*ЖУРНАЛ ВЫЧИСЛИТЕЛЬНОЙ МАТЕМАТИКИ И МАТЕМАТИЧЕСКОЙ ФИЗИКИ, 2020, том 60, № 9, с. 1453–1461*

*УДК 519.853.6, 519.853.4*

# МЕТОД ПРОЕКЦИИ ГРАДИЕНТА НА МАТРИЧНЫХ МНОГООБРАЗИЯХ[1)]

**© 2020 г. М. В. Балашов**

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

*e-mail: balashov73@mail.ru*

Поступила в редакцию 26.11.2019 г.
Переработанный вариант 24.12.2019 г.
Принята к публикации 09.04.2020 г.

Рассматривается задача минимизации функции с непрерывным по Липшицу градиентом на проксимально гладком подмножестве конечномерного евклидова пространства. При выполнении условия RSI (Restricted Secant Inequality) метод проекции градиента для рассматриваемой задачи сходится с линейной скоростью. В определенных случаях доказывается линейная скорость сходимости метода проекции градиента на вещественном многообразии Штифеля или Грассмана. Библ. 21.

**Ключевые слова:** непрерывный по Липшицу градиент, проксимальная гладкость, метод проекции градиента, метрическая проекция, невыпуклая экстремальная задача, Restricted Secant Inequality, многообразие Штифеля, многообразие Грассмана.

**DOI:** 10.31857/S0044466920090070

## 1. ВВЕДЕНИЕ И ОСНОВНЫЕ ОБОЗНАЧЕНИЯ

### *1.1. Введение*

Метод проекции градиента (МПГ) (известный также как проекционно-проксимальный метод и др.) для решения задачи

$$\min_A f(x) \tag{1}$$

был предложен в [1], [2]. Если функция в (1) сильно выпуклая с непрерывным по Липшицу градиентом и множество выпукло и замкнуто, то МПГ сходится со скоростью геометрической прогрессии (или линейной скоростью). МПГ доказал свою эффективность для решения различных экстремальных задач [2], [3].

Примером актуальной невыпуклой задачи (1) является задача минимизации гладкой функции $f$ на некотором матричном многообразии $A$ без края. Например, $A$ может быть многообразием Штифеля или Грассмана [4].

В случае, если множество $A$ в задаче (1) является гладким многообразием, традиционно используются варианты метода проекции градиента с шагами, связанными с локальными геодезическими на многообразии [5]–[9]. Метод Ньютона также находит применение [4], [6] для решения упомянутой задачи, главным образом с квадратичными функциями или ограничениями. Однако метод Ньютона сходится в общем случае локально и существует очень немного работ с явными оценками для его области сходимости в случае невыпуклой задачи (1). Последнее верно и для МПГ.

В данной статье мы планируем сформулировать МПГ для задачи минимизации гладкой невыпуклой в общем случае функции на вещественном многообразии Штифеля или Грассмана. Ключевым в работе является тот факт, что вещественное многообразие Штифеля есть *проксимально гладкое* множество с константой проксимальной гладкости 1, а вещественное многообразие Грассмана также проксимально гладкое множество с константой $1/\sqrt{2}$. Заметим, что в силу соображений компактности любое компактное гладкое многообразие без края является проксимально гладким. Таким образом, речь идет о нахождении наилучшей (т.е. наибольшей) константы.

[1)]Работа выполнена при финансовой поддержке РНФ, проект 16-11-10015.

Кроме того, в теоремах 2 и 4 работы сформулированы достаточные условия сходимости МПГ с линейной скоростью на многообразиях Штифеля и Грассмана соответственно.

В отличие от упомянутых выше работ, свойства проксимально гладких множеств позволяют предъявить некоторый вариант МПГ с простой и явной оценкой скорости сходимости. Кроме того, наши результаты не зависят от геодезических на многообразиях. Это является еще одним преимуществом предложенного подхода. За это мы платим погружением задачи (1) в евклидово пространство большей (по сравнению с размерностью многообразия) размерности.

### *1.2. Основные обозначения*

Обозначим евклидово пространство размерности $n$ через $\mathbb{R}^n$ и скалярное произведение $(\cdot,\cdot)$. Определим $B_R(a) = \{x \in \mathbb{R}^n \,|\, \|x - a\| \le R\}$, $a \in \mathbb{R}^n$, $R > 0$. Функция $f : \mathbb{R}^n \to \mathbb{R}$ называется *сильно выпуклой с константой* $\varkappa > 0$, если функция $f(x) - \frac{\varkappa}{2}\|x\|^2$ выпуклая. Мы будем говорить, что функция $f$ *гладкая,* если ее градиент Фреше $f'$ непрерывен по Липшицу с константой $L_1$. Для функции $f : \mathbb{R}^n \to \mathbb{R}$ и $\beta \in \mathbb{R}$ определим нижнее множество уровня $\mathcal{L}_f(\beta) = \{x \in \mathbb{R}^n \,|\, f(x) \le \beta\}$. Определим функцию расстояния $\varrho_A(x) = \inf_{a \in A}\|x - a\|$. Метрическая проекция точки $x \in \mathbb{R}^n$ на множество $A \subset \mathbb{R}^n$ есть

$$P_A x = \left\{a \in A \,|\, \|x - a\| = \varrho_A(x)\right\}.$$

Замкнутое множество $A \subset \mathbb{R}^n$ называется *проксимально гладким* (или *прокс-регулярным, слабо выпуклым*) с константой $R$, если функция расстояния $\varrho_A(x)$ непрерывно дифференцируема на множестве $U_A(R) = \{x \in \mathbb{R}^n \,|\, 0 < \varrho_A(x) < R\}$. Свойства таких множеств рассматривались разными авторами [10], [11]. Мы хотим подчеркнуть, что в нашей работе проксимально гладкие множества по определению замкнуты. Для проксимально гладкого множества $A$ и точки $x \in A$ обозначим через $N(A, x)$ конус проксимальных нормалей

$$N(A, x) = \{p \in \mathbb{R}^n \,|\, \exists t = t(p) > 0 \quad P_A(x + tp) = x\}.$$

Эквивалентным свойством для проксимальной гладкости замкнутого множества $A$ с константой $R$ является *опорный принцип* [10], [11]: для любой точки $x \in \partial A$ и всякого единичного вектора нормали $p(x) \in N(A, x)$ мы имеем $A \cap \operatorname{int} B_R(x + Rp(x)) = \emptyset$. В конечномерном пространстве $\mathbb{R}^n$ последнее эквивалентно единственности метрической проекции $P_A x$ для каждой точки $x \in U_A(R)$, так как одноточечная метрическая проекция в конечномерном пространстве непрерывна [12, Ch. 3, § 1, Proposition 23].

Для проксимально гладкого множества $A \subset \mathbb{R}^n$ с константой $R$ и любых точек $x_1, x_2 \in A \cup U_A(r)$, $r < R$, для проекций $a_i = P_A x_i$ мы имеем оценку (см. [11])

$$\|a_1 - a_2\| \le \frac{R}{R - r}\|x_1 - x_2\|. \tag{2}$$

Скажем, что $x \in A$ является стационарной точкой функции $f$ на проксимально гладком множестве $A \subset \mathbb{R}^n$, если существует число $t > 0$ со свойством $P_A(x - tf'(x)) = x$. Последнее эквивалентно включению $f'(x) \in -N(A, x)$. Данное условие является необходимым условием экстремума с проксимально гладким множеством [13, Appendix 5.1].

Пусть $I_k \in \mathbb{R}^{k\times k}$ – единичная матрица и $O(k)$ есть множество ортогональных матриц размера $k \times k$. Определим скалярное произведение $(X, Y) = \operatorname{tr} X^{\mathrm{T}} Y$ для матриц $X, Y \in \mathbb{R}^{n\times k}$. Нормой Фробениуса матрицы $X \in \mathbb{R}^{n\times k}$ называется $\|X\|_F = \sqrt{\operatorname{tr} X^{\mathrm{T}} X}$. Далее мы будем считать, что $\|X\| = \|X\|_F$ для всякой матрицы $X$. Обозначим через $S_{n,k}$ вещественное многообразие Штифеля с натуральными параметрами $n \ge k$, т.е. $S_{n,k} = \{X \in \mathbb{R}^{n\times k} \,|\, X^{\mathrm{T}} X = I_k\}$. Очевидно, что $S_{n,k}$ вложено в пространство $\mathbb{R}^{n\times k}$ матриц с фробениусовой нормой. Обозначим через $G_{n,k}$ вещественное многообразие

Грассмана с натуральными параметрами $n \geq k$, т.е. множество всех $k$-мерных линейных подпространств в $\mathbb{R}^n$. Известно [14], что $G_{n,k}$ может быть вложено (в определенном смысле изометрично) в евклидово пространство симметричных $n \times n$ матриц $\mathrm{Sym}(n)$ размерности $\frac{1}{2}n(n+1)$ с нормой $\frac{1}{\sqrt{2}}\|\cdot\|_F$. Следуя [14], мы будем рассматривать указанную реализацию $G_{n,k}$ вида $\{XX^{\mathrm{T}} \mid X \in S_{n,k}\}$ с нормой $\|\cdot\|_F$. Заметим, что $XX^{\mathrm{T}}$ есть метрический проектор на подпространство $\mathrm{lin}\{X_1, \ldots, X_k\}$, где $X_i$ есть $i$-й столбец матрицы $X \in S_{n,k}$.

Легко видеть, что для всякой матрицы $X \in S_{n,k}$ имеем

$$\left\|XX^{\mathrm{T}} - \frac{k}{n} I_n\right\| = \sqrt{\frac{k(n-k)}{n}}.$$

Таким образом, константа проксимальной гладкости для $G_{n,k}$ не более чем $\sqrt{\frac{k(n-k)}{n}}$.

**Предложение 1.** *Для всякой матрицы* $X \in \mathbb{R}^{n\times k}$ *и* $Q \in O(n)$ $(Q \in O(k))$ *имеем* $\|QX\| = \|X\|$ $(\|XQ\| = \|X\|)$.

## 2. МИНИМИЗАЦИЯ ГЛАДКОЙ ФУНКЦИИ НА ПРОКСИМАЛЬНО ГЛАДКОМ МНОЖЕСТВЕ

### *2.1. Основная теорема о сходимости МПГ в невыпуклом случае*

**Определение 1.** Предположим, что $f : \mathbb{R}^n \to \mathbb{R}$ – дифференцируемая функция, $A \subset \mathbb{R}^n$ – замкнутое подмножество и $\Omega \subset A$, $\Omega \neq \emptyset$, есть множество решений для (1). Будем говорить, что условие Restricted Secant Inequality (RSI) выполняется для функции $f$ на множестве $A$, если существует такое число $\mu > 0$, что для любого $x \in A$ и любого $x_0 \in P_\Omega x$ выполняется неравенство

$$(f'(x) - f'(x_0), x - x_0) \geq \mu\|x - x_0\|^2.$$

Условие RSI хорошо известно в безусловной оптимизации [15]. Отметим, что указанное свойство выполнено для всякой сильно выпуклой функции с константой сильной выпуклости $\mu$ [15].

**Теорема 1.** *Пусть* $A \subset \mathbb{R}^n$ *– проксимально гладкое с константой* $R > 0$ *множество. Предположим, что функция* $f : \mathbb{R}^n \to \mathbb{R}$ *удовлетворяет условию RSI на множестве A с константой* $\mu > 0$, *и* $f$ *— гладкая функция с константой* $L_1 > 0$. *Положим* $L = \sup\limits_{x \in A \cap \mathcal{L}_f(\beta)} \|f'(x)\|$ *для некоторого* $\beta \in \mathbb{R}$. *Допустим, что* $\frac{L}{\mu} < R$ *и множество решений* $\Omega$ *в* (1) *непусто. Определим*

$$t_0 = \frac{\mu - \frac{L}{R}}{L_1^2 - \frac{\mu L}{R}}.$$

*Тогда для любой начальной точки* $x_0 \in A \cap \mathcal{L}_f(\beta)$ *и всякого* $t \in (0, t_0]$ *итерационный процесс* $x_{k+1} = P_A\left(x_k - tf'(x_k)\right)$ *сходится к решению задачи* (1) *с линейной скоростью* $\varrho_\Omega(x_k) \leq q^k(t)\varrho_\Omega(x_0)$, *причем* $\|x_{k+1} - x_k\| \leq q^k(t)(1 + q(t))\varrho_\Omega(x_0)$,

$$q(t) = \frac{R}{R - tL}\sqrt{1 - 2t\mu + t^2L_1^2} \in (0,1).$$

*Последовательность* $\{f(x_k)\}$ *нестрого монотонно убывает.*

**Доказательство.** Пусть $x_k^0 \in \Omega$ такая точка, что $\left\|x_k - x_k^0\right\| = \varrho_\Omega(x_k)$. Используя включения $x_k - tf'(x_k) \in U_A(R)$, $x_k^0 - tf'(x_k^0) \in U_A(R)$ (заметим, что $t < \frac{R}{L}$) имеем в силу условия Липшица для метрической проекции (2)

$$\left\|P_A\left(x_k - tf'(x_k)\right) - P_A(x_k^0 - tf'(x_k^0))\right\| \leq \frac{R}{R - tL}\left\|(x_k - tf'(x_k)) - (x_k^0 - tf'(x_k^0))\right\|,$$

и в силу условия (RSI) находим

$$\left\|x_{k+1} - x_{k+1}^0\right\|^2 \le \left\|x_{k+1} - x_k^0\right\|^2 = \left\|P_A(x_k - tf'(x_k)) - P_A(x_k^0 - tf'(x_k^0))\right\|^2 =$$
$$= \frac{R^2}{(R - tL)^2}\left(\left\|x_k - x_k^0\right\|^2 - 2t(f'(x_k) - f'(x_k^0), x_k - x_k^0) + t^2\left\|f'(x_k) - f'(x_k^0)\right\|^2\right) \le \tag{3}$$
$$\le \left\|x_k - x_k^0\right\|^2 \frac{R^2}{(R - tL)^2}(1 - 2\mu t + t^2L_1^2).$$

Таким образом, мы получаем оценку

$$\varrho_\Omega(x_{k+1}) = \left\|x_{k+1} - x_{k+1}^0\right\| \le \left\|x_k - x_k^0\right\|q(t) = \varrho_\Omega(x_k)q(t).$$

Используя неравенство треугольника

$$\|x_{k+1} - x_k\| \le \left\|x_{k+1} - x_k^0\right\| + \left\|x_k^0 - x_k\right\|,$$

и формулу (3), $\left\|x_{k+1} - x_k^0\right\| \le q(t)\left\|x_k - x_k^0\right\|$, мы получаем, что

$$\|x_{k+1} - x_k\| \le \left\|x_k - x_k^0\right\|(1 + q(t)) \le \ldots \le q^k(t)(1 + q(t))\varrho_\Omega(x_0).$$

Следовательно, $\{x_k\}$ есть последовательность Коши, которая сходится к некоторой точке $x_* \in \Omega$.

Из равенства $\|x_{k+1} - x_k + tf'(x_k)\| = \varrho_A(x_k - tf'(x_k))$ и оценки $t \le \frac{1}{L_1}$ находим

$$\left\|x_{k+1} - \left(x_k - \frac{1}{L_1}f'(x_k)\right)\right\| \le \|x_{k+1} - x_k + tf'(x_k)\| + \left(\frac{1}{L_1} - t\right)\|f'(x_k)\| \le \frac{1}{L_1}\|f'(x_k)\|.$$

Используя условие Липшица для $f'$, получаем, что

$$f(x_{k+1}) \le f(x_k) + (f'(x_k), x_{k+1} - x_k) + \frac{L_1}{2}\|x_{k+1} - x_k\|^2 =$$
$$= f(x_k) + \frac{L_1}{2}\left(\left\|x_{k+1} - \left(x_k - \frac{1}{L_1}f'(x_k)\right)\right\|^2 - \frac{1}{L_1^2}\|f'(x_k)\|^2\right) \le f(x_k) \le \beta$$

и $x_{k+1} \in A \cap \mathcal{L}_f(\beta)$. Таким образом, по индукции, $\{f(x_k)\}$ монотонно не возрастает и $x_k \in A \cap \mathcal{L}_f(\beta)$ для всех $k$. Теорема доказана.

Теорема 1 была доказана для случая сильно выпуклой функции в [16]. Включение $q(t) \in (0,1)$ для всех $t \in (0, t_0]$ также обосновано в указанной статье.

Если функция $f$ непрерывна по Липшицу на множестве $A$ с константой Липшица $L$, то

$$f(x_{k+1}) - \min_A f = f(x_{k+1}) - f(x_{k+1}^0) \le L\left\|x_{k+1} - x_{k+1}^0\right\| \le Lq^k(t)\,\mathrm{diam}\,A.$$

Пусть $A$ — гладкое многообразие без края и $T_x$ — касательное подпространство к $A$ в точке $x \in A$. В этом случае МПГ с промежуточной фазой проектирования на касательное подпространство $T_{x_k}$: $x_0 \in A$, $z_k = P_{T_{x_k}}(x_k - t_kf'(x_k))$, $t_k > 0$, $x_{k+1} = P_Az_k$, см. [13], в некоторых случаях может быть вычислительно экономичнее, чем алгоритм с прямым проектированием на множество $A$. Например, это выгоднее для матриц малого ранга [17, § 2.3]. С другой стороны, в этом случае мы должны вычислять проекцию $z_k$ точки на подпространство $T_{x_k}$, и это должно приниматься во внимание.

**Пример 1.** Рассмотрим дифференцируемую функцию $f$ с не равным нулю градиентом $\|f'(x)\|$ на множестве $A$: существует $l > 0$ такое, что $\|f'(x)\| \ge l$ для всех $x \in A$. Пусть $\beta \in \mathbb{R}$. Предположим также, что $f$ сильно квазивыпукла с константой $r > 0$ на множестве $\mathcal{L}_f(\beta)$, т.е. для всякого $\alpha \le \beta$ множество уровня $\mathcal{L}_f(\alpha)$ может быть представлено как пересечение замкнутых шаров радиуса $r$ (либо пусто). Допустим также, что множество $A$ есть гладкое и проксимально гладкое с константой $R$ многообразие и $R > \frac{rL}{l}$. Тогда функция $f$ удовлетворяет условию RSI на множестве $A \cap \mathcal{L}_f(\beta)$ с константой $\mu = \frac{1}{2}\left(\frac{l}{r} - \frac{L}{R}\right)$.

Рассмотрим точку $x \in A \cap \mathcal{L}_f(\beta)$ и $x_0 \in P_\Omega x$. Тогда в силу опорного принципа для сильно выпуклых множеств [10] имеем

$$\left\| x - \frac{f'(x)}{\|f'(x)\|} r - x_0 \right\| \le r$$

и

$$(f'(x), x - x_0) \ge \frac{\|f'(x)\|}{2r} \|x - x_0\|^2 \ge \frac{l}{2r}\|x - x_0\|^2.$$

Из включения $\pm f'(x_0) \in N(A, x_0)$ ($A$ – гладкое многообразие) мы имеем в силу опорного принципа для проксимально гладких множеств, что

$$\left\| x_0 - x + \frac{f'(x_0)}{\|f'(x_0)\|} R \right\| \ge R$$

и, следовательно,

$$-(f'(x_0), x - x_0) \ge -\frac{\|f'(x_0)\|}{2R}\|x - x_0\|^2 \ge -\frac{L}{2R}\|x - x_0\|^2.$$

Таким образом, для всякой точки $x \in A \cap \mathcal{L}_f(\beta)$ имеем

$$(f'(x) - f'(x_0), x - x_0) \ge \frac{1}{2}\left(\frac{l}{r} - \frac{L}{R}\right)\|x - x_0\|^2.$$

Условие $\frac{L}{\mu} < R$ существенно в теореме 1.

**Пример 2.** Рассмотрим в евклидовой плоскости $\mathbb{R}^2$ функцию $f(x, y) = x^2 + \left(y - \frac{1}{2}\right)^2$ и множество $S = \{x^2 - y = 0\}$. Функция $f$ сильно выпукла с константой $\mu = 2$ и непрерывна по Липшицу в окрестности начала координат с константой $L \ge 1$. Минимальный радиус кривизны параболы $S$ равен $\frac{1}{2}$ в точке $(0, 0)$. Следовательно, $S$ проксимально гладкое множество с константой $R = \frac{1}{2}$. Таким образом, условие $\frac{L}{\mu} < R$ нарушено.

Выберем начальную точку МПГ $a_0 = (\tau, \tau^2)$, где $\tau > 0$. Рассмотрим две функции: $y_1 = x^2$ и $x^2 + \left(y_2 - \frac{1}{2}\right)^2 = \tau^2 + \left(\tau^2 - \frac{1}{2}\right)^2$. Тогда $y_1' = y_1'(\tau) = 2\tau$,

$$2x + 2y_2'\left(y_2 - \frac{1}{2}\right) = 0, \qquad y_2' = y_2'(\tau) = \frac{2\tau}{1 - 2\tau^2}.$$

Имеем $y_2' - y_1' \sim 4\tau^3$, $\tau \to +0$. Последнее означает, что угол $\varphi$ между касательными к графикам $y = y_1$ и $y = y_2$ в точке $(\tau, \tau^2)$ равен $\varphi \sim 4\tau^3$, $\tau \to +0$. Пусть $n_0$ – единичная нормаль к параболе $y = x^2$ в точке $a_0$ с условием $(n_0, e_2) > 0$, $e_2 = (0, 1)$. Для достаточно малых чисел $t > 0$ и для точки $a_1 = P_S(a_0 - tf'(a_0))$ получаем, что $\varrho_S(a_1) \le t\|f'(a_0)\| < \frac{1}{4}$. Используя (2), получаем

$$\|a_0 - a_1\| \le \frac{\frac{1}{2}}{\frac{1}{2} - \frac{1}{4}}\|a_0 + tn_0 - (a_0 - tf'(a_0))\| = 2t\|n_0 - (-f'(a_0))\| \sim 8t\tau^3, \qquad \tau \to +0.$$

Таким образом, на $k$-м шаге точка $a_k$ имеет абсциссу больше $\tau_k$, где $\tau_0 = \tau$, $\tau_k = \tau_{k-1} - 8t\tau_{k-1}^3$. Последовательность $\{\tau_k\}$ имеет асимптотику $\tau_k \asymp \frac{1}{\sqrt{k}}$. Линейная скорость сходимости отсутствует.

### 2.2. *Сравнение теоремы* 1 *с известными результатами*

Наиболее сильные результаты для линейной скорости сходимости МПГ на гладких многообразиях, известные автору, содержатся в [18, Theorem 2.3]. Однако указанный результат дает только асимптотику для скорости сходимости, но не точную оценку. При этом существенным для сходимости является условие Лежанского–Поляка–Лоясевича [13, § 3.2, Definition 1], которое достаточно трудно проверить.

Сходимость МПГ в теореме 1 локальна. Однако указанная теорема верна не только для гладкого многообразия, но для любого проксимально гладкого множества. Кроме того, в теореме получена гарантированная оценка скорости сходимости, а не асимптотика.

Автору известно крайне мало работ, в которых МПГ рассматривается на произвольном невыпуклом замкнутом множестве, которое не является гладким многообразием.

В работе [19, Theorem 3.3] доказана линейная сходимость МПГ для произвольного замкнутого множества $A$, но с очень сильными ограничениями на функцию $f$, определенную на множестве $A$. Большинство функций не удовлетворяют таким ограничениям.

Наиболее сильные известные автору результаты о МПГ (для множеств, не являющихся обязательно многообразиями) содержатся в работе [20]. Работа [20] изобилует разными результатами, но мы сосредоточимся на основной теоерме 3 из [20] о сходимости МПГ для случая евклидова пространства.

Если переписать для случая евклидова пространства формулу (14) работы [20] в наших обозначениях, то получится, что на множестве $A$ для сходимости алгоритма должно выполняться условие $\frac{L}{\mu} < R(1 - c_0) < R$, т.е. в точности условие теоремы 1. Более того, множества из работы [20] являются проксимально гладкими (см. [20, п. 2.3]).

Однако от функции в работе [20] требуется так называемое условие ограниченной сильной выпуклости (RSC), т.е. для некоторого $\alpha > 0$ и любых точек $x, y \in A$ должно выполняться неравенство

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

Легко видеть, что из условия RSC вытекает условие RSI. Обратное очевидно неверно. Более того, легко привести примеры задач, где теорема 1 работает, а [20, Theorem 3] нет.

Рассмотрим пример в $\mathbb{R}^2$. Пусть $a \in \left(0, \frac{1}{4}\right)$ и $A = \{(x, y) \mid x \in [0, a],\ 0 \leq y \leq x^2 + 1\}$. Множество $A$ проксимально гладкое с константой $R = \frac{1}{2}$ (это минимальный радиус кривизны параболы $y = x^2 + 1$ при $x \in [0, a]$). Рассмотрим $f(x, y) = x^2$. Очевидно, множество решений задачи (1) есть $\Omega = \{(0, y) \mid y \in [0, 1]\}$. Функция $f$ не удовлетворяет на множестве $A$ условию RSC, так как она постоянна на вертикальных отрезках. Для точки $(x, y) \in A$ имеем при $y \in [0, 1]$ $P_\Omega(x, y) = \{(0, y)\}$, при $y > 1$ $P_\Omega(x, y) = \{(0, 1)\}$. Из соотношений $f'(x, y) = (2x, 0)$, $f'(0, y) = (0, 0)$ и с учетом неравенства $|x| \geq |y - 1|$ для всех $(x, y) \in A$, $y \geq 1$, получаем, что

$$(f'(x, y) - f'(P_\Omega(x, y)), (x, y) - P_\Omega(x, y)) = 2x^2 \geq \|(x, y) - P_\Omega(x, y)\|^2 \quad \forall (x, y) \in A,$$

т.е. функция удовлетворяет условию RSI с константой $\mu = 1$. В силу выбора $a$ условие $\frac{L}{\mu} < R$ также выполнено для функции $f$ и множества $A$.

## 3. МПГ НА МНОГООБРАЗИИ ШТИФЕЛЯ

В случае многообразия Штифеля мы можем получить наилучшую константу проксимальной гладкости $R$. Следующее предложение можно найти в [21, Proposition 7].

**Предложение 2.** *Пусть* $X_0 \in S_{n,k}$. *Тогда для всякой матрицы* $X \in \mathbb{R}^{n\times k}$ *такой, что* $\|X - X_0\| < \sigma_k(X_0)$, *проекция* $X$ *на* $S_{n,k}$ *единственна и имеет вид* $P_{S_{n,k}}X = \sum_{i=1}^{k} U_i V_i^{\text{т}} = UI_{n,k}V^{\text{т}}$, *который задается сингулярным разложением матрицы* $X = U\Sigma V^{\text{т}}$. *Здесь* $\sigma_k(X_0)$ *есть наименьшее сингулярное число* $X_0$, $U_i$ ($V_i$) *есть* $i$*-й столбец* $U \in O(n)$ ($V \in O(k)$) *и* $I_{n,k} = (I_k \vdots 0)^{\text{т}} \in \mathbb{R}^{n\times k}$.

Данное предложение позволяет находить значение $P_{S_{n,k}}X$ для матрицы $X \in \mathbb{R}^{n\times k}$ со свойством $\varrho_{S_{n,k}}(X) < \sigma_k(X_0)$. По какой-то причине авторы [21] не обратили внимание на то, что сингулярные числа любой матрицы $X_0 \in S_{n,k}$ равны 1. Следовательно, $\sigma_k(X_0) = 1$ и для всякой матрицы $X$ со свойством $\varrho_{S_{n,k}}(X) < 1$ существует ровно одна метрическая проекция на $S_{n,k}$. Таким образом, многообразие Штифеля – проксимально гладкое множество с константой $R = 1$. Рассмотрим матрицу $X_0 = [e_1 \vdots e_2 \ldots e_k] \in S_{n,k}$, где $\{e_i\}_{i=1}^{n} \in \mathbb{R}^n$ – стандартный ортонормированный базис, а нормаль задана как $P = [e_1 \vdots 0 \ldots 0]$. Для любого $t \in \left(1, \frac{3}{2}\right)$ имеем $P_{S_{n,k}}(X_0 - tP) = [-e_1 \vdots e_2 \ldots e_k]$. Следовательно, $R = 1$ наилучшая (наибольшая) константа для всех $S_{n,k}$.

**Теорема 2.** *Пусть* $\beta \in \mathbb{R}$ *и* $n, k \in \mathbb{N}$, $n \geq k$. *Предположим, что функция* $f : \mathbb{R}^{n\times k} \to \mathbb{R}$ *– гладкая с константой* $L_1$, $f$ *удовлетворяет условию RSI с константой* $\mu$ *(например,* $f$ *сильно квазивыпуклая) и градиент* $f$ *ограничен по норме на множестве* $A \cap \mathscr{L}_f(\beta)$ *константой* $L$. *Допустим, что* $\frac{L}{\mu} < 1$. *Тогда для всякой точки* $X_0 \in S_{n,k} \cap \mathscr{L}_f(\beta)$ *МПГ (из теоремы* 1*) сходится к решению задачи* $\min_{X\in S_{n,k}} f(X)$ *с линейной скоростью.*

Доказательство вытекает из проксимальной гладкости многообразия Штифеля с константой $R = 1$ и теоремы 1.

## 4. МПГ НА МНОГООБРАЗИИ ГРАССМАНА

Пусть $X \in S_{n,k}$. Тогда $X = U\Sigma V^{\text{т}}$, $U \in O(n)$, $V \in O(k)$, $\Sigma = (I_k \vdots 0)^{\text{т}} \in \mathbb{R}^{n\times k}$ (сингулярное разложение матрицы). Таким образом, любой элемент $XX^{\text{т}} \in G_{n,k}$ можно представить в виде

$$XX^{\text{т}} = UZU^{\text{т}}, \; Z = \begin{pmatrix} I_k & 0 \\ 0 & 0 \end{pmatrix} \in \mathbb{R}^{n\times n}.$$

**Теорема 3.** *Пусть* $Y \in \mathrm{Sym}(n)$ *и* $Y = W^{\text{т}}\Lambda W$, $W \in O(n)$, $\Lambda = \mathrm{diag}\,\{\lambda_1, \lambda_2, \ldots, \lambda_n\}$ *со свойством* $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_k > \lambda_{k+1} \geq \ldots \geq \lambda_n$, *т.е.* $\lambda_k > \lambda_{k+1}$. *Тогда* $P_{G_{n,k}}Y = \left\{W^{\text{т}}ZW\right\}$.

**Доказательство.** В силу предложения 1 имеем

$$\left\|W^{\text{т}}\Lambda W - UZU^{\text{т}}\right\| = \left\|\Lambda - WUZU^{\text{т}}W^{\text{т}}\right\| = \left\|\Lambda - QZQ^{\text{т}}\right\|,$$

где $Q = WU$. Из равенства

$$\left\|\Lambda - QZQ^{\text{т}}\right\|^2 = \|\Lambda\|^2 + \left\|QZQ^{\text{т}}\right\|^2 - 2(\Lambda, QZQ^{\text{т}}) = \|\Lambda\|^2 + k - 2(\Lambda, QZQ^{\text{т}})$$

мы получаем, что метрическая проекция $Y$ на $G_{n,k}$ достигается для таких матриц $Q \in O(n)$, которые дают максимум выражения

$$(\Lambda, QZQ^{\text{т}}) = \mathrm{tr}\Lambda QZQ^{\text{т}} = \sum_{m=1}^{n} \lambda_m \sum_{i=1}^{k} Q_{mi}^2,$$

т.е. когда $\sum_{i=1}^{k} Q_{mi}^2 \leq 1$ для всех $m$ и $\sum_{m=1}^{n}\sum_{i=1}^{k} Q_{mi}^2 = k$. Таким образом, максимум $(\Lambda, QZQ^{\text{т}})$ достигается *тогда и только тогда* когда матрица $Q \in O(n)$ дает $\sum_{i=1}^{k} Q_{mi}^2 = 1$ для всех $1 \leq m \leq k$ (напомним, что $\lambda_1 \geq \ldots \lambda_k > \lambda_{k+1} \geq \ldots \lambda_n$). Значит, $\max_Q(\Lambda, QZQ^{\mathrm{T}}) = \sum_{m=1}^{k}\lambda_m$ имеет место только для матриц вида $Q = \begin{pmatrix} S & 0 \\ 0 & F \end{pmatrix} \in \mathbb{R}^{n\times n}$, где $S \in O(k)$, $F \in O(n-k)$. Из предыдущих рассуждений получаем

$$\varrho_{G_{n,k}}(Y) = \min_Q \left\|\Lambda - QZQ^{\mathrm{T}}\right\| = \left\|\Lambda - \begin{pmatrix} S & 0 \\ 0 & F \end{pmatrix} Z \begin{pmatrix} S^{\mathrm{T}} & 0 \\ 0 & F^{\mathrm{T}} \end{pmatrix}\right\| = \|\Lambda - Z\|.$$

Выбирая $U = W^{\mathrm{T}}\begin{pmatrix} S & 0 \\ 0 & F \end{pmatrix}$ для всякого $S \in O(k)$ и $F \in O(n-k)$ мы находим $\varrho_{G_{n,k}}(Y) = \left\|W^{\mathrm{T}}\Lambda W - W^{\mathrm{T}}ZW\right\|$. Следовательно, $W^{\mathrm{T}}ZW \in P_{G_{n,k}}Y$. Предположим, что $Y = W^{\mathrm{T}}\Lambda W$, $W \in O(n)$, и $U^{\mathrm{T}}ZU \in P_{G_{n,k}}Y$ для некоторого $U \in O(n)$. Тогда $\varrho_{G_{n,k}}(Y) = \left\|\Lambda - QZQ^{\mathrm{T}}\right\|$ для $Q = WU^{\mathrm{T}} = \begin{pmatrix} S & 0 \\ 0 & F \end{pmatrix}$ и некоторых матриц $S \in O(k)$ и $F \in O(n-k)$. Из равенства $W = \begin{pmatrix} S & 0 \\ 0 & F \end{pmatrix} U$ находим $W^{\mathrm{T}}ZW = U^{\mathrm{T}}ZU$.

Допустим, что $Y = W^{\mathrm{T}}\Lambda W = W_1^{\mathrm{T}}\Lambda_1 W_1$, $\Lambda$ и $\Lambda_1$ – диагональные матрицы, верхний левый $k \times k$ и правый нижний $(n-k)\times(n-k)$ блоки $\Lambda_1$ состоят из тех же собственных значений, что и соответствующие блоки матрицы $\Lambda$ (но в другом порядке); $W, W_1 \in O(n)$. Аналогично с предыдущими рассуждениями легко видеть, что $W_1^{\mathrm{T}}ZW_1 = W^{\mathrm{T}}ZW$. Значит, множество $P_{G_{n,k}}Y$ одноточечно. Теорема доказана.

Заметим, что для матрицы $Y \in \mathrm{Sym}(n)$ со свойством $\varrho_{G_{n,k}}(Y) = \|\Lambda - Z\| < 1/\sqrt{2}$ мы имеем $\Lambda = \mathrm{diag}\{1+\varepsilon_1, \ldots, 1+\varepsilon_k, \varepsilon_{k+1}, \ldots, \varepsilon_n\}$, где $\sum_{i=1}^{n}\varepsilon_i^2 < \frac{1}{2}$. Для всех $i$, $j$ выполняются неравенства $\varepsilon_i^2 + \varepsilon_j^2 < \frac{1}{2}$, $1+\varepsilon_i > 0$ и

$$(1+\varepsilon_i)^2 - \varepsilon_j^2 = 1 + 2\varepsilon_i + 2\varepsilon_i^2 - (\varepsilon_i^2 + \varepsilon_j^2) > 1 + 2\varepsilon_i + 2\varepsilon_i^2 - \frac{1}{2} \geq 0.$$

Следовательно, $1+\varepsilon_i = |1+\varepsilon_i| > |\varepsilon_j| \geq \varepsilon_j$ для всех $1 \leq i \leq k$, $j \geq k+1$. В силу теоремы 3 множество $P_{G_{n,k}}Y$ одноточечно. Таким образом, для любой матрицы $Y \in \mathrm{Sym}(n)$ со свойством $\varrho_{G_{n,k}}(Y) < 1/\sqrt{2}$ существует единственная метрическая проекция $P_{G_{n,k}}Y$. Значит, $G_{n,k}$ проксимально гладкое множество с константой $R = 1/\sqrt{2}$.

Рассмотрим $\Lambda = \mathrm{diag}\{\underbrace{1, \ldots, 1}_{k-1\ \text{раз}}, 1/2, 1/2, 0, \ldots 0\} \in \mathrm{Sym}(n)$. Для блочно диагональной матрицы $Q(\varphi)$, состоящей из блоков $I_{k-1}$, $\begin{pmatrix} \cos\varphi & \sin\varphi \\ -\sin\varphi & \cos\varphi \end{pmatrix}$ ($\varphi \in [0, 2\pi)$) и $I_{n-k-1}$, мы получаем, что матрица

$$P(\varphi) = Q(\varphi)ZQ^{\mathrm{T}}(\varphi) = \begin{pmatrix} I_{k-1} & 0 & 0 \\ 0 & \begin{pmatrix} \cos^2\varphi & -\sin\varphi\cos\varphi \\ -\sin\varphi\cos\varphi & \sin^2\varphi \end{pmatrix} & 0 \\ 0 & 0 & 0_{n-k-1} \end{pmatrix}$$

дает равенство $\|\Lambda - P(\varphi)\| = 1/\sqrt{2}$. Таким образом, константа $R = 1/\sqrt{2}$ неулучшаема. (Неулучшаемость константы проксимальной гладкости для $G_{n,k}$ получена совместно со студентом Р. Камаловым.)

Отметим, что рассмотренное вложение многообразия Грассмана в пространство $\mathrm{Sym}(n)$ не является единственно возможным и другие вложения (см. детали в [14]) могут давать другие оценки для константы $R$.

Теорема 3 позволяет сформулировать полный аналог теоремы 2 в случае минимизации на множестве $G_{n,k}$.

**Теорема 4.** *Пусть* $\beta \in \mathbb{R}$ *и* $n, k \in \mathbb{N}$, $n \geq k$. *Предположим, что функция* $f : \mathrm{Sym}(n) \to \mathbb{R}$ *гладкая с константой* $L_1$, $f$ *удовлетворяет условию RSI с константой* $\mu$ (*например,* $f$ *сильно квазивыпуклая*) *и градиент функции* $f$ *ограничен по норме на множестве* $A \cap \mathscr{L}_f(\beta)$ *константой* $L$. *Допустим, что* $\frac{L}{\mu} < 1/\sqrt{2}$. *Тогда для всякой точки* $X_0 \in G_{n,k} \cap \mathscr{L}_f(\beta)$ *МПГ* (*из теоремы* 1) *сходится к решению задачи* $\min_{X \in G_{n,k}} f(X)$ *с линейной скоростью.*

## СПИСОК ЛИТЕРАТУРЫ


1. *Goldstein A.A.* Convex programming in Hilbert space // Bull. Amer. Math. Soc. 1964. V. 70. № 5. P. 709–710.
2. *Левитин Е.С., Поляк Б.Т.* Методы минимизации при наличии ограничений // Ж. вычисл. матем. и матем. физ. 1966. V. 6. № 5. С. 787–823.
3. *Нестеров Ю.Е.* Введение в выпуклую оптимизацию. М.: МЦМНО, 2010. 278 с.
4. *Edelman A., Arias T.A., Smith S.T.* The geometry of algorithms with orthogonality constraints // J. Matrix Anal. Appl. 1998. V. 20. № 2. P. 303–353.
5. *Luenberger D.G.* The gradient projection methods along geodesics // Management Science. 1972. V. 18. № 11. P. 620–631.
6. *Hager W.W.* Minimizing a quadratic over a sphere // SIAM J. Optim. and Contr. 2001. V. 12. № 1. P. 188–208.
7. *Absil P.-A., Mahony R., Sepulchre R.* Matrix Manifolds. Princeton University Press, Princeton and Oxford. 2008. 240 p.
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 // Mathematics and Its Applications series. Dordrecht: Springer. 1998. V. 297.
10. *Vial J.-Ph.* Strong and weak convexity of sets and functions // Mathematics of Operations Research. 1983. V. 8. № 2. P. 231–259.
11. *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.
12. *Обен Ж.-П., Экланд И.* Прикладной нелинейный анализ. М.: Мир, 1988.
13. *Balashov M., Polyak B., Tremba A.* Gradient projection and conditional gradient methods for constrained nonconvex minimization // Numerical Functional Analysis and Optimization. 2020. V. 41. № 7. P. 822–849.
14. *Conway J.H., Hardin R.H., Sloane N.J.A.* Packing lines, planes, etc.: packings in grassmannian spaces // Experimental Mathematics. 1996. V. 5. P. 139–159.
15. *Karimi H., Nutini J., Schmidt M.* Linear convergence of gradient and proximal-gradient methods under the Polyak-Lojasiewicz condition // In: Frasconi P., Landwehr N., Manco G., Vreeken J. (eds.) Machine Learning and Knowledge Discovery in Databases. ECML PKDD 2016. Lecture Notes in Computer Science. V. 9851.
16. *Balashov M.V.* About the gradient projection algorithm for a strongly convex function and a proximally smooth set // J. Convex Anal. 2017. V. 24. № 2. P. 493–500.
17. *Wei K., Cai J.-F., Chan T.F., Leung Sh.* Guarantees of Riemannian optimization for low rank matrix recovery // 2016. arXiv: 1511.01562v8.
18. *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.
19. *Jain P., Kar P.* Non-convex Optimization for Machine Learning. Now Foundations and Trends. 2017. 154 p.
20. *Barber R.F., Ha W.* Gradient descent with nonconvex constraints: local concavity determines convergence // 2017. arXiv: 1703.07755v3.
21. *Absil P.-A., Malick J.* Projection-like retraction on matrix manifolds // SIAM J. OPTIM. 2012. V. 22. № 1. P. 135–158.