(x_k) \leq -E(\alpha - \alpha_1)\|\xi_k\|^2. \tag{13}$$

По условию ЛПЛ $\|\xi_k\|^2 \geq \mu(f(x_k) - f_*)$ и для $\varphi(x) = f(x) - f_*$ имеем

$$\varphi(x_{k+1}) \leq (1 - E(\alpha - \alpha_1)\mu)\varphi(x_k), \qquad E = \min\left\{d; \frac{2\beta(1-\alpha)}{L_1}\right\}. \tag{14}$$

С учетом формулы (11)

$$\|x_{k+1} - x_k\|^2 = \|x_k - z_k\|^2 + \|z_k - x_{k+1}\|^2 \leq t_k^2\|\xi_k\|^2 + \frac{\|x_k - z_k\|^4}{R^2} \leq$$
$$\leq D\|\xi_k\|^2, \quad \text{где} \quad D = d^2 + \frac{d^4L^2}{R^2}.$$

Применяя (13), получаем

$$\|x_{k+1} - x_k\|^2 \leq D\|\xi_k\|^2 \leq \frac{D}{E(\alpha - \alpha_1)}(f(x_k) - f(x_{k+1})) \leq \frac{D}{E(\alpha - \alpha_1)}\varphi(x_k). \tag{15}$$

Заметим, что для ретракции P_S оценка (14) также остается верной в силу теоремы А.

Для сходимости по точке с учетом формулы (11) и неравенства $\|\xi_k\| \leq L$

$$\|x_{k+1} - x_k\| \leq \|x_{k+1} - z_k\| + \|z_k - x_k\| \leq \|R_S z_k - z_k\| + \|z_k - x_k\| \leq \frac{\|z_k - x_k\|^2}{R} + \|z_k - x_k\|,$$
$$\|x_{k+1} - x_k\| \leq t_k\|\xi_k\| + \frac{t_k^2\|\xi_k\|^2}{R}.$$

Отсюда с учетом формулы (13) получаем

$$\|x_{k+1} - x_k\|^2 \leq \|\xi_k\|^2 d^2\left(1 + \frac{dL}{R}\right)^2 \leq \frac{d^2\left(1 + \frac{dL}{R}\right)^2}{E(\alpha - \alpha_1)}\varphi(x_k). \tag{16}$$

Теорема доказана.

ПРИЛОЖЕНИЕ

Лемма 1. *Пусть $S \subset \mathbb{R}^n$, $x \in \mathbb{R}^n$, $x \notin S$. Пусть существует $y \in P_S(x)$ и существует число $\lambda > 0$ такое, что $y \in P_S(x + \lambda(x - y))$. Тогда множество $P_S(x)$ является одноточечным.*

Зафиксируем вещественное число $\sigma_0 > 0$.

Теорема 3. *Константа проксимальной гладкости множества $\mathfrak{M}_r$ ($r < k$) в точности равна $\frac{\sigma_0}{\sqrt{2}}$.*

Доказательство. Рассмотрим произвольную матрицу $Y \in \mathbb{R}^{n\times k}$. Пусть ее сингулярное разложение задается формулой $Y = U_Y \Sigma V_Y^{\mathrm{T}}$. Определим матрицу $\Lambda \in \mathbb{R}^{n\times k}$ следующим образом: $\Lambda_{ii} = \max(\sigma_0, \sigma_i(Y))$ $\quad \forall i = \overline{1, r}$, остальные элементы матрицы Λ положим равными 0. Докажем, что матрица $Z \in \mathbb{R}^{n\times k}$, сингулярное разложение которой задано формулой $Z = U_Y \Lambda V_Y^{\mathrm{T}}$, принадлежит множеству $P_{\mathfrak{M}_r}(Y)$. Рассмотрим произвольную матрицу $X \in \mathfrak{M}_r$, имеющую сингулярное разложение $X = U\Lambda_1 V^{\mathrm{T}}$. Тогда с учетом формулы (6) верна следующая цепочка равенств и неравенств:

$$\|Y - Z\|^2 = \left\|U_Y\Sigma V_Y^{\mathrm{T}} - U_Y\Lambda V_Y^{\mathrm{T}}\right\|^2 = \left\|\Sigma - U_Y^{\mathrm{T}}U_Y\Lambda V_Y^{\mathrm{T}}V_Y\right\|^2 = \|\Sigma - \Lambda\|^2 =$$
$$= \sum_{i=1}^{k}(\sigma_i(\Sigma) - \sigma_i(\Lambda))^2 \leq \sum_{i=1}^{k}(\sigma_i(\Sigma) - \sigma_i(\Lambda_1))^2 = \sum_{i=1}^{k}(\sigma_i(Y) - \sigma_i(X))^2 \leq \|Y - X\|^2.$$

Таким образом, $Z \in P_{\mathfrak{M}_r}(Y)$, а $\varrho^2(Y, \mathfrak{M}_r) = \|Y - Z\|^2 = \sum_{i=1}^{k}(\sigma_i(\Sigma) - \sigma_i(\Lambda))^2$.

Пусть верно $\varrho(Y, \mathfrak{M}_r) < \frac{\sigma_0}{\sqrt{2}}$. Докажем, что $\sigma_r(Y) \neq \sigma_{r+1}(Y)$. Предположим противное, т.е., что $\sigma_r(Y) = \sigma_{r+1}(Y)$. Рассмотрим две альтернативы. Если $\sigma_r(Y) \geq \sigma_0$, то $\sum_{i=1}^{k}(\sigma_i(\Sigma) - \sigma_i(\Lambda))^2 \geq \sigma_0^2$, а значит, $\varrho(Y, \mathfrak{M}_r) \geq \sigma_0$, что противоречит $\varrho(Y, \mathfrak{M}_r) < \frac{\sigma_0}{\sqrt{2}}$. Если $\sigma_r(Y) < \sigma_0$, то $\sum_{i=1}^{k}(\sigma_i(\Sigma) - \sigma_i(\Lambda))^2 \geq (\sigma_r(Y) - \sigma_0)^2 + (\sigma_{r+1}(Y) - 0)^2 = (\sigma_r(Y) - \sigma_0)^2 + (\sigma_r(Y) - 0)^2 \geq \frac{\sigma_0^2}{2}$, что противоречит $\varrho(Y, \mathfrak{M}_r) < \frac{\sigma_0}{\sqrt{2}}$. Таким образом, $\sigma_r(Y) \neq \sigma_{r+1}(Y)$. Рассмотрим матрицы $Y(\lambda) = Y + \lambda(Y - Z) = U_Y(\Sigma + \lambda(\Sigma - \Lambda))V_Y^{\mathrm{T}}$ для любого $\lambda \in \mathbb{R}$. В силу непрерывности операций сложения и умножения на скаляр существует такое число $\hat{\lambda} > 0$, что выражение $U_Y(\Sigma + \hat{\lambda}(\Sigma - \Lambda))V_Y^{\mathrm{T}}$ является сингулярным разложением матрицы $Y(\hat{\lambda})$ (это следует из того, что $\sigma_r(Y) \neq \sigma_{r+1}(Y)$), и при этом $\max(\sigma_0, \sigma_i(Y(\hat{\lambda}))) = \max(\sigma_0, \sigma_i(Y) + \hat{\lambda}(\sigma_i(Y) - \max(\sigma_0, \sigma_i(Y)))$ $\forall i = \overline{1, r}$. Итак, $Z \in P_{\mathfrak{M}_r}(Y(\hat{\lambda}))$. По лемме 1 получаем, что множество $P_{\mathfrak{M}_r}(Y)$ является одноточечным. Таким образом, множество $\mathfrak{M}_r$ $(r < k)$ является проксимально гладким с $R_{\mathfrak{M}_r} = \frac{\sigma_0}{\sqrt{2}}$.

Докажем, что константа проксимальной гладкости множества $\mathfrak{M}_r$ $(r < k)$ неулучшаема. Рассмотрим матрицу Σ, заданную следующим образом: $\Sigma_{ii} = \sigma_0$ $\forall i = \overline{1, r-1}$, $\Sigma_{ii} = \frac{\sigma_0}{2}$ $\forall i = \overline{r, r+1}$, остальные элементы матрицы Σ положим равными 0. Аналогично приведенным выше рассуждениям (используя формулу (6)) получаем, что $\varrho(\Sigma, \mathfrak{M}_r) = \frac{\sigma_0}{\sqrt{2}}$. Рассмотрим матрицы M и N, заданные следующим образом: $M_{ii} = \sigma_0$ $\forall i = \overline{1, r}$, остальные элементы матрицы M положим равными 0; $N_{ii} = \sigma_0$ $\forall i = \overline{1, r-1}$, $N_{r+1,r+1} = \sigma_0$, остальные элементы матрицы N положим равными 0. Заметим (пользуясь формулой (6)), что $M, N \in P_{\mathfrak{M}_r}(\Sigma)$. Таким образом, $R_{\mathfrak{M}_r} = \frac{\sigma_0}{\sqrt{2}}$, причем эта константа неулучшаема. Теорема доказана.

Теорема 4. *Константа проксимальной гладкости множества* $\mathfrak{M}_k$ *в точности равна* σ_0.

Доказательство. Рассмотрим произвольную матрицу $Y \in \mathbb{R}^{n \times k}$, для которой верно $\varrho(Y, \mathfrak{M}_k) < \sigma_0$. Пусть ее сингулярное разложение задается формулой $Y = U_Y \Sigma V_Y^{\mathrm{T}}$. Аналогично доказательству теоремы 3 получаем, что верно неравенство $\sigma_k(Y) > 0$, которое обеспечивает одноточечность множества $P_{\mathfrak{M}_k}(Y)$. Таким образом, множество $\mathfrak{M}_k$ является проксимально гладким с $R_{\mathfrak{M}_k} = \sigma_0$.

Покажем, что константа проксимальной гладкости множества $\mathfrak{M}_k$ неулучшаема. Рассмотрим матрицу Σ, заданную следующим образом: $\Sigma_{ii} = \sigma_0$ $\forall i = \overline{1, k-1}$, остальные элементы матрицы Σ положим равными 0. Аналогично доказательству теоремы 3 получаем, что $\varrho(Y, \mathfrak{M}_k) = \sigma_0$. Рассмотрим матрицы M и N, заданные следующим образом: $M_{ii} = \sigma_0$ $\forall i = \overline{1, k}$, остальные элементы матрицы M положим равными 0; $N_{ii} = \sigma_0$ $\forall i = \overline{1, k-1}$, $N_{k,k} = -\sigma_0$, остальные элементы матрицы N положим равными 0. Заметим (пользуясь формулой (6)), что $M, N \in P_{\mathfrak{M}_k}(\Sigma)$. Таким образом, $R_{\mathfrak{M}_k} = \sigma_0$, причем эта константа неулучшаема. Теорема доказана.

Перейдем к рассмотрению множеств $\mathfrak{L}_r$. Заметим, что $\mathfrak{L}_1 = \mathfrak{M}_1$.

Теорема 5. *Константа проксимальной гладкости множества* $\mathfrak{L}_r$ $(r > 1)$ *в точности равна* $\frac{\sigma_0}{2}$.

Доказательство. Рассмотрим произвольную матрицу $Y \in \mathbb{R}^{n \times k}$, для которой верно $\varrho(Y, \mathfrak{L}_r) < \frac{\sigma_0}{2}$. Пусть ее сингулярное разложение задается формулой $Y = U_Y \Sigma V_Y^{\mathrm{T}}$. Вследствие того, что $\mathfrak{L}_r = \bigcup_{i=1}^r \mathfrak{M}_i$, множество $P_{\mathfrak{L}_r}(Y)$ может оказаться неодноточечным только в следующих случаях:

(*а*) Существует $i \in \{1, 2, \ldots, r\}$ такое, что множество $P_{\mathfrak{M}_i}(Y)$ является неодноточечным и при этом $P_{\mathfrak{M}_i}(Y) \subset P_{\mathfrak{L}_r}(Y)$. Из теорем 3 и 4 $\varrho(Y, \mathfrak{M}_i) \geq \frac{\sigma_0}{\sqrt{2}}$, а следовательно, $\varrho(Y, \mathfrak{L}_r) \geq \frac{\sigma_0}{\sqrt{2}}$, что противоречит $\varrho(Y, \mathfrak{L}_r) < \frac{\sigma_0}{2}$.

(*б*) Существуют $i, j \in \{1, 2, \ldots, r\}$, $i < j$, такие, что $\varrho(Y, \mathfrak{M}_i) = \varrho(Y, \mathfrak{M}_j) = \varrho(Y, \mathfrak{L}_r)$. Из доказательства теоремы 3 матрицы $Z_l = U_Y \Lambda_l V_Y^{\mathrm{T}}$, где матрицы Λ_l заданы следующим образом: $(\Lambda_l)_{mm} = \max(\sigma_0, \sigma_m(Y))\ \forall m = \overline{1, l}$, а их остальные элементы равны 0, принадлежат $P_{\mathfrak{M}_l}(Y)$, $l \in \{i, j\}$. Покажем, что $\max(\sigma_m, \sigma_0) = \sigma_0\ \ \forall m \in i+1, \ldots, j$. Действительно, иначе $\varrho(Y, \mathfrak{M}_i) = \|Y - Z_i\| \geq \sigma_0$, что противоречит $\varrho(Y, \mathfrak{L}_r) < \frac{\sigma_0}{2}$. Учитывая это, получаем

$$
\begin{aligned}
0 = \|Y - Z_j\|^2 - \|Y - Z_i\|^2 &= \sum_{m=1}^{k} \left(\sigma_m(\Sigma) - \sigma_m(\Lambda_j)\right)^2 - \sum_{m=1}^{k} \left(\sigma_m(\Sigma) - \sigma_m(\Lambda_i)\right)^2 = \\
&= \sum_{m=i+1}^{j} \left((\sigma_m(\Sigma) - \sigma_m(\Lambda_j))^2 - \sigma_m^2(\Sigma)\right) = \sum_{m=i+1}^{j} \left(\sigma_m^2(\Lambda_j) - 2\sigma_m(\Sigma)\sigma_m(\Lambda_j)\right) = \\
&= \sum_{m=i+1}^{j} \left((\sigma_0^2 - 2\sigma_m(\Sigma)\sigma_0)\right).
\end{aligned}
$$

Таким образом, $\sum_{m=i+1}^{j} \sigma_m(\Sigma) = \frac{N\sigma_0}{2}$, где $N = j - i$. По неравенствам между средними $\sum_{m=i+1}^{j} \sigma_m^2(\Sigma) \geq \frac{\left(\sum_{m=i+1}^{j} \sigma_m(\Sigma)\right)^2}{N} = \frac{N\sigma_0^2}{4}$. Используя это, получаем

$$
\begin{aligned}
\varrho^2(Y, \mathfrak{L}_r) = \varrho^2(Y, \mathfrak{M}_i) = \sum_{m=1}^{k} (\sigma_m(\Sigma) - \sigma_m(\Lambda_i))^2 &= \sum_{m=1}^{i} (\sigma_m(\Sigma) - \max(\sigma_m(\Sigma), \sigma_0))^2 + \\
&+ \sum_{m=i+1}^{j} \sigma_m^2(\Sigma) + \sum_{j+1}^{k} \sigma_m^2(\Sigma) \geq \frac{N\sigma_0^2}{4} \geq \frac{\sigma_0^2}{4},
\end{aligned}
$$

что противоречит $\varrho(Y, \mathfrak{L}_r) < \frac{\sigma_0}{2}$. Итак, для любой матрицы $Y \in \mathbb{R}^{n \times k}$, для которой верно $\varrho(Y, \mathfrak{L}_r) < \frac{\sigma_0}{2}$, множество $P_{\mathfrak{L}_r}(Y)$ является одноточечным. Таким образом, множество $\mathfrak{L}_r$ $(r > 1)$ является проксимально гладким с $R_{\mathfrak{M}_r} = \frac{\sigma_0}{2}$.

Докажем, что константа проксимальной гладкости множества $\mathfrak{L}_r$ $(r > 1)$ неулучшаема. Рассмотрим матрицу Σ, заданную следующим образом: $\Sigma_{ii} = \sigma_0\ \forall i = \overline{1, r-1}$, $\Sigma_{rr} = \frac{\sigma_0}{2}$, остальные элементы положим равными 0. Аналогично доказательству теоремы 3 получаем

$$
\varrho^2(\Sigma, \mathfrak{M}_i) = \begin{cases} (r - 1 - i)\sigma_0^2 + \dfrac{\sigma_0^2}{4}, & i \in \{1, 2, \ldots, r-1\}, \\ \dfrac{\sigma_0^2}{4}, & i = r. \end{cases}
$$

Итак, $\varrho(\Sigma, \mathfrak{L}_r) = \varrho(\Sigma, \mathfrak{M}_{r-1}) = \varrho(\Sigma, \mathfrak{M}_r) = \frac{\sigma_0}{2}$. Таким образом, $R_{\mathfrak{L}_r} = \frac{\sigma_0}{2}$, причем эта константа неулучшаема. Теорема доказана.

Пусть $Y \in \mathbb{R}^{n\times k}$ – произвольная матрица. Пусть ее сингулярное разложение задается формулой $Y = U_Y \Sigma V_Y^{\mathrm{T}}$. Определим $\forall l = \overline{1,k}$ матрицы Λ_l следующим образом: $(\Lambda_l)_{mm} = \max(\sigma_0, \sigma_m(Y))$ $\forall m = \overline{1,l}$, а их остальные элементы положим равными 0. Определим матрицы Z_l следующим образом: $Z_l = U_Y \Lambda_l V_Y^{\mathrm{T}}$ $\forall l = \overline{1,k}$.

Следствие 1. Матрица $Z_l \in \mathbb{R}^{n\times k}$ принадлежит множеству $P_{\mathfrak{M}_l}(Y)$ $\forall l = \overline{1,k}$.

Следствие 2. Матрица $Z_l \in \mathbb{R}^{n\times k}$, для которой $\|Y - Z_l\|$ минимально ($l = \overline{1,r}$), принадлежит множеству $P_{\mathfrak{L}_r}(Y)$.

СПИСОК ЛИТЕРАТУРЫ

1. *Goldstein A.A.* Convex programming in Hilbert space // Bull. Amer. Math. Soc. 1964. V. 70. № 5. P. 709–710.
2. *Levitin E.S., Polyak B.T.* Constrained minimization methods // Zh. Vychisl. Mat. Mat. Fiz. 1966. V. 6. № 5. P. 787–823.
3. *Absil P.-A., Malick J.* Projection-like retraction on matrix manifolds // SIAM J. Optim. 2012. V. 22. № 1. P. 135–158.
4. *Edelman A., Arias T., Smith S.T.* The geometry of algorithms with orthogonality constraints // J. Matrix Anal. Appl. 1998. V. 20. № 2. P. 303–353.
5. *Absil P.-A., Mahony R., Sepulchre R.* Matrix Manifolds. Princeton Univ. Press, Princeton and Oxford, 2008. 240 p.
6. *Luenberger D.G.* The gradient projection methods along geodesics // Management Sci. 1972. V. 18. № 11. P. 620–631.
7. *Hager W.W.* Minimizing a quadratic over a sphere // SIAM J. Optim. and Contr. 2001. V. 12. № 1. P. 188–208.
8. *Neto J.X. da Cruz, De Lima J.X. da Cruz, Oliveira P.R.* Geodesic algorithms on Riemannian manifolds // Balkan J. of Geom. and its Appl. 1998. V. 3. № 2. P. 89–100.
9. *Udrişte C.* Convex Functions and Optimization Methods on Riemannian Manifolds // Math. and Its Appl. Ser. Springer, 1998. V. 297.
10. *Schneider R., Uschmajew A.* Convergence results for projected line search methods on varieties of low-rank matricies via Lojasiewicz inequality // SIAM J. Optim. 2015. V. 25. № 1. P. 622–646.
11. *Merlet B., Nguyen T.N.* Convergence to equilibrium for discretizations of gradient-like flows on Riemannian manifolds // Different. Integral Equat. 2013. V. 26. P. 571–602.
12. *Birgin E.G., Martinez J.M., Raydan M.* Nonmonotone Spectral Projected Gradient Methods on Convex Sets // SIAM J. Optim. 2000. V. 10. № 4. P. 1196–1211.
13. *Balashov M.V.* The Gradient Projection Algorithm for Smooth Sets and Functions in Nonconvex Case // Set-Valued and Variat. Anal. 2021.V. 29.P. 341–360. https://doi.org/10.1007/s11228-020-00550-4
14. *Huikang Liu, Anthony Man-Cho So, Weijie Wu* Quadratic optimization with orthogonality constraint: explicit Lojasiewicz exponent and linear convergence of retraction-based line-search and stochastic variance-reduced gradient methods // Math. Program. 2018. V. 178. P. 215–262.
15. *Nesterov Yu.* Introductory lectures on convex optimization. Abasic course basic course. Berlin: Springer, 2004.
16. *Balashov M.V., Polyak B.T., Tremba A.A.* Gradient Projection and Conditional Gradient Methods for Constrained Nonconvex Minimization // Numerical Function. Anal. and Optimizat. 2020. V. 41. № 7. P. 822–849.
17. *Horn R., Johnson C.* Matrix Analysis. New York, NY, USA: Cambridge Univ. Press, 2009. 643 p.
18. *Vial J.-Ph.* Strong and weak convexity of sets and functions // Math. of Operat. Res. 1983. V. 8. № 2. P. 231–259.
19. *Clarke F.H., Stern R.J., Wolenski P.R.* Proximal smoothness and lower–C^2 property // J. Convex Anal. 1995. V. 2. № 1–2. P. 117–144.
20. *Conway J.H., Hardin R.H., Sloane N.J.A.* Packing Lines, Planes, etc.: Packings in Grassmannian Spaces // Experiment. Math. 1996. V. 5. P. 139–159.
21. *Балашов М.В.* Метод проекции градиента на матричных многообразиях // Ж. вычисл. матем. и матем. физ. 2020. Т. 60. № 9. С. 1453–1461.