



Primal-dual gradient methods for searching network equilibria in combined models with nested choice structure and capacity constraints

Meruza Kubentayeva¹ · Demyan Yarmoshik¹ · Mikhail Persiiianov^{1,2} · Alexey Kroshnin^{2,3} · Ekaterina Kotliarova¹ · Nazarii Tupitsa¹ · Dmitry Pasechnyuk¹ · Alexander Gasnikov^{1,2,3} · Vladimir Shvetsov^{1,4} · Leonid Baryshev⁴ · Alexey Shurupov⁴

Received: 1 July 2023 / Accepted: 28 November 2023 / Published online: 4 January 2024
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2024

Abstract

We consider a network equilibrium model (i.e. a combined model), which was proposed as an alternative to the classic four-step approach for travel forecasting in transportation networks. This model can be formulated as a convex minimization program. We extend the combined model to the case of the stable dynamics model in the traffic assignment stage, which imposes strict capacity constraints in the network. We propose a way to solve corresponding dual optimization problems with accelerated gradient methods and give theoretical guarantees of their convergence. We conducted numerical experiments with considered optimization methods on Moscow and Berlin networks.

Keywords Forecasting · Combined model · Trip distribution · Traffic assignment · Capacity constraints · Gradient method

1 Introduction

One of the most popular approaches to travel forecasting in transportation networks is the four-step procedure (Dios Ortúzar and Willumsen 2011): sequential run of trip generation, trip distribution, modal split, and traffic assignment stages. However,

✉ Meruza Kubentayeva
kubentay@gmail.com

¹ Moscow Institute of Physics and Technology, Dolgoprudny, Russia

² Institute for Information Transmission Problems RAS, Moscow, Russia

³ Higher School of Economics, Moscow, Russia

⁴ Russian University of Transport, Moscow, Russia

this approach has a number of limitations, e.g. there is no convergence guarantee (Oppenheim 1995; Boyce et al. 1994; Boyce 2002).

To overcome this issue, there were proposed network equilibrium models (NE / combined models) which can be formulated as an optimization or, more generally, a variational inequality problem (Beckmann et al. 1956; De Cea et al. 2005). In particular, Evans (1976) reduced the problem of searching equilibrium in the case of one transport mode to a convex optimization problem, combining trip distribution and route assignment models. Authors of Florian and Nguyen (1978) made an extension to the multi-modal case, where destination and mode are chosen simultaneously with the same value of a calibration parameter. The first mathematical formulation of a network equilibrium model with hierarchical destination and mode choices was proposed in Fernández et al. (1994) — the approach was presented for modelling nested choice structure of trips using several modes (e.g. park'n ride trips). Abrahamsson and Lundqvist (1999) formulated a nested combined model where mode choice is conditioned by destination choice and demonstrated its application for the Stockholm region. The recent works (Chu 2018), Liu et al. (2018), and Gao et al. (2022) proposed the extensions of the combined models for the cases of modeling trip frequency, remote park-and-ride, and tourism demand, respectively.

Finding a solution in trip distribution and traffic assignment problems — whether they are considered separately in the four-step approach or combined into one network equilibrium problem — relies on numerical methods for convex optimization. E.g., a classic choice for the traffic assignment problem (which is the most computationally expensive part) is the Frank–Wolfe algorithm (Frank and Wolfe 1956), and for the trip distribution problem it is the Sinkhorn algorithm (Sinkhorn 1974). A class of path-based algorithms can be an alternative to the link-based Frank–Wolfe algorithm for solving traffic assignment problem: Chen et al. (2020), Xie et al. (2017), Babazadeh et al. (2020). A popular choice for solving an optimization problem in the above-mentioned combined models is a partial linearization algorithm of Evans (1976) and its modifications for multi-modal and multi-user cases (Abrahamsson and Lundqvist 1999; Boyce et al. 1983). Recently, in Yang et al. (2013); Fan et al. (2022); Zarrinmehr et al. (2019); Cabannes et al. (2019); Wang et al. (2022), improvements of these algorithms were presented. Also, in subsequent years there have been developed a lot of new optimization methods, in particular, accelerated gradient methods (Nesterov 2004, 2009, 2015), which can be applied to the described problems.

Another direction of research on travel modelling in recent years is related to capacitated transportation networks, which allow to overcome some limitations of the standard Beckmann traffic assignment model (Nesterov and De Palma 2003; Zokaei Aashtiani et al. 2021; De Cea et al. 2005; Wang et al. 2019; Smith et al. 2019; Anikin et al. 2020; Zhu et al. 2020).

For the best of our knowledge, there is no works considering the application of accelerated gradient methods to combined models.

In this paper, we consider:

- An entropy-based trip distribution model with hierarchical choice structure (Wilson 1969; Fernández et al. 1994; Abrahamsson and Lundqvist 1999);

- The Beckmann traffic assignment model with inelastic demand (Beckmann et al. 1956);
- The stable dynamics traffic assignment model, where resulting flow distribution satisfy the network's capacity constraints (Nesterov and De Palma 2003);
- An NE model combining all the models mentioned above.

In the last case, we consider the nested combined model proposed in Abrahamsson and Lundqvist (1999), where transit and road networks are independent, and the transit network has constant travel costs. We extend it to the case of the stable dynamics model for traffic assignment.

We employ accelerated primal-dual gradient methods to solve corresponding optimization problems and compare their performances to the classic Sinkhorn, Frank–Wolfe, and generalised Evans algorithms. Also, we provide theoretical guarantees for their convergence rate.

The **main contributions** of the paper are the following:

- We propose a way to solve the dual problem of the nested combined model of Abrahamsson and Lundqvist (1999) with a universal accelerated gradient method USTM (Gasnikov and Nesterov 2018);
- We extend the nested combined model of Abrahamsson and Lundqvist (1999) to the case of capacitated networks: namely, we propose a way to solve the dual problem for searching equilibrium in combined trip distribution model with the nested choice structure and the stable dynamics traffic assignment model;
- We provide theoretical upper bounds on the complexity of searching network equilibrium by the USTM algorithm.
- We conducted numerical experiments comparing different algorithms on Moscow and Berlin transportation networks.

The paper is organized as follows. In Sect. 2, we give a general problem statement for a combined trip distribution, modal split, and traffic assignment model. In Sect. 3, we describe the primal-dual accelerated method to solve the NE problem and provide its convergence analysis. In Sects. 4 and 5, we describe optimization algorithms that we consider for separate traffic assignment and trip distribution models. Section 6 presents numerical experiments conducted on Moscow and Berlin transportation networks.

2 Problem statement

We start with the description of the Beckmann and the stable dynamics models for searching the road network user equilibrium. Similarly to Abrahamsson and Lundqvist (1999), we assume the road and the transit networks are independent, and there is no congestion effects in the transit network (its travel costs are constant and defined as the costs of the shortest routes). Then, in Sect. 2.2, we describe the trip distribution model with a hierarchical choice structure of destination and travel mode (by car, public

transport, or on foot). And finally, in 2.3, we consider the combined trip distribution-modal split-assignment problem and formulate its dual problem.

2.1 Route assignment models

Let the urban road network be represented by a directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where vertices \mathcal{V} correspond to intersections or centroids (Sheffi 1985) and edges \mathcal{E} correspond to roads, respectively. Suppose we are given the travel demands: namely, let d_{ij} (veh/h) be a trip rate from origin i to destination j . We denote by P_{ij} the set of all simple paths from i to j . Respectively, $P = \bigcup_{(i,j) \in OD} P_{ij}$ is the set of all possible routes for all origin–destination pairs OD . Agents traveling from node i to node j are distributed among paths from P_{ij} , i.e. for any $p \in P_{ij}$ there is a flow $x_p \in \mathbb{R}_+$ along the path p , and $\sum_{p \in P_{ij}} x_p = d_w$. Flows from origin nodes to destination nodes create the traffic in the entire network \mathcal{G} , which can be represented by an element of

$$X = X(d) = \left\{ x \in \mathbb{R}_+^{|P|} : \sum_{p \in P_{ij}} x_p = d_{ij}, (i, j) \in OD \right\}. \tag{1}$$

Note that the dimension of X can be extremely large: e.g. for $n \times n$ Manhattan network $\log |P| = \Omega(n)$. To describe a state of the network we do not need to know an entire vector x , but only flows on arcs:

$$f_e(x) = \sum_{p \in P} \delta_{ep} x_p \quad \text{for } e \in \mathcal{E},$$

where $\delta_{ep} = \mathbb{1}\{e \in p\}$. Let us introduce a matrix Θ such that $\Theta_{e,p} = \delta_{ep}$ for $e \in \mathcal{E}$, $p \in P$, so in vector notation we have $f = \Theta x$. To describe an equilibrium we use both path- and link-based notations (x, t) or (f, t) .

Beckmann model (Beckmann et al. 1956; Patriksson 2015). One of the key ideas behind the Beckmann model is that the cost (e.g. travel time, gas expenses, etc.) of passing a link e is the same for all agents and depends solely on the flow f_e along it. In what follows, we denote this cost for a given flow f_e by $t_e = \tau_e(f_e)$. In practice the BPR functions are usually employed (US Bureau of Public Roads 1964):

$$\tau_e(f_e) = \bar{t}_e \left(1 + \rho \left(\frac{f_e}{\bar{f}_e} \right)^\mu \right), \tag{2}$$

where \bar{t}_e are free flow times, and \bar{f}_e are road capacities of a given network’s link e . We take these functions with parameters $\rho = 0.15$ and $\mu = 0.25$.

Another essential point is a behavioral assumption on agents called the first Wardrop’s principle: we suppose that each of them knows the state of the whole network and chooses a path p minimizing the total cost

$$T_p(t) = \sum_{e \in p} t_e.$$

The cost functions are supposed to be continuous, non-decreasing, and non-negative. Then (x^*, t^*) , where $t^* = (t_e^*)_{e \in \mathcal{E}}$, is an equilibrium state, i.e. it satisfies conditions

$$t_e^* = \tau_e(f_e^*), \quad \text{where } f^* = \Theta x^*,$$

$$x_{p_w}^* > 0 \implies T_{p_{ij}}(t^*) = T_{ij}(t^*) = \min_{p \in P_{ij}} T_p(t^*),$$

if and only if x^* is a minimum of the potential function:

$$\Psi(x) = \sum_{e \in \mathcal{E}} \underbrace{\int_0^{f_e} \tau_e(z) dz}_{\sigma_e(f_e)} \longrightarrow \min_{f = \Theta x, x \in X} \tag{B}$$

$$\iff \Psi(f) = \sum_{e \in \mathcal{E}} \sigma_e(f_e) \longrightarrow \min_{f = \Theta x: x \in X},$$

and $t_e^* = \tau_e(f_e^*)$ (Beckmann et al. 1956).

Another way to find an equilibrium numerically is by solving a dual problem. We can construct it according to Theorem 4 from Nesterov and De Palma (2003), the solution of which is t^* :

$$Q(t) = \sum_{ij \in OD} d_{ij} T_{ij}(t) - \underbrace{\sum_{e \in \mathcal{E}} \sigma_e^*(t_e)}_{h(t)} \longrightarrow \max_{t \geq \bar{t}} \tag{DualB}$$

where

$$\sigma_e^*(t_e) = \sup_{f_e \geq 0} \{t_e f_e - \sigma_e(f_e)\} = \bar{f}_e \left(\frac{t_e - \bar{t}_e}{\bar{t}_e \rho} \right)^\mu \frac{(t_e - \bar{t}_e)}{1 + \mu}$$

is the conjugate function of $\sigma_e(f_e)$, $e \in \mathcal{E}$.

When we search for the solution to this problem numerically, on every step of an applied method we can reconstruct primal variable f from the current dual variable t : $f \in \partial \sum_{(i,j) \in OD} d_{ij} T_{ij}(t)$. Then we can use the duality gap — which is always nonnegative — for the estimation of the method’s accuracy:

$$\Delta(f, t) = \Psi(f) - Q(t).$$

It vanishes only at the equilibrium (f^*, t^*) .

Stable dynamics model. Nesterov and De Palma (2003) proposed an alternative model called the stable dynamics model, which takes an intermediate place between static and dynamic network assignment models. Namely, its equilibrium can be interpreted as the stationary regime of some dynamic process. Its key assumption is that we no longer introduce a complex dependence of the travel cost on the flow (as in the standard static models), but only pose capacity constraints, i.e. the flow value on each link imposes the feasible set of travel times

$$\tau_e(f_e) = \begin{cases} \bar{t}_e, & 0 \leq f_e < \bar{f}_e, \\ [\bar{t}_e, \infty], & f_e = \bar{f}_e, \\ +\infty, & f_e > \bar{f}_e. \end{cases} \tag{3}$$

Unlike in the Beckmann model, there is no one-to-one correspondence between equilibrium travel times and flows on the links of the network. There are examples in Nesterov and De Palma (2003) illustrating the difference. Also, one can find in Chudak et al. (2007) a detailed comparison of equilibria in these two models conducted for large and small networks.

Hence, an equilibrium state (x^*, t^*) of the stable dynamics model satisfies the next conditions:

$$t_e^* \in \tau_e(f_e^*), \\ x_{p_{ij}}^* > 0 \implies T_{p_{ij}}(t^*) = T_{ij}(t^*).$$

The above formula can be reformulated in terms of an optimization problem. The pair (f^*, t^*) is an equilibrium if and only if it is a solution of the saddle-point problem

$$\sum_{e \in E} [t_e f_e - (t_e - \bar{t}_e) \bar{f}_e] \longrightarrow \min_{f = \Theta x} \max_{t_e \geq \bar{t}_e} \tag{SaddleSD} \\ x \in X$$

where its primal problem is

$$\Psi(f) = \sum_{e \in \mathcal{E}} f_e \bar{t}_e \longrightarrow \min_{f = \Theta x} \tag{SD} \\ x \in X, f_e \leq \bar{f}_e$$

and its dual problem is

$$Q(t) = \sum_{(i,j) \in OD} d_{ij} T_{ij}(t) - \underbrace{\langle t - \bar{t}, \bar{f} \rangle}_{h(t)} \longrightarrow \max_{t_e \geq \bar{t}_e} \tag{DualSD}$$

In contrast with the Beckmann model, the equilibrium state in the stable dynamics model is defined by pair (f^*, t^*) (in particular, it differs from the system optimum (f^*, \bar{t}) in the model only by the time value).

In both cases the dual problem has form

$$Q(t) = \sum_{(i,j) \in OD} d_{ij} T_{ij}(t) - h(t) \longrightarrow \max_{t \geq \bar{t}}$$

The optimization problem is convex, non-smooth and composite.

2.2 Trip distribution with modal split (D-MS)

Let us further assume that there are several trip purposes (demand layers), travel modes (transportation modes), and agents types. We use the logit model with

calibration parameters α_{am}, β_{am} corresponding to choices of travel mode m by agents of the type a . Further, we consider the case when α_{am} values are the same for all travel modes of agent type a :

$$\alpha_{am} := \alpha_a.$$

Necessity of this condition will be explained below.

For example, if we want to make travelling by car (travel mode m_1) unavailable for non-car-owners a_1 , we can set $\beta_{a_1 m_1} := \inf$ to get zero trips $d^{a_1 m_1} = 0$. Thus, for every agent type a we can implicitly set its group (nest) of available travel modes.

To define destination choice model, we use the entropy-based trip distribution model of Wilson (1969). For every trip purpose r (e.g., home-work, home-other) we define calibration parameter γ_r . This parameter defines the sensitivity of agents with the trip purpose r to trip length.

According to Abrahamsson and Lundqvist (1999), Fernández et al. (1994), we consider the following problem:

$$\sum_{i,j,r,a,m} d_{ij}^{ram} T_{ij}^m + \underbrace{\sum_{i,j,r,a} \frac{1}{\gamma_r} d_{ij}^{ra} \ln d_{ij}^{ra} + \sum_{i,j,r,a,m} \frac{1}{\alpha_a} d_{ij}^{ram} \left(\ln \left(\frac{d_{ij}^{ram}}{d_{ij}^{ra}} \right) + \beta_{am} \right)}_{H(d)} \rightarrow \min_{d \in \Pi'(l,w)}, \tag{P1}$$

where

$$\Pi'(l, w) = \left\{ d_{ij}^{ram} \geq 0 : \sum_{j,m} d_{ij}^{ram} = l_i^r, \sum_{i,a,m} d_{ij}^{ram} = w_j^r \right\},$$

d_{ij}^{ram} is a number of trips from zone i to j by travel mode m of agents type a with trip purpose r and $d_{ij}^{ra} = \sum_m d_{ij}^{ram}$, l_i^r is a number of production from zone i of agents type a with trip purpose r ; w_j^r is a number of attractions to zone j of the trip purpose r .

This is the combined trip distribution-modal split (D-MS) problem, where the choice structure is nested: travel mode choice is conditioned by destination choice (Abrahamsson and Lundqvist 1999). If γ_r and α_a are equal, then (P1) reduces to the problem that also corresponds to D-MS model with simultaneous choices of destination and travel mode with the same calibration parameters (Florian and Nguyen 1978; Abrahamsson and Lundqvist 1999).

For fixed values d_{ij}^{ra} , it is straightforward to check that the optimal d_{ij}^{ram} satisfy the following relation:

$$\text{prob}_{ij}^{ram} = \frac{d_{ij}^{ram}}{d_{ij}^{ra}} = \frac{\exp(-\alpha_a T_{ij}^m - \beta_{am})}{\sum_{m'} \exp(-\alpha_a T_{ij}^{m'} - \beta_{am'})}, \tag{4}$$

where $T_{ij}^{am} = T_{ij}^m + \frac{\beta_{am}}{\alpha_a}$, i.e., the modal split corresponds to the logit model. Moreover,

$$\min_{d_{ij}^{ram}} \left(\sum_{i,j,r,a,m} d_{ij}^{ram} T_{ij}^{am} + \frac{1}{\alpha_a} \sum_{i,j,r,a,m} d_{ij}^{ram} \ln \frac{d_{ij}^{ram}}{d_{ij}^{ra}} \right) = d_{ij}^{ra} \underbrace{\left(-\frac{1}{\alpha_a} \ln \sum_m \exp(-\alpha_a T_{ij}^{am}) \right)}_{T_{ij}^a},$$

where T_{ij}^a is a composite travel cost for agents of type a .

Substituting $d_{ij}^{ram} = \text{prob}_{ij}^{ram} d_{ij}^{ra}$ we reduce the problem (P1) to

$$E(d, T) = \sum_{i,j,r,a} d_{ij}^{ra} T_{ij}^a + \sum_{i,j,r,a} \frac{1}{\gamma_r} d_{ij}^{ra} \ln d_{ij}^{ra} \rightarrow \min_{d \in \Pi(l,w)}, \tag{E}$$

where

$$\Pi(l, w) = \left\{ d \geq 0 : \sum_j d_{ij}^{ra} = l_i^r, \sum_{i,a} d_{ij}^{ra} = w_j^r \right\},$$

and $T_{ij}^a = -\frac{1}{\alpha_a} \ln \sum_m \exp(-\alpha_a T_{ij}^m - \beta_{am})$.

Let us derive its dual problem. In our problem statement, the system of constraints $\Pi^l(l, w)$ is consistent $\sum_{i,a} l_i^{r,a} = \sum_j w_j^r$. Therefore Gasnikov et al. (2015), we can introduce a tautological constraint

$$\sum_{i,j,a} d_{ij}^{ra} = \sum_{i,a} l_i^{r,a} = \sum_j w_j^r = N_r. \tag{5}$$

We will utilize tautological constraint (5) to obtain dual function with bounded sub-gradient norm.

$$\begin{aligned} & \min_{d \in \Pi(l,w)} \sum_{i,j,r,a} d_{ij}^{ra} T_{ij}^a + \sum_{i,j,r,a} \frac{1}{\gamma_r} d_{ij}^{ra} \ln d_{ij}^{ra} \\ &= \min_{d_{ij}^{ra} \geq 0} \max_{\lambda \geq 0} \sum_r \frac{1}{\gamma_r} \sum_{i,j,a} d_{ij}^{ra} \ln d_{ij}^{ra} + \sum_{i,j,a} d_{ij}^{ra} T_{ij}^a + \sum_{i,a} \lambda_{rai}^l \left(\sum_j d_{ij}^{ra} - l_i^r \right) \\ & \quad \sum_{i,j,a} d_{ij}^{ra} = N_r \\ & + \sum_j \lambda_{rj}^w \left(\sum_{i,a} d_{ij}^{ra} - w_j^r \right) \\ &= \max_{\lambda \geq 0} \sum_r \min_{d_{ij}^{ra} \geq 0} \frac{1}{\gamma_r} \sum_{i,j,a} d_{ij}^{ra} \ln d_{ij}^{ra} + \sum_{i,j,a} d_{ij}^{ra} T_{ij}^a + \sum_{i,a} \lambda_{rai}^l \left(\sum_j d_{ij}^{ra} - l_i^r \right) \\ & \quad \sum_{i,j,a} d_{ij}^{ra} = N_r \\ & + \sum_j \lambda_{rj}^w \left(\sum_{i,a} d_{ij}^{ra} - w_j^r \right) \\ &= - \min_{\lambda \geq 0} \varphi(\lambda^l, \lambda^w, T), \end{aligned} \tag{6}$$

where by $\varphi(\lambda^l, \lambda^w, T)$ we denoted the negative of the dual function.

Let y be the dual variable for the tautological constraint $\sum_{i,j,a} d_{ij}^{ra} = N_r$. Then taking the gradient by d we get one of the optimality conditions for the inner minimization problem:

$$\frac{1}{\gamma_r} \left(\ln d_{ij}^{ra} + 1 \right) + T_{ij}^a + \lambda_{rai}^l + \lambda_{rj}^w + y = 0,$$

therefore

$$d_{ij}^{ra} = \exp \left(-1 - \gamma_r \left(T_{ij}^a + \lambda_{rai}^l + \lambda_{rj}^w + y \right) \right).$$

By choosing y such that d_{ij}^{ra} satisfies $\sum_{i,j,a} d_{ij}^{ra} = N_r$ we obtain

$$d_{ij}^{ra}(\lambda_r^l, \lambda_r^w, T) = \frac{N_r \exp \left(-\gamma_r \left(T_{ij}^a + \lambda_{rai}^l + \lambda_{rj}^w \right) \right)}{\sum_{i,j,a} \exp \left(-\gamma_r \left(T_{ij}^a + \lambda_{rai}^l + \lambda_{rj}^w \right) \right)}. \tag{7}$$

Substituting this into (6) yields

$$\varphi(\lambda^l, \lambda^w, T) = \sum_r \frac{N_r}{\gamma_r} \ln \frac{1}{N_r} \sum_{i,j,a} \exp \left(-\gamma_r \left(T_{ij}^a + \lambda_{rai}^l + \lambda_{rj}^w \right) \right) + \sum_{i,a} \lambda_{rai}^l P_i^a + \sum_j \lambda_{rj}^w W_j^r. \tag{8}$$

2.3 Combined distribution-modal split-assignment problem (D-MS-A)

Now, we combine the road, the transit, and the pedestrian networks into one multi-modal network, which we denote again by $\mathcal{G} = (\mathcal{V}, \mathcal{E})$. Slightly abusing notations, in the same way as in Sect. 2.1 we can define the set of path flows $X(d)$ corresponding to an interzonal trip matrix $d \in \Pi'(l, w)$, and link flows f_e^{ram} , $f_e^m = \sum_{r,a} f_e^{ram}$, and $f_e = \sum_m f_e^m$.

According to Abrahamsson and Lundqvist (1999), the combined distribution-modal split-assignment problem can be formulated as follows:

$$P_3(f, d) = \Psi(f) + H(d) \rightarrow \begin{matrix} \min \\ f = \Theta x, x \in X(d) \\ d \in \Pi'(l, w) \end{matrix}, \tag{P3}$$

where

$$\Psi(f) = \sum_{e \in \mathcal{E}} \left(\sigma_e(f_e) + \sum_{m \in V} c_e^m f_e^m \right).$$

Similarly to Sect. 2.1 we obtain from (P3) the saddle-point problem

$$S'_3(d, t) = \sum_{i,j,r,a,m} d_{ij}^{ram} T_{ij}^m(t) - h(t) + H(d) \rightarrow \min_{d \in \Pi(l,w)} \max_{t \geq \bar{t}} \tag{S3'}$$

where $T_{ij}^m(t)$ is the minimal cost of the path from $i \in O$ to $j \in D$ with the links cost $t_e + c_e^m$. According to Sect. 2.2, the above problem reduces to

$$S_3(d, t) = \underbrace{\sum_{i,j,r,a} d_{ij}^{ra} T_{ij}^a(t) + \sum_{i,j,r,a} \frac{1}{\gamma_r} d_{ij}^{ra} \ln d_{ij}^{ra}}_{E(d,T(t))} - h(t) \rightarrow \min_{d \in \Pi(l,w)} \max_{t \geq \bar{t}} \tag{S3}$$

where $T_{ij}^a(t) = -\frac{1}{\alpha_a} \ln \left(\sum_m \exp \left(-\alpha_a T_{ij}^m(t) - \beta_{am} \right) \right)$.

Respectively, the dual problem is

$$D_3(\lambda, t) = -\varphi(\lambda^l, \lambda^w, T(t)) - h(t) \longrightarrow \max_{t \geq \bar{t}, \lambda^l, \lambda^w} \tag{D3}$$

Thus, there are several ways to formulate an optimization problem. In this paper, we consider the following particular formulation of the problem and further provide convergence analysis of the accelerated gradient method application to it:

$$D'_3(t) = -\varphi_3(t) - h(t) \longrightarrow \max_{t \geq \bar{t}} \tag{D3'}$$

where

$$\varphi_3(t) = \min_{\lambda} \varphi(\lambda, T(t)) = - \min_{d \in \Pi(l,w)} E(d, T(t)).$$

3 Dual approach for solving the combined model

In Sect. 3.1, we describe the universal gradient method of similar triangles (USTM) for solving the dual problem (D3') of the described combined model. And we provide its convergence analysis in Sect. (3.2).

3.1 Dual method for NE problem

Algorithm 1 Universal Method of Similar Triangles

```

Require:  $L_0 > 0$ , starting point  $t^0$ , accuracy  $\varepsilon > 0$ 
1:  $u^0 := t^0, A_0 := 0, k := 0$ 
2: repeat
3:    $L_{k+1} := L_k/2$ 
4:   while true do
5:      $\alpha_{k+1} := \frac{1}{2L_{k+1}} + \sqrt{\frac{1}{4L_{k+1}^2} + \frac{A_k}{L_{k+1}}}, A_{k+1} := A_k + \alpha_{k+1}$ 
6:      $y^{k+1} := \frac{\alpha_{k+1}u^k + A_k t^k}{A_{k+1}}$ 
7:      $u^{k+1} := \arg \min_{t \in \text{dom } h} \phi_{k+1}(t)$ 
8:      $t^{k+1} := \frac{\alpha_{k+1}u^{k+1} + A_k t^k}{A_{k+1}}$ 
9:     if  $\tilde{\Phi}(t^{k+1}) \leq \tilde{\Phi}(y^{k+1}) + \langle \tilde{\nabla} \Phi(y^{k+1}), t^{k+1} - y^{k+1} \rangle + \frac{L_{k+1}}{2} \|t^{k+1} - y^{k+1}\|_2^2 + \frac{\alpha_{k+1}}{2A_{k+1}} \varepsilon$  then
10:      break
11:     else
12:        $L_{k+1} := 2L_{k+1}$ 
13:     end if
14:   end while
15:    $k := k + 1$ 
16: until Stopping criterion is fulfilled

```

A popular approach for searching equilibrium in combined models is the partial linearization algorithm of Evans (1976) and its modifications for multi-modal and multi-user cases (Abrahamsson and Lundqvist 1999; Boyce et al. 1983). The approach is further developed in Yang et al. (2013) by incorporating better line-search procedures. Note that the algorithm can be viewed as partly dual, because it is formulated in terms external to the primal problem: it includes cost matrices T_{ij} , which are the dual variables of the saddle-point problem (S3) (or (S3')) or the dual problem (D3). But still, the algorithm is essentially primal, since it optimizes (P3) by flow and trip distribution pair.

Here we propose an alternative approach based on solving the dual problem (D3') with the universal method of similar triangles (USTM), and afterwards we prove the convergence rates for it.

Algorithm 1 provides the pseudocode of USTM with an inexact oracle and the euclidean prox-structure. Here we used the following notations:

$$\phi_0(t) = \frac{1}{2} \|t - t^0\|_2^2, \phi_{k+1}(t) = \phi_k(t) + \alpha_{k+1} [\tilde{\Phi}(y^{k+1}) + \langle \tilde{\nabla} \Phi(y^{k+1}), t - y^{k+1} \rangle + h(t)].$$

Note that we did not specify the stopping criterion as it can be different for different models (Fan et al. 2022).

To find a network equilibrium in D-MS-A model, we apply USTM to minimize the composite objective $-D'_3(t) = \varphi_3(t) + h(t)$ in (D3'), thus we set $\Phi(t) := \varphi_3(t)$ in Algorithm 1. Recall that

$$\varphi_3(t) = - \min_{d \in \Pi(l,w)} E(d, T(t)).$$

Note that at each iteration we need to compute φ_3 and $\nabla \varphi_3$ with travel costs y^{k+1} , t^{k+1} . Since this itself is done by an iterative procedure, we cannot expect to find the exact solution of the subproblem. Instead, we use the following inexact oracle for φ_3 :

$$\tilde{\varphi}_3(t) = \varphi_{3,\delta}(t) = -E(\tilde{d}(t), T(t)), \quad \tilde{\nabla} \varphi_3(t) = \nabla_\delta \varphi_3(t) = -\nabla_t E(\tilde{d}(t), T(t)),$$

where $\delta \geq 0$ and $\tilde{d}(t) = d_\delta(t) \in \Pi(l, w)$ (see Sect. 5.4) is a δ -solution of $\min_d E(d, T(t))$, i.e.

$$E(d_\delta(t), T(t)) \leq \min_{d \in \Pi(l,w)} E(d, T(t)) + \delta.$$

Recall that $E(d, T(t))$ is concave w.r.t. t , and its superdifferential $\partial_t E(d, T(t))$ is given by

$$\partial_t E(d, T(t)) = \sum_{i,j,r,a} d_{ij}^{ra} \partial T_{ij}^a(t).$$

Further, $d_{ij}^{ra} \partial T_{ij}^a(t) = \sum_m d_{ij}^{ram} \partial T_{ij}^m(t)$, and

$$\partial T_{ij}^m(t) = \partial \min_{a_p^m \in P_{ij}^m} \langle t, a_{pm} \rangle = \text{Conv} \left\{ a_p^m \in P_{ij}^m : \langle t, a_{pm} \rangle = T_{ij}^m(t) \right\},$$

where $a_p^m \in \{0, 1\}^{|\mathcal{E}|}$ is a binary vector encoding a path p for the travel mode m . Note that several shortest paths may exist. Finally, we get that

$$\partial_t E(d, T(t)) = \sum_{i,j,r,a,m} d_{ij}^{ram} \text{Conv} \left\{ a_p^m \in P_{ij}^m : \langle t, a_{pm} \rangle = T_{ij}^m(t) \right\}, \tag{9}$$

and any supergradient $\nabla_t E(d, T(t))$ is a vector of link flows by shortest paths, corresponding to the trip distribution d .

Since we solve the dual problem (D3'), we need a way to recover an approximate solution (d, f) of the primal problem (P3). For any $t \geq 0$, given $\tilde{d}_{ij}^{ra}(t)$ and $T_{ij}^m(t)$, define $d'(t) \in \Pi'(l, w)$ by formula (4). Then we reconstruct a full correspondence matrix after K iterations of Algorithm 1 as

$$\hat{d}^K = \frac{1}{A_K} \sum_{k=1}^K \alpha_k d'(y^k) \in \Pi'(l, w). \tag{10}$$

Corresponding link flows can be recovered as (see Kubentayeva and Gasnikov 2021, f. (18))

$$[\hat{f}^K]_e^m = \frac{1}{A_K} \sum_{k=1}^K \alpha_k [f^k]_e^m. \tag{11}$$

where f^k are link flows by shortest paths for times y^k and correspondence matrix $d'(y^k)$, such that

$$\sum_m [f^k]_e^m = -[\tilde{\nabla} \varphi_3(y^k)]_e, \quad e \in \mathcal{E}.$$

3.2 Convergence analysis

Below we derive some properties of the problem and then use them to apply the USTM convergence theorem, what gives us the convergence rate of our dual algorithm for searching equilibria in combined model.

The next lemma is a trivial counterpart of f. (5) in Kubentayeva and Gasnikov (2021), following from (9).

Lemma 1 *For any $d, d' \in \Pi(l, w)$, $t, t' \geq 0$, and supergradients $\nabla_t E(d, T(t))$, $\nabla_t E(d', T(t'))$ it holds that $\|\nabla_t E(d, T(t)) - \nabla_t E(d', T(t'))\|_2 \leq M = \sqrt{2}HN$, where $H \leq |\mathcal{V}| - 1$ is the maximum simple path length in the network, and $N = \sum_r N_r$ is the total number of active agents.*

Typically (e.g. for a Manhattan network) $H = O(\sqrt{|\mathcal{V}|})$.

Recall the following standard result concerning inexact oracles.

Lemma 2 *For any $t, t' \geq 0$*

$$\tilde{\varphi}_3(t') + \delta \geq \varphi_3(t') \geq \tilde{\varphi}_3(t) + \langle \tilde{\nabla} \varphi_3(t), t' - t \rangle.$$

Proof Since $E(d, T(t))$ is concave w.r.t. t ,

$$E(\tilde{d}(t), T(t')) \leq E(\tilde{d}(t), T(t)) + \langle \nabla_t E(\tilde{d}(t), T(t)), t' - t \rangle = -\tilde{\varphi}_3(t) - \langle \tilde{\nabla} \varphi_3(t), t' - t \rangle.$$

Thus,

$$\varphi_3(t') \geq -E(\tilde{d}(t), T(t')) \geq \tilde{\varphi}_3(t) + \langle \tilde{\nabla} \varphi_3(t), t' - t \rangle.$$

The claim follows. □

The following bound is the main tool to prove the convergence rate. It is an analogue of Lemma 2 in Gasnikov and Nesterov (2016) adapted to the case of an inexact oracle.

Lemma 3 *Assume at the k -th iteration of Algorithm 1 we call an inexact oracle $(\tilde{\Phi}, \tilde{\nabla} \Phi)$ satisfying*

$$\tilde{\Phi}(t') + \delta_k \geq \Phi(t') \geq \tilde{\Phi}(t) + \langle \tilde{\nabla} \Phi(t), t' - t \rangle \quad \forall t, t' \in \text{dom } h,$$

with $\delta_k = \frac{\alpha_{k+1}\epsilon}{4A_{k+1}}$. Then for any $k > 0$

$$\Phi(t^k) + h(t^k) \leq \frac{1}{A_k} \phi_k(u^k) + \frac{3\varepsilon}{4} \leq \frac{1}{A_k} \phi_k(t) - \frac{1}{2A_k} \|t - u^k\|_2^2 + \frac{3\varepsilon}{4} \quad \forall t \in \text{dom } h.$$

Moreover,

$$\Phi(t^k) + h(t^k) + \frac{1}{2A_k} \|t^* - u^k\|_2^2 \leq \Phi(t^*) + h(t^*) + \frac{1}{2A_k} \|t^* - t^0\|_2^2,$$

where $t^* = \arg \min \Phi(t) + h(t)$.

Proof We are going to prove by induction that

$$A_k(\Phi(t^k) + h(t^k)) \leq \phi_k(u^k) + A_k \frac{3\varepsilon}{4}.$$

Note that since ϕ_k is 1-strongly convex,

$$\phi_k(u^k) \leq \phi_k(t) - \frac{1}{2} \|t - u^k\|_2^2 \quad \forall t \in \text{dom } h.$$

The inner stopping criterion yields that

$$\begin{aligned} \tilde{\Phi}(t^{k+1}) &\leq \tilde{\Phi}(y^{k+1}) + \langle \tilde{\nabla} \Phi(y^{k+1}), t^{k+1} - y^{k+1} \rangle + \frac{L_{k+1}}{2} \|t^{k+1} - y^{k+1}\|_2^2 + \frac{\alpha_{k+1}}{2A_{k+1}} \varepsilon \\ &= \tilde{\Phi}(y^{k+1}) + \frac{\alpha_{k+1}}{A_{k+1}} \left[\langle \tilde{\nabla} \Phi(y^{k+1}), u^{k+1} - u^k \rangle + \frac{\varepsilon}{2} \right] + \frac{1}{2A_{k+1}} \|u^{k+1} - u^k\|_2^2. \end{aligned}$$

By the assumptions of the lemma,

$$\tilde{\Phi}(y^{k+1}) \leq \Phi(t^k) - \langle \tilde{\nabla} \Phi(y^{k+1}), t^k - y^{k+1} \rangle = \Phi(t^k) + \frac{\alpha_{k+1}}{A_{k+1}} \langle \tilde{\nabla} \Phi(y^{k+1}), u^k - t^k \rangle,$$

thus

$$\begin{aligned} A_{k+1} \tilde{\Phi}(t^{k+1}) &\leq A_{k+1} \tilde{\Phi}(y^{k+1}) + \alpha_{k+1} \left[\langle \tilde{\nabla} \Phi(y^{k+1}), u^{k+1} - u^k \rangle + \frac{\varepsilon}{2} \right] + \frac{1}{2} \|u^{k+1} - u^k\|_2^2 \\ &\leq A_k \Phi(t^k) + \alpha_{k+1} \left[\tilde{\Phi}(y^{k+1}) + \langle \tilde{\nabla} \Phi(y^{k+1}), u^{k+1} - u^k \rangle + \frac{A_k}{A_{k+1}} (u^k - t^k) + \frac{\varepsilon}{2} \right] \\ &\quad + \frac{1}{2} \|u^{k+1} - u^k\|_2^2 \\ &= A_k \Phi(t^k) + \alpha_{k+1} \left[\tilde{\Phi}(y^{k+1}) + \langle \tilde{\nabla} \Phi(y^{k+1}), u^{k+1} - y^{k+1} \rangle + \frac{\varepsilon}{2} \right] + \frac{1}{2} \|u^{k+1} - u^k\|_2^2. \end{aligned}$$

Now, using the convexity of h and the definition of δ_k we obtain that

$$\begin{aligned}
 & A_{k+1} [\Phi(t^{k+1}) + h(t^{k+1})] \\
 & \leq A_{k+1} [\tilde{\Phi}(t^{k+1}) + \delta_k] + A_k h(t^k) + \alpha_{k+1} h(u^{k+1}) \\
 & \leq A_k [\Phi(t^k) + h(t^k)] + \alpha_{k+1} [\tilde{\Phi}(y^{k+1}) + \langle \tilde{\nabla} \Phi(y^{k+1}), u^{k+1} - y^{k+1} \rangle + h(u^{k+1}) + \frac{3\varepsilon}{4}] \\
 & \quad + \frac{1}{2} \|u^{k+1} - u^k\|_2^2 \\
 & \leq \phi_k(u^{k+1}) + \alpha_{k+1} [\tilde{\Phi}(y^{k+1}) + \langle \tilde{\nabla} \Phi(y^{k+1}), u^{k+1} - y^{k+1} \rangle + h(u^{k+1})] + A_{k+1} \frac{3\varepsilon}{4} \\
 & = \phi_{k+1}(u^{k+1}) + A_{k+1} \frac{3\varepsilon}{4}.
 \end{aligned}$$

The last claim of the lemma follows from the inequality

$$\Phi(t^*) \geq \tilde{\Phi}(t) + \langle \tilde{\nabla} \Phi(t), t^* - t \rangle,$$

what implies

$$\frac{1}{A_k} \phi_k(t^*) \leq \Phi(t^*) + h(t^*) + \frac{1}{2A_k} \|t^* - t^0\|_2^2.$$

□

Now we are ready to prove the main result of this section: a primal-dual convergence rate for USTM in the combined model. The complexity analysis in the next theorem is similar to Theorems 3 and 4 in Kubentayeva and Gasnikov (2021), where USTM was applied to the route assignment problem.

Theorem 1 Assume $t^0 = \bar{t}$, $L_0 \leq \frac{M^2}{\varepsilon}$, and at the k -th iteration problem (E) is solved with accuracy $\delta_k = \frac{\alpha_{k+1}\varepsilon}{4A_{k+1}}$. Define

$$R = \|t^* - \bar{t}\|, \quad \tilde{R}^2 = \rho^2 N^{2/\mu} \sum_{e \in E} \frac{\tilde{t}_e^2}{\tilde{f}_e^{2/\mu}}.$$

Then in the case of Beckmann’s model, after at most

$$K = 4 \left(\frac{M\tilde{R}}{\varepsilon} \right)^2$$

iterations it holds that

$$0 \leq P_3(\hat{d}^K, \hat{f}^K) - D'_3(t^K) \leq \varepsilon.$$

In the case of Stable Dynamics model, after at most

$$K = 4 \left(\frac{MR}{\varepsilon} \right)^2$$

iterations it holds that

$$0 \leq P_3(\hat{d}^K, \hat{f}^K) - D'_3(t^K) + \langle (\hat{f}^K - \bar{f})_+, t^* - \bar{t} \rangle \leq \varepsilon,$$

$$\|(\hat{f}^K - \bar{f})_+\|_2 \leq \frac{2\varepsilon}{R}.$$

Proof By Lemma 1,

$$\tilde{\Phi}(t^{k+1}) \leq \tilde{\Phi}(y^{k+1}) + \langle \tilde{\nabla}\Phi(y^{k+1}), t^{k+1} - y^{k+1} \rangle + M\|t^{k+1} - y^{k+1}\|_2.$$

Then f. (A3) from the proof of Theorem 3 in Kubentayeva and Gasnikov (2021) ensures that for all k

$$A_k \geq \frac{\varepsilon k}{2M^2}.$$

Recall that $f = f^k$ are link flows by shortest paths for times $y = y^k$ and interzonal trips $d' = d'(y^k)$. Then, according to Sect. 2.3,

$$E(\tilde{d}(y), T(y)) = \sum_{i,j,r,a,m} d_{ij}^{ram} T_{ij}^m(y) + H(d') = \sum_{e,m} f_e^m(y_e + c_e^m) + H(d'),$$

and for any t

$$\begin{aligned} \tilde{\varphi}_3(y) + \langle \tilde{\nabla}\varphi_3(y), t - y \rangle &= -E(\tilde{d}(y), T(y)) - \langle \nabla_t E(\tilde{d}(y), T(y)), t - y \rangle \\ &= -\sum_{e,m} f_e^m(y_e + c_e^m) - H(d') + \sum_{e,m} f_e^m(y_e - t_e) \\ &= -\sum_{e,m} f_e^m(t_e + c_e^m) - H(d'). \end{aligned}$$

Therefore, due to the convexity of the entropy,

$$\begin{aligned} \phi_K(t) &= \sum_{k=1}^K \alpha_k (\tilde{\varphi}_3(y^k) + \langle \tilde{\nabla}\varphi_3(y^k), t - y^k \rangle + h(t)) + \frac{1}{2} \|t - t^0\|_2^2 \\ &= -A_K \sum_{e,m} [\hat{f}^K]_e^m (t_e + c_e^m) - \sum_{k=1}^K \alpha_k H(d'(y^k)) + A_K h(t) + \frac{1}{2} \|t - \bar{t}\|_2^2 \\ &\leq A_K \sum_e \left(\sigma_e^*(t_e) - [\hat{f}^K]_e t_e - \sum_m [\hat{f}^K]_e^m c_e^m \right) - A_K H(\hat{d}^K) + \frac{1}{2} \|t - \bar{t}\|_2^2. \end{aligned}$$

Then by Lemma 3,

$$-D'_3(t^K) \leq \sum_e \left(\sigma_e^*(t_e) - [\hat{f}^K]_e t_e - \sum_m [\hat{f}^K]_e^m c_e^m \right) - H(\hat{d}^K) + \frac{1}{2A_K} \|t - \bar{t}\|_2^2 + \frac{3\varepsilon}{4}.$$

The rest of the proof repeats the proofs of Lemmata 1 and 2 in Kubentayeva and Gasnikov (2021), mutatis mutandis. In the case of the Beckmann model, we substitute $t_e = \tau_e([\hat{f}^K]_e)$, what gives us the bound

$$\begin{aligned}
 -D'_3(t^K) &\leq -\sum_e \left(\sigma_e([\hat{f}^K]_e) + \sum_m [\hat{f}^K]_e^m c_e^m \right) - H(\hat{d}^K) + \frac{\tilde{R}^2}{2A_K} + \frac{3\epsilon}{4} \\
 &= -\Psi(\hat{f}^K) - H(\hat{d}^K) + \frac{\tilde{R}^2}{2A_K} + \frac{3\epsilon}{4} \leq -P_3(\hat{d}^K, \hat{f}^K) + \frac{(M\tilde{R})^2}{K\epsilon} + \frac{3\epsilon}{4}.
 \end{aligned}$$

In the case of the Stable Dynamics model,

$$\begin{aligned}
 -D'_3(t^K) &\leq \min_{t \geq \bar{t}} \left\{ \sum_e \left(\bar{f}_e(t_e - \bar{t}_e) - [\hat{f}^K]_e t_e - \sum_m [\hat{f}^K]_e^m c_e^m \right) + \frac{1}{2A_K} \|t - \bar{t}\|_2^2 \right\} \\
 &\quad - H(\hat{d}^K) + \frac{3\epsilon}{4} \\
 &= -\Psi(\hat{f}^K) - H(\hat{d}^K) + \min_{t \geq \bar{t}} \left(\langle \bar{f} - \hat{f}^K, t - \bar{t} \rangle + \frac{1}{2A_K} \|t - \bar{t}\|_2^2 \right) + \frac{3\epsilon}{4} \\
 &= -P_3(\hat{d}^K, \hat{f}^K) - \frac{A_K}{2} \|(\hat{f}^K - \bar{f})_+\|_2^2 + \frac{3\epsilon}{4} \\
 &\leq -P_3(\hat{d}^K, \hat{f}^K) - \frac{K\epsilon}{4M^2} \|(\hat{f}^K - \bar{f})_+\|_2^2 + \frac{3\epsilon}{4}.
 \end{aligned}$$

Since optimal $t^* - \bar{t} \geq 0$ are Lagrange multipliers for the problem (P3),

$$P_3(\hat{d}^K, \hat{f}^K) \geq P_3(d^*, f^*) - \langle (\hat{f}^K - \bar{f})_+, t^* - \bar{t} \rangle = D'_3(t^*) - \langle (\hat{f}^K - \bar{f})_+, t^* - \bar{t} \rangle,$$

and thus

$$-\langle (\hat{f}^K - \bar{f})_+, t^* - \bar{t} \rangle \leq P_3(\hat{d}^K, \hat{f}^K) - D'_3(t^K) \leq -\frac{K\epsilon}{4M^2} \|(\hat{f}^K - \bar{f})_+\|_2^2 + \frac{3\epsilon}{4}.$$

Therefore,

$$\frac{K\epsilon}{4M^2} \|(\hat{f}^K - \bar{f})_+\|_2^2 \leq R \|(\hat{f}^K - \bar{f})_+\|_2 + \frac{3\epsilon}{4}$$

and, finally,

$$\|(\hat{f}^K - \bar{f})_+\|_2 \leq \frac{4M^2R}{K\epsilon} + M\sqrt{\frac{3}{K}},$$

what yields the result. □

4 Frank–Wolfe variations and USTM in traffic assignment

Here, we consider several numerical methods for solving a separate problem of searching user equilibrium with inelastic demands. The Frank–Wolfe method and its variations with different line search strategies effectively solve the Beckmann traffic assignment problem, but due to its primal nature it cannot be applied to the stable

dynamics model. Meanwhile, the primal-dual USTM method can be applied to both problems. Further, we conduct the experiments for these methods.

4.1 Frank–Wolfe variations

In the Beckmann model, searching equilibria reduces to minimization of the potential function (B). One of the most popular and effective approaches to solve this problem numerically is the famous Frank–Wolfe method (Frank and Wolfe 1956; Jaggi 2013) as well as its numerous modifications (Fukushima 1984; LeBlanc et al. 1985; Arezki and Van Vliet 1990; Chen et al. 2002). Also, one can apply the primal-dual subgradient methods to optimize the dual problem and then reconstruct a solution to the primal one. However, our research (Kubentayeva and Gasnikov 2021) showed that this approach demands more parameter adjustments to reach the Frank–Wolfe algorithm’s performance with standard step size strategy.

In this paper, we test various step size strategies of Frank–Wolfe method. Namely, we consider some simple decaying step size schedules like standard choice of step size $\gamma_k = \frac{2}{k+1}$ and $\gamma_k = \frac{1}{k}$ leading to the averaging of f^k , and a number of approaches based on a choice of the optimal step size by solving auxiliary one-dimensional problem

$$\gamma_k := \arg \min_{\gamma \in [\gamma_{\min}, \gamma_{\max}]} \Psi((1 - \gamma)f^k + \gamma s^k).$$

The variety of these approaches corresponds to the different one-dimensional optimization methods. We consider the Brent method on a segment $\gamma_k \in [0, 1]$ (Brent 1971) and exponential decreasing of γ_k until Armijo rule is satisfied (Armijo 1966). The last modification considered is the backtracking line-search method developed for specific use in Frank–Wolfe algorithms proposed in Pedregosa et al. (2020).

The Frank–Wolfe method’s theoretical convergence rate for convex objective (with Lipschitz-continuous gradient) is $O(1/\varepsilon)$ (Pedregosa et al. 2020; Jaggi 2013)

Algorithm 2 Frank–Wolfe algorithm

Require: accuracy $\varepsilon > 0$

- 1: $t^0 := \bar{t}$, $f^0 := \arg \min_{s \in \{\Theta x : x \in X\}} \langle t^0, s \rangle$, $k := 0$
 - 2: **repeat**
 - 3: $s^k := \arg \min_{s \in \{\Theta x : x \in X\}} \langle t^k, s \rangle$, $t_e^k := \frac{\partial \Psi(f^k)}{\partial f_e} = \tau_e(f^k)$
 - 4: $\gamma_k := \frac{2}{k+2}$, $f^{k+1} := (1 - \gamma_k)f^k + \gamma_k s^k$
 - 5: $k := k + 1$
 - 6: **until** Stopping criterion is fulfilled
-

4.2 Primal-dual universal similar triangles method

Let us remind that (DualB) and (DualSD) dual problems of Beckmann and stable dynamics traffic assignment models have the same structure:

$$Q(t) = \sum_{ij} d_{ij}T_{ij}(t) - h(t) \longrightarrow \max_{t \geq \bar{t}}.$$

The optimization problems are convex, non-smooth and composite. We apply the USTM method described to minimize the composite objective $Q(t)$. Here, in the Algorithm 1, we set $\Phi(t) := \sum_{i,j} d_{ij}T_{ij}(t)$. As in Sect. 3.1, for both models, flows (primal variables) are reconstructed in the following way:

$$\hat{f}^K = -\frac{1}{A_K} \sum_{k=1}^K \alpha_k \nabla \Phi(y^k), \tag{12}$$

where α_k is a coefficient of the USTM method on iteration k , and $A_K = \sum_{k=1}^K \alpha_k$. Note that any element from the set $\partial \Phi(t)$ has form $\nabla \Phi(t) = -f$, where $f = \Theta x$ is a flow distribution on links induced by $x \in X$ concentrated on the shortest paths for given times t (and vice versa: any such f corresponds to a subgradient of $\Phi(t)$). Hence, weighted \hat{f}^K are also induced by flows on the paths.

For the Beckmann model, we can also use the duality gap to estimate the method's accuracy:

$$\Delta^K = Q(t^K) + \Psi(\hat{f}^K).$$

For the stable dynamics model, flows reconstruction according to (12) keeps feasibility of \hat{f}^K (i.e. they are induced by flows on the paths), but can violate the networks capacity constraints — so the duality gap Δ^K can be negative. To solve the SD traffic assignment problem with inelastic demand, Kubentayeva and Gasnikov (2021) proposed a novel way to reconstruct admissible flows — which also meet capacity constraints – together with a novel computable duality gap, which can be used in a stopping criterion.

The USTM method requires $O(1/\varepsilon^2)$ iterations to obtain an ε -solution of primal and dual problems of Beckmann and SD models (Nesterov 2015; Kubentayeva and Gasnikov 2021).

5 Sinkhorn's variations for trip distribution

Optimization problem of entropy-based trip distribution model of Wilson (1969) coincides with optimal transport (OT) problem with entropy regularizer (Cuturi 2013). To solve the problem, celebrated Sinkhorn's algorithm is used (Sect. 5.1). In Sects. 5.2 and 5.3, we consider accelerated gradient methods adapted for solving OT problems. These methods achieve better theoretical convergence rates

compared to Sinkhorn-like methods in some regimes. Later, in Sect. 6.3, we conduct experiments to compare performances of the mentioned methods.

5.1 Sinkhorn’s algorithm

In this section, for the sake of formulas simplicity, we assume a single agent type and travel mode. Since the problem (8) is separable, without loss of generality, we consider only one trip purpose and suppose $\sum_{i,j} d_{ij} = 1$. Thus, eq. (8) takes form

$$\varphi(\lambda^l, \lambda^w) = \frac{1}{\gamma} \ln \sum_{i,j} \exp \left(-\gamma \left(T_{ij} + \lambda_i^l + \lambda_j^w \right) \right) + \sum_i \lambda_i^l l_i + \sum_j \lambda_j^w w_j. \tag{13}$$

Following Guminov et al. (2021), we perform a change of variables $\mu^l = -\gamma \lambda^l$, $\mu^w = -\gamma \lambda^w$ in (13) and obtain an equivalent formulation

$$\phi(\mu^l, \mu^w) = \frac{1}{\gamma} \left[\ln \left(\mathbf{1}^T d(\mu^l, \mu^w) \mathbf{1} \right) - \langle \mu^l, l \rangle - \langle \mu^w, w \rangle \right] \rightarrow \min_{\mu^l, \mu^w}, \tag{14}$$

where

$$[d(\mu^l, \mu^w)]_{i,j} = \exp \left(\mu_i^l + \mu_j^w - \gamma T_{ij}(t) \right), \tag{15}$$

with the primal-dual coupling

$$d = d(\mu^l, \mu^w) / \mathbf{1}^T d(\mu^l, \mu^w) \mathbf{1}. \tag{16}$$

Similarly to the well-known Sinkhorn algorithm, the objective in (14) can be alternatively minimized (see Algorithm 3).

Algorithm 3 Sinkhorn’s Algorithm (dual objective with the tautological constraint)

```

Require:  $\mu_0^l := \mu_0^w := 0$ 
1:  $k := 0$ 
2: repeat
3:   if  $k \bmod 2 = 0$  then
4:      $\mu_{k+1}^l := \arg \min_{\xi} \phi(\mu_k^l, \xi)$ 
5:      $\mu_{k+1}^w := \mu_k^w$ 
6:   else
7:      $\mu_{k+1}^l := \mu_k^l$ 
8:      $\mu_{k+1}^w := \arg \min_{\xi} \phi(\xi, \mu_k^w)$ 
9:   end if
10:   $k := k + 1$ 
11: until Stopping criterion is fulfilled
    
```

Note that, according to Lemma 9 in Guminov et al. (2021) for the problem (14), partial explicit minimization is possible via the same formulas as for classical entropy-regularized OT problem (Cuturi 2013) without tautological constraint:

$$\psi(\mu^l, \mu^w) = \mathbf{1}^T d(\mu^l, \mu^w) \mathbf{1} - \langle \mu^l, l \rangle - \langle \mu^w, w \rangle \rightarrow \min_{\mu^l, \mu^w}, \tag{17}$$

but the primal-dual coupling formulas are different: (16) for the problem (14) and (15) for the problem (17).

The argminima of (14) should be implemented using numerically stable computation of the logarithm of the sum of exponents (logsumexp trick), but analytically the argminima are given by

$$\ln \mu_{k+1}^l := \ln \mu_k^l + \ln l - \ln (d(\mu_k^l, \mu_k^w) \mathbf{1}), \tag{18}$$

$$\ln \mu_{k+1}^w := \ln \mu_k^w + \ln w - \ln (\mathbf{1}^T d(\mu_k^l, \mu_k^w)), \tag{19}$$

where logarithm is taken element-wisely.

The authors of Gasnikov et al. (2015) pointed out that the objective (14), its gradient

$$\nabla_{\mu^l} \phi(\mu) = \partial \phi(\mu^l, \mu^w) / \partial \mu^l = \frac{1}{\gamma} \left(\frac{d(\mu^l, \mu^w) \mathbf{1}}{\mathbf{1}^T d(\mu^l, \mu^w) \mathbf{1}} - l \right), \tag{20}$$

$$\nabla_{\mu^w} \phi(\mu) = \partial \phi(\mu^l, \mu^w) / \partial \mu^w = \frac{1}{\gamma} \left(\frac{d(\mu^l, \mu^w)^T \mathbf{1}}{\mathbf{1}^T d(\mu^l, \mu^w) \mathbf{1}} - w \right), \tag{21}$$

and eq. (15) are invariant under transformations

$$\mu^l \rightarrow \mu^l + t_{\mu^l} \mathbf{1} \tag{22}$$

$$\mu^w \rightarrow \mu^w + t_{\mu^w} \mathbf{1}, \tag{23}$$

with $t_{\mu^l}, t_{\mu^w} \in \mathbb{R}$. That leads to better numerical stability. In our experiments, we present a variant of Algorithm 3 (labeled as SINKHORN-TAUT-SHIFT) with such invariant transformations, that provide maximum of the dual variables equals 1, and with numerically stable computations of the logarithm of the sum of exponents.

5.2 Accelerated Sinkhorn’s algorithm

Besides the Sinkhorn’s algorithm, accelerated gradient methods are adapted for solving OT problems. These methods achieve better theoretical convergence rates compared to Sinkhorn-like methods in some regimes. To the best available knowledge,

the first such method was proposed in Gasnikov et al. (2016), where the authors proposed non-adaptive Accelerated Gradient Descent (AGD) method for a more general class of entropy-linear programming problems. The algorithmic idea is to run AGD for solving (14) and equip it with some primal updates to guarantee the convergence rate also for the primal problem.

In this subsection, Algorithm 1 and Algorithm 2 (its adaptation for trip distribution problem listed as Algorithm 4) from Guminov et al. (2021) are described. The authors proposed to replace in the classical AGD methods the gradient step with a step of explicit minimization w.r.t. one of the blocks of variables. To formalize the latter, suppose that the vector of dual variables can be divided into m block s.t. $\mu = (\mu_1^T, \dots, \mu_m^T)^T$. So that, notations $\phi(\mu)$ and $\phi(\mu_1, \dots, \mu_m)$ are equivalent. And suppose that it is possible to minimize the dual objective (14) over i -th block holding the others variables fixed:

$$\arg \min_i \phi(\mu) := \arg \min_{\xi} \phi(\mu_1, \dots, \mu_{i-1}, \xi, \mu_{i+1}, \dots, \mu_m). \tag{24}$$

Introduce also a notation for block gradient

$$\nabla_i \phi(\mu) = \frac{\partial \phi(\mu_1, \dots, \mu_{i-1}, \xi, \mu_{i+1}, \dots, \mu_m)}{\partial \xi}. \tag{25}$$

The resulting algorithms m times theoretically slower than its gradient counterpart, where m is the number of blocks of variables used in alternating minimization. But in practice the algorithms work faster (Guminov et al. 2021).

Algorithm 4 AGM-NONPD

-
- 1: Set $A_0 := a_0 := 0, \eta_0 := \zeta_0 := \kappa_0$.
 - 2: $k := 0$
 - 3: **repeat**
 - 4: Set $\beta_k := \arg \min_{\beta \in [0,1]} \phi(\eta_k + \beta(\zeta_k - \eta_k))$
 - 5: Set $\kappa_k := \eta_k + \beta_k(\zeta_k - \eta_k)$ {Extrapolation step}
 - 6: Choose $i = \arg \max_{i \in \{1, \dots, m\}} \|\nabla_i \phi(\kappa_k)\|_2^2$
 - 7: Set $\eta_{k+1} = \arg \min_i \phi(\kappa_k)$
 - 8: Find largest $a_{k+1}, A_{k+1} := A_k + a_k$ from

$$\phi(\kappa_k) - \frac{a_{k+1}^2}{2(A_k + a_{k+1})} \|\nabla \phi(\kappa_k)\|^2 := \phi(\eta_{k+1})$$

- 9: Set $\zeta_{k+1} = \zeta_k - a_{k+1} \nabla \phi(\kappa_k)$ {Update momentum term}
 - 10: $k := k + 1$
 - 11: **until** Stopping criterion is fulfilled
-

In practice variable transformations (22, 23) (with $t_\xi = -\|\xi\|_\infty$, $\xi \in \{\mu^l, \mu^w\}$) performed after steps 5 and 7 of Algorithm 4 and can lead to a better numerical stability when γ is big.

For Algorithm 2 from Guminov et al. (2021) constraints residual $\|((d\mathbf{1} - l)^T, (\mathbf{1}^T d - w)^T)\|_2 = \tilde{O}\left(\frac{1}{k^2}\right)$, but in our experiment it was observed that constraints residual decrease faster for $d = d(\kappa_k)$ (16) than for the theoretically obtained primal variable d using primal-dual property of the algorithm. We present experiments only on the best performing modifications with $d = d(\kappa_k)$ primal variable reconstruction, labeled as NONPD (since it does not utilizes primal-dual property of the algorithms considered in this subsection).

According to (Guminov et al. 2021, Theorem 3) the objective of the form (14) can be minimized with the following rate

$$\phi(\kappa_k) - \phi^* = \tilde{O}\left(\frac{\gamma^{1/2}\|T\|_\infty}{k^2}\right).$$

5.3 MIXED AGM

One more natural modification of Algorithm 4 can be obtained by performing several steps of explicit minimization instead of one. The natural number of steps seems to be equal to the number of blocks m . But the proof of Algorithm 4 utilizes the following property of a step explicit minimization

$$\phi(\eta^{k+1}) \leq \phi(\mu^k) - \frac{1}{2L} \|\nabla_{i^k} \phi(\mu^k)\|_2^2$$

in order to obtain

$$\phi(\eta^{k+1}) \leq \phi(\mu^k) - \frac{1}{2nL} \|\nabla \phi(\mu^k)\|_2^2. \tag{26}$$

The latter is true since $i^k = \arg \max_{i \in \{1, \dots, m\}} \|\nabla_{i^k} \phi(\mu^k)\|_2^2$.

But the inequality (26) can be satisfied if one replaces lines 6 and 7 in Algorithm 4 with the following Algorithm 5.

Algorithm 5 k -th step

```

1:  $\zeta^0 = \mu_k$ .
2: while  $j \leq m$  and  $\|\nabla_{i^k} \phi(\mu^k)\|_2^2 \leq \sum_j \|\nabla \phi(\zeta^j)\|_2^2$  do
3:   Choose  $i^j = \arg \max_{i \in \{1, \dots, m\}} \|\nabla_{i^j} \phi(\zeta_j)\|_2^2$ 
4:   Set  $\zeta^{j+1} = \arg \min_i \phi(\mu^k)$ 
5:    $j = j + 1$ 
6: end while
Ensure:  $\eta^{k+1} = \zeta^j$ 

```

Despite the practical performance, this modification has no theoretical guarantees because $\|\nabla_{\mu^k} \phi(\mu^k)\|_2^2$ can be greater than $\sum_{j=1}^J \|\nabla \phi(\xi^j)\|_2^2$ for any $J > m$.

Moreover, Algorithm 4 is non-increasing. But non-increasing property of the algorithm can be violated (with either the exact minimization given by lines 6 and 7 of Algorithm 4 replaced with Algorithm 5 or not) due to numerical instabilities. Once it happened, Algorithm 4 is stopped. The computations can be proceeded from the last obtained η^k with Sinkhorn’s iterations. In fact, these numerical instabilities break monotonicity of Sinkhorn’s iterations too, but in practice the proceeding of computations with Sinkhorn’s iterations allows to find better minima.

The modification, named MIXED-AGM-NONPD, combines Sinkhorn’s iterations after reaching the stability limit and the exact minimization given by Algorithm 5.

5.4 Reconstruction of correspondence matrix

Finally, let us discuss a reconstruction of a solution to the primal problem (E). Assume we reconstruct a solution d_k of the primal problem (E) by formula (16) with $\mu^l = \mu_k^l, \mu^w = \mu_k^w$. However, since the dual problem is only approximately solved, d^k in general does not satisfy the marginal constraints. So obtain a feasible solution, one can use projection Algorithm 2 from Altschuler et al. (2017). According to (Altschuler et al. 2017, Theorem 4), it has complexity $O(|O| \cdot |D|)$ and returns a correspondence matrix $\hat{d}_k \in \Pi(l, w)$ such that

$$\|d_k - \hat{d}_k\|_1 \leq \delta_k = \|d_k \mathbf{1} - l\|_1 + \|d_k^T \mathbf{1} - w\|_1.$$

The error can be estimated following Theorem 8 in Stonyakin et al. (2019):

$$E(\hat{d}_k, T) \leq \min_{d \in \Pi(l, w)} E(d, T) + 2\delta_k \|T\|_\infty + \frac{4\delta_k}{\gamma} \log \left(\frac{|O| \cdot |D|}{\delta_k} \right). \tag{27}$$

Consider μ_k^l, μ_k^w obtained with Sinkhorn’s algorithm. Using (20) and the convexity of ϕ we get

$$\begin{aligned} \phi(\mu_k^l, \mu_k^w) - \phi^* &\leq \langle \mu_k^l - \mu_*^l, \nabla_{\mu^l} \phi(\mu_k^l, \mu_k^w) \rangle + \langle \mu_k^w - \mu_*^w, \nabla_{\mu^w} \phi(\mu_k^l, \mu_k^w) \rangle \\ &= \frac{1}{\gamma} (\langle \mu_k^l - \mu_*^l, d_k \mathbf{1} - l \rangle + \langle \mu_k^w - \mu_*^w, d_k^T \mathbf{1} - w \rangle), \end{aligned}$$

where (μ_*^l, μ_*^w) is the solution of (15). Then Lemma 3 and Theorem 7 from Stonyakin et al. (2019) ensure that

$$\phi(\mu_k^l, \mu_k^w) - \phi(\mu_*^l, \mu_*^w) \leq \frac{1}{2} \|T\|_\infty (\|d_k \mathbf{1} - l\|_1 + \|d_k^T \mathbf{1} - w\|_1).$$

Combining the above bounds, we obtain the following bound on the duality gap:

$$E(\hat{d}_k, T) + \phi(\mu_k^l, \mu_k^w) \leq \frac{5}{2} \delta_k \|T\|_\infty + \frac{4\delta_k}{\gamma} \log \left(\frac{|O| \cdot |D|}{\delta_k} \right). \tag{28}$$

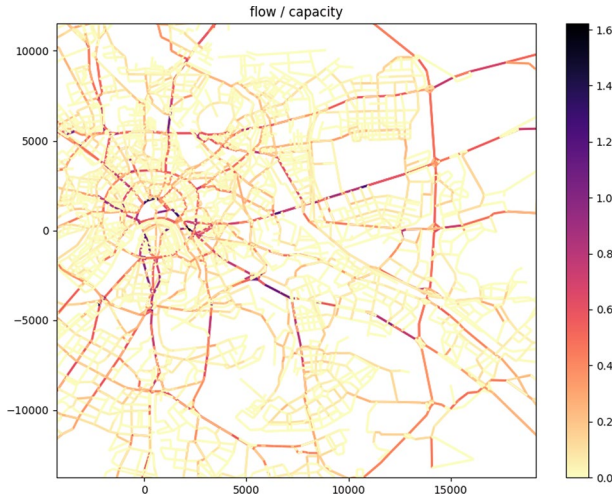


Fig. 1 Moscow network link loads, obtained for the combined Beckmann model

6 Numerical experiments

In our experiments, we consider the morning peak-hour in Moscow transportation network. The city’s data are provided by Russian University of Transport.

The city and its suburbs are split into 1420 zones. Moscow road network consists of 12970 nodes and 36905 links, a part of it is visualized on Fig. 1. We model the crossings by inserting auxiliary links for each allowed turn between road links. Resulting graph contains 63073 nodes and 94546 links.

In our four-stage model of Moscow we consider

- Two demand layers: home-to-work, and home-to-others;
- Two agent types: car owners and non-car-owners;
- And three travel modes: public transport, pedestrian and car.

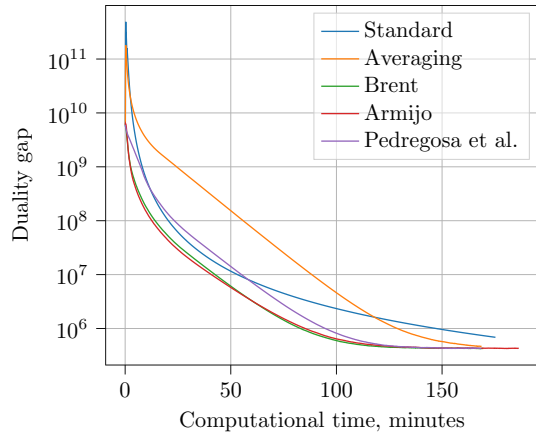
6.1 Parallel computing

Calculation of flows f is the most expensive part, since we have to find the shortest paths for all pairs $w \in OD$. We use Dijkstra’s algorithm (Dijkstra 1959) to find the shortest paths, which runs in $O(|\mathcal{E}| + |\mathcal{V}| \log |\mathcal{V}|)$ time; Given the shortest paths tree, flows aggregation have linear performance $O(|\mathcal{V}|)$. Hence, the total complexity of flows calculation is $O(|O|(|\mathcal{E}| + |\mathcal{V}| \log |\mathcal{V}|))$. Moreover, flows

Table 1 Effect of CPU parallelism

# cores	1	4	8	16	24	32
Total time (sec)	2794	810	470	335	293	274

Fig. 2 Convergence rate for the different Frank–Wolfe modifications for Moscow network



reconstruction for every source $o \in O$ can be computed in parallel. Table 1 shows the result of running 100 iterations of the Frank–Wolfe method on Moscow road network with the various number of cores involved (processor’s speed is 3092,72 MHz).

6.2 Frank–Wolfe algorithm’s variations

Each of the considered modifications of the Frank–Wolfe algorithm was run up to 2000 iterations for the traffic assignment task of the classic four-stage model for the Moscow road network. The results are shown in Fig. 2.

6.3 Sinkhorn algorithm’s variations

Experiments were run for the Trip distribution stage with dual function adjustment for gradient methods described in Sect. 5 for the Moscow road network. The results are shown in Fig. 3.

Different formulations of minimized targets for Sinkhorn’s method were considered (for example, formulation (14) or (17)), but conceptual differences were not identified, therefore only (14) formulation is shown as SINKHORN-TAUT-SHIFT. Label AAM-NONPD corresponds (Guminov et al. 2021, Algorithm 3), that can be easily adapted similarly as Algorithm 4 was adapted from (Guminov et al. 2021, Algorithm 1). One should note that utilized Sinkhorn’s variation has comparable to gradient methods convergence rate, hence common approach is suitable for solving Trip Distribution problem.

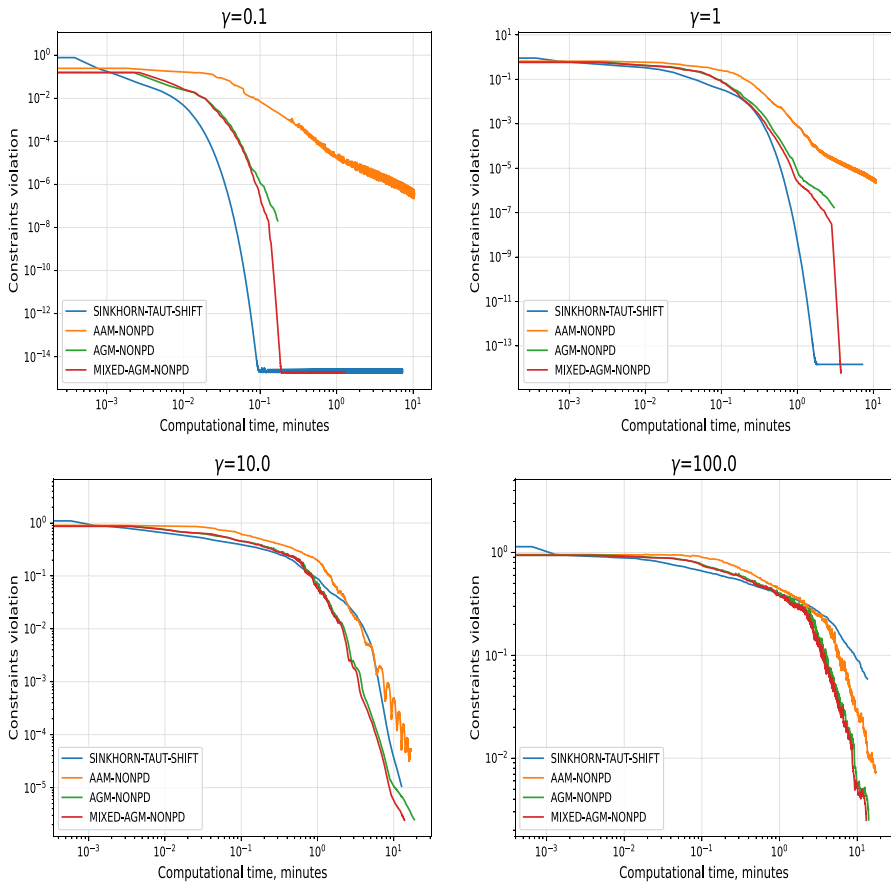
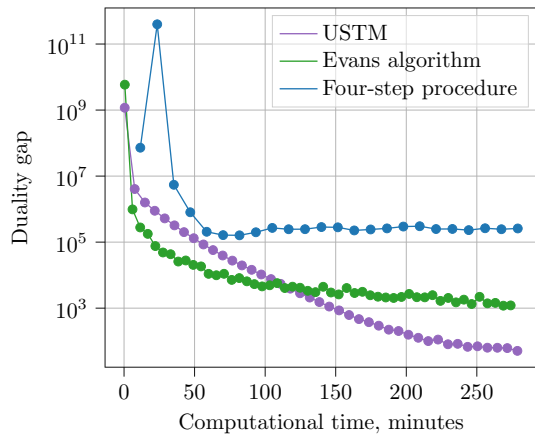


Fig. 3 Sinkhorn’s algorithm modifications for the trip distribution stage

Fig. 4 Duality gap convergence



6.4 Combined model, Beckmann

Here we compare three algorithms for finding a fixed-point of the four-stage Beckmann traffic model, namely Four-stage procedure, Evans algorithm and our dual approach via USTM. The difference with the two-stage model is the addition of mode split and mode cost averaging steps. Mode split step usually cause wobbling between public transport and car modes when using straightforward Four-stage procedure: if the road network is free at the first iteration agents start alternating between these two modes at each iteration. So we applied exponential averaging of modes cost matrices to handle this problem: $T_{ij}^m[k + 1] = \frac{1}{2} (T_{ij}^m[\text{new}] + T_{ij}^m[k])$.

Figure 4 shows the convergence of the duality gap for all three algorithms considered. It can be seen that the Four-stage procedure does not tend to converge to zero duality gap: after 5-6 iterations (about 70 min) it reaches its lower value of the duality gap, then it starts to fluctuate around this value. In order to increase the accuracy of the approximate solution found by Four-stage procedure, one has to increase the number of inner iterations, which will make each outer iteration slower.

In contrast, Evans method steadily converge to zero duality gap.

Some intuition about the behavior of the methods can be given by the Fig. 5, where two-dimensional projections of d_{ij}^m trajectories are depicted. The projections were made by multidimensional scaling method, which tries to preserve pairwise distances while matching points from high-dimensional space (in our case — correspondence matrices) to points on the plane. As one can see, the trajectories start from the same point, since the calculation of the correspondence matrices and the modal splitting in both methods is the same. After a few iterations the trajectories of

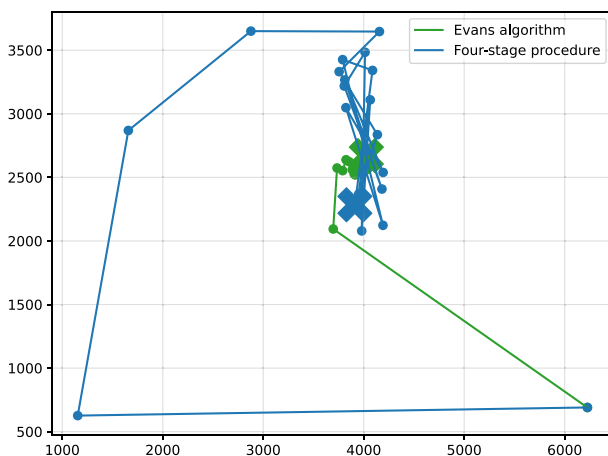


Fig. 5 2-dimensional projections of d_{ij}^m trajectories for the Evans algorithm and the four-stage procedure, obtained by multidimensional scaling. The trajectory of the Evans method is sparsified to 50 points. The last point is marked with a large cross

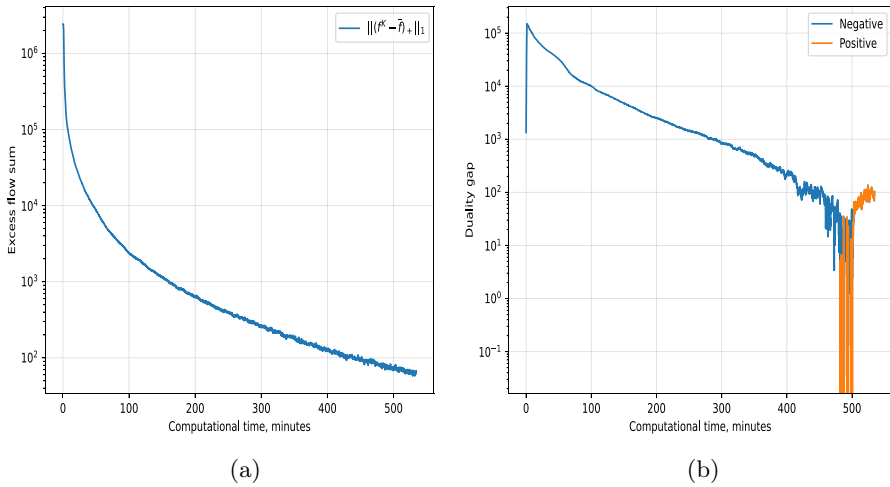


Fig. 6 Convergence of USTM on the Stable Dynamics model: **a** total flow above the link capacity limits, **b** absolute value of the duality gap, the sign is marked by color

the methods are in the same region again, but the Evans method proceeds with small steps, while the Four-stage procedure makes long jumps around the point to which Evans method converges. The trajectory of USTM is similar to the trajectory of the Evans algorithm and is omitted for the sake of readability.

6.5 Combined model, stable dynamics

Here we compare the results obtained for the Beckman and the Stable Dynamics models on the Moscow city transportation model. We use the USTM algorithm to search for the equilibria because other algorithms are not applicable since the link travel times are not functions of the link flows in the Stable Dynamics model.

We used the same Moscow network as in previous experiment, but, since Stable Dynamics model is usually infeasible for peak-hours correspondences, we divided the peak-hour departures l_i^ra and arrivals w_j^r by two.

Convergence trajectories for Stable Dynamics model are shown in Fig. 6. We discuss convergence of Beckmann model in more representative case of peak-hour departures and arrivals in Sect. 6.4, therefore convergence trajectories for Beckmann model are omitted in this subsection.

We assess the convergence by monitoring two values: constraints violation and function suboptimality. Since the dual approach allows the flows to exceed the link capacities, the primal variables stay outside of the feasible region, thus the duality gap could be negative, as shown in Fig. 6b. Then duality gap is negative, the objective function value at that approximate solution (of the minimization problem) is less than the optimal function value, but the approximate solution is infeasible.

The comparison of the approximate solutions is given in Fig. 7. It is evident that Beckmann's model is more likely to exceed the link capacity. Figure 7b shows that

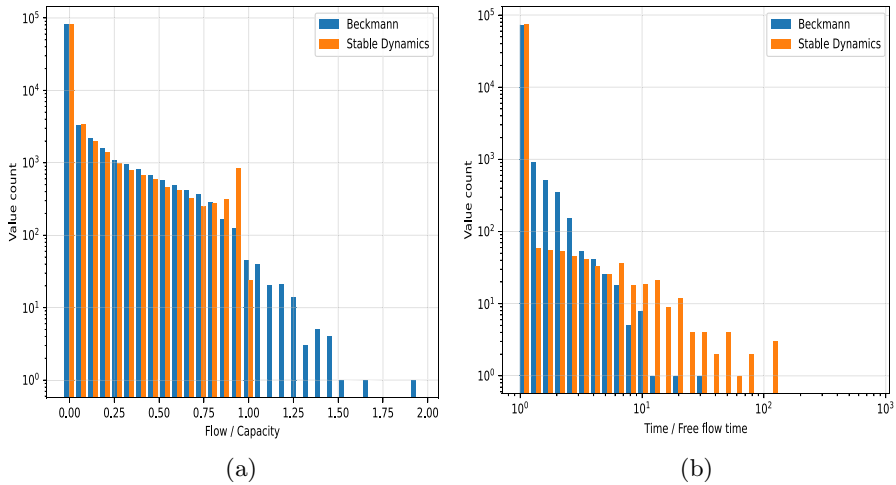


Fig. 7 Histograms of the network load: **a** histogram of the ratio of the amount of flow on the link to its capacity, **b** histogram of the ratio of the travel time on the link to the travel time on the same link when it is free

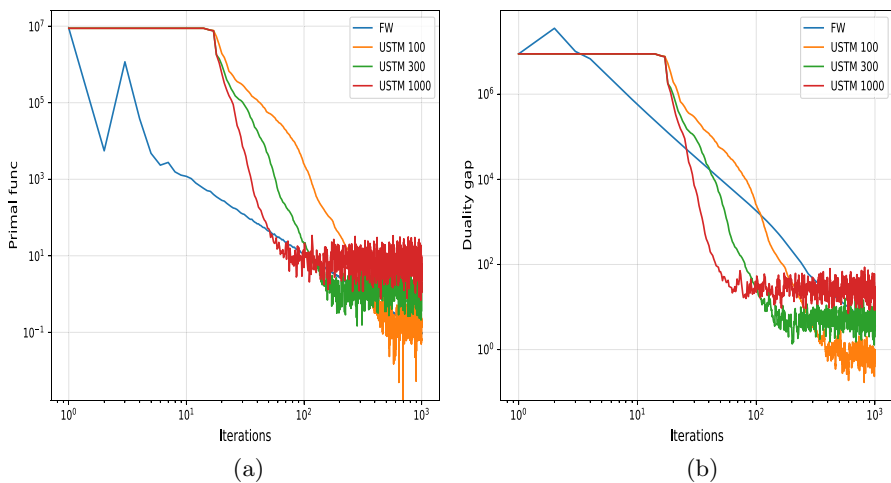


Fig. 8 Convergence of Frank–Wolfe and USTM with different ϵ for Beckmann model for Berlin–Center network: **a** primal function, **b** duality gap

the travel time on some links in Stable Dynamics model exceeds the free-flow time by several hundred times. This implies that some zones are connected to the rest of the network only by low-capacity links, leading to huge traffic congestion at equilibrium. This result is likely due to inaccuracies in the input data, but if not, these bottleneck links should be prioritized in the transportation network improvement process.

6.6 Traffic assignment model: Frank–Wolfe vs USTM for Beckmann model

Experiments were conducted for single trip purpose, agent type and travel mode (by car) for the Berlin–Center network split into 865 zones with 12981 nodes and 28376 links (for more details see Transportation Networks for Research Core Team 2023). As it was shown in the article (Kubentayeva and Gasnikov 2021), performance of the USTM method is better than UGD (Nesterov 2015) and other variations of accelerated gradient descent, thus only USTM and conventional Frank–Wolfe methods are considered. Convergence by primal function and duality gap is presented in Fig. 8. It is necessary to emphasize that the bigger ε , the faster USTM converges to ε accuracy and oscillates. Thereby, it makes sense to use restarting technique for faster convergence — run method with ε' and then with final desired accuracy $\varepsilon < \varepsilon'$.

Acknowledgements This work was supported by a grant for research centers in the field of artificial intelligence, provided by the Analytical Center for the Government of the Russian Federation in accordance with the subsidy agreement (agreement identifier 000000D730321P5Q0002) and the agreement with the Moscow Institute of Physics and Technology dated November 1, 2021 No. 70-2021-00138.

Author Contributions AG conceptualized the research topic. In Sect. 2, AG, VS, MK, DY, and AK made a review of the existing results and formulated optimization problems of the proposed NE model. AK, DY, and MK provided in Sects. 3 and 5.4 convergence analysis for the USTM method for solving the considered NE problem with SD and Beckmann models in the traffic assignment stage. DY, MP, and MK implemented combined models and conducted the experiments of Sects. 6.1, 6.4, and 6.5. DP made a review of numerical methods in Sect. 4 for the traffic assignment problem and conducted the experiments of Sects. 6.2. and 6.6. NT and EK made a review of numerical methods for the trip distribution problem in Sects. 5.1–5.3 and conducted the experiments of the Sect. 6.3 VS, LB, and AS aggregated travel demand and transportation network data of Moscow, designed the network structure, participated in designing the experiments for the Moscow network, and provided computing resources.

Declarations

Conflict of interest The authors declare no competing interests.

References

- Abrahamsson T, Lundqvist L (1999) Formulation and estimation of combined network equilibrium models with applications to stockholm. *Transp Sci* 33(1):80–100
- Altschuler J, Weed J, Rigollet P (2017) Near-linear time approximation algorithms for optimal transport via Sinkhorn iteration. In: Guyon, I., Luxburg, U.V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., Garnett, R. (eds.) *Advances in neural information processing systems* 30, pp. 1961–1971. Curran Associates, Inc., [arXiv:1705.09634](https://arxiv.org/abs/1705.09634). <http://papers.nips.cc/paper/6792-near-linear-time-approximation-algorithms-for-optimal-transport-via-sinkhorn-iteration.pdf>
- Anikin A, Dorn Y, Nesterov Y (2020) Computational methods for the stable dynamic model, pp 280–294. https://doi.org/10.1007/978-3-030-38603-0_21
- Arezki Y, Van Vliet D (1990) A full analytical implementation of the PARTAN/Frank-Wolfe algorithm for equilibrium assignment. *Transp Sci* 24(1):58–62
- Armijo L (1966) Minimization of functions having Lipschitz continuous first partial derivatives. *Pac J Math* 16(1):1–3
- Babazadeh A, Javani B, Gentile G, Florian M (2020) Reduced gradient algorithm for user equilibrium traffic assignment problem. *Transp A: Transp Sci* 16(3):1111–1135. <https://doi.org/10.1080/23249935.2020.1722279>
- Beckmann MJ, McGuire CB, Winsten CB (1956) *Studies in the economics of transportation*. Technical report

- Boyce D (2002) Is the sequential travel forecasting paradigm counterproductive? *J Urban Plan Develop* 128(4):169–183
- Boyce DE, Chon KS, Lee YJ, Lin K, LeBlanc LJ (1983) Implementation and computational issues for combined models of location, destination, mode, and route choice. *Environ Plan A* 15(9):1219–1230
- Boyce DE, Zhang Y-F, Lupa MR (1994) Introducing feedback into four-step travel forecasting procedure versus equilibrium solution of combined model. *Transport Res Record* 1443:65
- Brent RP (1971) An algorithm with guaranteed convergence for finding a zero of a function. *Comput J* 14(4):422–425
- Cabannes T, Glista E, Dwarakanath K, Rao X, Veeravalli T, Bayen AM (2019) Sensitivity analysis and relaxation of the static traffic assignment problem with capacity constraints. In: 2019 IEEE 58th conference on decision and control (CDC), pp 2214–2219. <https://doi.org/10.1109/CDC40024.2019.9030149>
- De Cea J, Fernández JE, Dekock V, Soto A (2005) Solving network equilibrium problems on multimodal urban transportation networks with multiple user classes. *Transp Rev* 25(3):293–317
- Chen A, Jayakrishnan R, Tsai W (2002) Faster Frank-Wolfe traffic assignment with new flow update scheme. *J Transp Eng ASCE*. [https://doi.org/10.1061/\(ASCE\)0733-947X\(2002\)128:1\(31\)](https://doi.org/10.1061/(ASCE)0733-947X(2002)128:1(31))
- Chen X, Liu Z, Zhang K, Wang Z (2020) A parallel computing approach to solve traffic assignment using path-based gradient projection algorithm. *Transp Res Part C: Emerg Technol* 120:102809. <https://doi.org/10.1016/j.trc.2020.102809>
- Chu Y-L (2018) Implementation of a new network equilibrium model of travel choices. *J Traffic Transp Eng (English Edition)* 5(2):105–115. <https://doi.org/10.1016/j.jtte.2017.05.014>
- Chudak FA, Dos Santos Eleuterio V, Nesterov Y (2007) Static traffic assignment problem: a comparison between Beckmann (1956) and Nesterov & de Palma (1998) models. In: 7th swiss transport research conference. ETH
- Cuturi M (2013) Sinkhorn distances: lightspeed computation of optimal transportation distances
- Dijkstra EW et al (1959) A note on two problems in connexion with graphs. *Numer Math* 1(1):269–271
- Dios Ortúzar J, Willumsen LG (2011) *Modelling Transport*, (4th edn). Wiley
- Evans SP (1976) Derivation and analysis of some models for combining trip distribution and assignment. *Transp Res* 10(1):37–57. [https://doi.org/10.1016/0041-1647\(76\)90100-3](https://doi.org/10.1016/0041-1647(76)90100-3)
- Fan Y, Ding J, Liu H, Wang Y, Long J (2022) Large-scale multimodal transportation network models and algorithms-part i: the combined mode split and traffic assignment problem. *Transp Res Part E: Logist Transp Rev* 164:102832
- Fernández E, De Cea J, Florian M, Cabrera E (1994) Network equilibrium models with combined modes. *Transp Sci* 28(3):182–192
- Florian M, Nguyen S (1978) A combined trip distribution modal split and trip assignment model. *Transp Res* 12(4):241–246
- Frank M, Wolfe P (1956) An algorithm for quadratic programming. *Naval Res Logist Q* 3(1–2):95–110
- Fukushima M (1984) A modified Frank-Wolfe algorithm for solving the traffic assignment problem. *Transp Res Part B: Methodol* 18(2):169–177
- Gao J, Wang Y, Zhou J (2022) A study on two-stage selection model of tourism destination at the scale of urban agglomerations. *Arch Transp* 63:143–157. <https://doi.org/10.5604/01.3001.0016.0020>
- Gasnikov AV, Gasnikova EV, Nesterov YE, Chernov AV (2016) Efficient numerical methods for entropy-linear programming problems. *Comput Math Math Phys* 56(4):514–524. <https://doi.org/10.1134/S0965542516040084>
- Gasnikov AV, Nesterov YE (2018) Universal method for stochastic composite optimization problems. *Comput Math Math Phys* 58(1):48–64
- Gasnikov A, Dvurechensky P, Kamzolov D, Nesterov Y, Spokoiny V, Stetsyuk P, Suvorikova A, Chernov A (2015) Universal method with inexact oracle and its applications for searching equilibriums in multistage transport problems. *arXiv preprint arXiv:1506.00292*
- Gasnikov A, Nesterov Y (2016) Universal fast gradient method for stochastic composite optimization problems. *arXiv:1604.05275*
- Guminov S, Dvurechensky P, Tupitsa N, Gasnikov A (2021) On a combination of alternating minimization and Nesterov's momentum. In: International conference on machine learning, pp 3886–3898. PMLR
- Jaggi M (2013) Revisiting frank-wolfe: projection-free sparse convex optimization. In: Proceedings of the 30th international conference on machine learning, pp 427–435
- Kubentayeva M, Gasnikov A (2021) Finding equilibria in the traffic assignment problem with primal-dual gradient methods for stable dynamics model and Beckmann model. *Mathematics* 9(11):1217
- LeBlanc LJ, Helgason RV, Boyce DE (1985) Improved efficiency of the Frank-Wolfe algorithm for convex network programs. *Transp Sci* 19(4):445–462

- Liu Z, Chen X, Meng Q, Kim I (2018) Remote park-and-ride network equilibrium model and its applications. *Transp Res Part B: Methodol* 117:37–62. <https://doi.org/10.1016/j.trb.2018.08.004>
- Nesterov Y (2009) Primal-dual subgradient methods for convex problems. *Math Program* 120(1):221–259. <https://doi.org/10.1007/s10107-007-0149-x>
- Nesterov Y (2015) Universal gradient methods for convex optimization problems. *Math Program* 152(1):381–404. <https://doi.org/10.1007/s10107-014-0790-0>
- Nesterov Y, De Palma A (2003) Stationary dynamic solutions in congested transportation networks: summary and perspectives. *Netw Spat Econ* 3(3):371–395
- Nesterov Y (2004) *Introductory lectures on convex optimization: a basic course*. Kluwer Academic Publishers, Massachusetts
- Oppenheim N et al (1995) *Urban travel demand modeling: from individual choices to general equilibrium*. Wiley, New York
- Patriksson M (2015) *The traffic assignment problem: models and methods*. Courier Dover Publications
- Pedregosa F, Negiar G, Askari A, Jaggi M (2020) Linearly convergent Frank–Wolfe with backtracking line-search. In: international conference on artificial intelligence and statistics. Proceedings of machine learning research
- Sheffi Y (1985) *Urban transportation networks*, vol 6, Prentice-Hall, Englewood Cliffs, NJ
- Sinkhorn R (1974) Diagonal equivalence to matrices with prescribed row and column sums II. *Proc Am Math Soc* 45:195–198. <https://doi.org/10.2307/2040061>
- Smith M, Huang W, Viti F, Tampère CMJ, Lo HK (2019) Quasi-dynamic traffic assignment with spatial queueing, control and blocking back. *Transp Res Part B: Methodol* 122:140–166. <https://doi.org/10.1016/j.trb.2019.01.018>
- Stonyakin FS, Dvinskikh D, Dvurechensky P, Kroshnin A, Kuznetsova O, Agafonov A, Gasnikov A, Tyurin A, Uribe CA, Pasechnyuk D, Artamonov S (2019) Gradient methods for problems with inexact model of the objective. In: Khachay M, Kochetov Y, Pardalos P (eds) *Math Optim Theor Oper Res*. Springer, Cham, pp 97–114
- Transportation Networks for Research Core Team (2023) *Transportation networks for research*. <https://github.com/bstabler/TransportationNetworks>. Accessed: 2023-04-30
- US Bureau of Public Roads (1964) *Traffic assignment manual*. Department of Commerce, urban planning division, Washington D.C
- Wang X, Shahidehpour M, Jiang C, Li Z (2019) Coordinated planning strategy for electric vehicle charging stations and coupled traffic-electric networks. *IEEE Transact Power Syst* 34(1):268–279. <https://doi.org/10.1109/TPWRS.2018.2867176>
- Wang Z, Zhang K, Chen X, Wang M, Liu R, Liu Z (2022) An improved parallel block coordinate descent method for the distributed computing of traffic assignment problem. *Transp A: Transp Sci* 18(3):1376–1400. <https://doi.org/10.1080/23249935.2021.1942303>
- Wilson AG (1969) The use of entropy maximising models, in the theory of trip distribution, mode split and route split. *J Transp Econ Polic* 108–126
- Xie J, Nie Y, Liu X (2017) A greedy path-based algorithm for traffic assignment. *Transp Res Record*. <https://doi.org/10.13140/RG.2.2.18661.09447>
- Yang C, Chen A, Xu X (2013) Improved partial linearization algorithm for solving the combined travel-destination-mode-route choice problem. *J Urban Plan Develop* 139(1):22–32
- Zarrinmehr A, Aashtiani H, Nie Y, Azizian H (2019) Complementarity formulation and solution algorithm for auto-transit assignment problem. *Transp Res Record J Transp Res Board* 2673:036119811983795. <https://doi.org/10.1177/0361198119837956>
- Zhu J-X, Luo Q-Y, Guan X-Y, Yang J-L, Bing X (2020) A traffic assignment approach for multi-modal transportation networks considering capacity constraints and route correlations. *IEEE Access* 8:158862–158874
- Zokaeei Aashtiani H, Poorzahedy H, Nourinejad M (2021) Wardrop's first principle: extension for capacitated networks. *Sci Iran* 28(1):175–191

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.