

# МЕТОДЫ ОПТИМИЗАЦИИ И ОПТИМАЛЬНОЕ УПРАВЛЕНИЕ

---

УДК 519.853.6

*В.А. Григоренко, Д.А. Ключин, С.И. Ляшко*

## МНОГООБЛОЧНЫЙ МЕТОД ADMM С УСКОРЕНИЕМ НЕСТЕРОВА

**Ключевые слова:** ADMM (alternating direction method of multipliers), многооблочный ADMM, ускоренный ADMM, ускорение Нестерова.

### Введение

Стандартный метод ADMM для решения задачи оптимизации имеет вид:

$$\begin{aligned} & \text{минимизировать } f_1(x_1) + f_2(x_2) \\ & \text{при условии } A_1x_1 + A_2x_2 = b, \end{aligned}$$

где  $x_i \in \mathbb{R}^{m_i}$ ,  $A_i \in \mathbb{R}^{n \times m_i}$ ,  $b \in \mathbb{R}^n$  и  $f_i: \mathbb{R}^{m_i} \rightarrow \mathbb{R}$  — замкнутые выпуклые функции,  $i = 1, 2$ . Часто  $f_1(x)$  представляет собой гладкую функцию, а  $f_2(x)$  — негладкую функцию, например штрафную функцию в задаче LASSO, которая часто используется для выбора переменных, где всего несколько ненулевых регрессионных параметров. Такие задачи часто возникают при обучении нейронных сетей [1], в биоинформатике [2], вычислительной статистике [3] и т.д.

Главной идеей ADMM является разделение исходной оптимизационной задачи на подзадачи, каждую из которых гораздо легче решить [4]. Это справедливо как для классического двухблочного ADMM, так и для многоблочного подхода [5–9]. На каждой итерации базового алгоритма минимизируется расширенный лагранжиан отдельно по каждой переменной  $x_i$ .

$$L(x_1, x_2, \lambda, \rho) = \sum_{i=1}^2 f_i(x_i) - \lambda^T \left( \sum_{i=1}^2 A_i x_i - b \right) + \frac{\rho}{2} \left\| \sum_{i=1}^2 A_i x_i - b \right\|_2^2.$$

Итерационная схема ADMM приведена ниже:

$$\begin{cases} x_1^{k+1} = \arg \min_{x_1} L(x_1, x_2^k, \lambda^k, \rho), \\ x_2^{k+1} = \arg \min_{x_2} L(x_1^{k+1}, x_2, \lambda^k, \rho), \\ \lambda^{k+1} = \lambda^k - \rho(A_1 x_1^{k+1} + A_2 x_2^{k+1} - b), \end{cases}$$

где  $x_2^0 \in \mathbb{R}^{m_2}$ ,  $\lambda^0 \in \mathbb{R}^n$ ,  $\rho > 0$ .

Метод Нестерова часто используется для ускорения градиентного метода.

Итерационная схема Нестерова имеет порядок сходимости  $O(1/k^2)$ , что является

© В.А. ГРИГОРЕНКО, Д.А. КЛЮШИН, С.И. ЛЯШКО, 2021

*Международный научно-технический журнал*

*«Проблемы управления и информатики», 2021, № 4*

оптимальной скоростью сходимости для класса липшицевых функций [10]. Ускорение достигается за счет вычисления релаксационного шага на каждой итерации. Итерационная схема ускорения Нестерова для градиентного метода минимизации целевой функции  $F(x)$ ,  $x \in \mathbb{R}^n$ ,  $F: \mathbb{R}^n \rightarrow \mathbb{R}$  приведена ниже:

$$\begin{cases} x_k = y_k - \tau \nabla F(y_k), \\ \alpha_{k+1} = (1 + \sqrt{4\alpha_k^2 + 1}) / 2, \\ y_{k+1} = x_k + (\alpha_k - 1)(x_k - x_{k-1}) / \alpha_{k+1} \end{cases}$$

и параметры удовлетворяют условиям  $\alpha_0 = 1$ ,  $x_0 = y_1 \in \mathbb{R}^n$ ,  $\tau < 1/K$ , где  $K$  — константа Липшица для  $\nabla F$ .

Этот подход получил широкое применение для ускорения проксимальных методов [11]. После этого подход Нестерова был успешно применен для итерационных схем с разбиением (например, FISTA) [12]. В [13] этот подход был обобщен для ADMM-метода, где было предложено к обычной итерационной схеме ADMM прибавить коррекционный параметр, предложенный Нестеровым в [10], на каждом шаге итерации при вычислении  $x_1$ ,  $x_2$ ,  $\lambda$  для ускорения. Метод сходится, если оба слагаемых целевой функции являются выпуклыми.

Итерационная схема для ускоренного ADMM с расширенным лагранжианом  $L$  приведена ниже:

$$\begin{cases} x_1^{k+1} = \arg \min_{x_1} L(x_1, x_2^k, \lambda^k, \rho), \\ x_2^{k+1} = \arg \min_{x_2} L(x_1^{k+1}, x_2, \lambda^k, \rho), \\ \lambda^{k+1} = \lambda^k - \rho(A_1 x_1^{k+1} + A_2 x_2^{k+1} - c), \\ \alpha_{k+1} = \frac{1 + \sqrt{1 + 4\alpha_k^2}}{2}, \\ \hat{x}_{k+1}^{(i)} = x_k^{(i)} + \frac{\alpha_k - 1}{\alpha_{k+1}} (x_k^{(i)} - x_{k-1}^{(i)}), \\ \hat{\lambda}_{k+1} = \lambda_k + \frac{\alpha_k - 1}{\alpha_{k+1}} (\lambda_k - \lambda_{k-1}), \end{cases}$$

где  $x_2^0 \in \mathbb{R}^{m_2}$ ,  $\lambda^0 \in \mathbb{R}^n$ ,  $\tau > 0$ ,  $\alpha_0 = 1$ .

Скорость сходимости  $O(1/k^2)$  данного метода для гладких выпуклых функций была доказана в [13], и он является более эффективным, чем классический ADMM.

В [14] было рассмотрено решение выпуклой мультиблочной оптимизационной задачи с использованием ADMM и предложена его модификация, которая расширяет данный подход путем разбиения существующей задачи на  $N$  подзадач и решения их отдельно на каждой итерации. Модифицированный метод имеет лучшую производительность, чем классический двухблочный подход.

Целью данной статьи является разработка и исследование новой модификации метода ADMM с более высокой скоростью сходимости по сравнению с уже

существующими методами. Для достижения поставленной задачи предлагаем одновременно использовать мультиблочный ADMM и ускоренный ADMM (ускорение Нестерова).

### Многоблочная выпуклая оптимизация

Рассмотрим оптимизационную задачу с  $N$  блоками, где целевая функция является суммой  $N > 2$  выпуклых функций:

$$\begin{aligned} & \text{минимизировать } \sum_{i=1}^N f_i(x_i) \\ & \text{при условии } \sum_{i=1}^N A_i x_i = b. \end{aligned}$$

Многоблочная выпуклая оптимизационная задача имеет много практических применений.

Например, задача RPCA (Robust Principal Component Analysis) [15] возникает, когда необходимо восстановить разреженный компонент матрицы  $L_0$  по сильно возмущенным измерениям  $M = L_0 + S_0$ .

Классическая итерационная схема для многоблочного ADMM имеет вид [14]:

$$\begin{cases} x_i^{k+1} = \arg \min_{x_i \in X_i} f_i(x_i) + \frac{\beta}{2} \left\| \sum_{j=1}^{i-1} A_j x_j^{k+1} + A_i x_i + \sum_{j=i+1}^m A_j x_j^k - b - \frac{\lambda^k}{\beta} \right\|^2, \\ \lambda^{k+1} = \lambda^k - \beta \left( \sum_{i=1}^m A_i x_i^{k+1} - b \right), \end{cases}$$

где  $x_i^0 \in \mathbb{R}^{m_i}$ ,  $\lambda^0 \in \mathbb{R}^n$ ,  $\beta > 0$ .

Сходимость ADMM была доказана для случая двух переменных. Для многоблочного ADMM сходимость была доказана в [16–18]. Метод является сходящимся, если выполняются три условия: для каждого  $i$  матрица  $A_i$  имеет полный ранг и матрицы  $A_i$  являются ортогональными; параметр  $\beta$  является достаточно маленьким; для каждой  $N$  подзадачи добавлен проксимальный шаг (в этом случае нет необходимости накладывать ограничения на  $A_i$ ) [18, 19].

Предлагается новая модификация многоблочного ADMM, которая основывается на добавлении ускорения Нестерова, подобного модификации двухблочного ADMM, описанного в [13]. На каждой итерации вычисляем шаг Нестерова и используем для коррекции результатов:

$$\begin{aligned} \alpha_{k+1} &= \frac{1 + \sqrt{1 + 4\alpha_k^2}}{2}, \quad \alpha_0 = 1, \\ \hat{x}_i^{(k+1)} &= x_i^{(k)} + \frac{\alpha_k - 1}{\alpha_{k+1}} (x_i^{(k)} - x_i^{(k-1)}), \\ \hat{\lambda}_{k+1} &= \lambda_k + \frac{\alpha_k - 1}{\alpha_{k+1}} (\lambda_k - \lambda_{k-1}), \end{aligned}$$

где  $x_i^0 \in \mathbb{R}^{m_i}$ ,  $\lambda^0 \in \mathbb{R}^n$ ,  $\beta > 0$ .

Это позволяет увеличить скорость сходимости. Предложенная итерационная схема приведена ниже:

$$\left\{ \begin{array}{l} x_i^{(k)} = \arg \min_{x_i \in X_i} f_i(x_i) + \frac{\beta}{2} \left\| \sum_{j=1}^{i-1} A_j \hat{x}_j^{(k)} + A_i x_i + \sum_{j=i+1}^m A_j x_j^{(k-1)} - b - \frac{\lambda_k}{\beta} \right\|^2, \\ \lambda^k = \hat{\lambda}^k - \beta \left( \sum_{i=1}^m A_i x_i^{(k)} - b \right), \\ \alpha_{k+1} = \frac{1 + \sqrt{1 + 4\alpha_k^2}}{2}, \\ \hat{x}_i^{(k+1)} = x_i^{(k)} + \frac{\alpha_k - 1}{\alpha_{k+1}} (x_i^{(k)} - x_i^{(k-1)}), \\ \hat{\lambda}_{k+1} = \lambda_k + \frac{\alpha_k - 1}{\alpha_{k+1}} (\lambda_k - \lambda_{k-1}). \end{array} \right.$$

где  $x_i^0 \in \mathbb{R}^{m_i}$ ,  $\lambda^0 \in \mathbb{R}^n$ ,  $\beta > 0$ ,  $\alpha_0 = 1$ .

### Доказательство сходимости многоблочного ADMM с ускорением Нестерова

Рассмотрим итерационную схему:

$$\begin{aligned} x_i^k &= \arg \min_{x_i \in X_i} f_i(x_i) + \frac{\beta}{2} \left\| \sum_{j=1}^{i-1} A_j \hat{x}_j^{(k)} + A_i x_i + \sum_{j=i+1}^m A_j x_j^{(k-1)} - b - \frac{\lambda_k}{\beta} \right\|^2, \\ \lambda^k &= \hat{\lambda}^k - \beta \left( \sum_{i=1}^m A_i x_i^{(k)} - b \right), \\ \alpha_{k+1} &= \frac{1 + \sqrt{1 + 4\alpha_k^2}}{2}, \\ \hat{x}_i^{(k+1)} &= x_i^{(k)} + \frac{\alpha_k - 1}{\alpha_{k+1}} (x_i^{(k)} - x_i^{(k-1)}), \\ \hat{\lambda}_i^{(k+1)} &= \lambda_i^{(k)} + \frac{\alpha_k - 1}{\alpha_{k+1}} (\lambda_i^{(k)} - \lambda_i^{(k-1)}). \end{aligned}$$

Введем обозначения:

$$x = \begin{pmatrix} x_1 \\ \dots \\ x_m \end{pmatrix}, \quad f(x) = \sum_{i=1}^m f_i(x_i), \quad w = \begin{pmatrix} x_1 \\ \dots \\ x_m \\ \lambda \end{pmatrix}, \quad F(w) = \begin{pmatrix} -A_1^T \lambda \\ \dots \\ -A_m^T \lambda \\ \sum_{i=1}^m A_i x_i - b \end{pmatrix},$$

где  $x_1 \in X_1, \dots, x_m \in X_m, \lambda \in \mathbb{R}^n$ .

$$W = X_1 \times X_2 \times \dots \times X_m \subset \mathbb{R}^n, \quad W^* = \bigcap_{w \in W} \{ \tilde{w} \in W : f(x) - f(\tilde{x}) + (w - \tilde{w})^T F(w) \geq 0 \}.$$

Положим

$$S = \begin{pmatrix} 0 & -A_1^T H A_2 & \dots & -A_1^T H A_m & 0 \\ -A_2^T H A_1 & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & -A_{m-1}^T H A_m & \dots \\ -A_m^T H A_1 & \dots & -A_m^T H A_{m-1} & 0 & 0 \\ 0 & \dots & \dots & 0 & H^{-1} \end{pmatrix},$$

$$G = \begin{pmatrix} 2A_1^T H A_1 & A_1^T H A_2 & \dots & A_1^T H A_m & 0 \\ A_2^T H A_1 & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & A_{m-1}^T H A_m & \dots \\ A_m^T H A_1 & \dots & A_m^T H A_{m-1} & 2A_m^T H A_m & 0 \\ 0 & \dots & \dots & 0 & H^{-1} \end{pmatrix},$$

$$P = (A_1, A_2, \dots, A_m, 0)^T H (A_1, A_2, \dots, A_m, 0),$$

где  $H$  — скалярная матрица,  $H = \beta I$ , где  $I$  — единичная матрица,  $\beta > 0$ .

Отсюда следует, что  $G = S + 2P$ .

Рассмотрим доказательство сходимости предложенного алгоритма.

**Лемма 1.**

$$\tilde{w}^k \in W, f(x) - f(\tilde{x}^k) + (w - \tilde{w}^k)^T \{F(\tilde{w}^k) + S(\tilde{w}^k - w^k)\} \geq 0 \quad \forall w \in W,$$

$$\tilde{x}_i^k \in X_i, f_i(x_i - \tilde{x}_i^k)^T \left\{ -A_i^T \lambda^k + A_i^T H \left( A \tilde{x}_i^k + \sum_{j=1, j \neq i}^m A_j x_j^k - b \right) \right\} \geq 0 \quad \forall x_i \in X_i.$$

*Доказательство.* Сделаем подстановку  $\tilde{\lambda}^k = \lambda^k - H \left( \sum_{j=1}^m A_j \tilde{x}_j^k - b \right)$ .

Получим:

$$\tilde{x}_i^k \in X_i,$$

$$f_i(x_i) - f_i(\tilde{x}_i^k) + (x_i - \tilde{x}_i^k)^T \left\{ -A_i^T \tilde{\lambda}^k - A_i^T H \left( \sum_{j=1, j \neq i}^m A_j (\tilde{x}_j^k - x_j^k) \right) \right\} \geq 0 \quad \forall x_i \in X_i.$$

(1)

$$f(x) - f(\tilde{x}^k) + \begin{pmatrix} x_1 - \tilde{x}_1^k \\ \vdots \\ x_i - \tilde{x}_i^k \\ \vdots \\ x_m - \tilde{x}_m^k \end{pmatrix}^T \left\{ \begin{pmatrix} -A_1^T \tilde{\lambda}^k \\ \vdots \\ -A_i^T \tilde{\lambda}^k \\ \vdots \\ -A_m^T \tilde{\lambda}^k \end{pmatrix} - \begin{pmatrix} A_1^T H \left( \sum_{j=2}^m A_j (\tilde{x}_j^k - x_j^k) \right) \\ \vdots \\ A_1^T H \left( \sum_{j=1, j \neq i}^m A_j (\tilde{x}_j^k - x_j^k) \right) \\ \vdots \\ A_m^T H \left( \sum_{j=1}^{m-1} A_j (\tilde{x}_j^k - x_j^k) \right) \end{pmatrix} \right\} \geq 0 \quad \forall w \in W.$$

Уравнение  $\lambda^k = \hat{\lambda} - \beta \left( \sum_{i=1}^m A_i x_i^{(k)} - b \right)$  можем переписать в виде:

$$\left( \sum_{j=1}^m A_j \tilde{x}_j^k - b \right) + H^{-1}(\tilde{\lambda}^k - \lambda^k) = 0,$$

что в вариационной форме эквивалентно:

$$\tilde{\lambda}^k \in \mathbb{R}^l, (\lambda - \tilde{\lambda}^k)^T \left\{ \left( \sum_{j=1}^m A_j \tilde{x}_j^k - b \right) + H^{-1}(\tilde{\lambda}^k - \lambda^k) \right\} \geq 0 \quad \forall \lambda \in \mathbb{R}^l. \quad (2)$$

Объединяя (1) и (2), получим  $\tilde{w}^k \in W$  и

$$f(x) - f(\tilde{x}^k) + \begin{pmatrix} x_1 - \tilde{x}_1^k \\ \vdots \\ x_i - \tilde{x}_i^k \\ \vdots \\ x_m - \tilde{x}_m^k \\ \lambda^k - \tilde{\lambda}^k \end{pmatrix}^T \left\{ \begin{pmatrix} -A_1^T \tilde{\lambda}^k \\ \vdots \\ -A_i^T \tilde{\lambda}^k \\ \vdots \\ -A_m^T \tilde{\lambda}^k \\ \sum_{j=1}^m A_j \tilde{x}_j^k - b \end{pmatrix} - \begin{pmatrix} A_1^T H \left( \sum_{j=2}^m A_j (\tilde{x}_j^k - x_j^k) \right) \\ \vdots \\ A_i^T H \left( \sum_{j=1, j \neq i}^m A_j (\tilde{x}_j^k - x_j^k) \right) \\ \vdots \\ A_m^T H \left( \sum_{j=1}^{m-1} A_j (\tilde{x}_j^k - x_j^k) \right) \\ H^{-1}(\tilde{\lambda}^k - \lambda^k) \end{pmatrix} \right\} \geq 0 \quad \forall w \in W.$$

Используя введенные выше обозначения  $F$  и  $S$ , полученное равенство можно записать в виде:

$$\tilde{w}^k \in W, f(x) - f(\tilde{x}^k) + (w - \tilde{w}^k)^T \{F(\tilde{w}^k) + S(\tilde{w}^k - w^k)\} \geq 0 \quad \forall w \in W.$$

**Лемма 2.**

$$(\tilde{w}^k - w^*)^T S(w^k - \tilde{w}^k) \geq 0 \quad \forall w^* \in W^*.$$

*Доказательство.* Справедливость неравенства следует из леммы 1 и монотонности функционала  $F$ . Для фиксированного  $w^* \in W^*$  имеем

$$(\tilde{w}^k - w^*)^T S(w^k - \tilde{w}^k) \geq (w^k - w^*)^T F(\tilde{w}^k) + f(\tilde{x}^k) - f(x^*) \quad \forall w^* \in W^*.$$

Используя монотонность функционала  $F$  и оптимальность  $w^*$ , имеем:

$$(\tilde{w}^k - w^*)^T F(\tilde{w}^k) + f(\tilde{x}^k) - f(x^*) \geq (\tilde{w}^k - w^*)^T F(w^*) + f(\tilde{x}^k) - f(x^*) \geq 0.$$

Объединив два приведенных выше неравенства, получаем утверждение леммы 2.

Здесь и далее  $\|x\|_H = \sqrt{x^T H x}$ , где  $H$  — положительно-определенная матрица

**Лемма 3.**

$$(w^k - w^*)^T G(w^k - \tilde{w}^k) \geq \varphi(w^k, \tilde{w}^k) \quad \forall w^* \in W^*,$$

$$\text{где } \varphi(w^k, \tilde{w}^k) = \|w^k - \tilde{w}^k\|_G^2 + 2(\lambda^k - \tilde{\lambda}^k)^T \left( \sum_{i=1}^m A_i (x_i^k - \tilde{x}_i^k) \right).$$

*Доказательство.* Используя утверждение  $G = S + 2P$ , докажем сначала, что

$$(\tilde{w}^k - w^*)^T P(w^k - \tilde{w}^k) = (\lambda^k - \tilde{\lambda}^k)^T \left( \sum_{i=1}^m A_i (x_i^k - \tilde{x}_i^k) \right) \quad \forall w^* \in W^*.$$

Поскольку  $P = (A_1, A_2, \dots, A_m, 0)^T H(A_1, A_2, \dots, A_m, 0)$ , имеем

$$(\tilde{w}^k - w^*)^T P(w^k - \tilde{w}^k) = \left( \sum_{i=1}^m A_i(\tilde{x}_i^k - x_i^*) \right)^T H \left( \sum_{i=1}^m A_i(x_i^k - \tilde{x}_i^k) \right).$$

Используя тот факт, что  $\sum_{i=1}^m A_i x_i^* = b$  и  $H \left( \sum_{i=1}^m A_i \tilde{x}_i^k - b \right) = \lambda^k - \tilde{\lambda}^k$ , получим

$$\left( \sum_{i=1}^m A_i(\tilde{x}_i^k - x_i^*) \right)^T H \left( \sum_{i=1}^m A_i(x_i^k - \tilde{x}_i^k) \right) = (\lambda^k - \tilde{\lambda}^k)^T \left( \sum_{i=1}^m A_i(x_i^k - \tilde{x}_i^k) \right).$$

Прибавим к обеим частям уравнения

$$(\tilde{w}^k - w^*)^T (2P)(w^k - \tilde{w}^k) = 2(\lambda^k - \tilde{\lambda}^k)^T \left( \sum_{i=1}^m A_i(x_i^k - \tilde{x}_i^k) \right)$$

и используем условие, что  $G = S + 2P$ . Имеем

$$(\tilde{w}^k - w^*)^T G(w^k - \tilde{w}^k) \geq 2(\lambda^k - \tilde{\lambda}^k)^T \left( \sum_{i=1}^m A_i(x_i^k - \tilde{x}_i^k) \right).$$

Утверждение леммы следует из приведенного неравенства и определения  $\varphi(w^k, \tilde{w}^k)$ .

Дальше необходимо доказать, что  $\varphi(w^k, \tilde{w}^k) \geq \delta \|w^k - \tilde{w}^k\|_G^2$  для некоторого  $\delta > 0$ .

**Лемма 4.**

$$\varphi(w^k, \tilde{w}^k) \geq \frac{\alpha_k - 1}{\alpha_{k+1}} \|w^k - \tilde{w}^k\|_G^2.$$

*Доказательство.* Воспользовавшись определением  $G$ , имеем

$$\|w^k - \tilde{w}^k\|_G^2 = \sum_{i=1}^m \|A_i(x_i^k - \tilde{x}_i^k)\|_H^2 + \left\| \sum_{i=1}^m A_i(x_i^k - \tilde{x}_i^k) \right\|_H^2 + \|\lambda^k - \tilde{\lambda}^k\|_{H^{-1}}^2.$$

Подставим полученное значение в определение  $\varphi(w^k, \tilde{w}^k)$ :

$$\varphi(w^k, \tilde{w}^k) = \sum_{i=1}^m \|A_i(x_i^k - \tilde{x}_i^k)\|_H^2 + \left\| \sum_{i=1}^m A_i(x_i^k - \tilde{x}_i^k) + H^{-1}(\lambda^k - \tilde{\lambda}^k) \right\|_H^2.$$

Следовательно, имеем

$$\frac{\varphi(w^k, \tilde{w}^k)}{\|w^k - \tilde{w}^k\|_G^2} = \frac{\sum_{i=1}^m \|A_i(x_i^k - \tilde{x}_i^k)\|_H^2 + \left\| \sum_{i=1}^m A_i(x_i^k - \tilde{x}_i^k) + H^{-1}(\lambda^k - \tilde{\lambda}^k) \right\|_H^2}{\sum_{i=1}^m \|A_i(x_i^k - \tilde{x}_i^k)\|_H^2 + \left\| \sum_{i=1}^m A_i(x_i^k - \tilde{x}_i^k) \right\|_H^2 + \|\lambda^k - \tilde{\lambda}^k\|_{H^{-1}}^2}.$$

Необходимо доказать неравенство

$$\left\| \sum_{i=1}^m A_i(x_i^k - \tilde{x}_i^k) \right\|_H^2 + \|\lambda^k - \tilde{\lambda}^k\|_{H^{-1}}^2 \geq \left\| \sum_{i=1}^m A_i(x_i^k - \tilde{x}_i^k) + H^{-1}(\lambda^k - \tilde{\lambda}^k) \right\|_H^2.$$

Введем обозначения:

$$\begin{aligned}
a_1 &= \sum_{i=1}^m \left\| A_i(x_i^k - \tilde{x}_i^k) \right\|_H^2, \quad b_1 = \sum_{i=1}^m \left\| A_i(x_i^k - \tilde{x}_i^k) \right\|_H^2 + \left\| \lambda^k - \tilde{\lambda}^k \right\|_{H^{-1}}^2, \\
b_2 &= \sum_{i=1}^m \left\| A_i(x_i^k - \tilde{x}_i^k) + H^{-1}(\lambda^k - \tilde{\lambda}^k) \right\|_H^2, \quad a_2 = \frac{1}{m} \left\| \sum_{i=1}^m A_i(x_i^k - \tilde{x}_i^k) \right\|_H^2, \\
&= \frac{\varphi(w^k, \tilde{w}^k)}{\left\| w^k - \tilde{w}^k \right\|_G^2} = \frac{a_1 + b_2}{a_1 + b_1} \geq \frac{a_2 + b_2}{a_2 + b_1} = \\
&= \frac{\frac{1}{m} \left\| \sum_{i=1}^m A_i(x_i^k - \tilde{x}_i^k) \right\|_H^2 + \sum_{i=1}^m \left\| A_i(x_i^k - \tilde{x}_i^k) + H^{-1}(\lambda^k - \tilde{\lambda}^k) \right\|_H^2}{\frac{1}{m} \left\| \sum_{i=1}^m A_i(x_i^k - \tilde{x}_i^k) \right\|_H^2 + \sum_{i=1}^m \left\| A_i(x_i^k - \tilde{x}_i^k) \right\|_H^2 + \left\| \lambda^k - \tilde{\lambda}^k \right\|_{H^{-1}}^2}, \\
\frac{\varphi(w^k, \tilde{w}^k)}{\left\| w^k - \tilde{w}^k \right\|_G^2} &\geq \frac{\frac{m+1}{m} \left\| \sum_{i=1}^m A_i(x_i^k - \tilde{x}_i^k) \right\|_H^2 + 2(\lambda^k - \tilde{\lambda}^k)^T \left( \sum_{i=1}^m A_i(x_i^k - \tilde{x}_i^k) \right) + \left\| \lambda^k - \tilde{\lambda}^k \right\|_{H^{-1}}^2}{\frac{m+1}{m} \left\| \sum_{i=1}^m A_i(x_i^k - \tilde{x}_i^k) \right\|_H^2 + \left\| \lambda^k - \tilde{\lambda}^k \right\|_{H^{-1}}^2} \geq \\
&\geq \frac{\left( \frac{\alpha_{k+1}}{\alpha_{k+1} - \alpha_k + 1} \right)^2 \left\| \sum_{i=1}^m A_i(x_i^k - \tilde{x}_i^k) \right\|_H^2 + 2(\lambda^k - \tilde{\lambda}^k)^T \left( \sum_{i=1}^m A_i(x_i^k - \tilde{x}_i^k) \right) + \left\| \lambda^k - \tilde{\lambda}^k \right\|_{H^{-1}}^2}{\left( \frac{\alpha_{k+1}}{\alpha_{k+1} - \alpha_k + 1} \right)^2 \left\| \sum_{i=1}^m A_i(x_i^k - \tilde{x}_i^k) \right\|_H^2 + \left\| \lambda^k - \tilde{\lambda}^k \right\|_{H^{-1}}^2}.
\end{aligned}$$

Воспользуемся неравенством

$$\frac{\tau \left\| p \right\|_H^2 + 2p^T q + \left\| q \right\|_{H^{-1}}^2}{\tau \left\| p \right\|_H^2 + \left\| q \right\|_{H^{-1}}^2} \geq 1 - \sqrt{\frac{1}{\tau}},$$

где  $H$  — положительно-определенная матрица,  $\tau > 0$ .

Положим

$$\tau = \left( \frac{\alpha_{k+1}}{\alpha_{k+1} - \alpha_k + 1} \right)^2, \quad p = \sum_{i=1}^m A_i(x_i^k - \tilde{x}_i^k), \quad q = \lambda^k - \tilde{\lambda}^k.$$

Утверждение леммы 4 доказано.

**Лемма 5.**

$$\psi(\alpha) \geq q(\alpha),$$

$$\text{где } \psi(\alpha) = \left\| w^k - w^* \right\|_G^2 - \left\| w^{k+1}(\alpha) - w^* \right\|_G^2, \quad q(\alpha) = 2\alpha\varphi(w^k, \tilde{w}^k) - \alpha^2 \left\| w^k - \tilde{w}^k \right\|_G^2.$$

*Доказательство.* Используя лемму 3 и определение  $q(\alpha)$ , получим:

$$\begin{aligned}
\psi(\alpha) &= \left\| w^k - w^* \right\|_G^2 - \left\| w^{k+1}(\alpha) - w^* \right\|_G^2 = \left\| w^k - w^* \right\|_G^2 - \left\| (w^k - w^*) - \alpha(w^k - \tilde{w}^k) \right\|_G^2 = \\
&= 2\alpha(w^k - w^*)^T G(w^k - \tilde{w}^k) - \alpha^2 \left\| w^k - \tilde{w}^k \right\|_G^2 \geq 2\alpha\varphi(w^k, \tilde{w}^k) - \alpha^2 \left\| w^k - \tilde{w}^k \right\|_G^2 = q(\alpha).
\end{aligned}$$

Используем квадратичную функцию  $q(\alpha)$ , которая является минорантой  $\psi(\alpha)$ :

$$\alpha_k^* = \frac{\varphi(w^k, \tilde{w}^k)}{\|w^k - \tilde{w}^k\|_G^2}, \alpha_k = \gamma \alpha_k^*.$$

Из леммы 4 имеем:

$$\alpha_k^* \geq \frac{\alpha_k - 1}{\alpha_{k+1}}.$$

Докажем, что

$$\|w^{k+1} - w^*\|_G^2 \leq \|w^k - w^*\|_G^2 - \gamma(2-\gamma) \left( \frac{\alpha_k - 1}{\alpha_{k+1}} \right)^2 \|w^k - \tilde{w}^k\|_G^2, \gamma \in (0, 2).$$

Для любого  $\alpha > 0$  из леммы 2 следует:

$$\|w^{k+1} - w^*\|_G^2 \leq \|w^k - w^*\|_G^2 - q(\alpha) \forall w^* \in W^*,$$

$$q(\alpha) = 2\gamma \left( \frac{\alpha_k - 1}{\alpha_{k+1}} \right) \varphi(w^k, \tilde{w}^k) - \gamma^2 \left( \frac{\alpha_k - 1}{\alpha_{k+1}} \right)^2 \|w^k - \tilde{w}^k\|_G^2.$$

Из леммы 4 получаем:

$$q(\alpha) \geq \gamma(2-\gamma) \left( \frac{\alpha_k - 1}{\alpha_{k+1}} \right)^2 \|w^k - \tilde{w}^k\|_G^2.$$

**Лемма 6.** Пусть  $\{w^k\}$  — последовательность, сгенерированная многоблочным ADMM с ускорением Нестерова. Докажем, что

$$\|w^{k+1} - w^*\|_G^2 \leq \|w^k - w^*\|_G^2 - \frac{2-\gamma}{\gamma} \|w^k - w^{k+1}\|_G^2 \forall w^* \in W^*.$$

Поскольку  $\alpha(w^k - \tilde{w}^k) = (w^k - w^{k+1})$  и  $\alpha = \gamma \left( \frac{\alpha_k - 1}{\alpha_{k+1}} \right)$ , то неравенство следует из леммы 5.

Введем следующие обозначения:

$$y_i = A_i x_i, i = 1, \dots, m, v = (y_1, y_2, \dots, y_m, \lambda),$$

$$V^* = \{(A_1 x_1^*, A_2 x_2^*, \dots, A_m x_m^*, \lambda^*) | (x_1^*, x_2^*, \dots, x_m^*, \lambda^*) \in W^*\}.$$

**Теорема.** Пусть  $\{w^k\}$  — последовательность, сгенерированная многоблочным ADMM с ускорением Нестерова, тогда:

- если все  $A_i$  имеют полный ранг, то  $w^k \rightarrow w^*, w^* \in W^*$ ;
- иначе —  $v^k \rightarrow v^*, v^* \in V^*$ .

*Доказательство.* Если все  $A_i$  имеют полный ранг, то матрица  $G$  положительно-определенная, и для последовательности  $\{w^k\}$  выполняются условия леммы 6, следовательно  $w^k \rightarrow w^*, w^* \in W^*$ . В противном случае сходимость следует из неравенства:

$$\|v^{k+1} - v^*\|_H^2 \leq \|v^k - v^*\|_H^2 - \frac{2-\gamma}{\gamma} \|v^k - v^{k+1}\|_H^2 \forall v^* \in V^*,$$

$$\text{где } H = \begin{pmatrix} 2H & H & \dots & H & 0 \\ H & \ddots & \ddots & \vdots & \vdots \\ \vdots & \ddots & \ddots & H & \vdots \\ H & \dots & H & 2H & 0 \\ 0 & \dots & \dots & 0 & H^{-1} \end{pmatrix}_{(m+1) \times (m+1)}.$$

### Вычислительные эксперименты

В этом разделе сравним эффективность модифицированного многоблочного ADMM с ускорением Нестерова с обычным многоблочным и классическим двухблочным ADMM-методами. Для этого рассмотрим задачу обмена (exchange problem), задачу выбора базиса (basis pursuit) и задачу LASSO с ограничениями [20].

Подобные задачи часто возникают при исследовании оптимального управления классическими и неклассическими уравнениями математической физики [21–30].

**Задача обмена.** Пусть имеем сеть с  $N$  агентами, которые обмениваются  $n$  продуктами. Каждый агент  $i$  имеет некоторую функцию стоимости  $f_i: \mathbb{R}^n \rightarrow \mathbb{R}$ . Задача обмена имеет вид

$$\min_{\{x_i\}} \sum_{i=1}^N f_i(x_i), \quad \sum_{i=1}^N x_i = 0.$$

Рассмотрим квадратичные функции стоимости  $f_i(x_i) = \frac{1}{2} \|C_i x_i - d_i\|^2$ , где  $C_i \in \mathbb{R}^{p \times n}$ ,  $d_i \in \mathbb{R}^p$ . Тогда на каждом шаге итерации имеем:

$$x_i^{k+1} = \arg \min_{x_i} \frac{1}{2} \|C_i x_i\|^2 + \frac{\rho}{2} \|x_i - b_i^k\|^2 \quad (b_i^k \text{ — не зависящий от } x \text{ вектор}).$$

Для эксперимента сгенерируем случайный вектор  $x_i^*$  по стандартному нормальному распределению,  $C_i$  — случайные гауссовы матрицы,  $d_i = C_i x_i^*$ . Размерность тестовой задачи была установлена как  $n=100$ ,  $N=100$ ,  $p=80$ . Для проверки работы алгоритма вычислено 400 итераций и построены графики зависимости целевой функции и погрешности от количества итераций (рис. 1).

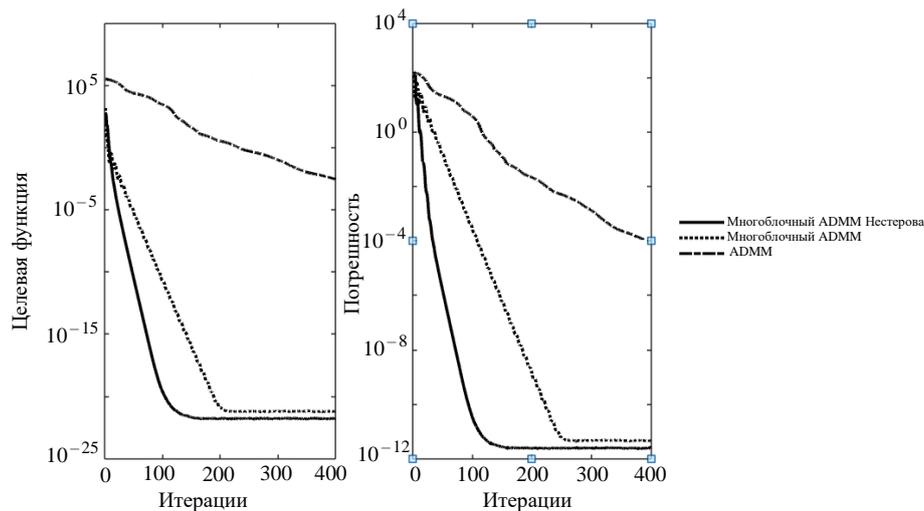


Рис. 1

**Задача выбора базиса.** Формулировка этой задачи выглядит так:

$$\begin{aligned} & \text{минимизировать } \|x\|_1 \\ & \text{при условии } Ax = b, \end{aligned}$$

где  $x \in \mathbb{R}^m$ ,  $A \in \mathbb{R}^{n \times m}$ ,  $b \in \mathbb{R}^n$ . Данная задача обычно возникает, когда имеем неопределенную систему линейных уравнений и необходимо восстановить вектор  $x$  по небольшому количеству наблюдений  $b$  ( $n \ll m$ ) [14]. Разделим данные на  $m$  блоков:  $x = (x_1, x_2, \dots, x_m)$  и  $A = (A_1, A_2, \dots, A_m)$ . Тогда можно представить задачу в виде (1) и применить итерационную схему многоблочного ADMM.

Каждая итерация может быть вычислена как:

$$\begin{aligned} x_i^{(k)} &= \arg \min_{x_i \in X_i} \|x_i\|_1 + \frac{\beta}{2} \left\| \sum_{j=1}^{i-1} A_j \hat{x}_j^{(k)} + A_i x_i + \sum_{j=i+1}^m A_j x_j^{(k-1)} - b - \frac{\lambda_k}{\beta} \right\|^2 = \\ &= \text{shrinkage} \left( \frac{\beta A_i^T}{\sqrt{a_i^T a_i}} \left[ b + \frac{\lambda^k}{\beta} - \sum_{j=1}^{i-1} A_j \hat{x}_j^{(k)} - \sum_{j=i+1}^m A_j x_j^{(k-1)} \right], \frac{1}{\beta(a_i^T a_i)} \right). \end{aligned}$$

Для тестирования сгенерируем случайный вектор  $x^*$  с  $k$  ( $k \ll n$ ) ненулевыми элементами со стандартным нормальным распределением. Матрица  $A$  также сгенерирована в соответствии со стандартным нормальным распределением и разделена на  $m$  блоков. Сгенерированный набор данных имеет  $n = 500$  строк и  $m = n \times 2,5$  столбцов. Было вычислено 300 итераций алгоритма и показана зависимость средней погрешности от количества итераций для 100 запусков алгоритма. Для многоблочного ADMM и многоблочного ADMM с ускорением Нестерова целевая функция была представлена в виде суммы пяти функций. График зависимости значений целевой функции от количества итераций представлен ниже (рис. 2).

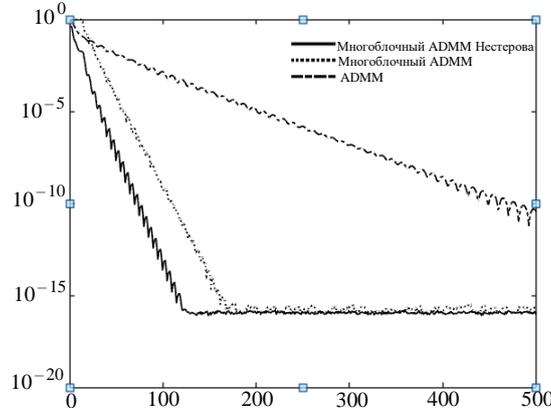


Рис. 2

**Задача LASSO с ограничениями.** Формулировка этой задачи выглядит следующим образом:

$$\begin{aligned} & \text{минимизировать } \frac{1}{2} \|Ax - b\|_2^2 + \lambda \|x\|_1 \\ & \text{при условии } Cx = d, x \geq 0, \end{aligned}$$

где  $x \in \mathbb{R}^m$ ,  $A \in \mathbb{R}^{n \times m}$ ,  $b \in \mathbb{R}^n$ . Задача часто возникает в статистике, машинном обучении и распознавании образов. Для тестирования сгенерируем случайный вектор  $x^*$  в соответствии со стандартным нормальным распределением. Матрицы  $A$  и  $C$  также

*Международный научно-технический журнал  
«Проблемы управления и информатики», 2021, № 4*

сгенерированы по стандартному нормальному распределению и разделены на  $m$  блоков. Сгенерированный набор данных имеет  $n = 1000$  строк и 500 столбцов. Всего было вычислено 200 итераций алгоритма. Проведенный эксперимент вновь показал, что мультиблочный ADMM с ускорением Нестерова сходится быстрее (рис. 3).

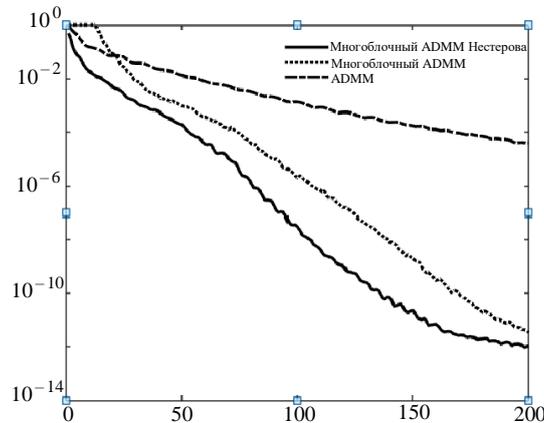


Рис. 3

### Заключение

Многоблочный ADMM является значительно более эффективным, чем стандартный двухблочный ADMM, а применение ускорения Нестерова на каждой итерации многоблочного ADMM существенно увеличивает скорость сходимости метода. ADMM является очень удобным инструментом для решения большого количества оптимизационных задач в математической статистике, машинном обучении, распознавании образов и очень хорошо подходит для распределенных вычислений. В статье рассмотрены несколько существующих методов и приемов, что позволило увеличить эффективность ADMM. Предложен новый подход, который основывается на комбинации многоблочного ADMM и вычислении ускорения Нестерова. Для предложенного метода приведено теоретическое доказательство сходимости. Для демонстрации эффективности в сравнении с существующими методами решено несколько задач оптимизации, в частности задача обмена (exchange problem), задача выбора базиса (basis pursuit) и задача LASSO с ограничениями.

*В.А. Григоренко, Д.А. Ключин, С.І. Ляшко*

### БАГАТОБЛОЧНИЙ МЕТОД ADMM З ПРИСКОРЕННЯМ НЕСТЕРОВА

Метод ADMM (alternating direction methods of multipliers) широко використовується для розв'язання багатьох оптимізаційних задач за допомогою паралельних обчислень. Цей метод є особливо важливим при розв'язанні задач, які виникають в машинному навчанні, математичній статистиці, розпізнаванні образів, відновленні сигналів, обробці даних значного об'єму. ADMM дає змогу розв'язувати оптимізаційні задачі, цільова функція яких являє собою суму гладкої та негладкої функцій, що є характерним для задач з штрафною функцією. Метою цієї статті є розробка покращеної версії ADMM з кращою швидкістю збіжності, ніж вже існуючі методи. В статті описано існуючі підходи для покращення ефективності ADMM-методу, наведено основні роботи з даної тематики та запропоновано новий метод, який базується на комбінації двох вже існуючих підходів — розбиття початкової оптимізаційної задачі на  $N$  підзадач та застосування багатоблочного підходу для розв'язання і обчислення прискорення Нестерова на кожній ітерації. Наведено теоретичне обґрунтування збіжності даного методу та встановлено умови збіжності. Реалізовано запропонований алгоритм мовою програмування Python і застосовано для розв'язання задачі

обміну з генерованими випадковим чином даними, задачі пошуку базису та задачі LASSO з обмеженнями. Наведено результати порівняння ефективності багатоблочного ADMM з прискоренням Нестерова та існуючих багатоблочного і стандартного двоблочного ADMM. Багатоблочний ADMM з прискоренням Нестерова показав кращу обчислювальну ефективність, ніж вже існуючі методи. Ще однією перевагою запропонованого методу є його зручність для проведення паралельних обчислень із застосуванням сучасних багатопроекторних систем. В зв'язку з великими об'ємами даних, обробка яких вимагає значного часу при розв'язанні оптимізаційних задач, запропонований метод має важливе практичне значення, оскільки він значно перевищує за швидкістю відомі аналоги. Використання запропонованого методу дасть можливість розв'язати практично важливі задачі великого обсягу, застосувавши паралельні обчислення.

**Ключові слова:** ADMM (alternating direction method of multipliers), багатоблочний ADMM, прискорений ADMM, прискорення Нестерова.

V.A. Hrygorenko, D.A. Klyushin, S.I. Lyashko

## MULTI-BLOCK ADMM METHOD WITH NESTEROV ACCELERATION

ADMM (alternating direction methods of multipliers) is widely used to solve many optimization problems. This method is especially important for solving problems arising in great variety of fields, especially in machine learning, mathematical statistics and pattern recognition, signal denoising and big data analysis using parallel computations. ADMM also useful for solving optimization problems in cases when objective function presented as sum of smooth and non-smooth functions. Standard two block ADMM can be extended for solving problems where objective function can be represented as sum of N functions (multiblock approach). In this paper we described some most common used technics used for acceleration of ADMM and reviewed most significant works related to this topic. The aim of this paper is to develop an improved version of the ADMM with better convergence rate. For this, we used a combination of two already existing approaches: splitting the initial optimization problem into subtasks and solving them in parallel using multiblock approach and calculating the Nesterov acceleration at each iteration step. We provided a theoretical justification for the convergence of this method, defined necessary for convergence conditions, and also implemented the proposed algorithm in the Python programming language and applied it to solve the problem of exchange with random data, basis pursuit problem and LASSO with restrictions problem. The article presents the results of comparing the effectiveness of the multiblock ADMM method with Nesterov acceleration and the existing multiblock and standard two-block ADMM method. Multiblock ADMM with Nesterov acceleration demonstrates better performance that already existing methods and also can be easily adopted for parallel calculation. Proposed method has great practical value due to necessity to solve optimization problems with great volumes of data, which requires high performance, because it works much more faster than well-known analogies. The use of the proposed method will make it possible to solve practically important problems of large volume using parallel calculations.

**Keywords:** ADMM (alternating direction method of multipliers), multiblock ADMM, fast ADMM, Nesterov acceleration.

1. Scardapone S., Comminiello D., Hussain A., Uncini A. Group sparse regularization for deep neural networks. *Neurocomputing*. 2017. **241**. P. 81–89. <https://doi.org/10.1016/j.neucom.2017.02.029>.
2. Guo Y., Wu C., Guo M., Zou Q., Liu X., Keinan A. Combining sparse group lasso and linear mixed model improves power to detect genetic variants underlying quantitative traits. *Frontiers in Genetics*. 2019. **10**. Article 271. (<https://www.frontiersin.org/articles/10.3389/fgene.2019.00271/full>) <https://doi.org/10.3389/fgene.2019.00271>.
3. Song, Q. An overview of reciprocal L1-regularization for high dimensional regression data. *Wiley Interdisciplinary Reviews: Computational Statistics*. 2018. **10**. e1416. <https://doi.org/10.1002/wics.1416>.
4. Boyd S., Parikh N., Chu E., Peleato B., Ekstein J. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends in Machine Learning*. 2011. **3**(1). P. 1–122. <https://doi.org/10.1561/22000000016>.

5. Lin T.Y., Ma S.Q., Zhang S.Z. On the sublinear convergence rate of multi-block ADMM. *Operations Research Society of China*. 2015. **3**, N 3. P. 251–274. <https://doi.org/10.1007/s40305-015-0092-0>.
6. Lin T.Y., Ma S.Q., Zhang S.Z. On the global linear convergence of the ADMM with multiblock variables. *SIAM Journal on Optimization*. 2015. **25**, N 3. P. 1478–1497. <https://doi.org/10.1137/140971178>.
7. Lin, T.Y., Ma, S.Q., Zhang, S.Z. Iteration complexity analysis of multi-block ADMM for a family of convex minimization without strong convexity. *Scientific Computing*. 2016. **69**, N 1. P. 52–81. <https://doi.org/10.1007/s10915-016-0182-0>.
8. Sun D.F., Toh K.C., Yang L.Q. A convergent 3-block semiproximal alternating direction method of multipliers for conic programming with 4-type constraints. *SIAM Journal on Optimization*. 2015. **25**, N 2. P. 882–915. <https://doi.org/10.1137/140964357>.
9. Tao M., Yuan X. Convergence analysis of the direct extension of ADMM for multiple-block separable convex minimization. *Advances in Computational Mathematics*. 2018. **44**, N 3. P. 773–813. <https://doi.org/10.1007/s10444-017-9560-x>.
10. Nesterov Y. A method of solving a convex programming problem with convergence rate  $O(1/k^2)$ . *Soviet Math. Dokl.* 1983. **27**, P. 372–376.
11. Nesterov Y. Gradient methods for minimizing composite functions. *Mathematical Programming*. 2013. **140**, N 1. P. 125–161. <https://doi.org/10.1007/s10107-012-0629-5>.
12. Beck A., Teboulle M. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM Journal of Imaging Sciences*. 2009. **2**. P. 183–202. <https://doi.org/10.1137/080716542>.
13. Goldstein T., O'Donoghue B., Setzer S. Fast alternating direction optimization methods. *SIAM Journal on Imaging Sciences*. 2014. **7**, N 3. P. 1588–1623. <https://doi.org/10.1137/120896219>.
14. Wang X., Hong M., Ma S. Solving multiple-block separable convex minimization problems using two-block alternating direction method of multipliers. *Pacific Journal of Optimization*. 2014. **11**, N 4. P. 645–667.
15. Candès E., Li X., Ma Y., Wright J. Robust principal component analysis? *Journal of the ACM*. 2011. **58**, N 1. P. 1–37. <https://doi.org/10.1145/1970392.1970395>.
16. Chen C., He B., Ye Y. The direct extension of ADMM for multi-block convex minimization problems is not necessarily convergent. *Mathematical Programming*. 2016. **155**, N 1. P. 57–79. <https://doi.org/10.1007/s10107-014-0826-5>.
17. Chen C., Shen Y., You Y. On the convergence analysis of the alternating direction method of multipliers with three blocks. *Abstract and Applied Analysis*. 2013. **183961**. <https://doi.org/10.1155/2013/183961>.
18. Hong M., Luo Z.-Q. On the linear convergence of the alternating direction method of multipliers. *Mathematical Programming*. 2017. **162**. P. 165–199. <https://doi.org/10.1007/s10107-016-1034-2>.
19. Deng W., Lai M., Peng Z., Yin W. Parallel Multi-Block ADMM with  $o(1/k)$  Convergence. *Journal of Scientific Computations*. 2017. **71**. P. 712–736. <https://doi.org/10.1007/s10915-016-0318-2>.
20. Goldstein T., Setzer S. High-order methods for basis pursuit. *UCLA CAM Report*. 2010. P. 10–41.
21. Klyushin D.A., Lyashko S.I., Lyashko N.I., Bondar O.S., Tymoshenko A.A. Generalized optimization of processes of drug transport in tumors. *Cybernetics and Systems Analysis*. 2020. **56**, N 5. P. 758–765. <https://doi.org/10.1007/s10559-020--00296-0>.
22. Sandrakov G.V., Lyashko S.I., Bondar E.S., Lyashko N.I. Modeling and optimization of micro-needle systems. *Journal of Automation and Information Sciences*. 2019. **51**, N 6. P. 1–11. <https://doi.org/10.1615/JAutomatInfScien.v51.i6.10>.
23. Lyashko S.I., Klyushin D.A., Tymoshenko A.A., Lyashko N.I., Bondar E.S. Optimal control of intensity of water point sources in unsaturated porous medium. *Journal of Automation and Information Sciences*. 2019. **51**, N 7. P. 24–33. <https://doi.org/10.1615/JAutomatInfScien.v51.i7.20>.
24. Tymoshenko A., Klyushin D., Lyashko S. Optimal control of point sources in Richards-Klute equation. *Advances in Intelligent Systems and Computing*. 2019. **754**. P. 194–203. [https://doi.org/10.1007/978-3-319-91008-6\\_20](https://doi.org/10.1007/978-3-319-91008-6_20).
25. Sergienko I.V., Lyashko S.I., Voitsekhovskii S.A. An approximate solution for a class of second-order elliptic variational inequalities in arbitrary-form domains. *Cybernetics and Systems Analysis*. 2004. **40**, N 4. P. 486–490. <https://doi.org/10.1023/B:CASA.0000047870.13325.c2>.
26. Lyashko S.I., Nomirovskii D.A. The generalized solvability and optimization of parabolic systems in domains with thin low-permeable inclusions. *Cybernetics and Systems Analysis*. 2003. **39**, N 5. P. 737–745. <https://doi.org/10.1023/B:CASA.0000012094.62199.de>.
27. Lyashko S.I., Semenov V.V. Controllability of linear distributed systems in classes of generalized actions. *Cybernetics and Systems Analysis*. 2001. **37**, N 1. P. 13–32. <https://doi.org/10.1023/A:1016607831284>.
28. Lyashko S.I., Nomirovskii D.A., Sergienko T.I. Trajectory and final controllability in hyperbolic and pseudohyperbolic systems with generalized actions. *Cybernetics and Systems Analysis*. 2001. **37**, N 4. P. 756–763. <https://doi.org/10.1023/A:1013871026026>.
29. Lyashko S.I., Klyushin D.A., Palienko L.I. Simulation and generalized optimization in pseudohyperbolic systems. *Journal of Automation and Information Sciences*. 2000. **32**, N 5. P. 108–117. <https://doi.org/10.1615/JAutomatInfScien.v32.i5.80>.
30. Lyashko S.I. Numerical solution of pseudoparabolic equations. *Cybernetics and Systems Analysis*. 1995. **31**, N 5. P. 718–722. <https://doi.org/10.1007/BF02366321>.

Получено 09.03.2021