

Finding Equilibria in the Traffic Assignment Problem with Primal-Dual Gradient Methods for Stable Dynamics Model and Beckmann Model

Meruza Kubentayeva

Institute for Information Transmission Problems RAS, kubentay@gmail.com

Alexander Gasnikov

Moscow Institute of Physics and Technology,
Institute for Information Transmission Problems RAS,
Higher School of Economics, gasnikov@yandex.ru

Abstract

In this paper we consider the application of several gradient methods to the traffic assignment problem: we search equilibria in the stable dynamics model (Nesterov and De Palma, 2003) and the Beckmann model. Unlike the celebrated Frank–Wolfe algorithm widely used for the Beckmann model, these gradients methods solve the dual problem and then reconstruct a solution to the primal one. We deal with the universal gradient method, the universal method of similar triangles, and the method of weighted dual averages, and estimate their complexity for the problem. Due to the primal-dual nature of these methods, we use a duality gap in a stopping criterion. In particular, we present a novel way to reconstruct admissible flows in the stable dynamics model, which provides us with a computable duality gap.

Keywords: stable dynamics model, Beckmann model, traffic equilibrium, universal gradient method, universal method of similar triangles, method of weighted dual averages, duality gap

1 Introduction

The Beckmann model for searching static traffic equilibria in road networks is among the most widely used models by transportation planners (Beckmann et al., 1956; Patriksson, 2015). The equilibria found are practical for evaluating the network efficiency and distribution of business centers and residential areas, and establishing urban development plans, etc. This model introduces a cost function on every link of a transportation network, which defines a dependence of the travel cost on the flow along the link. In practice the BPR functions are usually employed (US Bureau of Public Roads, 1964):

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

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 $\mu = 0.25$.

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

$$\tau_e(f_e) = \begin{cases} \bar{t}_e, & 0 \leq f_e < \bar{f}_e, \\ [\bar{t}_e, \infty], & f_e = \bar{f}_e, \\ +\infty, & f_e > \bar{f}_e. \end{cases} \quad (2)$$

Unlike in the Beckmann model, there is no one-to-one correspondence between equilibrium travel times and flows on the links of the network. We can illustrate the difference on a simple example of two parallel routes (Figure 1).

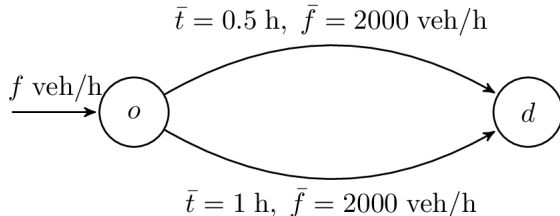


Figure 1: Parallel routes

Let the input flow take values 1000, 2000, and 3000 veh/h. For the stable dynamics model, in the first and second cases, all drivers choose the upper route; the equilibrium travel time simply equals the upper route’s free flow time (0.5 h) in the first case and varies from 0.5 h to 1 h (according to the model) in the second. In the third case, the input flow exceeds the upper route’s capacity, so the upper route’s flow is 2000 veh/h, the lower one’s is 1000 veh/h, and the equilibrium travel time is 1 h. All these equilibria can be interpreted as stationary regimes of some dynamic processes, e.g. the last case can be viewed as the result of the queue at the beginning of the upper route (since this route’s capacity is smaller than the input flow) created by drivers who wanted to take this route until the waiting time plus the route’s travel time reached the lower route’s travel time (Nesterov and De Palma, 2003). In the Beckmann model equilibria are as follows: *for all three cases* only the upper route is used, and the equilibrium travel times are approximately 0.5 h, 0.6 h, and 0.9 h, respectively. Chudak et al. (2007) conducted a detailed comparison — for large and small networks — of equilibria in these two models.

In the Beckmann model, searching equilibria reduces to minimization of a potential function. 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).

In the case of the stable dynamics model, one cannot directly apply the Frank–Wolfe method. However, an equilibrium can be found as a solution of a pair of primal and dual optimization problems. The same holds also for the Beckmann model, so in both cases we can apply primal-dual (sub)gradient methods.

In this work, we compare several primal-dual gradient methods for searching equilibria in both the Beckmann and the stable dynamics models, namely, the universal gradient method (UGM) (Nesterov, 2015), the universal method of similar triangles (UMST) (Gasnikov and Nesterov, 2018), and the method of weighted dual averages (WDA) (Nesterov, 2009). The main advantage of the above universal methods is an automatic adjustment to a local (Hölder) smoothness of a minimized function, which is especially important since the dual problems we are dealing with are essentially non-smooth. Due to the primal-dual nature of these methods, one can use an adaptive stopping criterion guaranteeing required accuracy.

The **main contributions** of this paper include the following:

- We propose a novel way to reconstruct admissible flows (i.e., meeting the capacity constraints and induced by flows on the paths) in the stable dynamics model and a novel computable duality gap, which can be used in a stopping criterion.
- We provide theoretical upper bounds on the complexity of searching equilibria by the considered algorithms: UMST, UGD, and WDA.
- We conducted numerical experiments comparing these algorithms on the Anaheim transportation network—the source code is available for use and can be found in Kubentayeva (2021).

The paper is organized as follows. In Section 2 we give a problem statement, define equilibria in the Beckmann and the stable dynamics models and corresponding optimization problems. Section 3 is devoted to the complexity analysis of UGM, UMST, and WDA. We show that the number of iterations required to obtain an ε -solution of primal and dual problems is $O(1/\varepsilon^2)$ for UGM and

UMST. In Section 4 results of experiments on Anaheim transportation network are presented. Finally, some conclusions are drawn in Section 5.

2 Problem statement

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

$$X = \left\{ x \in \mathbb{R}_+^{|P|} : \sum_{p \in P_w} x_p = d_w, w \in OD \right\}.$$

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:

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

where $\delta_{ep} = \mathbb{1}\{e \in p\}$. Let us introduce a matrix Θ such that $\Theta_{e,p} = \delta_{ep}$ for $e \in E, p \in P$, 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. 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 $t_e = \tau_e(f_e)$. 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 $t^* = (t_e^*)_{e \in E}$, is an equilibrium state, i.e. it satisfies conditions

$$\begin{aligned} t_e^* &= \tau_e(f_e^*), \quad \text{where } f^* = \Theta x^*, \\ x_{p_w}^* > 0 &\implies T_{p_w}(t^*) = T_w(t^*) = \min_{p \in P_w} T_p(t^*), \end{aligned}$$

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

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

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

Another way to find an equilibrium numerically is by solving a dual problem. According to Theo-

rem 4 from [Nesterov and De Palma \(2003\)](#), we can construct it in the following way:

$$\begin{aligned}
\min_{f=\Theta x: x \in X} \Psi(f) &= \min_{x \in X, f} \left[\Psi(f) + \sup_{t \in \mathbb{R}^{|E|}} \langle t, \Theta x - f \rangle \right] = \sup_{t \in \mathbb{R}^{|E|}} \min_{x \in X, f} [\Psi(f) + \langle t, \Theta x - f \rangle] \\
&= \sup_{t \in \mathbb{R}^{|E|}} \left[- \sum_{e \in E} \max_{f_e} \{t_e f_e - \sigma_e(f_e)\} + \min_{x \in X} \sum_p \sum_{e \in E} t_e \delta_{ep} x_p \right] \\
&= \max_{t \in \text{dom } \sigma^*} - \left[\sum_{e \in E} \sigma_e^*(t_e) - \sum_{w \in OD} d_w T_w(t) \right] = - \min_{t \geq \bar{t}} Q(t),
\end{aligned}$$

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 $\sigma_e(f_e)$, $e \in E$. Finally, we obtain the dual problem, the solution of which is t^* :

$$Q(t) = - \sum_{w \in OD} d_w T_w(t) + \sum_{e \in E} \sigma_e^*(t_e) \longrightarrow \min_{t \geq \bar{t}}. \quad (\text{DualB})$$

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 : $f \in \partial \sum_{w \in OD} d_w T_w(t)$ (see Subsection 3.1). 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](#)). An equilibrium state (x^*, t^*) of the stable dynamics model satisfies the next conditions:

$$\begin{aligned}
t_e^* &\in \tau_e(f_e^*), \\
x_{pw}^* > 0 &\implies T_{pw}(t^*) = T_w(t^*),
\end{aligned}$$

where $\tau(f)$ is defined earlier by (2). The above formula can be reformulated in terms of an optimization problem:

$$\begin{aligned}
x^* &= \arg \min_{x \in X} \sum_{w \in OD} \sum_{p \in P_w} x_p T_p(t^*) \\
&= \arg \min_{x \in X} \sum_{e \in E} t_e^* f_e(x) \\
&= \arg \min_{x \in X} \sum_{e \in E} [t_e^* f_e(x) - \underline{(t_e^* - \bar{t}_e) \bar{f}_e}],
\end{aligned}$$

$$\begin{aligned}
t_e^* \in \tau_e(f_e^*) &\iff t_e^* = \arg \max_{t_e \geq \bar{t}_e} t_e (f_e^* - \bar{f}_e) \\
&= \arg \max_{t_e \geq \bar{t}_e} [t_e (f_e^* - \bar{f}_e) + \underline{\bar{t}_e \bar{f}_e}].
\end{aligned}$$

Here, we add underlined constant terms to show that the pair (f^*, t^*) is an equilibrium if and only if it is a solution of the saddle-point problem

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

where its primal problem is

$$\begin{aligned}
\Psi(x) &= \sup_{t_e \geq \bar{t}_e} \sum_{e \in E} [t_e f_e - (t_e - \bar{t}_e) \bar{f}_e] = \sum_{e \in E} \bar{t}_e f_e + \sum_{e \in E} \sup_{t_e \geq \bar{t}_e} (t_e - \bar{t}_e) (f_e - \bar{f}_e) \longrightarrow \min_{f=\Theta x: x \in X} \\
&\iff \Psi(f) = \sum_{e \in E} f_e \bar{t}_e \longrightarrow \min_{\substack{f=\Theta x: \\ x \in X, f_e \leq \bar{f}_e}} , \quad (\text{SD})
\end{aligned}$$

and its dual problem is

$$\begin{aligned} Q(t) &= - \inf_{f=\Theta x: x \in X} \sum_{e \in E} [t_e f_e - (t_e - \bar{t}_e) \bar{f}_e] \\ &= - \sum_{w \in OD} d_w T_w(t) + \langle t - \bar{t}, \bar{f} \rangle \longrightarrow \min_{t_e \geq \bar{t}_e}. \end{aligned} \quad (\text{DualSD})$$

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).

3 Numerical methods

We have the following objective functions

- The stable dynamics model:

$$Q(t) = - \underbrace{\sum_{w \in OD} d_w T_w(t)}_{\Phi(t)} + \underbrace{\langle t - \bar{t}, \bar{f} \rangle}_{h(t)},$$

- The Beckmann model:

$$Q(t) = - \underbrace{\sum_{w \in OD} d_w T_w(t)}_{\Phi(t)} + \underbrace{\sum_{e \in E} \bar{f}_e \left(\frac{t_e - \bar{t}_e}{\bar{t}_e \rho} \right)^\mu \frac{(t_e - \bar{t}_e)}{1 + \mu}}_{h(t)}.$$

In both cases it has form

$$Q(t) = \Phi(t) + h(t) \longrightarrow \min_{t \geq \bar{t}}. \quad (3)$$

The optimization problem (3) is convex, non-smooth and composite. We use all these properties to identify the best optimization method to solve the considered problem.

3.1 Subgradient

In our research, we consider first-order methods, i.e. they require a subgradient of $\Phi(t)$, the properties and effective computation of which we discuss in this section.

To get the subdifferential $\partial \Phi(t)$ let us re-write $\Phi(t)$ in the following way:

$$\Phi(t) = - \sum_{w \in OD} d_w T_w(t) = - \sum_{w \in OD} d_w \min_{p \in P_w} \langle t, a_p \rangle,$$

where the vector $a_p = (\delta_{ep})_{e \in E}$ encodes a path p . Obviously, the shortest path may not be unique. Using the rules of subgradient calculus (Rockafellar, 2015) we get the following expression:

$$\partial \Phi(t) = - \sum_{w \in OD} d_w \partial \left(\min_{p \in P_w} \langle t, a_p \rangle \right) = - \sum_{w \in OD} d_w \text{Conv} \{ a_p : p \in P_w, T_p(t) = T_w(t) \},$$

i.e. the subdifferential $\partial \Phi(t)$ is a sum of convex hulls of binary vectors that encode the shortest length paths. An important consequence is that for any $t_1, t_2 \in \mathbb{R}_+^{|E|}$ and $\nabla \Phi(t_1) \in \partial \Phi(t_1)$, $\nabla \Phi(t_2) \in \partial \Phi(t_2)$ the following bound holds:

$$\|\nabla \Phi(t_1) - \nabla \Phi(t_2)\|_2 \leq M = \sqrt{2H} \sum_{w \in OD} d_w, \quad (4)$$

where H is the diameter of the graph G .

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 $\Phi(t)$).

In practice, the 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 the following algorithm 1. We use Dijkstra's algorithm (Dijkstra et al., 1959) to find the shortest paths in line 3, which runs in $O(|E| + |V| \log |V|)$ time; finding the traversal order with topological sort (Section 22.4 in Cormen et al. (2009)) and further flows aggregation have linear performance $O(|V|)$. Hence, the total complexity of Algorithm 1 is $O(|O|(|E| + |V| \log |V|))$. When the transportation network is an (almost) planar graph or another sparse graph, $|E| = O(|V|)$ and the complexity is $O(|O| \cdot |V| \log |V|)$. Moreover, flows reconstruction for every source $o \in O$ can be computed in parallel. And Dijkstra's algorithm also can be parallelized and has efficient implementations (Crauser et al., 1998; Chao and Hongxia, 2010).

Algorithm 1 Flows reconstruction

Input: times t

```

1:  $f := \mathbf{0}_{|E|}$  {flows on edges}
2: for origin  $o$  in  $O$  do
3:   Get a shortest-path tree  $\mathcal{T}_o$  from  $o$  to all destinations in  $D$  with weights  $t$ 
4:    $\text{traversal\_order} := \text{TopologicalSort}(\mathcal{T}_o)$  {sorting from furthest to closest vertices}
5:    $f_{\text{out}} := \mathbf{0}_{|V|}$  {total output flow from each vertex}
6:    $f_{\text{out}}[v] := d_w$  for  $w = (o, v) \in OD$ 
7:   for  $v$  in  $\text{traversal\_order}$  do
8:     Get predecessor  $p$  of  $v$  in  $\mathcal{T}_o$ 
9:      $e := (p, v)$ 
10:     $f[e] := f[e] + f_{\text{out}}[v]$ 
11:     $f_{\text{out}}[p] := f_{\text{out}}[p] + f_{\text{out}}[v]$ 
12:   end for
13: end for
14: return flows  $f$ 

```

3.2 Reconstruction of admissible flows in SD model

For given times t considered Algorithm 1 reconstructs feasible flows f , i.e. $f = \Theta x$ for some $x \in X$. These flows meet all the constraints in the Beckmann model, but they can violate the capacity constraints in the stable dynamics model. In the latter case, an additional step is required to obtain admissible flows from f . Note that we could instead find flows that meet capacity constraints first (Theorem 8 from Nesterov and De Palma (2003)), but to reconstruct feasible flows from them is a more complex problem.

Suppose we are given some flows $g = \Theta x$ such that

$$\xi = 1 - \max_{e \in E} g_e / \bar{f}_e > 0. \quad (5)$$

Then for any $f = \Theta x$ we can construct admissible flows $\pi(f)$ in the following way: let $\eta = \max_{e \in E} f_e / \bar{f}_e - 1$, then

$$\pi(f) = \begin{cases} f, & \eta \leq 0, \\ \frac{\xi f + \eta g}{\xi + \eta}, & \eta > 0. \end{cases}$$

In practice, we propose the following procedure to find admissible flows g : run some optimization method (e.g. UGM) for a small number of iterations for the same problem but with decreased capacities: $\frac{1}{2}\bar{f}$ instead of \bar{f} ; if obtained feasible flows \hat{f}^N satisfy $\hat{f}^N \leq \frac{3}{4}\bar{f}$, then take $g = \hat{f}^N$; otherwise, run it again with capacities $\frac{3}{4}\bar{f}$ and check $\hat{f}^N \leq \frac{7}{8}\bar{f}$, etc.

Stopping criterion. The stopping criterion we use for the stable dynamics model is based on a duality gap

$$Q(\hat{t}^N) + \Psi(\pi(\hat{f}^N)) \leq \varepsilon, \quad (6)$$

where $\hat{f}^N \in \{\Theta x : x \in X\}$, $\hat{t}^N \geq \bar{t}$ are estimates of an equilibrium (f^*, t^*) after N iterations of the applied method. Note that here the duality gap with \hat{f}^N is not applicable.

3.3 Universal gradient method

The method for solving non-smooth problems with smooth techniques was proposed by Nesterov (2015) and was called *the universal gradient method*. The pseudocode of UGM for the considered problem (3) is provided in Algorithm 2. Here the euclidean prox-structure is used. Note that we did not specify the stopping criterion as it can be different for different models.

Algorithm 2 Universal gradient method

Input: $L_0 > 0$, accuracy $\varepsilon > 0$

- 1: Set $t^0 := \bar{t}$, $k := 0$
 - 2: **repeat**
 - 3: $L_{k+1} := L_k/2$
 - 4: **while true do**
 - 5: $t^{k+1} := \arg \min_{t \in \text{dom } h} \langle \nabla \Phi(t^k), t - t^k \rangle + h(t) + L_{k+1} \frac{\|t - t^k\|_2^2}{2}$
 - 6: **if** $\Phi(t^{k+1}) \leq \Phi(t^k) + \langle \nabla \Phi(t^k), t^{k+1} - t^k \rangle + L_{k+1} \frac{\|t^{k+1} - t^k\|_2^2}{2} + \frac{\varepsilon}{2}$ **then**
 - 7: **break**
 - 8: **else**
 - 9: $L_{k+1} := 2L_{k+1}$
 - 10: **end if**
 - 11: **end while**
 - 12: $k := k + 1$
 - 13: **until** Stopping criterion is fulfilled
-

Now let us define

$$\hat{f}^N = -\frac{1}{S_N} \sum_{k=0}^{N-1} \frac{\nabla \Phi(t^k)}{L_{k+1}}, \quad \hat{t}^N = \frac{1}{S_N} \sum_{k=1}^N \frac{t^k}{L_k}, \quad S_N = \sum_{k=1}^N \frac{1}{L_k}, \quad (7)$$

where L_k are the estimates of the local Lipschitz constant in UGM and UMST methods.

Convergence of the UGM was proved in Nesterov (2015) and is summarized in the following lemma and theorem.

Lemma 3.1. *After N iterations of UGM for the stable dynamics model it holds that*

$$Q(\hat{t}^N) - Q(t^*) \leq \frac{R^2}{S_N} + \frac{\varepsilon}{2}, \quad (8)$$

$$0 \leq Q(\hat{t}^N) + \Psi(\hat{f}^N) + \langle t^* - \bar{t}, (\hat{f}^N - \bar{f})_+ \rangle \leq \frac{R^2}{S_N} + \frac{\varepsilon}{2}, \quad (9)$$

$$\|(\hat{f}^N - \bar{f})_+\|_2 \leq \frac{4R}{S_N} + \sqrt{\frac{2\varepsilon}{S_N}}, \quad (10)$$

where \hat{f}^N , \hat{t}^N , and S_N are defined by (7), and $R = \|t^* - \bar{t}\|_2$ is the distance from the starting point to a solution.

Theorem 3.2. *Let $L_0 \leq \frac{M^2}{\varepsilon}$, where M comes from (4). Then after at most*

$$N_Q = 2 \left(\frac{RM}{\varepsilon} \right)^2 \quad (11)$$

iterations of UGM for the stable dynamics model it holds that $Q(\hat{t}^N) - Q(t^) \leq \varepsilon$. Moreover, the stopping criterion (6) is fulfilled after at most*

$$N_{stop} = O \left(\left(\frac{RM}{\varepsilon} \right)^2 \max \left\{ 1, \left(\frac{\langle g - f^*, \bar{t} \rangle}{\xi R \min_e \bar{f}_e} \right)^2 \right\} \right) \quad (12)$$

iterations, where ξ comes from (5).

Now we provide results on the rate of convergence for the Beckmann model. The stopping criterion in this case is the following:

$$Q(\hat{t}^N) + \Psi(\hat{f}^N) \leq \varepsilon. \quad (13)$$

Lemma 3.3. *After N iterations of UGM for the Beckmann model it holds that*

$$\begin{aligned} Q(\hat{t}^N) - Q(t^*) &\leq \frac{R^2}{S_N} + \frac{\varepsilon}{2}, \\ 0 \leq Q(\hat{t}^N) + \Psi(\hat{f}^N) &\leq \frac{\|\tau(\hat{f}^N) - \bar{t}\|_2^2}{S_N} + \frac{\varepsilon}{2}, \end{aligned}$$

where \hat{f}^N , \hat{t}^N , S_N are defined by (7), and $R = \|t^* - \bar{t}\|_2$

Theorem 3.4. *Let $L_0 \leq \frac{M^2}{\varepsilon}$, where M comes from (4). Then after at most*

$$N_Q = 2 \left(\frac{RM}{\varepsilon} \right)^2 \quad (14)$$

iterations of UGM for the Beckmann model it holds that $Q(\hat{t}^N) - Q(t^*) \leq \varepsilon$. Moreover, the stopping criterion (13) is fulfilled after at most

$$N_{stop} = 2 \left(\frac{\tilde{R}M}{\varepsilon} \right)^2 \quad (15)$$

iterations, where

$$\tilde{R}^2 = \rho^2 \sum_{e \in E} \frac{\bar{t}_e^2}{\bar{f}_e^{2/\mu}} \left(\sum_{w \in OD} d_w \right)^{2/\mu}. \quad (16)$$

3.4 Universal Method of Similar Triangles

Let us introduce the following notations:

$$\phi_0(t) = \frac{1}{2} \|t - t^0\|_2^2,$$

$$\phi_{k+1}(t) = \phi_k(t) + \alpha_{k+1} [\Phi(y^{k+1}) + \langle \nabla \Phi(y^{k+1}), t - y^{k+1} \rangle + h(t)].$$

Flows are reconstructed in the following way:

$$\hat{f}^N = -\frac{1}{A_N} \sum_{k=1}^N \alpha_k \nabla \Phi(y^k) \quad (17)$$

Lemma 3.5. *After N iterations of UMST for the stable dynamics model it holds that*

$$Q(t^N) - Q(t^*) \leq \frac{R^2}{A_N} + \frac{\varepsilon}{2}, \quad (18)$$

$$0 \leq Q(t^N) + \Psi(\hat{f}^N) + \langle t^* - \bar{t}, (\hat{f}^N - \bar{f})_+ \rangle \leq \frac{R^2}{A_N} + \frac{\varepsilon}{2}, \quad (19)$$

$$\|(\hat{f}^N - \bar{f})_+\|_2 \leq \frac{4R}{A_N} + \sqrt{\frac{2\varepsilon}{A_N}}, \quad (20)$$

where \hat{f}^N is defined by (17) and $R = \|t^* - \bar{t}\|_2$ is the distance from the starting point to a solution.

Algorithm 3 Universal Method of Similar Triangles

Input: $L_0 > 0$, accuracy $\varepsilon > 0$

```

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

```

Theorem 3.6. Let $L_0 \leq \frac{4M^2}{\varepsilon}$, where M comes from (4). Then after at most

$$N_Q = 4 \left(\frac{RM}{\varepsilon} \right)^2 \quad (21)$$

iterations of UGM for the stable dynamics model it holds that $Q(t^N) - Q(t^*) \leq \varepsilon$. Moreover, the stopping criterion (6) with $\hat{t}^N = t^N$ is fulfilled after at most

$$N_{stop} = O \left(\left(\frac{RM}{\varepsilon} \right)^2 \max \left\{ 1, \left(\frac{(g - f^*, \bar{t})}{\xi R \min_e f_e} \right)^2 \right\} \right) \quad (22)$$

iterations, where ξ comes from (5).

Theorem 3.7. Let $L_0 \leq \frac{4M^2}{\varepsilon}$, where M comes from (4). Then after at most

$$N_Q = 4 \left(\frac{RM}{\varepsilon} \right)^2 \quad (23)$$

iterations of UMST for the Beckmann model it holds that $Q(t^N) - Q(t^*) \leq \varepsilon$. Moreover, the stopping criterion (13) with $\hat{t}^N = t^N$ is fulfilled after at most

$$N_{stop} = 4 \left(\frac{\tilde{R}M}{\varepsilon} \right)^2 \quad (24)$$

iterations, where \tilde{R} is defined by (16).

3.5 Method of Weighted Dual Averages

Convergence of WDA-method was proved in Nesterov (2009) and is summarized in the following theorem.

Theorem 3.8. Non-composite WDA-method satisfies the following bounds

- for the stable dynamics model:

$$Q(\hat{t}^k) - Q(t^*) = O \left(\frac{M + \|\bar{f}\|_2}{\sqrt{k}} \left(\frac{R^2}{\chi} + \chi \right) \right),$$

- for the Beckmann model if $\mu \leq 1$:

$$Q(\hat{t}^k) - Q(t^*) = O \left(\frac{1}{\sqrt{k}} \left(M + \max_e \bar{f}_e \left[\frac{2R + \chi}{\bar{t}_e \rho} \right]^\mu \right) \left(\frac{R^2}{\chi} + \chi \right) \right).$$

Algorithm 4 Method of Weighted Dual Averages

Input: accuracy $\varepsilon > 0$, constant $\chi > 0$

1: $s^0 := \bar{0}$, $t^0 := \bar{t}$, $k := 0$

2: **repeat**

3: Compute subgradient g^k , set $s^{k+1} := s^k + \frac{1}{\|g^k\|_2} g^k$

- non-composite case: $g^k := \nabla\Phi(t^k) + \nabla h(t^k)$
- composite case: $g^k := \nabla\Phi(t^k)$

4: Set $\beta_{k+1} := \frac{\hat{\beta}_{k+1}}{\chi}$, where $\hat{\beta}_{k+1} = \sum_{i=0}^k \frac{1}{\beta_i}$, $\hat{\beta}_0 = 1$

5: Set t^{k+1}

- non-composite case: $t^{k+1} := \arg \min_{t \in \text{dom } h} \langle s^{k+1}, t \rangle + \frac{\beta_{k+1}}{2} \|t - t^0\|_2^2$
- composite case: $t^{k+1} := \arg \min_{t \in \text{dom } h} \langle s^{k+1}, t \rangle + \frac{\beta_{k+1}}{2} \|t - t^0\|_2^2 + \sum_{i=0}^k \frac{1}{\|g^i\|_2} h(t)$

6: $k := k + 1$

7: **until** Stopping criterion is fulfilled

4 Numerical experiments

This section presents numerical results for the algorithms described above, namely, composite variants of UMST and UGM, both composite and non-composite WDA-method, on the Anaheim network (Transportation Networks for Research Core Team, 2021; Chudak et al., 2007). The network consists of 38 zones, 416 nodes, and 916 links. Experiments and the source code in Python 3 (Van Rossum and Drake, 2009) can be found in Kubentayeva (2021). We used Dijkstra’s algorithm for finding the shortest paths in the network from the graph-tool library (Peixoto, 2014), where it is implemented in C++. We also used the Numpy library (Harris et al., 2020) for all vector operations.

Stable dynamics model. Parameters of the network are adjusted to the Beckmann model, so we have to increase the capacities to ensure the existence of an equilibrium for the stable dynamics model. In our experiments, the capacities are multiplied by 2.5. In Figure 2, we plot the number of (inner) iterations of the algorithms required to fulfill the stopping criterion (6) against $1/\varepsilon$. We consider the number of inner iterations for Alg. 3 and Alg. 2 since the complexity of an inner iteration in this case is similar to the complexity of an iteration of the other algorithms. Note that according to (Nesterov, 2015, formula (2.23)) the number $N(k)$ of inner iterations of UGM or UMST at step k is bounded as

$$N(k) \leq 2k + \log_2 \left(\frac{M^2}{\varepsilon L_0} \right),$$

so asymptotic rates from Theorems 3.2, 3.4, 3.6, and 3.7 are still valid.

As we can see, the best results are shown by UMST, followed by UGM having similar performance. Both composite and non-composite WDA-method are much slower.

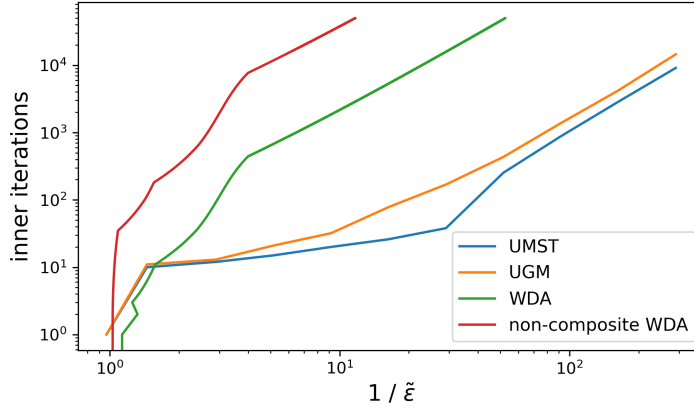


Figure 2: Convergence rates of UMST, UGM, composite and non-composite WDA-methods for the stable dynamics model with the stopping criterion (6). Here $\tilde{\epsilon}$ is the relative accuracy ε/Δ_0 , where Δ_0 is the duality gap at the start point.

Beckmann model. For the Beckmann model we also compare our methods with the Frank–Wolfe algorithm (Alg. 5) — which theoretical convergence rate for convex objective (with Lipschitz-continuous gradient) is $O(1/\varepsilon)$ (Pedregosa et al., 2020; Jaggi, 2013).

Algorithm 5 Frank-Wolfe algorithm

Input: 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
-

Figure 3 shows the convergence rates of the methods for the Beckmann model. The Frank–Wolfe method demonstrates the best results and is followed by UMST. Unlike the stable dynamics case, composite WDA-method is faster than UGM. However, the non-composite WDA-method has the worst performance again.

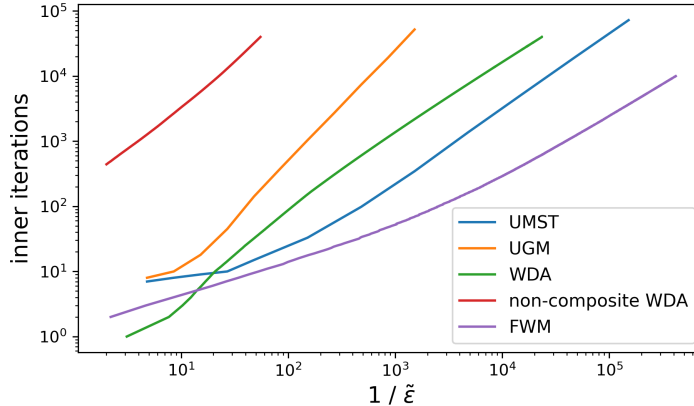


Figure 3: Convergence rates of UMST, UGM, composite and non-composite WDA-methods, and the Frank–Wolfe method for the Beckmann model with the stopping criterion (13). Here $\tilde{\epsilon}$ is the relative accuracy ϵ/Δ_0 , where Δ_0 is the duality gap at the start point.

5 Conclusion

We considered several primal-dual subgradient methods for finding equilibria in the stable dynamics and the Beckmann models. We suggested a way to reconstruct admissible flows in the stable dynamics model, which provides us with a novel computable duality gap. Complexity bounds for UMST and UGM were presented in terms of the iterations number required to achieve a desired accuracy in the dual function value or the duality gap. Finally, we conducted numerical experiments comparing convergence of the considered algorithms on the Anaheim transportation network: UMST is the best one for optimization of the dual problems in both models. Also, using the duality gap as a stopping criterion, we compared these methods with the Frank–Wolfe algorithm for the Beckmann model — which, as expected, remains the most suitable approach in this case (but it is not applicable for the stable dynamics model).

The reader may be interested in another related topic, searching stochastic traffic equilibria. In Gasnikov and Kubentayeva (2018); Baimurzina et al. (2019) we with our colleagues studied the application of the UMST for finding Nash–Wardrop stochastic equilibria in the Beckmann model. In this case, a driver selects a route randomly according to Gibbs’ distribution taking into account current time costs on the links of the network. It leads to iteration complexity $O(\frac{1}{\sqrt{\gamma\epsilon}})$, where $\gamma > 0$ is a stochasticity parameter (when $\gamma \rightarrow 0$ the model boils down to the ordinary Beckmann model). However, the great decrease in the number of iterations comes along with a more expensive calculation of the objective function’s gradient.

Acknowledgements

We would like to thanks Yu. Nesterov for fruitful discussions.

The research of M. Kubentayeva was supported by Russian Science Foundation (project 18-71-10108).

The research of A. Gasnikov was partially supported by RFBR, project number 18-29-03071 mk, and was partially supported by the Ministry of Science and Higher Education of the Russian Federation (Goszadaniye) no 075-00337-20-03.

References

- Arezki, Y. and Van Vliet, D. (1990). A full analytical implementation of the PARTAN/Frank–Wolfe algorithm for equilibrium assignment, *Transportation Science* **24**(1): 58–62.

- Baimurzina, D. R., Gasnikov, A. V., Gasnikova, E. V., Dvurechensky, P. E., Ershov, E. I., Kubentaeva, M. B. and Lagunovskaya, A. A. (2019). Universal method of searching for equilibria and stochastic equilibria in transportation networks, *Computational Mathematics and Mathematical Physics* **59**(1): 19–33.
- Beckmann, M. J., McGuire, C. B. and Winsten, C. B. (1956). Studies in the economics of transportation, *Technical report*.
- Chao, Y. and Hongxia, W. (2010). Developed Dijkstra shortest path search algorithm and simulation, *2010 International Conference on Computer Design and Applications*, Vol. 1, IEEE, pp. V1–116.
- Chen, A., Jayakrishnan, R. and Tsai, W. (2002). Faster Frank–Wolfe traffic assignment with new flow update scheme, *Journal of Transportation Engineering-asce - J TRANSP ENG-ASCE* **128**.
- Chudak, F. A., Dos Santos Eleuterio, V. and Nesterov, Y. (2007). Static traffic assignment problem: A comparison between Beckmann (1956) and Nesterov & de Palma (1998) models, *7th Swiss Transport Research Conference*, ETH.
- Cormen, T. H., Leiserson, C. E., Rivest, R. L. and Stein, C. (2009). *Introduction to algorithms, 3rd-edition*, MIT press.
- Crauser, A., Mehlhorn, K., Meyer, U. and Sanders, P. (1998). A parallelization of Dijkstra’s shortest path algorithm, *International Symposium on Mathematical Foundations of Computer Science*, Springer, pp. 722–731.
- Dijkstra, E. W. et al. (1959). A note on two problems in connexion with graphs, *Numerische mathematik* **1**(1): 269–271.
- Frank, M. and Wolfe, P. (1956). An algorithm for quadratic programming, *Naval research logistics quarterly* **3**(1-2): 95–110.
- Fukushima, M. (1984). A modified Frank–Wolfe algorithm for solving the traffic assignment problem, *Transportation Research Part B: Methodological* **18**(2): 169–177.
- Gasnikov, A. V. and Kubentayeva, M. B. (2018). Searching stochastic equilibria in transport networks by universal primal-dual gradient method, *Computer research and modeling* **10**(3): 335–345.
- Gasnikov, A. V. and Nesterov, Y. E. (2018). Universal method for stochastic composite optimization problems, *Computational Mathematics and Mathematical Physics* **58**(1): 48–64.
- Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M. H., Brett, M., Haldane, A., del Río, J. F., Wiebe, M., Peterson, P., Gérard-Marchant, P., Sheppard, K., Reddy, T., Weckesser, W., Abbasi, H., Gohlke, C. and Oliphant, T. E. (2020). Array programming with NumPy, *Nature* **585**(7825): 357–362.
URL: <https://doi.org/10.1038/s41586-020-2649-2>
- Jaggi, M. (2013). Revisiting Frank–Wolfe: Projection-free sparse convex optimization., *Proceedings of the 30th international conference on machine learning*, pp. 427–435.
- Kubentayeva, M. (2021). TransportNet, <https://github.com/Meruzakub/TransportNet>. Accessed: 2021-04-30.
- LeBlanc, L. J., Helgason, R. V. and Boyce, D. E. (1985). Improved efficiency of the Frank–Wolfe algorithm for convex network programs, *Transportation Science* **19**(4): 445–462.
- Nesterov, Y. (2009). Primal-dual subgradient methods for convex problems, *Mathematical programming* **120**(1): 221–259.
- Nesterov, Y. (2015). Universal gradient methods for convex optimization problems, *Mathematical Programming* **152**(1-2): 381–404.

- Nesterov, Y. and De Palma, A. (2003). Stationary dynamic solutions in congested transportation networks: summary and perspectives, *Networks and Spatial Economics* **3**(3): 371–395.
- Patriksson, M. (2015). *The traffic assignment problem: models and methods*, Courier Dover Publications.
- Pedregosa, F., Negiar, G., Askari, A. and Jaggi, M. (2020). Linearly convergent frank-wolfe with backtracking line-search, *Proceedings of the 23rd International Conference on Artificial Intelligence and Statistics*.
URL: <https://arxiv.org/pdf/1806.05123.pdf>
- Peixoto, T. P. (2014). The graph-tool python library, *figshare* .
URL: http://figshare.com/articles/graph_tool/1164194
- Rockafellar, R. T. (2015). *Convex analysis*, Princeton university press.
- Sheffi, Y. (1985). *Urban transportation networks*, Vol. 6, Prentice-Hall, Englewood Cliffs, NJ.
- Transportation Networks for Research Core Team (2021). Transportation networks for research, <https://github.com/bstabler/TransportationNetworks>. Accessed: 2021-04-30.
- US Bureau of Public Roads (1964). *Traffic Assignment Manual*, Department of Commerce, Urban Planning Division, Washington D.C.
- Van Rossum, G. and Drake, F. L. (2009). *Python 3 Reference Manual*, CreateSpace, Scotts Valley, CA.

6 Appendix

6.1 Proofs for UGM

Proof of Lemma 3.1. Note that function $\Phi(t)$ satisfies (4). Then according to Theorem 1 in Nesterov (2015) applied with $\nu = 0$ one has

$$\begin{aligned} Q(\hat{t}^N) &\leq \frac{1}{S_N} \sum_{k=1}^N \frac{1}{L_k} Q(t^k) \\ &\leq \min_{t \geq \bar{t}} \left\{ \frac{1}{S_N} \sum_{k=0}^{N-1} \frac{1}{L_{k+1}} [\Phi(t^k) + \langle \nabla \Phi(t^k), t - t^k \rangle] + h(t) + \frac{\|t - t^0\|_2^2}{S_N} \right\} + \frac{\varepsilon}{2}. \end{aligned} \quad (25)$$

Equation (8) follows immediately if one substitutes $t = t^*$. Now let us estimate the first term on the r.h.s.

$$\begin{aligned} &\min_{t \geq \bar{t}} \left\{ \frac{1}{S_N} \sum_{k=0}^{N-1} \frac{1}{L_{k+1}} [\Phi(t^k) + \langle \nabla \Phi(t^k), t - t^k \rangle] + h(t) + \frac{\|t - \bar{t}\|_2^2}{S_N} \right\} \\ &= \min_{t \geq \bar{t}} \left\{ \frac{1}{S_N} \sum_{k=0}^{N-1} \frac{1}{L_{k+1}} \underbrace{[\Phi(t^k) + \langle \nabla \Phi(t^k), 0 - t^k \rangle]}_{\leq \Phi(0)} - \langle \hat{f}^N, t \rangle + \langle \bar{f}, t - \bar{t} \rangle + \frac{\|t - \bar{t}\|_2^2}{S_N} \right\} \\ &\leq \Phi(0) - \langle \hat{f}^N, \bar{t} \rangle + \min_{t \geq \bar{t}} \left\{ \langle \bar{f} - \hat{f}^N, t - \bar{t} \rangle + \frac{\|t - \bar{t}\|_2^2}{S_N} \right\} \\ &= -\Psi(\hat{f}^N) - \frac{S_N \|\hat{f}^N - \bar{f}\|_2^2}{4}. \end{aligned}$$

Here we used that $\Phi(0) = -\sum_{w \in OD} d_w T_w(0) = 0$. Therefore,

$$Q(\hat{t}^N) + \Psi(\hat{f}^N) + \frac{S_N \|\hat{f}^N - \bar{f}\|_2^2}{4} \leq \frac{\varepsilon}{2}.$$

Now notice that, since the flow \hat{f}^N is induced by some traffic distribution $x \in X$, we have

$$\begin{aligned}
0 &\leq \Phi(t^*) + \langle t^*, \hat{f}^N \rangle \\
&= Q(t^*) - \langle t^* - \bar{t}, \bar{f} \rangle + \Psi(\hat{f}^N) - \langle \bar{t}, \hat{f}^N \rangle + \langle t^*, \hat{f}^N \rangle \\
&= Q(t^*) + \Psi(\hat{f}^N) + \langle t^* - \bar{t}, \hat{f}^N - \bar{f} \rangle \\
&\leq Q(\hat{t}^N) + \Psi(\hat{f}^N) + \langle t^* - \bar{t}, (\hat{f}^N - \bar{f})_+ \rangle,
\end{aligned}$$

hence

$$Q(\hat{t}^N) + \Psi(\hat{f}^N) \geq -\langle t^* - \bar{t}, (\hat{f}^N - \bar{f})_+ \rangle \geq -R\|(\hat{f}^N - \bar{f})_+\|_2.$$

This yields

$$\frac{S_N\|(\hat{f}^N - \bar{f})_+\|_2^2}{4} - R\|(\hat{f}^N - \bar{f})_+\|_2 \leq \frac{\varepsilon}{2},$$

and thus

$$\|(\hat{f}^N - \bar{f})_+\|_2 \leq \frac{2R}{S_N} \left(1 + \sqrt{1 + \frac{\varepsilon S_N}{2R^2}} \right) \leq \frac{4R}{S_N} + \sqrt{\frac{2\varepsilon}{S_N}}.$$

□

Proof of Theorem 3.2. Theorem 1 in [Nesterov \(2015\)](#) ensures that $L_k \leq \frac{M^2}{\varepsilon}$ for all $k \geq 0$, thus $S_N \geq \frac{\varepsilon N}{M^2}$. Then the first bound (11) follows immediately from (8).

Now let us prove the second bound. First, suppose $\hat{f}_e^N \leq \bar{f}_e$ for all $e \in E$. Then $\pi(\hat{f}^N) = \hat{f}^N$, thus by (9) for $N = N_Q$

$$Q(\hat{t}^N) + \Psi(\pi(\hat{f}^N)) = Q(\hat{t}^N) + \Psi(\hat{f}^N) \leq \frac{R^2}{S_N} + \frac{\varepsilon}{2} \leq \frac{(RM)^2}{\varepsilon N} + \frac{\varepsilon}{2} \leq \varepsilon.$$

Otherwise, if $\hat{f}_e^N \not\leq \bar{f}_e$, one has $\pi(\hat{f}^N) = \frac{\xi \hat{f}^N + \eta g}{\xi + \eta}$, where $\eta = \max_{e \in E} \hat{f}_e^N / \bar{f}_e - 1$, hence (8) and (9) yield

$$\begin{aligned}
Q(\hat{t}^N) + \Psi(\pi(\hat{f}^N)) &\leq \frac{\xi}{\xi + \eta} \left(Q(\hat{t}^N) + \Psi(\hat{f}^N) \right) + \frac{\eta}{\xi + \eta} \left(Q(\hat{t}^N) + \Psi(g) \right) \\
&= \frac{\xi}{\xi + \eta} \left(Q(\hat{t}^N) + \Psi(\hat{f}^N) \right) + \frac{\eta}{\xi + \eta} \left(Q(\hat{t}^N) - Q(t^*) \right) + \frac{\eta}{\xi + \eta} \left(\Psi(g) - \Psi(f^*) \right) \\
&\leq \frac{R^2}{S_N} + \frac{\varepsilon}{2} + \frac{\eta}{\xi} \langle g - f^*, \bar{t} \rangle.
\end{aligned}$$

Finally, according to (10)

$$\eta = \max_{e \in E} \hat{f}_e^N / \bar{f}_e - 1 = \left\| \frac{(\hat{f}^N - \bar{f})_+}{\bar{f}} \right\|_{\infty} \leq \frac{1}{\min_e \bar{f}_e} \|(\hat{f}^N - \bar{f})_+\|_2 \leq \frac{1}{\min_e \bar{f}_e} \left(\frac{4R}{S_N} + \sqrt{\frac{2\varepsilon}{S_N}} \right).$$

Combining all bounds together we obtain

$$Q(\hat{t}^N) + \Psi(\pi(\hat{f}^N)) \leq \frac{R^2 M^2}{\varepsilon N} + \frac{\langle g - f^*, \bar{t} \rangle}{\xi \min_e \bar{f}_e} \left(\frac{4RM^2}{\varepsilon N} + \sqrt{\frac{2M^2}{N}} \right) + \frac{\varepsilon}{2},$$

and substituting $N = N_{stop}$, we conclude that the stopping criterion (6) is fulfilled. □

Proof of Lemma 3.3. First of all, note that

$$\max_{t \geq \bar{t}} \left\{ \langle \hat{f}^N, t \rangle - \sum_{e \in E} \sigma_e^*(t_e) \right\} = \sum_{e \in E} \sigma_e(\hat{f}_e^N) = \Psi(\hat{f}^N),$$

and maximum is attained at point $t = \nabla\Psi(\hat{f}^N) = \tau(\hat{f}^N)$. As in the proof of Theorem 3.2, the inequality (25) holds in Beckmann's model case. Then the first term in the r.h.s. can be estimated as follows

$$\begin{aligned}
& \min_{t \geq \bar{t}} \left\{ \frac{1}{S_N} \sum_{k=0}^{N-1} \frac{1}{L_{k+1}} [\Phi(t^k) + \langle \nabla\Phi(t^k), t - t^k \rangle] + h(t) + \frac{\|t - \bar{t}\|_2^2}{S_N} \right\} \\
&= \min_{t \geq \bar{t}} \left\{ \frac{1}{S_N} \sum_{k=0}^{N-1} \frac{1}{L_{k+1}} \underbrace{[\Phi(t^k) + \langle \nabla\Phi(t^k), 0 - t^k \rangle]}_{\leq \Phi(0)} - \langle \hat{f}^N, t \rangle + \sum_{e \in E} \sigma_e^*(t_e) + \frac{\|t - \bar{t}\|_2^2}{S_N} \right\} \\
&\leq \Phi(0) + \left\{ \sum_{e \in E} \sigma_e^*(t_e(\hat{f}_e^N)) - \langle \hat{f}^N, \tau(\hat{f}^N) \rangle + \frac{\|\tau(\hat{f}^N) - \bar{t}\|_2^2}{S_N} \right\} \\
&= -\Psi(\hat{f}^N) + \frac{1}{S_N} \|\tau(\hat{f}^N) - \bar{t}\|_2^2,
\end{aligned}$$

and we finally get an upper bound on the duality gap:

$$0 \leq Q(\hat{t}^N) + \Psi(\hat{f}^N) \leq \frac{\|\tau(\hat{f}^N) - \bar{t}\|_2^2}{S_N} + \frac{\varepsilon}{2}.$$

In the same time, substituting $t = t^*$ one obtains

$$Q(\hat{t}^N) \leq Q(t^*) + \frac{\|t^* - \bar{t}\|_2^2}{S_N} + \frac{\varepsilon}{2}.$$

□

Proof of Theorem 3.4. By construction, $\hat{t}_e^N \leq \sum_{w \in OD} d_w$ for all $e \in E$, thus $\|\tau(\hat{f}^N) - \bar{t}\|_2 \leq \tilde{R}$. According to Theorem 1 in Nesterov (2015) $S_N \geq \frac{\varepsilon^N}{M^2}$, thus the statement follows immediately from Lemma 3.3. □

6.2 Proofs for UMST

Proof of Lemma 3.5. According to the inequality (30) in Gasnikov and Nesterov (2018)

$$Q(t^N) \leq \min_{t \geq \bar{t}} \left\{ \frac{1}{A_N} \sum_{k=1}^N \alpha_k [\Phi(y^k) + \langle \nabla\Phi(y^k), t - y^k \rangle] + h(t) + \frac{\|t - t^0\|_2^2}{2A_N} \right\} + \frac{\varepsilon}{2}. \quad (26)$$

Note that the above inequality has the same form as (25), if one replaces S_N with A_N , $\frac{1}{L_{k+1}}$ with α_k , y^k with t^k , and $\frac{\|t - t^0\|_2^2}{S_N}$ with $\frac{\|t - t^0\|_2^2}{2A_N}$. Then the claim follows by the same reasoning as in the proof of Lemma 3.1. □

Proof of Theorem 3.6. Due to (4) one has

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

From Young's inequality we get that

$$M \|t^{k+1} - y^{k+1}\|_2 \leq \frac{\alpha_{k+1}}{2A_{k+1}} \varepsilon + \frac{A_{k+1}M^2}{2\alpha_{k+1}\varepsilon} \|t^{k+1} - y^{k+1}\|_2^2.$$

If $L_{k+1} \geq \frac{A_{k+1}M^2}{\alpha_{k+1}\varepsilon}$, then the stopping condition for inner iterations is fulfilled. Therefore, at the end of the k -th iteration either $L_{k+1} < \frac{2A_{k+1}M^2}{\alpha_{k+1}\varepsilon}$ or $L_{k+1} = \frac{L_k}{2}$.

Now we are going to prove by induction that $\alpha_k \geq \frac{\varepsilon}{2M^2}$, which is equivalent to $L_k \leq \frac{2M^2}{\varepsilon} + \frac{4M^4}{\varepsilon^2} A_k$, for all $k \geq 1$. For $k = 1$ it follows from $A_1 = \alpha_1$ and $L_0 \leq \frac{4M^2}{\varepsilon}$. In case where $L_{k+1} < \frac{2A_{k+1}M^2}{\alpha_{k+1}\varepsilon}$

equation $A_{k+1} = L_{k+1}\alpha_{k+1}^2$ immediately yields $\alpha_{k+1} \geq \frac{\varepsilon}{2M^2}$. If $L_{k+1} = \frac{L_k}{2}$, then by the induction hypothesis and monotonicity of the sequence $\{A_k\}_{k \in \mathbb{N}}$ we obtain

$$L_{k+1} \leq \frac{M^2}{\varepsilon} + \frac{2M^4}{\varepsilon^2} A_{k-1} < \frac{2M^2}{\varepsilon} + \frac{4M^4}{\varepsilon^2} A_k.$$

Therefore,

$$A_N \geq \frac{\varepsilon N}{2M^2}. \quad (27)$$

Arguing in the same way as in the proof of Theorem 3.2 we obtain that

$$Q(t^N) - Q(t^*) \leq \frac{2R^2 M^2}{\varepsilon N} + \frac{\varepsilon}{2}$$

and

$$Q(\hat{t}^N) + \Psi(\pi(\hat{f}^N)) \leq \frac{2R^2 M^2}{\varepsilon N} + \frac{\langle g - f^*, \bar{t} \rangle}{\xi \min_e \bar{f}_e} \left(\frac{8RM^2}{\varepsilon N} + \sqrt{\frac{4M^2}{N}} \right) + \frac{\varepsilon}{2}.$$

After substitution $N = N_Q$ or $N = N_{stop}$ the claim follows. \square

Proof of Theorem 3.7. Repeating the proof of Theorem 3.4 we obtain that

$$Q(t^N) - Q(t^*) \leq \frac{R^2}{A_N} + \frac{\varepsilon}{2}, \quad Q(t^N) + \Psi(\hat{f}^N) \leq \frac{\tilde{R}^2}{A_N} + \frac{\varepsilon}{2}.$$

Then we conclude applying (27). \square

6.3 Proof of Theorem 3.8 (WDA)

According to eq. (3.5) from Nesterov (2009),

$$Q(\hat{t}^k) - Q(t^*) = O\left(\frac{L}{\sqrt{k}} \left(\frac{R^2}{\chi} + \chi\right)\right),$$

whenever $\|g_k\|_2 \leq L$ for all k .

In case of the stable dynamics model $\nabla h(t) = \bar{f}$, thus we can take $L = M + \|\bar{f}\|_2$.

For the Beckmann model

$$\frac{\partial h(t)}{\partial t_e} = \bar{f}_e \left(\frac{t_e - \bar{t}_e}{\bar{t}_e \rho} \right)^\mu.$$

Theorem 3 in Nesterov (2009) yields that $\|t^k - t^*\|_2^2 \leq R^2 + \chi^2$ for all k , thus $\|t^k - \bar{t}\|_2 \leq 2R + \chi$. Then using $\mu \leq 1$ one obtains

$$\|\nabla h(t^k)\|_2 \leq \|t^k - \bar{t}\|_2^\mu \max_e \frac{\bar{f}_e}{(\bar{t}_e \rho)^\mu} \leq (2R + \chi)^\mu \max_e \frac{\bar{f}_e}{(\bar{t}_e \rho)^\mu},$$

thus we can take

$$L = M + \max_e \bar{f}_e \left(\frac{2R + \chi}{\bar{t}_e \rho} \right)^\mu.$$