2015 American Control Conference
Palmer House Hilton
July 1-3, 2015. Chicago, IL, USA

# A Multi-Convex approach to Latency Inference and Control in Traffic Equilibria from Sparse data

Jérôme Thai[1] and Rim Hariss[2] and Alexandre Bayen[3]

***Abstract*— A common behavioral assumption in the modeling of traffic networks is the user equilibrium. Since traffic volumes, resulting from the rational behavior of agents, are easily but sparsely observable, and delay functions are not directly observable, we present a mathematical program with equilibrium constraint (MPEC) framework to impute the delay functions and centrally control the system from partial observations of equilibria. We also develop a novel method for solving MPECs using multi-convex optimization. Our block descent method has an intuitive interpretation, and numerical experiments demonstrate its accuracy for structural estimation, and highlight the importance of sensor placement for toll pricing.**

## I. Introduction

*Wardrop equilibrium*, or *user equilibrium* (UE), is used as a common solution concept in the study of traffic models, typically in transportation and telecommunications networks. It assumes that each agent has access to the *delay function* on each arc and chooses the shortest path from origin to destination [34]. In reality, delay functions are not known and are often difficult to estimate. But traffic flows are easily measurable. For example, they are the most common types of data on road networks, obtained from loop detectors and radars. Hence previous work has studied the *inverse problem*, which aims at estimating the delay functions given observations of (approximate) equilibria.

In practice, data often suffers from *missing values* due to mistakes in data collection, or limitation from experimental design [23]. For example, there is often a shortage of information in transportation networks due to the cost of deploying sensors in large metropolitan areas. With limited infrastructure, an optimal placement of the sensors is then crucial for the accurate estimation of the delay functions of the underlying traffic model. This problem is particularly important if a *leader* wants to "centrally control" the system to achieve a desirable objective. For instance, inaccurate delay function estimates can lead to *toll strategies* [6], [19] for which the induced response is far from the *system-optimum* (SO), or even worse than the UE solution of the uncontrolled system problem.

Hence the goal of the article is threefold. In the first part, we are interested in the following (more difficult) inverse equilibrium problem: given *partially-observed* equilibria, how can we estimate the delay functions and so impute the missing data? The estimation problem is posed as a *Mathematical Program with Variational Inequality* (MPVI) constraint. In the second part, we develop a novel and efficient technique to solve Mathematical Programs with Equilibrium constraints (MPEC) in which constraints are defined as a *Variational Inequality* (VI), and we apply the proposed methodology for the estimation and control of traffic networks in equilibrium. In the third part, we show that a good sensor placement is essential to design toll strategies that minimize the social cost of the system.

### A. Literature review on inverse problems

Our work contributes to various disciplines. The problem of estimating the parameters of the underlying process based on available observations has been addressed before in many fields. Iyengar and Kang [20] and Keshavarz, Wang and Boyd [22] consider the problem of imputing the objective of a parametric optimization problem from nearly optimal points. Inverse reinforcement learning in robotics has been studied by Ng and Russell [26] and Abbeel and Ng [1] which consists in learning the reward function based on information about the optimal policy. Burton et al. [11] studied the inverse shortest path problem of imputing arc weights from observations of shortest path costs. In control, Boyd et al. focused on recovering the parameters of the Lyapunov function given a linear control policy [9, §10.6]. The difference between our work and [20], [22] is that we focus on the context of VI, which combines generality and computational efficiency; see [18], [17] for details on VI.

In econometrics, techniques for estimating the parameters of models given equilibria are generally referred to as *structural estimation*. Many of them focused on imputing demand and production functions [30], [2], [4]. Recent work of Bertsimas et al. [8] focused on equilibrium models described by a VI and studied the inverse VI problem from full observations of (approximate) equilibria. One difference between our approach and [8] is that we allow partial observations through a linear observation model. Furthermore, the imputation problem is posed as a MPEC in which the parameters of the model are selected such that the induced response minimizes the observation residual; see [24] for an overview on the MPEC.

### B. Related work in VI and convex optimization

Besides the novel use of the MPEC for estimation, our work includes key contributions to the MPEC literature: (i) the reformulation of the MPEC as a penalized single-level

[1]Jérôme Thai is with the department of Electrical Engineering and Computer Sciences, University of California at Berkeley. jerome.thai@berkeley.edu
[2]Rim Hariss is with the Operations Research Center, Massachusetts Institute of Technology. rhariss@mit.edu
[3] Alexandre Bayen is with the department of Electrical Engineering and Computer Sciences, and the department of Civil and Environmental Engineering, University of California at Berkeley. bayen@berkeley.edu

978-1-4799-8684-2/$31.00 ©2015 AACC

optimization program via LP duality, (ii) sufficient conditions for the *multi-convex* structure (as defined in [35]) of the reformulation, (iii) a novel algorithmic approach exploiting the structure of the reformulation. Although the application of LP duality to VI is common, *e.g.* for toll pricing [6], for solving the VI problem [3], and for solving inverse VI problems [8], our analysis extends the works outlined above to propose a unified framework for the study, estimation, and control of models described as a VI.

Convex optimization is also closely related to VI; see review from Scutari et al. [32]. When the map that describes the VI can be expressed as the gradient of some convex *potential function*, the VI problem coincides with the convex optimization problem. Under some milder assumptions, VI also coincides with coupled convex optimization problems, which arise in the study of Nash equilibria. The use of iterative algorithms such as the Gauss-Seidel scheme [16, §5], or more specifically coordinate descent methods [35], is a natural approach to solve such problems. Under different conditions, Aghassi et al. [3] show that the reformulation of the VI via LP duality is convex. Building on these previous works, we decompose our single-level formulation into blocks of variables and we define sufficient conditions for which the problem is convex in each block. We also give an intuitive interpretation of the proposed decomposition, and apply block coordinate descent to solve estimation and toll pricing models on the highway network near Los Angeles.

## II. Traffic equilibria and Variational Inequality

In the traffic equilibrium problem we are given a directed network $(\mathcal{N}, \mathcal{A})$ with $\mathcal{N}$ being the set of nodes indexed by $i$ and $\mathcal{A}$ being the set of directed arcs $a = (i, j) \in \mathcal{N} \times \mathcal{N}$. Agents travelling from an origin $s \in \mathcal{N}$ to a destination $t \in \mathcal{N}$ are associated to an origin-destination (OD) pair $(s, t)$ called a commodity. We are given a set $\mathcal{C} \subseteq \mathcal{N} \times \mathcal{N}$ of commodities and for each $k = (s_k, t_k) \in \mathcal{C}$, a flow of demand rate $d_k$ must be routed from $s_k$ to $t_k$. The $k$-th *commodity flow vector*, denoted $\mathbf{x}^k = (x_a^k)_{a\in\mathcal{A}}$, describes the flow in commodity $k$ on each arc. Then a commodity flow vector $\mathbf{x}^k \in \mathbb{R}_+^{\mathcal{A}}$ is feasible if it satisfies the flow conservation at every node $i \in \mathcal{N}$:

$$\sum_{j:(j,i)\in\mathcal{A}} x_{(j,i)}^k - \sum_{j:(i,j)\in\mathcal{A}} x_{(i,j)}^k = \begin{cases} -d_k & \text{if} \quad i = s_k \\ d_k & \text{if} \quad i = t_k \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

With $\mathbf{N} \in \{-1, 0, 1\}^{|\mathcal{N}|\times|\mathcal{A}|}$ the node-arc incidence matrix of the network and $\mathbf{b}^k \in \mathbb{R}^{|\mathcal{N}|}$ the demand vector associated to commodity $k$ with entries such that $b_{s_k}^k = -d_k$, $b_{t_k}^k = d_k$, and $b_i^k = 0, \forall\, i \neq s_k, t_k$, the system of linear equations (1) can be cast in matrix format: $\mathbf{N}\mathbf{x}^k = \mathbf{b}^k$, $\mathbf{x}^k \succeq 0, \forall\, k \in \mathcal{C}$. If we let $\mathbf{x} := (\mathbf{x}^k)_{k\in\mathcal{C}} \in \mathbb{R}^{|\mathcal{C}|\cdot|\mathcal{A}|}$ and $\mathbf{b} := (\mathbf{b}^k)_{k\in\mathcal{C}} \in \mathbb{R}^{|\mathcal{C}|\cdot|\mathcal{N}|}$ be the overall vectors of commodity flows and demand vectors, and $\mathbf{A} = \mathrm{diag}(\mathbf{N}, \cdots, \mathbf{N}) \in \{-1, 0, 1\}^{|\mathcal{C}|\cdot|\mathcal{N}|\times|\mathcal{C}|\cdot|\mathcal{A}|}$ the block-diagonal matrix with $\mathbf{N}$ at each of its $|\mathcal{C}|$ blocks, the system of flow conservation equations can be written in general format: $\mathbf{A}\mathbf{x} = \mathbf{b}$, $\mathbf{x} \succeq 0$.

We define the *aggregate flow vector* $\mathbf{v} = \sum_k \mathbf{x}^k$ as the sum of all commodity flows, with entries $v_a = \sum_k x_a^k$ being the aggregate flow on each arc. We are also given *continuous positive nondecreasing* delay functions $s_a : \mathbb{R}_+^{\mathcal{A}} \to \mathbb{R}, \forall\, a \in \mathcal{A}$ such that the travel time on arc $a$ for an aggregate flow $\mathbf{v}$ is $s_a(\mathbf{v})$. Each delay $s_a$ can include additional travel costs such as the toll incurred on arc $a$. The functions $s_a$ are also called *latency functions* in algorithmic game theory [27]. Beckmann et al. [5] considered the *separable case* in which the delay $s_a(\cdot)$ only depends on the aggregate flow $v_a$ on arc $a$, and proved UE commodity flows always exist and are *optimal solutions* of the convex program:

$$\min_{\mathbf{x}} z(\mathbf{Z}\mathbf{x}) \quad \text{s.t.} \quad \mathbf{A}\mathbf{x} = \mathbf{b},\, \mathbf{x} \succeq 0 \quad (2)$$

where $\mathbf{Z} \in \{0, 1\}^{|\mathcal{A}|\times|\mathcal{C}|\cdot|\mathcal{A}|}$ maps $\mathbf{x}$ to $\mathbf{v}$, *i.e.* $\mathbf{v} = \mathbf{Z}\mathbf{x} = \sum_k \mathbf{x}_k$, and $f : \mathbb{R}_+^{\mathcal{A}} \to \mathbb{R}$ is the *Beckmann function* on $\mathbf{v}$:

$$z(\mathbf{v}) = \sum_{a\in\mathcal{A}} \int_0^{v_a} s_a(u)du \quad (3)$$

This is known as the *arc-flow or link-flow formulation* in traffic assignment [28]. A common alternative is the *path-flow formulation* in which we enumerate the set $\mathcal{P}_k$ of all simple paths from $s_k$ to $t_k$ for $k \in \mathcal{C}$ and let $\mathcal{P} := \cup_{k\in\mathcal{C}}\mathcal{P}_k$. A path-flow vector $\boldsymbol{\phi} := (\phi_p)_{p\in\mathcal{P}}$ describing the flow on each path is feasible if $\phi_p \geq 0, \forall\, p \in \mathcal{P}$ and $\sum_{p\in\mathcal{P}_k} \phi_p = d_k, \forall\, k \in \mathcal{C}$. Let $\mathbf{P} \in \{0, 1\}^{|\mathcal{A}|\times|\mathcal{P}|}$ and $\mathbf{U} \in \{0, 1\}^{|\mathcal{C}|\times|\mathcal{R}|}$ be the arc-route and OD-route incidence matrices respectively, then the path-flow equilibria are optimal solutions of:

$$\min_{\boldsymbol{\phi}} z(\mathbf{P}\boldsymbol{\phi}) \quad \text{s.t.} \quad \mathbf{U}\boldsymbol{\phi} = \mathbf{d},\, \boldsymbol{\phi} \succeq 0 \quad (4)$$

where the aggregate flow is given by $\mathbf{v} = \mathbf{P}\boldsymbol{\phi}$. In game theory, (4) is known as a *potential game*, the map $\boldsymbol{\phi} \mapsto z(\mathbf{P}\boldsymbol{\phi})$ is called a *potential function*, and the equilibria are given by the *Karush-Kuhn-Tucker* (KKT) conditions [31].

From our assumptions on $s_a$, (2) and (4) have same solutions in terms of $\mathbf{v}$ [7, §8.6], and the Beckmann function $z$ is convex [5]. Hence (2) and (4) can be expressed as a *convex program*, denoted OP($\mathcal{K}, f$):

$$\min\, f(\mathbf{x}) \quad \text{s.t.} \quad \mathbf{x} \in \mathcal{K} \quad (5)$$

where $\mathcal{K} \subseteq \mathbb{R}^n$ is a convex set, $f : \mathcal{K} \to \mathbb{R}$ a convex function, and $n$ the dimension of the problem. From the optimality conditions in convex optimization [10, §4.2.3]:

***Theorem 1****. With $f$ differentiable and $\nabla f$ its gradient, a feasible point $\mathbf{x}^\star \in \mathcal{K}$ solves OP($\mathcal{K}, f$) if and only if*

$$\nabla f(\mathbf{x}^\star)^T(\mathbf{u} - \mathbf{x}^\star) \geq 0, \forall\, \mathbf{u} \in \mathcal{K} \quad (6)$$

The VI problem can be seen as a generalization of (6) where $\nabla f$ is replaced by a general map $F$. Given a set $\mathcal{K} \subseteq \mathbb{R}^n$ and a map $F : \mathcal{K} \to \mathbb{R}^n$, the VI problem, denoted VI($\mathcal{K}, F$), consists in finding $\mathbf{x}^\star \in \mathcal{K}$ such that:

$$F(\mathbf{x}^\star)^T(\mathbf{u} - \mathbf{x}^\star) \geq 0, \forall\, \mathbf{u} \in \mathcal{K} \quad (7)$$

Let $S : \mathbb{R}_+^{\mathcal{A}} \to \mathbb{R}_+^{\mathcal{A}}$ be the overall delay map such that its entries $S := (s_a)_{a\in\mathcal{A}}$ are the arc delay functions of the aggregate flow $\mathbf{v} \in \mathbb{R}_+^{\mathcal{A}}$. Since $S$ is the gradient of the Beckmann function, *i.e.* $\nabla z = S$, (2) is equivalent to VI($\mathcal{K}, F$) where the pair $(\mathcal{K}, F)$ is given by:

$$\begin{aligned} &\mathcal{K} = \{\mathbf{x} \,|\, \mathbf{A}\mathbf{x} = \mathbf{b},\, \mathbf{x} \succeq 0\} \\ &F(\mathbf{x}) = \mathbf{Z}^T S(\mathbf{Z}\mathbf{x}) \end{aligned} \tag{8}$$

This is another important characterization of traffic equilibria; see Dafermos [13]. We conclude the section with the notion of *system-optimiality* (SO), where each agent tries to minimize the *total delay* (9); see [28, §2.4]. SO flow solutions are obtained by solving (2) or (4) where the Beckmann function $z$ is substituted for the total delay function $c$:

$$c(\mathbf{v}) = \mathbf{v}^T S(\mathbf{v}) = \sum_{a \in \mathcal{A}} v_a s_a(v_a) \tag{9}$$

We note that $\mathbf{x}^T F(\mathbf{x}) = S^T(\mathbf{Z}\mathbf{x})\mathbf{Z}\mathbf{x} = c(\mathbf{Z}\mathbf{x})$.

## III. Structural estimation with missing data

We consider the traffic equilibrium model described by VI($\mathcal{K}, F$) described by (8), with an unknown delay map $S = (s_a)_{a \in \mathcal{A}}$. We are given the matrices $\mathbf{Z}$, $\mathbf{A}$ which encode the geometry of the network, and a linear observation model $\mathbf{H}$ on the aggregate flow $\mathbf{v}$. For instance, in urban networks, traffic sensors (such as loop detectors [12]) are usually placed on a small arc subset $\mathcal{A}^{\text{obs}} \subset \mathcal{A}$, mostly along highways, hence the associated observation matrix $\mathbf{H} \in \{0,1\}^{|\mathcal{A}^{\text{obs}}| \times |\mathcal{A}|}$ has entries such that for all $a \in \mathcal{A}^{\text{obs}}$, $\mathbf{H}_{aa'} = 1$ if $a' = a$ and $\mathbf{H}_{aa'} = 0$ if $a' \neq a$.

Changes in the demand vector $\mathbf{b}$ affect the set of feasible flows $\mathcal{K}$, and induce different equilibria $\mathbf{x}$ on the network. However, equilibria are not directly observable, we only observe linear measurements $\mathbf{z} := \mathbf{H}\mathbf{v} = \mathbf{H}\mathbf{Z}\mathbf{x} \in \mathbb{R}^{\mathcal{A}^{\text{obs}}}$.

Our goal is to impute the delay map $S$, or more broadly the map $F$ from a set of observations consisting of pairs $(\mathbf{b}^j, \mathbf{z}^j)$ of different demands $\mathbf{b}^j$ and different aggregate flow measurements $\mathbf{z}^j$ for $j = 1, \cdots, N$. The feasible sets $\mathcal{K}_j$ associated to demands $\mathbf{b}^j$ are given by:

$$\mathcal{K}_j = \{\mathbf{x}^j \,|\, \mathbf{A}\mathbf{x}^j = \mathbf{b}^j,\, \mathbf{x}^j \succeq 0\} \tag{10}$$

Following standard techniques, we assume for tractability reasons, that $F(\cdot|\boldsymbol{\theta})$ is restricted to be a finite dimensional affine parametric model, *i.e.*, there exists basis maps $F_i$, $i = 0, \cdots, d$ and a *convex* set $\Theta \subseteq \mathbb{R}^d$, such that:

$$F(\cdot|\boldsymbol{\theta}) = F_0 + \sum_{i=1}^{d} \theta_i F_i(\cdot), \quad \forall\, \theta \in \Theta \tag{11}$$

where the fixed part $F_0$ of the parametric map is non null. The shift $F_0$ imposes a normalization on $F(\cdot|\boldsymbol{\theta})$ that excludes trivial solutions from $\Theta$, *e.g.* null maps for which the entire set $\mathcal{K}$ is solution to the VI problem.

We also want the basis $F_i$, $i = 0, \cdots, d$ and $\Theta$ to include our prior knowledge on the delay function. For example, if we know that the true map $F$ is convex, then having $\Theta \subseteq \mathbb{R}_+^d$ and $F_i$ convex for all $i$ is sufficient to ensure that the parametric map $F(\cdot|\boldsymbol{\theta})$ is convex.

Based on the measurements $\mathbf{z}^j$, one wants to find $\boldsymbol{\theta}^\star \in \Theta$ such that, *for all* $j$, there exists a solution $\mathbf{x}^j$ of VI($\mathcal{K}_j, F(\cdot|\boldsymbol{\theta}^\star)$) that corresponds to the measurements, *i.e.* $\mathbf{H}\mathbf{Z}\mathbf{x}^j = \mathbf{z}^j$. Because of noisy data and approximations in the parametric model, it is generally not possible to fit the model perfectly to the data and one hopes to minimize the measurement residuals $\|\mathbf{H}\mathbf{Z}\mathbf{x}^j - \mathbf{z}^j\|$, $\forall\, j$. This leads to the following MPEC, where $\tilde{\mathbf{H}} := \mathbf{H}\mathbf{Z}$

$$\begin{aligned} \min_{\boldsymbol{\theta}, \mathbf{v}} \quad & \tfrac{1}{2} \textstyle\sum_j \|\tilde{\mathbf{H}}\mathbf{x}^j - \mathbf{z}^j\|^2 \\ \text{s.t.} \quad & \mathbf{x}^j \text{ is a solution to VI}(\mathcal{K}_j, F(\cdot|\boldsymbol{\theta})),\ \forall\, j \\ & \boldsymbol{\theta} \in \Theta \end{aligned} \tag{12}$$

The MPEC formulation (12) is new to the best of our knowledge. It aims at estimating general equilibrium models with map $F$. In the case of traffic equilibrium, it is standard to assume $s_a$ to be continuous, positive, nondecreasing, and separable, *i.e.* only dependent on the aggregate flow $v_a$. Hence $S = (s_a)_{a \in \mathcal{A}}$ can be expressed as the gradient of the convex *potential function* $z$ defined in (3). We include the prior information on $F$ with the following restrictions on the choice of the parametric model defined in (11):

(i) $\Theta$ is a convex subset of $\mathbb{R}_+^d$.
(ii) $\forall\, i$, $\exists\, S_i = (s_{ia})_{a \in \mathcal{A}} : \mathbb{R}_+^{\mathcal{A}} \to \mathbb{R}_+^{\mathcal{A}}$ continuous, positive, nondecreasing, separable such that $F_i(\mathbf{x}) = \mathbf{Z}^T S_i(\mathbf{Z}\mathbf{x})$.

Hence, $F(\cdot|\boldsymbol{\theta})$ can also be expressed as the gradient of a parametric convex potential function $f(\cdot|\boldsymbol{\theta}) = f_0 + \sum_{i=1}^d \theta_i f_i$, where the basis functions are $f_i(\mathbf{Z}\mathbf{x}) = z_i(\mathbf{Z}\mathbf{x})$ with $z_i(\mathbf{v}) = \sum_a \int_0^{v_a} s_{ia}(u)du$. By convexity, the VI problems in (12) are equivalent to the convex programs MP($\mathcal{K}_j, f(\cdot|\boldsymbol{\theta})$), and are also equivalent to the KKT optimality conditions; we refer, *e.g.*, to [18], [17] for details.

## IV. Constraints as a variational inequality

The program (12) is a MPEC, which is a special class of optimization problems in which a subset of the decision variables satisfy an equilibrium condition. MPECs include programs with VI constraint, *bilevel programs* in which the lower level is an optimization process, and programs with *complementary constraints*. They arise in many practical applications such as Stackelberg games in economics, network design in transportation; see [24].

In the present work, we are interested in solving MPECs in the context of VIs. We consider the general program over the sets $\mathcal{K} \subseteq \mathbb{R}^n$ and $\Theta \subseteq \mathbb{R}^d$, with parametric map $F(\cdot|\boldsymbol{\theta}) : \mathcal{K} \to \mathbb{R}^n$, and upper-level objective $g : \mathbb{R}^n \times \mathbb{R}^d \to \mathbb{R}$:

$$\min_{\boldsymbol{\theta}, \mathbf{x}} g(\mathbf{x}, \boldsymbol{\theta}) \quad \text{s.t.} \quad \begin{aligned} & \mathbf{x} \text{ is a solution to VI}(\mathcal{K}, F(\cdot|\boldsymbol{\theta})) \\ & \boldsymbol{\theta} \in \Theta \end{aligned} \tag{13}$$

A *mathematical program with VI constraint* (MPVI) is a particular case of semi-infinite programs, which are closely related to bilevel programs [33]. When $F(\cdot|\boldsymbol{\theta})$ is the gradient of a potential $f(\cdot|\boldsymbol{\theta})$, a related problem is:

$$\min_{\boldsymbol{\theta}, \mathbf{x}} g(\mathbf{x}, \boldsymbol{\theta}) \quad \text{s.t.} \quad \begin{aligned} & \mathbf{x} \text{ is a solution to OP}(\mathcal{K}, f(\cdot|\boldsymbol{\theta})) \\ & \boldsymbol{\theta} \in \Theta \end{aligned} \tag{14}$$

Both (13) and (14) are difficult to solve because they are in general not convex (so one only hopes to find a local optimum) and it is required to solve a mathematical program just to check feasibility at one point. They are also NP-hard problems even in the linear case [15]. Most of existing approaches utilize the KKT conditions to develop highly specialized algorithms; see [14] and references therein. Using LP duality, we present a simple (approximate) single-level reformulation of (13) with theoretical guarantees, when $\mathcal{K}$ is *polyhedral*, with standard form:

$$\mathcal{K} = \{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{A}\mathbf{x} = \mathbf{b},\, \mathbf{x} \succeq 0\} \tag{15}$$

where $\mathbf{A} \in \mathbb{R}^{m\times n}$ and $\mathbf{b} \in \mathbb{R}^m$. We first consider the system:

$$\begin{aligned} F(\mathbf{x}|\boldsymbol{\theta})^T\mathbf{x} &= \mathbf{b}^T\mathbf{y} \\ \mathbf{A}^T\mathbf{y} &\preceq F(\mathbf{x}|\boldsymbol{\theta}) \end{aligned} \tag{16}$$

***Lemma 1***. *Let $\boldsymbol{\theta} \in \Theta$ and $\mathcal{K}$ polyhedral given by (15). Then a feasible point $\mathbf{x}^\star \in \mathcal{K}$ solves VI($\mathcal{K}, F(\cdot|\boldsymbol{\theta})$) if and only if there exists $\mathbf{y}^\star \in \mathbb{R}^m$ such that $(\mathbf{x}^\star, \mathbf{y}^\star)$ satisfies (16).*

The lemma above is a consequence of LP duality. The proof can be found in [3, Th. 1] and in another form in [17, §1.2.1]. It leads to the single-level formulation of (13) below, which has appeared in [19, §1.3] for toll pricing:

***Theorem 2***. *Suppose $\mathcal{K}$ is polyhedral given by (15). Then the program with VI constraint (13) is equivalent to:*

$$\min_{\boldsymbol{\theta},\mathbf{x},\mathbf{y}} g(\mathbf{x},\boldsymbol{\theta}) \quad \text{s.t.} \quad \begin{aligned} &F(\mathbf{x}|\boldsymbol{\theta})^T\mathbf{x} = \mathbf{b}^T\mathbf{y} \\ &\mathbf{A}^T\mathbf{y} \preceq F(\mathbf{x}|\boldsymbol{\theta}) \\ &\mathbf{x} \in \mathcal{K},\, \boldsymbol{\theta} \in \Theta \end{aligned} \tag{17}$$

To circumvent the disjunctive nature[1] of the constraint $F(\mathbf{x}|\boldsymbol{\theta})^T\mathbf{x} = \mathbf{b}^T\mathbf{y}$, and since models described by VIs only approximate the reality, we define the notion of *approximate equilibria*, when the equality $F(\mathbf{x}|\boldsymbol{\theta})^T\mathbf{x} = \mathbf{b}^T\mathbf{y}$ holds approximately in (16). More precisely, we define the residual:

$$r(\boldsymbol{\theta},\mathbf{x},\mathbf{y}) = F(\mathbf{x}|\boldsymbol{\theta})^T\mathbf{x} - \mathbf{b}^T\mathbf{y} \tag{18}$$

More importantly, we have the following relation between the residual $r$ and the VI problem, due to [8, Th.2]:

***Theorem 3.*** *Let $\boldsymbol{\theta} \in \Theta$, $\epsilon > 0$ and suppose $\mathcal{K}$ is a polyhedron given by (15). Then a point $\hat{\mathbf{x}} \in \mathcal{K}$ satisfies:*

$$F(\hat{\mathbf{x}}|\boldsymbol{\theta})^T(\mathbf{u} - \hat{\mathbf{x}}) \geq -\epsilon,\, \forall\, \mathbf{u} \in \mathcal{K} \tag{19}$$

*if and only if there exists $\mathbf{y}$ with $\mathbf{A}^T\mathbf{y} \preceq F(\hat{\mathbf{x}}|\boldsymbol{\theta})$ such that:*

$$r(\boldsymbol{\theta},\hat{\mathbf{x}},\mathbf{y}) \leq \epsilon \tag{20}$$

*In addition, if $F(\cdot|\boldsymbol{\theta})$ is the gradient of a convex potential function $f(\cdot|\boldsymbol{\theta})$, a point $\hat{\mathbf{x}} \in \mathcal{K}$ satisfying (19) is such that:*

$$f(\hat{\mathbf{x}}) - \min_{\mathbf{x}\in\mathcal{K}} f(\mathbf{x}) \leq \epsilon \tag{21}$$

Hence, a point $\hat{\mathbf{x}} \in \mathcal{K}$ satisfying (19) or (20) is called an $\epsilon$-approximate solution to the VI problem, or an $\epsilon$-approximate equilibrium, and it can be seen as a generalization of the notion of approximate solution to convex programs. This leads us to consider the following *Mathematical program with Dualized VI constraint* (MDVI), which is a penalized version of (17):

$$\begin{aligned} \min_{\boldsymbol{\theta},\mathbf{x},\mathbf{y}} \quad & g(\mathbf{x},\boldsymbol{\theta}) + \lambda\, r(\boldsymbol{\theta},\mathbf{x},\mathbf{y}) \\ \text{s.t.} \quad & \mathbf{A}^T\mathbf{y} \preceq F(\mathbf{x}|\boldsymbol{\theta}) \\ & \mathbf{x} \in \mathcal{K},\, \boldsymbol{\theta} \in \Theta \end{aligned} \tag{22}$$

where $\lambda > 0$ is the weight on the residual $r(\boldsymbol{\theta},\mathbf{x},\mathbf{y})$ to be minimized to obtain an approximate equilibrium. (22) is a reformulation of the standard (13) (or MPEC) and is in general not convex. However, (22) is easier to solve.

## V. A MULTI-CONVEX APPROACH

Convex programs can be solved efficiently by high-quality software packages such at MOSEK, SeDuMi, CPLEX etc.; see [10]. This motivates us to cast (22) into a (multi)-convex optimization framework.

As discussed in Section III, it is preferable to limit ourselves to affine parametric model $F(\cdot|\boldsymbol{\theta})$ (11), where the parameters $\boldsymbol{\theta}$ belong to a convex subset $\Theta \subseteq \mathbb{R}^d$. The computational benefits of these restrictions become clear in our latency inference approach, in which the objective function in (12) is given by:[2]

$$g(\mathbf{x},\boldsymbol{\theta}) = \frac{1}{2}\|\tilde{\mathbf{H}}\mathbf{x} - \mathbf{z}\|^2 + \frac{\mu}{2}\|\boldsymbol{\theta}\|^2 \tag{23}$$

where $\mu\|\boldsymbol{\theta}\|^2$ with $\mu > 0$ is a regularization term. Then (22) is convex with respect to block $(\boldsymbol{\theta},\mathbf{y})$, denoted $\text{MP}_{\boldsymbol{\theta}|\mathbf{x}}$:

$$\begin{aligned} \min_{\boldsymbol{\theta},\mathbf{y}} \quad & g(\mathbf{x},\boldsymbol{\theta}) + \lambda(F(\mathbf{x}|\boldsymbol{\theta})^T\mathbf{x} - \mathbf{b}^T\mathbf{y}) \\ \text{s.t.} \quad & \mathbf{A}^T\mathbf{y} \preceq F(\mathbf{x}|\boldsymbol{\theta}) \\ & \boldsymbol{\theta} \in \Theta \end{aligned} \tag{24}$$

In addition, if $\Theta$ is polyhedral and $\|\cdot\|$ is the Euclidian norm, $\text{MP}_{\boldsymbol{\theta}|\mathbf{x}}$ is a quadratic program. $\text{MP}_{\boldsymbol{\theta}|\mathbf{x}}$ can be seen as an *inverse VI problem* with an *additional cost* $g$, where $(\boldsymbol{\theta},\mathbf{y})$ is computed such that the solution of VI($\mathcal{K}, F(\cdot|\boldsymbol{\theta})$) is approximately $\mathbf{x}$ and such that $\boldsymbol{\theta}$ minimizes some objective $g$. With $g = 0$, the above program has been proposed in [8] for the parametric estimation of the VI problem from *complete observations* of equilibria.

The optimization in $(\mathbf{x},\mathbf{y})$ is the program, denoted $\text{MP}_{\mathbf{x}|\boldsymbol{\theta}}$:

$$\begin{aligned} \min_{\mathbf{x},\mathbf{y}} \quad & g(\mathbf{x},\boldsymbol{\theta}) + \lambda(F(\mathbf{x}|\boldsymbol{\theta})^T\mathbf{x} - \mathbf{b}^T\mathbf{y}) \\ \text{s.t.} \quad & \mathbf{A}^T\mathbf{y} \preceq F(\mathbf{x}|\boldsymbol{\theta}) \\ & \mathbf{x} \in \mathcal{K} \end{aligned} \tag{25}$$

In the context of latency inference, where $g$ is given by (23), $\text{MP}_{\mathbf{x}|\boldsymbol{\theta}}$ is convex if the maps $\mathbf{x} \mapsto F(\mathbf{x}|\boldsymbol{\theta})^T\mathbf{x}$ and $\mathbf{x} \mapsto F(\mathbf{x}|\boldsymbol{\theta})$ are respectively convex and concave over $\mathcal{K}$. The above program has been proposed in [3, Th. 2] as a reformulation of the VI (without the additional cost $g$).

We summarize all the assumptions on $F$ and $g$ and give sufficient conditions for the multi-convexity of (22):

(a) the map $F(\mathbf{x}|\cdot)$ is concave over $\Theta$ for all $\mathbf{x} \in \mathcal{K}$.
(b) the function $g(\mathbf{x},\cdot)$ is convex over $\Theta$ for all $\mathbf{x} \in \mathcal{K}$.
(c) the function $g(\cdot,\boldsymbol{\theta})$ is convex over $\mathcal{K}$ for all $\boldsymbol{\theta} \in \Theta$.

[1]The constraint $F(\mathbf{x}|\boldsymbol{\theta})^T\mathbf{x} = \mathbf{b}^T\mathbf{y}$ prevents the usual constraint qualifications in nonlinear programming to be satisfied, which reduces considerably the applicability of standard nonlinear programming techniques [21], [24]. Numerical issues have been observed in practice in [19].

[2]We suppose we only observe one equilibrium for clarity, but the analysis naturally extends to multiple partially-observed equilibria.

(d) $\mathbf{x} \mapsto F(\mathbf{x}|\boldsymbol{\theta})^T\mathbf{x}$ is convex over $\mathcal{K}$ for all $\boldsymbol{\theta} \in \Theta$.
(e) the map $F(\cdot|\boldsymbol{\theta})$ is concave over $\mathcal{K}$ for all $\boldsymbol{\theta} \in \Theta$.

***Theorem 4.*** *Suppose $\mathcal{K}$ is polyhedral given by (15) and $\Theta$ is a convex subset of $\mathbb{R}^d$. If conditions (a),(b) are satisfied, then $MP_{\boldsymbol{\theta}|\mathbf{x}}$ is a convex program. If conditions (c),(d),(e) are satisfied, then $MP_{\mathbf{x}|\boldsymbol{\theta}}$ is a convex program. If conditions (a) to (e) are satisfied, then (22) is a multi-convex program with blocks $(\mathbf{x}, \mathbf{y})$ and $(\boldsymbol{\theta}, \mathbf{y})$.*

This motivates us to solve (22) via *block coordinate descent* (BCD) with blocks $(\boldsymbol{\theta}, \mathbf{y})$ and $(\mathbf{x}, \mathbf{y})$. Similarly to the so-called Expectation-Maximization algorithm, we alternatively update the estimation or control parameters $\boldsymbol{\theta}$ via $\mathrm{MP}_{\boldsymbol{\theta}|\mathbf{x}}$, and computes the induced response $\mathbf{x}$ via $\mathrm{MP}_{\mathbf{x}|\boldsymbol{\theta}}$, while minimizing the upper-level objective $g$. With the multi-convex structure of the problem, each block can be solved efficiently. Even though convergence of BCD is not well-understood or requires restrictive assumptions, such as the multi-convex structure [35], BCD works well in practice and has been studied extensively [16], [35], [29].

BCD also solves a sequence of convex programs and it is particular desirable in latency inference where the block $\mathrm{MP}_{\mathbf{x}|\boldsymbol{\theta}}$ in the overall problem (26) is separable.

$$\begin{array}{ll} \min & \frac{\lambda}{2}\|\boldsymbol{\theta}\|^2 + \sum_{j=1}^N \{\frac{\mu}{2}\|\tilde{\mathbf{H}}\mathbf{x}^j - \mathbf{z}^j\|^2 + r(\boldsymbol{\theta}, \mathbf{x}^j, \mathbf{y}^j)\} \\ \text{s.t.} & \mathbf{A}^T\mathbf{y}^j \preceq F(\mathbf{x}^j|\boldsymbol{\theta}) \qquad j = 1, \cdots, N \\ & \mathbf{A}\mathbf{x}^j = \mathbf{b}^j,\, \mathbf{x}^j \succeq 0 \qquad j = 1, \cdots, N \\ & \boldsymbol{\theta} \in \Theta \end{array} \tag{26}$$

With $\boldsymbol{\theta}$ fixed, the block $(\mathbf{x}, \mathbf{y}) := ((\mathbf{x}^j)_{j=1}^N, (\mathbf{y}^j)_{j=1}^N)$ is separable into $N$ blocks $(\mathbf{x}^j, \mathbf{y}^j)$, which can be updated *in parallel*. The associated $N$ programs are denoted $\mathrm{MP}_{\mathbf{x}^j|\boldsymbol{\theta}}$.

## VI. Discussion on the multi-convexity

Our approach favors the availability of convex programming solvers instead of developing complex specialized solvers, which has been the focus of the MPEC literature; see [24], [14]. However, Theorem 4 seems to require restrictive conditions. In the context of traffic equilibria, we argue that in our latency inference approach, conditions (a) to (d) rely on standard assumptions. In control (or design), we argue that conditions (b),(c),(d) hold in general. Besides, Theorem 4 gives valuable insights on the applicability of convex programming to MPECs via BCD.

In traffic equilibrium models, it is standard to have positive nondecreasing convex delays $s_a$, hence the map $\mathbf{x}^T F(\mathbf{x}) = S(\mathbf{Z}\mathbf{x})^T\mathbf{Z}\mathbf{x}$ is convex over $\mathcal{K}$. Following our structural estimation methodology, the affine parametric delay $S(\cdot|\boldsymbol{\theta})$ is restricted to be positive nondecreasing convex, which ensures the convexity of the affine parametric map $\mathbf{x}^T F(\mathbf{x}|\boldsymbol{\theta})$. Hence conditions (a),(d) are in general satisfied for latency inference.

We also need conditions (b),(c) to be satisfied. They hold in many applications, *e.g.* in latency inference with $g$ given by (23), in control where one wants to minimize the average delay $\mathbf{x}^T F(\mathbf{x})$ (which is convex in general) plus an additional convex cost $h$ depending on $\boldsymbol{\theta}$; see, *e.g.*, [25], [19].

However, even though condition (a) holds for latency inference, it may not hold for control or design. For example, $S(\mathbf{v}|\cdot)$ is not concave in network optimization [25], where $S(\mathbf{v}|\mathbf{m}) = (s(v_a/m_a))_{a\in\mathcal{A}}$ is parametrized by the arc capacities $\mathbf{m} = (m_a)_{a\in\mathcal{A}}$.

Unfortunately, condition (e) does not hold in general because $F(\cdot|\boldsymbol{\theta})$ is convex. Hence, the constraint $\mathbf{A}^T\mathbf{y} \preceq F(\mathbf{x}|\boldsymbol{\theta})$ in $\mathrm{MP}_{\mathbf{x}|\boldsymbol{\theta}}$ is in general not convex. A solution consists in updating the block $(\mathbf{x}, \mathbf{y})$ by solving approximately $\mathrm{MP}_{\mathbf{x}|\boldsymbol{\theta}}$, with a few steps of, *e.g.*, projected gradient descent, or sequential LPs with piecewise linear approximation of the concave constraint.

## VII. Latency inference and toll pricing

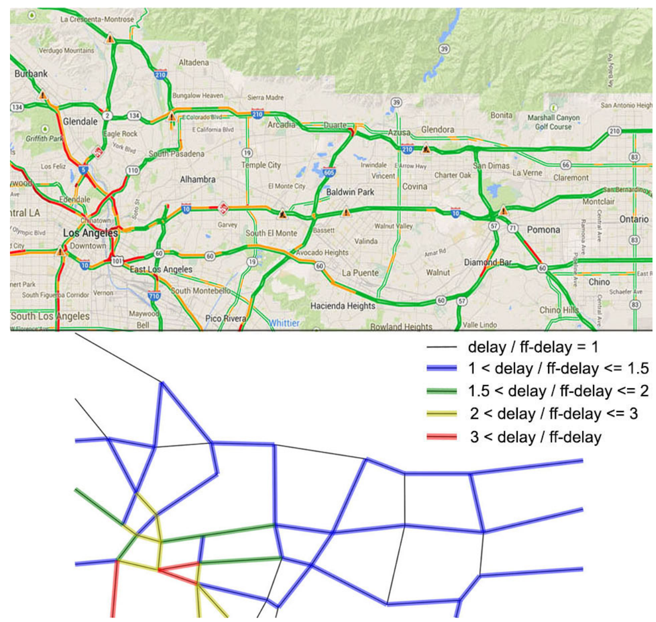


Fig. 1. **Top: Highway network of L.A. in morning rush hour on 2014-06-12 at 9:14 AM from Google Maps; bottom: The network in UE with the resulting delays under demand $\mathbf{b}^4$=1.2*b. The congested area is near central L.A. since it has the higher concentration of employment.**

We consider the highway network of the I-210 corridor near Los Angeles. The roads' characteristics (geometry, capacity, free flow delay) are obtained from OpenStreetMaps. The resulting network has 44 nodes and 122 directed arcs; see Figure 1. The OD demands are based on data from the Census Bureau. They represent a quasi-static morning rush hour model, and are encoded in the demand vector $\mathbf{b}$. We consider for our parametric model the polynomial representation $s_a(\cdot|\boldsymbol{\theta})$ given by (27), with $d_a$ and $m_a$ the free flow delay and capacity on arc $a$:

$$\text{Parametric}: \quad s_a(v_a|\boldsymbol{\theta}) = d_a(1 + \textstyle\sum_{i=1}^6 \theta_i (v_a/m_a)^i) \tag{27}$$

$$\text{True}: \quad s_a^{\text{true}}(v_a) = d_a(1 - \tfrac{3.5}{3} + \tfrac{3.5}{3 - v_a/m_a}) \tag{28}$$

$$\text{BPR}: \quad s_a^{\text{BPR}}(v_a) = d_a(1 + 0.15(v_a/m_a)^4) \tag{29}$$

We consider $N = 4$ partial observations of UE aggregate flows $\mathbf{z}^1, \mathbf{z}^2, \mathbf{z}^3, \mathbf{z}^4 \in \mathbb{R}_+^{\mathcal{A}^{\text{obs}}}$ associated to 4 demand vectors $\mathbf{b}^1, \mathbf{b}^2, \mathbf{b}^3, \mathbf{b}^4$ which are scaled versions of $\mathbf{b}$ with respective factors 0.5, 0.8, 1, 1.2. The measurements are generated by solving (2) with the 'true' delay function (28), chosen to be hyperbolic and similar to the delay (29) estimated by the Bureau of Public Roads (BPR), but not polynomial. The resulting true UE flows $\mathbf{x}^1, \mathbf{x}^2, \mathbf{x}^3, \mathbf{x}^4$, with delays represented in Figure 1, are given as inputs to the observation model $\mathbf{z} = \tilde{\mathbf{H}}\mathbf{x} = \mathbf{H}\mathbf{Z}\mathbf{x}$.

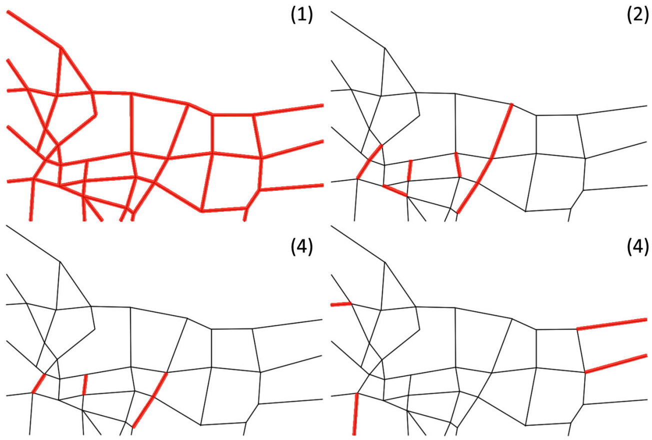


Fig. 2. **The 4 sensor configurations. 1) top left: all arcs are observed, hence $\tilde{\mathbf{H}} = \mathbf{I}\mathbf{Z} = \mathbf{Z}$ and we observe the full aggregate flow $\mathbf{v}$; 2) top right: 10 directed arcs are observed in the congested area; 3) bottom left: 4 directed arcs are observed in the congested area; 4) bottom right: 4 directed arcs are observed at the boundaries of the region, where the inflows are already known from the OD demands.**

We consider the 4 sensor configurations in Figure 2 and solve (MDVI-est) to infer delays for the 4 observation models $\tilde{\mathbf{H}}$, with $F$ verifying conditions (a) to (d) given by:

$$\begin{aligned}
&F(\mathbf{x}|\boldsymbol{\theta}) = \mathbf{Z}^T S_0(\mathbf{Z}\mathbf{x}) + \textstyle\sum_{i=1}^{6} \theta_i \mathbf{Z}^T S_i(\mathbf{Z}\mathbf{x}),\ \boldsymbol{\theta} \in \Theta := \mathbb{R}_+^6\\
&S_0(\mathbf{v}) = (d_a)_{a\in\mathcal{A}} \qquad \text{(constant map)}\\
&S_i(\mathbf{v}) = (d_a(v_a/m_a)^i)_{a\in\mathcal{A}},\ i = 1, \cdots, 6
\end{aligned}$$

where $\mathbf{v} = \mathbf{Z}\mathbf{x}$ is the aggregate flow. We apply the proposed block descent method where each one of the $(\boldsymbol{\theta}, \mathbf{y})$, $(\mathbf{x}^j, \mathbf{y}^j)$, $j = 1, \cdots, N$ blocks is cyclically solved using the Python software package CVXOPT.[3]

*Treatment of the concave constraint:* We first considered different linear approximations of the constraint $\mathbf{A}^T\mathbf{y}^j \preceq F(\mathbf{x}^j|\boldsymbol{\theta})$ in $\mathrm{MP}_{\mathbf{x}^j|\boldsymbol{\theta}}$. However, due to the strong convexity of the hyperbolic delay $s_a$ (28), especially for large values of $v_a$, BCD combined with these techniques failed to converge to satisfactory solutions. The constraint was then relaxed altogether, and we proceeded to update only block $\mathbf{x}^j$ in each $(\mathrm{MP}_{\mathbf{x}^j|\boldsymbol{\theta}})$, independently of $\mathbf{y}^j$, the feasibility of the overall iterates being enforced by the $(\boldsymbol{\theta}, \mathbf{y})$ update via $\mathrm{MP}_{\boldsymbol{\theta}|\mathbf{x}}$.

*Tuning of the parameters:* We solved (26) for $\lambda = 10^{-2}, 10^0, 10^2, 10^4, 10^6$, $\mu = 10^3$ and obtained 5 candidate maps $\hat{S}$, one for each $\lambda$. The set of equilibrium flows $\{\hat{\mathbf{x}}^j\}_{j=1}^4$ induced by each of our candidates was computed, and the best estimate was selected such that the measurement residual $\sum_{j=1}^4 \|\tilde{\mathbf{H}}\hat{\mathbf{x}}^j - \mathbf{z}^j\|^2$ is minimized. The best candidate for each of the 4 sensor configurations, the relative error in aggregate flows $\sum_{j=1}^4 \|\hat{\mathbf{v}}^j - \mathbf{v}^j\|_1 / \sum_{j=1}^4 \|\mathbf{v}^j\|_1$, and the associated $\lambda$ are shown in Figure 3.

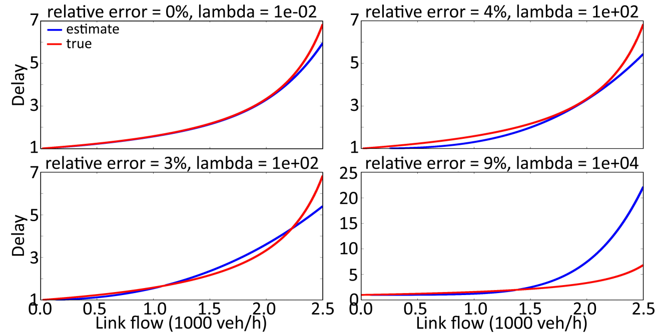


Fig. 3. **Best candidate parameters $\boldsymbol{\theta}$ with the associated graph $1 + \sum_{i=1}^{6} \theta_i x^i$ for each of the 4 sensor configurations shown in Figure 2. 1) top left: when all arcs are observed, the estimated $\boldsymbol{\theta}$ fit 'perfectly' to the true one; 2) top right: when 10 arcs are observed in the congested area, the fit is very good; 3) bottom left: when 4 arcs are observed in the congested area, the fit is very good; 4) bottom right: when 4 arcs at the boundaries are observed, the fit is bad. In case 4, the measurements are 'useless' because they are already contained in the given OD demands, which are part of our prior information, thus having no measurements gives the same results as case 4. The small relative error in case 4 is due to the fact that at least 23% of the aggregate arc flows (the ones at the boundaries) are fixed by the flow conservation equations. Hence comparing the relative errors to the worst case gives another measure of the accuracy of the estimated $\boldsymbol{\theta}$. In our experiments, the relative accuracies are 0/9=%, 4/9=44%, 3/9=33%, 9/9=100%.**

*Toll pricing:* The estimated delay function is used to find a toll $\boldsymbol{\tau} \in \mathbb{R}_+^{\mathcal{A}}$ such that the resulting *tolled UE flow is an un-tolled SO flow*; see eq. (9). This can be formulated as a (MPVI) with parametric map and objective; see [19]:

$$\begin{aligned}
&\hat{F}(\mathbf{x}|\boldsymbol{\tau}) = \mathbf{Z}^T\hat{S}(\mathbf{Z}\mathbf{x}|\boldsymbol{\tau}) \quad \text{with} \quad \hat{S}(\cdot|\boldsymbol{\tau}) = \boldsymbol{\tau} + \hat{S}\\
&\hat{g}(\mathbf{x}, \boldsymbol{\tau}) = \mathbf{x}^T\hat{F}(\mathbf{x}) + \boldsymbol{\tau}^T\mathbf{Z}\mathbf{x} = \hat{c}(\mathbf{Z}\mathbf{x}) + \boldsymbol{\tau}^T\mathbf{Z}\mathbf{x}\\
&\boldsymbol{\tau} \in T := \mathbb{R}_+^{\mathcal{A}}
\end{aligned} \tag{30}$$

where $\hat{F} = F(\cdot|\hat{\boldsymbol{\theta}})$ or $\hat{S} = S(\cdot|\hat{\boldsymbol{\theta}})$ is the estimated map with imputed parameters $\hat{\boldsymbol{\theta}}$ and $\hat{c}(\mathbf{v}) = \hat{S}^T(\mathbf{v})^T\mathbf{v}$ is the estimated total delay function defined in (9) with $\hat{S} = \hat{S}(\cdot|\boldsymbol{\tau} = \mathbf{0})$ is the estimated un-tolled delay. The parametric function $\hat{S}(\mathbf{v}|\boldsymbol{\tau}) = (\tau_a + \hat{s}_a(v_a))_{a\in\mathcal{A}}$ shifts the imputed delays $\hat{s}_a$ by $\tau_a$, the toll incurred on arc $a$. The objective $g$ minimizes the estimated total delay $\hat{c}$ and the toll collected $\boldsymbol{\tau}^T\mathbf{v}$. The system (30) represents our *estimated toll pricing model*.

We reformulate the (13) as a (22) and apply the same methodology as the one used for latency inference, since conditions (a) to (d) also hold. With $c^{\text{true}}(\mathbf{v}) = S^{\text{true}}(\mathbf{v})^T\mathbf{v}$ the true total delay, $\mathbf{v}^{\text{SO}}$ and $\mathbf{v}^{\text{UE}}$ the true SO and UE aggregate flows, we compute the relative loss:

$$\text{relative loss} = \frac{c^{\text{true}}(\mathbf{v}^{\text{res}}) - c^{\text{true}}(\mathbf{v}^{\text{SO}})}{c^{\text{true}}(\mathbf{v}^{\text{UE}}) - c^{\text{true}}(\mathbf{v}^{\text{SO}})} \tag{31}$$

where $\mathbf{v}^{\text{res}}$ is the realized flow under the toll vector $\hat{\boldsymbol{\tau}}$ obtained from our estimated toll pricing model. We also use the estimated equilibrium model to predict the (un-tolled) UE total delay and compute the relative error to the true value. The numerical results are provided in Figure 4:

[3] CVXOPT is free and available at `http://cvxopt.org`. Implementation of the block descent is open source and available at `https://github.com/jeromethai/traffic-estimation-wardrop`.

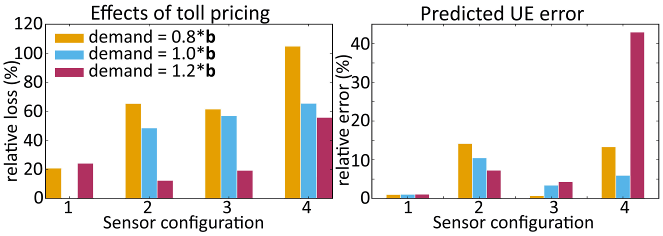


Fig. 4. **Prediction under different OD demands $\mathrm{b}^2, \mathrm{b}^3, \mathrm{b}^4$ using the estimated model based on observations provided by each of the 4 sensor configurations. Because of missing data, the imputed model with delay $\hat{S}$ is different from the true model with delay $S^{\text{true}}$. Left: relative loss for different demands and different sensor configurations; right: relative total delay error for the predicted UE. With full observations (case 1), the estimated model perform well, with relative losses less than 20% and relative errors close to 0%. With partial observations in the congested area (cases 2 and 3), the estimated model performs moderately well, the relative losses and relative errors are under 60% and 15% respectively. With observations at the boundaries (case 4), the estimated model provides a toll (for demand 0.8*b) with relative loss above 100%, *i.e.* the tolled UE is *worse* than the un-tolled UE.**

## VIII. Conclusions

The structural estimation of equilibrium models based on partial observations of equilibria is formulated as a MPEC. Then, we propose a simple single-level reformulation of the general MPEC that can be solved efficiently via a block descent method. The proposed algorithm alternatively updates the parameters of the parametric model and the induced equilibrium to minimize a common objective function. In the context of traffic, the proposed reformulation is in general convex in each block of variable, except for a concave constraint that can be relaxed, thus the block updates can be performed with high-quality convex optimization solvers. This methodology is applied for both the latency inference and control of traffic equilibria to illustrate the sensitivity of toll strategies to errors in estimates.

## Ackowledgement

The authors would like to thank Professor Anil Aswani, Samitha Samaranayake, and Walid Krichene for insightful discussions.

## References

[1] P. Abbeel and A. Ng. Apprenticeship learning via inverse reinforcement learning. *In Proceedings of 21th International Conference on Machine Learning*, 2004.
[2] D. Ackerberg, C. Benkard, S. Berry, and A. Pakes. Econometric Tools for Analyzing Market Outcomes. *Handbook of Econometrics*, 6, 2007.
[3] M. Aghassi, D. Bertsimas, and G. Perakis. Solving asymmetric variational inequalities via convex optimization. *Operations Research Letters 34*, 5:481–490, 2006.
[4] P. Bajari, C. Benkard, and J. Levin. Estimating dynamic models of imperfect competition. *Econometrica*, 75:1331–1370, 2007.
[5] M. Beckmann, C. B. McGuire, and C. B. Winsten. *Studies in the Economics of Transportation*. Cowles Commission Monograph, 1956.
[6] P. Bergendorff, D. W. Hearn, and M. V. Ramana. Congestion Toll Pricing of Traffic Networks. *Network Optimization*, 450:51–71, 1997.
[7] D. P. Bertsekas. *Network Optimization: Continuous and Discrete Methods*. Athena Scientific, Belmont, MA 02178, 1998.
[8] D. Bertsimas, V. Gupta, and I. Ch. Paschalidis. Data-Driven Estimation in Equilibrium Using Inverse Optimization. *Mathematical Programming*, 2014.
[9] S. Boyd, L. El Ghaoui, E. Feron, and V. Balakrishnan. *Linear Matrix Inequalities in Systems and Control Theory*. Philadelphia: SIAM books, 1994.
[10] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, March 8 2004.
[11] D Burton, W. Pulleyblank, and P. Toint. The inverse shortest path problem with upper bounds on shortest path costs. *Network Optimization*, 450:156–171, 1997.
[12] T. Choe, A. Skabardonis, and P. Varaiya. Freeway performance measurement system (PeMS): an operation tool. *81st Annual Meeting Transportation Research Board, Washington, DC*, 2002.
[13] S. C. Dafermos. Traffic equilibrium and variational inequalities. *Transportation Science*, 14:42–54, 1980.
[14] S. Dempe. *Foundations of Bilevel Programming*. Springer, 2002.
[15] X. Deng. Complexity issues in bilevel linear programming. *Multilevel Optimization: Algorithms and Applications*, 20:149–164, 1998.
[16] F. Facchinei and C. Kanzow. Generalized Nash equilibrium problems. *Quart. J. Oper. Res. (4 OR)*, 5:173–210, 2007.
[17] F. Facchinei and J. Pang. *Finite-Dimensional Variational Inequalities and Complementarity Problems*. Springer, New York, 2003.
[18] P. Harker and J. Pang. Finite-dimensional variational inequality and nonlinear complementarity problems: a survey of theory, algorithms and applications. *Mathematical programming*, 48:161–220, 1990.
[19] D. W. Hearn and M. V. Ramana. Solving congestion toll pricing models. *Springer US*, 1998.
[20] G. Iyengar and W. Kang. Inverse conic programming with applications. *Operations Research Letters 33*, 2005.
[21] H. Jiang, D. Ralph, and J. Pang. QPECgen, a MATLAB generator for mathematical programs with quadratic objectives and affine variational inequality constraints. *Computational Optimization and Applications*, 13:25–59, 1999.
[22] A. Keshavarz, Y. Wang, and S. Boyd. Imputing a Convex Objective Function. *IEEE International Symposium on Intelligent Control (ISIC)*, 2011.
[23] R. Little and D. Rubin. *Statistical analysis with missing data*. Wiley, New York, 1987.
[24] Z. Q. Luo, J. S. Pang, and D. Ralph. *Mathematical Programs with Equilibrium Constraints*. Cam, 1996.
[25] P. Marcotte. Network design problem with congestion effects: a case of bilevel programming. *Mathematical Programming*, 34:142–162, 1986.
[26] A. Ng and S. Russell. Algorithms for inverse reinforcement learning. *In Proceedings of 17th International Conference on Machine Learning*, pages 663–670, 2000.
[27] N. Nisam, T. Roughgarden, E. Tardos, and V. V. Vazirani. *Algorithmic Game Theory*. Cambridge University Press, 2007.
[28] M. Patriksson. *The Traffic Assignment Problem - Models and Methods*. VSP, Utrecht, 1994.
[29] M. Razaviyayn, M. Hong, and Z.-Q. Luo. A Unified Convergence Analysis of Block Successive Minimization Methods for Nonsmooth Optimization. *SIAM Journal on Optimization*, 2013.
[30] J. Rust. Structural Estimation of Markov decision processes. *Handbook of Econometrics*, 4:3081–3143, 1994.
[31] W. H. Sandholm. Potential Games with Continuous Player Sets. *Journal of Economic Theory*, 97:81–108, 2001.
[32] Gesualdo Scutari, Daniel P. Palomar, Francisco Facchinei, and Jong-Shi Pang. Convex Optimization, Game Theory, and Variational Inequality Theory. *IEEE Signal Processing Magazine*, 35, 2010.
[33] O. Stein. *Bi-level strategies in semi-infinite programming.* Nonconvex Optimization and its Applications, 71. Academic Publishers, Boston, MA, 2003.
[34] J. G. Wardrop and J. I. Whitehead. Correspondence. Some Theoretical Aspects of Road Traffic Research. *ICE Proceedings: Engineering Divisions 1*, 1952.
[35] Y. Xu and W. Yin. A block coordinate descent method for multi-convex optimization with applications to nonnegative tensor factorization and completion. *SIAM Journal on Imaging Sciences*, 2013.