Primal-Dual Method for Searching Equilibrium in Hierarchical Congestion Population Games

Pavel Dvurechensky[1,3], Alexander Gasnikov[2,3], Evgenia Gasnikova[2], Sergey Matsievsky[4], Anton Rodomanov[5], and Inna Usik[4]

[1] Weierstrass Institute for Applied Analysis and Stochastics, Berlin 10117, Germany,
`pavel.dvurechensky@wias-berlin.de`
[2] Moscow Institute of Physics and Technology,
Dolgoprudnyi 141700, Moscow Oblast, Russia,
`gasnikov@yandex.ru`, `egasnikova@yandex.ru`
[3] Institute for Information Transmission Problems, Moscow 127051, Russia,
[4] Immanuel Kant Baltic Federal University, Kaliningrad 236041, Russia,
`matsievsky@newmail.ru`, `lavinija@mail.ru`
[5] Higher School of Economics National Research University, Moscow 125319, Russia
`anton.rodomanov@gmail.com`

Abstract. In this paper, we consider a large class of hierarchical congestion population games. One can show that the equilibrium in a game of such type can be described as a minimum point in a properly constructed multi-level convex optimization problem. We propose a fast primal-dual composite gradient method and apply it to the problem, which is dual to the problem describing the equilibrium in the considered class of games. We prove that this method allows to find an approximate solution of the initial problem without increasing the complexity.

Keywords: convex optimization, algorithm complexity, dual problem, primal-dual method, logit dynamics, multistage model of traffic flows, entropy, equilibrium

1 Problem Statement

In this subsection, we briefly describe a variational principle for equilibrium description in hierarchical congestion population games. In particular, we consider a multistage model of traffic flows. Further details can be found in [1].

We consider the traffic network described by the directed graph $\Gamma^1 = \langle V^1, E^1 \rangle$. Some of its vertices $O^1 \subseteq V^1$ are sources (origins), and some are sinks (destinations) $D^1 \subseteq V^1$. We denote a set of source-sink pairs by $OD^1 \subseteq O^1 \otimes D^1$. Let us assume that for each pair $w^1 \in OD^1$ there is a flow of network users of the amount of $d_{w^1}^1 := d_{w^1}^1 \cdot M$, where $M \gg 1$, per unit time who moves from the origin of w^1 to its destination. We call the pair $w^1, d_{w^1}^1$ as correspondence.

Let edges Γ^1 be partitioned into two types $E^1 = \tilde{E}^1 \coprod \bar{E}^1$. The edges of type $\tilde{E}^1$ are characterized by non-decreasing functions of expenses $\tau_{e^1}^1(f_{e^1}^1) := \tau_{e^1}^1(f_{e^1}^1/M)$. Expenses $\tau_{e^1}^1(f_{e^1}^1)$ are incurred by those users who use in their path

an edge $e^1 \in \tilde{E}^1$, the flow of users on this edge being equal to $f_{e^1}^1$. The pairs of vertices setting the edges of type $\bar{E}^1$ are in turn a source-sink pairs OD^2 (with correspondences $d_{w^2}^2 = f_{e^1}^1$, $w^2 = e^1 \in \bar{E}_1$) in a traffic network of the second level $\Gamma^2 = \langle V^2, E^2 \rangle$ whose edges are partitioned in turn into two types $E^2 = \tilde{E}^2 \coprod \bar{E}^2$. The edges having type $\tilde{E}^2$ are characterized by non-decreasing functions of expenses $\tau_{e^2}^2(f_{e^2}^2) := \tau_{e^2}^2(f_{e^2}^2/M)$. Expenses $\tau_{e^2}^2(f_{e^2}^2)$ are incurred by those users who use in their path an edge $e^2 \in \tilde{E}^2$, the flow of users on this edge being equal to $f_{e^2}^2$.

The pairs of vertices setting the edges having type $\bar{E}^2$ are in turn source-sink pairs OD^3 (with correspondences $d_{w^3}^3 = f_{e^2}^2$, $w^3 = e^2 \in \bar{E}^2$) in a traffic network of a higher level $\Gamma^3 = \langle V^3, E^3 \rangle$, etc. We assume that in total there are m levels: $\tilde{E}^m = E^m$. Usually, in applications, the number m is small and varies from 2 to 10.

Let $P_{w^1}^1$ be the set of all paths in Γ^1 which correspond to a correspondence w^1. Each user in the graph Γ^1 chooses a path $p_{w^1}^1 \in P_{w^1}^1$ (a consecutive set of the edges passed by the user) corresponding to his correspondence $w^1 \in OD^1$. Having defined a path $p_{w^1}^1$, it is possible to restore unambiguously the edges having type $\bar{E}^1$ which belong to this path. On each of these edges $w^2 \in \bar{E}^1$, user can choose a path $p_{w^2}^2 \in P_{w^2}^2$ ($P_{w^2}^2$ is a set of all paths corresponding in the graph Γ^2 to the correspondence w^2), etc. Let us assume that each user have made the choice.

We denote by $x_{p^1}^1$ the size of the flow of users on a path $p^1 \in P^1 = \coprod_{w^1 \in OD^1} P_{w^1}^1$, $x_{p^2}^2$ the size of the flow of users on a path $p^2 \in P^2 = \coprod_{w^2 \in OD^2} P_{w^2}^2$, etc. Let us notice that

$$x_{p_{w^k}^k}^k \geq 0, \quad p_{w^k}^k \in P_{w^k}^k, \quad \sum_{p_{w^k}^k \in P_{w^k}^k} x_{p_{w^k}^k}^k = d_{w^k}^k, \quad w^k \in OD^k, \quad k = 1, ..., m$$

and that

$$w^{k+1} \left(= e^k\right) \in OD^{k+1} \left(= \bar{E}^k\right), \quad d_{w^{k+1}}^{k+1} = f_{e^k}^k, \quad k = 1, ..., m-1.$$

For all $k = 1, ..., m$, we introduce for the graph Γ^k and the set of paths P^k a matrix

$$\Theta^k = \left\| \delta_{e^k p^k} \right\|_{e^k \in E^k, p^k \in P^k}, \quad \delta_{e^k p^k} = \begin{cases} 1, & e^k \in p^k \\ 0, & e^k \notin p^k \end{cases}.$$

Then, for all $k = 1, ..., m$, the vector f^k of flows on the edges of thegraph Γ^k is defined in a unique way by the vector of flows on the paths $x^k = \left\{ x_{p^k}^k \right\}_{p^k \in P^k}$:

$$f^k = \Theta^k x^k.$$

We introduce the following notation

$$x = \left\{ x^k \right\}_{k=1}^m, \quad f = \left\{ f^k \right\}_{k=1}^m, \quad \Theta = \mathrm{diag} \left\{ \Theta^k \right\}_{k=1}^m.$$

Let us now describe the probabilistic model for the choice of the path by a network user. We assume that each user l of a traffic network who uses a

correspondence $w^k \in OD^k$ at a level k (and simultaniously the edge $e^{k-1}(= w^k) \in \bar{E}^{k-1}$ at the level $k-1$) chooses to use a path $p^k \in P_{w^k}^k$ if

$$p^k = \arg \max_{q^k \in P_{w^k}^k} \{-g_{q^k}^k(t) + \xi_{q^k}^{k,l}\},$$

where $\xi_{q^k}^{k,l}$ are iid random variables with double exponential distribution (also known as Gumbel's distribution) with cumulative distribution function

$$P(\xi_{q^k}^{k,l} < \zeta) = \exp\{-e^{-\zeta/\gamma^k - E}\},$$

where $E \approx 0.5772$ is Euler–Mascheroni constant. In this case

$$M[\xi_{q^k}^{k,l}] = 0, \quad D[\xi_{q^k}^{k,l}] = (\gamma^k)^2 \pi^2/6.$$

Also, it turns out that, when the number of agents on each correspondence $w^k \in OD^k$, $k = 1, ..., m$ tends to infinity, i. e. $M \to \infty$, the limiting distribution of users among paths is the Gibbs's distribution (also known as logit distribution)

$$x_{p^k}^k = d_{w^k}^k \frac{\exp(-g_{p^k}^k(t)/\gamma^k)}{\sum\limits_{\tilde{p}^k \in P_{w^k}^k} \exp(-g_{\tilde{p}^k}^k(t)/\gamma^k)}, p^k \in P_{w^k}^k, w^k \in OD^k, k = 1, ..., m. \quad (1)$$

It is worth noting here that (see Theorem 1 below)

$$\gamma^k \psi_{w^k}^k(t/\gamma^k) = M_{\{\xi_{p^k}^k\}_{p^k \in P_{w^k}^k}} \left[\max_{p^k \in P_{w^k}^k} \{-g_{p^k}^k(t) + \xi_{p^k}^k\} \right].$$

For the sake of convenience we introduce the graph

$$\Gamma = \coprod_{k=1}^m \Gamma^k = \left\langle V, E = \coprod_{k=1}^m \tilde{E}^k \right\rangle$$

and denote $t_e = \tau_e(f_e)$, $e \in E$.

Assume that, for a given vector of expenses t on edges E, which is identical to all users, each user chooses the shortest path at each level based on noisy information and averaging of the information from the higher levels. Then, in the limit number of users tending to infinity, such behavior of users leads to the description of distribution of users on paths/edges given in (1) and the equilibrium configuration in the system is characterized by the vector t for which the vector x, obtained from (1), leads to the vector $f = \Theta x$ satisfying $t = \{\tau_e(f_e)\}_{e \in E}$.

Theorem 1 (Variational principle). *The described above fixed point equilibrium t can be found as a solution of the following problem (here and below we denote by dom σ^* the effective domain of the function conjugated to a function σ)*

$$\min_{f,x}\{\Psi(x, f): \ f = \Theta x, \ x \in X\} = - \min_{t \in \mathrm{dom}\, \sigma^*} \left\{\gamma^1 \psi^1(t/\gamma^1) + \sum_{e \in E} \sigma_e^*(t_e)\right\}, \quad (2)$$

where

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

$$\Psi^2(x) = \sum_{e^2 \in \tilde{E}^2} \sigma_{e^2}^2(f_{e^2}^2) + \Psi^3(x) + \gamma^2 \sum_{w^2 \in \bar{E}^1} \sum_{p^2 \in P_{w^2}^2} x_{p^2}^2 \ln(x_{p^2}^2/d_{w^2}^2), d_{w^2}^2 = f_{w^2}^1,$$

$$\cdots$$

$$\Psi^k(x) = \sum_{e^k \in \tilde{E}^k} \sigma_{e^k}^k(f_{e^k}^k) + \Psi^{k+1}(x) + \gamma^k \sum_{w^k \in \bar{E}^{k-1}} \sum_{p^k \in P_{w^k}^k} x_{p^k}^k \ln(x_{p^k}^k/d_{w^k}^k),$$

$$d_{w^{k+1}}^{k+1} = f_{w^{k+1}}^k,$$

$$\cdots$$

$$\Psi^m(x) = \sum_{e^m \in E^m} \sigma_{e^m}^m(f_{e^m}^m) + \gamma^m \sum_{w^m \in \bar{E}^{m-1}} \sum_{p^m \in P_{w^m}^m} x_{p^m}^m \ln(x_{p^m}^m/d_{w^m}^m),$$

$$d_{w^m}^m = f_{w^m}^{m-1},$$

$$\sigma_e^*(t_e) = \max_{f_e}\left\{ f_e t_e - \int_0^{f_e} \tau_e(z)dz \right\},$$

$$\frac{d\sigma_e^*(t_e)}{dt_e} = \frac{d}{dt_e} \max_{f_e}\left\{ f_e t_e - \int_0^{f_e} \tau_e(z)dz \right\} = f_e : \quad t_e = \tau_e(f_e), \quad e \in E,$$

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

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

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

$$\psi^1(t) = \sum_{w^1 \in OD^1} d_{w^1}^1 \psi_{w^1}^1(t).$$

2   General Numerical Method

In this subsection, we describe one of our contributions made by this paper, namely a general accelerated primal-dual gradient method for composite minimization problems.

We consider the following convex composite optimization problem [3]:

$$\min_{x \in Q} [\phi(x) := f(x) + \Psi(x)]. \tag{3}$$

Here $Q \subseteq E$ is a closed convex set, the function f is differentiable and convex on Q, and function Ψ is closed and convex on Q (not necessarily differentiable).

In what follows we assume that f is L_f-smooth on Q:

$$\|\nabla f(x) - \nabla f(y)\|_* \leq L_f \|x - y\|, \qquad \forall x, y \in Q. \tag{4}$$

We stress that the constant $L_f > 0$ arises only in theoretical analysis and not in the actual implementation of the proposed method. Moreover, we assume that the set Q is unbounded and that L_f can be unbounded on the set Q.

The space E is endowed with a norm $\|\cdot\|$ (which can be arbitrary). The corresponding dual norm is $\|g\|_* := \max_{x \in E}\{\langle g, x \rangle : \|x\| \leq 1\}$, $g \in E^*$. For mirror descent, we need to introduce the Bregman divergence. Let $\omega : Q \to \mathbb{R}$ be a distance generating function, i.e. a 1-strongly convex function on Q in the $\|\cdot\|$-norm:

$$\omega(y) \geq \omega(x) + \langle \omega'(w), y - x \rangle + \frac{1}{2}\|y - x\|^2, \qquad \forall x, y \in Q. \tag{5}$$

Then, the corresponding Bregman divergence is defined as

$$V_x(y) := \omega(y) - \omega(x) - \langle \omega'(x), y - x \rangle, \qquad x, y \in Q. \tag{6}$$

Finally, we generalize the Grad and Mirr operators from [2] to composite functions:

$$\mathrm{Grad}_L(x) := \operatorname*{argmin}_{y \in Q} \left\{ \langle \nabla f(x), y - x \rangle + \frac{L}{2}\|y - x\|^2 + \Psi(y) \right\}, \quad x \in Q,$$

$$\mathrm{Mirr}_z^\alpha(g) := \operatorname*{argmin}_{y \in Q} \left\{ \langle g, y - z \rangle + \frac{1}{\alpha}V_z(y) + \Psi(y) \right\}, \qquad g \in E^*, \ z \in Q. \tag{7}$$

2.1   Algorithm description

Below is the proposed scheme of the new method. The main differences between this algorithm and the algorithm of [2] are as follows: 1) now the Grad and Mirr operators contain the $\Psi(y)$ term inside; 2) now the algorithm does not require the actual Lipschitz constant L_f, instead it requires an arbitrary number L_0[1] and automatically adapts the Lipschitz constant in iterations; 3) now we need to use a different formula for α_{k+1} to guarantee convergence (see next section).

Note that Algorihtm 1 if well-defined in the sense that it is always guaranteed that $\tau_k \in [0, 1]$ and, hence, $x_{k+1} \in Q$ as a convex combination of points from Q. Indeed, from the formula for α_{k+1} we have

$$\alpha_{k+1}L_{k+1} \geq \left(\sqrt{\frac{1}{4L_{k+1}^2} + \frac{1}{2L_{k+1}}} \right) L_{k+1} = 1, \tag{8}$$

therefore $\tau_k = \frac{1}{\alpha_{k+1}L_{k+1}} \leq 1$.

---

[1] The number L_0 can be always set to 1 with virtually no harm to the convergence rate of the method.

Algorithm 1 Accelerated gradient method.

Require: $x_0 \in Q$: initial point; T: number of iterations; L_0: initial estimate of L_f.

$y_0 \leftarrow x_0$, $z_0 \leftarrow x_0$, $\alpha_0 \leftarrow 0$

for $k = 0, \ldots, T - 1$ **do**

    $L_{k+1} \leftarrow \max\{L_0, L_k/2\}$

    while True **do**

        $\alpha_{k+1} \leftarrow \sqrt{\alpha_k^2 \frac{L_k}{L_{k+1}} + \frac{1}{4L_{k+1}^2}} + \frac{1}{2L_{k+1}}$, and $\tau_k \leftarrow \frac{1}{\alpha_{k+1} L_{k+1}}$.

        $x_{k+1} \leftarrow \tau_k z_k + (1 - \tau_k) y_k$

        $y_{k+1} \leftarrow \mathrm{Grad}_{L_{k+1}}(x_{k+1})$

        if $f(y_{k+1}) \le f(x_{k+1}) + \langle \nabla f(x_{k+1}), y_{k+1} - x_{k+1} \rangle + \frac{L_{k+1}}{2} \|y_{k+1} - x_{k+1}\|^2$ **then**

break

        $L_{k+1} \leftarrow 2L_{k+1}$

    end while

    $z_{k+1} \leftarrow \mathrm{Mirr}_{z_k}^{\alpha_{k+1}}(\nabla f(x_{k+1}))$

end for

    return y_T

2.2   Convergence rate

First we prove the analogues of Lemma 4.2 and Lemma 4.3 from [2].

Lemma 1. *For any $u \in Q$ and $\tau_k = \frac{1}{\alpha_{k+1} L_{k+1}}$ we have*

$$\alpha_{k+1} \langle \nabla f(x_{k+1}), z_k - u \rangle \le \alpha_{k+1}^2 L_{k+1}(\phi(x_{k+1}) - \phi(y_{k+1})) + (V_{z_k}(u) - V_{z_{k+1}}(u))$$
$$+ \alpha_{k+1} \Psi(u) - (\alpha_{k+1}^2 L_{k+1}) \Psi(x_{k+1}) + (\alpha_{k+1}^2 L_{k+1} - \alpha_{k+1}) \Psi(y_k). \quad (9)$$

Proof. From the first order optimality condition for $z_{k+1} = \mathrm{Mirr}_{z_k}^{\alpha_{k+1}}(\nabla f(x_{k+1}))$ we get

$$\left\langle \nabla f(x_{k+1}) + \frac{1}{\alpha_k} V_{z_k}'(z_{k+1}) + \Psi'(z_{k+1}), z_{k+1} - u \right\rangle \le 0, \qquad \forall u \in Q. \quad (10)$$

Therefore

$$\begin{aligned}
\alpha_{k+1} &\langle \nabla f(x_{k+1}), z_k - u \rangle \\
&= \alpha_{k+1} \langle \nabla f(x_{k+1}), z_k - z_{k+1} \rangle + \alpha_{k+1} \langle \nabla f(x_{k+1}), z_{k+1} - u \rangle \\
&\le \alpha_{k+1} \langle \nabla f(x_{k+1}), z_k - z_{k+1} \rangle + \langle V_{z_k}'(z_{k+1}), u - z_{k+1} \rangle \\
&\quad + \alpha_{k+1} \langle \Psi'(z_{k+1}), u - z_{k+1} \rangle \\
&\le (\alpha_{k+1} \langle \nabla f(x_{k+1}), z_k - z_{k+1} \rangle - \alpha_{k+1} \Psi(z_{k+1})) \\
&\quad + \langle V_{z_k}'(z_{k+1}), u - z_{k+1} \rangle + \alpha_{k+1} \Psi(u),
\end{aligned} \quad (11)$$

where the second inequality follows from the convexity of Ψ.

Using the triangle equality of the Bregman divergence,

$$\langle V_x'(y), u - y \rangle \;=\; V_x(u) - V_y(u) - V_x(y),$$

we get

$$\langle V'_{z_k}(z_{k+1}), u - z_{k+1} \rangle = V_{z_k}(u) - V_{z_{k+1}}(u) - V_{z_k}(z_{k+1})$$
$$\leq V_{z_k}(u) - V_{z_{k+1}}(u) - \frac{1}{2} \|z_{k+1} - z_k\|^2, \tag{12}$$

where we have used $V_{z_k}(z_{k+1}) \geq \frac{1}{2} \|z_{k+1} - z_k\|^2$ in the last inequality.

So we have

$$\alpha_{k+1} \langle \nabla f(x_{k+1}), z_k - u \rangle$$
$$\leq \left(\alpha_{k+1} \langle \nabla f(x_{k+1}), z_k - z_{k+1} \rangle - \frac{1}{2} \|z_{k+1} - z_k\|^2 - \alpha_{k+1}\Psi(z_{k+1}) \right) \tag{13}$$
$$+ (V_{z_k}(u) - V_{z_{k+1}}(u)) + \alpha_{k+1}\Psi(u)$$

Define $v := \tau_k z_{k+1} + (1 - \tau_k)y_k \in Q$. Then we have $x_{k+1} - v = \tau_k(z_k - z_{k+1})$ and $\tau_k \Psi(z_{k+1}) + (1 - \tau_k)\Psi(y_k) \geq \Psi(v)$ due to convexity of Ψ. Using this and the formula for τ_k, we get

$$\left(\alpha_{k+1} \langle \nabla f(x_{k+1}), z_k - z_{k+1} \rangle - \frac{1}{2} \|z_{k+1} - z_k\|^2 - \Psi(z_{k+1}) \right)$$
$$\leq - \left(\frac{\alpha_{k+1}}{\tau_k} \langle \nabla f(x_{k+1}), v - x_{k+1} \rangle + \frac{1}{2\tau_k^2} \|v - x_{k+1}\|^2 + \frac{\alpha_{k+1}}{\tau_k}\Psi(v) \right)$$
$$+ \frac{\alpha_{k+1}(1 - \tau_k)}{\tau_k}\Psi(y_k)$$
$$\leq -(\alpha_{k+1}^2 L_{k+1}) \left(\langle \nabla f(x_{k+1}), v - x_{k+1} \rangle + \frac{L_{k+1}}{2} \|v - x_{k+1}\|^2 + \Psi(v) \right)$$
$$+ (\alpha_{k+1}^2 L_{k+1} - \alpha_{k+1})\Psi(y_k)$$
$$\leq -(\alpha_{k+1}^2 L_{k+1}) \left(\langle \nabla f(x_{k+1}), y_{k+1} - x_{k+1} \rangle + \frac{L_{k+1}}{2} \|y_{k+1} - x_{k+1}\|^2 + \Psi(y_{k+1}) \right)$$
$$+ (\alpha_{k+1}^2 L_{k+1} - \alpha_{k+1})\Psi(y_k)$$

$$\tag{14}$$

Here the last inequality follows from the definition of y_{k+1}.

Note that by the termination condition for choosing L_{k+1} we have

$$\phi(y_{k+1}) = f(y_{k+1}) + \Psi(y_{k+1})$$
$$\leq f(x_{k+1}) + \langle \nabla f(x_{k+1}), y_{k+1} - x_{k+1} \rangle$$
$$+ \frac{L_{k+1}}{2} \|y_{k+1} - x_{k+1}\|^2 + \Psi(y_{k+1}) \tag{15}$$
$$= \phi(x_{k+1}) + \langle \nabla f(x_{k+1}), y_{k+1} - x_{k+1} \rangle$$
$$+ \frac{L_{k+1}}{2} \|y_{k+1} - x_{k+1}\|^2 + \Psi(y_{k+1}) - \Psi(x_{k+1}).$$

After rearranging:

$$- \left(\langle \nabla f(x_{k+1}), y_{k+1} - x_{k+1} \rangle + \frac{L_{k+1}}{2} \|y_{k+1} - x_{k+1}\|^2 + \Psi(y_{k+1}) \right) \tag{16}$$
$$\leq \phi(x_{k+1}) - \phi(y_{k+1}) - \Psi(x_{k+1}).$$

Hence,

$$
\begin{aligned}
&\left(\alpha_{k+1}\langle\nabla f(x_{k+1}), z_k - z_{k+1}\rangle - \frac{1}{2}\|z_{k+1} - z_k\|^2 - \Psi(z_{k+1})\right) \\
&\leq (\alpha_{k+1}^2 L_{k+1})(\phi(x_{k+1}) - \phi(y_{k+1})) - (\alpha_{k+1}^2 L_{k+1})\Psi(x_{k+1}) \\
&+ (\alpha_{k+1}^2 L_{k+1} - \alpha_{k+1})\Psi(y_k).
\end{aligned}
\tag{17}
$$

Finally, combining the previous estimates, we get

$$
\begin{aligned}
\alpha_{k+1}\langle\nabla f(x_{k+1}), z_k - u\rangle \leq\ & (\alpha_{k+1}^2 L_{k+1})(\phi(x_{k+1}) - \phi(y_{k+1})) \\
& + (V_{z_k}(u) - V_{z_{k+1}}(u)) - (\alpha_{k+1}^2 L_{k+1})\Psi(x_{k+1}) \\
& + (\alpha_{k+1}^2 L_{k+1} - \alpha_{k+1})\Psi(y_k) + \alpha_{k+1}\Psi(u).
\end{aligned}
\tag{18}
$$

$\square$

Lemma 2. *For any $u \in Q$ and $\tau_k = \frac{1}{\alpha_{k+1} L_{k+1}}$ we have*

$$
\begin{aligned}
&(\alpha_{k+1}^2 L_{k+1})\phi(y_{k+1}) - (\alpha_{k+1}^2 L_{k+1} - \alpha_{k+1})\phi(y_k) \\
&\leq \alpha_{k+1}\left(f(x_{k+1}) + \langle\nabla f(x_{k+1}), u - x_{k+1}\rangle + \Psi(u)\right) + (V_{z_k}(u) - V_{z_{k+1}}(u)).
\end{aligned}
\tag{19}
$$

Proof. Using convexity of f and relation $\tau_k(x_{k+1} - z_k) = (1 - \tau_k)(y_k - x_{k+1})$, we obtain

$$
\begin{aligned}
&\alpha_{k+1}(\Psi(x_{k+1}) - \Psi(u)) + \alpha_{k+1}\langle\nabla f(x_{k+1}), x_{k+1} - u\rangle \\
&= \alpha_{k+1}(\Psi(x_{k+1}) - \Psi(u)) + \alpha_{k+1}\langle\nabla f(x_{k+1}), x_{k+1} - z_k\rangle \\
&\quad + \alpha_{k+1}\langle\nabla f(x_{k+1}), z_k - u\rangle \\
&\leq \alpha_{k+1}(\Psi(x_{k+1}) - \Psi(u)) + \frac{\alpha_{k+1}(1 - \tau_k)}{\tau_k}\langle\nabla f(x_{k+1}), y_k - x_{k+1}\rangle \\
&\quad + \alpha_{k+1}\langle\nabla f(x_{k+1}), z_k - u\rangle \\
&\leq \alpha_{k+1}(\Psi(x_{k+1}) - \Psi(u)) + (\alpha_{k+1}^2 L_{k+1} - \alpha_{k+1})(f(y_k) - f(x_{k+1})) \\
&\quad + \alpha_{k+1}\langle\nabla f(x_{k+1}), z_k - u\rangle \\
&\leq \alpha_{k+1}\phi(x_{k+1}) - \alpha_{k+1}\Psi(u) + (\alpha_{k+1}^2 L_{k+1} - \alpha_{k+1})f(y_k) \\
&\quad - (\alpha_{k+1}^2 L_{k+1})f(x_{k+1}) + \alpha_{k+1}\langle\nabla f(x_{k+1}), z_k - u\rangle.
\end{aligned}
\tag{20}
$$

Now we apply Lemma 1 to bound the last term, group the terms and get

$$
\begin{aligned}
&\alpha_{k+1}(\Psi(x_{k+1}) - \Psi(u)) + \alpha_{k+1}\langle\nabla f(x_{k+1}), x_{k+1} - u\rangle \\
&\leq \alpha_{k+1}\phi(x_{k+1}) - (\alpha_{k+1}^2 L_{k+1})\phi(y_{k+1}) \\
&+ (\alpha_{k+1}^2 L_{k+1} - \alpha_{k+1})\phi(y_k) + (V_{z_k}(u) - V_{z_{k+1}}(u)).
\end{aligned}
\tag{21}
$$

After rearranging, we obtain (19). $\square$

Now we are ready to prove the convergence theorem for Algorithm 1.

Theorem 2. *For the sequence $\{y_k\}_{k\geq 0}$ in Algorithm 1 we have*

$$(\alpha_T^2 L_T)\phi(y_T) \leq \min_{x \in Q} \left\{ \sum_{k=1}^{T} \alpha_k \left(f(x_k) + \langle \nabla f(x_k), u - x_k \rangle + \Psi(u) \right) + V_{z_0}(u) \right\}$$

$$(22)$$

and, hence, the following rate of convergence:

$$\phi(y_T) - \phi(x^*) \leq \frac{4 L_f R^2}{T^2}. \tag{23}$$

Proof. Note that the special choice of $\{\alpha_k\}_{k\geq 0}$ in Algorithm 1 gives us

$$\alpha_{k+1}^2 L_{k+1} - \alpha_{k+1} = \alpha_k^2 L_k, \qquad k \geq 0. \tag{24}$$

Therefore, taking the sum over $k = 0, \ldots, T-1$ in (19) and using that $\alpha_0 = 0$, $V_{z_T}(u) \geq 0$ we get, for any $u \in Q$,

$$(\alpha_T^2 L_T)\phi(y_T) \leq \sum_{k=1}^{T} \alpha_k \left(f(x_k) + \langle \nabla f(x_k), u - x_k \rangle + \Psi(u) \right) + V_{z_0}(u) \tag{25}$$

and (22) is straightforward. At the same time, using the convexity of $f(x)$, the definition of $\phi(x)$, and $u = x^* = \operatorname{argmin}_{x \in Q} \phi(x)$, we obtain

$$(\alpha_T^2 L_T)\phi(y_T) \leq \min_{x \in Q} \left\{ \sum_{k=1}^{T} \alpha_k \left(f(x_k) + \langle \nabla f(x_k), u - x_k \rangle + \Psi(u) \right) + V_{z_0}(u) \right\}$$

$$\leq \left(\sum_{k=1}^{T} \alpha_k \right) \phi(x^*) + V_{z_0}(x^*).$$

$$(26)$$

From (24) it follows that $\sum_{k=1}^{T} \alpha_k = \alpha_T^2 L_T$, so

$$\phi(y_T) \leq \phi(x^*) + \frac{1}{\alpha_T^2 L_T} V_{z_0}(x^*). \tag{27}$$

Now it remains to estimate the rate of growth of coefficients $A_k := \alpha_k^2 L_k$. For this we use the technique from [3]. Note that from (24) we have

$$A_{k+1} - A_k = \sqrt{\frac{A_{k+1}}{L_{k+1}}} \tag{28}$$

Rearranging and using $(a + b)^2 \leq 2a^2 + 2b^2$ and $A_k \leq A_{k+1}$, we get

$$A_{k+1} = L_{k+1}(A_{k+1} - A_k)^2 = L_{k+1} \left(\sqrt{A_{k+1}} + \sqrt{A_k} \right)^2 \left(\sqrt{A_{k+1}} - \sqrt{A_k} \right)^2$$

$$\leq 4 L_{k+1} A_{k+1} \left(\sqrt{A_{k+1}} - \sqrt{A_k} \right)^2$$

$$(29)$$

From this it follows that

$$\sqrt{A_{k+1}} \geq \frac{1}{2} \sum_{i=0}^{k} \frac{1}{\sqrt{L_i}}. \tag{30}$$

Note that according to (4) and the stopping criterion for choosing L_{k+1} in Algorithm (1), we always have $L_i \leq 2L_f$. Hence,

$$\sqrt{A_{k+1}} \geq \frac{k+1}{2\sqrt{2L_f}} \qquad \Longleftrightarrow \qquad A_{k+1} \geq \frac{(k+1)^2}{8L_f}. \tag{31}$$

Thus, combining (31) and (27) with $V_{z_0}(x^*) =: \frac{R^2}{2}$, we have proved (23). $\square$

Using the same arguments to [3], it is also possible to prove that the average number of evaluations of the function f per iteration in Algorithm 1 equals 4.

Theorem 3. *Let N_k be the total number of evaluations of the function f in Algorithm 1 after the first k iterations. Then for any $k \geq 0$ we have*

$$N_k \leq 4(k+1) + 2\log_2 \frac{L_f}{L_0}. \tag{32}$$

3 Application to the Equilibrium Problem

In this section, we apply Algorithm 1 to solve the dual problem in (2)

$$\min_{t \in \operatorname{dom} \sigma^*} \left\{ \gamma^1 \psi^1(t/\gamma^1) + \sum_{e \in E} \sigma_e^*(t_e) \right\}.$$

with t in the role of x, $\gamma^1 \psi^1(t/\gamma^1)$ in the role of $f(x)$, and $\sum_{e \in E} \sigma_e^*(t_e)$ in the role of $\Psi(x)$.

The inequality (22) leads to the fact that Algorithm 1 is primal-dual [6–9], which means that the sequences $\{t^i\}$ (which is in the role of $\{x_k\}$) and $\{\tilde{t}^i\}$ (which is in the role of $\{y_k\}$) generated by this method have the following property:

$$\gamma^1 \psi^1(\tilde{t}^T/\gamma^1) + \sum_{e \in E} \sigma_e^*(\tilde{t}_e^T)$$

$$- \min_{t \in \operatorname{dom} \sigma^*} \left\{ \frac{1}{A_T} \sum_{i=0}^{T} \left[\alpha_i(\gamma^1 \psi^1(t^i/\gamma^1) + \langle \nabla \psi^1(t^i/\gamma^1), t - t^i \rangle) \right] + \sum_{e \in E} \sigma_e^*(t_e) \right\} \tag{33}$$

$$\leq \frac{4L_2 R_2^2}{T^2},$$

where

$$L_2 \leq (1/\min_{k=1,\ldots,m} \gamma^k) \sum_{w^1 \in OD^1} d_{w^1}^1 \cdot (l_{w^1})^2,$$

with l_{w^1} being the total number of edges (among all of the levels) in the longest path for correspondence w^1,

$$R_2^2 = \max\{\tilde{R}_2^2, \hat{R}_2^2\}, \quad \tilde{R}_2^2 = (1/2)\left\|\bar{t} - t^*\right\|_2^2, \quad \hat{R}_2^2 = (1/2)\sum_{e \in E} \left(\tau_e\left(\bar{f}_e^N\right) - t_e^*\right)^2,$$

$\bar{f}^N$ is defined in Theorem 2, the method starts from $t^0 = \bar{t}$, t^* is a solution of the problem (2).

Theorem 4. *Let the problem (2) be solved by Algorithm 1 generating sequences* $\{t^i\}, \{\tilde{t}^i\}$. *Then. after T iterations one has*

$$0 \le \left\{\gamma^1 \psi^1(\tilde{t}^T/\gamma^1) + \sum_{e \in E} \sigma_e^*(\tilde{t}_e^T)\right\} + \Psi(\bar{x}^T, \bar{f}^T) \le \frac{4L_2 R_2^2}{T^2},$$

where

$$f^i = \Theta x^i = -\nabla \psi^1(t^i/\gamma^1), \quad x^i = \left\{x_{p^k}^{k,i}\right\}_{p^k \in P_{w^k}^k, w^k \in OD^k}^{k=1,\dots,m},$$

$$x_{p^k}^{k,i} = d_{w^k}^k \frac{\exp(-g_{p^k}^k(t^i)/\gamma^k)}{\sum\limits_{\tilde{p}^k \in P_{w^k}^k} \exp(-g_{\tilde{p}^k}^k(t^i)/\gamma^k)}, \quad p^k \in P_{w^k}^k, \quad w^k \in OD^k, \quad k = 1, \dots, m,$$

$$\bar{f}^T = \frac{1}{A_T}\sum_{i=0}^{T} \alpha_i f^i, \quad \bar{x}^T = \frac{1}{A_T}\sum_{i=0}^{T} \alpha_i x^i.$$

Theorem 2 provides the bound for the number of iterations in order to solve the problem (2) with given accuracy. Nevertheless, on each iteration it is necessary to calculate $\nabla \psi^1(t/\gamma^1)$ and also $\psi^1(t/\gamma^1)$. Similarly to [9–11] it is possible to show, using the smoothed version of Bellman–Ford method, that for this purpose it is enough to perform $O(|O^1||E| \max\limits_{w^1 \in OD^1} l_{w^1})$ arithmetic operations.

In general, it is worth noting that the approach of adding some artificial vertices, edges, sources, sinks is very useful in different applications [12–14].

References

1. Gasnikov, A., Gasnikova, E., Matsievsky, S., Usik, I.: Searching of equilibriums in hierarchical congestion population games. Trudy MIPT. 28, 129–142 (2015)
2. Allen-Zhu, Z., Orecchia, L. Linear coupling: An ultimate unification of gradient and mirror descent. ArXiV preprint:1407.1537v4 (2014)
3. Nesterov, Yu.: Gradient methods for minimizing composite functions. Math. Prog. 140, 125–161 (2013)
4. Nesterov, Yu., Nemirovski, A.: On first order algorithms for l_1/nuclear norm minimization. Acta Numerica. 22, 509–575 (2013)
5. Nesterov, Yu.: Universal gradient methods for convex optimization problems. CORE Discussion Paper 2013/63 (2013)

6. Nesterov, Yu.: Primal-dual subgradient methods for convex problems. Math. Program. Ser. B. 120, 261–283 (2009)
7. Nemirovski, A., Onn, S., Rothblum, U. G.: Accuracy certificates for computational problems with convex structure. Mathematics of Operation Research. 35, 52–78 (2010)
8. Gasnikov, A., Dvurechensky, P., Kamzolov, D., Nesterov, Yu., Spokoiny, V., Stetsyuk, P., Suvorikova, A., Chernov, A.: Searching for equilibriums in multistage transport models. Trudy MIPT. 28, 143–155 (2015)
9. Gasnikov, A., Gasnikova, E., Dvurechensky, P., Ershov, E., Lagunovskaya, A.: Search for the stochastic equilibria in the transport models of equilibrium flow distribution. Trudy MIPT. 28, 114–128 (2015)
10. Ed. Gasnikov, A.: Introduction to mathematical modelling of traffic flows. MCCME, Moscow (2013)
11. Nesterov, Yu.: Characteristic functions of directed graphs and applications to stochastic equilibrium problems. Optim. Engineering. 8, 193–214 (2007)
12. Gasnikov, A.: About reduction of searching competetive equillibrium to the mini-max problem in application to different network problems. Math. Mod. 27, 121–136 (2015)
13. Babicheva, T., Gasnikov, A., Lagunovskaya, A., Mendel, M.: Two-stage model of equilibrium distributions of traffic flows. Trudy MIPT. 27, 31–41 (2015)
14. Vaschenko, V., Gasnikov, A., Molchanov, E., Pospelova, L., Shananin, A.: Analysis of tariff policy of a railway cargo transportation. Preprint of CCAS RAS (2014)