

УДК 519.688

А. В. Гасников^{1,2,3,8}, П. Е. Двуреченский^{1,2,4}, Д. И. Камзолов^{1,2},
Ю. Е. Нестеров^{3,5}, В. Г. Спокойный^{1,2,3,4}, П. И. Стецюк⁶, А. Л. Суворикова^{1,2,7},
А. В. Чернов⁸

¹Лаборатория структурных методов анализа данных в предсказательном моделировании при
МФТИ(ГУ) (ПреМоЛаб)

²Институт проблем передачи информации РАН

³Национальный исследовательский университет Высшая школа экономики

⁴Weierstrass Institute for Applied Analysis and Stochastics, Berlin

⁵Universite catholique de Louvain, CORE, Belgium

⁶Институт кибернетики им. В.М. Глушкова НАН Украины

⁷The International Research Training Group 1792, Berlin

⁸Московский физико-технический институт (государственный университет)

Поиск равновесий в многостадийных транспортных моделях

В работе предлагается оригинальный способ поиска равновесий в многостадийных моделях транспортных потоков. В основе подхода лежит сочетание метода Синхорна и универсального градиентного метода.

Ключевые слова: седловая задача, энтропия, метод Синхорна, универсальный метод, неточный оракул.

*A. V. Gasnikov^{1,2,3,8} P. E. Dvurechensky^{1,2,4} D. I. Kamzolov^{1,2} Y. E. Nesterov^{3,5}
V. G. Spokoiny^{1,2,3,4} P. I. Stetsyuk⁶ A. L. Suvorikova^{1,2,7} A. V. Chernov⁸*

¹Research laboratory in Predictive Modeling and Optimization at PhysTech (PreMoLab)

²Institute for Information transmission problems RAS

³Higher School of Economics National Research University

⁴Weierstrass Institute for Applied Analysis and Stochastics, Berlin

⁵Universite catholique de Louvain, CORE, Belgium

⁶V.M. Glushkov Institute of Cybernetics of NAS of Ukraine

⁷The International Research Training Group 1792, Berlin

⁸Moscow Institute of Physics and Technology (State University)

Searching for equilibria in multistage transport models

We propose an original way to find an equilibria in multistage traffic flow models. The basis of our approach is the combination of the balancing method and the universal gradient method.

Key words: saddle problem, entropy, balancing method, Synchorn method, universal method, inexact oracle.

1. Введение

Поиск (стохастических) равновесий в многостадийных моделях транспортных потоков приводит к решению следующей седловой задачи с правильной (выпукло-вогнутой) структурой [1, 2]:

$$\min_{\substack{\sum_{j=1}^n z_{ij}=L_i, \sum_{i=1}^n z_{ij}=W_j \\ z_{ij} \geq 0, i, j=1, \dots, n}} \max_{x \in Q} \left\{ \sum_{i,j=1}^n z_{ij} \ln z_{ij} + \sum_{i,j=1}^n c_{ij}(x) z_{ij} + g(x) \right\} =$$

$$= \max_{x \in Q} \max_{\lambda, \mu \in \mathbb{R}^n} \left\{ \langle \lambda, L \rangle + \langle \mu, W \rangle - \sum_{i,j=1}^n \exp(-c_{ij}(x) - 1 + \lambda_i + \mu_j) + g(x) \right\}, \quad (1)$$

где $g(x)$ и $c_{ij}(x)$ — вогнутые гладкие функции (если ищутся не стохастические равновесия, то $c_{ij}(x)$ могут быть негладкими), $Q \subset \mathbb{R}^n$ — выпуклое ограниченное замкнутое множество простой структуры. Легко понять, что система балансовых ограничений в (1) либо несовместна $\sum_{i=1}^n L_i \neq \sum_{j=1}^n W_j$, либо вырождена (имеет не полный ранг). В последнем случае это приводит к тому, что двойственные переменные (λ, μ) определены с точностью до произвольной постоянной $\tilde{\alpha}$:

$$(\lambda + \tilde{\alpha}e, \mu - \tilde{\alpha}e), \quad e = (1, \dots, 1) \in \mathbb{R}^n.$$

Без ограничения общности положим $\sum_{i=1}^n L_i = \sum_{j=1}^n W_j = 1$. Тогда задачу (1) можно переписать следующим образом:

$$\begin{aligned} & \min_{\substack{\sum_{j=1}^n z_{ij}=L_i, \sum_{i=1}^n z_{ij}=W_j \\ z_{ij} \geq 0, i, j=1, \dots, n}} \max_{x \in Q} \left\{ \sum_{i,j=1}^n z_{ij} \ln z_{ij} + \sum_{i,j=1}^n c_{ij}(x) z_{ij} + g(x) \right\} = \\ & = \max_{x \in Q} \max_{\tilde{\lambda}, \tilde{\mu} \in \mathbb{R}^n} \left\{ \langle \tilde{\lambda}, L \rangle + \langle \tilde{\mu}, W \rangle - \sum_{i,j=1}^n \exp(-c_{ij}(x) - 1 + \tilde{\lambda}_i + \tilde{\mu}_j) + g(x) \right\} = \\ & = \max_{x \in Q} \max_{\lambda, \mu \in \mathbb{R}^n} \left\{ \langle \lambda, L \rangle + \langle \mu, W \rangle + 1 - \sum_{i,j=1}^n \exp(-c_{ij}(x) + \lambda_i + \mu_j) + g(x) \right\} = \\ & = - \min_{x \in Q} (f(x) - g(x)), \end{aligned} \quad (2)$$

где выпуклая функция $f(x)$ определяется следующим образом:

$$\begin{aligned} f(x) & = - \min_{\substack{\sum_{j=1}^n z_{ij}=L_i, \sum_{i=1}^n z_{ij}=W_j \\ z_{ij} \geq 0, i, j=1, \dots, n;}} \left\{ \sum_{i,j=1}^n z_{ij} \ln z_{ij} + \sum_{i,j=1}^n c_{ij}(x) z_{ij} \right\} = \\ & = - \max_{\lambda, \mu \in \mathbb{R}^n} \left\{ \langle \lambda, L \rangle + \langle \mu, W \rangle + 1 - \sum_{i,j=1}^n \exp(-c_{ij}(x) + \lambda_i + \mu_j) \right\}. \end{aligned} \quad (3)$$

Кроме того, задачу (1) можно представить так:

$$\begin{aligned} & \min_{\substack{\sum_{j=1}^n z_{ij}=L_i, \sum_{i=1}^n z_{ij}=W_j \\ z_{ij} \geq 0, i, j=1, \dots, n; \sum_{i,j=1}^n z_{ij}=1}} \max_{x \in Q} \left\{ \sum_{i,j=1}^n z_{ij} \ln z_{ij} + \sum_{i,j=1}^n c_{ij}(x) z_{ij} + g(x) \right\} = \\ & = \max_{x \in Q} \max_{\lambda, \mu \in \mathbb{R}^n} \left\{ \langle \lambda, L \rangle + \langle \mu, W \rangle - \ln \left(\sum_{i,j=1}^n \exp(-c_{ij}(x) + \lambda_i + \mu_j) \right) + g(x) \right\} = \\ & = - \min_{x \in Q} (\hat{f}(x) - g(x)), \end{aligned} \quad (4)$$

где выпуклая функция $\hat{f}(x)$ определяется как

$$\begin{aligned} \hat{f}(x) = & - \min_{\substack{\sum_{j=1}^n z_{ij} = L_i, \sum_{i=1}^n z_{ij} = W_j \\ z_{ij} \geq 0, i, j = 1, \dots, n; \sum_{i,j=1}^{n,n} z_{ij} = 1}} \left\{ \sum_{i,j=1}^n z_{ij} \ln z_{ij} + \sum_{i,j=1}^n c_{ij}(x) z_{ij} \right\} = \\ = & - \max_{\lambda, \mu \in \mathbb{R}^n} \left\{ - \ln \left(\sum_{i,j=1}^n \exp(-c_{ij}(x) + \lambda_i + \mu_j) \right) + \langle \lambda, L \rangle + \langle \mu, W \rangle \right\}. \end{aligned} \quad (5)$$

Поскольку мы добавили в ограничения условие $\sum_{i,j=1}^{n,n} z_{ij} = 1$, являющееся следствием балансовых уравнений, то это привело к тому, что двойственные переменные (λ, μ) определены с точностью до двух произвольных постоянных C_λ, C_μ : $(\lambda + C_\lambda e, \mu + C_\mu e)$.

В данной работе показано, как можно решать задачи (2) и (4) с помощью методов первого порядка. Мы ограничимся изучением только полноградиентных методов для этих задач, т.е. не будем рассматривать, например, рандомизацию при вычислении градиента. Планируется отдельно исследовать вопрос о возможности ускорения вычислений за счет введения рандомизации для внешней задачи. Мы предполагаем, что это может уменьшить общую трудоемкость метода только в случае, когда вспомогательная задача вычисления $\nabla c_{ij}(x)$ тоже достаточно сложная. Следует отметить, что для большинства приложений умение вычислять $\nabla c_{ij}(x)$ для конкретной пары (i, j) позволяет одновременно рассчитать и все $\nabla c_{ij}(x)$, $j = 1, \dots, n$, без дополнительных затрат. Кроме того, планируется отдельно исследовать вопросы о том, какие подходы и в какой степени могут быть распараллелены. Оказывается, что целесообразность рандомизации тесно связана с именно возможностью использовать параллельные вычисления.

Структура статьи следующая. В п. 2 рассматриваются алгоритмы приближенного вычисления субградиента функций $f(x)$ и $\hat{f}(x)$. В п. 3 мы приводим алгоритм для решения задач (2) и (4). В основе оригинального подхода п. 3 лежит сочетание метода Синхорна для решения внутренней задачи оптимизации по двойственным множителям и универсального метода с неточным оракулом для внешней задачи.

2. Приближенное вычисление субградиента целевой функции

Заметим, что расчет градиента $\nabla f(x)$ (в ряде транспортных приложений [1]) вогнутые функции $c_{ij}(x)$ – негладкие, тогда вместо градиентов стоит понимать суперградиенты $c_{ij}(x)$ и субградиент $f(x)$ осуществляется по следующей формуле (Демьянова–Данскина–Рубинова, см., например, [1, 3]):

$$\nabla f(x) = - \sum_{i,j=1}^n z_{ij}^*(x) \nabla c_{ij}(x), \quad (6)$$

где $z_{ij}^*(x)$ – единственное решение задачи (3). Аналогично

$$\nabla \hat{f}(x) = - \sum_{i,j=1}^n z_{ij}^*(x) \nabla c_{ij}(x), \quad (7)$$

где $z_{ij}^*(x)$ – единственное решение задачи (5).

Лемма 1. Пусть $z^*(x) \in \Sigma_{n^2}(1) = \left\{ z \in \mathbb{R}_+^{n^2} : \sum_{i,j=1}^n z_{ij} = 1 \right\}$ – единственное решение задачи (3), тогда справедливо неравенство

$$\|z^*(x_1) - z^*(x_2)\|_1 \leq \|c(x_1) - c(x_2)\|_\infty. \quad (8)$$

Пусть $z^*(x) \in \Sigma_{n^2}(1) = \left\{ z \in \mathbb{R}_+^{n^2} : \sum_{i,j=1}^n z_{ij} = 1 \right\}$ – единственное решение задачи (5), тогда справедливо неравенство (8).

Доказательство. Обозначим $d(z) = \sum_{i,j=1}^n z_{ij} \ln z_{ij}$. Из условий оптимальности в задаче (3) получаем

$$\begin{aligned} \langle \nabla d(z(x_1)) + c(x_1), z(x_2) - z(x_1) \rangle &\geq 0, \\ \langle \nabla d(z(x_2)) + c(x_2), z(x_1) - z(x_2) \rangle &\geq 0. \end{aligned}$$

Складывая эти неравенства и используя сильную выпуклость энтропии $d(z)$ в 1-норме с константой 1 на симплексе $\Sigma_{n^2}(1)$, получаем

$$\begin{aligned} &\|c(x_1) - c(x_2)\|_\infty \|z(x_2) - z(x_1)\|_1 \geq \\ &\geq \langle c(x_1) - c(x_2), z(x_2) - z(x_1) \rangle \geq \\ &\geq \langle \nabla d(z(x_1)) - \nabla d(z(x_2)), z(x_1) - z(x_2) \rangle \geq \\ &\geq \|z(x_2) - z(x_1)\|_1^2. \end{aligned}$$

Отсюда получаем неравенство (8). Второе утверждение леммы доказывается аналогично. ■

Обозначим $c(x) = \{c_{ij}(x)\}_{i,j}^n \in \mathbb{R}_+^{n^2}$. Тогда $\nabla f(x) = \left(\frac{dc(x)}{dx^T} \right)^T z^*(x)$, где $z^*(x)$ – единственное решение задачи (3), это матрица $dc(x)/dx^T$, состоящая из строк $(\nabla c_{ij}(x))^T$, $i, j = 1, \dots, n$. Пусть в пространстве векторов x выбрана некоторая норма $\|\cdot\|_x$. Через $\|\cdot\|_{x,*}$ обозначим сопряженную к ней. С учетом леммы 1, выполнено

$$\begin{aligned} \|\nabla f(x_1) - \nabla f(x_2)\|_{x,*} &= \left\| \left(\frac{dc(x_1)}{dx^T} \right)^T z^*(x_1) - \left(\frac{dc(x_2)}{dx^T} \right)^T z^*(x_2) \right\|_{x,*} \leq \\ &\leq \left\| \left(\frac{dc(x_1)}{dx^T} \right)^T (z^*(x_1) - z^*(x_2)) \right\|_{x,*} + \left\| \left(\left(\frac{dc(x_1)}{dx^T} \right)^T - \left(\frac{dc(x_2)}{dx^T} \right)^T \right) z^*(x_2) \right\|_{x,*} \leq \\ &\leq \left\| \left(\frac{dc(x_1)}{dx^T} \right)^T \right\|_{1 \rightarrow x,*} \|z^*(x_1) - z^*(x_2)\|_1 + \left\| \left(\left(\frac{dc(x_1)}{dx^T} \right)^T - \left(\frac{dc(x_2)}{dx^T} \right)^T \right) z^*(x_2) \right\|_{x,*} \leq \\ &\leq \left\| \left(\frac{dc(x_1)}{dx^T} \right)^T \right\|_{1 \rightarrow x,*} \|c(x_1) - c(x_2)\|_\infty + \left\| \left(\left(\frac{dc(x_1)}{dx^T} \right)^T - \left(\frac{dc(x_2)}{dx^T} \right)^T \right) z^*(x_2) \right\|_{x,*}. \end{aligned}$$

Здесь $\|\cdot\|_{1 \rightarrow x,*}$ обозначает норму линейного оператора, согласованную с 1-нормой в пространстве прообразов и нормой $\|\cdot\|_{x,*}$ в пространстве образов. Для любого $x \in Q$ включение $z^*(x) \in \Sigma_{n^2}(1)$, т.е. каждая компонента вектора $z^*(x)$ принадлежит отрезку $[0, 1]$. Значит, если каждая компонента вектора $c(x)$ зависит от x Липшицевым образом и ее субградиент ограничен, а также множество Q ограничено, то субградиент функции $f(x)$ имеет ограниченную вариацию. Если субградиент функции $-g(x)$ также имеет ограниченную вариацию субградиента, то вся целевая функция задачи (2) имеет ограниченную вариацию субградиента. Если дополнительно известно, что градиент каждой компоненты вектора $c(x)$ и градиент функции $-g(x)$ удовлетворяют условию Липшица, то градиент функции $f(x)$ также будет удовлетворять условию Липшица. Однако полученная оценка константы, ограничивающей вариацию субградиента, и константы Липшица градиента получатся слишком завышенными. Тем не менее ниже предложен метод, которому не требуются оценки этих констант. Достаточно знать, что хотя бы одна из них является конечной. Аналогичные рассуждения приводят к тем же заключениям относительно субградиента функции $f(x)$.

Опишем теперь алгоритм приближенного вычисления субградиента и значения функции $f(x)$. Проводя рассуждения, аналогичные лемме 2 работы [4] с использованием теоремы Синхорна (1967), получаем, что существуют единственные с точностью до умножения на константу векторы $u, v \in \mathbb{R}_+^n$ такие, что для матрицы $Z^*(x) \in \mathbb{R}^{n \times n}$, являющейся решением задачи (3), справедливо равенство

$$Z^*(x) = \text{diag}(u) \exp(-C(x)) \text{diag}(v),$$

где $\text{diag}(u)$, $\text{diag}(v)$ – матрицы с диагональными элементами, равными соответствующим элементам векторов u, v , $\exp(-C(x))$ означает поэлементную экспоненту от матрицы $-C(x)$. При этом векторы u, v могут быть вычислены с помощью итерационного процесса (метод Синхорна [4, 5]):

$$\begin{aligned} 1) \quad u_0 &= \frac{L}{\exp(-C(x))e}, \quad v_0 = \frac{W}{(\exp(-C(x)))^T e}, \\ 2) \quad u_{k+1} &= \frac{L}{\exp(-C(x))v_k}, \quad v_{k+1} = \frac{W}{(\exp(-C(x)))^T u_{k+1}}. \end{aligned}$$

Здесь используется покомпонентное деление вектора на вектор.

Отметим, что каждая итерация этого процесса требует $O(n^2)$ арифметических операций (а.о.). Пусть с помощью этого процесса получены приближения u_N, v_N . Определим приближенное решение задачи (3) следующим образом:

$$\tilde{Z}(x) = \text{diag}(u_{N-1}) \exp(-C(x)) \text{diag}(v_{N-1}).$$

Следуя работе [6], будем использовать проективную метрику $\rho(\cdot, \cdot)$ для пары векторов $a, b \in \mathbb{R}_{++}^n$:

$$\rho(a, b) = \ln \left(\max_{i,k} \frac{a_i b_k}{a_k b_i} \right).$$

Для матрицы $A \in \mathbb{R}_{++}^{n \times n}$ определим числа

$$\vartheta(A) = \max_{i,j,k,l} \frac{a_{ik} a_{jl}}{a_{jk} a_{il}} \quad \text{и} \quad \gamma(A) = \left(\frac{\sqrt{\vartheta(A)} - 1}{\sqrt{\vartheta(A)} + 1} \right)^2.$$

Тогда, согласно оценке (3.12) работы [6], если верно

$$\frac{1}{1 - \gamma(\exp(-C(x)))} \times \left[\rho(\text{diag}(u_{N-1}) \exp(-C(x)) v_{N-1}, L) + \rho(\text{diag}(u_N) \exp(-C(x)) v_{N-1}, W) \right] \leq \delta, \quad (9)$$

то выполняется

$$\begin{aligned} \exp(-\delta) &\leq \frac{z_{ij}^*(x)}{\tilde{z}_{ij}(x)} \leq \exp(\delta), \\ \exp(-\delta) &\leq \frac{\tilde{z}_{ij}(x)}{z_{ij}^*(x)} \leq \exp(\delta) \quad \forall i, j = 1, \dots, n. \end{aligned} \quad (10)$$

Здесь z_{ij}^* – элементы матрицы $Z^*(x)$, $\tilde{z}_{ij}(x)$ – элементы матрицы $\tilde{Z}(x)$. Кроме того:

$$\begin{aligned} &\rho(\text{diag}(u_{N-1}) \exp(-C(x)) v_{N-1}, L) + \rho(\text{diag}(u_N) \exp(-C(x)) v_{N-1}, W) \leq \\ &\leq (\gamma(\exp(-C(x))))^N \left(\rho(\exp(-C(x))e, L) + \rho(\text{diag}(u_0) \exp(-C(x))e, W) \right), \end{aligned}$$

то есть метод имеет геометрическую скорость сходимости.

Из неравенства (10), так как $z_{ij}^*(x) \in [0, 1]$ для всех $i, j = 1, \dots, n$, следует, что

$$|z_{ij}^*(x) - \tilde{z}_{ij}(x)| \leq z_{ij}^*(x)(\exp(\delta) - 1) \leq 2\delta \quad \forall i, j = 1, \dots, n,$$

при $\delta \leq 1$.

Пусть получено приближенное решение задачи (3):

$$\tilde{Z}(x) = \text{diag}(u_{N-1}) \exp(-C(x)) \text{diag}(v_{N-1})$$

такое, что справедливо неравенство (9). Определим приближенное значение функции $f(x)$ и приближенный градиент этой функции следующим образом:

$$\tilde{f}(x) = - \sum_{i,j=1}^n \tilde{z}_{ij}(x) \ln \tilde{z}_{ij}(x) - \sum_{i,j=1}^n c_{ij}(x) \tilde{z}_{ij}(x),$$

$$\tilde{\nabla} f(x) = - \sum_{i,j=1}^n \tilde{z}_{ij}(x) \nabla c_{ij}(x).$$

Тогда

$$\begin{aligned} & |\tilde{f}(x) - f(x)| = \\ & = \left| - \sum_{i,j=1}^n \tilde{z}_{ij}(x) \ln \tilde{z}_{ij}(x) - \sum_{i,j=1}^n c_{ij}(x) \tilde{z}_{ij}(x) + \sum_{i,j=1}^n z_{ij}^*(x) \ln z_{ij}^*(x) + \sum_{i,j=1}^n c_{ij}(x) z_{ij}^*(x) \right| \leq \\ & \leq \left| \sum_{i,j=1}^n (z_{ij}^*(x) - \tilde{z}_{ij}(x)) \ln \tilde{z}_{ij}(x) \right| + \left| \sum_{i,j=1}^n z_{ij}^*(x) (\ln z_{ij}^*(x) - \ln \tilde{z}_{ij}(x)) \right| + \\ & \quad + \left| \sum_{i,j=1}^n c_{ij}(x) (z_{ij}^*(x) - \tilde{z}_{ij}(x)) \right| \leq \\ & \leq 2\delta \sum_{i,j=1}^n |\ln \tilde{z}_{ij}(x)| + n^2\delta + 2\delta \sum_{i,j=1}^n |c_{ij}(x)|. \end{aligned} \quad (11)$$

Эта оценка дает критерий остановки метода Синхорна для получения заданной точности вычисления значения функции $f(x)$.

Аналогично можно получить оценку

$$|\tilde{f}(x) - f(x)| \leq 2\delta \sum_{i,j=1}^n |\ln z_{ij}^*(x)| + n^2\delta + 2\delta \sum_{i,j=1}^n |c_{ij}(x)|.$$

Из этой оценки можно получить оценку на общую трудоемкость вычисления значения функции $f(x)$ с заданной точностью. Также имеем

$$\begin{aligned} \left\| \tilde{\nabla} f(x) - \nabla f(x) \right\|_{x',*} &= \left\| - \sum_{i,j=1}^n \nabla c_{ij}(x) \tilde{z}_{ij}(x) + \sum_{i,j=1}^n \nabla c_{ij}(x) z_{ij}^*(x) \right\|_{x',*} \leq \\ &\leq 2\delta \sum_{i,j=1}^n \left\| \nabla c_{ij}(x) \right\|_{x',*}, \end{aligned}$$

где $\|\cdot\|_{x',*}$ означает выбранную норму, которая может не совпадать с нормой $\|\cdot\|_{x,*}$, выбранной для измерения скорости изменения субградиента функции $f(x)$.

Теперь мы можем построить алгоритм, позволяющий приближенно вычислить значение $\tilde{f}(x)$ функции $f(x)$ и ее субградиент $\tilde{\nabla} f(x)$, удовлетворяющие условиям

$$|\tilde{f}(x) - f(x)| \leq \delta_1 \quad \text{и} \quad \left\| \tilde{\nabla} f(x) - \nabla f(x) \right\|_{x',*} \leq \delta_2.$$

Алгоритм 1. «Вычисление значения и производной целевой функции»**Data:** Матрица $C(x)$, векторы $L, W \in \Sigma_n(1)$, точности $\delta_1, \delta_2 > 0$ **Result:** $\tilde{f}(x), \tilde{\nabla} f(x)$

- 1) Вычислим матрицу $A \leftarrow \exp(-C(x))$.
- 2) Вычислим число $\gamma(A)$.
- 3) Вычислим $C_1 \leftarrow \sum_{i,j=1}^n |c_{ij}(x)|$, $C_2 \leftarrow \sum_{i,j=1}^n \|\nabla c_{ij}(x)\|_{x',*}$.
- 4) Вычислим $u_0 \leftarrow \frac{L}{Ae}$, $v_0 \leftarrow \frac{W}{A^T e}$.
- 5) Вычислим $\alpha \leftarrow \rho(Ae, L) + \rho(\text{diag}(u_0)Ae, W)$.
- 6) **foreach** $k \geq 0$ **do**
 - а) Вычислим вектор $a \leftarrow Av_k$.
 - б) Вычислим $u_{k+1} \leftarrow \frac{L}{a}$, $v_{k+1} \leftarrow \frac{W}{A^T u_{k+1}}$.
 - в) Вычислим матрицу $Z \leftarrow \text{diag}(u_k)A\text{diag}(v_k)$.
 - г) Вычислим число $\beta \leftarrow \frac{1}{1-\gamma(A)} (\rho(\text{diag}(u_k)a, L) + \rho(\text{diag}(u_{k+1})a, W))$.

if $\delta_1 \geq \beta \left(\sum_{i,j=1}^n |\ln \tilde{z}_{ij}(x)| + n^2 + 2C_1 \right)$ **и** $\delta_2 \geq 2\beta C_2$ **then**

$$\tilde{f}(x) \leftarrow - \sum_{i,j=1}^n z_{ij} \ln z_{ij} - \sum_{i,j=1}^n c_{ij}(x) z_{ij}, \quad \text{и} \quad \tilde{\nabla} f(x) \leftarrow - \sum_{i,j=1}^n z_{ij} \nabla c_{ij}(x)$$

end

else

$k \leftarrow k + 1$

 Переходим к шагу а)

end

Algorithm 1: Вычисление значения и производной целевой функции**Лемма 2.** Общая трудоемкость приведенного алгоритма составляет

$$O(n^4 + mn^2 + n^2N),$$

где

$$N = \max \left\{ \left[\frac{1}{1-\gamma(A)} \ln \frac{\alpha \left(\sum_{i,j=1}^n |\ln z_{ij}^*(x)| + n^2 + 2C_1 \right)}{\delta_1(1-\gamma(A))} \right], \left[\frac{1}{1-\gamma(A)} \ln \frac{2\alpha C_2}{\delta_2(1-\gamma(A))} \right] \right\}.$$

Доказательство. Из геометрической скорости алгоритма следует, что после N шагов 6 справедливо неравенство

$$\beta \leq \frac{\alpha \gamma(A)^N}{1-\gamma(A)}.$$

Это означает, что после N итераций неравенства, проверяемые на шаге 6г, выполняются. Шаги 1, 4, 5, 6а, 6б, 6в, 6г требуют $O(n^2)$ а.о. Шаг 2 требует $O(n^4)$ а.о., но может быть

параллелизован. Шаг 3 требует $O(mn^2)$ а.о., где m – размерность вектора x . Наконец, последний шаг вычисления результата работы алгоритма требует $O(mn^2)$ а.о. Используя эти оценки, получаем результат леммы. ■

Аналогичные результаты можно получить для приближенного вычисления функции $\hat{f}(x)$ и ее субградиента.

К сожалению, критерий останова алгоритма 1 имеет недостатки с практической точки зрения. Это вызвано двумя причинами. Во-первых, критерий останова требует вычисления числа $\gamma(\exp(-C(x)))$, что требует порядка n^4 арифметических операций и при высокой размерности задачи трудно выполнимо. Во-вторых, число $\gamma(\exp(-C(x)))$ может оказаться очень близким к единице для матриц большой размерности, что приведет к необходимости делать очень много итераций для достижения требуемой точности. При этом теоретическая оценка скорости сходимости метода Синхорна является довольно сильно завышенной, как показывает практика. То есть можно ожидать, что эмпирически приближенное значение целевой функции и ее субградиента можно найти с требуемой точностью гораздо быстрее, чем это обеспечивается теоретической оценкой. Планируется посвятить отдельную публикацию аспектам практической реализации предложенного алгоритма поиска значения целевой функции и ее субградиента.

3. Универсальный метод с неточным оракулом

Как было указано в предыдущем пункте, функция $\varphi(x) = f(x) - g(x)$ принадлежит классу функций с субградиентом, удовлетворяющим условию Гельдера с параметрами $M_\nu > 0, \nu \in [0, 1]$:

$$\|\nabla\varphi(x) - \nabla\varphi(y)\|_{x,*} \leq M_\nu \|x - y\|_x^\nu, x, y \in Q,$$

из которого следует, что

$$\varphi(y) \leq \varphi(x) + \langle \nabla\varphi(x), y - x \rangle + \frac{M_\nu}{1 + \nu} \|x - y\|_x^{1+\nu}, x, y \in Q.$$

Кроме того, существует алгоритм, который для заданных $\delta_1, \delta_2 > 0$ позволяет вычислить приближенно значение функции $\varphi(x)$ и ее субградиент $\nabla\varphi(x)$, удовлетворяющие неравенствам

$$|\tilde{\varphi}(x) - \varphi(x)| \leq \delta_1 \quad \text{и} \quad \|\tilde{\nabla}\varphi(x) - \nabla\varphi(x)\|_{x',*} \leq \delta_2.$$

Предположим, что множество Q ограничено и имеет диаметр D в норме $\|\cdot\|_{x'}$. Отметим, что норма, выбранная для измерения константы в условии Гельдера, и норма, выбранная для измерения диаметра множества Q , могут не совпадать, что дает дополнительную гибкость. Проводя рассуждения, аналогичные п. 4.1.3.е [7] и лемме 2 [8], получим следующую лемму.

Лемма 3. Для любого $\delta > 0$ и $x \in Q$ мы можем вычислить такие $\tilde{\varphi}_\delta(x), \tilde{\nabla}_\delta\varphi(x)$, что для любого

$$M \geq \left[\frac{1 - \nu}{1 + \nu} \cdot \frac{6}{\delta} \right]^{\frac{1-\nu}{1+\nu}} M_\nu^{\frac{2}{1+\nu}}$$

и любого $y \in Q$ справедливо неравенство

$$0 \leq \varphi(y) - \left[\tilde{\varphi}_\delta(x) + \langle \tilde{\nabla}_\delta\varphi(x), y - x \rangle \right] \leq \frac{M}{2} \|x - y\|_x^2 + \frac{\delta}{4}. \quad (12)$$

Доказательство. Для этого нужно вызвать алгоритм 1 с параметрами $\delta_1 = \frac{\delta}{24}, \delta_2 = \frac{\delta}{24D}$ и положить

$$\begin{aligned} \tilde{\varphi}_\delta(x) &= \tilde{f}(x) - g(x) - \delta_1 - \delta_2 D = \tilde{f}(x) - g(x) - \frac{\delta}{12}, \\ \tilde{\nabla}_\delta\varphi(x) &= \tilde{\nabla}f(x) - \nabla g(x), \end{aligned}$$

где $\tilde{f}(x), \tilde{\nabla}f(x)$ – выход алгоритма 1. ■

Из определения сразу следует, что для любого $x \in Q$ справедливы неравенства

$$0 \leq \varphi(x) - \tilde{\varphi}_\delta(x) \leq \frac{\delta}{4}.$$

В силу ограниченности множества Q , с использованием леммы 2 трудоемкость метода для получения $\tilde{\varphi}_\delta(x), \tilde{\nabla}_\delta \varphi(x)$ может быть равномерно по $x \in Q$ оценена как

$$O\left(n^4 + mn^2 + n^2\theta_1 \ln \frac{\theta_2}{\delta}\right),$$

где θ_1, θ_2 – некоторые константы. Следуя [8], введем сильно выпуклую с константой 1 на множестве Q прокс-функцию $d(x)$. Мы предполагаем, что эта функция достигает минимум на Q в точке x_0 и что $d(x_0) = 0$. Определим также «расстояние» Брегмана:

$$\xi(x, y) = d(y) - d(x) - \langle \nabla d(x), y - x \rangle.$$

Приведем модификацию универсального быстрого градиентного метода, позволяющую работать с неточным оракулом $\tilde{\varphi}_\delta(x), \tilde{\nabla}_\delta \varphi(x)$.

Алгоритм 2 «Универсальный быстрый градиентный метод для задачи (2)».

Data: Число $M_0 > 0$, точность $\varepsilon > 0$

Result: Точка y_k

Определим $\varphi_0(x) \leftarrow \xi(x_0, x), y_0 \leftarrow x_0, A_0 \leftarrow 0$

foreach $k \geq 0$ **do**

1) Найдем $z_k = \arg \min_{x \in Q} \varphi_k(x)$.

2) Найдем минимальное $i_k \geq 0$ такое, что для числа $a_{k+1, i_k} > 0$, найденного из уравнения $a_{k+1, i_k}^2 = \frac{1}{2^{i_k} M_k} (A_k + a_{k+1, i_k})$ и используемого в определениях

$$A_{k+1, i_k} = A_k + a_{k+1, i_k}, \tau_{k, i_k} = \frac{a_{k+1, i_k}}{A_{k+1, i_k}}, \delta_{k+1, i_k} = \varepsilon \tau_{k, i_k} \quad x_{k+1, i_k} = \tau_{k, i_k} z_k + (1 - \tau_{k, i_k}) y_k,$$

$$\hat{x}_{k+1, i_k} = \arg \min_{y \in Q} \{ \xi(z_k, y) + a_{k+1, i_k} \langle \tilde{\nabla}_{\delta_{k+1, i_k}} \varphi(x_{k+1, i_k}), y \rangle \},$$

$$y_{k+1, i_k} = \tau_{k, i_k} \hat{x}_{k+1, i_k} + (1 - \tau_{k, i_k}) y_k, \text{ выполнено неравенство}$$

$$\tilde{\varphi}_{\delta_{k+1, i_k}}(y_{k+1, i_k}) \leq \tilde{\varphi}_{\delta_{k+1, i_k}}(x_{k+1, i_k}) + \langle \tilde{\nabla}_{\delta_{k+1, i_k}} \varphi(x_{k+1, i_k}), y_{k+1, i_k} - x_{k+1, i_k} \rangle + 2^{i_k-1} M_k \|y_{k+1, i_k} - x_{k+1, i_k}\|_x^2 + \frac{\varepsilon \tau_{k, i_k}}{4}.$$

3) Положим $x_{k+1} = x_{k+1, i_k}, y_{k+1} = y_{k+1, i_k}, a_{k+1} = a_{k+1, i_k}, \tau_k = \tau_{k, i_k}$. Определим

$$A_{k+1} = A_k + a_{k+1}, M_{k+1} = 2^{i_k-1} M_k, \delta_{k+1} = \delta_{k+1, i_k} \text{ и}$$

$$\varphi_{k+1}(x) = \varphi_k(x) + a_{k+1} [\tilde{\varphi}_{\delta_{k+1}}(x_{k+1}) + \langle \tilde{\nabla}_{\delta_{k+1}} \varphi(x_{k+1}), x - x_{k+1} \rangle].$$

end

Algorithm 2: Универсальный быстрый градиентный метод для задачи (2)

Теорема 1. Алгоритм 2 корректен. Кроме того, для любого $k \geq 0$ выполнено

$$A_k \left(\varphi(y_k) - \frac{\varepsilon}{2} \right) \leq \varphi_k^* = \min_{x \in Q} \varphi_k(x),$$

где

$$\varphi_k(x) = \sum_{i=0}^k a_i [\tilde{\varphi}_{\delta_i}(x_i) + \langle \tilde{\nabla}_{\delta_i} \varphi(x_i), x - x_i \rangle] + \xi(x_0, x).$$

Причем

$$A_k \geq \left[\frac{1}{2^{2+4\nu} 6^{1-\nu} M_\nu^2} \varepsilon^{1-\nu} k^{1+3\nu} \right]^{\frac{1}{1+\nu}}.$$

Значит, для всех $k \geq 1$ верно

$$\varphi(y_k) - \varphi^* \leq \left[\frac{2^{2+4\nu} 6^{1-\nu} M_\nu^2}{\varepsilon^{1-\nu} k^{1+3\nu}} \right]^{\frac{1}{1+\nu}} \xi(x_0, x^*) + \frac{\varepsilon}{2}.$$

Если известно число $R^2 > 0$, такое, что $\xi(x_0, x^*) \leq R^2$, то при $A_k \geq \frac{2R^2}{\varepsilon}$ выполнено $\varphi(y_k) - \varphi^* \leq \varepsilon$.

Доказательство является понятной модификацией доказательства теоремы 3 работы [8]. ■

Опишем теперь, как можно восстанавливать решение двойственной задачи по информации, генерируемой алгоритмом 2. Для этого предположим, что $Q = \{x \in \mathbb{R}^m : x \geq \bar{x}\}$. В этой ситуации удобно выбрать $d(x) = \frac{1}{2} \|x - \bar{x}\|_2^2$. Тогда $x_0 = \bar{x}$ и $\xi(x, y) = \frac{1}{2} \|x - y\|_2^2$. На каждой итерации алгоритма 2 необходимо решить задачу вида

$$\min_{x \in Q} \{ \xi(\hat{x}, y) + \alpha \langle g, y \rangle \} = \min_x \max_{s \geq 0} \left\{ \frac{1}{2} \|x - \hat{x}\|_2^2 + \alpha \langle g, y \rangle + \langle s, \bar{x} - x \rangle \right\}$$

при заданных \hat{x}, g . Эта задача легко решается в явном виде. В частности, для поиска точки z_k необходимо решить задачу

$$z_k = \arg \min_{x \geq \bar{x}} \left\{ \frac{1}{A_k} \xi(x_0, x) + \sum_{i=0}^k \frac{a_i}{A_k} [\tilde{\varphi}_{\delta_i}(x_i) + \langle \tilde{\nabla}_{\delta_i} \varphi(x_i), x - x_i \rangle] \right\}.$$

Эта задача эквивалентна следующей:

$$\min_x \max_{s \geq 0} \left\{ \frac{1}{A_k} \xi(x_0, x) + \sum_{i=0}^k \frac{a_i}{A_k} [\tilde{\varphi}_{\delta_i}(x_i) + \langle \tilde{\nabla}_{\delta_i} \varphi(x_i), x - x_i \rangle] + \langle s, \bar{x} - x \rangle \right\}.$$

Пусть s_k – оптимальное значение переменной s в последней задаче. Тогда в силу условий оптимальности в этой задаче имеем

$$\frac{1}{A_k} \nabla \xi(x_0, z_k) + \sum_{i=0}^k \frac{a_i}{A_k} \tilde{\nabla}_{\delta_i} \varphi(x_i) - s_k = 0,$$

$$\langle s_k, \bar{x} - z_k \rangle = 0.$$

Используя эти равенства и определение неточного оракула (12), получаем

$$\begin{aligned} & \varphi(x) + \langle s_k, \bar{x} - x \rangle \geq \\ & \geq \sum_{i=0}^k \frac{a_i}{A_k} [\tilde{\varphi}_{\delta_i}(x_i) + \langle \tilde{\nabla}_{\delta_i} \varphi(x_i), x - z_k + z_k - x_i \rangle] + \langle s_k, \bar{x} - z_k + z_k - x \rangle = \\ & = \sum_{i=0}^k \frac{a_i}{A_k} [\tilde{\varphi}_{\delta_i}(x_i) + \langle \tilde{\nabla}_{\delta_i} \varphi(x_i), z_k - x_i \rangle] + \frac{1}{A_k} \langle \nabla \xi(x_0, z_k), z_k - x \rangle \geq \\ & \geq \sum_{i=0}^k \frac{a_i}{A_k} [\tilde{\varphi}_{\delta_i}(x_i) + \langle \tilde{\nabla}_{\delta_i} \varphi(x_i), z_k - x_i \rangle] + \frac{1}{A_k} (\xi(x_0, z_k) - \xi(x_0, x)) = \\ & = \frac{1}{A_k} \varphi_k^* - \frac{1}{A_k} \xi(x_0, x). \end{aligned}$$

Предположим, что $\xi(x_0, x^*) \leq R^2$, тогда исходная задача эквивалентна задаче

$$\max_{s \geq 0} \min_x \{ \varphi(x) + \langle s, \bar{x} - x \rangle : \xi(x_0, x) \leq R^2 \} = \max_{s \geq 0} \psi_R(s).$$

Тогда из полученной оценки на $\varphi(x) + \langle s_k, \bar{x} - x \rangle$ и теоремы 1 получим

$$\psi_R(s_k) = \min_x \{ \varphi(x) + \langle s, \bar{x} - x \rangle : \xi(x_0, x) \leq R^2 \} \geq \frac{1}{A_k} \varphi_k^* - \frac{R}{A_k} \geq \varphi(y_k) - \frac{\varepsilon}{2} - \frac{R^2}{A_k}.$$

Отсюда, используя слабую двойственность, имеем

$$0 \leq \varphi(y_k) - \psi_R(s_k) \leq \frac{\varepsilon}{2} + \frac{R^2}{A_k}.$$

Отметим, что предложенный подход легко распространяется на задачи композитной оптимизации, возникающие при поиске стохастических равновесий в транспортных моделях, см. [9].

4. Заключительные замечания

В приложениях часто возникают задачи, имеющие следующий вид (см., например, [10, 11]):

$$f(x) = \Phi(x, y(x)) \rightarrow \min_x, \quad (13)$$

при этом $y(x)$ и $\nabla y(x)$ могут быть получены из решения отдельной подзадачи лишь приближенно. Довольно типично, что существует способ, который выдает ε -приближенное решение за время, зависящее от ε , логарифмическим образом $\sim \ln(\varepsilon^{-1})$. В данной работе намечен общий способ решения таких задач. Его наиболее важной отличительной чертой является адаптивность (самонастраиваемость), т.е. методу на вход не надо подавать никаких констант Липшица (более того, метод будет работать и в негладком случае). Метод сам настраивается локально на оптимальную гладкость функции. Это свойство метода делает его привилегированным, поскольку в реальных приложениях, чтобы что-то знать о свойствах $f(x)$, нужно что-то знать о свойствах зависимости $y(x)$, а это часто не доступно по постановке задачи, или при попытке оценить приводит к сильно завышенным оценкам.

В задаче (13) важно уметь эффективно пересчитывать значения $y(x)$, а не рассчитывать их каждый раз заново (на каждой итерации внешнего цикла). Поясним сказанное. Предположим, что мы уже как-то посчитали, скажем, $y(x)$, решив, например, с какой-то точностью соответствующую задачу оптимизации. Тогда для вычисления $y(x + \Delta x)$ (на следующей итерации внешнего цикла) у нас будет хорошее начальное приближение $y(x)$. А, как известно (см., например, [11]), расстояние от точки старта до решения (не в случае сходимости метода со скоростью геометрической прогрессии или быстрее) существенным образом определяет время работы алгоритма. Тем не менее известные нам приложения (см., [10, 11]) пока как раз всецело соответствуют сходимости процедуры поиска $y(x)$ со скоростью геометрической прогрессии. Связано это с тем, что если расчет $y(x)$ с точностью ε осуществляется за $O(\ln(\varepsilon^{-1}))$ операций, то для внешней задачи можно выбирать самый быстрый метод (а стало быть, и самый требовательный к точности). Как правило, такое сочетание оказывается недоминируемым. В частности, для рассматриваемой нами задачи планируется в отдельной статье сравнить описанный здесь подход с возможными альтернативными. Ведь, как известно [12], выпукло-вогнутые седловые задачи, к которым, безусловно, относятся задачи (2), (4), можно по-разному решать, и способ, описанный в п. 3, требует дополнительного сопоставительного анализа, которому и планируется посвятить отдельную статью. Здесь же ограничимся ссылкой на пример 4 и последующий текст из работы [11] и общим тезисом, который пока неплохо подтверждался на практике:

Если есть возможность в задаче оптимизации (в седловой задаче) явно прооптимизировать по части переменных (или как-то эффективно это сделать с хорошей точностью), то, как правило, это и надо сделать и строить итерационный метод исходя из этого.

В реальных транспортных приложениях [1, 2, 9, 13] достаточно сложным является расчет $c_{ij}(x)$ и их градиентов (особенно при поиске стохастических равновесий). Тем не менее

эти задачи имеют вполне четкую привязку к решению некоторых задач на графах типа поиска кратчайших путей. Так же, как и в предыдущем абзаце для внутренней задачи, для внешней задачи можно не рассчитывать $c_{ij}(x)$ и их градиенты каждый раз заново, а пересчитывать; также можно допускать неточность в их вычислении (и ненулевую вероятность ошибки), надеясь на ускорение (благо метод работает в концепции неточного оракула, природа которой не принципиальна, см. [7]). Также здесь оказываются полезными идеи БАД (быстрого автоматического дифференцирования) [14, 15], которые позволяют практически за то же время, что занимает вычисление самих функций, вычислять их градиенты.

Исследование А.В. Гасникова, П.Е. Двуреченского, В.Г. Спокойного и А.Л. Сувориковой в части 2 выполнено в ИППИ РАН за счет гранта Российского научного фонда (проект № 14-50-00150), исследование А.В. Гасникова и П.Е. Двуреченского в части 3 выполнено при поддержке гранта РФФИ 15-31-20571-мол_а_вед, исследование Ю.Е. Нестерова в части 3 выполнено при поддержке гранта РФФИ 13-01-12007-офи_м.

Литература

1. Гасников А.В., Дорн Ю.В., Нестеров Ю.Е., Шпирко С.В. О трехстадийной версии модели стационарной динамики транспортных потоков // Математическое моделирование. 2014. Т. 26:6. С. 34–70. [arXiv:1405.7630](https://arxiv.org/abs/1405.7630)
2. Гасников А.В. Об эффективной вычислимости конкурентных равновесий в транспортно-экономических моделях // Математическое моделирование. 2015. Т. 27, № 12. С. 121–136. [arXiv:1410.3123](https://arxiv.org/abs/1410.3123)
3. Гасников А.В., Гасникова Е.В., Нестеров Ю.Е., Чернов А.В. Об эффективных численных методах решения задач энтропийно-линейного программирования // ЖВМ и МФ. 2016. Т. 56, № 4. (принята к печати) [arXiv:1410.7719](https://arxiv.org/abs/1410.7719)
4. Cuturi M. Sinkhorn Distances: Lightspeed Computation of Optimal Transport // NIPS, 2013.
5. Benamou J.D., Carlier G., Cuturi M., Nenna L., Peyré G. Iterative Bregman Projections for Regularized Transportation Problems. e-print, 2015. [arXiv:1412.5154](https://arxiv.org/abs/1412.5154) (to appear in SISC).
6. Franklin J., Lorenz J. On the scaling of multidimensional matrices // Linear Algebra and its applications. 1989. V. 114. P. 717–735.
7. Devolder O. Exactness, inexactness and stochasticity in first-order methods for large-scale convex optimization. CORE UCL, PhD thesis, March 2013. http://www.ecore.be/DPs/dp_1327057920.pdf
8. Nesterov Yu. Universal gradient methods for convex optimization problems // Mathematical Programming Series A. 2015. V. 152, I. 1, P. 381–404.
9. Гасников А.В., Гасникова Е.В., Двуреченский П.Е., Ершов Е.И., Лагуновская А.А. Поиск стохастических равновесий в транспортных моделях равновесного распределения потоков // Труды МФТИ, 2015. Т. 7, № 4. С. 114–128. [arXiv:1505.07492](https://arxiv.org/abs/1505.07492)
10. Bogolubsky L., Dvurechensky P., Gasnikov A., Gusev G., Nesterov Yu., Raigorodskii A., Tikhonov A., Zhukovskii M. Learning supervised PageRank with gradient-free optimization methods. e-print, 2014. [arxiv:1411.4282](https://arxiv.org/abs/1411.4282)
11. Гасников А.В., Двуреченский П.Е., Нестеров Ю.Е. Стохастические градиентные методы с неточным оракулом. e-print, 2015. [arXiv:1411.4218](https://arxiv.org/abs/1411.4218)
12. Nemirovski A. Lectures on modern convex optimization analysis, algorithms, and engineering applications. Philadelphia: SIAM, 2013. http://www2.isye.gatech.edu/~nemirovs/Lect_ModConvOpt.pdf

13. *Гасников А.В., Двуреченский П.Е., Дорн Ю.В., Максимов Ю.В.* Численные методы поиска равновесного распределения потоков в моделях Бэкмана и стабильной динамики // Математическое моделирование. 2016. Т. 28. (принята к печати) [arXiv:1506.00293](https://arxiv.org/abs/1506.00293)
14. *Ким К., Нестеров Ю., Скоков В., Черкасский Б.* Эффективные алгоритмы для дифференцирования и задачи экстремали // Экономика и математические методы. 1984. Т. 20. С. 309–318.
15. *Евтушенко Ю.Г.* Оптимизация и быстрое автоматическое дифференцирование. М.: ВЦ РАН, 2013.

References

1. *Gasnikov A.V., Dorn Y.V., Nesterov Y.E., Shpirko S.V.* On the three-stage version of stable dynamic model. Math modeling. 2014. V. 26:6. P. 34–70. [arXiv:1405.7630](https://arxiv.org/abs/1405.7630) (in Russian).
2. *Gasnikov A. V.* On the effective computability of competitive equilibrium in the transport-economic models. Math modeling. 2015. V. 27, N 12. P. 121–136. [arXiv:1410.3123](https://arxiv.org/abs/1410.3123) (in Russian).
3. *Gasnikov A.V., Gasnikova E.V., Nesterov Y.E., Chernov A.V.* On effective numerical methods for solving linear-entropy programming // JCM and MP. 2016. V. 56, N 4. (in print) [arXiv:1410.7719](https://arxiv.org/abs/1410.7719) (in Russian).
4. *Cuturi M.* Sinkhorn Distances: Lightspeed Computation of Optimal Transport. NIPS, 2013.
5. *Benamou J.D., Carlier G., Cuturi M., Nenna L., Peyré G.* Iterative Bregman Projections for Regularized Transportation Problems. e-print, 2015. [arXiv:1412.5154](https://arxiv.org/abs/1412.5154) (to appear in SISC).
6. *Franklin J., Lorenz J.* On the scaling of multidimensional matrices. Linear Algebra and its applications. 1989. V. 114. P. 717–735.
7. *Devolder O.* Exactness, inexactness and stochasticity in first-order methods for large-scale convex optimization. CORE UCL, PhD thesis, March 2013. http://www.ecore.be/DPs/dp_1327057920.pdf
8. *Nesterov Yu.* Universal gradient methods for convex optimization problems. Mathematical Programming Series A. 2015. V. 152, I. 1, P. 381–404.
9. *Gasnikov A.V., Gasnikova E.V., Dvurechensky P.E., Ershov E.I., Lagunovskaya A.A.* Search for the stochastic equilibria in the transport models of equilibrium flow distribution. Trudy MIPT. 2015. V. 7, N 4. P. 114–128. [arXiv:1505.07492](https://arxiv.org/abs/1505.07492) (in Russian).
10. *Bogolubsky L., Dvurechensky P., Gasnikov A., Gusev G., Nesterov Yu., Raigorodskii A., Tikhonov A., Zhukovskii M.* Learning supervised PageRank with gradient-free optimization methods. e-print, 2014. [arxiv:1411.4282](https://arxiv.org/abs/1411.4282)
11. *Gasnikov A.V., Dvurechensky P.E., Nesterov Y.E.* Stochastic gradient method with inexact oracle. e-print, 2015. [arXiv:1411.4218](https://arxiv.org/abs/1411.4218) (in Russian).
12. *Nemirovski A.* Lectures on modern convex optimization analysis, algorithms, and engineering applications. Philadelphia: SIAM, 2013. http://www2.isye.gatech.edu/~nemirovs/Lect_ModConvOpt.pdf
13. *Gasnikov A.V., Dvurechensky P.E., Dorn Y.V., Maximov Y.V.* Numerical methods for finding an equilibrium flow distribution in Beckmann model and stable dynamics. Math modelling. 2016. V. 28. (in print) [arXiv:1506.00293](https://arxiv.org/abs/1506.00293) (in Russian).
14. *Kim K., Nesterov Yu., Skokov V., Cherkassky B.* Effective algorithms for the derivation and extremal problem. Economics and Mathematical Methods. 1984. V. 20. P. 309–318.
15. *Evtushenko Y.G.* Optimization and fast automatic differentiation. M.: CC RAS, 2013. (in Russian).

Поступила в редакцию 10.11.2015.