Гасников Александр Владимирович

Эффективные численные методы поиска равновесий в больших транспортных сетях

Диссертации на соискание степени д.ф.-м.н. по специальности 05.13.18

Научный консультант: чл.-корр. РАН, проф. А.А. Шананин

Московский физико-технический институт; дисс. совет Д 212.156.05, 1 декабря 2016

План выступления

- Поиск равновесий в транспортных сетях
- Универсальный метод (подобных) треугольников (УМТ)
- Основные операции над алгоритмами:
mini-batch, регуляризация, рестарты, выпуклая комбинация
- УМТ с неточным оракулом. Суперпозиция алгоритмов. Примеры
- Прямо-двойственность УМТ
- Условный вариант метода зеркального спуска
- Google problem. Число итераций *vs* Стоимость итерации
- Безградиентные и покомпонентные методы
- Результаты, выносимые на защиту

Поиск равновесий в транспортных сетях

Необходимо решать задачи выпуклой оптимизации вида

$$\boxed{\min_{f,x}\left\{\Psi\left(x,f\right):\ f=\Theta x,\ x\in X\right\}=-\min_{t\in\operatorname{dom}\sigma^*}\left\{\gamma\psi\left(t/\gamma\right)+\sum_{e\in E}\sigma_e^*\left(t_e\right)\right\},}\quad (1)$$

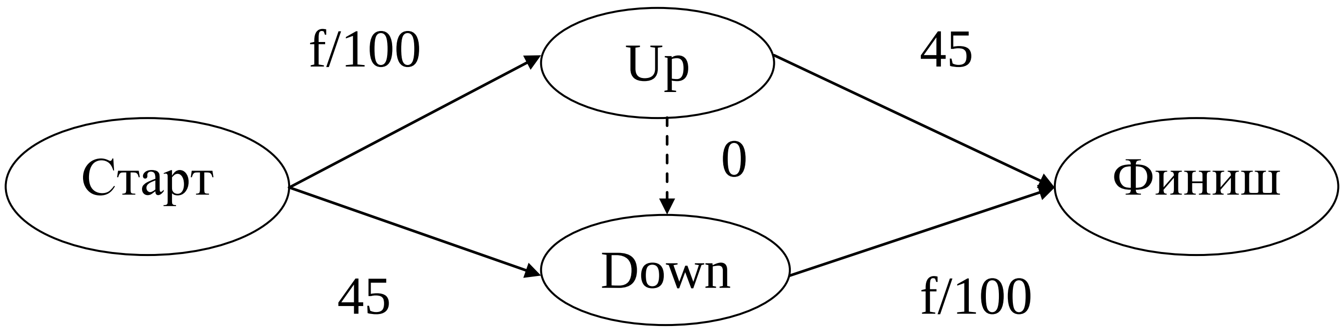
$$\Psi\left(x,f\right):=\Psi^1\left(x\right)=\sum_{e^1\in\tilde{E}^1}\sigma_{e^1}\left(f_{e^1}^1\right)+\Psi^2\left(x\right)+\gamma^1\sum_{w^1\in OD^1}\sum_{p^1\in P_{w^1}^1}x_{p^1}^1\ln\left(x_{p^1}^1\big/d_{w^1}^1\right),\ d_{w^2}^2=f_{w^2}^1,\ldots$$

$$\psi_{w^k}^k\left(t\right)=\ln\left(\sum_{p^k\in P_{w^k}^k}\exp\left(-g_{p^k}^k\left(t\right)\right)\right),\ k=1,\ldots,m,\ \psi^1\left(t\right)=\sum_{w^1\in OD^1}d_{w^1}^1\psi_{w^1}^1\left(t\right),$$

$$g_{p^m}^m\left(t\right)=\sum_{e^m\in\tilde{E}^m}\delta_{e^m p^m}t_{e^m}=\sum_{e^m\in E^m}\delta_{e^m p^m}t_{e^m},$$

$$g_{p^k}^k\left(t\right)=\sum_{e^k\in\tilde{E}^k}\delta_{e^k p^k}t_{e^k}-\sum_{e^k\in\bar{E}^k}\delta_{e^k p^k}\gamma^{k+1}\psi_{e^k}^{k+1}\left(t/\gamma^{k+1}\right).$$

Парадокс Браеса (модель Бэкмана, $m=1, \gamma=0$)



Корреспонденция: $d = 4000\ авт/час,\ x_{up} + x_{down} = d.$

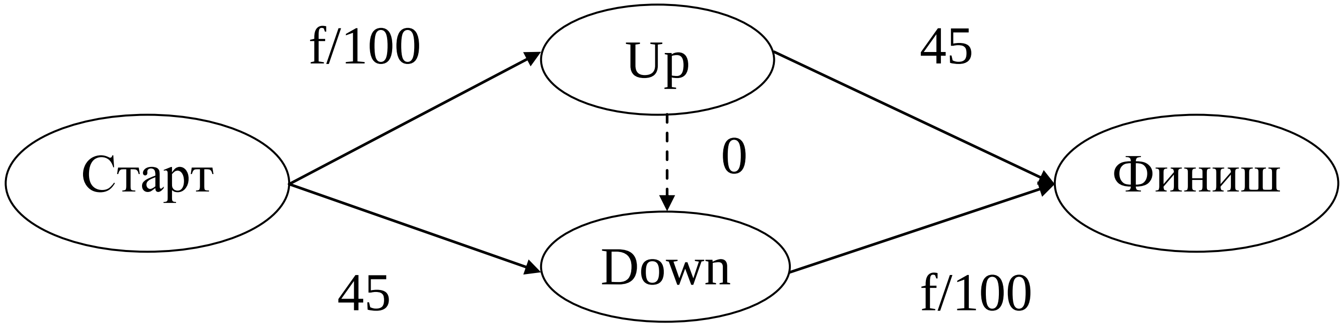
Равновесие: $x_{up} = 2000\ авт/час, x_{down} = 2000\ авт/час,$

$$\tilde{T}_{up}(x) = \tilde{T}_{down}(x) = 65\ мин.$$

Равновесие (если есть ребро Up->Down): $x_{up,down} = 4000\ авт/час,$

$$\tilde{T}_{up,down}(x) = 80\ мин.$$

Парадокс Браеса (модель Бэкмана, $m=1$, $\gamma=0$)

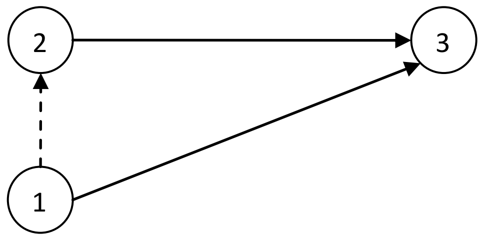


$$\boxed{\Psi=\sum_{e\in E}\sigma_e\left(f_e\right)+\gamma\sum_{p\in P_w}x_p\ln\left(x_p/d_w\right)\to\min_{f=\Theta x,\ x\in X}},\ \sigma_e\left(f_e\right)=\int_0^{f_e}\tau_e\left(z\right)dz.\ (2)$$

Платные дороги: $\overline{\tau}\left(f\right)=\tau\left(f\right)+\underbrace{f\tau'\left(f\right)}_{штраф}$. (VCG – механизм);

$$x_{up}=x_{down}=1750\ авт/час,\ x_{up,down}=500\ авт/час.$$

Парадокс Браеса (модель Стабильной Динамики, $m = 1,\ \gamma = 0$)



Корреспонденции: $d_{13} = d_{23} = 1500\ авт/час,\ d_{12} = 0;$

$$\tau_e^{\mu}(f_e) \xrightarrow[\mu \to 0+]{} \begin{cases} \bar{t}_e, & 0 \le f_e < \bar{f}_e \\ [\bar{t}_e, \infty), & f_e = \bar{f}_e \end{cases},\ \text{BPR}:\ \tau_e(f_e) = \bar{t}_e \cdot \left(1 + \eta \cdot \left(f_e / \bar{f}_e\right)^{\frac{1}{\mu}}\right),$$

$$(\bar{f}_{12} = \bar{f}_{13}) = \bar{f}_{23} = 2000\ авт/час,\ \bar{t}_{13} = 1\ час,\ \bar{t}_{23} = 30\ мин,\ \bar{t}_{12} = 15\ мин.$$

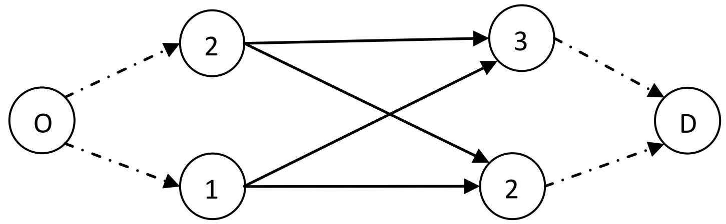
Равновесие (без 1-2): $x_{13} = x_{23} = 1500\ авт/час,\ T_{13} = 1\ час,\ T_{23} = 30\ мин.$

Равновесие (с 1-2): $x_{13} = 1000\ авт/час,\ x_{123} = 500\ авт/час,$

$x_{23} = 1500\ авт/час,\ T_{13} = 1\ час,\ T_{23} = 45\ мин.$

Если в (2) перейти к пределу $\mu \to 0+$, то: $\boxed{\sum_{e \in E} f_e \bar{t}_e \to \min_{f = \Theta x,\ f \le \bar{f},\ x \in X}}.$

Энтропийная модель расчета матрицы корреспонденций ($m=1$)



$$\sum_{i=1,\ j=1}^{n,\ n} T_{ij} d_{ij} + \sum_{i=1}^{n} \left(\lambda_i^L \sum_{j=1}^{n} d_{ij} \right) + \sum_{i=1}^{n} \left(\lambda_j^W \sum_{i=1}^{n} d_{ij} \right) + \gamma \sum_{i,\ j=1}^{n} d_{ij} \ln d_{ij} \ \to \min_{\sum_{i,\ j=1}^{n} d_{ij} = N,\ \{d_{ij}\} \geq 0},$$

$\left\{ \lambda_i^L, \lambda_j^W \right\}$ – потенциалы “притяжения” районов (Канторович–Гавурин).

К сожалению, они не известны. Зато известны такие $\left\{ L_i, W_j \right\}$, что

$$\sum_{j=1}^{n} d_{ij} = L_i,\ \sum_{i=1}^{n} d_{ij} = W_j\ \left(N = \sum_{i=1}^{n} L_i = \sum_{j=1}^{n} W_j \right). \qquad \text{(A)}$$

$$\boxed{\sum_{i=1,\ j=1}^{n,\ n} T_{ij} d_{ij} + \gamma \sum_{i,\ j=1}^{n} d_{ij} \ln d_{ij} \ \to \min_{d \in (\mathrm{A}),\ \{d_{ij}\} \geq 0}.}$$

Проблема (возникающая при попытке построить многостадийную модель): $d(T)$ возникает в модели расчета матрицы корреспонденций, а $T\left(\left\{\tau_e\left(f_e(d)\right)\right\}\right)$, где $f_e(d)$ рассчитывается согласно модели Бэкмана или модели Стабильной Динамики. Получается порочный круг!

На практике проблему решали с помощью метода простых итераций (искали неподвижную точку). Однако не было установлено, что такая процедура сходится и, тем более, не было оценок скорости сходимости. В диссертации предложен новый способ записи этой задачи!

Ключевое наблюдение:

$$T\left(\left\{\tau_e\left(f_e(d)\right)\right\}\right) = \partial_d\left(\min_{f=\Theta x,\ x\in X(d)} \Psi(f)\right) \left(\nabla_x \Psi\left(f(x)\right) = \tilde{T}(x), T_{ij} = \min_{p\in P_{ij}} \tilde{T}_p\right).$$

$$\boxed{\Psi(f) + \gamma \sum_{i,j=1}^{n} d_{ij} \ln d_{ij} \ \to \min_{d\in(\mathrm{A}),\ \{d_{ij}\}\geq 0;\ f=\Theta x,\ x\in X(d)}}.$$

$$\min_{x\in X}\Psi\big(f(x)\big)=\min_{f,x}\left\{\sum_{e\in E}\sigma_e\left(f_e\right):\ f=\Theta x,\ x\in X\right\}=$$

$$=\min_{f,x}\left\{\sum_{e\in E}\max_{t_e\in\mathrm{dom}\ \sigma_e^*}\left[f_e t_e-\sigma_e^*\left(t_e\right)\right]:\ f=\Theta x,\ x\in X\right\}=$$

$$=\max_{t\in\mathrm{dom}\ \sigma^*}\left\{\min_{f,x}\left[\sum_{e\in E}f_e t_e:\ f=\Theta x,\ x\in X\right]-\sum_{e\in E}\sigma_e^*\left(t_e\right)\right\}=$$

$$=\max_{t\in\mathrm{dom}\ \sigma^*}\left\{\sum_{w\in OD}d_w T_w\left(t\right)-\sum_{e\in E}\sigma_e^*\left(t_e\right)\right\},$$

где $T_w\left(t\right)=\min_{p\in P_w}\sum_{e\in E}\delta_{ep}t_e$ – длина кратчайшего пути из i в j (Дейкстра).

$$T\left(\left\{\tau_e\left(f_e\left(d\right)\right)\right\}\right)=\partial_d\max_{t\in\mathrm{dom}\ \sigma^*}\left\{\sum_{w\in OD}d_w T_w\left(t\right)-\sum_{e\in E}\sigma_e^*\left(t_e\right)\right\}=\partial_d\min_{f=\Theta x,\ x\in X\left(d\right)}\Psi\left(f\left(x\right)\right).$$

Необходимые сведения о сложности решения задач выпуклой оптимизации

$$F(x) \to \min_{x\in Q}.$$

Под решением понимаем

$$E\left[F\left(x^N\right)\right] - F_* \le \varepsilon.$$

N – число вычислений (стохастического) градиента $F(x)$.

R – “расстояние” от точки старта до (ближайшего) решения.

N	$E\left[\lVert\nabla F(x,\xi)\rVert_*^2\right] \le M^2$	$\lVert\nabla F(y) - \nabla F(x)\rVert_* \le L\lVert y - x\rVert$	$E\left[\lVert\nabla F(x,\xi) - \nabla F(x)\rVert_*^2\right] \le D$
$F(x)$ выпуклая	$\frac{M^2R^2}{\varepsilon^2}$	$\sqrt{\frac{LR^2}{\varepsilon}}$	$\max\left\{\sqrt{\frac{LR^2}{\varepsilon}}, \frac{DR^2}{\varepsilon^2}\right\}$
$F(x)$ μ-сильно выпуклая в норме $\lVert\ \rVert$	$\frac{M^2}{\mu\varepsilon}$	$\sqrt{\frac{L}{\mu}}\left\lceil \ln\left(\frac{\mu R^2}{\varepsilon}\right)\right\rceil$	$\max\left\{\sqrt{\frac{L}{\mu}}\left\lceil \ln\left(\frac{\mu R^2}{\varepsilon}\right)\right\rceil, \frac{D}{\mu\varepsilon}\right\}$

Универсальный метод

Рассматривается задача выпуклой композитной оптимизации (размерность вектора x велика)

$$F(x) = f(x) + h(x) \to \min_{x \in Q}. \qquad (3)$$

Положим $R^2 = V(x_*, y^0)$, где прокс-расстояние (расстояние Брэгмана) определяется формулой

$$V(x,z) = d(x) - d(z) - \langle \nabla d(z), x - z \rangle;$$

прокс-функция $d(x) \geq 0$ ($d(y^0) = 0$, $\nabla d(y^0) = 0$) считается сильно выпуклой относительно выбранной нормы $\| \ \|$, с константой сильной выпуклости ≥ 1; x_* – решение задачи (3) (если решение не единственно, то выбирается то, которое доставляет минимум $V(x_*, y^0)$).

Предположение 1. *Пусть*

$$\left\|\nabla f\left(y\right)-\nabla f\left(x\right)\right\|_{*}\le L_{\nu}\left\|y-x\right\|^{\nu},\ \nu\in\left[0,1\right].$$

Предположение 2. *Пусть* $f\left(x\right)$ – μ*-сильно выпуклая функция в норме* $\|\ \|$*, т.е. для любых* $x, y\in Q$ *имеет место неравенство*

$$f\left(y\right)+\left\langle\nabla f\left(y\right),x-y\right\rangle+\frac{\mu}{2}\left\|x-y\right\|^{2}\le f\left(x\right).$$

Введем (в евклидовом случае $\tilde{\omega}_n=1$)

$$\tilde{\omega}_n=\sup_{x,y\in Q}\frac{2V\left(x,y\right)}{\left\|y-x\right\|^{2}},\ \tilde{\mu}=\mu/\tilde{\omega}_n.$$

$$\varphi_0(x) = V(x, y^0) + \alpha_0\left[f(y^0) + \langle \nabla f(y^0), x - y^0 \rangle + \tilde{\mu} V(x, y^0) + h(x) \right],$$

$$\varphi_{k+1}(x) = \varphi_k(x) + \alpha_{k+1}\left[f(y^{k+1}) + \langle \nabla f(y^{k+1}), x - y^{k+1} \rangle + \tilde{\mu} V(x, y^k) + h(x) \right].$$

Универсальный Метод Треугольника (УМТ)

Положим $A_0 = \alpha_0 = 1/L_0^0$, $k = 0$, $j_0 = 0$. До тех пор пока

$$f(x^0) > f(y^0) + \langle \nabla f(y^0), x^0 - y^0 \rangle + \frac{L_0^{j_0}}{2}\left\| x^0 - y^0 \right\|^2 + \frac{\alpha_0}{2A_0}\varepsilon,$$

где $x^0 := u^0 := \arg\min_{x \in Q} \varphi_0(x)$, $(A_0 :=)\, \alpha_0 := \dfrac{1}{L_0^{j_0}}$, выполнять

$$j_0 := j_0 + 1;\ L_0^{j_0} := 2^{j_0} L_0^0.$$

1. $L_{k+1}^{0} = L_{k}^{j_k}/2,\ j_{k+1} = 0.$

2. $\alpha_{k+1} := \dfrac{1+A_k\tilde{\mu}}{2L_{k+1}^{j_{k+1}}} + \sqrt{\dfrac{1+A_k\tilde{\mu}}{4\left(L_{k+1}^{j_{k+1}}\right)^2} + \dfrac{A_k\cdot\left(1+A_k\tilde{\mu}\right)}{L_{k+1}^{j_{k+1}}}},\ A_{k+1} := A_k + \alpha_{k+1},$ (*)

$$y^{k+1} := \frac{\alpha_{k+1}u^k + A_k x^k}{A_{k+1}},\ u^{k+1} := \arg\min_{x\in Q}\varphi_{k+1}(x),\ x^{k+1} := \frac{\alpha_{k+1}u^{k+1} + A_k x^k}{A_{k+1}}.\ (**)$$

До тех пор пока

$$f\left(y^{k+1}\right) + \left\langle \nabla f\left(y^{k+1}\right), x^{k+1} - y^{k+1}\right\rangle + \frac{L_{k+1}^{j_{k+1}}}{2}\left\|x^{k+1} - y^{k+1}\right\|^2 + \frac{\alpha_{k+1}}{2A_{k+1}}\varepsilon < f\left(x^{k+1}\right),$$

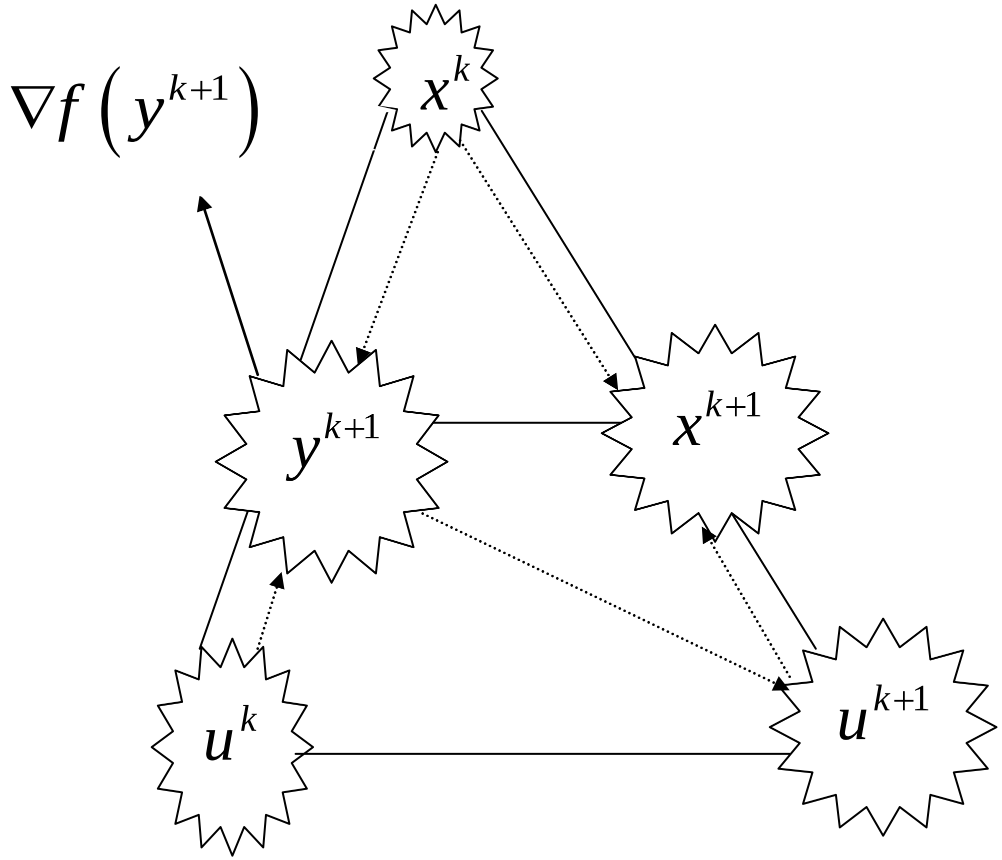
выполнять $j_{k+1} := j_{k+1} + 1;\ L_{k+1}^{j_{k+1}} = 2^{j_{k+1}} L_k^{j_k};$ (*), (**).

3. Если не выполнен критерий останова, то $k := k+1$ и **go to** 1.

Геометрическая интерпретация

$$\boxed{\begin{aligned} y^{k+1} &= \frac{\alpha_{k+1}u^k + A_k x^k}{A_{k+1}}, \\ u^{k+1} &= \arg\min_{x\in Q}\varphi_{k+1}\left(x\right), \\ x^{k+1} &= \frac{\alpha_{k+1}u^{k+1} + A_k x^k}{A_{k+1}}. \end{aligned}}$$

для простоты: $h\left(x\right)=0,\ \mu=0$



$$u^{k+1} = \mathrm{Mirr}_{u^k}\left(\alpha_{k+1}\nabla f\left(y^{k+1}\right)\right) \qquad \boxed{\mathrm{Mirr}_{x^k}\left(\mathrm{v}\right) = \arg\min_{x\in Q}\left\{\left\langle \mathrm{v}, x - x^k\right\rangle + V\left(x, x^k\right)\right\}.}$$

Теорема 2. *Пусть предположение 1 выполняется хотя бы для* $\nu = 0$*, и справедливо предположение 2 с* $\mu \geq 0$ *(допускается брать* $\mu = 0$*). Тогда УМТ для задачи (3) сходится согласно оценке*

$$F\left(x^N\right) - \min_{x \in Q} F\left(x\right) \leq \varepsilon,$$

$$N\left(\varepsilon\right) \approx \min\left\{ \inf_{\nu \in [0,1]} \left(\frac{L_\nu \cdot \left(16R\right)^{1+\nu}}{\varepsilon} \right)^{\frac{2}{1+3\nu}}, \right.$$

$$\left. \inf_{\nu \in [0,1]} \left\{ \left(\frac{8 L_\nu^{\frac{2}{1+\nu}} \tilde{\omega}_n}{\mu \varepsilon^{\frac{1-\nu}{1+\nu}}} \right)^{\frac{1+\nu}{1+3\nu}} \ln^{\frac{2+2\nu}{1+3\nu}} \left(\frac{16 L_\nu^{\frac{4+6\nu}{1+\nu}} R^2}{\left(\mu / \tilde{\omega}_n\right)^{\frac{1+\nu}{1+3\nu}} \varepsilon^{\frac{5+7\nu}{2+6\nu}}} \right) \right\} \right\}. \quad (4)$$

При этом среднее число вычислений значения функции на одной итерации будет ≈ 4*, а градиента функции* ≈ 2.

Заметим, что при этом для любого $k = 0,1,2,...$ имеют место неравенства

$$\left\|u^k - x_*\right\|^2 \le 2R^2, \ \max\left\{\left\|x^k - x_*\right\|^2, \left\|y^k - x_*\right\|^2\right\} \le 4R^2 + 2\left\|x^0 - y^0\right\|^2.$$

Если $\inf$ достигается при $\nu = 0$, то УМТ соответствует (с точностью до логарифмического множителя) по оценке скорости сходимости методу зеркального спуска, а если при $\nu = 0$, то быстрому градиентному методу.

Предположим теперь, что вместо настоящих градиентов нам доступны только стохастические градиенты $\nabla f(x) \to \nabla f(x,\xi)$ (можно обобщить на случай, когда доступны только реализации значения функции $f(x) \to f(x,\xi)$).

Предположение 3. *Пусть для всех* $x \in Q$

$$E_{\xi}\left[\nabla f(x,\xi)\right] = \nabla f(x) \text{ и } E_{\xi}\left[\left\| f(x,\xi) - \nabla f(x)\right\|_*^2\right] \le D.$$

Введём обозначение (mini-batch'инг)

$$\overline{\nabla}^m f(x) = \frac{1}{m}\sum_{k=1}^{m} \nabla f\left(x, \xi^k\right),$$

где ξ^k – независимые одинаково распределенные (так же как ξ) случайные величины. Переопределим последовательность

$$\varphi_0(x) = V\left(x, y^0\right) + \alpha_0\left[f\left(y^0\right) + \left\langle \overline{\nabla}^m f\left(y^0\right), x - y^0 \right\rangle + \tilde{\mu} V\left(x, y^0\right) + h(x)\right],$$

$$\varphi_{k+1}(x) = \varphi_k(x) + \alpha_{k+1}\left[f\left(y^{k+1}\right) + \left\langle \overline{\nabla}^m f\left(y^{k+1}\right), x - y^{k+1} \right\rangle + \tilde{\mu} V\left(x, y^k\right) + h(x)\right].$$

Тогда если дополнительно в условиях теоремы 2 имеет место предположение 3 и на шаге 2 УМТ ввести $m_{k+1} := 8DA_{k+1} / L_{k+1}^{j_{k+1}} \alpha_{k+1} \varepsilon$, заменив условие выхода из цикла следующим

$$f\left(y^{k+1}\right)+\left\langle \bar{\nabla}^{m_{k+1}} f\left(y^{k+1}\right), x^{k+1}-y^{k+1}\right\rangle+\frac{L_{k+1}^{j_{k+1}}}{2}\left\|x^{k+1}-y^{k+1}\right\|^{2}+\frac{\alpha_{k+1}}{2A_{k+1}}\varepsilon< f\left(x^{k+1}\right),$$

то оценка (4) видоизменится следующим образом: $N\left(\varepsilon\right)\rightarrow 2N\left(\varepsilon/4\right)$. При этом среднее число вычислений значения функции на одной итерации будет ≈ 4. А оценка среднего числа обращений за стохастическим градиентом примет вид

$$2\cdot\min\left\{\frac{64DR^{2}}{\varepsilon^{2}}, \frac{8D\tilde{\omega}_{n}}{\mu\varepsilon}\ln\left(\frac{8L_{0}^{j_{0}}R^{2}}{\varepsilon}\right)\right\}+4N\left(\varepsilon/4\right). \tag{5}$$

Оценки (4), (5) сохранят свой вид (немного увеличатся числовые коэффициенты), если считать, что вместо оракула, выдающего градиент (несмещенную оценку градиента) и значение функции, у нас есть доступ только к введенному Деволдером–Нестеровым–Глинёром (δ, L, μ)-оракулу с

$$\boxed{\delta = \mathrm{O}\left(\varepsilon / N(\varepsilon)\right)} \text{ и } L = \mathrm{O}\left(\max_{k=0,\ldots,N} L_k^{j_k}\right).$$

Этот оракул на запрос, в котором указывается только одна точка x, выдает такую пару $\left(f_\delta(x), g_\delta(x,\xi)\right)$ (можно обобщить на случай, когда $f_\delta(x) \to f_\delta(x,\xi)$), что для всех $x \in Q$

$$E_\xi\left[\left\|g_\delta\left(x,\xi\right)-E_\xi\left[g_\delta\left(x,\xi\right)\right]\right\|_*^2\right]\le D,$$

и для любых $x, y \in Q$

$$\frac{\mu}{2}\left\|y-x\right\|^2 \le f\left(y\right)-f_\delta\left(x\right)-\left\langle E_\xi\left[g_\delta\left(x,\xi\right)\right], y-x\right\rangle \le \frac{L}{2}\left\|y-x\right\|^2+\delta.$$

С точностью до числовых и логарифмических множителей оценки (4), (5) оптимальны (А.С. Немировский, 1979). Оптимальна и оценка $\delta = \mathrm{O}\left(\varepsilon/N\left(\varepsilon\right)\right)$ на уровень допустимого шума (Деволдер–Нестеров–Глинёр, 2013).

Рассмотрим конкретные примеры приложений описанных результатов. В этих примерах известно, что существует $L = L_1 < \infty$ (см. обозначения в предположении 1).

Пример 1 (min max задачи). Рассматривается седловая задача (говорят также, что функционал представим в форме Лежандра)

$$f(x) = \max_{\|y\|_2 \le R_y} \left\{ G(y) + \langle By, x \rangle \right\} \to \min_{\|x\|_2 \le R_x},$$

где функция $G(y)$ – сильно вогнутая с константой μ относительно 2-нормы и константой Липшица градиента L_G (также в 2-норме). Тогда функция $f(x)$ будет гладкой, с константой Липшица градиента в 2-

норме $L_f = \sigma_{\max}(B)/\mu$. Казалось бы, что можно решить задачу минимизации функции $f(x)$ за

$$\mathrm{O}\left(\sqrt{\sigma_{\max}(B)R_x^2/(\mu\varepsilon)}\right)$$

итераций, где ε – желаемая точность по функции. Но это возможно, только если мы можем абсолютно точно находить

$$\nabla f(x) = By^*(x),$$

где $y^*(x)$ – решение вспомогательной задачи максимизации по y при заданном x. В действительности, мы можем решать эту задачу (при различных x) лишь приближенно. Если мы решаем вспомогательную задачу УМТ с точностью $\delta/2$ (на это потребуется $\mathrm{O}\left(\sqrt{L_G/\mu}\ln\left(L_G R_y^2/\delta\right)\right)$ итераций), то пара

$$\left(G\left(y_{\delta/2}(x)\right)+\left\langle By_{\delta/2}(x),x\right\rangle,By_{\delta/2}(x)\right),$$

где $y_{\delta/2}(x)$ – $\delta/2$-решение вспомогательной задачи, будет $\left(\delta,2L_f,0\right)$-оракулом (Деволдер–Нестеров–Глинёр, 2013). Выбирая

$$\delta=\mathrm{O}\left(\varepsilon\sqrt{\varepsilon/\left(L_f R_x^2\right)}\right),$$

получим после

$$\mathrm{O}\left(\sqrt{\frac{L_G\sigma_{\max}(B)R_x^2}{\mu^2\varepsilon}}\ln\left(\frac{L_f L_G R_x^2 R_y^2}{\varepsilon}\right)\right)$$

итераций (на итерациях производится умножение матрицы B на вектор/строчку и вычисление градиента $G(y)$) ε-решение задачи минимизации $f(x)$. Отметим, что если не использовать сильную вогнутость функции $G(y)$, то для получения пары $\left(x^N, y^N\right)$, удовлетворяющей неравенству (что, фактически, отвечает решению исходной задачи с точностью ε)

$$\max_{\|y\|_2 \le R_y} \left\{ G(y) + \left\langle By, x^N \right\rangle \right\} - \min_{\|x\|_2 \le R_x} \left\{ G\left(y^N\right) + \left\langle By^N, x \right\rangle \right\} \le \varepsilon,$$

потребуется

$$\Omega\left(\max\left\{ L_G R_y^2, \sigma_{\max}(B) R_x R_y \right\} \big/ \varepsilon \right)$$

итераций. □

Пример 2 (min min задачи). Пусть

$$f(x) = \min_{y \in Q} \Phi(y, x),$$

где Q – ограниченное выпуклое множество, а $\Phi(y, x)$ – такая достаточно гладкая, выпуклая по совокупности переменных функция, что при $y, y' \in Q$

$$\left\| \nabla \Phi(y', x') - \nabla \Phi(y, x) \right\|_2 \le L \left\| (y', x') - (y, x) \right\|_2.$$

Пусть для произвольного x (для простоты здесь считаем, что $x \in \mathbb{R}^n$) можно найти такой $\tilde{y} = \tilde{y}(x) \in Q$, что

$$\max_{z \in Q} \left\langle \nabla_y \Phi(\tilde{y}, x), \tilde{y} - z \right\rangle \le \delta.$$

Тогда

$$\Phi(\tilde{y},x)-f(x)\le\delta,$$

$$\left\|\nabla f(x')-\nabla f(x)\right\|_2\le L\left\|x'-x\right\|_2,$$

и

$$\left(\Phi(\tilde{y},x)-2\delta,\nabla_y\Phi(\tilde{y},x)\right)$$

будет $(6\delta,2L,0)$-оракулом для $f(x)$. □

Этот пример, оказался полезным для решения двойственной задачи (1) с помощью УМТ (с $\mu=0$).

Пример 3 (восстановление матрицы корреспонденций в компьютерной сети по замерам потоков на линках). а) Рассмотрим следующую задачу выпуклой композитной оптимизации (вместо ограничения $\sum_{k=1}^{n} x_k = 1$ можно рассматривать ограничение $\sum_{k=1}^{n} x_k \le 1$)

$$F(x) = \frac{1}{2}\|Ax - b\|_2^2 + \mu \sum_{k=1}^{n} x_k \ln x_k \to \min_{\sum_{k=1}^{n} x_k = 1,\ x \ge 0}.$$

Разберем два случая а) $0 < \mu \ll \varepsilon/(2\ln n)$ – мало (сильную выпуклость композита в 1-норме можно не учитывать); б) $\mu \gg \varepsilon/(2\ln n)$ – достаточно большое (сильную выпуклость композита в 1-норме необходимо учитывать).

Выберем норму в прямом пространстве $\| \; \| = \| \; \|_1$. Положим

$$f(x) = \frac{1}{2}\|Ax - b\|_2^2,\ h(x) = \mu \sum_{k=1}^{n} x_k \ln x_k,\ Q = S_n(1) = \left\{ x \geq 0 : \sum_{k=1}^{n} x_k = 1 \right\},$$

$L = \max\limits_{k=1,\dots,n} \left\| A^{\langle k \rangle} \right\|_2^2$, где $A^{\langle k \rangle}$ – k-й столбец матрицы A. Для случая а) можно выбирать $d(x) = \ln n + \sum\limits_{k=1}^{n} x_k \ln x_k$. Тогда $V(x,z) = \sum\limits_{k=1}^{n} x_k \ln(x_k / z_k)$, $R^2 \leq \ln n$.

Здесь имеет место ситуация, когда композит совпадает по форме с прокс-расстоянием (энтропийного типа), и шаг итерации УМТ (с $\mu = 0$) осуществим по явным формулам. Таким образом, стоимость итерации будет $\mathrm{O}(nnz(A))$, где $nnz(A)$ – число ненулевых элементов в матрице A (считаем, что это число $\geq n$). Оценка числа итераций будет иметь вид (см. теорему 2)

$$N = \mathrm{O}\left(\sqrt{\frac{\max\limits_{k=1,\dots,n}\left\|A^{\langle k\rangle}\right\|_2^2 \ln n}{\varepsilon}}\right). \ \square$$

К сожалению, использовать УМТ в случае б) нельзя, поскольку сильно выпуклым является композит, а не гладкая часть функционала $f(x)$. Однако можно на базе УМТ (с $\mu = 0$) построить с помощью рестартов оптимальный метод. Описываемые далее (во многом известные) конструкции позволяют переносить алгоритмы с сильно выпуклых задач на не сильно выпуклые задачи и наоборот, сохраняя при этом оптимальность. Эти конструкции (наряду с mini-batch'ингом и переходом к двойственной задаче) являются основными операциями, которые активно используются на протяжении всей диссертационной работы.

Введем семейство μ-сильно выпуклых в норме $\| \ \|$ задач ($\mu > 0$)

$$F^{\mu}(x) = F(x) + \mu V\left(x, y^{0}\right) \to \min_{x \in Q}. \tag{6}$$

Пусть F_*^{μ} – оптимальное значение в задаче (6), а F_* – в задаче (3).

Утверждение 1 (регуляризация). *Пусть*

$$\mu \leq \frac{\varepsilon}{2V\left(x_*, y^{0}\right)} = \frac{\varepsilon}{2R^{2}},$$

и удалось найти $\varepsilon/2$-решение задачи (6), т.е. нашелся такой $x^{N} \in Q$, что

$$F^{\mu}\left(x^{N}\right) - F_*^{\mu} \leq \varepsilon/2.$$

Тогда

$$F\left(x^{N}\right) - F_* \leq \varepsilon.$$

Утверждение 2 (рестарты). *Пусть справедливо предположение 1 с* $\nu = 1$ *(*$L = L_1$*), функция* $F(x)$ *–* μ*-сильно выпуклая в норме* $\| \ \|$*. Пусть точка* $x^{\bar{N}}(y^0)$ *выдается УМТ (с* $\mu = 0$*), стартующим с точки* y^0*, после*

$$\bar{N} = \sqrt{\frac{16L\omega_n}{\mu}}$$

итераций, где (следует сравнить с введенным ранее $\tilde{\omega}_n$*)*

$$\omega_n = \sup_{x \in Q} \frac{2V(x, y^0)}{\|x - y^0\|^2}.$$

Положим $\left[x^{\bar{N}}(y^0)\right]^1 = x^{\bar{N}}(y^0)$ *и определим по индукции*

$$\left[x^{\bar{N}}\left(y^0\right)\right]^{k+1} = x^{\bar{N}}\left(\left[x^{\bar{N}}\left(y^0\right)\right]^k\right),\ k = 1, 2, \ldots .$$

При этом на $k+1$ *перезапуске (рестарте) также корректируется прокс-функция (считаем, что так определенная функция корректно определена на Q с сохранением свойства сильной выпуклости)*

$$d^{k+1}(x) = d\left(x - \left[x^{\bar{N}}\left(y^0\right)\right]^k + y^0\right) \geq 0,$$

чтобы $d^{k+1}\left(\left[x^{\bar{N}}\left(y^0\right)\right]^k\right) = 0,\ \nabla d^{k+1}\left(\left[x^{\bar{N}}\left(y^0\right)\right]^k\right) = 0.$ *Тогда*

$$F\left(\left[x^{\bar{N}}\left(y^0\right)\right]^k\right) - F_* \leq \frac{\mu \left\| y^0 - x_* \right\|^2}{2^{k+1}}.$$

Пример 3. б) В этом случае следует использовать рестарт-технику (см. утверждение 2). Но для выбранной в п. а) функции $V(x,z)$ (расстояние Кульбака–Лейблера) процедура рестартов некорректна. Однако существует другой способ выбора прокс-функции

$$d(x)=\frac{1}{2(a-1)}\|x\|_a^2,\ a=\frac{2\ln n}{2\ln n-1}.$$

В этом случае имеем $R^2=\mathrm{O}(\ln n)$, $\omega_n=\mathrm{O}(\ln n)$. Сложность выполнения одной итерации (дополнительная к вычислению градиента гладкой части функционала $\mathrm{O}(nnz(A))$) определяется тем, насколько эффективно можно решить задачу следующего вида

$$\tilde{F}(x) = \langle c, x\rangle + \|x\|_a^2 + \bar{\mu}\sum_{k=1}^{n} x_k \ln x_k \to \min_{x\in S_n(1)}. \tag{7}$$

В конце второй главы показывается (на базе концепции неточного оракула) с помощью перехода к двойственной задаче и ее (приближенного) решения с помощью прямо-двойственной версии метода эллипсоидов, что задачу (7) можно решить за $\mathrm{O}\left(n\ln^2(n/\varepsilon)\right)$ арифметических операций, что в типичных ситуациях много меньше $\mathrm{O}\left(nnz(A)\right)$. При этом (см. утверждение 2 и теорему 2) необходимое число итераций можно оценить следующим образом

$$N = \mathrm{O}\left(\sqrt{\frac{\max\limits_{k=1,\ldots,n}\left\|A^{\langle k\rangle}\right\|_2^2 \ln n}{\mu}}\left\lceil \log_2\left(\frac{\mu}{\varepsilon}\right)\right\rceil\right).$$

Заметим, что из этой формулы с помощью утверждения 1 можно получить (с точностью до $\sim\sqrt{\ln n}$) оценку примера 3 а). □

Этот пример имел одной из своих целей продемонстрировать, что для большого класса задач отсутствие явных формул для шага итерации – не есть сколько-нибудь сдерживающие обстоятельство для использования метода.

Регуляризация $\mu \sim \varepsilon / R^2$

N	$E\left[\left\|\partial_x f(x,\xi)\right\|_*^2\right] \le M^2$	$\left\|\nabla f(y) - \nabla f(x)\right\|_* \le L\left\|y - x\right\|$	$E\left[\left\|\nabla_x f(x,\xi) - \nabla f(x)\right\|_*^2\right] \le D$
$F(x)$ μ-сильно выпуклая в норме $\\|\ \\|$	$\frac{M^2}{\mu\varepsilon}$	$\sqrt{\frac{L}{\mu}}\left\lceil \ln\left(\frac{\mu R^2}{\varepsilon}\right)\right\rceil$	$\max\left\{\sqrt{\frac{L}{\mu}}\left\lceil \ln\left(\frac{\mu R^2}{\varepsilon}\right)\right\rceil, \frac{D}{\mu\varepsilon}\right\}$
$F(x)$ выпуклая	$\frac{M^2R^2}{\varepsilon^2}$	$\sqrt{\frac{LR^2}{\varepsilon}}$	$\max\left\{\sqrt{\frac{LR^2}{\varepsilon}}, \frac{DR^2}{\varepsilon^2}\right\}$

Техника рестартов

N	$E\left[\left\|\partial_x f(x,\xi)\right\|_*^2\right] \le M^2$	$\left\|\nabla f(y) - \nabla f(x)\right\|_* \le L\left\|y - x\right\|$	$E\left[\left\|\nabla_x f(x,\xi) - \nabla f(x)\right\|_*^2\right] \le D$
$F(x)$ выпуклая	$\frac{M^2R^2}{\varepsilon^2}$	$\sqrt{\frac{LR^2}{\varepsilon}}$	$\max\left\{\sqrt{\frac{LR^2}{\varepsilon}}, \frac{DR^2}{\varepsilon^2}\right\}$
$F(x)$ μ-сильно выпуклая в норме $\\|\ \\|$	$\frac{M^2}{\mu\varepsilon}$	$\sqrt{\frac{L}{\mu}}\left\lceil \ln\left(\frac{\mu R^2}{\varepsilon}\right)\right\rceil$	$\max\left\{\sqrt{\frac{L}{\mu}}\left\lceil \ln\left(\frac{\mu R^2}{\varepsilon}\right)\right\rceil, \frac{D}{\mu\varepsilon}\right\}$

Прямо-двойственные методы

Пусть требуется решать задачу

$$g(x) \to \min_{Ax=b,\ x\in Q}, \tag{8}$$

где функция $g(x)$ – 1-сильно выпуклая в p-норме $(1 \le p \le 2)$. Построим двойственную задачу (см. также пример 1)

$$f(y) = \max_{x\in Q}\left\{\langle y, b - Ax\rangle - g(x)\right\} \to \min_{y}. \tag{9}$$

Во многих важных приложениях основной вклад в вычислительную сложность внутренней задачи максимизации дает умножение Ax, $A^T y$.

Это так, например, для сепарабельных функционалов

$$g(x)=\sum_{k=1}^{n} g_k(x_k)$$

и параллелепипедных ограничениях Q. В частности, это имеет место для задач энтропийно-линейного программирования (ЭЛП), возникающих при расчете матрицы корреспонденций, в которых имеется явная формула $x(y)$.

Положим $L=\max_{\|x\|_p\le 1}\|Ax\|_2^2$. В частности, для задачи ЭЛП $p=1$, т.е. $L=\max_{k=1,\dots,n}\left\|A^{\langle k\rangle}\right\|_2^2$, где $A^{\langle k\rangle}$ – k-й столбец матрицы $A^{\langle k\rangle}$. Для задачи поиска вектора PageRank (см. ниже) $p=2$, т.е. $L=\lambda_{\max}\left(A^T A\right)=\sigma_{\max}(A)$.

Пусть УМТ (с $\mu = 0$) с $\| \; \| = \| \; \|_2$, $d(y) = \frac{1}{2}\|y\|_2^2$, $y^0 = 0$, для задачи (9) генерирует “модельные” точки $\{y^k\}$ (на основе которых строятся, по Ю.Е. Нестерову, модели функционала задачи $\varphi_k(y)$), а выдает в итоге $\tilde{y}^N$. Положим

$$x^N = \sum_{k=0}^{N} \lambda_k x\left(y^k\right),\ \lambda_k = \alpha_k / A_N .$$

Поскольку (x_* – решение задачи (8))

$$g\left(x^N\right) - g\left(x_*\right) \le f\left(\tilde{y}^N\right) + g\left(x^N\right),$$

то следующий результат, позволяет с контролируемой точность восстановить решение задачи (8).

Теорема 3. *Пусть нужно решить задачу (8) посредством перехода к задаче (9), исходя из выписанных выше формул. Выбираем в качестве критерия останова УМТ (с $\mu = 0$) выполнение следующих условий*

$$f\left(\tilde{y}^N\right) + g\left(x^N\right) \le \varepsilon, \ \left\|Ax^N - b\right\|_2 \le \tilde{\varepsilon}.$$

Тогда метод гарантированно остановится, сделав не более чем

$$6 \cdot \max\left\{\sqrt{\frac{LR^2}{\varepsilon}}, \sqrt{\frac{LR}{\tilde{\varepsilon}}}\right\}$$

итераций, где $R^2 = \left\|y_*\right\|_2^2$, y_* – решение задачи (9) (если решение не единственно, то можно считать, что выбирается то y_*, которое доставляет минимум R^2).

Метод зеркального спуска

Рассмотрим задачу выпуклой условной оптимизации

$$f\left(x\right) \to \min_{g(x)\le 0,\ x\in Q}. \tag{10}$$

Под решением этой задачи будем понимать такой $\bar{x}^N \in Q \subseteq \mathbb{R}^n$, что с вероятностью $\ge 1-\sigma$ имеет место неравенство (определение M_f, M_g см. ниже)

$$f\left(\bar{x}^N\right) - f_* \le \varepsilon_f = \frac{M_f}{M_g}\varepsilon_g,\ g\left(\bar{x}^N\right) \le \varepsilon_g, \tag{11}$$

где $f_* = f\left(x_*\right)$ – оптимальное значение функционала в задаче (10), x_* – решение задачи (10). Выберем точку старта $x^1 = \arg\min_{x\in Q} d\left(x\right)$. Считаем, что $d\left(x^1\right)=0$, $\nabla d\left(x^1\right)=0$. Как и для УМТ определим $R^2=V\left(x_*,x^1\right)$, где x_* – решение задачи (10) (если решение не единственно, то выбирается то, которое доставляет минимум $V\left(x_*,x^1\right)$). Рассмотрим сначала случай ограниченного множества. Тогда можно также определить $\bar{R}^2=\max_{x,y\in Q}V\left(y,x\right)$. Будем считать, что имеется такая последовательность независимых случайных величин $\left\{\xi^k\right\}$ и последовательности $\left\{\nabla_x f\left(x,\xi^k\right)\right\}$, $\left\{\nabla_x g\left(x,\xi^k\right)\right\}$, $k=1,...,N$, что имеют место следующие соотношения

$$E_{\xi^k}\left[\nabla_x f\left(x,\xi^k\right)\right]=\nabla f\left(x\right),\ E_{\xi^k}\left[\nabla_x g\left(x,\xi^k\right)\right]=\nabla g\left(x\right); \tag{12}$$

$$\left\|\nabla_x f\left(x,\xi^k\right)\right\|_*^2 \le M_f^2,\ \left\|\nabla_x g\left(x,\xi^k\right)\right\|_*^2 \le M_g^2 \tag{13}$$

или

$$E_{\xi^k}\left[\left\|\nabla_x f\left(x,\xi^k\right)\right\|_*^2\right] \le M_f^2,\ E_{\xi^k}\left[\left\|\nabla_x g\left(x,\xi^k\right)\right\|_*^2\right] \le M_g^2\ . \tag{14}$$

На каждой итерации $k = 1,\ldots,N$ нам доступен стохастический (суб-)градиент $\nabla_x f\left(x,\xi^k\right)$ или $\nabla_x g\left(x^k,\xi^k\right)$ в одной, выбранной нами (методом), точке x^k. Определим оператор "проектирования" согласно введенному прокс-расстоянию

$$\mathrm{Mirr}_{x^k}\left(\mathrm{v}\right) = \arg\min_{y\in Q}\left\{\left\langle \mathrm{v}, y - x^k\right\rangle + V\left(y, x^k\right)\right\}.$$

МЗС для задачи (10) будет иметь вид

Метод зеркального спуска (МЗС)

$$\boxed{\begin{aligned} x^{k+1} &= \mathrm{Mirr}_{x^k}\left(h_f \nabla_x f\left(x^k,\xi^k\right)\right), \text{ если } g\left(x^k\right) \le \varepsilon_g, \\ x^{k+1} &= \mathrm{Mirr}_{x^k}\left(h_g \nabla_x g\left(x^k,\xi^k\right)\right), \text{ если } g\left(x^k\right) > \varepsilon_g, \end{aligned}} \tag{15}$$

где $h_g = \varepsilon_g / M_g^2$, $h_f = \varepsilon_g / \left(M_f M_g\right)$, $k = 1,\ldots,N$. Обозначим через I множество индексов k, для которых $g\left(x^k\right) \le \varepsilon_g$. Введем также обозначения

$$\left[N\right] = \left\{1,\ldots,N\right\},\ J = \left[N\right] \setminus I,\ N_I = \left|I\right|,\ N_J = \left|J\right|,\ \bar{x}^N = \frac{1}{N_I}\sum_{k\in I} x^k .$$

В сформулированной далее теореме, предполагается, что последовательность $\left\{x^k\right\}_{k=1}^{N+1}$ генерируется методом (15).

Теорема 4. *Пусть справедливы условия (12),* **(***14). Тогда при любом*

$$N \ge \frac{2M_g^2 R^2}{\varepsilon_g^2} + 1$$

выполняются неравенства $E\left[N_I\right] \ge 1$ *и*

$$E\left[f\left(\overline{x}^N\right)\right] - f_* \le \varepsilon_f,\; g\left(\overline{x}^N\right) \le \varepsilon_g\ .$$

Пусть справедливы условия (12), (13). Тогда при любом

$$N \ge \frac{81 M_g^2 \overline{R}^2}{\varepsilon_g^2} \ln\left(\frac{1}{\sigma}\right)$$

с вероятностью $\ge 1 - \sigma$ *выполняются неравенства* $N_I \ge 1$ *и (11):*

$$f\left(\overline{x}^N\right) - f_* \le \varepsilon_f,\; g\left(\overline{x}^N\right) \le \varepsilon_g.$$

В случае неограниченного множества Q и отсутствия ограничения $g(x) \leq 0$ приведенная теорема останется верной, если заменить в ее формулировке $\bar{R}$ на $\max\left\{\bar{R}, \tilde{R}\right\}$, где $\tilde{R} = \sup_{x \in \tilde{Q}} \|x - x_*\|$,

$$\tilde{Q} = \left\{x \in Q:\ \|x - x_*\|^2 \leq 65R^2 \ln\left(4N/\sigma\right)\right\}.$$

При этом с вероятностью $\geq 1 - \sigma/2$ имеет место включение

$$\left\{x^k\right\}_{k=1}^{N+1} \in \tilde{Q}.$$

Google problem (PageRank)

Описанный МЗС (в рандомизированном, но не условном варианте, с $\|\ \| = \|\ \|_1$ и прокс-функцией энтропией) лежит в основе ряда подходов поиска вектора PageRank (Назина–Поляка, Юдицкого–Лана–Немировского–Шапиро, Григориадиса–Хачияна, см. таблицу 1 ниже – примечательно, что в последнем случае рандомизация осуществляется при “проектировании” на симплекс, а не при замене градиента его несмещенной случайной оценкой, как в других двух подходах).

Известно, что поисковая система Google была создана в качестве учебного проекта студентов Л. Пейжда и С. Брина из Стэнфордского университета. В 1998 г. Лари Пейждом и Сергеем Брином был предложен специальный способ ранжирования web-страниц. Этот способ, также как и довольно большой класс задач ранжирования, возникающих, например, при вычислении индексов цитирования ученых или журналов, сводится к нахождению левого собственного вектора p_* (нормированного на единицу: $\sum_{k=1}^{n} p_k = 1,\ \ p_k \geq 0$), отвечающего собственному значению 1, некоторой стохастической (по строкам) матрицы

$P = \left\| p_{ij} \right\|_{i,j=1}^{n,n}$, т.е. p_* – решение в классе распределений вероятности системы $\boxed{p^T = p^T P}$, $n \gg 1$ (предполагаем, что имеется всего один класс сообщающихся состояний, поэтому решение единственно).

Приведем краткое резюме сложностных оценок основных алгоритмов поиска вектора PageRank. “Сложность” понимается как количество арифметических операций типа умножения двух чисел, которые достаточно осуществить, чтобы с вероятностью не меньше $1-\sigma$ достичь точности решения ε по “Целевому” функционалу (во всех приводимых оценках сложности мы опускаем аддитивное слагаемое $\mathrm{O}(n)$, отвечающее за препроцессинг).

Метод	**Условие**	**Сложность**	**Цель (min)**
Назина–Поляка	нет	$\mathrm{O}\left(\dfrac{n\ln(n/\sigma)}{\varepsilon^2}\right)$	$\left\Vert P^T p - p\right\Vert_2^2$
методы Ю.Е. Нестерова	$\bar{S}$	$\mathrm{O}\left(\dfrac{s^2\ln n}{\varepsilon^2}\right)$	$\left\Vert P^T p - p\right\Vert_2$
вариация алгоритма Юдицкого–Лана–Немировского–Шапиро	нет	$\mathrm{O}\left(\dfrac{n\ln(n/\sigma)}{\varepsilon^2}\right)$	$\left\Vert P^T p - p\right\Vert_\infty$
Нестерова–Немировского	G, S	$\dfrac{sn}{\alpha}\ln\left(\dfrac{2}{\varepsilon}\right)$	$\left\Vert p - p_*\right\Vert_1$

Поляка–Трембы	S	$\frac{2sn}{\varepsilon}$	$\left\| P^T p - p \right\|_1$
Д. Спилмана	G, S	$O\left(\frac{s^2}{\alpha\varepsilon} \ln\left(\frac{1}{\varepsilon} \right) \right)$	$\left\| p - p_* \right\|_\infty$
вариация метода условного градиента из параграфа 2 главы 4	$\bar{S}$	$O\left(\frac{s^2 \ln\left(n/s^2 \right)}{\varepsilon^2} \right)$	$\left\| P^T p - p \right\|_2$
вариация метода Григориадиса–Хачияна из параграфов 1, 2 главы 4	$\bar{S}$	$O\left(\frac{s \ln n \ln\left(n/\sigma \right)}{\varepsilon^2} \right)$	$\left\| P^T p - p \right\|_\infty$

MCMC в варианте параграфа 1 главы 4 (см. также Приложение)	SG	$\mathrm{O}\left(\frac{\ln n \ln(n/\sigma)}{\alpha\varepsilon^2}\right)$	$\lVert p - p_* \rVert_2$
прямой ACRCD* из главы 5 с $\beta = 1/2$	S	$\mathrm{O}\left(\frac{sn}{\sqrt{\varepsilon}}\right)$	$\lVert Ap - b \rVert_2^2$
двойственный ACRCD* из главы 5 с $\beta = 1/2$ (см. также главу 3)	S	$\mathrm{O}\left(sn\sqrt{\frac{\bar{L}R}{\varepsilon}}\right)$	$\lVert Ap - b \rVert_2$

Таблица 1

В таблице 1 использовались следующие обозначения

$$A = \begin{pmatrix} \left(P^T - I\right) \\ 1 \; \ldots \; 1 \end{pmatrix}, \; b = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}, \; \bar{L} = \left(\frac{1}{n+1} \sum_{k=1}^{n+1} \left\| A_k \right\|_2 \right)^2,$$

где A_k – вектор, находящийся в k-й строке матрицы A, а R определяется в теореме 3 для пары задач (8), (9), в которых

$$x = p, \; g(x) = \frac{1}{2}\|x\|^2, \; Q = \mathbb{R}^n.$$

Покомпонентные методы

В развитие примера 3, в котором использовалась Minimal Mutual Information Model для восстановления матрицы корреспонденций по замерам потоков на линках (ребрах) в большой компьютерной сети, изучена и другая модель восстановления матрицы корреспонденций (Tomogravity Model), приводящая к задаче

$$\frac{L}{2}\|Ax-b\|_2^2+\frac{\mu}{2}\|x-x_g\|_2^2\to\min_{x\in\mathbb{R}^n}.$$

Для этой задачи можно построить двойственную

$$\frac{1}{2\mu}\left(\|x_g-A^T y\|_2^2-\|x_g\|_2^2\right)+\frac{1}{2L}\left(\|y+b\|_2^2-\|b\|_2^2\right)\to\min_{y\in\mathbb{R}^m}.$$

В реальных приложениях битовая матрица Кирхгофа A ($m \times n \sim 10^5 \times 10^8$) является сильно разреженной (в среднем $s \ll m$ ненулевых элементов в каждом столбце и в среднем $\tilde{s} = sn/m$ ненулевых элементов в каждом столбце). Используя эту специфику, в главе 5 предлагаются ускоренные покомпонентные методы решения прямой и двойственной задачи, общее время работы которых (с точностью до множителя $\sim \ln\left(\varepsilon^{-1}\right)$ – это отражено волной в $\tilde{O}(\ )$), соответственно,

$$T^{прям} = \tilde{O}\left(sn\sqrt{\frac{Ls}{\mu}} \right), T^{двойств} = \tilde{O}\left(sn\sqrt{\frac{L\tilde{s}}{\mu}} \right).$$

Отсюда можно сделать довольно неожиданный вывод: при $m \ll n$ стоит решать прямую задачу, а в случае $m \gg n$ – двойственную. Первый случай соответствует приложениям к изучению больших (компьютерных) сетей. Второй случай соответствует задачам, приходящим из анализа данных.

К сожалению, в общем случае для рандомизаций, отличных от покомпонентных, отмеченные выше преимущества исчезают (во всяком случае, с точки зрения имеющейся сейчас теории). Зато, используя рандомизацию на евклидовой сфере, можно существенно сократить число итераций за счет евклидовой асимметрии множества, на котором происходит оптимизация (эти результаты представляются не только новыми, но и весьма неожиданными). Пусть рассматривается задача минимизации μ_p-сильно выпуклой в p-норме функции $f(x)$

на множестве Q, например, единичном шаре в p-норме с оракулом, выдающим производную по случайно выбранному на евклидовой сфере направлению или ее дискретную аппроксимацию – для безградиентного оракула. Пусть выбрана норма $\| \; \| = \| \; \|_p$, $1 \le p \le 2$, $1/p + 1/q = 1$, R^2 – “расстояние” Брэгмана от точки старта до решения, согласованное с этой нормой (см. главу 2). Пусть в предположении 1 $\| \; \| = \| \; \|_2$ и M_2 соответствует L_0, а L_2 соответствует L_1. Тогда существуют такие модификации МЗС и метода треугольников, которые приводят к следующим (оптимальным) оценкам (следует сравнить эти оценки с оценкой (4) в двух наиболее интересных ситуациях $\nu = 0$ и $\nu = 1$), соответственно,

$$\min\left\{ \mathrm{O}\left(\frac{M_2^2 R^2}{\varepsilon^2} \right) \tilde{\mathrm{O}}\left(n^{2/q} \right), \tilde{\mathrm{O}}\left(\frac{M_2^2}{\mu_p \varepsilon} \right) \tilde{\mathrm{O}}\left(n^{2/q} \right) \right\},$$

$$\min\left\{ \mathrm{O}\left(\sqrt{\frac{L_2 R^2}{\varepsilon}} \right) \tilde{\mathrm{O}}\left(n^{1/q+1/2} \right), \tilde{\mathrm{O}}\left(\sqrt{\frac{L_2}{\mu_p}} \right) \tilde{\mathrm{O}}\left(n^{1/q+1/2} \right) \right\}.$$

Получены оценки на уровень допустимого шума не случайной природы, при котором выписанные оценки сохраняют вид. Оценки можно распространить (подобно (4), (5)) и на задачи стохастической оптимизации.

Результаты, выносимые на защиту

- Показано, что на модель расчета матрицы корреспонденций можно смотреть как на (энтропийно-регуляризованную) разновидность модели Бэкмана.
- Получен оригинальный вывод модели Стабильной Динамики из модели Бэкмана с помощью вырождения функций затрат на прохождения ребер методом внутренних штрафов.
- Предложен общий способ формирования вариационных принципов для поиска равновесий в многостадийных транспортных моделях.

- Предложена вариация универсального быстрого градиентного метода Ю.Е. Нестерова, самонастраивающегося на гладкость задачи для сильно выпуклых задач композитной оптимизации. Предложенную вариацию удалось распространить и на задачи стохастической оптимизации. Получены оценки скорости сходимости, показывающие, что предложенный метод является равномерно оптимальным (по числу итераций, с точностью до числового множителя) для общего класса задач выпуклой оптимизации. Описываемая линейка универсальных методов активно использовалась в диссертации при разработке комплекса программ для расчета различных блоков многостадийной транспортной модели.

Особая благодарность

Нестерову Юрию Евгеньевичу (CORE UCL, Belgium)

Основные публикации

1. Аникин А.С., Гасников А.В., Горнов А.Ю. О неускоренных эффективных методах решения разреженных задач квадратичной оптимизации // Труды МФТИ. – 2016. – Т. 8. № 2. – С. 44–59.
2. Аникин А.С., Гасников А.В., Горнов А.Ю. Рандомизация и разреженность в задачах huge-scale оптимизации на примере работы метода зеркального спуска // Труды МФТИ. – 2016. – Т. 8. № 1. – С. 11–24.
3. Аникин А.С., Гасников А.В., Горнов А.Ю., Камзолов Д.И., Максимов Ю.В., Нестеров Ю.Е. Эффективные численные методы решения задачи PageRank для дважды разреженных матриц // Труды МФТИ. – 2015. – Т. 7. № 4. – С. 74–94.
4. Бабичева Т.С., Гасников А.В., Лагуновская А.А., Мендель М.А. Двухстадийная модель равновесного распределения транспортных потоков // Труды МФТИ. – 2015. Т. 7. № 3. – С. 31–41.
5. Баймурзина Д.Р., Гасников А.В., Гасникова Е.В. Теория макросистем с точки зрения стохастической химической кинетики // Труды МФТИ. – 2015. – Т. 7. № 4. – С. 95–103.
6. Блинкин М.Я., Гасников А.В., Омельченко С.С., Усик И.В. Эволюционный вывод простейшей модели бимодального расщепления спроса на городские передвижения // Труды МФТИ. – 2016. – Т. 8. № 1. – С. 25–31.
7. Гасников А.В. Об эффективной вычислимости конкурентных равновесий в транспортно-экономических моделях // Математическое моделирование. – 2015. – Т. 27. № 12. – С. 121–136.
8. Гасников А.В., Гасникова Е.В. Об энтропийно-подобных функционалах, возникающих в стохастической химической кинетике при концентрации инвариантной меры и в качестве функций Ляпунова динамики квазисредних // Матем. Заметки. – 2013. – Т. 94:6. – С. 819–827.
9. Гасников А.В., Гасникова Е.В. О возможной динамике в модели расчета матрицы корреспонденций (А.Дж. Вильсона) // Труды МФТИ. –2010. – Т. 2. № 4 (8). – С. 45–54.
10. 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.
11. Гасников А.В., Гасникова Е.В., Мациевский С.В., Усик И.В. О связи моделей дискретного выбора с разномасштабными по времени популяционными играми загрузок // Труды МФТИ. – 2015. – Т. 7. № 4. – С. 129–142.
12. Гасников А.В., Гасникова Е.В., Мендель М.А., Чепурченко К.В. Эволюционные выводы энтропийной модели расчета матрицы корреспонденций // Математическое моделирование. – 2016. – Т. 28. № 4. – С. 111–124.

13. 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. – С. 523–534.
14. Гасников А.В., Гасникова Е.В., Федько О.С. О возможной динамике в модели ранжирования web-страниц PageRank и модернизированной модели расчета матрицы корреспонденций // Труды МФТИ. – 2012. – Т. 4. № 2(14). – С. 101–120.
15. Гасников А.В., Двуреченский П.Е. Стохастический промежуточный метод для задач выпуклой оптимизации // ДАН РАН. – 2016. – Т. 467. № 2. – С. 131–134.
16. Гасников А.В., Двуреченский П.Е., Дорн Ю.В., Максимов Ю.В. Численные методы поиска равновесного распределения потоков в модели Бэкмана и в модели стабильной динамики // Математическое моделирование. 2016. – Т. 28. № 10. – С. 40–64.
17. Гасников А.В., Двуреченский П.Е., Камзолов Д.И., Нестеров Ю.Е., Спокойный В.Г., Стецюк П.И., Суворикова А.Л., Чернов А.В. Поиск равновесий в многостадийных транспортных моделях // Труды МФТИ. – 2015. – Т. 7. № 4. – С. 143–155.
18. Гасников А.В., Двуреченский П.Е., Нестеров Ю.Е. Стохастические градиентные методы с неточным оракулом // Труды МФТИ. – 2016. – Т. 8. № 1. – С. 41–91.
19. Гасников А.В., Двуреченский П.Е., Спокойный В.Г., Стецюк П.И., Суворикова А.Л. Суперпозиция метода балансировки и универсального градиентного метода для поиска энтропийно-сглаженного барицентра Вассерштейна и равновесий в многостадийных моделях транспортных потоков // Труды МФТИ. – 2016. – Т. 8. № 3. – С. 5–24.
20. Гасников А.В., Двуреченский П.Е., Усманова И.Н. О нетривиальности быстрых (ускоренных) рандомизированных методов // Труды МФТИ. – 2016. – Т. 8. № 2. – С. 67–100.
21. Гасников А.В., Дмитриев Д.Ю. Об эффективных рандомизированных алгоритмах поиска вектора PageRank // ЖВМ и МФ. – 2015. – Т. 55. № 3. – С. 355–371.
22. 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.
23. Гасников А.В., Камзолов Д.И., Мендель М.С. 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.

24.Гасников А.В., Лагуновская А.А., Морозова Л.Э. О связи имитационной логит динамики в популяционной теории игр и метода зеркального спуска в онлайн оптимизации на примере задачи выбора кратчайшего маршрута // Труды МФТИ. – 2015. – Т. 7. № 4. – С. 104–113.
25. Гасников А.В., Лагуновская А.А., Усманова И.Н., Федоренко Ф.А. Безградиентные прокс-методы с неточным оракулом для негладких задач выпуклой стохастической оптимизации на симплексе // Автоматика и телемеханика. – 2016. – № 10. – С. 57–77.
26.Гасников А.В., Нестеров Ю.Е., Спокойный В.Г. Об эффективности одного метода рандомизации зеркального спуска в задачах онлайн оптимизации // ЖВМ и МФ. – 2015. – Т. 55. № 4. – С. 582–598.
27.Бузун Н.О., Гасников А.В., Гончаров Ф.О. Горбачев О.Г., Гуз С.А., Крымова Е.А., Натан А.А., Черноусова Е.О. Стохастический анализ в задачах. Учебное пособие. Часть 1. Под ред. А.В. Гасникова. М.: МФТИ, 2016. – 212 с.
28.Гасников А.В., Кленов С.Л., Нурминский Е.А., Холодов Я.А., Шамрай Н.Б. Введение в математическое моделирование транспортных потоков. Под ред. А.В. Гасникова. 2-е изд. М.: МЦНМО, 2013. – 427 с. – Режим доступа: http://www.mou.mipt.ru/gasnikov1129.pdf
29.Аникин А.С., Гасников А.В., Горнов А.Ю., Двуреченский П.Е., Семенов В.В. Параллелизуемые двойственные методы поиска равновесий в смешанных моделях распределения потоков в больших транспортных сетях. Труды 40-й международной школы-конференции “Информационные технологии и системы – 2016”. Россия, Санкт-Петербург (Репино), 25–30 сентября 2016. – 8 с. – Режим доступа: https://arxiv.org/ftp/arxiv/papers/1604/1604.08183.pdf
30.Аникин А.С., Гасников А.В., Двуреченский П.Е., Тюрин А.И., Чернов А.В. Двойственные подходы к задачам минимизации сильно выпуклых функционалов простой структуры при аффинных ограничениях // e-print, 2016. – 16 с. – Режим доступа: https://arxiv.org/ftp/arxiv/papers/1602/1602.01686.pdf
31.Ващенко М.П., Гасников А.В., Молчанов Е.Г., Поспелова Л.Я., Шананин А.А. Вычислимые модели и численные методы для анализа тарифной политики железнодорожных грузоперевозок. М.: ВЦ РАН, 2014. – 52 с. – Режим доступа: https://arxiv.org/ftp/arxiv/papers/1501/1501.02205.pdf
32.Гасников А.В. Марковские модели макросистем // e-print, 2014. – 34 с. arXiv:1412.2720
33. Гасников А.В., Гасникова Е.В., Мациевский С.В. Прямо-двойственный метод зеркального спуска для условных задач стохастической композитной оптимизации. Труды 40-й международной школы-конференции “Информационные технологии и системы – 2016”. Россия, Санкт-Петербург (Репино), 25–30 сентября 2016. – 6 с. – Режим доступа:

https://arxiv.org/ftp/arxiv/papers/1604/1604.08194.pdf
34. Гасников А.В., Двуреченский П.Е., Камзолов Д.И. Градиентные и прямые методы с неточным оракулом для задач стохастической оптимизации. Сборник трудов Международной конференции, посвящено 90-летию со дня рождения академика Н.Н. Красовского. Екатеринбург, 15 – 20 сентября 2014. Екатеринбург: Изд-во института математики и механики УрО РАН им. Н.Н. Красовского, 2015. – С. 111–117.
35. Гасников А.В., Крымова Е.А., Лагуновская А.А., Усманова И.Н., Федоренко Ф.А. Стохастическая онлайн оптимизация. Одноточечные и двухточечные нелинейные многорукие бандиты. Выпуклый и сильно выпуклый случаи // e-print, 2015. – 12 с. – Режим доступа: https://arxiv.org/ftp/arxiv/papers/1509/1509.01679.pdf
36. Гасников А.В., Нестеров Ю.Е. Универсальный метод для задач стохастической композитной оптимизации // e-print, 2016. – 25 с. – Режим доступа: https://arxiv.org/ftp/arxiv/papers/1604/1604.05275.pdf
37. Anikin A., Dvurechensky P., Gasnikov A., Golov A., Gornov A., Maximov Yu., Mendel M., Spokoiny V. 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. – 16 p. – Режим доступа: https://arxiv.org/ftp/arxiv/papers/1508/1508.00858.pdf
38. 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. – 11 p. arXiv:1411.4282
39. Bogolubsky L., Dvurechensky P., Gasnikov A., Gusev G., Nesterov Yu., Raigorodskii A., Tikhonov A., Zhukovskii M. Learning Supervised PageRank with Gradient-Based and Gradient-Free Optimization Methods // NIPS, 2016. – 34 p. – Режим доступа: http://arxiv.org/pdf/1603.00717v1.pdf
40. Chernov A., Dvurechensky P., Gasnikov A. Fast Primal-Dual Gradient Method for Strongly Convex Minimization Problems with Linear Constraints // In: Kochetov, Yu. et all (eds.) DOOR-2016. LNCS, V. 9869. P. 391–403. Springer, Heidelberg, 2016.
41. Dvurechensky P., Gasnikov A. Stochastic Intermediate Gradient Method for Convex Problems with Inexact Stochastic Oracle // J. Optim. Theory Appl. – 2016. – V. 171. no. 1. – P. 121–145.
42. Dvurechensky P., Gasnikov A., Gasnikova E., Matsievski S., Rodomanov A., Usik I. Primal-Dual Method for Searching Equilibrium in Hierarchical Congestion Population Games. Supplementary Proceedings of the 9th International Conference on Discrete Optimization and Operations Research and Scientific School (DOOR-2016). Vladivostok, 19–23 September, 2016. P. 584–595.