УДК 517.9

НЕВЫПУКЛАЯ ОПТИМИЗАЦИЯ НА ГЛАДКИХ МНОГООБРАЗИЯХ[1]

М. В. Балашов (Москва, Россия)
balashov73@mail.ru

Рассматривается задача минимизации невыпуклой в общем случае функции с непрерывным по Липшицу градиентом на вещественном компактном многообразии без края. Обсуждаются условия ограничения ошибки в указанной задаче, которые могут заменить условие выпуклости. В качестве примера многобразий рассмотрены вещественном многообразии Штифеля или Грассмана. Многобразие Штифеля рассматривается как вложение в пространство матриц. Многобразие Грассмана рассмотрено как вложение в пространство симметричных матриц. В этой ситуации доказана проксимальная гладкость многообразия Штифеля с константой 1 и многообразия Грассмана с константой $1/\sqrt{2}$ в Фробениусовской норме. Получены достаточные условия сходимости метода проекции градиента в определенных случаях. Получены формулы для вычисления метрической проекции на указанные многобразия для точек, достаточно близких к многообразию (не далее константы проксимальной гладкости).

Ключевые слова: невыпуклая оптимизация, условие ограничения ошибки, проксимальная гладкость, гладкое многообразие, многообразие Штифеля, многообразие Грассмана.

NONCONVEX OPTIMIZATION ON SMOOTH MANIFOLDS[1]

M. V. Balashov (Moscow, Russia)
balashov73@mail.ru

The problem of minimization for a nonconvex in general function with Lipschitz continuous gradient on a smooth manifold without egde is considered. Error bound conditions which can replace the convexity condition are discussed in this problem. As an example of manifolds the real Stiefel or Grassmann manifolds are considered. The Stiefel manifold is embedded in a matrix space. The Grassmann manifold is embedded in a space of symmetric matrices. Proximal smoothness of the Stiefel manifold with constant 1 and the Grassmann manifold with constant $1/\sqrt{2}$ in the Frobenius norm is proved in this situation. Sufficient conditions of convergence for the gradient projection algorithm are obtained in certain cases. Formulas for calculating the metric projection onto considered manifolds for sufficiently close points (not further than proximal smoothness constant) are obtained.

Keywords: nonconvex optimization, error bound condition, proximal smoothness, smooth manifold, Stiefel manifold, Grassnmann manifold.

Рассмотрим задачу

$$\min_S f(x), \tag{1}$$

где S — компактное гладкое многообразие без края, а функция f имеет липшицев градиент f' с константой $L_1 > 0$. Мы будем предполагать, что S задаётся системой $S = \{x \in \mathbb{R}^n \mid g(x) = 0\}$, $g : \mathbb{R}^n \to \mathbb{R}^m$, $m < n$. Для матрицы Якоби $g'(x) = \left(\frac{\partial g_i(x)}{\partial x_j}\right)_{i=1\ldots m, j=1\ldots n} \in \mathbb{R}^{m\times n}$ ранг $\operatorname{rank} g'(x) = m$

[1]Работа выполнена при финансовой поддержке РНФ (проект № 16-11-10015).
[1]The article is done with the financial support of RSF (project № 16-11-10015).

для всех $x \in S$. Определим $\varrho_S(X) = \varrho(x, S) = \inf\limits_{a \in S} \|x - a\|$ — расстояние от точки x до множества S. Мы будем предполагать проксимальную гладкость S [1,2] с константой $R > 0$. Последнее эквивалентно тому, что для каждой точки $x \in U_S(R) = \{x \in \mathbb{R}^n \mid 0 < \varrho_S(x) < R\}$ существует единственная метрическая проекция $P_S x$ на множество S, которая непрерывна по $x \in U_S(R)$.

Пусть Ω — множество стационарных точек в (1), т.е. $\Omega = \{x_* \in S \mid f'(x_*) \in -N(S, x_*)\}$. Для проксимально гладкого множества в качестве $N(S, x_*)$ можно понимать любой нормальный конус (они все совпадают). Рассмотрим алгоритм поиска стационарной точки в задаче (1), а также точек минимума.

Обычно при решении задачи (1) используются шаги метода проекции градиента, ассоциированные с геодезическими на многообразии S [3].

Мы предлагаем подход [4, 5], в котором метод проекции градиента записывается в стандартном виде

$$x_0 \in S, \qquad x_{k+1} = P_S(x_k - tf'(x_k)), \quad k = 0, 1, \ldots . \tag{2}$$

Пусть $\mathcal{G}_t(x) = \frac{1}{t}(x - P_S(x - tf'(x)))$, $x \in S$, — градиентное отображение [6], а $P_{T_x} f'(x)$ — проекция $f'(x)$ на касательное подпространство T_x ко множеству S в точке $x \in S$, т.е.

$$P_{T_x} f'(x) = \left(I_n - g'(x)^T \left(g'(x) g'(x)^T\right)^{-1} g'(x)\right) f'(x),$$

I_n — единичная $n \times n$ матрица.

Рассмотрим некоторые условия *ограничения ошибки:* существует такое $\mu > 0$, что $\mu \varrho_\Omega(x) \le \|P_{T_x} f'(x)\|$ (будем называть это условие *касательное ограничение ошибки* или tEB) или $\mu \varrho_\Omega(x) \le \|\mathcal{G}_t(x)\|$ для всех $x \in S$ (будем называть это условие *градиентное ограничение ошибки* или gEB). Условие tEB предложил А. Тремба.

Свойства tEB и gEB позволяют получить сходимость метода проекции градиента со скоростью геометрической прогрессии в ряде случаев.

Пусть в задаче $\min\limits_{x \in Q} f(x)$ выполнено условие gEB и Ω есть множество глобальных минимумов функции f на множестве $Q \cap \{x \in \mathbb{R}^n \mid f(x) \le f(x_0)\}$, где $x_0 \in Q$. В работе [5] доказано, что если множество $Q \subset \mathbb{R}^n$ проксимально гладкое с константой $R > 0$ (Q не обязательно является многообразием), а функция $f : \mathbb{R}^n \to \mathbb{R}$ липшицева с константой L_0 и ее градиент f' липшицев с константой L_1, то при выборе $0 < t \le \frac{1}{L_1 + \frac{3L_0}{R}}$ метод проекции градиента (2) для задачи $\min\limits_{x \in Q} f(x)$ с начальным условием x_0 сходится со скоростью геометрической прогрессии.

Пусть в задаче (1) множество Ω есть множество глобальных минимумов функции f на множестве $S \cap \{x \in \mathbb{R}^n \mid f(x) \leq f(x_0)\}$, где $x_0 \in S$. В работе [4] для гладкого и проксимально гладкого многообразия $S \subset \mathbb{R}^n$ размерности $n-1$ предложен алгоритм в духе метода проекции градиента: $z_k = x_k + tP_{T_{x_k}} f'(x_k)$, $x_{k+1} = P_S z_k$. При выборе $0 < t \leq \frac{1}{L_1 + \frac{2L_0}{R}}$ и выполнении условия Лежанского-Поляка-Лоясевича [4], т.е.

$$\exists \mu > 0: \ \mu(f(x) - f(\Omega)) \leq \|P_{T_x} f'(x)\|^2, \quad \forall x \in S,$$

которое эквивалентно условию gEB, указанный метод также сходится со скоростью геометрической прогрессии.

Аналогичный алгоритм можно предложить и в случаем многообразия S любой размерности.

Рассмотрим $z = (x, \lambda) \in \mathbb{R}^{n+m}$ и $F(z) = \begin{pmatrix} f'(x) + g'(x)^T \lambda \\ g(x) \end{pmatrix}$. Пусть $\lambda(x) = -\left(g'(x) g'(x)^T\right)^{-1} g'(x) f'(x)$ для всех $x \in S$. Заметим, что $f'(x) + g'(x)^T \lambda(x) = P_{T_x} f'(x)$ для всех $x \in S$.

Центральным моментом является следующий результат.

Теорема 3. *Пусть в задаче (1) функции f и g дважды непрерывно дифференцируемы и вторые производные имеют липшицев градиент. Пусть множество стационарных точек Ω конечно и в каждой из точек $x_0 \in \Omega$ матрица*

$$F'_z(x_0, \lambda(x_0)) = \begin{pmatrix} f''(x_0) + \sum\limits_{i=1}^{m} \lambda_i(x_0) g''_i(x_0) & g'(x_0)^T \\ g'(x_0) & 0 \end{pmatrix}$$

невырождена. Тогда выполнено tEB: существует такое число $\mu > 0$, что

$$\mu \varrho(x, \Omega) \leq \|P_{T_x} f'(x)\| \qquad \forall x \in S.$$

Доказательство. В невырожденной задаче множество стационарных точек конечно ($\Omega = \{x_j\}_{j=1}^{J}$). Это легко доказать от противного. Заметим также, что функция $S \ni x \to \lambda(x)$ липшицева, обозначим её константу Липшица через L_λ.

По определению стационарных точек $F(x_j, \lambda(x_j)) = 0$ для всякого $x_j \in \Omega$. Тогда из дифференцируемости $F(z)$ по формуле Тейлора имеем для любого $j \in \{1, \ldots, J\}$ при $\varrho \to +0$ асимптотическое равенство

$$F_x(x) = F(x, \lambda(x)) - F(x_j, \lambda(x_j)) = F'(x_j, \lambda(x_j)) \begin{bmatrix} x - x_j \\ \lambda(x) - \lambda(x_j) \end{bmatrix} + o_j(\varrho),$$

где $\varrho = \sqrt{\|x-x_j\|^2 + \|\lambda(x)-\lambda(x_j)\|^2} \le \|x-x_j\|\sqrt{1+L_\lambda^2}$.

Пусть $\sigma_0 > 0$ удовлетворяет условию $\|F'(x_j, \lambda(x_j))^{-1}\| \le \sigma_0, \; j = 1, ..., J$. По теореме Банаха об обратном операторе это эквивалентно условию

$$\|F'(x_j, \lambda(x_j))h\| \ge \frac{1}{\sigma_0}\|h\|, \;\; \forall h \in \mathbb{R}^{n+m}, \; j = 1, ..., J.$$

Выберем число $\ell > 0$ настолько малое, что для всех $j = 1, ..., J$ и $h \in \mathbb{R}^{n+m}$, $\|h\| \le \ell$, выполняется $\|o_j(\|h\|)\| \le \frac{1}{2\sigma_0}\|h\|$.

Зафиксируем $x \in S$, $\rho(x, \Omega) \le \frac{\ell}{\sqrt{1+L_\lambda^2}}$. Используя разложение Тейлора относительно *ближайшей* к x стационарной точки x_j с учетом $\varrho \le \|x - x_j\|\sqrt{1+L_\lambda^2} \le \ell$ имеем

$$\|F_x(x)\| \ge \left\| F'(x_j, \lambda(x_j)) \begin{bmatrix} x - x_j \\ \lambda(x) - \lambda(x_j) \end{bmatrix} \right\| - \|o_j(\varrho)\| \ge \frac{1}{\sigma_0}\varrho - \frac{1}{2\sigma_0}\varrho = \frac{1}{2\sigma_0}\varrho$$
$$\ge \frac{1}{2\sigma_0}\|x - x_j\| = \frac{1}{2\sigma_0}\rho(x, \Omega).$$

Таким образом, выполняется условие tEB

$$\|P_{T_x} f'(x)\| = \|F_x(x)\| \ge \mu\rho(x, \Omega), \quad \forall x \in \cup_{j=1}^{J} \mathrm{int} B_r(x_j), \tag{3}$$

где $\mu = 1/(2\sigma_0)$, $r = \frac{\ell}{\sqrt{1+L_\lambda^2}}$.

На оставшемся компактном множестве $S_1 = \left\{ x \in S : \rho(x, \Omega) \ge \frac{\ell}{\sqrt{1+L_\lambda^2}} \right\}$ функция $F_x(x)$ непрерывная, и, в силу отсутствия в S_1 стационарных точек, $\|F_x(x)\| > 0$ для всех $x \in S_1$. По теореме Вейерштрасса существует число $b > 0$ такое, что $\|F_x(x)\| \ge b > 0$ для всех $x \in S_1$. Посокольку $\operatorname{diam} S = \sup_{x,y \in S} \|x - y\| \ge \rho(x, \Omega)$, то

$$\|F_x(x)\| \ge b \ge \frac{b}{\operatorname{diam} S}\rho(x, \Omega), \;\; \forall x \in S : \rho(x, \Omega) > \frac{\ell}{\sqrt{1+L_\lambda^2}}. \tag{4}$$

Комбинируя неравенства (3) и (4), получаем условие tEB на всём множестве S:

$$\|F_x(x)\| \ge \min\left\{\frac{1}{2\sigma_0}, \frac{b}{\operatorname{diam} S}\right\}\rho(x, \Omega), \;\; \forall x \in S.$$

□

Будем называть задачу (1) невырожденной, если выполнены условия теоремы 3. Например, условие теоремы выполнено для квадратичной функции $f(x) = (x, \Lambda x)$ на единичной евклидовой сфере, $\Lambda =$

$\mathrm{diag}\{\lambda_1, \lambda_2, ..., \lambda_n\}$, $\lambda_i < \lambda_j$ для всех $i < j$, с $\mu = \min_{1 \le i \ne j \le n} |\lambda_i - \lambda_j|$, а также, более общо, для квадратичной функции на многообразии Штифеля $S_{n,k} = \{X \in \mathbb{R}^{n \times k} \mid X^T X = I_k\}$, $k \le n$ [7].

Заметим, что условие невырожденности задачи не обязательно для выполнения условия tEB. Пусть функция $x \to \lambda(x)$ непрерывно дифференцируема в некоторой окрестности множества S вида $U_S(\delta)$, $\delta > 0$, и существует такое число $\mu > 0$, что для любой точки $x_* \in \Omega$ выполнено условие

$$\left\| F'_z(x_*, \lambda(x_*)) \begin{pmatrix} I_n \\ \lambda'(x_*) \end{pmatrix} h \right\| \ge \mu \|h\|, \quad \forall h \in T_{x_*} \subset \mathbb{R}^n.$$

Тогда в соответствующей задаче (1) выполнено условие tEB. Доказательство повторяет доказательство теоремы 3. Заметим, однако, что обратимость матрицы $F'(x, \lambda(x))$ в точках $x \in \Omega$ полезна, например, для применения метода Ньютона.

Условия tEB и gEB эквивалентны. Покажем, что условие tEB влечет условие gEB. Действительно, пусть $x \in S$, $x_1 = P_S(x - tf'(x))$. Тогда известно, что $\|P_{T_{x_1}} f'(x_1)\| \le (1 + L_1 t)\, \|\mathcal{G}_t(x)\|$ [4, стр. 7]. Отсюда

$$\varrho(x, \Omega) - \|x - x_1\| \le \varrho(x_1, \Omega) \le \frac{1}{\mu} \|P_{T_{x_1}} f'(x_1)\| \le \frac{1 + tL_1}{\mu} \|\mathcal{G}_t(x)\|,$$

$$\varrho(x, \Omega) \le \left(\frac{1 + tL_1}{\mu} + t \right) \|\mathcal{G}_t(x)\|.$$

Рассмотрим применение полученных результатов в случае многообразий Штифеля и Грассмана. Мы ограничимся вопросами оценки константы проксимальной гладкости для указанных многообразий и формулами для нахождения метрической проекции.

Мы можем вычислить наибольшую константу проксимальной гладкости для этих многообразий. Многообразие Штифеля определено выше, а многообразие Грассмана можно определить по формуле $G_{n,k} = \{XX^T \mid X \in S_{n,k}\}$ [8]. Заметим, что многообразие Грассмана вложено (в определенном смысле изометрично) в пространство симметричных $n \times n$ матриц $Sym(n)$.

Далее мы будем рассматривать указанные реализации многообразий $S_{n,k}$ и $G_{n,k}$ в соответствующих евклидовых пространствах $\mathbb{R}^{n \times k}$ и $Sym(n)$ со скалярным произведением $(X, Y) = \mathrm{tr} X^T Y$, $X, Y \in \mathbb{R}^{n \times k}$, и соответствующей Фробениусовской нормой $\|X\| = \sqrt{(X, X)}$.

Из [9, Proposition 7] вытекает, что многообразие Штифеля проксимально гладкое с константой $R = 1$ во Фробениусовской норме. При этом $P_{S_{n,k}}X = UI_{n,k}V^T$, где $I_{n,k} = (I_k \vdots 0)^T \in \mathbb{R}^{n\times k}$, а $X = U\Sigma V^T$ — сингулярное разложение матрицы X. Здесь $\Sigma \in \mathbb{R}^{n\times k}$, а U и V ортогональные матрицы соответствующих размеров.

Пусть $X \in S_{n,k}$. Тогда $X = U\Sigma V^T$, $U \in O(n)$, $V \in O(k)$, $\Sigma = (I_k \vdots 0)^T \in \mathbb{R}^{n\times k}$ (сингулярное разложение). Таким образом, любой элемент $XX^T \in G_{n,k}$ может быть представлен в ввиде

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

Теорема 4. *Пусть $Y \in Sym(n)$ и $Y = W^T\Lambda W$, $W \in O(n)$, $\Lambda = diag\{\lambda_1, \lambda_2, \dots, \lambda_n\}$ со свойством $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_k > \lambda_{k+1} \geq \cdots \geq \lambda_n$, то есть $\lambda_k > \lambda_{k+1}$. Тогда метрическая проекция $P_{G_{n,k}}Y = \{W^TZW\}$.*

Доказательство. Легко проверить, что для произвольной ортогональной матрицы $Q \in O(n)$ $(Q \in O(k))$ и для произвольной матрицы $X \in \mathbb{R}^{n\times k}$ выполнено равенство $\|QX\| = \|X\|$ $(\|XQ\| = \|X\|)$. Отсюда следует, что

$$\|W^T\Lambda W - UZU^T\| = \|\Lambda - WUZU^TW^T\| = \|\Lambda - QZQ^T\|.$$

Положим $Q = WU$. Из равенства $\|\Lambda - QZQ^T\|^2 =$

$$\|\Lambda\|^2 + \|QZQ^T\|^2 - 2(\Lambda, QZQ^T) = \|\Lambda\|^2 + k - 2(\Lambda, QZQ^T)$$

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

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

т.е. когда $\sum\limits_{i=1}^{k} Q_{mi}^2 \leq 1$ для всех m и $\sum\limits_{m=1}^{n}\sum\limits_{i=1}^{k} Q_{mi}^2 = k$. Следовательно максимум (Λ, QZQ^T) достигается *тогда и только тогда* когда матрица $Q \in O(n)$ такая, что $\sum\limits_{i=1}^{k} Q_{mi}^2 = 1$ для всех $1 \leq m \leq k$ (напомним, что $\lambda_1 \geq \dots \lambda_k > \lambda_{k+1} \geq \dots \lambda_n$). Таким образом, равенство $\max_Q(\Lambda, QZQ^T) = \sum\limits_{m=1}^{k} \lambda_m$ имеет место только для матриц вида $Q = \begin{pmatrix} S & 0 \\ 0 & F \end{pmatrix} \in \mathbb{R}^{n\times n}$ с компонентами $S \in O(k)$, $F \in O(n-k)$. Из предыдущих рассуждений имеем $\varrho_{G_{n,k}}(Y) = \min_Q \|\Lambda - QZQ^T\| = \left\| \Lambda - \begin{pmatrix} S & 0 \\ 0 & F \end{pmatrix} Z \begin{pmatrix} S^T & 0 \\ 0 & F^T \end{pmatrix} \right\| =$ $\|\Lambda - Z\|$. Выбирая $U = W^T \begin{pmatrix} S & 0 \\ 0 & F \end{pmatrix}$ для всех $S \in O(k)$ и $F \in O(n-k)$ мы получаем $\varrho_{G_{n,k}}(Y) = \|W^T\Lambda W - W^T ZW\|$. Следовательно $W^T ZW \in P_{G_{n,k}}Y$. Предположим $Y = W^T\Lambda W$, $W \in O(n)$, и $U^T ZU \in P_{G_{n,k}}Y$ для некоторой $U \in O(n)$. Тогда $\varrho_{G_{n,k}}(Y) = \|\Lambda - QZQ^T\|$ для $Q = WU^T = \begin{pmatrix} S & 0 \\ 0 & F \end{pmatrix}$ для некоторых $S \in O(k)$ и $F \in O(n-k)$. Из равенства $W = \begin{pmatrix} S & 0 \\ 0 & F \end{pmatrix} U$ мы получаем $W^T ZW = U^T ZU$.

Предположим, что $Y = W^T\Lambda W = W_1^T\Lambda_1 W_1$, Λ и Λ_1 — диагональные матрицы, верхний-левый $k \times k$ и правый-нижний $(n-k)\times(n-k)$ блоки матрицы Λ_1 состоят из тех же собственных значений, что и сооответствующие блоки Λ (но взятых в другом порядке); $W, W_1 \in O(n)$. Аналогично с предыдущими рассуждениями легко показать, что $W_1^T ZW_1 = W^T ZW$. Следовательно $P_{G_{n,k}}Y$ одноточечно. □

Дальнейшие результаты были доказаны вместе со студентом Р. Камаловым.

Заметим, что для матрицы $Y \in \mathrm{Sym}(n)$ со свойством $\varrho_{G_{n,k}}(Y) = \|\Lambda - Z\| < 1/\sqrt{2}$ мы получаем $\Lambda = \mathrm{diag}\{1+\varepsilon_1, \dots, 1+\varepsilon_k, \varepsilon_{k+1}, \dots \varepsilon_n\}$, где $\sum\limits_{i=1}^{n} \varepsilon_i^2 < \frac{1}{2}$. Для всех i, j имеем $\varepsilon_i^2 + \varepsilon_j^2 < \frac{1}{2}$, $1 + \varepsilon_i > 0$ 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 предыдущих рассуждений множество $P_{G_{n,k}}Y$ одноточечно. Итак, для всякого $Y \in \mathrm{Sym}(n)$ с $\varrho_{G_{n,k}}(Y) < 1/\sqrt{2}$ существует единственная метрическая проекция $P_{G_{n,k}}Y$. Значит, $G_{n,k}$ проксимально гладкое множество с константой $R = 1/\sqrt{2}$.

Рассмотрим $\Lambda = \mathrm{diag}\{\underbrace{1, \dots, 1}_{k-1 \text{ раз}}, 1/2, 1/2, 0, \dots 0\} \in \mathrm{Sym}(n)$. 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 $P(\varphi) =$

$Q(\varphi)ZQ^T(\varphi) =$

$$= \begin{pmatrix} I_{k-1} & 0 & 0 \\ 0 & \begin{pmatrix} \cos^2\varphi & -\sin\varphi\cos\varphi \\ -\sin\varphi\cos\varphi & \sin^2\varphi \end{pmatrix} & 0 \\ 0 & 0 & 0_{n-k-1} \end{pmatrix}$$

даёт равенство $\|\Lambda - P(\varphi)\| = 1/\sqrt{2}$. Поэтому константа $R = 1/\sqrt{2}$ неулучшаема (нельзя увеличить). $\square$

Мы хотим обратить внимание на то, что рассмотренное вложение многообразия Грассмана в пространство $\mathrm{Sym}(n)$ не единственно возможное и другие вложения (см. детали в [8]) могут давать иные оценки для константы проксимальной гладкости R.

СПИСОК ЛИТЕРАТУРЫ

[1] *Vial J.-Ph.* Strong and weak convexity of sets and functions. // Mathematics of Operations Research. 1983. Vol. 8, № 2. P. 231–259.

[2] *Clarke F. H., Stern R. J., Wolenski P. R.* Proximal smoothness and lower–C^2 property. // J. Convex Anal. 1995. Vol. 2, № 1-2. P. 117–144.

[3] *Edelman A., Arias T. A., Smith S. T.* The geometry of algorithms with orthogonality constraints. // J. Matrix Anal. Appl. 1998. Vol. 20, № 2. P. 303–335.

[4] *Balashov M., Polyak B., Tremba A.* Gradient projection and conditional gradient methods for constrained nonconvex minimization. 2019. arXiv:1906.11580.

[5] *Балашов М. В.* Метод проекции градиента для проксимально гладкого множества и функции с непрерывным по Липшицу градиентом. // В печати.

[6] *Nesterov Yu.* Introductory lectures on convex optimization. A basic course basic course. Springer, 2004.

[7] *Liu H., Wu W., So A. M.-Ch.* Quadratic Optimization with Orthogonality Constraints: Explicit Lojasiewicz Exponent and Linear Convergence of Line-Search Methods. 2015. arXiv:1510.01025.

[8] *Conway J. H., Hardin R. H., Sloane N. J. A.* Packing Lines, Planes, etc.: Packings in Grassmannian Spaces // Experimental Mathematics. 1996. Vol. 5. P. 139–159.

[9] *Absil P.-A., Malick J.* Projection-like retraction on matrix manifolds // SIAM J. OPTIM. 2012. Vol 22, № 1. P. 135–158.