On Distributed Optimal Control of Traffic Flows in Transportation Networks

Christian Rosdahl[1], and Gustav Nilsson[1], Giacomo Como[1,2]

Abstract— We propose and analyze distributed computation algorithms for finite-horizon optimal control problems in transportation networks. We model traffic flow dynamics by the cell-transmission model and focus on two problems: system-optimum dynamic traffic assignment (where the routing is part of the optimization) and freeway network control (where the routing is exogenous and the optimization is confined to speed limits and ramp-metering controls). While these are non-convex problems, we focus on some recently proposed provably exact convex relaxations and apply Alternating Direction Method of Multipliers techniques. We present fully distributed iterative algorithms and implement them on some transportation network testbeds, testing their convergence speed and accuracy.

I. Introduction

With the pervasive diffusion of interconnected GPS-based devices and novel intelligent traffic control actuators, and with connected and autonomous vehicles around the corner, it is now possible to both obtain traffic state data and provide real-time route guidance to drivers. Ideally, this data, together with traffic control and route guidance possibilities, should be used to reduce congestion, in order to decrease both travel times for the users and pollution. This has renewed the interest of the control systems community for the analysis and synthesis of transportation networks.

Two classical optimal traffic flow control problems are the *Dynamic Traffic Assignment (DTA)* and the *Freeway Network Control (FNC)* problems. The former, originally introduced in [2], [3], has been widely studied by the transportation research community [4]. In its *system-optimum* version (as opposed to the user-optimum framework), it entails the minimization of a global cost of the whole network assuming that one has the ability to control the drivers' route choices (something that might be achievable, e.g., through route guidance with GPS unit, or proper incentives and pricing). In the latter, see, e.g., [5], [6], [7], the traffic flow in the network is controlled by variable speed limits on the freeways and ramp-metering to achieve a system optimal flow in the network.

In this paper, we follow the approach in [8] and consider formulations of the DTA and FNC problems in network flow dynamics modeled by the *Cell Transmission Model (CTM)* [9], [10] and their exact convex relaxations generalizing [6] and [11]. Our main contributions are distributed optimization algorithms solving such convexified DTA and FNC problems, based on the Alternating Direction Method of Multipliers (ADMM) method [12]. The proposed algorithms are fully distributed, in the sense that each road segment only needs information from its neighboring segments to compute the optimal speed limit, ramp-metering signal or route choices for the outgoing vehicles. Moreover, our method is scaleable, so when the controlled network grows, only segments connected to the newly built area need to update their algorithms. We then present numerical implementations of these algorithms for both the DTA and the FNC problem.

Another distributed algorithm for solving the considered DTA problem is proposed in [13]. Except for that both methods are inspired by ADMM, the approaches are quite different. In [13], the constraints on the decision variables are not included in the Lagrangian, and must thus instead be taken into account when solving the distributed subproblems. This allows for dividing the the decision variables into two groups and performing the Lagrangian minimization in two steps. In contrast, our method includes the constraints in the Lagrangian. By performing the Lagrangian minimization in more steps, this still allows the algorithm to be distributed, while each subproblem can be solved easily and efficiently.

The paper is organized as follows. In the rest of this section we will introduce some basic notation. In Section II we present the dynamical model we use for the transportation network together with the optimization problems we want to solve and their convex relaxations. In Section III, the distributed algorithm is presented, and simulations of it for one DTA problem and one FNC problem are shown in Section IV. The paper is concluded in Section V.

We let $\mathbb{R}_{(+)}$ denote the (non-negative) reals. For a set $\mathcal{A}$, we denote the vector indexed by $\mathcal{A}$ as $\mathbb{R}_{(+)}^{\mathcal{A}}$. We let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ denote a directed multigraph, where $\mathcal{V}$ is the set of nodes and $\mathcal{E}$ the set of directed links. For a link $i = (n, m) \in \mathcal{E}$, we denote its head $\omega_i = m \in \mathcal{V}$ and its tail $\tau_i = n \in \mathcal{V}$.

II. Problem Formulation

In this section, we first introduce the controlled traffic flow dynamics model as a discrete-time control system that can be thought of as a generalized version of Daganzo's CTM. We then formulate the FNC and DTA problems as finite-horizon optimal control problems for this model. Finally,

we present exact convex relaxations of these problems and prove their tightness, i.e., that every optimal solution of the relaxed problem can be mapped into an optimal solution of the original problem. Both the presented setup and tightness results are to be considered the discrete-time analogue of [8].

We model the transportation network topology as a directed multigraph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ is the set of nodes and $\mathcal{E}$ is the set of links. Every link $i \in \mathcal{E}$ is directed from its tail node $\tau_i \in \mathcal{V}$ to its head node $\omega_i \in \mathcal{V} \setminus \{\tau_i\}$. Notice that we allow for the possibility of parallel links, i.e., links with the same tail node and head node, hence the prefix in 'multigraph', but we do not allow for selfloops, i.e., links whose head node coincides with its tail node. Each link $i \in \mathcal{E}$ represents a *cell*, i.e., a portion of road. One particular node $w \in \mathcal{V}$ represents the external world, with cells i such that $\tau_i = w$ representing onramps and cells i such that $\omega_i = w$ representing offramps. We shall denote by

$$\mathcal{R} = \{i \in \mathcal{E} : \tau_i = w\}, \qquad \mathcal{S} = \{i \in \mathcal{E} : \omega_i = w\},$$

the sets of onramps and offramps, respectively. Throughout, we shall assume that every link $i \in \mathcal{E}$ lies on a cycle in $\mathcal{G}$ that passes through node w. This assumption amounts to saying that every cell is reachable from at least one onramp and that from every cell at least one offramp can be reached. The network topology is typically illustrated by omitting the external world node w and letting sources have no tail node and sinks have no head node. We shall denote the set of adjacent pairs of cells by $\mathcal{L} = \{(i, j) \in \mathcal{E}^2 : \tau_j = \omega_i \neq w\}$, so that direct flow from a cell i to another cell j is possible only if $(i, j) \in \mathcal{L}$. We let the exogenous inflow to an onramp $i \in \mathcal{R}$ be denoted by $\lambda_i \geq 0$ and let $\mu_i \geq 0$ denote the outflow from an offramp $i \in \mathcal{S}$ towards the external world. Conventionally, we shall set $\lambda_i = 0$ for every non-onramp cell $i \in \mathcal{E} \setminus \mathcal{R}$, and $\mu_i = 0$ for all non-offramp cells $i \in \mathcal{E} \setminus \mathcal{S}$.

We consider a controlled traffic flow dynamics model describing the evolution of the traffic volume among the different cells. The traffic volume in every cell $i \in \mathcal{E}$ is denoted by the variable $x_i \geq 0$, while we use the notation $u_i \in [0, 1]$ to denote a local control variable. Every cell $i \in \mathcal{E}$ is equipped with a *supply function* $s_i(x_i)$ that returns the maximum possible inflow to the cell when the current traffic volume on it is equal to $x_i \geq 0$, as well as with a controlled *demand function* $d_i(x_i, u_i)$ that returns the maximum possible outflow from the cell when the current traffic volume on it is equal to $x_i \geq 0$ and the local control variable is set to the value u_i.

On every non-onramp cell $i \in \mathcal{E} \setminus \mathcal{R}$, the supply function $s_i(x_i)$ is assumed to be continuous, non-increasing, and concave for values of the traffic volume x_i in the interval $[0, x_i^{\mathrm{jam}}]$, where $x_i^{\mathrm{jam}} = \sup\{x_i \geq 0 : s_i(x_i) > 0\}$ is the jam traffic volume. Conventionally, for all onramp cells $i \in \mathcal{R}$, we set $s_i(x_i) = +\infty$. On the other hand, we assume that the controlled demand functions have the following structure. Every cell $i \in \mathcal{E}$ is equipped with a flow capacity $C_i > 0$ and a continuous, strictly increasing, concave function $d_i(x_i)$ such that $d_i(0) = 0$, to be referred to as the *uncontrolled* demand function. The uncontrolled demand function, flow
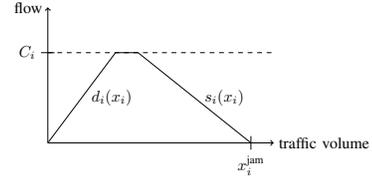

Fig. 1. Trapezoidal fundamental diagram in Example 1.

capacity, and supply function of a cell can be interpreted as the rising, constant, and decreasing parts of a fundamental diagram. A standard case is illustrated below.

Example 1: Consider a non-onramp cell $i \in \mathcal{E} \setminus \mathcal{R}$ with linear uncontrolled demand function $d_i(x_i) = v_i x_i$ and affine supply function $s_i(x_i) = -w_i x_i + x_i^{\mathrm{jam}} w_i$. This leads to a standard trapezoidal fundamental diagram (see Fig. 1). Here, the constants $v_i > 0$ and $w_i > 0$ are referred to as the free-flow speed and the shock-wave speed, respectively.

Then, the controlled demand function is set equal to

$$d_i(x_i, u_i) = \min\{d_i(x_i), u_i C_i\}, \qquad i \in \mathcal{R} \tag{1}$$

on the onramp cells and to

$$d_i(x_i, u_i) = \min\{u_i d_i(x_i), C_i\}, \qquad i \in \mathcal{E} \setminus \mathcal{R} \tag{2}$$

on the non-onramp cells. Equation (1) is to be interpreted as the possibility of implementing ramp-metering by setting up the maximum outflow from an onramp $i \in \mathcal{R}$ to an arbitrary value $u_i C_i$ between 0 and the maximum flow capacity C_i. On the other hand, (2) is to be interpreted as the possibility to control the speed limit in a non-onramp cell i by rescaling the uncontrolled demand function $d_i(x_i)$. Indeed, for trapezoidal fundamental diagrams as in Example 1, (2) is equivalent to the modulation of the free-flow speed $u_i v_i$.

We assume that the system is sampled at times $0 = t_0 < t_1 < \ldots$ that are equally spaced at distance $h > 0$ from each other, so that $t_k = hk$ for $k = 0, 1, \ldots$. At the k-th time instant t_k, the state of the network is a nonnegative vector $x^k \in \mathbb{R}_+^{\mathcal{E}}$ whose entries x_i^k represent the current traffic volumes in the cells i in $\mathcal{E}$, while the control is the vector $u^k \in [0, 1]^{\mathcal{E}}$, whose entries u_i^k represent the the local control variables currently actuated at the cells i. Moreover, the split rates at the k-th time instant are reported in the routing matrix $R^k \in \mathbb{R}_+^{\mathcal{E} \times \mathcal{E}}$ whose entries R_{ij}^k represent the fraction of outflow from cell i that moves towards cell j. To satisfy mass-conservation and topological constraints, the routing matrix R^k is assumed be such that

$$R_{ij}^k \geq 0, \quad i, j \in \mathcal{E}, \qquad \sum_j R_{ij}^k = 1, \quad i \in \mathcal{E} \setminus \mathcal{S}, \tag{3}$$

supported on the set of adjacent pairs $\mathcal{L}$, i.e., such that

$$R_{ij}^k = 0, \qquad \forall (i, j) \notin \mathcal{L}. \tag{4}$$

We denote the exogenous inflow vector at the k-th time instant by $\lambda^k \in \mathbb{R}_+^{\mathcal{E}}$, with the property that $\lambda_i^k = 0$ for all non-onramp cells $i \in \mathcal{E} \setminus \mathcal{R}$ for all $k \geq 0$. For a given initial state x^0, the traffic flow dynamics update rule then reads

$$x_i^{k+1} = x_i^k + h \left(\lambda_i^k + \sum_j R_{ji}^k z_j^k - z_i^k \right), \tag{5}$$

for every $i \in \mathcal{E}$ and $k \geq 0$, where

$$z_i^k = \beta_i^k d_i(x_i^k, u_k^i) \qquad (6)$$

is the total outflow from cell i and

$$\beta_i^k = \sup \Big\{ \beta \in [0,1] : \\ \beta \cdot \max_{\substack{l \in \mathcal{E}: \\ R_{il}^k > 0}} \sum_{h \in \mathcal{E}} R_{hl}^k d_h(x_h^k, u_h^k) - s_l(x_l^k) \leq 0 \Big\}. \qquad (7)$$

We are now ready to formulate the DTA and the FNC as optimization problems. In the former, we assume that, given the initial traffic volume and the exogenous inflows, we can control both the demand functions as in (1)–(2) and the routing matrix within the constraints (3)–(4). In the latter, we assume that the routing matrix is exogenous and the control action is limited to the demand functions, i.e., onramp metering and speed limits. We shall consider a finite time horizon $k_{\max} > 0$ and convex separable costs $\psi_i(x_i)$ of the traffic volumes. Given initial traffic volumes x^0 and exogenous inflows $\{\lambda^k\}_{k=0}^{k_{\max}}$, the DTA problem then reads

$$\min_{\substack{\{x^k, u^k, z^k, R^k\}_{k=0}^{k_{\max}} : \\ (3),(4),(5),(6),(7)}} \sum_{k=0}^{k_{\max}} \sum_{i \in \mathcal{E}} \psi_i(x_i^k). \qquad (8)$$

On the other hand, given initial traffic volumes x^0, exogenous inflows $\{\lambda^k\}_{k=0}^{k_{\max}}$, and an exogenous routing matrix $\{R^k\}_{k=0}^{k_{\max}}$ satisfying (3)–(4), the FNC problem reads

$$\min_{\substack{\{x^k, u^k, z^k\}_{k=0}^{k_{\max}} : \\ (5),(6),(7)}} \sum_{k=0}^{k_{\max}} \sum_{i \in \mathcal{E}} \psi_i(x_i^k). \qquad (9)$$

We now present convex relaxations of the problems above. For this, we introduce the flow variables f_{ij}^k and μ_i^k for $i, j \in \mathcal{E}$, $k = 0, \ldots, k_{\max}$, that satisfy the constraints

$$\mu_i^k \geq 0, \qquad i \neq \mathcal{S} \implies \mu_i^k = 0, \qquad (10)$$

$$f_{ij}^k \geq 0, \qquad \tau_j \neq \omega_i \implies f_{ij}^k = 0, \qquad (11)$$

$$\sum_j f_{ji}^k \leq s_i(x_i^k), \qquad \sum_j f_{ij}^k \leq \min\{d_i(x_i^k), C_i\}. \qquad (12)$$

We then rewrite the dynamics as

$$x_i^{k+1} = x_i^k + h \left(\lambda_i^k + \sum_j f_{ji}^k - \sum_j f_{ij}^k \right). \qquad (13)$$

For given x^0 and $\{\lambda^k\}_{k=0}^{k_{\max}}$, we consider the following relaxation of the DTA problem

$$\min_{\substack{\{x^k, \mu^k, f^k\}_{k=0}^{k_{\max}} : \\ (10),(11),(12),(13)}} \sum_{k=0}^{k_{\max}} \sum_{i \in \mathcal{E}} \psi_i(x_i^k). \qquad (14)$$

Analogously, given x^0, $\{\lambda^k\}_{k=0}^{k_{\max}}$, and $\{R^k\}_{k=0}^{k_{\max}}$ satisfying (3)–(4), we consider the additional constraint

$$f_{ij}^k = R_{ij}^k \left(\mu_i^k + \sum_l f_{il}^k \right) \qquad (15)$$

and the following relaxation of the FNC problem

$$\min_{\substack{\{x^k, \mu^k, f^k\}_{k=0}^{k_{\max}} : \\ (10),(11),(12),(13),(15)}} \sum_{k=0}^{k_{\max}} \sum_{i \in \mathcal{E}} \psi_i(x_i^k). \qquad (16)$$

We then have the following result that mirrors the continuous time results in [8, Proposition 1].

Proposition 1: Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be a network topology, $k = 0, 1, \ldots, k_{\max}$ where $k_{\max} > 0$ is the time horizon, $x^0 \in \mathbb{R}_+^{\mathcal{E}}$ a vector of initial traffic volumes, and λ_i^k exogenous inflows to the onramps $i \in \mathcal{R}$ at time k. Then:

(i) for every feasible solution $\{x^k, f^k, \mu^k\}_{k=0}^{k_{\max}}$ of the convex optimization problem (14), let

$$z_i^k = \mu_i^k + \sum_j f_{ij}^k, \quad u_i^k = \begin{cases} z_i^k / d_i(x_i^k) & \text{if } i \notin \mathcal{R} \\ z_i^k / C_i & \text{if } i \in \mathcal{R}, \end{cases} \qquad (17)$$

and $R_{ij}^k = f_{ij}^k / (\mu_i^k + \sum_j f_{ij}^k)$, for all $i, j \in \mathcal{E}$, with the convention that $u_i^k = 1$ if $d_i(x_i^k) = z_i^k = 0$ on a non-onramp cell $i \in \mathcal{E} \setminus \mathcal{R}$, and that, if $z_i^k = 0$, then $R_{ij}^k = |\{l \in \mathcal{E} : \tau_l = \omega_i\}|^{-1}$ for all $j \in \mathcal{E}$ such that $\tau_j = \omega_i$. Then, for all $k = 0, \ldots, k_{\max}$, the matrix R^k satisfies the constraints (3)–(4) and x^k satisfies the controlled traffic dynamics (5)–(7), so that $\{x^k, u^k, z^k, R^k\}$ is a feasible solution of the DTA problem (8).

Moreover, let $\{R^k\}_{k=0}^{k_{\max}}$ be routing matrices satisfying (3)-(4). Then:

(ii) for every feasible solution $\{x^k, f^k, \mu^k\}_{k=0}^{k_{\max}}$ of the convex optimization (16), let z^k and u^k be as in (17). Then, x^k satisfies the controlled traffic dynamics (5)–(7), so that $\{x^k, u^k, z^k\}_{k=0}^{k_{\max}}$ is a feasible solution of the DTA problem (9).

III. DISTRIBUTED ALGORITHM

If the standard augmented Lagrangian method is used to solve the relaxed DTA and FNC problems, the algorithm can not be implemented in a fully distributed manner. However, as we will now show, it is possible to obtain a distributed solution method by introducing copies of the variables, and force those copies to be equal through additional constraints.

In [13], another algorithm inspired by ADMM for solving the DTA problem distributively is presented. Our approach solves the same problem and is also inspired by ADMM, but differs significantly from the one in [13]. There, the constraints on the decision variables are not included in the Lagrangian, and must be handled when solving the distributed subproblems. This allows, by a suitable partitioning of the variables, for the Lagrangian minimization to be carried out in two steps in each subproblem. In contrast, our approach starts with a Lagrangian which directly includes the constraints for the optimization problem. By choosing variables suitably and carrying out the Lagrangian minimization in a few more steps, this still yields a distributed method. Also, in our case, each subproblem becomes very simple.

For the problems, we separate the outflows from each cell and the inflows to each cell. This is done for the DTA by introducing the matrix $g \in \mathbb{R}_+^{\mathcal{L} \times (k_{\max}+1)}$ and imposing the

additional constraint $f = g$. Each variable g_{ij}^k is considered as an *outflow* from cell i to cell j at time k, and is associated with cell i, while each variable f_{ij}^k is thought of as an *inflow* to cell j from cell i at time k, and is associated with cell j. By updating the variables in f and g separately, we obtain a cell-wise decoupling of the optimization problem, where only the variables associated with a particular cell and its neighboring cells are needed in order to update the variables for the cell in question. To decouple the equations in time, we introduce the matrix $y \in \mathbb{R}_+^{\mathcal{E} \times (k_{\max}+1)}$ whose entries are required to fulfill $y_i^k = x_i^{k+1}$ for $0 \le k < k_{\max}$ and $i \in \mathcal{E}$. By updating x and y separately, only variables associated to the previous and next time points are needed to update the variables associated with a specific time point.

For the DTA problem, the optimization problem becomes

$$\min_{\substack{\{x^k\}_{k=1}^{k_{\max}} \\ \{y^k, \mu^k, f^k, g^k\}_{k=0}^{k_{\max}}}} \sum_{k=0}^{k_{\max}} \sum_{i \in \mathcal{E}} \psi_i(y_i^k)$$

subject to $f^k = g^k$, $x^{k+1} = y^k$,

$$y_i^k = x_i^k + h\left(\lambda_i^k - \mu_i^k + \sum_j f_{ji}^k - \sum_j g_{ij}^k\right),$$
$$\lambda_i^k + \sum_j f_{ji}^k \le s_i(x_i^k), \qquad \mu_i^k + \sum_j g_{ij}^k \le d_i(x_i^k),$$

for $i \in \mathcal{E}$, and $\mu_i^k = 0$ for $i \notin \mathcal{S}$.

For the FNC problem, the turning ratios R_{ij} are pre-determined. This can be taken into account by adding the extra constraint $g_{ij}^k = R_{ij}^k \sum_j g_{ij}^k$. The corresponding augmented Lagrangian at time k is then

$$L_\rho^k(x^k, x^{k+1}, y^k, f^k, g^k, \mu^k; \nu^k, \sigma^k, \xi^k, \eta^k) = \sum_i \psi_i(x_i^k)$$
$$+ \sum_i \gamma_i^k \left(y_i^k - x_i^k - h\left(\lambda_i^k - \mu_i^k + \sum_j f_{ji}^k - \sum_j g_{ij}^k\right)\right)$$
$$+ \sum_{(i,j)} \nu_{ij}^k (f_{ij}^k - g_{ij}^k) + \sum_i \sigma_i^k (y_i^k - x_i^{k+1})$$
$$+ \sum_i \xi_i^k \left(\lambda_i^k + \sum_j f_{ji}^k - s_i(x_i^k)\right)$$
$$+ \sum_i \eta_i^k \left(\mu_i^k + \sum_j g_{ij}^k - d_i(x_i^k)\right) + \frac{\rho}{2} M^k,$$

where M^k consists of penalty terms which have been added to the Lagrangian. These are zero when the constraints are satisfied and positive otherwise. This procedure is described in [12] for equality-constrained problems, but here we have applied the analogous idea for inequality constraints as well. Note that the penalty terms are squared, so that the augmented Lagrangian is differentiable. Then,

$$M^k = M^k(x^k, x^{k+1}, y^k, f^k, g^k, \mu^k)$$
$$= \sum_i \left(y_i^k - x_i^k - h\left(\lambda_i^k - \mu_i^k + \sum_j f_{ji}^k - \sum_j g_{ij}^k\right)\right)^2$$
$$+ \sum_{(i,j)} (f_{ij}^k - g_{ij}^k)^2 + \sum_i (y_i^k - x_i^{k+1})^2$$
$$+ \sum_i \left(\max\left\{0, \lambda_i^k + \sum_j f_{ji}^k - s_i(x_i^k)\right\}\right)^2$$
$$+ \sum_i \left(\max\left\{0, \mu_i^k + \sum_j g_{ij}^k - d_i(x_i^k)\right\}\right)^2.$$

Furthermore, $\gamma^k, \sigma^k \in \mathbb{R}^{\mathcal{E}}$, $\nu^k \in \mathbb{R}^{\mathcal{L}}$, and $\xi^k, \eta^k \in \mathbb{R}_+^{\mathcal{E}}$ are dual variables for the problem and $\rho > 0$ is an *penalty parameter* to be chosen. Note that for the last time step, $k = k_{\max}$, the two sum terms containing x_i^{k+1} must be removed from the augmented Lagrangian. For the FNC problem, the additional terms

$$\sum_i \theta_i^k \left(g_{ij}^k - R_{ij}^k \sum_j g_{ij}^k\right) + \sum_i \frac{\rho}{2}\left(g_{ij}^k - R_{ij}^k \sum_j g_{ij}^k\right)^2$$

are added to the augmented Lagrangian, where $\theta^k \in \mathbb{R}^{\mathcal{E}}$ are dual variables for the extra constraint. The augmented Lagrangian for the whole optimization problem is then, in the DTA case, given by

$$L_\rho(x, x, y, f, g, \mu; \nu, \sigma, \xi, \eta) =$$
$$\sum_{k=0}^{k_{\max}} L_\rho^k(x^k, x^{k+1}, y^k, f^k, g^k, \mu^k; \nu^k, \sigma^k, \xi^k, \eta^k).$$

In the augmented Lagrangian method, constrained optimization problems are solved by iteratively minimizing the augmented Lagrangian for given dual variables and then updating the dual variables by taking a step in the gradient direction of the dual function. In ADMM, as described in [12], a similar approach is employed, with the difference that the primal variables are divided into two sets and that the Lagrangian is minimized with respect to one of these sets at a time, while keeping the remaining variables constant. It is this idea that allows the ADMM to be used distributively, in difference form the augmented Lagrangian method. Since the problems we are considering yield couplings between variables both for different cells and different time points, we are making use of the idea to minimize the augmented Lagrangian in several steps, but generalize it to five steps instead of two. This enables a decoupling both between cells and between time points. The resulting optimization algorithm for the DTA problem thus consists in an iterative procedure in which first the following steps are performed

$$f^+ := \operatorname*{argmin}_f L_\rho(x, y, f, g, \mu; \gamma, \nu, \sigma, \xi, \eta),$$
$$g^+ := \operatorname*{argmin}_g L_\rho(x, y, f^+, g, \mu; \gamma, \nu, \sigma, \xi, \eta),$$
$$\mu^+ := \operatorname*{argmin}_\mu L_\rho(x, y, f^+, g^+, \mu; \gamma, \nu, \sigma, \xi, \eta),$$
$$y^+ := \operatorname*{argmin}_y L_\rho(x, y, f^+, g^+, \mu^+; \gamma, \nu, \sigma, \xi, \eta),$$
$$x^+ := \operatorname*{argmin}_x L_\rho(x, y^+, f^+, g^+, \mu^+; \gamma, \nu, \sigma, \xi, \eta).$$

In each iteration, these steps are then followed by dual variable updates according to

$$(\gamma_i^k)^+ := \gamma_i^k + \rho\left(y_i^k - x_i^k - h\left(\lambda_i^k - \mu_i^k + \sum_j (f_{ji}^k - g_{ij}^k)\right)\right)$$
$$(\nu_{ij}^k)^+ := \nu_{ij}^k + \rho(f_{ij}^k - g_{ij}^k), \quad (\sigma_i^k)^+ := \sigma_i^k + \rho(y_i^k - x_i^{k+1}),$$
$$(\xi_i^k)^+ := \max\left\{0, \xi_i^k + \rho\left(\lambda_i^k + \sum_j f_{ji}^k - s_i(x_i^k)\right)\right\},$$
$$(\eta_i^k)^+ := \max\left\{0, \eta_i^k + \rho\left(\mu_i^k + \sum_j g_{ij}^k - d_i(x_i^k)\right)\right\}.$$

Note that the dual variables ξ_i^k and η_i^k always are non-negative, which is required in the solution of the dual problem. For the FNC problem, the augmented Lagrangian is a function of the extra dual variables $\theta^k \in \mathbb{R}^{\mathcal{E}}$ as well, and also these variables must be updated in the end of each iteration according to

$$(\theta_i^k)^+ := \theta_i^k + \rho \left(g_{ij}^k - R_{ij}^k \sum_j g_{ij}^k \right).$$

The fact that the algorithm is distributed can, e.g. in the case of updating the variable $f_{j_1 i}^k$ (an inflow to cell i), be seen as

$$\left(f_{j_1 i}^k \right)^+ = \underset{f_{j_1 i}^k}{\arg\min}\, L_\rho(x, y, f, g, \mu; \gamma, \nu, \sigma, \xi, \eta) =$$
$$\underset{f_{j_1 i}^k}{\arg\min}\, L_\rho^k(x_i^k, y_i^k, f_{ji}^k, (g_{ji}^k, g_{ij}^k), \mu_i^k; \gamma_i^k, \nu_{ji}^k, 0, \xi_i^k, 0)\,.$$

It follows that in order to compute the new flow from cell j_1 to cell i at time k, only information about state variables and Lagrange multipliers associated with cell i and its neighboring cells are needed. Furthermore, only variables at time step k are needed. In general, for all the variable updates, variable values for the adjacent time points $k-1$ and $k+1$ are also needed, but not for any other times. Thus, only information associated to neighboring cells and time points is needed to update the primal variables.

The total number of variables needed in the optimization problem is proportional to the number of time steps as well as the number of cells or the number of adjacent cells. The number of adjacent cells are in practice not increasing fast with the size of the network, since each cell typically has at most two or three adjacent cells in each direction.

IV. SIMULATION RESULTS

In this section, we present simulations for both the DTA and FNC problem, on two different networks.

A. DTA

The algorithm for solving the DTA problem is tested on a setup obtained from [8]. In this example, the single-source single-sink network in Fig. 2 is considered. The network is initially assumed to be empty, and the time horizon is chosen as 250 seconds with time discretization interval $h = 10$. The exogenous inflow at cell 1 is prescribed to be $\lambda_1^1 = 0.8$, $\lambda_1^2 = 1.6$, $\lambda_1^3 = 0.8$ and $\lambda_1^k = 0$ for $k \geq 4$, and exogenous outflow is only allowed at cell 10. Furthermore, the supply and demand functions are given by $s_i(x_i, k) = \min \left\{ w_i(x_i^{\text{jam}} - x_i)/L_i, C_i^k \right\}$ and $d_i(x_i, k) = \min \left\{ v_i x_i/L_i, C_i^k \right\}$, where v_i, w_i, C_i^k, L_i and x_i^{jam} are the free-flow speed, the speed of the congestion wave, the capacity (at time step k), the cell length and the jam traffic volume for cell i respectively. In the simulations we set $v_i = w_i = 50$ feet/s, $L_i = 500$ feet for all cells i. Moreover, for cell 1, 2, 9 and 10 we let the $C_i^k = 1.2$ vehicles/s for all k and $x_i^{\text{jam}} = 20$ vehicles. For all other cells, $x_i^{\text{jam}} = 10$ vehicles and $C_i^k = 0.6$ for all k, apart from cell 4 where $C_4^5 = C_4^6 = 0$ vehicles/s and $C_4^7 = C_4^8 = 0.3$ vehicles/s.
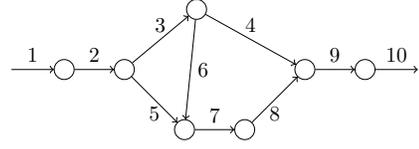

Fig. 2. Network used for test of the DTA solving algorithm.

TABLE I
RESULTS FOR TESTS OF THE DTA ALGORITHM.

ρ	0.1	1	10	100
nbr. of itr.	1 000 000	100 000	11 285	100 000
comp time. [s]	84 835	6 962	765	6 837
rel. cost err. ε_ψ	$3.5 \cdot 10^{-6}$	$34 \cdot 10^{-6}$	$13 \cdot 10^{-6}$	$168 \cdot 10^{-6}$
mean err. $\bar{\varepsilon}_y$	$69 \cdot 10^{-6}$	$674 \cdot 10^{-6}$	$6.7 \cdot 10^{-3}$	$12 \cdot 10^{-3}$
max err. $\varepsilon_y^{\max}$	$497 \cdot 10^{-6}$	$4.8 \cdot 10^{-3}$	$47 \cdot 10^{-3}$	$170 \cdot 10^{-3}$
feasibility res.	$2.9 \cdot 10^{-6}$	$29 \cdot 10^{-6}$	$5.6 \cdot 10^{-6}$	$10 \cdot 10^{-3}$
duality gap	$5.1 \cdot 10^{-3}$	$5.2 \cdot 10^{-3}$	$1.0 \cdot 10^{-3}$	$16 \cdot 10^{-3}$

This is in order to simulate a time dependent bottleneck in the traffic network. The cost function associated with each cell at each time point is chosen to be $\psi_i(x_i) = x_i^2$.

An optimal solution is found when a set of feasible decision variables are found such that the duality gap, i.e., the difference between the cost function and the dual function of the constrained optimization problem, is zero. Thus, the algorithm iterations should continue until these criteria are fulfilled within some small error tolerance. To check the duality gap criterion we consider an approximation of the duality gap obtained by approximating the dual function as the Lagrangian evaluated at the primal variables obtained in the last iteration, and then evaluating the cost function and the dual function approximation for the primal and dual variables obtained in the last iteration. Both the feasibility and duality tolerances were chosen to 10^{-3}.

In order to verify that the correct results are obtained from the algorithm, the optimization problem is also solved in a centralized manner by CVX [14]. Values that are compared are the relative error in the cost function, ε_ψ, the mean error (over all time points and cells) in the cell traffic volumes, $\bar{\varepsilon}_y$, as well as the maximal error in any cell traffic volume at any time, $\varepsilon_y^{\max}$. Tab. I shows the performance of the algorithm for different values of ρ. The algorithm was iterated either until the stopping criteria were fulfilled or until a maximal threshold (10^6 for $\rho = 0.1$ and 10^5 otherwise) of the number of iterations was reached. Changes of the cost function, of the feasibility residual, and of the duality gap with number of iterations for different values of ρ are presented in Fig. 3. From these results, we can conclude that the algorithm manages to find the optimal decision variables for the tested DTA problem with high accuracy, as long as ρ is sufficiently small.

B. FNC

In order to test the FNC algorithm for a realistic transportation network, a network inspired by the freeway system in Los Angeles is used. The topology is a slightly modified version of the one described in [15].
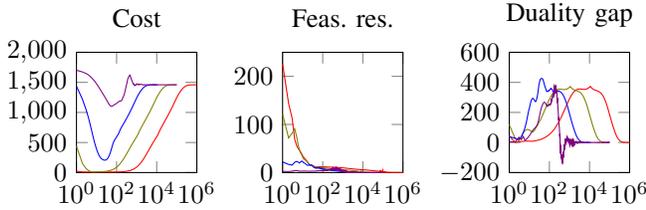
Fig. 3. Cost function, feasibility residual and duality gap for the DTA problem with different penalty parameters ρ. In the plots $\rho = 0.1$ is red, $\rho = 1$ is green, $\rho = 10$ is blue and $\rho = 100$ is violet.
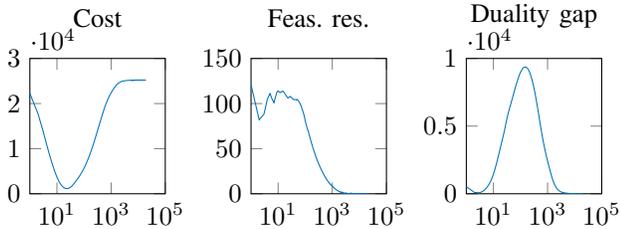

Fig. 4. Cost function, feasibility residual and duality gap for the FNC problem. After an initial decrease, the cost increases until it corresponds to a feasible solution.

The supply and demand functions are chosen as the affine functions $s_i(x_i) = w_i(x_i^{\text{jam}} - x_i)/L_i$ and $d_i(x_i) = v_i x_i / L_i$, where w_i, L_i, v_i, x_i^{jam} are the wave speed, the cell length, the free-flow vehicle speed and the jam mass on cell i, respectively. We let the intersection between the supply and demand function determine the maximum flow capacity. The routing matrix is chosen such that 10% of the vehicles leave the network at each offramp, while the traffic is equally distributed among the other out-links in each intersection.

The initial state, consisting of the initial cell traffic volumes, is chosen as the equilibrium obtained by running CTM without any controls applied for the constant inflow $\lambda_i = 0.05$ vehicles per second at each source cell. For the optimization horizon, the inflow at each source cell is assumed to be $\lambda_i = 0.1$ vehicles per second. The time horizon is chosen as 1 minute and the time discretization interval as $h = 10$ seconds, fulfilling the CFL-condition $\max_e \frac{v_e h}{L_e} < 1$. The used penalty parameter is $\rho = 10$.

When running the algorithm on this setup, the resulting evolution of the cost function, feasibility residual and duality gap are as shown in Fig. 4. Note that the cost function increases after an initial decrease until the obtained cost corresponds to a feasible solution. The obtained cost, with the cell cost functions $\psi_i(x) = x^2$, is 25 204. When the same setup is simulated with CTM without any control, the corresponding cost is 30 434. Thus, the optimal control manages to achieve a 17% decrease of the cost.

V. CONCLUSION

We presented a distributed algorithm for the optimal traffic flow control in transportation networks. By applying the algorithm to test scenarios, we demonstrated that it can be used for solving both the DTA and the FNC problems.

For future development, it would be desirable to formally prove under which conditions the algorithm converges by, e.g., giving an upper limit of the penalty ρ and to further examine is how many iterations are necessary to yield a solution with sufficient accuracy. Finally, it would be interesting to study stability and robustness of the resulting optimal controls with respect to dynamic routing [16], [17].

REFERENCES

[1] C. Rosdahl. Distributed control of dynamic flows in traffic networks. Master's thesis, Lund University, 2017.
[2] D.K. Merchant and G.L. Nemhauser. A model and an algorithm for the dynamic traffic assignment problem. *Transportation Science*, 12:183–199, 1978.
[3] D.K. Merchant and G.L. Nemhauser. Optimality conditions for a dynamic traffic assignment model. *Transportation Science*, 12:200–207, 1978.
[4] S. Peeta and A. K. Ziliaskopoulos. Foundations of dynamic traffic assignment: The past, the present and the future. *Networks and Spatial Economics*, 1(3-4):233–265, 2001.
[5] G. Gomes and R. Horowitz. Optimal freeway ramp metering using the asymmetric cell transmission model. *Transportation Research Part C*, 14(4):244–268, 2006.
[6] A. Muralidharan and R. Horowitz. Optimal control of freeway networks based on the link node cell transmission model. In *Proceedings of the American Control Conference (ACC)*, pages 5769–5774, 2012.
[7] M. Schmitt, C. Ramesh, and J. Lygeros. Sufficient optimality conditions for distributed, non-predictive ramp metering in the monotonic cell transmission model. *Transportation Research Part B: Methodological*, 105:401 – 422, 2017.
[8] G. Como, E. Lovisari, and K. Savla. Convexity and robustness of dynamic traffic assignment and freeway network control. *Transportation Research Part B: Methodological*, 91:446–465, 2016.
[9] C. F. Daganzo. The cell transmission model: A dynamic representation of highway traffic consistent with the hydrodynamic theory. *Transportation Research Part B: Methodological*, 28B(4):269–287, 1994.
[10] C. F. Daganzo. The cell transmission model, part II: network traffic. *Transportation Research B: Methodological*, 29B(2):79–93, 1995.
[11] A.K. Ziliaskopoulos. A linear programming model for the single destination system optimum dynamic traffic assignment problem. *Transportation science*, 34(1):37–49, 2000.
[12] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Found. Trends Mach. Learn.*, 3(1):1–122, January 2011.
[13] Q. Ba and K. Savla. On distributed computation of optimal control of traffic flow over networks. *Fifty-fourth Annual Allerton Conference, Allerton House, UIUC, Illinois, USA*, September 2016.
[14] CVX Research Inc. CVX: Matlab software for disciplined convex programming, version 2.1. http://cvxr.com/cvx, March 2017.
[15] G. Como, E. Lovisari, and K. Savla. Convexity and robustness of dynamic network traffic assignment for control of freeway networks. In *Proceedings of 14th IFAC Symposium on Control in Transportation Systems (CTS 2016)*, 2016.
[16] G. Como, K. Savla, D. Acemoglu, M. A. Dahleh, and E. Frazzoli. Stability analysis of transportation networks with multiscale driver decisions. *SIAM Journal on Control and Optim.*, 51(1):230–252, 2013.
[17] G. Como. On resilient control of dynamical flow networks. *Annual Reviews in Control*, 43:80–90, 2017.