Об одной модификации быстрого градиентного метода решения задачи энтропийно-линейного программирования

Чернов А.В.

В работе рассмотривается модификация быстрого градиентного метода (БГМ). Показана его прямо-двойственность как способность восстановить решение прямой задачи по решению двойственной. Получены теоретические результаты о его сходимости как для задач безусловной минимизации, так и для задач условной минимизации с линейными ограничениями-равенствами и ограничениями-неравенствами на примере задачи энтропийно-линейного программирования (задача ЭЛП). Доказаны строгая и сильная выпуклость двойственного функционала последней, а также показано, что градиент двойственного функционала удовлетворяет условию Липшица.

Ключевые слова: быстрый градиентый метод, задача энтропийно-линейного программирования, условная минимизация, безусловная минимизация, прямо-двойственные методы.

Постановка задачи

Рассмотрим на n-мерном вероятностном симплексе

$$S_n(1) = \{x \in R^n_+ : \sum_{i=1}^{n} x_i = 1\}$$

задачу ЭЛП (1).

$$
\begin{gathered}
f(x)=\sum_{i=1}^{n} g_i(x_i)\to\min_{x\in G};\\
g_i(x_i)=\begin{cases} x_i\ln\frac{x_i}{\xi_i}, \text{ если } x_i>0;\\ 0, \text{ иначе};\end{cases};\\
G=\{x\in S_n(1): C_1x-b_1\le 0; C_2x-b_2=0\}.
\end{gathered}\tag{1}
$$

Здесь $\xi_i, i=1,..,n$ – параметры задачи, определяющие целевую функцию, $C_1\in R^{m_1\times n}$, $b_1\in R^{m_1}$, $C_2\in R^{m_2\times n}$, $b_2\in R^{m_2}$ – параметры (матрицы и вектора соответственно), определяющие допустимое множество задачи.

Введем дополнительные обозначения (2) для упрощения дальнейших выкладок.

$$
C=[C_1;C_2]\in R^{(m_1+m_2)\times n};\quad b=[b_1;b_2]\in R^{m_1+m_2}.\tag{2}
$$

Пусть $D=D(x)$ – диагональная матрица размерности $n\times n$, на диагонали которой расположены компоненты вектора x, т.е. $d_{i,i}(x)=x_i$ и $d_{i,j}(x)=0$ при $i\neq j$. Видно что для такой задачи градиент функции $\nabla f(x)$, функция Лагранжа $L(x,y)$, её градиент $\nabla L(x,y)$ и гессиан принимают вид:

$$
\begin{gathered}
\nabla f(x)_i=\ln\frac{x_i}{\xi_i}+1,\quad i=\overline{1,n};\\
L(x,y)=f(x)+\langle y, Cx-b\rangle,\quad x\in S_n(1), y\in R_+^{m_1}\times R^{m_2};\\
\nabla L(x,y)_i=\ln\frac{x_i}{\xi_i}+1+[C^Ty]_i,\quad i=\overline{1,n};\\
L''(x,y)=f''(x)=D^{-1}(x),\quad x\in R_{++}^n.
\end{gathered}\tag{3}
$$

Несложно показать, что верна следующая лемма (см. например критерии выпуклости первого и второго порядков [1]):

Лемма 1. *Целевая функция задачи ЭЛП (1) выпукла на R_+^n строго выпукла на R_{++}^n и сильно выпукла на любом ограниченном подмножестве R_{++}^n.*

В силу того, что множество G – выпуклый компакт и функция $f(x)$ непрерывна и строго выпукла, задача (1) имеет единственное решение [1].

В рамках изучения различных методов решения задачи ЭЛП в работе подразумевается поиск точки в пространстве R^n, достаточно близкой к решению поставленной задачи. Формально критерий поиска такой точки задается следующим определением:

Определение 1. *Под $(\varepsilon_f, \varepsilon_g)$-решением задачи ЭЛП (1) будем понимать такую точку $x_t \in R^n$, что $|f(x_t) - f^*| < \varepsilon_f$ и $\Delta(x_t, G) < \varepsilon_g$, где f^* - точное решение задачи, а $\Delta(x, G) = ||(C_1 x - b_1)_+|| + ||C_2 x - b_2||$ - невязка точки x для множества G.*

Определение 1 совместно с определением задачи минимизации ЭЛП (1) дают постановку изучаемой задачи как поиск $(\varepsilon_f, \varepsilon_g)$-решения.

О двойственной задаче

Для задачи ЭЛП возможно в явном виде построить двойственную задачу. Переход к двойственной задаче позволяет перейти от задачи условной минимизации к задаче безусловной минимизации или условной минимизации, но на простом множестве вида $R_+^{m_1} \times R^{m_2}$ и тем самым открывается возможность использования методов решения задач минимизации на множестве простой структуры. Утверждения сформулированные ниже описывают постановку задачи двойственной к задаче ЭЛП и её основные свойства.

Теорема 1. *Задача, двойственная к задаче ЭЛП (1), может быть записана в виде:*

$$\begin{gathered} \psi(y) \to \max_{y \in Q}, \text{ где } Q = R_+^{m_1} \times R^{m_2}; \\ \psi(y) = -\langle y, b \rangle - \ln \left[\sum_{i=1}^{n} \xi_i \exp\left(-\left[C^T y \right]_i \right) \right]. \end{gathered} \tag{4}$$

При этом функция $x(y)$, полученная при построении двойственной задачи, вычисляется по формуле (5).

$$x_i(y) = \frac{\xi_i \exp\left(-[C^T y]_i \right)}{\sum\limits_{j=0}^{n} \xi_j \exp\left(-[C^T y]_j \right)}. \tag{5}$$

Доказательство. Используя представление функции Лагранжа (3) для задачи (1), двойственная задачу можно записать в виде:

$$\psi(u, v) = \min_{x \in S_n(1)} L(x, u, v),$$

где $y = (u, v)$. Решение такой задачи существует, единственно и находится явно. Действительно, выпишем "расширенную"функцию Лагранжа для такой задачи, полагая u и v параметрами:

$$\overline{L}(x,u,v,\mu,\nu) = f(x)+\langle u, C_1x - b_1\rangle+\langle v, C_2x - b_2\rangle+\mu(\sum_{i=1}^{n} x_i-1)-\langle \nu, x\rangle\,,$$

где $\mu \in R^1$, $\nu \in R^n_+$. Тогда условия Каруша-Куна-Таккера [1] принимают вид:

$$[\nabla \overline{L}(x,u,v,\mu,\nu)]_i = 1 + \ln\frac{x_i}{\xi_i} + [C_1^T u + C_2^T v]_i + \mu - \nu_i = 0, \quad i = \overline{1,n}.$$

.

$$\nu_i x_i = 0, \quad i = \overline{1,n}; \quad \sum_{i=1}^{n} x_i - 1 = 0.$$

Заметим, что функция $g(z), z \geq 0$ недифференцируема в точке 0, но непрерывна. Кроме того субдифференциал функции $g(z)$ в этой точке является пустым множеством. В силу сказанного и выпуклости функции $g(z)$ при $z \geq 0$ точка минимума функции не может быть на границе положительного ортанта (это также подтверждается тем фактом, что в окрестности нуля $g'(z) < 0$, т.е. функция убывает). Поэтому, исходя из условия дополняющей нежесткости $\nu_i x_i = 0$, можно положить $\nu_i = 0$ при всех значениях i[1]. Следовательно:

$$x_i = \xi_i \cdot \exp(-[C_1^T u + C_2^T v]_i) \cdot \exp(-\mu - 1).$$

Учитывая, что точка $x \in S_n(1)$ находим:

$$\mu = \ln\left[\sum_{i=1}^{n} \xi_i \cdot \exp(-[C_1^T u + C_2^T v]_i)\right] - 1.$$

$$x_i = \frac{\xi_i \cdot \exp(-[C_1^T u + C_2^T v]_i)}{\sum\limits_{j=1}^{n} \xi_j \cdot \exp(-[C_1^T u + C_2^T v]_j)}, \quad i = \overline{1,n}.$$

[1] максимум функции $g_i(x)$ вполне может достигаться на границе ортанта

Подставив получившиеся выражения x_i и μ в функцию Лагранжа $\overline{L}(x,u,v,\mu,\nu)$ с учетом $\nu_i = 0$ легко получить требуемое выражение целевой функции двойственной задачи:

$$\psi(u,v) = -\langle u, b_1\rangle - \langle v, b_2\rangle - \ln\left[\sum_{i=1}^{n} \xi_i \cdot \exp(-[C_1^T u + C_2^T v]_i)\right].$$

Полагая $y = [u; v] \in R_+^{m_1} \times R^{m_2}$ и $C = [C_1; C_2]$, приходим к утверждению теоремы

$\square$

Отметим, что получившуюся двойственную задачу максимизации функции $\psi(y)$, легко переписать в виде задачи минимизации функции $\phi(y) = -\psi(y)$

$$\phi(y) = \langle y, b\rangle + \ln\left[\sum_{i=1}^{n} \xi_i \exp\left(-\left[C^T y\right]_i\right)\right] \to \min_{y\in Q}, \tag{6}$$

где $Q = R_+^{m_1} \times R^{m_2}$.

В дальнейших выкладках будет испопользоваться значение градиента функции и матрицы вторых производных для задачи минимизации, определить которые позволяет теорема 2.

Теорема 2. *Для двойственной функции $\phi(y)$ градиент и матрица вторых производных могут быть записаны в виде*

$$\nabla\phi(y) = b - Cx(y), \quad \phi''(y) = CD(x(y))C^T - Cx(y)(Cx(y))^T, \tag{7}$$

Данная теорема доказывается тривиально путем прямого вычисления.

Следующая лемма 2 определяет важное свойство двойственной функции, которое требуется во многих теоремах о сходимости методов:

Лемма 2. *Градиент двойственной функции $\nabla\phi(y)$ удовлетворяет условию Липшица:*

$$||\nabla\phi(y^1) - \nabla\phi(y^2)|| \le L||y^1 - y^2||.$$

Пусть $\langle C\rangle_i$ – i-ый столбец матрицы C, тогда константу Липшица можно взять как $L = \max\limits_{i=\overline{1,n}} ||\langle C\rangle_i||^2$.

Доказательство. Так как функция $\phi(y)$, $y \in Q$ дважды непрерывно дифференцируема, то [2]

$$||\nabla\phi(y^1) - \nabla\phi(y^2)|| \leq ||\phi''(y)|| \cdot ||y^1 - y^2||, \quad y^1, y^2, y \in Q.$$

Используя формулу (7), в силу того, что матрица $Cx(y)(Cx(y))^T$ неотрицательно определена, получим

$$||\phi''(y)|| = ||CD(x(y))C^T - (Cx(y))(Cx(y))^T|| \leq ||CD(x(y))C^T||.$$

Обозначим через $D(x(y))^{0,5}$ корень квадратный из диагональной матрицы $D(x(y))$. На диагонали этой матрицы стоят элементы $d_{ii} = \sqrt{x_i}, i = \overline{1,n}$. Пусть $A = C \times D^{0,5}; D^{0,5} = D(x(y))^{0,5}$. Тогда $A_{j,i} = \sqrt{x_i} \cdot c_{j,i}$, $i = \overline{1,n}$, $j = \overline{1,m}$, и

$$||CDC^T|| \leq ||A||||A^T||.$$

Будем далее в качестве нормы матрицы понимать её евклидову норму

$$||A|| = \sqrt{\sum_{i=1}^{n}\sum_{j=1}^{n} A_{i,j}^2}.$$

Для такой нормы $||A|| = ||A^T||$, поэтому

$$||CDC^T|| \leq ||A||^2 = \sum_{i=1}^{n}\sum_{j=m}^{n} x_i c_{j,i} = \sum_{i=1}^{n} x_i \sum_{j=m}^{n} c_{j,i}^2 \leq$$
$$\leq \sum_{i=1}^{n} x_i \cdot \max_{l=\overline{1,n}} \sum_{j=1}^{m} c_{j,l}^2 = \max_{l=\overline{1,n}} \sum_{j=1}^{m} c_{j,l}^2 = \max_{i=\overline{1,n}} ||\langle C\rangle_i||^2.$$

□

Замечание 1. Полученное значение константы Липшица L приведено в статье [3], однако доказательство выглядит в ней более сложным и громоздким.

Пусть $Q_2 = \left\{y \in Q | c_1^T y = c_2^T y = \ldots = c_n^T y = \alpha, \alpha \in R\right\}; Q_1 = Q \setminus Q_2$.

Лемма 3. *Двойственная функция $\phi(y)$, $y \in Q$ выпукла на Q, строго выпукла на любом выпуклом множестве $G \in Q_1$, сильно выпукла на любом выпуклом компакте $G \in Q_1$.*

Доказательство. Двойственная функция $\phi(y)$ выпукла по теореме двойственности [1, 4], следовательно, квадратичная форма $F(a) = a^T\phi''(y)a \geq 0$ при $y \in Q$ и $a \in R^m$.
Введем новую переменную $z = C^Ta \in R^n$. Тогда

$$F(a) = a^TCDC^Ta - a^TCx(Cx)^Ta = z^TDz - (z^Tx)^2 \equiv F_1(z).$$

Рассмотрим задачу

$$F_1(z) \to \min_{z\in R^n}.$$

Очевидно, что $\nabla F_1(z) = 2Dz - 2(z^Tx)x$ и в точке минимума выполняется равенство $\nabla F_1(z)_i = 0$, т.е. $x_iz_i = (z^Tx)x_i$ или $z_i = z^Tx \equiv \alpha \quad \forall i = \overline{1,n}$, где α – некоторое число.
Пусть теперь $a = y$, тогда $z_i = c_i^Ty = \alpha$, $i = \overline{1,n}$ и по формуле (5) $x_i = \xi_i/\sum\limits_{j=1}^{n} \xi_j$, $i = \overline{1,n}$.
Тогда значение квадратичной формы $F(y)$ при $c_i^Ty = \alpha$ равно нулю. Таким образом квадратичная форма $F(y)$ обращается в нуль на множестве Q_2. Множество Q_2 может быть пусто и непусто, что зависит от структуры матрицы C. Пусть $Q_2 \neq \emptyset$, тогда на множестве $Q_1 = Q \setminus Q_2$ квадратичная форма $F(y)$ строго положительна. Значит на любом выпуклом подмножестве $G \in Q_1$ функция $\phi(y)$ строго выпукла.
Если же подмножество G – выпуклый компакт, то минимальное собственное число матрицы $\phi''(y)$ при $y \in G$ больше нуля и на множестве G функция сильно выпукла.
Пусть теперь $Q_2 = \emptyset$, тогда $Q_1 = Q$. Поэтому на любом выпуклом множестве $G \in Q$ функция $\phi(y)$ строго выпукла и сильно выпукла на любом выпуклом компакте $G \in Q$.

□

Замечание 2. Если множество G не пусто, то решение x^* существует и единственно. По теореме двойственности [4] решение двойственной задачи существует и может быть как единственно так и не единственно. Признаком неединственности решения является, очевидно, условие $Q_2 \neq \emptyset$ и значение компонент $x_i = \xi_i/\sum\limits_{j=1}^{n} \xi_j$ оказывается решением прямой задачи.

Быстрый Градиентный метод

В настоящее время все более широкое распространение получает метод, предложенный Ю.Е. Нестеровым и называемый быстрым градиентным методом (БГМ). Данному методу посвящен ряд публикаций, однако во всех этих публикациях он применяется к ограниченной на простом множестве задаче. Иными словами задача безусловной минимизации, решаемая методом, ограничивается, например, шаром радиуса R. В случае если решение задачи не находится в этом шаре, то метод перезапускается на шаре большего радиуса. Далее будет показано, что для широкого класса задач, включающего задачу ЭЛП, это делать не нужно, а также получены оценки сходимости БГМ – верхняя оценка количества итераций, необходимого для поиска $(\varepsilon_f, \varepsilon_g)$-решения.
Для дальнейшего изложения потребуется следующая лемма

Лемма 4. *[5] Пусть $f(x)$ – выпуклая функция, $f^* = \min f(x), x \in R^m$ причем $f^* = f(x^*)$, $x^{k+1} = x^k - h\nabla f(x^k)$ – итерационный процесс, стартующий из точки x^0, $R = ||x^0 - x^*||$ – расстояние от стартовой точки до решения задачи, существует такая константа M, что $||\nabla f(x)|| \le M \quad \forall x \in U_{\sqrt{2}R}(x^*)$, градиент функции $f(x)$ при $x \in U_{\sqrt{2}R}(x^*)$ удовлетворяет условию Липшица с константой L.*
Тогда при $h = 1/2L$ имеют место неравенства

$$f(\overline{x}^N) - f^* \le \frac{2LR^2}{N}; \quad ||x^N - x^*|| \le ||x^0 - x^*||. \tag{8}$$

Рассмотрим задачу выпуклой безусловной минимизации (9) гладкой функции $\phi(y)$ в пространстве R^m, градиент которой удовлетворяет условию Липшица с константой L.

$$\phi(y) \to \min_{y \in R^m} \tag{9}$$

Следуя [6], определим итерационный процесс (10) БГМ решения задачи (9), причем прокс-функция $d(y) = L||y - y^0||^2/2$.

$$
\begin{aligned}
\tilde{y}^{k+1} &= \arg\min_{y\in R^m}\left\{\tfrac{L}{2}||y-y^{k+1}||_2^2+\phi(y^k)+\left\langle\nabla\phi(y^{k+1}),y-y^{k+1}\right\rangle\right\};\\
\breve{y}^{k+1} &= \arg\min_{y\in R^m}\Big\{\tfrac{L}{2}||y-y^0||_2^2+\\
&\quad+\sum_{i=0}^{k}\alpha_{i+1}\left(\phi(y^{i+1})+\left\langle\nabla\phi(y^{i+1}),y-y^{i+1}\right\rangle\right)\Big\};\\
y^{k+1} &= \tau_k\breve{y}^k+(1-\tau_k)\tilde{y}^k.
\end{aligned}
\tag{10}
$$

Здесь вспомогательные неотрицательные числовые последовательности используемые в БГМ удовлетворяют условиям (11) ниже:

$$
\alpha_1\in(0,1];\quad A_k=\sum_{i=1}^{k}\alpha_i;\quad \alpha_k^2\le A_k;\quad \tau_k\le 1. \tag{11}
$$

Выражения (10) можно переписать в явном виде (28).

$$
\begin{aligned}
&\tilde{y}^{k+1}=y^{k+1}-\tfrac{1}{L}\nabla\phi(y^{k+1});\\
&\breve{y}^{k+1}=y^0-\tfrac{1}{L}\sum_{i=0}^{k}\alpha_{i+1}\nabla\phi(y^{i+1})=\breve{y}^k-\tfrac{1}{L}\alpha_{k+1}\nabla\phi(y^{k+1});\\
&y^{k+1}=\tau_k\breve{y}^k+(1-\tau_k)\tilde{y}^k;\\
&y^0=\breve{y}^0=\tilde{y}^0.
\end{aligned}
\tag{12}
$$

Для упрощения дальнейших выкладок введем функции $\Phi_k(u)$, $k\in\overline{0,\infty}$:

$$
\Phi_k(u)=\left\langle\nabla\phi(y^k),u-y^k\right\rangle+\phi(y^k).
$$

В случае выпуклой функции $\phi(y)$ при любом k выполняется неравенство $\Phi_k(u)\le\phi(u)$.

Теорема 3. *Пусть гладкая функция $\phi(y)$ выпукла в шаре $U_R(y^*)$, точка y^* – решение задачи минимизации $\phi(y)\to\min$ при $y\in U_R(y^*)$, градиент функции $\phi(y)$ удовлетворяет условию Липшица с контсантой L и последовательности α_k, τ_k удовлетворяют условиям*

$$
\tau_0=1;\quad \tau_k\le\frac{1}{\alpha_{k+1}};\quad \alpha_k^2\ge\alpha_{k+1}\frac{1-\tau_k}{\tau_k}. \tag{13}
$$

Тогда БГМ, стартующий из точки $y^0\in\partial U_R(y^*)$ и определяемый на последовательностях α_k, τ_k (13), генерирует последовательности точек y^k, $\tilde{y}^k$, $\breve{y}^k$, лежащие в шаре $U_R(y^*)$, В точке

$$\hat{y}^N = \frac{1}{A_N}\left[\sum_{k=1}^{N}\alpha_k^2\tilde{y}^k + \sum_{k=1}^{N-1}\alpha_{k+1}\breve{y}^k + \alpha_1 y^1 - \sum_{k=0}^{N-1}\alpha_{k+1}^2 y^{k+1}\right] \quad (14)$$

выполняются неравенства:

$$\begin{gathered}A_N\phi(\hat{y}^N) \le \min_u\left\{\sum_{k=0}^{N-1}\alpha_{k+1}\Phi_{k+1}(u) + \tfrac{L}{2}\left(||u-\breve{y}^0||^2 - ||u-\breve{y}^N||^2\right)\right\};\\ 0 \le \phi(\hat{y}^N) - \phi(y^*) \le \tfrac{L}{2A_N}\left(||y^*-\breve{y}^0||^2 - ||y^*-\breve{y}^N||^2\right).\end{gathered} \quad (15)$$

Доказательство. Покажем, что все последовательности точек, определяемые БГМ, будут лежать в шаре $U_R(x^*)$, где $R = ||x^0 - x^*||$. Действительно, следуя рассуждениям из [7] в доказательстве леммы 4.3, получаем

$$\begin{aligned}&\phi(y^{k+1}) - \phi(u) \le \left\langle\nabla\phi(y^{k+1}), y^{k+1}-u\right\rangle = \\ &\quad= \frac{1-\tau_k}{\tau_k}\left\langle\nabla\phi(y^{k+1}), \tilde{y}^k - y^{k+1}\right\rangle + \left\langle\nabla\phi(y^{k+1}), \breve{y}^k - u\right\rangle \le \\ &\qquad\le \frac{1-\tau_k}{\tau_k}(\phi(\tilde{y}^k) - \phi(y^{k+1})) + \left\langle\nabla\phi(y^{k+1}), \breve{y}^k - u\right\rangle.\end{aligned} \quad (16)$$

Следуя лемме 4.2 из [7], находим

$$\begin{aligned}&\left\langle\nabla\phi(y^{k+1}), \breve{y}^k - u\right\rangle = \left\langle\nabla\phi(y^{k+1}), \breve{y}^k - \breve{y}^{k+1}\right\rangle + \\ &\quad+ \left\langle\nabla\phi(y^{k+1}), \breve{y}^{k+1} - u\right\rangle = \left\langle\nabla\phi(y^{k+1}), \breve{y}^k - \breve{y}^{k+1}\right\rangle + \\ &\quad+ \frac{L}{2\alpha_{k+1}}\left(||u-\breve{y}^k||^2 - ||u-\breve{y}^{k+1}||^2 - ||\breve{y}^{k+1}-\breve{y}^k||^2\right) \le \\ &\le \frac{\alpha_{k+1}}{2L}||\nabla\phi(y^{k+1})||^2 + \frac{L}{2\alpha_{k+1}}\left(||u-\breve{y}^k||^2 - ||u-\breve{y}^{k+1}||^2\right).\end{aligned} \quad (17)$$

Следовательно, из (16) и (17) находим

$$
\begin{aligned}
\phi(y^{k+1}) - \phi(u) \leq \left\langle \nabla\phi(y^{k+1}), y^{k+1} - u \right\rangle \leq \\
\leq \frac{1-\tau_k}{\tau_k}(\phi(\tilde{y}^k) - \phi(y^{k+1})) + \frac{\alpha_{k+1}}{2L}||\nabla\phi(y^{k+1})||^2 + \\
+ \frac{L}{2\alpha_{k+1}}\left(||u - \breve{y}^k||^2 - ||u - \breve{y}^{k+1}||^2\right) \leq \\
\leq \frac{1-\tau_k}{\tau_k}(\phi(\tilde{y}^k) - \phi(y^{k+1})) + \alpha_{k+1}(\phi(y^{k+1}) - \phi(\tilde{y}^{k+1})) + \\
+ \frac{L}{2\alpha_{k+1}}\left(||u - \breve{y}^k||^2 - ||u - \breve{y}^{k+1}||^2\right).
\end{aligned}
$$

Полученное неравенство можно переписать в следующем виде

$$
\begin{gathered}
\alpha_{k+1}^2\phi(\tilde{y}^{k+1}) - \alpha_{k+1}\tfrac{1-\tau_k}{\tau_k}(\phi(\tilde{y}^k) \leq \alpha_{k+1}\left(\alpha_{k+1} - \tfrac{1-\tau_k}{\tau_k}\right)\phi(y^{k+1}) + \\
+\alpha_{k+1}\left\langle \nabla\phi(y^{k+1}), u - y^{k+1}\right\rangle + \tfrac{L}{2}\left(||u - \breve{y}^k||^2 - ||u - \breve{y}^{k+1}||^2\right).
\end{gathered}
$$

Просуммировав по индексу $k = \overline{0, N-1}$, находим

$$
\begin{gathered}
\alpha_N^2\phi(\tilde{y}^N) + \sum_{k=1}^{N-1}\left(\alpha_k^2 - \alpha_{k+1}\tfrac{1-\tau_k}{\tau_k}\right)\phi(\tilde{y}^k) - \alpha_1\tfrac{1-\tau_0}{\tau_0}\phi(\tilde{y}^0) \leq \\
\leq \sum_{k=0}^{N-1}\alpha_{k+1}\left\langle \nabla\phi(y^{k+1}), u - y^{k+1}\right\rangle + \\
+ \sum_{k=0}^{N-1}\phi(y^{k+1})\left(\alpha_{k+1}^2 - \alpha_{k+1}\tfrac{1-\tau_k}{\tau_k}\right) + \\
+\tfrac{L}{2}\left(||u - \breve{y}^0||^2 - ||u - \breve{y}^N||^2\right) = \\
= \sum_{k=0}^{N-1}\alpha_{k+1}\Phi_{k+1}(u) + \sum_{k=0}^{N-1}\phi(y^{k+1})\left(\alpha_{k+1}^2 - \tfrac{\alpha_{k+1}}{\tau_k}\right) + \\
+\tfrac{L}{2}\left(||u - \breve{y}^0||^2 - ||u - \breve{y}^N||^2\right) \leq \\
\leq \sum_{k=0}^{N-1}\alpha_{k+1}\phi(u) + \sum_{k=0}^{N-1}\phi(y^{k+1})\left(\alpha_{k+1}^2 - \tfrac{\alpha_{k+1}}{\tau_k}\right) + \\
+\tfrac{L}{2}\left(||u - \breve{y}^0||^2 - ||u - \breve{y}^N||^2\right).
\end{gathered} \tag{18}
$$

Выражение (14), используя $y^{k+1} = \tau_k\breve{y}^k + (1-\tau_k)\tilde{y}^k$, можно переписать в виде

$$\hat{y}^N = \frac{1}{A_N}\left[\alpha_N^2 \tilde{y}^N + \sum_{k=1}^{N-1}\left(\alpha_k^2 - \alpha_{k+1}\frac{1-\tau_k}{\tau_k}\right)\tilde{y}^k + \right. \\ \left. + \sum_{k=0}^{N-1}\left(\frac{\alpha_{k+1}}{\tau_k} - \alpha_{k+1}^2\right) y^{k+1}\right]. \tag{19}$$

Отметим, что

$$\alpha_N^2 + \sum_{k=1}^{N-1}\left(\alpha_k^2 - \alpha_{k+1}\frac{1-\tau_k}{\tau_k}\right) + \sum_{k=0}^{N-1}\left(\frac{\alpha_{k+1}}{\tau_k} - \alpha_{k+1}^2\right) = \sum_{k=0}^{N-1}\alpha_{k+1} = A_N.$$

Следовательно, в силу неравенства Йенсена

$$A_N\phi(\hat{y}^N) \leq \alpha_N^2\phi(\tilde{y}^N) + \sum_{k=1}^{N-1}\left(\alpha_k^2 - \alpha_{k+1}\frac{1-\tau_k}{\tau_k}\right)\phi(\tilde{y}^k) + \\ + \sum_{k=0}^{N-1}\left(\frac{\alpha_{k+1}}{\tau_k} - \alpha_{k+1}^2\right)\cdot\phi(y^{k+1}). \tag{20}$$

Таким образом, неравенство (18) преобразуется к следующему виду

$$\phi(\hat{y}^N) \leq \frac{1}{A_N}\left[\sum_{k=0}^{N-1}\alpha_{k+1}\Phi_{k+1}(u) + \frac{L}{2}\left(||u - \breve{y}^0||^2 - ||u - \breve{y}^N||^2\right)\right] \leq \\ \leq \phi(u) + \frac{L}{2A_N}\left(||u - \breve{y}^0||^2 - ||u - \breve{y}^N||^2\right).$$

Т.к. полученное неравенство выполнено при любом значении $u \in R^m$, а значит и при $u = y^*$, следует

$$A_N\phi(\hat{y}^N) \leq \min_u\left[\sum_{k=0}^{N-1}\alpha_{k+1}\Phi_{k+1}(u) + \frac{L}{2}\left(||u - \breve{y}^0||^2 - ||u - \breve{y}^N||^2\right)\right];$$

$$0 \leq \phi(\hat{y}^N) - \phi(y^*) \leq \frac{L}{2A_N}\left(||y^* - \breve{y}^0||^2 - ||y^* - \breve{y}^N||^2\right).$$

А значит утверждение теоремы (15) выполнено.
Из второго неравенства также следует

$$||y^* - \breve{y}^N|| \leq ||y^* - \breve{y}^0|| = ||y^* - y^0||.$$

В силу леммы 4 и выпуклости квадрата нормы находим

$$
\begin{gathered}
||\tilde{y}^{k+1}-y^*||^2 \le ||y^{k+1}-y^*||^2 = ||\tau_k(\breve{y}^k-y^*)+(1-\tau_k)(\tilde{y}^k-y^*)||^2 \le \\
\le \tau_k||\breve{y}^k-y^*||^2+(1-\tau_k)||\tilde{y}^k-y^*||^2 \le \\
\le \tau_k||y^0-y^*||^2+(1-\tau_k)||\tilde{y}^k-y^*||^2.
\end{gathered}
$$

Из полученно неравенства индукцией по k легко показать

$$||\tilde{y}^k-y^*|| \le ||y^k-y^*|| \le ||y^0-y^*||.$$

□

Следствие 1. *[5] Пусть ϕ^* – решение задачи (9), а последовательности для БГМ имеют вид*

$$\alpha_1=1; \quad \alpha_{k+1}=\alpha_k+\tfrac{1}{2}=\tfrac{k+2}{2}; \quad \tau_k=\tfrac{1}{\alpha_{k+1}}. \tag{21}$$

Тогда в условиях теоремы 3 для точки

$$\overline{y}^N=\frac{1}{N^2+3N}\left(\sum_{k=1}^{N-1}\tilde{y}^k+(N+1)^2\tilde{y}^N\right)$$

выполнено

$$
\begin{array}{rcl}
\phi(\overline{y}^N) & \le & \frac{4}{N^2+3N}\min\limits_u\left\{\sum\limits_{k=0}^{N-1}\frac{k+2}{2}\Phi_{k+1}(u)+\right. \\
 & & \left.+\frac{L}{2}\left(||u-\breve{y}^0||^2-||u-\breve{y}^N||^2\right)\right\}; \\
0 & \le & \phi(\overline{y}^N)-\phi^* \le \frac{4L}{N^2+3N}\left(||y^*-\breve{y}^0||^2-||y^*-\breve{y}^N||^2\right).
\end{array} \tag{22}
$$

Указанное следствие доказывается тривиальной подстановкой последовательностей (21) в формулы теоремы 3.

Приложение к задаче с аффинных ограничениями

Рассмотрим задачу (23) условной минимизации гладкой функции $f(x)$ на множестве $G \subset R^n$ с аффинными ограничениями $Ax=b$, где $b \in R^m$, A - матрица размерности $m \times n$.

$$f(x) \to \min_{x \in G; Ax=b}. \tag{23}$$

Для такой задачи можно выписать двойственную задачу (24)

$$\phi(y) = \max_{x\in G}\left\{\langle y, b - Ax\rangle - f(x)\right\} \to \min_{y\in R^m}. \qquad (24)$$

Под $x(y)$ будем понимать решение внутренней задачи максимизации при построении двойственной

$$x(y) = \arg\max_{x\in G}\left\{\langle y, b - Ax\rangle - f(x)\right\}.$$

При решении задач ЭЛП (1) или в более общем случае для сепарабельных функционалов, когда $f(x) = \sum f_i(x_i)$, в которых $x(y)$ можно получить явно, основной вклад в вычислительную сложность дает матричное произведение Ax и/или A^Ty при простых, например, параллелепипедных ограничениях G. К сожалению, значение $x(y)$ указать явно можно не всегда и оно известно лишь с определенной точностью. Однако в случае сильной выпуклости функции $f(x)$ или её сепарабельности такая неточность влияет на оценку сложности решения задачи лишь логарифмическим образом, и аккуратный учет этого приводит лишь к логарифмическим поправкам сложности исследуемого метода ([8, 9, 10, 11]).
Воспользуемся для решения двойственной задачи (24) БГМ в виде (10), для которого последовательности имеют вид (13). Тогда на N-м шаге согласно теореме 3 выполняется соотношение (15), из которого следует неравенство

$$\phi(\hat{y}^N) - \min_{u\in U_{2R}(y^*)}\left\{\sum_{k=0}^{N-1}\frac{\alpha_{k+1}}{A_N}\left[\phi(y^{k+1}) + \left\langle\nabla\phi(y^{k+1}), u - y^{k+1}\right\rangle\right]\right\} \leq$$
$$\leq \frac{8LR^2}{A_k} = \gamma_N. \qquad (25)$$

Положим

$$\lambda_k = \frac{\alpha_k}{A_N}; \quad x^N = \sum_{k=1}^{N}\lambda_k x(y^k) = \lambda_N x(y^N) + \frac{A_{N-1}}{A_N}x^{N-1}.$$

Тогда в силу $x(y)$ аналогично пункту 3 из [12] неравенство (25) преобразуется к виду

$$\phi(\hat{y}^N) - \sum_{k=1}^{N} \lambda_k \left\langle y^k, b - Ax(y^k) \right\rangle + \sum_{k=1}^{N} \lambda_k f(x(y^k)) -$$
$$- \min_{u \in U_{2R}(y^*)} \left\{ \sum_{k=0}^{N-1} \lambda_{k+1} \left\langle b - Ax(y^{k+1}), y - y^{k+1} \right\rangle \right\} \leq \gamma_N.$$

Легко видеть, что полученное неравенство легко преобразуется к виду

$$\phi(\hat{y}^N) + f\left(\sum_{k=1}^{N} \lambda_k x(y^k)\right) + \max_{u \in U_{2r}(y^*)} \left\{ \left\langle A \sum_{k=1}^{N} \lambda_k x(y^k) - b, y \right\rangle \right\} \leq \gamma_N.$$

Следовательно, выполняется неравенство

$$\phi(\hat{y}^N) + f\left(x^N\right) + 3R||Ax^N - b|| \leq \gamma_N.$$

В силу условия задачи $Ax^* = b$, а также в силу слабой двойственности [1] $-f(x^*) \leq \phi(y^*)$ выполнено:

$$f(x^N) - f(x^*) \leq f(x^N) + \phi(y^*) \leq f(x^N) + \phi(\hat{y}^N) \leq$$
$$\leq \phi(\hat{y}^N) + f\left(x^N\right) + 3R||Ax^N - b|| \leq \gamma_N.$$

Из этих же свойств находим

$$-f(x^*) = \langle y^*, b - Ax^* \rangle - f(x^*) = \phi(y^*) \geq \left\langle y^*, b - Ax^N \right\rangle - f(x^N);$$

$$-R||Ax^N - b|| \leq f(x^N) - f(x^*) \leq f(x^N) + \phi(\hat{y}^N).$$

Из неравенств выше легко видеть, что будет выполняться неравенство

$$R||Ax^N - b|| \leq \gamma_N/2.$$

Таким образом будут дополнительно выполнены неравенства

$$\|f(x^*) - f(x^N)\| \leq \gamma_N; \quad \|\phi(\hat{y}^N) + f(x^N)\| \leq \gamma_N.$$

Сказанное доказывает следующую теорему

Теорема 4. *БГМ при решении задачи с аффинными ограничениями (23) с помощью двойственной задачи (24) с критериями остановки $f(x^N) + g(y^N) \le \varepsilon_f$ и $||Ax - b|| \le \varepsilon_g$ гарантированно остановится, причем число итераций будет меньше чем*

$$\max\left\{\sqrt{\frac{18LR^2}{\varepsilon_f}}, \sqrt{\frac{18LR}{\varepsilon_g}}\right\}.$$

Приложение к задаче с линейными ограничениями-неравенствами

Ранее мы рассмотрели применение БГМ к задачам БМ и к задачам к ней сводимым. Однако, полученные результаты можно обобщить и на более широкий случай, а именно когда задача решается на множестве $W = R_+^{m_1} \times R^{m_2}$, при этом полагаем, что $m = m_1 + m_2$. Рассмотрим задачу условной минимизации вида

$$\phi(y) \to \min_{y \in W}; \quad W = R_+^{m_1} \times R^{m_2}. \tag{26}$$

Будем полагать, что $\nabla\phi(y)$ удовлетворяет условию Липшица с константой L.
Для такой задачи БГМ можно записать в следующем виде

$$\begin{array}{rcl}
\tilde{y}^{k+1} & = & \arg\min\limits_{y \in W} \left\{ \frac{L}{2}||y - y^{k+1}||_2^2 + \phi(y^k) + \left\langle \nabla\phi(y^{k+1}), y - y^{k+1} \right\rangle \right\}; \\
\breve{y}^{k+1} & = & \arg\min\limits_{y \in W} \Big\{ \frac{L}{2}||y - y^0||_2^2 + \\
 & & + \sum\limits_{i=0}^{k} \alpha_{i+1} \left(\phi(y^{i+1}) + \left\langle \nabla\phi(y^{i+1}), y - y^{i+1} \right\rangle \right) \Big\}; \\
y^{k+1} & = & \tau_k \breve{y}^k + (1 - \tau_k)\tilde{y}^k.
\end{array} \tag{27}$$

В дальнейшем воспользуемся также функцией-срезкой $(\cdot)_+$, которую можно записать как $(x)_+ = \max(x, 0)$, $x \in R^1$
Следующая лемма позволяет упростить запись БГМ.

Лемма 5. *[13] Если $Q = R_+^{m_1} \times R^{m_2}$, то последовательности точек $\tilde{y}^k$, $\breve{y}^k$, y^k можно записать в виде (28).*

$$\tilde{y}_j^k = \begin{cases} (y_j^k - \frac{1}{L}\nabla\phi(y^k)_j)_+ \text{ при } j \le m_1; \\ \ \ y_j^k - \frac{1}{L}\nabla\phi(y^k)_j \text{ при } j > m_1; \end{cases}$$
$$\breve{y}_j^k = \begin{cases} (y_j^0 - \frac{1}{L}\sum\limits_{i=0}^{k}\alpha_i\nabla\phi(y^i)_j)_+ \text{ при } j \le m_1; \\ \ \ y_j^0 - \frac{1}{L}\sum\limits_{i=0}^{k}\alpha_i\nabla\phi(y^i)_j \text{ при } j > m_1. \end{cases} \tag{28}$$

Введем дополнительную переменную

$$\check{y}^k = y^0 - \frac{1}{L}\sum_{i=0}^{k}\alpha_i\nabla\phi(y^i) = \check{y}^{k-1} - \frac{1}{L}\alpha_k\phi(y^k).$$

Тогда, используя утверждение леммы 5, выражение (27) принимает вид

$$\begin{aligned} &\tilde{y}_j^k = \begin{cases} (y_j^k - \frac{1}{L}\nabla\phi(y^k)_j)_+ \text{ при } j \le m_1; \\ \ \ y_j^k - \frac{1}{L}\nabla\phi(y^k)_j \text{ при } j > m_1; \end{cases} \\ &\check{y}^k = \check{y}^{k-1} - \tfrac{1}{L}\alpha_k\phi(y^k) \\ &\breve{y}_j^k = \begin{cases} (\check{y}_j^k)_+ \text{ при } j \le m_1; \\ \ \ \check{y}_j^k \text{ при } j > m_1. \end{cases} \\ &y^{k+1} = \tau_k\breve{y}^k + (1-\tau_k)\tilde{y}^k. \end{aligned} \tag{29}$$

Используя такое представление БГМ, запишем следующую теорему

Теорема 5. *Пусть* $y^* = \arg\min\phi(y)$ *при* $y \in U_R(y^*) \cap W$*, причем функция* $\phi(y)$ *является выпуклой в этом шаре, а её градиент удовлетворяет условию Липшица с константой* L*.*
Тогда БГМ (29), стартующий из точки $y^0 \in cl\{U_R(y^*) \cap W\}$ *и определяемый на последовательностях* τ_k*,* α_k *согласно условиям (13) генерирует последовательности точек* y^k*,* $\breve{y}^k$*,* $\tilde{y}^k$ *лежащие в* $U_R(y^*) \cap W$*, а в точке* $\hat{y}^N$ *(14), выполняется (15).*

Доказательство. Отметим, что последовательности точек $\tilde{y}^k$, $\breve{y}^k$, y^k в силу утверждения теоремы 3 принадлежат множеству $U_R(y^*)$. В силу (29) $\breve{y}^k \in U_R(y^*) \cap W$ и $\tilde{y}^k \in U_R(y^*) \cap W$. Поэтому, т.к. $y^k \in U_R(y^*) \cap W$ как выпуклая комбинация точек $\breve{y}^k$ и $\tilde{y}^k$. Аналогично доказанному ранее можно показать, что для точки (14) выполнено (15). □

Рассмотрим задачу минимизации (30) функции $f(x)$ на выпуклом множестве Q и двойственную к ней задачу минимизации функции $\phi(y)$

$$\begin{gathered} f(x) \to \min_{x \in G}; \quad G = \{x \in Q : C_1 x - b_1 \le 0; C_2 x - b_2 = 0\}; \\ \phi(y) = \max_{x \in Q}(\langle y, b - Cx \rangle - f(x)) \to \min_{y \in R_+^{m_1} \times R^{m_2}}. \end{gathered} \tag{30}$$

Для решения такой задачи БГМ можно переписать в виде алгоритма 1, который строит последовательности точек как в двойственном так и в прямом пространстве. Причем последовательность точек в прямо пространстве сходится к решению задачи.

Algorithm 1 Прямо-двойственный быстрый градиентный метод для задачи с аффинными ограничениями и ограничениями-неравенствами

Input: Последовательности $\{\alpha_i\}_{i=0}^{\infty}$ и $\{\tau_i\}_{i=0}^{\infty}$, точка y^0
Output: Последовательности точек $\tilde{y}^k$, x^k
repeat
Вычисляем

$$\tilde{y}_j^k = \begin{cases} (y_j^k - \frac{1}{L}\nabla\phi(y^k)_j)_+ \text{ при } j \le m_1; \\ y_j^k - \frac{1}{L}\nabla\phi(y^k)_j \text{ при } j > m_1; \end{cases}$$

$$\breve{y}_j^k = \begin{cases} (y_j^0 - \frac{1}{L}\sum\limits_{i=0}^{k}\alpha_i\nabla\phi(y^i)_j)_+ \text{ при } j \le m_1; \\ y_j^0 - \frac{1}{L}\sum\limits_{i=0}^{k}\alpha_i\nabla\phi(y^i)_j \text{ при } j > m_1; \end{cases}$$

$$y^{k+1} = \tau_k \tilde{y}^k + (1-\tau_k)\breve{y}^k; \quad x^k = \sum_{i=0}^{k}\frac{\alpha_i}{A_k}x(y^i).$$

until $f(x^k) + \phi(y^k) > \varepsilon_f$ или $\Delta(x^k, G) > \varepsilon_g$;

Отметим, что выход из алгоритма осуществляется при выполнении условий $f(x^k) + \phi(y^k) = f(x^k) - \psi(y^k) \le \varepsilon_f$, а величину $f(x^k) - \psi(y^k)$ можно понимать как зазор двойственности: разница между значениями прямой и двойственной функции в точке (x^k, y^k).
В указанных условиях сформулированную ранее для задачи БМ теорему можно обобщить:

Теорема 6. *[13] Пусть функция $f(x)$ выпукла на выпуклом множестве Q, а функция $\phi(y)$ определяется как (30), её градиент*

$\nabla\phi(y)$ удовлетворяет условию Липшица с константой L, $x(y) = \arg\max_{x\in Q}(\langle y, b - Cx\rangle - f(x))$.
Тогда для последовательностей точек, генерируемых согласно алгоритму 1, справедливы неравенства

$$\Delta(x^k, G) \leq \frac{2LR}{A_k}; \quad |f(x^*) - f(x^k)| \leq \frac{2LR^2}{A_k}; \quad f(x^k) + \phi(\tilde{y}^k) \leq \frac{2LR^2}{A_k}. \tag{31}$$

Следствие 2. *Пусть $\alpha_i = (i+1)/2$ (для такой последовательности выполняются соотношения (11)) и $y^0 = 0$, тогда количество итераций N, которое достаточно выполнить для достижения заданной точности $(\varepsilon_f, \varepsilon_g)$, будет определяться*

$$N = \max\left\{\sqrt{\frac{8LR}{\varepsilon_g}}, \sqrt{\frac{8LR^2}{\varepsilon_f}}\right\}. \tag{32}$$

При реализации алгоритма 1 целесообразно пользоваться формулами (29) и $x^{k+1} = (x^k \cdot A_k + \alpha_{k+1} \cdot x(y^{k+1}))/A_{k+1}$, что упрощает вычислительные затраты в рамках одной итерации.

Заключение

В силу теоремы 6 и леммы 2 алгоритм 1 может применяться для поиска $(\varepsilon_f, \varepsilon_g)$-решения задачи ЭЛП. Доказанные теоремы отражают прямо-двойственный характер БГМ: данный метод позволяет найти решение прямой задачи по последовательности точек в двойственном пространстве, что видно из формулировок доказанных теорем. Результаты численных экспериментов были отражены в работах [13, 14, 15, 16, 17]. Представленные в работе результаты уточнили результаты, изложенные в статье [5], расширяя область их применимости на более широкий класс задач с линейными ограничениями-неравенствами для широкого класса вспомогательных последовательностей α_i применяемых в БГМ.

Список литературы

[1] В.Г. Жадан. Методы оптимизации. Ч.1. Введение в выпуклый анализ и теорию оптимизации // М.: МФТИ – 2014.

[2] Б.Т. Поляк. Введение в оптимизацию. Изд. 2-е, испр. и доп. // М.: ЛЕНАНД – 2014.

[3] Y. Nesterov. Smooth minimization of non-smooth function. // Math. Program. Ser. A. – 2005 – V.103. No.1. P.127–152.

[4] А.Г. Сухарев, А.В. Тимохов, В.В. Федоров. Курс методов оптимизации // М.: ФИЗМАТЛИТ – 2005.

[5] А.С. Аникин, А.В. Гасников, А.И. Тюрин, А.В. Чернов. Двойственные подходы к задачам минимизации сильно выпуклых функционалов простой структуры при аффинных ограничениях. // ЖВМ и МФ – 2017 – Т.57. № 8. С. 28-42.

[6] Ю.Е. Нестеров. Метод минимизации выпуклых функций со скоростью сходимости $o(1/k^2)$. // Докл. АН СССР. – 1983 – Т.269, № 3, С. 543-547

[7] Z. Allen-Zhu and L. Orecchia. Linear coupling of Gradient and Mirror Descent: A Novel, Simple Interpretation of Nesterov's Accelerated Method and Mirror Descent // ITCS 2017: Innovations in Theoretical Computer Science – 2017.

[8] A. Anikin, P. Dvurechensky, A. Gasnikov, A. Golov, A. Gornov, Yu. Maximov, M. Mendel, V. Spokoiny. Modern efficient numerical approaches to regularized regression problems in application to traffic demands matrix calculation from link loads. // Proceedings of International conference ITAS-2015. Russia, Sochi, September – 2015.

[9] A. Nemirovski. Lectures on modern convex optimization analysis, algorithms, and engineering applications. // Philadelphia: SIAM – 2013.

[10] А.В. Гасников, П.Е. Двуреченский, Ю.Е. Нестеров. Стохастические градиентные методы с неточным оракулом. // Труды МФТИ – 2016 – Т.8. № 1. С. 41-91.

[11] А.В. Гасников, Д.И. Камзолов, М.А. Мендель. Основные конструкции над алгоритмами выпуклой оптимизации и их приложения к получению новых оценок для сильно выпуклых задач. // Труды МФТИ – 2016 – Т.8. № 3. С. 25-42.

[12] Yu. Nesterov. Complexity bounds for primal-dual methods minimizing the model of objective function. // CORE Discussion Papers. 2015/03 – 2015.

[13] А.В. Чернов. Прямо-двойственный метод решения задачи энтропийно-линейного программирования. // Интеллектуальные системы. – 2016. – Т. 20, вып. 1. С. 39-56

[14] A. Chernov, P. Dvurechencky, A. Gasnikov. Fast primal-dual gradient method for strongly convex minimization problems with linear constraints. // In: Kochetov, Yu. et all (eds.) DOOR-2016. LNCS – 2016 – V.9869.

[15] А.В. Гасников, П.Е. Двуреченский, А.Л. Суворикова, А.В. Чернов. Об энтропийной регуляризации транспортной задачи линейного программирования. // VIII Московская Международная конференция по исследованию операций (ORM2016). – 2016 – Т.2. С. 241-242.

[16] А.В. Чернов, А.И. Тюрин. Двойственный быстрый градиентный метод решения задач энтропийно-линейного программирования. // VIII Московская Международная конференция по исследованию операций (ORM2016). – 2016 – Т.2. С. 254-257.

[17] A. Chernov, P. Dvurechensky. A primal-dual first-order method for minimization problems with linear constraints. // The 40th Interdisciplinary Conference and School. – 2016.