Where Should Traffic Sensors Be Placed on Highways?

Sebastian A. Nugroho*, Suyash C. Vishnoi†, Ahmad F. Taha‡,◇, Christian G. Claudel†, and Taposh Banerjee‡

***Abstract*—This paper investigates the practical engineering problem of traffic sensors placement on stretched highways with ramps. Since it is virtually impossible to install bulky traffic sensors on each highway segment, it is crucial to find placements that result in optimized network-wide, traffic observability. Consequently, this results in accurate traffic density estimates on segments where sensors are *not* installed. The substantial contribution of this paper is the utilization of control-theoretic observability analysis—jointly with integer programming—to determine traffic sensor locations based on the nonlinear dynamics and parameters of traffic networks. In particular, the celebrated asymmetric cell transmission model is used to guide the placement strategy jointly with observability analysis of nonlinear dynamic systems through Gramians. Thorough numerical case studies are presented to corroborate the proposed theoretical methods and various computational research questions are posed and addressed. The presented approach can also be extended to other models of traffic dynamics.**

***Keywords*—Traffic sensor placement, highway traffic networks, asymmetric cell transmission model, robust observer, observability Gramian, convex integer programming.**

I. Motivation and Paper Contributions

With the development of intelligent transportation systems technologies, numerous sensing methods for traffic data collection have become popular [1], [2]. Fixed sensors such as induction loops and magnetometers allow network operators to obtain high-quality measurements of vehicle density besides other information such as vehicle speeds and flow [1]. Sensors, however, are generally expensive to install and maintain which makes them infeasible for installation throughout the network and covering all segments. This poses a problem for traffic management operations and controls that require knowledge of the traffic state on all segments. Such operations include traffic control tasks such as ramp metering [3], [4], and variable speed limit control [4], [5]. While it may only be feasible to collect data from a fixed number of segments in the network, it is still possible to obtain accurate estimates of the state of traffic on all the segments if the sensors are strategically placed.

The problem of placing traffic sensors on highway networks has been divided into two main categories in the literature, one involving estimation and the other involving observability. Under the former category, the objective is to determine the optimal placement of sensors that minimizes the estimation error for unmeasured quantities such as travel time [6]–[8], OD matrix [9], [10], link flows [11]. Under observability, the literature is further divided based on the type of observability that is considered, full or partial observability. A fully observable system is one in which all the states are observable given the available measurements from the sensors. Different observability problems considered in this category are link flow observability [12]–[16], route flow observability [17], [18] and OD flow observailibty [19]–[21]. The general idea is to determine a sensor placement configuration that ensures full observability of the system. Partial observability implies that not all states are observable given the set of available sensor measurements. Under this category, the literature focuses on determining such sensor locations that maximize the number of observable states while obeying some budget constraints. This concept is attractive in cases where the number of sensors required to achieve full observability is too large. Interested readers can refer to [2], [16], [22], [23] and references therein for a detailed literature review of the aforementioned categories of sensor placement problem. In the current work, we focus on determining the optimal sensor placement to achieve full observability of the system while maintaining a balance between the number of sensors and the degree of observability of the system. Note that the degree of observability is a quantitative measure of the quality of state estimates that can be obtained by using a certain sensor placement configuration and is not related to the idea of partial observability. Unlike the aforementioned studies that utilize the relationship between the various flows in the network such as link and path flows, this study uses a traffic dynamics model to determine the relationship between various state variables. Also, the states considered here are traffic densities instead of flows.

A study that is closer to the work presented in this paper is [24] which also considers a traffic dynamics model and applies a control theoretic approach to determine optimal sensor locations on unconnected highway segments. This approach is extended to study sensor placement on complex traffic networks in [25], and then adopted in [26] for studying the observability of highway traffic with Lagrangian sensors. These studies linearize the nonlinear traffic dynamics around a steady-state traffic flow. The major drawback of such approaches is that the linearized dynamics are valid only around the specific states. Furthermore, these studies consider the Greenshield's fundamental diagram [27] which is inferior in modeling traffic flows when compared with the triangular and trapezoidal fundamental diagrams which are often considered in the implementation

*Department of Electrical Engineering and Computer Science, University of Michigan, 1301 Beal Ave., Ann Arbor, MI 48109 (snugroho@umich.edu).

†Department of Civil, Architectural, and Environmental Engineering, The University of Texas at Austin, 301 E. Dean Keeton St. Stop C1700, Austin, TX 78712 (scvishnoi@utexas.edu, christian.claudel@utexas.edu).

‡Department of Electrical and Computer Engineering, The University of Texas at San Antonio, 1 UTSA Circle, San Antonio, TX 78249 (ahmad.taha@utsa.edu, taposh.banerjee@utsa.edu).

◇Corresponding author.

This work is partially supported by the Valero Energy Corporation and National Science Foundation (NSF) under Grants 1636154, 1728629, 1917164, and 1917056.

of the *cell transmission model* (CTM) [28], [29]. [1] does deal with optimal sensor placement and density reconstruction while considering the CTM model but to simplify things it still linearizes the model and thus has the same drawback as above.

There exist some other papers studying the observability properties of traffic systems considering a traffic dynamics model, albeit do not formally study the sensor placement. The observability properties related to the switching modes model [30], which is a piecewise-linearized version of the CTM [29] is studied in [31]. In [32], the authors propose a ramp-metering controller based on a switching modes scheme and study the observability properties of the considered freeway system. A new piecewise affine system model based on CTM is developed in [33], where its observability is investigated. These studies however, linearize the traffic dynamics and suffer the same drawback.

This paper investigates the sensor placement problem for highway networks with ramps having nonlinear traffic dynamics. The traffic model is built using the *asymmetric CTM* (ACTM) [28], [34], [35] (a close variant of the CTM) and the triangular fundamental diagram, and thus is also an improvement over [24]–[26] in terms of modeling traffic dynamics, to remove the dependence of optimal sensor placements on assumptions about traffic states. The traffic sensor placement is addressed via observability analysis based on the nonlinear traffic model, which is different from the ones used in the aforementioned papers.

There exist several approaches in the literature to quantify observability for nonlinear systems. The two most prevalent approaches include utilizing the concept from differential embedding and Lie derivative [36], [37], while the other is based on empirical observability Gramian [38]. In the context of sensor placement for nonlinear systems, four different approaches have been proposed recently. The first approach leverages empirical observability Gramian to find the best set of sensors that maximizes some metrics on the Gramian matrix—see [39]–[41]. Another approach is proposed in [42] where the observability Gramian around a certain initial state is constructed through a moving horizon estimation (MHE) framework. As such, the sensor selection problem is posed as an integer programming (IP) problem which objective is maximizing the logarithmic determinant of the Gramian. The third approach, developed in [43], introduces a randomized algorithm for dealing with the sensor placement problem and accordingly, theoretical bounds for eigenvalue and condition number of observability Gramian are proposed. The last and most recent approach is established in [44] where the authors make use of the numerous observer designs for some classes of nonlinear systems posed as semidefinite programs (SDP). The resulting problem, which combines the sensor selection *and* observer gain synthesis, is posed as a nonconvex mixed-integer SDP, which can be cast into a convex one through some relaxation/reformulation techniques.

Herein, based on the nonlinear traffic model developed using ACTM, we formulate the sensor placement problem using traffic observability analysis. The resulting problem—categorized as convex IP—is solved via an integer branch-and-bound (BnB) algorithm and therefore, optimal traffic sensor locations can be obtained. Notice that the use of nonlinear traffic model and observability analysis result in traffic sensor locations that are valid for various traffic conditions. The novelties made in this paper are listed as follows:

- We present a discrete-time nonlinear state-space model for stretched highway networks having multiple on- and off-ramps based on ACTM. In addition, we analytically compute the corresponding Lipschitz constant for the nonlinear counterparts. The formulated state-space model along with Lipschitz constant pave the way for implementing various control-theoretic approaches to solve control and estimation problems prevailing in highway traffic networks.
- We leverage the concept of observability for nonlinear systems to construct the traffic sensor placement problem, which is equivalent to maximizing the determinant and trace of observability Gramian matrix given the number of allocated traffic sensors. To the best of our knowledge, this is the first attempt to solve traffic sensor placement problem based on the degree of observability for the nonlinear traffic dynamics.
- We propose a corresponding robust observer framework for Lipschitz nonlinear discrete-time systems developed using the concept of $\mathcal{L}_\infty$ stability for traffic density estimation purpose. By using the optimal configuration of traffic sensors, the computation of stabilizing observer gain matrix can be cast as a SDP and as such, can be solved efficiently.
- We verify the effectiveness of our approach to solve the traffic sensor placement problem through various case studies. Specifically, we *(a)* study the performance and computational efficiency of determinant and trace functions in estimating the system's initial state, *(b)* assess their impact on traffic density estimation performance under noisy conditions, and *(c)* compare their results relative to randomized and uniform sensor placement strategies.

The remainder of the paper is organized as follows. In Section II, we provide the mathematical modeling of stretched highway with ramps using ACTM, which result in a nonlinear state-space form. Next, Section III discusses our strategy for addressing traffic sensor placement based on system's observability. Section IV develops a simple robust observer design for traffic density estimation purpose. The proposed approach is extensively tested in Section V for solving the traffic sensor placement problem and density estimation via numerous case studies. Finally, Section VI concludes the paper.

Paper's Notation: Let $\mathbb{N}$, $\mathbb{R}$, $\mathbb{R}_{++}$, $\mathbb{R}^n$, and $\mathbb{R}^{p\times q}$ denote the set of natural numbers, real numbers, positive real numbers, and real-valued row vectors with size of n, and p-by-q real matrices respectively. $\mathbb{S}^m$ denotes the set of symmetric matrices with the dimension of m. Specifically, $\mathbb{S}^m_+$ and $\mathbb{S}^m_{++}$ denotes the set of positive semi-definite and positive definite matrices respectively. For any vector $z \in \mathbb{R}^n$, $\|z\|_2$ denotes its Euclidean norm, i.e. $\|z\|_2 = \sqrt{z^\top z}$, where $z^\top$ is the transpose of z. The symbol $\otimes$ denotes the Kronecker product where $\det(\boldsymbol{A})$ and $\mathrm{trace}(\boldsymbol{A})$ return the determinant and trace of matrix $\boldsymbol{A}$. For simplicity, the notation '$*$' denotes terms induced by symmetry in symmetric block matrices. Tab. I provides the nomenclature utilized in this paper.

Note about Supplemental Document: The appendices of this paper, containing detailed mathematical derivations and proofs,

Table I
PAPER NOMENCLATURE: PARAMETER, VARIABLE, AND SET DEFINITIONS.

Notation	Description
Ω	the set of highway segments on the stretched highway $\Omega = \{1, 2, \ldots, N\}$, $N := \|\Omega\|$
Ω_I	the set of highway segments with on-ramps $\Omega_I = \{1, 2, \ldots, N_I\}$, $N_I := \|\Omega_I\|$
Ω_O	the set of highway segments with off-ramps $\Omega_O = \{1, 2, \ldots, N_O\}$, $N_O := \|\Omega_O\|$
$\hat{\Omega}$	the set of on-ramps, $\hat{\Omega} = \{1, 2, \ldots, N_I\}$, $N_I = \|\hat{\Omega}\|$
$\check{\Omega}$	the set of off-ramps, $\check{\Omega} = \{1, 2, \ldots, N_O\}$, $N_O = \|\check{\Omega}\|$
T	duration of each time step
l	length of each segment, on-ramp, and off-ramp
$\rho_i[k]$	traffic density on Segment $i \in \Omega$ at time kT, $k \in \mathbb{N}$
$q_i[k]$	traffic flow from Segment $i \in \Omega$ into the next segment
$\delta_i[k], \sigma_i[k]$	demand and supply functions for Segment $i \in \Omega$
$\hat{\rho}_i[k]$	traffic density on the on-ramp of Segment $i \in \Omega_I$
$r_i[k]$	traffic flow into Segment $i \in \Omega_I$ from the on-ramp
$\hat{r}_i[k]$	traffic flow into the on-ramp of Segment $i \in \Omega_I$
$\hat{\delta}_i[k], \hat{\sigma}_i[k]$	demand and supply functions for the on-ramp of Segment $i \in \Omega_I$
$\check{\rho}_i[k]$	traffic density on the off-ramp of Segment $i \in \Omega_O$
$s_i[k]$	traffic flow from Segment $i \in \Omega_O$ into the off-ramp
$\check{\delta}_i[k], \check{\sigma}_i[k]$	demand and supply functions for the off-ramp of Segment $i \in \Omega_O$
$\check{s}_i[k]$	traffic flow from the off-ramp of Segment $i \in \Omega_O$
$f_{in}[k]$	traffic wanting to enter Segment 1 of the highway
$f_{out}[k]$	traffic that can leave Segment N of the highway
$\hat{f}_i[k]$	traffic wanting to enter the on-ramp of Segment $i \in \Omega_I$
$\check{f}_i[k]$	traffic that can leave the off-ramp of Segment $i \in \Omega_O$
$\beta_i[k]$	split ratio for the off-ramp of Segment $i \in \Omega_O$, where $\beta_i[k] \in [0, 1]$
$\xi_i[k]$	occupancy parameter for the on-ramp of Segment $i \in \Omega_I$ where $\xi_i[k] \in [0, w_c]$
v_f	free-flow speed
w_c	congestion wave speed
ρ_m	maximum density
ρ_c	critical density

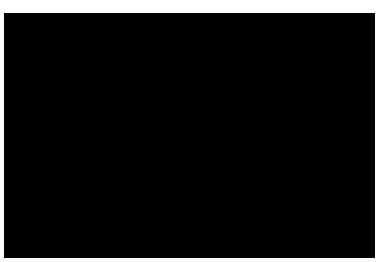


Figure 1. Triangular fundamental diagram.

are all included in a supplemental document attached with this manuscript submission.

II. NONLINEAR DISCRETE-TIME MODELING OF TRAFFIC NETWORKS WITH RAMPS

This section presents the discrete-time modeling of traffic dynamics on a stretched highway with arbitrary number and location of ramps. To that end, here we utilize the *Lighthill-Whitman-Richards* (LWR) Model [45], [46] for traffic flow. In this paper, the relationship between traffic density and flux is given by the triangular-shaped fundamental diagram which has been extensively used in the literature [28]. It is depicted in Fig. 1 and is constructed as

$$q\left(\rho(t,d)\right) = \begin{cases} v_f \rho(t,d), & \text{if } 0 \le \rho(t,d) \le \rho_c \\ w_c\left(\rho_m - \rho(t,d)\right), & \text{if } \rho_c \le \rho(t,d) \le \rho_m. \end{cases} \tag{1}$$

where t and d denote the time and distance; $\rho(t,d)$ denotes the traffic density (vehicles/distance) and $q(t,d)$ denotes the traffic flux (vehicles/time). To represent the traffic dynamics as a series of difference, state-space equations—a useful bookkeeping for the ensuing discussions—we discretize the LWR Model with respect to both space and time (this is also referred to as the Godunov discretization). This approach allows the highway of length L to be divided into segments (cells) of equal length l and the traffic networks model to be represented by discrete-time equations. These segments form both the highway and the attached ramps. Throughout the paper, the segments forming the highway are referred to as mainline segments. We assume that the highway is split into N mainline segments.

To ensure computational stability, the Courant-Friedrichs-Lewy condition (CFL) given as $v_f T l^{-1} \le 1$ has to be satisfied. Since each segment is of the same length l, then we have $\rho(t,d) = \rho(kT, l)$, where $k \in \mathbb{N}$ represents the discrete-time index. For simplicity of notation, from here on $\rho(kT, l)$ is simply written as $\rho[k]$.

As mentioned earlier, traffic is modeled using the ACTM which is originally given in [28], [34], [35]. The ACTM is a variant of the CTM that departs from the CTM in its treatment of asymmetric merge junctions such as the on-ramp-highway junctions. Unlike the CTM, it assumes separate allocations of the available space on the highway for traffic from each merging branch, which allows for comparatively simple flow conservation equations at those merges than the original CTM. A variation to the original ACTM is introduced in the modeling of the ramps which here are treated as normal segments rather than point queues as in the original approach. The approach in this paper is similar to that in [47]. Herein, we define new functions referred to as the demand function $\delta_i[\cdot]$ and the supply function $\sigma_i[\cdot]$ to simplify the ensuing expressions. The demand function equals the traffic flux leaving Segment i through the highway assuming that the next segment has infinite storage. The supply function equals the traffic flux that can enter Segment i through the highway assuming that the previous segment has infinite storage. These functions are constructed from the triangular FD given in (1). The demand function $\delta_i[\cdot]$, with and without off-ramp, can be written as

$$\delta_i[k] = \begin{cases} \min\left(\bar{\beta}_i[k] v_f \rho_i[k], \bar{\beta}_i[k] v_f \rho_c, \frac{\bar{\beta}_i[k]}{\beta_i[k]} \check{\sigma}_i[k]\right), & \text{if } i \in \Omega_O \\ \min\left(v_f \rho_i[k], v_f \rho_c\right), & \text{if } i \in \Omega \setminus \Omega_O. \end{cases} \tag{2}$$

Here $\check{\sigma}_i[k]$ can be given as

$$\check{\sigma}_i[k] = \min\left(w_c(\rho_m - \check{\rho}_i[k]), v_f \rho_c\right). \tag{3}$$

In (2), the split-ratio $\beta_i[k]$ relates the traffic flow from Segment i into its off-ramp with the traffic flow from Segment i into the

Figure 2. Consecutive segments of the highway with ramp connections.

next segment such that

$$s_i[k] = \beta_i[k](s_i[k] + q_i[k]) \implies s_i[k] = \frac{\beta_i[k]}{\bar{\beta}_i[k]} q_i[k]. \quad (4)$$

where we define $\bar{\beta}_i[k] \triangleq 1 - \beta_i[k]$ to simplify the equations. Since $\frac{\bar{\beta}_i[k]}{\beta_i[k]} v_f \rho_c > \bar{\beta}_i v_f \rho_c$, therefore in (2) we can also simply write $\check{\sigma}_i[k] = w_c(\rho_m - \check{\rho}_i[k])$.

Similarly, the supply function $\sigma_i[k]$, with and without on-ramp, can be represented as

$$\sigma_i[k] = \begin{cases} \min\left(w_c(\rho_m - \rho_i[k]), v_f \rho_c\right) - r_i[k], & \text{if } i \in \Omega_I \\ \min\left(w_c(\rho_m - \rho_i[k]), v_f \rho_c\right), & \text{if } i \in \Omega \setminus \Omega_I, \end{cases} \quad (5)$$

where $r_i[k]$ is described by the following equation

$$r_i[k] = \min\left(v_f \hat{\rho}_i[k], \xi_i(\rho_m - \rho_i[k]), \frac{\xi_i}{w_c} v_f \rho_c\right). \quad (6)$$

Here $v_f \hat{\rho}_i[k]$ is the demand of the on-ramp in free-flow, and $\frac{\xi_i[k]}{w_c}$ is the fraction of the flow that Segment i can receive from its on-ramp. The upstream and downstream flows for any mainline Segment i are given as

$$q_{i-1}[k] = \min\left(\delta_{i-1}[k], \sigma_i[k]\right) \quad (7a)$$

$$q_i[k] = \min\left(\delta_i[k], \sigma_{i+1}[k]\right), \quad (7b)$$

The discrete-time flow conservation equation for a mainline segment with both on- and off-ramp can then be written as

$$\rho_i[k+1] = \rho_i[k] + \frac{T}{l}\left(q_{i-1}[k] + r_i[k] - q_i[k] - s_i[k]\right) \quad (8)$$

Equations for segments with only one or no ramp can be written by removing the respective flow terms. The flows and conservation equations for the ramps can also be written similar to the mainline segments. The detailed illustration for the above model is given in Fig. 2.

In the present study, we assume an arbitrary structure of the highway with respect to the positioning of the ramps on different mainline segments. In that, we assume that every mainline segment having no ramps is followed by a segment having an on-ramp, which is followed by a segment having an off-ramp, and which is followed by a segment having no ramps and this structure repeats throughout the highway. Additionally, we assume that the first and last mainline segments have no ramps. For example, a highway split into seven segments has an on-ramp on Segments 2 and 5, an off-ramp on Segments 3 and 6, and no ramps on Segments 1, 4 and 7. Fig. 3 gives a schematic representation of the structure of the highway considered in this paper. Note that this structure of the highway is arbitrarily chosen for experimentation of the methodology presented in this work. It is easy to consider a different highway structure as per an actual site where the method may be used.

In this paper we assume that the upstream demand at Segment 1 and downstream supply at Segment N are known, denoted by

Figure 3. Schematic of the highway structure considered in this study.

$\delta_0[k] = f_{\text{in}}[k]$ and $\sigma_{N+1}[k] = f_{\text{out}}[k]$ respectively. Similarly, we assume that all the on-ramp demands $\hat{f}_i, i \in \Omega_I$ and off-ramp supplies $\check{f}_i, i \in \Omega_O$ are also known. These known demand and supply values are lumped into the input vector defined as

$$\boldsymbol{u}[k] = [f_{\text{in}} \ f_{\text{out}} \ \ldots \ \hat{f}_j \ \ldots \ \check{f}_l \ \ldots]^\top \in \mathbb{R}^{2+N_I+N_O},$$

where $j \in \Omega_I$ and $l \in \Omega_O$. The state vector can be defined as

$$\boldsymbol{x}[k] := [\rho_i[k] \ \ldots \ \hat{\rho}_j[k] \ \ldots \ \check{\rho}_l[k] \ \ldots]^\top \in \mathbb{R}^{N+N_I+N_O},$$

for which $i \in \Omega$, $j \in \Omega_I$ and $l \in \Omega_O$. Here, k is same as the time-index of the simulation where each time step is of duration T. The estimation interval is thus equal to T. The equations describing state vector evolution can be be divided into categories:

- $i \in \Omega \setminus \Omega_I \cup \Omega_O$

$$x_i[k+1] = x_i[k] + \frac{T}{l}\Big(\min\left(\delta_{i-1}[k], \sigma_i[k]\right) - \min\left(\delta_i[k], \sigma_{i+1}[k]\right)\Big) \quad (9a)$$

- $i \in \Omega_I$, $j = N + \bar{j}, \bar{j} \in \hat{\Omega}$

$$x_i[k+1] = x_i[k] + \frac{T}{l}\Big(\min\left(\delta_{i-1}[k], \sigma_i[k]\right) - \min\left(\delta_i[k], \sigma_{i+1}[k]\right) + \min\left(v_f x_j[k], \xi_i[k](\rho_m - x_i[k]), \frac{\xi_i[k]}{w_c} v_f \rho_c\right)\Big) \quad (9b)$$

- $i \in \Omega_O$, $j = N + N_I + \bar{j}, \bar{j} \in \check{\Omega}$

$$x_i[k+1] = x_i[k] + \frac{T}{l}\Big(\min\left(\delta_{i-1}[k], \sigma_i[k]\right) - \frac{1}{\bar{\beta}_i[k]} \min\left(\delta_i[k], \sigma_{i+1}[k]\right)\Big) \quad (9c)$$

- $\bar{i} \in \hat{\Omega}, i = N + \bar{i}, j \in \Omega_I$

$$x_i[k+1] = x_i[k] + \frac{T}{l}\Big(\hat{r}_j[k] - \min\left(v_f x_i[k], \xi_j[k](\rho_m - x_j[k]), \frac{\xi_j[k]}{w_c} v_f \rho_c\right)\Big) \quad (9d)$$

- $\bar{i} \in \check{\Omega}, i = N + N_I + \bar{i}, j \in \Omega_O$

$$x_i[k+1] = x_i[k] + \frac{T}{l}\Big(\frac{\beta_j[k]}{\bar{\beta}_j[k]} \min\left(\delta_j[k], \sigma_{j+1}[k]\right) - \check{s}_j[k]\Big) \quad (9e)$$

Proposition 1. *The evolution of traffic density described in (9) can be written in a compact state-space form as follows*

$$\boldsymbol{x}[k+1] = \boldsymbol{A}\boldsymbol{x}[k] + \boldsymbol{G}\boldsymbol{f}(\boldsymbol{x}, \boldsymbol{u}) + \boldsymbol{B}_{\mathbf{u}}\boldsymbol{u}[k], \quad (10)$$

where $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ for $n := N + N_I + N_O$ represents the linear dynamics of the system, $\boldsymbol{B}_{\mathbf{u}} \in \mathbb{R}^{n \times m}$ for $m := 2 + N_I + N_O$ represents the way external inputs affecting the system, $\boldsymbol{f} : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^g$ is a vector valued function representing nonlinearities in (9), and $\boldsymbol{G} \in \mathbb{R}^{n \times g}$ is a matrix representing the distribution of nonlinearities.

The proof and the structures of the matrices and function in (10)

are all provided in Appendix A of the supplemental document attached with this manuscript submission. The above result is important as it allows us to perform optimal sensor placement for the traffic dynamical system defined by (1)–(8), as well as to develop a robust state observer. This also allows for other control-theoretic studies to use this state-space model for other traffic engineering applications. The next section presents our strategy to address the optimal sensor placement problem for the nonlinear traffic dynamical system derived above.

III. Observability-based Sensor Placement

In this section we discuss our approach for addressing the traffic sensor placement problem. The traffic dynamics with measurements can be expressed as

$$\boldsymbol{x}[k+1] = \boldsymbol{A}\boldsymbol{x}[k] + \boldsymbol{G}\boldsymbol{f}(\boldsymbol{x},\boldsymbol{u},k) + \boldsymbol{B}_{\mathbf{u}}\boldsymbol{u}[k] \tag{11a}$$

$$\boldsymbol{y}[k] = \boldsymbol{\Gamma}\boldsymbol{C}\boldsymbol{x}[k]. \tag{11b}$$

In the above model, we introduce $\boldsymbol{y} \in \mathbb{R}^p$ to represent the vector of measurements, which corresponds to all of the highway segments equipped with sensors measuring the density. The matrix $\boldsymbol{C} \in \mathbb{R}^{p\times n}$ in (11b) is useful to determine the placement of traffic sensors, whereas $\boldsymbol{\Gamma} := \mathrm{Diag}(\boldsymbol{\gamma})$ with $\boldsymbol{\gamma} \in \{0,1\}^p$ represents the selection of sensors—that is, $\gamma_i = 1$ if the i-th highway segment is measured (this consequently leads to nonzero row i in $\boldsymbol{C}$) and $\gamma_i = 0$ otherwise. It is assumed throughout the paper that every trajectory of (11) lie in the domain of interest $\boldsymbol{\Omega} := \boldsymbol{\mathcal{X}} \times \boldsymbol{\mathcal{U}}$. In the context of traffic dynamics, $\boldsymbol{\mathcal{X}}$ represents the operating region of the states $\boldsymbol{x}$. Since $\boldsymbol{x}$ represents the densities, then $\boldsymbol{\mathcal{X}} := [0,\rho_m]^n$. Likewise, the set $\boldsymbol{\mathcal{U}}$ can be constructed as $\boldsymbol{\mathcal{U}} := [0, v_f\rho_c]^m$, which represents the set of admissible inputs.

Having described system (11) with sensor placement, we now state the paper's major computational objective: from a given possible set of traffic sensors $\mathcal{G}_\gamma$ such that $\boldsymbol{\gamma} \in \mathcal{G}_\gamma$, find the best (or *optimal*) sensor configuration $\boldsymbol{\gamma}^*$ such that system (11) is observable, i.e., the system's initial state $\boldsymbol{x}_0 := \boldsymbol{x}[0]$ can be uniquely determined from a finite set of measurements. In this paper, we opt to formulate the traffic sensor placement using the concept of observability through MHE framework developed in [42]. The framework utilizes a series of past measurement data to estimate the initial state at each time window. The reasons of pursuing this approach are two-fold. First, as argued in [42], this approach is considerably more scalable than using empirical observability Gramian and second, we experience numerical issues in applying the method from [44]. With that in mind, we consider the first N observation window (or N discrete measurements) of system (11) expressed in the equation below

$$\boldsymbol{h}(\boldsymbol{\gamma},\boldsymbol{x}_0) := \tilde{\boldsymbol{y}} - \boldsymbol{g}\left(\boldsymbol{\gamma},\boldsymbol{x}_0\right). \tag{12}$$

In (12), the mapping $\boldsymbol{g} : \mathbb{N}^p \times \mathbb{R}^n \to \mathbb{R}^{Np}$ is given as

$$\boldsymbol{g}^\top(\boldsymbol{\gamma},\boldsymbol{x}_0) := \Big[(\boldsymbol{\Gamma}\boldsymbol{C}\boldsymbol{x}[0])^\top \ (\boldsymbol{\Gamma}\boldsymbol{C}\boldsymbol{x}[1])^\top \ldots (\boldsymbol{\Gamma}\boldsymbol{C}\boldsymbol{x}[N-1])^\top\Big]. \tag{13}$$

Note that $\boldsymbol{g}(\cdot)$ defined above is indeed a function of $\boldsymbol{x}_0$ since for any k such that $0 < k \le N-1$ then we have

$$\boldsymbol{x}[k] = \boldsymbol{A}^k\boldsymbol{x}[0] + \sum_{j=0}^{k-1}\boldsymbol{A}^{k-1-j}\left(\boldsymbol{G}\boldsymbol{f}(\boldsymbol{x},\boldsymbol{u},j) + \boldsymbol{B}_{\mathbf{u}}\boldsymbol{u}[j]\right). \tag{14}$$

Remark 1. *The function $\boldsymbol{g}(\cdot)$ in (13) is in fact also dependent on input vector $\boldsymbol{u}$, since the k-th state $\boldsymbol{x}[k]$, as it is seen from (14), relies on $\boldsymbol{\mathcal{U}}_s(k-1)$ defined as*

$$\boldsymbol{\mathcal{U}}_s(k-1) := \{\boldsymbol{u}[j]\}_{j=0}^{k-1}, \text{ where } \boldsymbol{u}[j] \in \boldsymbol{\mathcal{U}}, \ \forall j = 0,\ldots,k-1,$$

i.e., the k-th sequence of control inputs. The reason why $\boldsymbol{g}(\cdot)$ only takes $\boldsymbol{\gamma}$ and $\boldsymbol{x}_0$ as arguments is due to the assumption that $\boldsymbol{\mathcal{U}}_s(i)$ is known for each time index i, thus the dependence of $\boldsymbol{g}(\cdot)$ towards input vector $\boldsymbol{u}$ can be ignored for simplicity.

The term $\tilde{\boldsymbol{y}} \in \mathbb{R}^{Np}$ in (12) denotes the stacked N measurements data constructed as

$$\tilde{\boldsymbol{y}}^\top := \left[\tilde{\boldsymbol{y}}[0]^\top \ \tilde{\boldsymbol{y}}[1]^\top \ \ldots \ \tilde{\boldsymbol{y}}[N-1]^\top\right].$$

Obviously, for every initial condition of the system $\boldsymbol{x}_0 \in \boldsymbol{\mathcal{X}}$, it holds that $\boldsymbol{h}(\boldsymbol{\gamma},\boldsymbol{x}_0) = 0$ such that $\tilde{\boldsymbol{y}} = \boldsymbol{g}\left(\boldsymbol{\gamma},\boldsymbol{x}_0\right)$. Hence, for a fixed $\boldsymbol{\gamma}$, the mapping $\boldsymbol{g}(\cdot)$ maps the initial state into the N measurements output. The observability of system (11) with respect to $\boldsymbol{g}(\cdot)$ is formally defined as follows [48].

Definition 1. *The system (11) with a prescribed $\boldsymbol{\gamma}$ is uniformly observable in $\boldsymbol{\mathcal{X}}$ if, for all admissible inputs $\boldsymbol{u}[k] \in \boldsymbol{\mathcal{U}}$, there exists a finite $N > 0$ such that the mapping $\boldsymbol{g}\left(\boldsymbol{\gamma},\boldsymbol{x}_0\right)$ defined in (13) is injective (one-to-one) with respect to $\boldsymbol{x}_0 \in \boldsymbol{\mathcal{X}}$.*

Note that, from Definition 1, if $\boldsymbol{g}\left(\boldsymbol{\gamma},\boldsymbol{x}_0\right)$ is injective with respect to $\boldsymbol{x}_0$, then $\boldsymbol{x}_0$ can be uniquely determined from the set of measurements $\tilde{\boldsymbol{y}}$. A sufficient condition for $\boldsymbol{g}(\cdot)$ to be injective is that the Jacobian of $\boldsymbol{g}(\cdot)$ around $\boldsymbol{x}_0$, denoted by $\boldsymbol{J}_w(\cdot)$, is full rank [48]. In fact, it is not difficult to show that, if system (11) has no nonlinear counterparts, then the Jacobian matrix $\boldsymbol{J}_w(\cdot)$ reduces to the N-step observability matrix for the linear dynamics.

In this work, we use the concept of observability Gramian to quantify the system's osbervability for a given set of sensors. The Gramian $\boldsymbol{G}_v$ (or Gram matrix) for some finite-dimensional real k vectors $\boldsymbol{v}_i \in \mathbb{R}^n$ where $i = 1,2,\ldots,k$ is constructed as $\boldsymbol{G}_v := \boldsymbol{V}^\top\boldsymbol{V}$ where $\boldsymbol{V} := \left[\boldsymbol{v}_1 \,|\, \boldsymbol{v}_2 \,\ldots\, |\, \boldsymbol{v}_k\right]$. The observability of a discrete-time linear time-invariant system of the form

$$\boldsymbol{x}[k+1] = \boldsymbol{A}\boldsymbol{x}[k], \quad \boldsymbol{y}[k] = \boldsymbol{C}\boldsymbol{x}[k], \tag{15}$$

where $\boldsymbol{x}$ is the state and $\boldsymbol{y}$ is the output can be determined from the associated Gramian. The corresponding N-step observability Gramian is given as

$$\boldsymbol{\mathcal{W}}_o := \boldsymbol{\mathcal{O}}_N^\top\boldsymbol{\mathcal{O}}_N = \sum_{i=0}^{N-1}(\boldsymbol{A}^i)^\top\boldsymbol{C}^\top\boldsymbol{C}\boldsymbol{A}^i,$$

where $\boldsymbol{\mathcal{O}}_N$ denotes the N-step observability matrix. The system (15) is then observable if and only if $\boldsymbol{\mathcal{W}}_o$ is of full rank. Using the same analogy, the observability Gramian for the nonlinear system (11) with respect to $\boldsymbol{\gamma}$ around $\boldsymbol{x}_0$ can be constructed as

$$\boldsymbol{W}_o(\boldsymbol{\gamma},\boldsymbol{x}_0) := \boldsymbol{J}_w^\top(\boldsymbol{\gamma},\boldsymbol{x}_0)\boldsymbol{J}_w(\boldsymbol{\gamma},\boldsymbol{x}_0), \tag{16}$$

where $\boldsymbol{W}_o(\cdot) \in \mathbb{R}^{n\times n}$ and $\boldsymbol{J}_w(\cdot) \in \mathbb{R}^{Np\times n}$ is given as [42]

$$\boldsymbol{J}_w(\boldsymbol{\gamma},\boldsymbol{x}_0) := \left[\boldsymbol{I} \otimes \boldsymbol{\Gamma}\boldsymbol{C}\right] \times \begin{bmatrix} \boldsymbol{I} \\ \dfrac{\partial\boldsymbol{x}[1]}{\partial\boldsymbol{x}[0]} \\ \vdots \\ \dfrac{\partial\boldsymbol{x}[N-1]}{\partial\boldsymbol{x}[0]} \end{bmatrix}, \tag{17}$$

where for system (11), by applying chain rule, for $0 < k \le$

$N-1$ the k-th partial derivative can be obtained from

$$\frac{\partial \boldsymbol{x}[k]}{\partial \boldsymbol{x}[0]} = \prod_{j=0}^{k-1} \boldsymbol{A} + \boldsymbol{G}\frac{\partial \boldsymbol{f}}{\partial \boldsymbol{x}}(\boldsymbol{x}, \boldsymbol{u}, j),$$

where the term $\frac{\partial \boldsymbol{f}}{\partial \boldsymbol{x}}(\cdot) \in \mathbb{R}^{g\times n}$ denotes the Jacobian of $\boldsymbol{f}(\cdot)$ with respect to $\boldsymbol{x}$ evaluated at discrete time j. The explicit form of this Jacobian matrix is detailed in Appendix B in the supplemental document attached with this manuscript submission. This requires $\boldsymbol{x}[j]$ and $\boldsymbol{u}[j]$ to be known, which can be obtained from simulating (11) for up to j-th discrete time. In the context of sensor placement problem, the objective is to find the best $\boldsymbol{\gamma}$ satisfying $\boldsymbol{\gamma} \in \mathcal{G}_\gamma$ which maximizes the observability of system (11) based on the observability Gramian (16).

There exist some measures that quantify observability based on Gramian matrix, some of which include the rank, smallest eigenvalue, condition number, trace, and determinant—see reference [40]. Each measures has a unique characterization vis-à-vis observability. The rank of observability Gramian quantifies the dimension of observable subspace. The smallest eigenvalue of the Gramian is the worst case metric of observability and reflects the difficulty of estimating the actual initial state $\boldsymbol{x}_0$ from the given set of measurements $\tilde{\boldsymbol{y}}$. That is, the estimation error can be very much sensitive to observation noise when the smallest eigenvalue of the Gramian matrix is relatively close to zero [49]. The condition number—which measures the ratio between the largest and smallest eigenvalues of the Gramian matrix—on the other hand, accentuates the sensitivity of observability.

We realize that the Gramian (16) quantifies the observability on a particular initial state $\boldsymbol{x}_0$. Large condition number (which implies *ill-conditioned* Gramian matrix) indicates that small change in initial state may produce a significant impact on observability [49]. The trace of observability Gramian measures the average observability of the system in all directions in state space. Therefore, a larger value of the trace indicates an increase in overall observability [41]. Similar to the trace, determinant of observability Gramian matrix also measures the system's observability in every directions in state space. Nonetheless, as pointed out in [39], determinant has the ability to capture information from all elements of the Gramian matrix as well as taking into account information redundancy for the case when multiple sensors are considered. It is also suggested in [40] that the determinant is a better measure for quantifying observability than the trace, since trace tends to overlook (near) zero eigenvalues.

Based on the above considerations, we settle upon maximizing the determinant and trace of the observability Gramian since, despite their differences discussed previously, both metrics are popular in sensor placement literature through empirical observability Gramian; see [39]–[41], as well as their easy implementation on standard optimization interfaces such as YALMIP [50] and CVX [51]. The resulting sensor placement problem is given as follows

$$(\mathbf{P1}) \quad \kappa = \underset{\boldsymbol{\gamma}}{\text{minimize}} \quad \begin{cases} -\det\left(\boldsymbol{W}_o(\boldsymbol{\gamma}, \hat{\boldsymbol{x}}_0)\right), \\ -\text{trace}\left(\boldsymbol{W}_o(\boldsymbol{\gamma}, \hat{\boldsymbol{x}}_0)\right), \end{cases} \tag{18a}$$

$$\text{subject to} \quad \boldsymbol{\gamma} \in \mathcal{G}_\gamma, \ \boldsymbol{\gamma} \in \{0,1\}^p. \tag{18b}$$

Notice that **P1** only takes the integer variables $\boldsymbol{\gamma}$ as the optimization variable while $\hat{\boldsymbol{x}}_0$ is fixed. Since the constraint and objective function are convex, then **P1** is categorized as a convex IP, which can be solved optimally via a BnB algorithm. After the observability of the system has been determined from the solutions of **P1**, then the measurement equation (11b) can be reformulate into $\boldsymbol{y}[k] = \tilde{\boldsymbol{C}}\boldsymbol{x}[k]$ where $\tilde{\boldsymbol{C}} \in \mathbb{R}^{n\times\tilde{p}}$ is the reduced state-to-output matrix that corresponds to the nonzero rows of $\boldsymbol{\Gamma}^*\boldsymbol{C}$, where $\boldsymbol{\gamma}^*$ is the optimal solution of **P1**. The traffic density estimation is then performed based on $\boldsymbol{y}[k] \in \mathbb{R}^{\tilde{p}}$.

IV. A Robust Observer for State Estimation

The previous section focuses on determining the traffic sensor placement based on the observability of the traffic dynamics through a finite observation window. As such, the obvious approach for performing traffic density estimation is via MHE. However, compared to other approaches such as observers or Kalman filter (KF)-based estimators, MHE for nonlinear systems is generally computationally more expensive since a nonlinear least-square problem has to be solved in every discrete time instance, which make it only suitable for slow dynamical systems or situations where high-powered computational resources are available. To that end, herein we employ a robust observer for estimating traffic density, since observer uses a single gain matrix for all discrete-time instances and only need to be computed once for a given traffic sensors configuration.

To develop the robust state observer, the nonlinearities in (10) have to be identified. In fact, we prove that these nonlinearities satisfy the Lipschitz continuity condition, which for a vector-valued function is defined as follows.

Definition 2 (Lipschitz Continuity)**.** *Let $\boldsymbol{f} : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$. Then, $\boldsymbol{f}$ is said to be Lipschitz continuous in $\boldsymbol{\Omega} \subseteq \mathbb{R}^n \times \mathbb{R}^m$ if there exists a constant $\gamma \in \mathbb{R}_+$ such that*

$$\|\boldsymbol{f}(\boldsymbol{x}, \boldsymbol{u}) - \boldsymbol{f}(\hat{\boldsymbol{x}}, \boldsymbol{u})\|_2 \le \gamma\|\boldsymbol{x} - \hat{\boldsymbol{x}}\|_2, \tag{19}$$

for all $\boldsymbol{x}, \hat{\boldsymbol{x}} \in \boldsymbol{\mathcal{X}}$ and $\boldsymbol{u} \in \boldsymbol{\mathcal{U}}$.

The next proposition shows that the nonlinear mapping $\boldsymbol{f}(\cdot)$ in (10) is Lipschitz continuous with Lipschitz constant γ_l.

Proposition 2. *The non-linear function $\boldsymbol{f}(\cdot)$ governing the traffic dynamics (10) and specified in Appendix A is Lipschitz continuous in $\boldsymbol{x}$ and $\boldsymbol{u}$ with the Lipschitz constant γ_l given as*

$$\gamma_l = h\left(v_f, \omega_c, l, N, N_I, N_O, \xi_1, \xi_2, \ldots, \xi_{N_I}, \beta_1, \beta_2, \ldots, \beta_{N_O}\right) \tag{20}$$

where the function $h(\cdot)$ is defined as in (30).

The proof of Proposition 2 is given in Appendix C of the supplemental document attached with this manuscript submission. In what follows, we consider the traffic dynamics (11) with the new measurement matrix $\tilde{\boldsymbol{C}}$ and the addition of *unknown inputs*—which may include unmodeled dynamics, disturbance, process noise, and measurement noise—described as

$$\boldsymbol{x}[k+1] = \boldsymbol{A}\boldsymbol{x}[k] + \boldsymbol{G}\boldsymbol{f}(\boldsymbol{x}, \boldsymbol{u}) + \boldsymbol{B}_{\mathbf{u}}\boldsymbol{u}[k] + \boldsymbol{B}_{\mathbf{w}}\boldsymbol{w}[k] \tag{21a}$$

$$\boldsymbol{y}[k] = \tilde{\boldsymbol{C}}\boldsymbol{x}[k] + \boldsymbol{D}_{\mathbf{w}}\boldsymbol{w}[k]. \tag{21b}$$

In the above model, the disturbance vector $\boldsymbol{w} \in \mathbb{R}^q$ lumps all unknown inputs into a single vector with the corresponding matrices $\boldsymbol{B}_{\mathbf{w}}$ and $\boldsymbol{D}_{\mathbf{w}}$ are of appropriate dimensions. These

matrices describe how the external disturbances are distributed in the system, where $\boldsymbol{B}_{\mathbf{w}}$ takes into account any disturbance affecting the system's dynamics while $\boldsymbol{D}_{\mathbf{w}}$ considers disturbance affecting the measurements data. Both can be constructed from observing empirical traffic data. The observer dynamics for (21) are derived from the classical Luenberger observer given as

$$\begin{aligned}\hat{\boldsymbol{x}}[k+1] &= \boldsymbol{A}\hat{\boldsymbol{x}}[k+1] + \boldsymbol{G}\boldsymbol{f}(\hat{\boldsymbol{x}},\boldsymbol{u}) \\ &\quad + \boldsymbol{B}_{\mathbf{u}}\boldsymbol{u}[k] + \boldsymbol{L}(\boldsymbol{y}[k] - \hat{\boldsymbol{y}}[k])\end{aligned} \tag{22a}$$

$$\hat{\boldsymbol{y}}[k] = \tilde{\boldsymbol{C}}\hat{\boldsymbol{x}}[k], \tag{22b}$$

where $\hat{\boldsymbol{x}}[k]$ is the state estimation vector, $\hat{\boldsymbol{y}}[k]$ is the measurement estimation vector, and $\boldsymbol{L}(\boldsymbol{y}-\hat{\boldsymbol{y}})$ is the Luenberger-type correction term with $\boldsymbol{L} \in \mathbb{R}^{n\times\tilde{p}}$. By defining the estimation error as $\boldsymbol{e}[k] := \boldsymbol{x}[k] - \hat{\boldsymbol{x}}[k]$, then from (21) and (22), the error dynamics can be shown to be in the following form

$$\begin{aligned}\boldsymbol{e}[k+1] &= \left(\boldsymbol{A} - \boldsymbol{L}\tilde{\boldsymbol{C}}\right)\boldsymbol{e}[k] + \boldsymbol{G}\Delta\boldsymbol{f}[k] \\ &\quad + (\boldsymbol{B}_{\mathbf{w}} - \boldsymbol{L}\boldsymbol{D}_{\mathbf{w}})\boldsymbol{w}[k]\end{aligned} \tag{23a}$$

$$\boldsymbol{z}[k] = \boldsymbol{Z}\boldsymbol{e}[k], \tag{23b}$$

where $\Delta\boldsymbol{f}[k] := \boldsymbol{f}(\boldsymbol{x},\boldsymbol{u}) - \boldsymbol{f}(\hat{\boldsymbol{x}},\boldsymbol{u})$ and in particular, $\boldsymbol{z} \in \mathbb{R}^z$ is the performance output of the error dynamics with respect to the user-defined performance matrix $\boldsymbol{Z} \in \mathbb{R}^{z\times n}$.

The goal of the observer is to provide asymptotic estimation error for estimation error dynamics given above. Since the presence of unknown inputs will not make the estimation error to be exactly zero, a robustness metric, referred to as and $\mathcal{L}_\infty$ stability, is employed. The purpose of this metric is to provide numerical assurance on the behavior of performance output $\boldsymbol{z}$ against nonzero, time-varying unknown inputs $\boldsymbol{w}$. Our prior work [52] deals with a robust observer design using the concept of $\mathcal{L}_\infty$ stability for traffic density estimation purpose assuming nonlinear continuous-time traffic dynamics model corresponding with the Greenshield's model. Furthermore, our recent work [53] develops a robust $\mathcal{L}_\infty$ observer for the discrete-time model corresponding with the ACTM. In the sequel, we reproduce a simple numerical procedure from [53] to find an observer gain $\boldsymbol{L}$ that, if successfully solved, renders the estimation error dynamics (23) to be $\mathcal{L}_\infty$ stable with performance level μ. Compared with [53], this paper includes the mathematical proof for the next result.

Proposition 3. *Consider the nonlinear dynamics (21) and observer (22) in which $\boldsymbol{w} \in \mathcal{L}_\infty$ and $\boldsymbol{f} : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$ is locally Lipschitz in $\boldsymbol{\Omega}$ with Lipschitz constant γ_l. If there exist $\boldsymbol{P} \in \mathbb{S}^n_{++}$, $\boldsymbol{Y} \in \mathbb{R}^{n\times p}$, $\epsilon, \mu_0, \mu_1, \mu_2 \in \mathbb{R}_+$, and $\alpha \in \mathbb{R}_{++}$ such that the following optimization problem is solved*

$$(\mathbf{P2}) \quad \underset{\boldsymbol{P},\boldsymbol{Y},\epsilon,\alpha,\mu_{0,1,2}}{\text{minimize}} \quad \mu_0\mu_1 + \mu_2 \tag{24a}$$

subject to

$$\begin{bmatrix} (\alpha-1)\boldsymbol{P} + \epsilon\gamma_l^2\boldsymbol{I} & * & * & * \\ \boldsymbol{O} & -\epsilon\boldsymbol{I} & * & * \\ \boldsymbol{O} & \boldsymbol{O} & -\alpha\mu_0\boldsymbol{I} & * \\ \boldsymbol{P}\boldsymbol{A} - \boldsymbol{Y}\boldsymbol{C} & \boldsymbol{P}\boldsymbol{G} & \boldsymbol{P}\boldsymbol{B}_{\mathbf{w}} - \boldsymbol{Y}\boldsymbol{D}_{\mathbf{w}} & -\boldsymbol{P} \end{bmatrix} \preceq 0 \tag{24b}$$

$$\begin{bmatrix} -\boldsymbol{P} & * & * \\ \boldsymbol{O} & -\mu_2\boldsymbol{I} & * \\ \boldsymbol{Z} & \boldsymbol{O} & -\mu_1\boldsymbol{I} \end{bmatrix} \preceq 0, \tag{24c}$$

then the error dynamics given in (23) is $\mathcal{L}_\infty$ stable with performance level $\mu = \sqrt{\mu_0\mu_1 + \mu_2}$ for performance output given as $\boldsymbol{z} = \boldsymbol{Z}\boldsymbol{e}$ and observer gain $\boldsymbol{L}$ computed as $\boldsymbol{L} = \boldsymbol{P}^{-1}\boldsymbol{Y}$.

The proof of Proposition 3 is provided in Appendix D of the supplemental document attached with this manuscript submission. Note that optimization problem described in (24) is nonconvex due to bilinearity appearing in the form of $(\alpha-1)\boldsymbol{P}$, $\alpha\mu_0$, and $\mu_0\mu_1$. To get a convex one, one can simply fix the values of α and μ_0 or μ_1, and solve for the other variables using any SDP solvers. Having described our strategies for placing traffic sensors and estimating traffic density via a robust observer framework, we then perform numerical tests of these strategies through various case studies—presented in the ensuing sections.

V. Case Study: Results and Analysis

This section demonstrates the proposed approach for determining traffic sensors location and estimating traffic density through $\mathcal{L}_\infty$ observer. Specifically, in this numerical study we attempt to answer the following questions:

- *Q1*: How does the observation window length relate to traffic observability and initial state estimation?
- *Q2*: How computationally efficient are the determinant and trace observability metrics in terms of solving the sensor placement problem?
- *Q3*: How do actual and presumed initial states affect the resulting sensors' location? Are sensor placements robust to changes in initial states?
- *Q4*: How reliable is the state estimation when using sensor locations obtained from utilizing determinant and trace observability metrics?
- *Q5*: How does the theory-driven placement of sensors compare to randomized and uniform sensor placement strategies?

All simulations are performed using MATLAB R2019a running on 64-bit Windows 10 with 3.4GHz Intel^{R} Core^{TM} i7-6700 CPU and 16 GB of RAM with YALMIP [50] as the interface to solve all IP and convex SDP. Throughout the section, all highways are configured with $v_f = 28.8889$ m/s (65 mph), $w_c = 6.6667$ m/s (15 mph), $\rho_c = 0.0249$ vehicles/m (40 vehicles/mile), $\rho_m = 0.1333$ vehicles/m, and $l = 400$ m. The discrete-time step is chosen to be $T = 1$ sec.

A. Observability Analysis for Traffic Sensor Placement

Herein, we perform a numerical analysis on sensor placement approach through traffic network's observability discussed in Section III. The highway, referred to the rest of the section as *Highway A*, stretches for approximately 3.2 miles with 13 segments on the mainline, 4 on-ramps, and 4 off-ramps such that there are 21 segments in total. The objective of traffic sensor placement problem translates to finding the set of r highway segments that must be equipped with traffic sensors such that the entire highway traffic is observable, which is carried out by solving **P1**. Realize that **P1** is classified as a convex *integer programming* (IP) since $\hat{\boldsymbol{x}}_0$, the presumed initial state, is fixed. **P1** is solved using YALMIP's branch-and-bound (BnB) algorithm [50] along with MOSEK [54] solver.

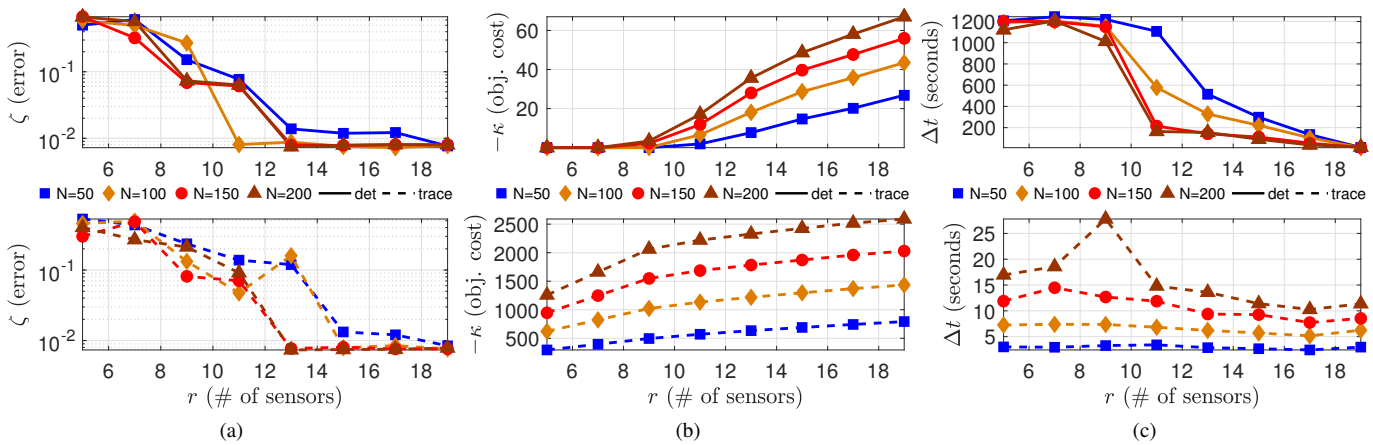


Figure 4. Results on observability analysis for different number of sensor allocations: (a) the relative error ζ, (b) inverse optimal value of **P1**, denoted by $-\kappa$, and (c) total computational time, which includes the overall time spent for solving **P1** and **P3**.

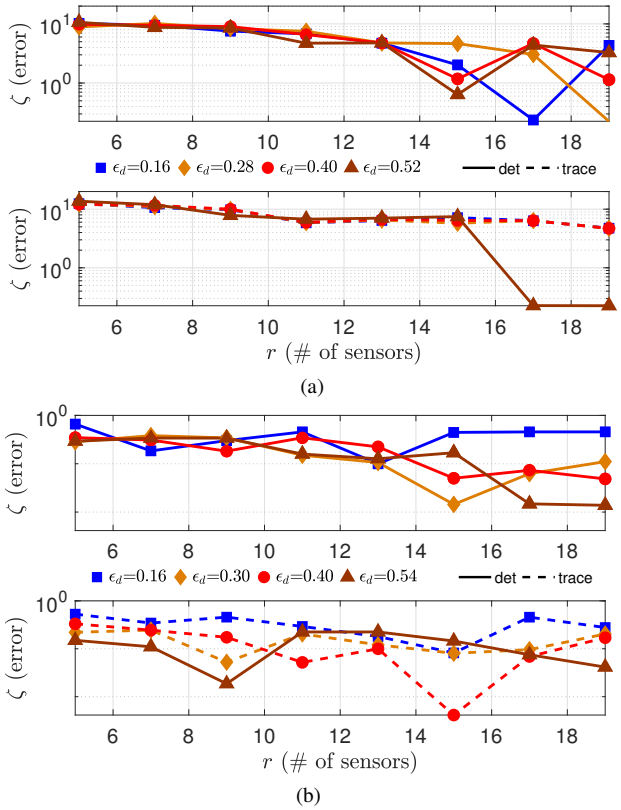


Figure 5. Relative error ζ for (a) various presumed initial state $\hat{\boldsymbol{x}}_0$ and (b) actual initial state $\boldsymbol{x}_0$.

In the first instance of observability analysis, problem **P1** is solved with fixed actual initial state $\boldsymbol{x}_0$ and presumed initial state $\hat{\boldsymbol{x}}_0$, which are generated randomly in $(0, \rho_m]^n$. **P1** is then solved with different observation windows N for varying number of allocated sensors r. In this case, $r = \lceil p \times r_p\% \rceil$ where $\lceil \cdot \rceil$ is the ceiling function, $p = n = 21$, and $r_p\%$ represents the percentage of sensor's allocation (from 20% to 90%). The number of allocated sensors is a convex constraint such that, with respect to (18b), $\boldsymbol{\gamma} \in \mathcal{G}_\gamma$ is equivalent to $\sum_i \gamma_i = r$.

We solve **P1** for both observability metrics—the trace and determined metrics discussed in Section III. It is suggested in [40] to use logarithmic determinant as opposed to pure determinant to avoid numerical problems. However, since YALMIP does not have option for realizing logarithmic determinant objective function particularly for IP, then **P1** is solved using `geomean` function, which is equivalent to $\sqrt[n]{\det(\cdot)}$. The maximum iteration of YALMIP's BnB algorithm is chosen to be 10^4. If the optimality gap is not satisfied with the default tolerance after the maximum iteration is achieved, the resulting solution is taken as the final outcome.

In this numerical study, we put our interest on comparing the relative error ζ, the inverse optimal value of **P1**, denoted by $-\kappa$, and total computational time Δt. The relative error ζ represents the difference between actual initial state $\boldsymbol{x}_0$ and the *computed* initial state $\tilde{\boldsymbol{x}}_0$, which is obtained from solving the following nonlinear least-square problem [42]

$$(\mathbf{P3}) \quad \underset{\tilde{\boldsymbol{x}}_0}{\text{minimize}} \quad \|\tilde{\boldsymbol{y}}_\gamma - \boldsymbol{g}_\gamma(\tilde{\boldsymbol{x}}_0)\|_2^2 \tag{25a}$$

$$\text{subject to} \quad \mathbf{0} \leq \tilde{\boldsymbol{x}}_0 \leq \mathbf{1} \times \rho_m. \tag{25b}$$

In (25a), $\tilde{\boldsymbol{y}}_\gamma$ is obtained from simulating the undisturbed traffic dynamics (11) with a prescribed set of sensors $\boldsymbol{\gamma}$ from $k = 0$ to $k = N$ and initial state $\boldsymbol{x}_0$. Likewise, the nonlinear mapping $\boldsymbol{g}_\gamma(\cdot)$ is constructed as in (13) with the same sensor combination $\boldsymbol{\gamma}$. Problem **P3** is solved by using the MATLAB data-fitting function `lsqnonlin`, which implements a *trust region reflective* algorithm [55]. The optimality tolerance is set to be 10^{-6} and an initial guess equals to $\hat{\boldsymbol{x}}_0$. After $\tilde{\boldsymbol{x}}_0$ is obtained as the solution of **P3**, ζ is computed as

$$\zeta := \frac{\|\tilde{\boldsymbol{x}}_0 - \boldsymbol{x}_0\|_2}{\|\boldsymbol{x}_0\|_2}.$$

Note that ζ appraises the quality of sensor placement, as smaller ζ suggests that the given sensor configuration is capable to provide a more accurate estimate of the initial state. Of course, this is reasonable if **P3** returns an *optimal* solution. Nonetheless, as **P3** is a nonconvex optimization problem, then it is highly unlikely that optimal solutions can be obtained. To reduce the chance of variability, the actual $\boldsymbol{x}_0$ and presumed $\hat{\boldsymbol{x}}_0$ initial states are fixed for this particular test.

The results of this numerical study are depicted in Fig. 4. In particular, it can be seen from Fig. 4a that larger observation window yields smaller relative error. Notice that this behavior is reflected better from the trace but less evident from the

Table II
SENSOR LOCATIONS FOR TWO DIFFERENT OBSERVATION LENGTHS AND OBSERVABILITY METRICS. THE NOTATION $(x) + \{y\}$ MEANS THAT THE SET OF SENSORS $\{y\}$ IS APPENDED TO THAT OF ROW (x) FROM THE SAME COLUMN. THE ITALICIZED NUMBERS INDICATE SUBOPTIMAL SENSOR LOCATIONS.

r	det		trace	
	$N = 100$	$N = 200$	$N = 100$	$N = 200$
5	$\{1, 6, 10, 16, 20\}$	$\{1, 6, 10, 14, 15\}$	$\{6, 14, 17, 19, 20\}$	$\{6, 14, 17, 19, 20\}$
7	$\{\mathit{1}, \mathit{6}, \mathit{15}, \mathit{16}, \mathit{17}, \mathit{20}, \mathit{21}\}$	$\{\mathit{1}, \mathit{6}, \mathit{14}, \mathit{16}, \mathit{17}, \mathit{20}, \mathit{21}\}$	$(5) + \{16, 21\}$	$(5) + \{15, 16\}$
9	$(7) + \{\mathit{14}, \mathit{19}\}$	$\{1, 14, 15, 16, 17, 18, 19, 20, 21\}$	$(7) + \{15, 18\}$	$(7) + \{18, 21\}$
11	$(9) + \{10, 18\}$	$(9) + \{6, 10\}$	$(9) + \{8, 10\}$	$(9) + \{8, 10\}$
13	$(11) + \{8, 12\}$	$(11) + \{8, 12\}$	$(11) + \{11, 13\}$	$(11) + \{1, 11\}$
15	$(13) + \{2, 5\}$	$(13) + \{2, 5\}$	$(13) + \{1, 2\}$	$(13) + \{7, 9\}$
17	$(15) + \{7, 11\}$	$(15) + \{9, 11\}$	$(15) + \{7, 9\}$	$(15) + \{2, 13\}$
19	$(17) + \{3, 13\}$	$(17) + \{3, 13\}$	$(17) + \{3, 12\}$	$(17) + \{3, 12\}$

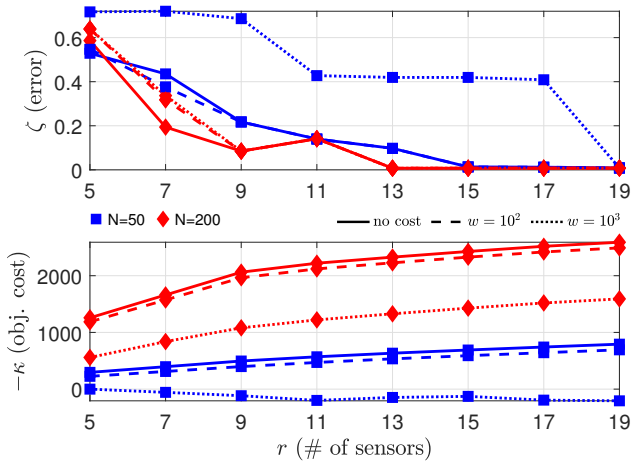


Figure 6. Comparison of relative error (a) and the trace of the Gramian (b) between uniform and different (randomized) sensor costs.

determinant—the variations are presumed to be caused by suboptimal solutions obtained from solving **P3**. Fig. 4b also indicates this behavior: larger observation window yields better optimal value **P1**, which in turn implies that the observability measure is directly proportional with observation window. This corroborates first principles in control theory; more available data (i.e., a larger observation window) results in a better system observability. The corresponding total computational time for both **P1** and **P3** are reported in Fig. 4c. It can be observed that, in this test where YALMIP's BnB algorithm is utilized, employing the trace in **P1** is more efficient than doing so with the determinant.

The resulting sensor's location for observation window $N = 100$ and $N = 200$ are shown in Tab. II. For the same number of sensors and observation windows, the sensor locations for determinant and trace objective functions are quite distinct. This observation is expected as both metrics measure different aspect on observability Gramian. Interestingly, as it is seen from Tab. II, the results from using trace as observability metric suggest a *modular-like behavior*[1]: the sensor combination from a particular row covers the ones from its previous rows. This modularity property is also observed from the determinant objective function, despite some irregularities for the first three rows, which is caused by suboptimal solutions. This phenomena is in tandem with the conjecture stated in [42], claiming that the logarithmic determinant objective function in **P1** is submodular.

Next, we asses the variability of sensor placement due to the differences in actual $\boldsymbol{x}_0$ and presumed $\hat{\boldsymbol{x}}_0$ initial states. The numerical experiment is carried out as follows. First, we set $\boldsymbol{x}_0$ to be fixed and use different values for $\hat{\boldsymbol{x}}_0$ which are randomly generated inside $(0, \rho_m]^n$ in such a way that their Euclidean distances, computed as $\epsilon_d := \|\hat{\boldsymbol{x}}_0 - \boldsymbol{x}_0\|_2$, are unique to each other. For this scenario, the results are given in Fig. 5a, which shows the resulting relative error. In another scenario, $\hat{\boldsymbol{x}}_0$ is fixed while $\boldsymbol{x}_0$ varies—see Fig. 5b for the results. It can be seen from both figures that, in general, varying $\hat{\boldsymbol{x}}_0$ while fixing $\boldsymbol{x}_0$ yields higher relative error magnitudes than fixing $\hat{\boldsymbol{x}}_0$ and varying $\boldsymbol{x}_0$. However, when $\boldsymbol{x}_0$ is fixed, ζ experiences minimal variations at least for $r < 14$, whereas more variations occur on the other scenario. Again, these variations are most likely attributed to the suboptimal solutions obtained from solving **P3** via `lsqnonlin`. It is also observed that larger distance ϵ_d does not necessarily result in larger relative error ζ, suggesting that the proposed method is rather resilient towards the values of actual and presumed initial states.

The corresponding sensor combinations for scenario when $\boldsymbol{x}_0$ is fixed and $\hat{\boldsymbol{x}}_0$ varies with two different Euclidean distances are provided in Tab. III. It can be seen that, for the trace, the sensor combinations are the same, indicating that it is robust to different actual and presumed initial states. The same pattern is also observed for the determinant, despite being less apparent. The first three rows are expected since they are suboptimal, hence the different sensor combinations. However, when the solutions are optimal, the resulting sensor combinations are similar except for $r = 15$. Similar results are also obtained for scenario when $\boldsymbol{x}_0$ varies and $\hat{\boldsymbol{x}}_0$ is fixed.

Lastly, in this section we investigate the impact of different sensor costs towards the quality of sensor's configuration. To that end, the following problem is solved

$$(\mathbf{P4}) \quad \underset{\boldsymbol{\gamma},\, t \geq 0}{\text{minimize}} \quad -\operatorname{trace}\left(\boldsymbol{W}_o(\boldsymbol{\gamma}, \hat{\boldsymbol{x}}_0)\right) + w \times t \tag{26a}$$

$$\text{subject to} \quad \boldsymbol{\gamma} \in \mathcal{G}_\gamma,\ \boldsymbol{\gamma} \in \{0,1\}^p,\ \boldsymbol{c}^\top \boldsymbol{\gamma} \leq t. \tag{26b}$$

Realize that **P4** is **P1** with the addition of sensor costs and only considers the trace function since YALMIP faces numerical errors while `geomean` function is fused with sensor costs. In this problem, $w \in \mathbb{R}_{++}$ is used to determine the priority

[1]A set function $f : 2^{\mathcal{V}} \to \mathbb{R}$, in which $\mathcal{V} = \{1, 2, \ldots, p\}$ for $p \in \mathbb{N}$, is modular if and only if for any $\mathcal{S} \subseteq \mathcal{V}$ it can be expressed as $f(\mathcal{S}) = w(\emptyset) + \sum_{s \in \mathcal{S}} w(s)$ for some weight function $w : \mathcal{V} \to \mathbb{R}$ [56].

Table III

SENSOR'S LOCATIONS FOR TWO DIFFERENT EUCLIDEAN DISTANCES BETWEEN ACTUAL $\boldsymbol{x}_0$ AND PRESUMED $\hat{\boldsymbol{x}}_0$ INITIAL STATES WHERE $\boldsymbol{x}_0$ IS FIXED AND $\hat{\boldsymbol{x}}_0$ VARIES. THE NOTATION $(x) + \{y\}$ MEANS THAT THE SET OF SENSORS $\{y\}$ IS APPENDED TO THAT OF ROW (x) FROM THE SAME COLUMN. THE ITALICIZED NUMBERS INDICATE SUBOPTIMAL SENSOR LOCATIONS.

r	det		trace	
	$\epsilon_d = 0.16$	$\epsilon_d = 0.40$	$\epsilon_d = 0.16$	$\epsilon_d = 0.40$
5	$\{\mathit{1}, \mathit{6}, \mathit{15}, \mathit{17}, \mathit{20}\}$	$\{\mathit{1}, \mathit{9}, \mathit{15}, \mathit{16}, \mathit{19}\}$	$\{14, 15, 18, 19, 20\}$	$\{14, 15, 18, 19, 20\}$
7	$\{\mathit{1}, \mathit{5}, \mathit{16}, \mathit{17}, \mathit{18}, \mathit{20}, \mathit{21}\}$	$\{\mathit{1}, \mathit{8}, \mathit{15}, \mathit{17}, \mathit{18}, \mathit{19}, \mathit{21}\}$	$(5) + \{16, 17\}$	$(5) + \{16, 17\}$
9	$\{\mathit{1}, \mathit{6}, \mathit{11}, \mathit{14}, \mathit{15}, \mathit{16}, \mathit{18}, \mathit{19}, \mathit{20}\}$	$\{\mathit{1}, \mathit{5}, \mathit{10}, \mathit{14}, \mathit{16}, \mathit{17}, \mathit{18}, \mathit{19}, \mathit{21}\}$	$(7) + \{3, 21\}$	$(7) + \{3, 21\}$
11	$(7) + \{9, 14, 15, 19\}$	$(5) + \{5, 14, 17, 18, 20, 21\}$	$(9) + \{6, 9\}$	$(9) + \{6, 9\}$
13	$(9) + \{3, 9, 17, 21\}$	$(5) + \{3, 6, 11, 14, 17, 18, 20, 21\}$	$(11) + \{2, 5\}$	$(11) + \{2, 5\}$
15	$(11) + \{3, 7, 11, 12\}$	$(9) + \{3, 6, 8, 12, 15, 20\}$	$(13) + \{1, 8\}$	$(13) + \{1, 8\}$
17	$(13) + \{2, 5, 8, 12\}$	$(13) + \{2, 5, 8, 12\}$	$(15) + \{4, 7\}$	$(15) + \{4, 7\}$
19	$(17) + \{4, 10\}$	$(17) + \{4, 10\}$	$(17) + \{10, 11\}$	$(17) + \{10, 11\}$

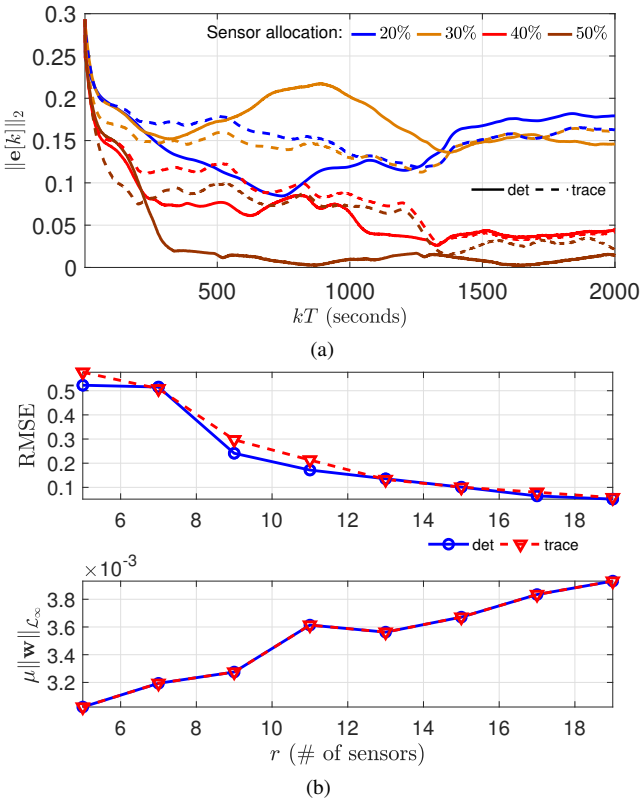


Figure 7. State estimation results via $\mathcal{L}_\infty$ observer for different allocations of sensor: (a) the estimation error and (b) the estimation error performance bounds.

between the two objectives and $\boldsymbol{c} \in \mathbb{R}_+^p$ represents the costs for all sensors. Two different values for w are used, where $w \geq 10^2$ to signify the sensor costs. The sensor costs are generated randomly such that $c_i \in [0, 1]$. The results are shown in Fig. 6, from which it can be seen that the relative errors and the degrees of observability for sensor locations determined by considering sensor costs are considerably worse than not using sensor costs at all. This happens because **P4** has to simultaneously maximizing system's observability and minimizing the sensor costs, thereby giving sensor locations that have lower degree of observability.

B. Traffic Density Estimation with Various Sensor Allocations

This section investigates different traffic sensor placements—obtained from the previous section—for robust traffic density

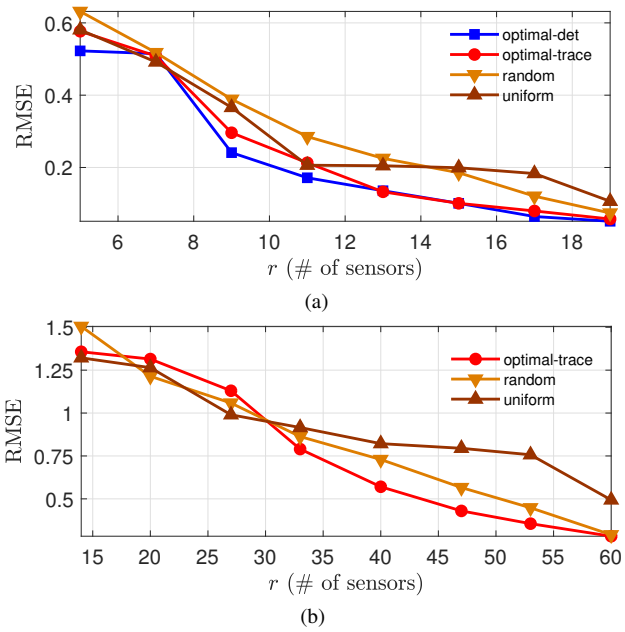
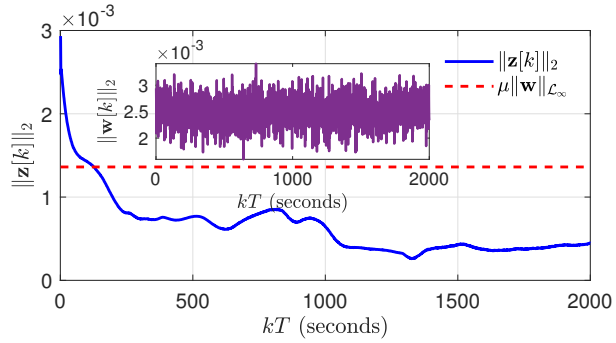


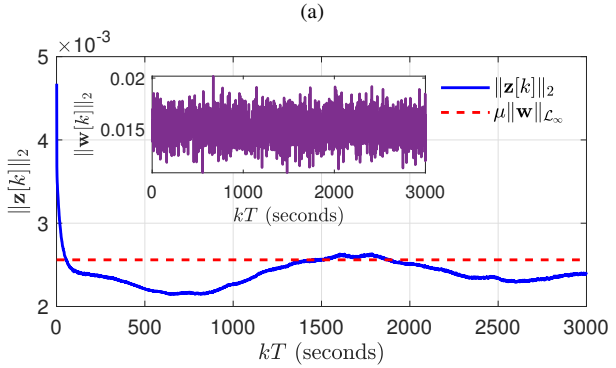
Figure 8. RMSE for optimal and randomized sensor placement for (a) Highway A and (b) Highway B. The simulations with randomized sensor placement are performed 10 times and the presented results correspond to the average values.

estimation purpose. To that end, we implement the $\mathcal{L}_\infty$ observer proposed in Section IV and set $\alpha = 0.1$ and $\mu_1 = 10^4$ to get a convex optimization problem out of **P2**, in which SDPNAL+ [57] is used to solve the problem since MOSEK returns numerical problems. The performance matrix is chosen to be $\boldsymbol{Z} = 0.01\boldsymbol{I}$. Herein we simulate Highway A with final time $k_f = 2000$ and invoke Gaussian noise with covariance matrices $Q = \nu\boldsymbol{I}$ and $R = \nu\boldsymbol{I}$ where $\nu = 10^{-3}$ with corresponding unknown input matrices $\boldsymbol{B}_\mathbf{w} = \begin{bmatrix}\boldsymbol{B}_\mathbf{u} & \boldsymbol{O}\end{bmatrix}$ and $\boldsymbol{D}_\mathbf{w} = \begin{bmatrix}\boldsymbol{O} & \boldsymbol{I}\end{bmatrix}$, which simulate process and measurement noise.

In this part of numerical study, we are focused on finding out whether different observability metrics used to solve **P1** have explicit impact on the quality of traffic density estimation, since according to Tab. II, different observability metrics return distinctive sensor locations. With that in mind, we compare the resulting estimation errors $\boldsymbol{e}[k]$, root-mean-square error



(a)



(b)

Figure 9. Performance of estimation error with 40% measurements for (a) Highway A (with determinant observability metric) and (b) Highway B (with trace observability metric).

(RMSE)—which is computed from

$$\text{RMSE} = \sum_{i=1}^{n} \sqrt{\frac{1}{k_f} \sum_{k=0}^{k_f} (e_i[k])^2},$$

and estimation error performance bounds $\mu \|\boldsymbol{w}\|_{\mathcal{L}_\infty}$, which is obtained from multiplying the worst case disturbance $\|\boldsymbol{w}\|_{\mathcal{L}_\infty}$ with performance index μ retrieved from solving **P2**. The sensor locations used in this test, for both observability metrics, are collected from Tab. II that correspond to observation window $N = 200$.

The results of this numerical test are illustrated in Fig. 7. It is indicated from Fig. 7b that, as more sensors are utilized, the estimation error is decreasing. However, it is of importance to note that using determinant as the observability metric yields better state estimation results than doing so with trace, since the RMSE for determinant is slightly smaller. This showcases the prominence of determinant as opposed to trace to quantify observability. Interestingly, the estimation error performance bounds are increasing as more sensors are utilized with no noticeable difference for determinant and trace. This suggests that, regardless of sensor locations, adding sensors increases the susceptibility of the entire network towards measurement noise. The trajectories of estimation error for different number of sensors are given in Fig. 7a, from which it can be observed that utilizing more sensors yields smaller and faster convergence on estimation error—notice that this observation is in accordance with RMSE shown in Fig. 7b. It is also observed from Fig. 7a that, in general, the determinant returns smaller estimation error compared to the trace. A slight exception occurs when the sensor allocation is at 30% (also see Fig. 7b). This is suspected to be caused by suboptimal sensor placement—see Tab. II.

C. Optimal, Randomized, and Uniform Sensor Placement

In the last part of numerical test, we analyze the results of traffic density estimation with optimal, randomized, and uniform sensor placement. We consider two highway networks: the first one is Highway A (detailed in Section V-A) while the other, referred to as *Highway B*, consists a larger network with 40 segments on the mainline, 13 on-ramps, and 13 off-ramps, giving 66 highway segments in total with length of 9.9 miles. The sensor placement and observer design for Highway A follow from two previous sections. For Highway B, the sensor placement is acquired from solving **P1** with trace observability metric since it is much more efficient than the determinant, especially for larger network. The observation window is again set to be $N = 200$ with varying number of sensor allocations. We set the covariant matrix R for Highway B to be $5 \times 10^{-3}\boldsymbol{I}$ to simulate a higher measurement noise, while the final time is set to be $k_f = 3000$. The remaining parameters for observer design are chosen to remain the same as in Highway A.

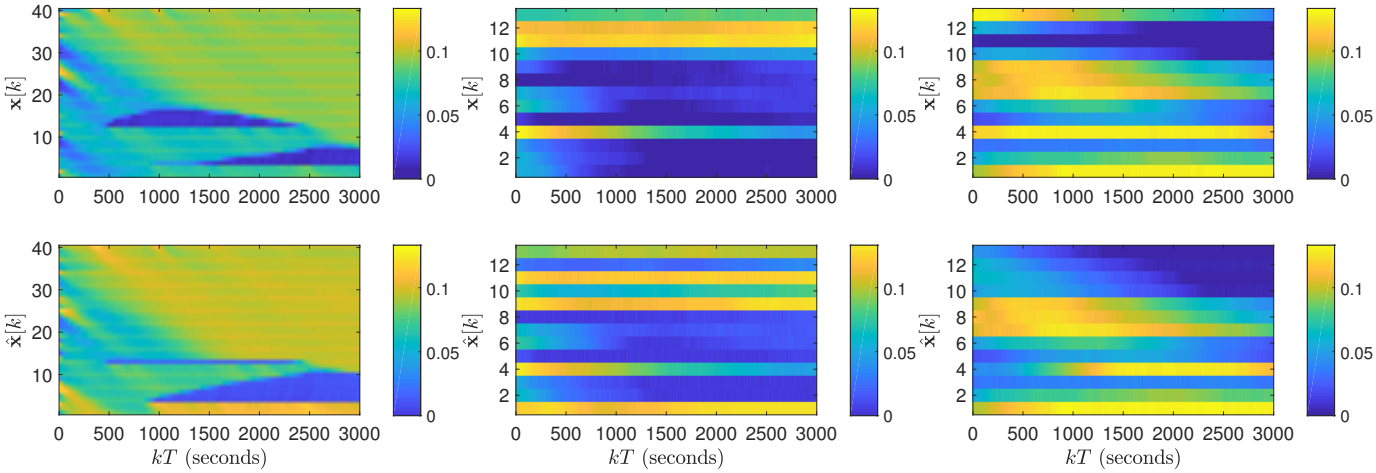
The first results of this numerical experiment are provided in Fig. 8. This figure showcases the RMSE of traffic density estimation with optimal, randomized, and uniform sensor placement for both highways. For the uniform case, the sensors are firstly placed at odd locations and will begin to be placed at even locations only when all odd locations are used up. Also note that, to compensate randomization on sensor placement, the simulations for this case (randomized sensor placement) are performed 10 times for every r and the results are averaged. For Highway A, the optimal sensor placements return significantly smaller RMSE than the randomized and uniform ones, regardless of observability metric being used. For Highway B, the RMSE for optimal sensor placement with trace is generally smaller than using randomized or uniform sensor placement. These results suggest that using optimal sensor placement, regardless of observability metric, gives better traffic density estimation than doing so with randomized sensor placement.

The second results of this numerical experiment are illustrated in Fig. 9, where the corresponding error performance norm $\|\boldsymbol{z}[k]\|_2$ are compared with the estimation error performance bound $\mu\|\boldsymbol{w}\|_{\mathcal{L}_\infty}$. Notice the larger value of error performance norm for Highway B due to higher measurement noise. Specifically for both highways, it can be seen that the norm of error performance is converging to a value below the estimation error performance bound within the given simulation window. These findings are in accordance with the definition of $\mathcal{L}_\infty$ stability. The resulting traffic density estimations for Highway B are depicted in Fig. 10.

VI. Conclusions, Paper Limitations, and Future Work

Given the thorough computational analysis in the previous section, the following observations are made, thereby answering the posed research questions in Section V:

- *A1:* Increasing observation window yields smaller initial state estimation errors while traffic network's quantifiable observability is directly proportional with observation window. This corroborates control-theoretic first principles.
- *A2:* Using trace as observability metric allows the sensor placement problem to be solved more efficiently than doing



(a) (b) (c)

Figure 10. Comparison between real $\boldsymbol{x}[k]$ and estimated $\hat{\boldsymbol{x}}[k]$ traffic densities on (a) mainline segments, (b) on-ramps, and (c) off-ramps for highway B with 40% measurements—sensor locations are obtained form solving **P1** with trace observability metric.

so with determinant metric for observability. This is critical if large-scale networks are considered.

– *A3:* The resulting sensor placement obtained from solving **P1** is robust towards actual and presumed initial states. This illustrates that the sensor placement to guarantee a wide range of initial operating conditions.
– *A4:* When used to determine optimal sensor locations, the determinant is better than the trace in term of state estimation error and hence offers better network-wide observability. This poses tradeoffs between state estimation quality and computational tractability for the two metrics; see A2 above.
– *A5:* The optimal sensor placement outperforms randomized and uniform sensor placement in estimating traffic density.

The approach in this paper has its own set of limitations. First, we consider a time-invariant traffic model where in reality, some parameters are in fact time-varying, which include critical density, split ratio, free-flow speed, and congestion wave speed. We also do not consider the capacity drop phenomenon or the stochasticity of traffic in the model presented in this paper. Second, the sensor placement strategy does not consider the effect of measurement noise. To that end, future work will include solving the traffic sensor placement problem while considering a time-varying nonlinear traffic model incorporating capacity drop and stochasticity and taking measurement noise into account, which is resulting in a robust sensor placement. Moreover, the submodularity properties of the determinant and trace—as well as other observability metrics—for sensor selection purpose will be investigated to assist in solving the placement problem for large networks. Finally, we point out that the presented placement approach in the paper can be extended or utilized to other models in transportation systems beyond stretched highways, assuming that a nonlinear state-space representation of the dynamics is possible.

Acknowledgments

We gratefully acknowledge the constructive comments from the reviewers—they have contributed positively to this work. We also acknowledge the financial support from the National Science Foundation through Grants 1636154, 1728629, 1917164, and 1917056.

References

[1] E. Lovisari, C. Canudas de Wit, and A. Kibangou, "Density/flow reconstruction via heterogeneous sources and optimal sensor placement in road networks," *Transportation Research Part C*, vol. 69, no. C, pp. 451–476, 2016.

[2] M. Gentili and P. Mirchandani, "Locating sensors on traffic networks: Models, challenges and research opportunities." *Transportation Research Part C*, vol. 24, 2012.

[3] M. Papageorgiou and A. Kotsialos, "Freeway ramp metering: An overview," *IEEE Transactions on Intelligent Transportation Systems*, vol. 3, no. 4, pp. 271–281, 2002.

[4] R.-C. Carlson, I. Papamichail, M. Papageorgiou, and A. Messmer, "Optimal motorway traffic flow control involving variable speed limits and ramp metering," *Transportation Science*, vol. 44, no. 2, pp. 238–253, 2010.

[5] A. Hegyi and S. Hoogendoorn, "Dynamic speed limit control to resolve shock waves on freeways - Field test results of the SPECIALIST algorithm," in *13th International IEEE Conference on Intelligent Transportation Systems*. IEEE, 2010, pp. 519–524.

[6] J. Kianfar and P. Edara, "Optimizing freeway traffic sensor locations by clustering global-positioning-system-derived speed patterns," *IEEE transactions on intelligent transportation systems*, vol. 11, no. 3, pp. 738–747, 2010.

[7] I. Jovanović, M. Šelmić, and M. Nikolić, "Metaheuristic approach to optimize placement of detectors in transport networks — case study of serbia," *Canadian journal of civil engineering*, vol. 46, no. 3, pp. 176–187, 2019.

[8] M. Gentili and P. B. Mirchandani, "Review of optimal sensor location models for travel time estimation," *Transportation research. Part C, Emerging technologies*, vol. 90, pp. 74–96, 2018.

[9] E. Mitsakis, E. Chrysohoou, J. M. Salanova Grau, P. Iordanopoulos, and G. Aifadopoulou, "The sensor location problem: methodological approach and application," *Transport (Vilnius, Lithuania)*, vol. 32, no. 2, pp. 113–119, 2017.

[10] M. Owais, G. S. Moussa, and K. F. Hussain, "Sensor location model for o/d estimation: Multi-criteria meta-heuristics approach," *Operations Research Perspectives*, vol. 6, 2019.

[11] N. Mehr and R. Horowitz, "A submodular approach for optimal sensor placement in traffic networks," in *2018 Annual American Control Conference (ACC)*, vol. 2018-. AACC, 2018, pp. 6353–6358.

[12] S.-R. Hu, S. Peeta, and C.-H. Chu, "Identification of vehicle sensor locations for link-based network traffic applications," *Transportation Research Part B*, vol. 43, no. 8, pp. 873–894, 2009.

[13] M. Ng, "Synergistic sensor location for link flow inference without path enumeration: A node-based approach," *Transportation Research, Part B*, vol. 46, no. 6, pp. 781–788, 2012.

[14] L. Bianco, G. Confessore, and P. Reverberi, "A network based model for traffic sensor location with implications on O/D matrix estimates," *Transportation Science*, vol. 35, no. 1, pp. 50–60, 2001.

[15] M. Shao, L. Sun, and X. Shao, "Sensor location problem for network traffic flow derivation based on turning ratios at intersection," *Mathematical Problems in Engineering*, vol. 2016, p. 10, 2016.

[16] M. Salari, L. Kattan, W. H. Lam, H. Lo, and M. A. Esfeh, "Optimization of traffic sensor location for complete link flow observability in traffic network considering sensor failure," *Transportation Research Part B*, vol. 121, pp. 216–251, 2019.

[17] E. Castillo, A. J. Conejo, J. M. Menéndez, and P. Jiménez, "The observability problem in traffic network models," *Computer-aided civil and infrastructure engineering*, vol. 23, no. 3, pp. 208–222, 2008.

[18] M. Rinaldi and F. Viti, "Exact and approximate route set generation for resilient partial observability in sensor location problems," *Transportation research. Part B: methodological*, vol. 105, pp. 86–119, 2017.

[19] H. Yang and J. Zhou, "Optimal traffic counting locations for origin–destination matrix estimation," *Transportation Research Part B*, vol. 32, no. 2, pp. 109–126, 1998.

[20] L. Gan, H. Yang, and S. Wong, "Traffic counting location and error bound in origin-destination matrix estimation problems," *Journal of Transportation Engineering*, vol. 131, no. 7, pp. 524–534, 2005.

[21] R. Mínguez, S. Sánchez-Cambronero, E. Castillo, and P. Jiménez, "Optimal traffic plate scanning location for od trip matrix and route estimation in road networks," *Transportation Research Part B*, vol. 44, no. 2, pp. 282–298, 2010.

[22] F. Viti, M. Rinaldi, F. Corman, and C. M. Tampère, "Assessing partial observability in network sensor location problems," *Transportation research. Part B: methodological*, vol. 70, pp. 65–89, 2014.

[23] E. Castillo, Z. Grande, A. Calviño, W. Y. Szeto, and H. K. Lo, "A state-of-the-art review of the sensor location, flow observability, estimation, and prediction problems in traffic networks," *Journal of Sensors*, vol. 2015, pp. 1–26, 2015.

[24] S. Contreras, P. Kachroo, and S. Agarwal, "Observability and sensor placement problem on highway segments: A traffic dynamics-based approach," *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, no. 3, pp. 848–858, 2015.

[25] S. Agarwal, P. Kachroo, and S. Contreras, "A dynamic network modeling-based approach for traffic observability problem," *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, no. 4, pp. 1168–1178, 2016.

[26] S. Contreras, S. Agarwal, and P. Kachroo, "Quality of traffic observability on highways with lagrangian sensors," *IEEE Transactions on Automation Science and Engineering*, vol. 15, no. 2, pp. 761–771, 2018.

[27] B. Greenshields, J. Bibbins, W. Channing, and H. Miller, "A study of traffic capacity," in *Highway research board proceedings*, vol. 1935. National Research Council (USA), Highway Research Board, 1935.

[28] G. Gomes, R. Horowitz, A. A. Kurzhanskiy, P. Varaiya, and J. Kwon, "Behavior of the cell transmission model and effectiveness of ramp metering," *Transportation Research Part C: Emerging Technologies*, vol. 16, no. 4, pp. 485 – 513, 2008.

[29] C. F. Daganzo, "The cell transmission model: A dynamic representation of highway traffic consistent with the hydrodynamic theory," *Transportation Research Part B: Methodological*, vol. 28, no. 4, pp. 269 – 287, 1994.

[30] L. Muñoz, X. Sun, R. Horowitz, and L. Alvarez, "Traffic density estimation with the cell transmission model," in *Proceedings of the 2003 American Control Conference, 2003.*, vol. 5. IEEE, 2003, pp. 3750–3755.

[31] L. Muñoz, X. Sun, R. Horowitz, and L. Alvarez, "Piecewise-linearized cell transmission model and parameter calibration methodology," *Transportation Research Record*, vol. 1965, no. 1, pp. 183–191, 2006.

[32] A. Ferrara, S. Sacone, S. Siri, C. Vivas, and F. R. Rubio, "Switched observer-based ramp metering controllers for freeway systems," in *2016 IEEE 55th Conference on Decision and Control (CDC)*. IEEE, 2016, pp. 6777–6782.

[33] Y. Guo, "Dynamic-model-based switched proportional-integral state observer design and traffic density estimation for urban freeway," *European Journal of Control*, vol. 44, pp. 103–113, 2018.

[34] G. C. Gomes, "Optimization and microsimulation of on -ramp metering for congested freeways," Ph.D. dissertation, 2004.

[35] G. Gomes and R. Horowitz, "Optimal freeway ramp metering using the asymmetric cell transmission model," *Transportation Research Part C: Emerging Technologies*, vol. 14, no. 4, pp. 244–262, 2006.

[36] C. Letellier, L. A. Aguirre, and J. Maquet, "Relation between observability and differential embeddings for nonlinear dynamics," *Phys. Rev. E*, vol. 71, p. 066213, Jun 2005.

[37] A. J. Whalen, S. N. Brennan, T. D. Sauer, and S. J. Schiff, "Observability and controllability of nonlinear networks: The role of symmetry," *Phys. Rev. X*, vol. 5, p. 011005, Jan 2015.

[38] S. Lall, J. E. Marsden, and S. Glavaški, "Empirical model reduction of controlled nonlinear systems," *IFAC Proceedings Volumes*, vol. 32, no. 2, pp. 2598 – 2603, 1999, 14th IFAC World Congress 1999, Beijing, Chia, 5-9 July.

[39] M. Serpas, G. Hackebeil, C. Laird, and J. Hahn, "Sensor location for nonlinear dynamic systems via observability analysis and max-det optimization," *Computers & Chemical Engineering*, vol. 48, pp. 105 – 112, 2013.

[40] J. Qi, K. Sun, and W. Kang, "Optimal pmu placement for power system dynamic state estimation by using empirical observability gramian," *IEEE Transactions on power Systems*, vol. 30, no. 4, pp. 2041–2054, 2015.

[41] A. K. Singh and J. Hahn, "Sensor location for stable nonlinear dynamic systems: Multiple sensor case," *Industrial & engineering chemistry research*, vol. 45, no. 10, pp. 3615–3623, 2006.

[42] A. Haber, F. Molnar, and A. E. Motter, "State observation and sensor selection for nonlinear networks," *IEEE Transactions on Control of Network Systems*, vol. 5, no. 2, pp. 694–708, June 2018.

[43] S. D. Bopardikar, O. Ennasr, and X. Tan, "Randomized sensor selection for nonlinear systems with application to target localization," *IEEE Robotics and Automation Letters*, vol. 4, no. 4, pp. 3553–3560, Oct 2019.

[44] S. A. Nugroho and A. F. Taha, "Sensor placement strategies for some classes of nonlinear dynamic systems via lyapunov theory," in *2019 IEEE 58th Conference on Decision and Control (CDC)*, 2019, pp. 4551–4556.

[45] M. J. Lighthill and G. B. Whitham, "On kinematic waves ii. a theory of traffic flow on long crowded roads," *Proc. R. Soc. Lond. A*, vol. 229, no. 1178, pp. 317–345, 1955.

[46] P. I. Richards, "Shock waves on the highway," *Operations research*, vol. 4, no. 1, pp. 42–51, 1956.

[47] S. C. Vishnoi, S. A. Nugroho, A. F. Taha, C. G. Claudel, and T. Banerjee, "Asymmetric cell transmission model-based, ramp-connected robust traffic density estimation under bounded disturbances," in *2020 American Control Conference, ACC 2020, Denver, CO, USA, July 1-3, 2020*. IEEE, 2020, pp. 1197–1202. [Online]. Available: https://doi.org/10.23919/ACC45564.2020.9147391

[48] S. Hanba, "On the "uniform" observability of discrete-time nonlinear systems," *IEEE Transactions on Automatic Control*, vol. 54, no. 8, pp. 1925–1928, 2009.

[49] A. J. Krener and K. Ide, "Measures of unobservability," in *Proceedings of the 48h IEEE Conference on Decision and Control (CDC) held jointly with 2009 28th Chinese Control Conference*, 2009, pp. 6401–6406.

[50] J. Löfberg, "YALMIP: A toolbox for modeling and optimization in matlab," in *Proc. IEEE Int. Symp. Computer Aided Control Syst. Design*, 2004, pp. 284–289.

[51] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming, version 2.2," http://cvxr.com/cvx, 2020.

[52] S. A. Nugroho, A. F. Taha, and C. G. Claudel, "A control-theoretic approach for scalable and robust traffic density estimation using convex optimization," *IEEE Transactions on Intelligent Transportation Systems*, pp. 1–15, November 2019.

[53] S. C. Vishnoi, S. A. Nugroho, A. F. Taha, C. Claudel, and T. Banerjee, "Asymmetric cell transmission model-based, ramp-connected robust traffic density estimation under bounded disturbances," in *2020 Annual American Control Conference (ACC)*, 2020, to appear.

[54] E. D. Andersen and K. D. Andersen, *The Mosek Interior Point Optimizer for Linear Programming: An Implementation of the Homogeneous Algorithm*. Boston, MA: Springer US, 2000, pp. 197–232.

[55] T. F. Coleman and Y. Li, "An interior trust region approach for nonlinear minimization subject to bounds," *SIAM Journal on Optimization*, vol. 6, no. 2, pp. 418–445, 1996.

[56] L. Lovász, *Submodular functions and convexity*. Berlin, Heidelberg: Springer Berlin Heidelberg, 1983, pp. 235–257.

[57] L. Yang, D. Sun, and K.-C. Toh, "Sdpnal+: a majorized semismooth newton-cg augmented lagrangian method for semidefinite programming with nonnegative constraints," *Mathematical Programming Computation*, vol. 7, no. 3, pp. 331–366, 2015.

[58] A. Chakrabarty, S. H. Żak, and S. Sundaram, "State and unknown input observers for discrete-time nonlinear systems," in *2016 IEEE 55th Conference on Decision and Control (CDC)*, Dec 2016, pp. 7111–7116.

Sebastian A. Nugroho received the B.S. and M.S. degrees in Electrical Engineering from Institut Teknologi Bandung (ITB), Indonesia in 2012 and 2014. He is a Ph.D. candidate at the Department of Electrical and Computer Engineering, University of Texas at San Antonio, USA. His research interests are control theory and engineering optimization with applications to cyber-physical systems. He received the Valero PhD Competitive Research Scholarship Awards from 2017 and 2019 and the Outstanding Graduate Research Awards in 2019.

Suyash C. Vishnoi is a graduate student in Civil, Architectural and Environmental Engineering at the University of Texas, Austin. He received the B.Tech. degree in Civil Engineering from the Indian Institute of Technology, Roorkee in 2018. He is a recipient of the Talbot S. Huff, Sr. Highway Engineering Graduate Fellowship from the Cockrell School of Engineering at UT Austin. His research interests include state estimation and control of traffic on highway networks.

Ahmad F. Taha is an assistant professor with the Department of Electrical and Computer Engineering at the University of Texas, San Antonio. He received the B.E. and Ph.D. degrees in Electrical and Computer Engineering from the American University of Beirut, Lebanon in 2011 and Purdue University, West Lafayette, Indiana in 2015. Dr. Taha is interested in understanding how complex cyber-physical systems (CPS) operate, behave, and *misbehave*. His research focus includes optimization, control, and security of CPSs with applications to power, water, and transportation networks. Dr. Taha is an editor of IEEE Transactions on Smart Grid and the editor of the IEEE Control Systems Society Electronic Letter (E-Letter).

Christian G. Claudel is an Assistant Professor of Civil, Architectural and Environmental Engineering at UT-Austin. He received the PhD degree in Electrical Engineering from UC-Berkeley in 2010, and the MS degree in Plasma Physics from Ecole Normale Superieure de Lyon in 2004. He received the Leon Chua Award from UC-Berkeley in 2010 for his work on the Mobile Millennium traffic monitoring system. His research interests include control and estimation of distributed parameter systems, wireless sensor networks and unmanned aerial vehicles.

Taposh Banerjee received the M.E. degree in electrical communication engineering from the Indian Institute of Science, Bangalore, Bengaluru, India, and the Ph.D. degree in electrical and computer engineering from the University of Illinois at Urbana-Champaign, Champaign, IL, USA. He is currently an Assistant Professor with the Department of Electrical and Computer Engineering, the University of Texas at San Antonio, San Antonio, TX, USA. He has previously held postdoctoral positions with Harvard University, MIT, Rutgers University, and the University of Michigan. His research interests include sequential analysis and high-dimensional statistics, with applications to biology, medicine, cyber-physical systems, and data science. He is a recipient of the Abraham Wald Prize in Sequential Analysis 2016.

Appendix A
State-Space Equation Parameters

In this section, we present the method for deriving the state-space equation (10). For ease of reading, the time parameter is omitted from the notations for all the discrete time variables. Also, the state variables are written in terms of the traffic density variables that they represent. To obtain the equation (10), we use the analytical expression for the non-linear $\min(\cdot,\cdot)$ function given as

$$\min(a,b) = \frac{1}{2}\left(\, a + b - |\, a - b \,|\right). \tag{27}$$

By applying (27) to the non-linear state evolution equations (9) and simplifying, we can break the equations down to linear and non-linear parts that are lumped into the state-space equation parameters which are given below:
$\boldsymbol{A} \in \mathbb{R}^{n\times n}$, $n = N + N_I + N_O$. The elements of $\boldsymbol{A}$ can be denoted by $a_{i,j}$ where $, i, j \in [1, n]$, and are given as follows

- $i = 1$
$$a_{1,1} = 1 - \frac{T}{4l}\left(v_f + w_c\right), a_{1,2} = \frac{T}{4l}\left(w_c - \frac{\xi_2}{2}\right),$$
$$a_{1,N+1} = \frac{T}{4l} v_f$$
- $i = N$
$$a_{N,N-1} = \frac{T\bar{\beta}_{N-1}}{8l} v_f, a_{N,N} = 1 - \frac{T}{4l}\left(v_f + w_c\right),$$
$$a_{N,n} = -\frac{T\bar{\beta}_{N-1}}{4l\beta_{N-1}} w_c$$
- $i \in \Omega \setminus \Omega_I \cup \Omega_O, i \neq \{1, N\}, j \in \hat{\Omega}, k \in \check{\Omega}$
$$a_{i,i-1} = \frac{T\bar{\beta}_{i-1}}{8l} v_f, a_{i,i} = 1 - \frac{T}{4l}\left(v_f + w_c\right),$$
$$a_{i,i+1} = \frac{T}{4l}\left(w_c - \frac{\xi_{i+1}}{2}\right), a_{i,N+j} = \frac{T}{4l} v_f,$$
$$a_{i,N+N_I+k} = -\frac{T\bar{\beta}_{i-1}}{4l\beta_{i-1}} w_c$$
- $i \in \Omega_I, j \in \hat{\Omega}$
$$a_{i,i-1} = \frac{T}{4l} v_f, a_{i,i} = 1 - \frac{T}{4l}\left(v_f + w_c + \frac{\xi_i}{2}\right),$$
$$a_{i,i+1} = \frac{T}{4l} w_c, a_{i,N+j} = \frac{T}{4l} v_f$$
- $i \in \Omega_O, j \in \check{\Omega}$
$$a_{i,i-1} = \frac{T}{4l} v_f, a_{i,i} = 1 - \frac{T}{4l}\left(\frac{1}{2} v_f + w_c\right),$$
$$a_{i,i+1} = \frac{T}{4l\bar{\beta}_i} w_c, a_{i,N+N_I+j} = \frac{T}{4l\beta_i} w_c$$
- $i = N + \bar{i}, \bar{i} \in \hat{\Omega}, j \in \Omega_I$
$$a_{i,i} = 1 - \frac{T}{2l}\left(v_f + \frac{w_c}{2}\right), a_{i,j} = \frac{T\xi_j}{4l}$$
- $i = N + N_I + \bar{i}, \bar{i} \in \check{\Omega}, j \in \Omega_O$
$$a_{i,i} = 1 - \frac{T}{4l}\left(v_f + w_c\right), a_{i,j} = \frac{T\beta_j}{8l} v_f,$$
$$a_{i,j+1} = -\frac{T\beta_j}{4l\bar{\beta}_j} w_c$$

The remaining elements of $\boldsymbol{A}$ that are not assigned a value above are equal to 0.

$\boldsymbol{f} = [f_1 \;\; f_2 \;\; \dots \;\; f_n]^T \in \mathbb{R}^g$ where $f_i, i \in [1, n]$ are column vectors of different but fixed lengths corresponding to the different cases of mainline sections and ramps, and g is the sum of lengths of all the column vectors. The elements of each column vector f_i can be denoted by $f_{j,i}$ where j is in the range of the length of f_i, and are defined as follows:

- $i = 1$

$$f_{1,1} = \frac{1}{w_c l} \mid f_{in} - \sigma_1 \mid, f_{2,1} = \frac{1}{v_f w_c l} \mid \delta_1 - \sigma_2 \mid,$$
$$f_{3,1} = \frac{1}{w_c l} \mid w_c\left(\rho_m - \rho_1\right) - v_f \rho_c \mid,$$
$$f_{4,1} = \frac{1}{v_f l} \mid v_f \rho_1 - v_f \rho_c \mid,$$
$$f_{5,1} = \frac{1}{w_c l} \mid w_c\left(\rho_m - \rho_2\right) - v_f \rho_c \mid,$$
$$f_{6,1} = \frac{1}{v_f w_c l} \mid v_f \hat{\rho}_2 - \frac{\xi_2}{w_c} \min\left(w_c\left(\rho_m - \rho_2\right), v_f \rho_c\right) \mid$$
$$f_{7,1} = \frac{T}{4l}\left(2 f_{in} + \left(\frac{\xi_2}{2 w_c} - 1\right) v_f \rho_c + \frac{\xi_2}{2} \rho_m\right)$$

- $i = N$

$$f_{1,N} = \frac{1}{v_f w_c l} \mid \delta_{N-1} - \sigma_N \mid, f_{2,N} = \frac{1}{v_f l} \mid \delta_N - f_{out} \mid,$$
$$f_{3,N} = \frac{1}{v_f w_c l} \mid \bar{\beta}_{N-1} \min\left(v_f \rho_{N-1}, v_f \rho_c\right)$$
$$- \frac{\bar{\beta}_{N-1}}{\beta_{N-1}} \check{\sigma}_{N-1} \mid, f_{4,N} = \frac{1}{w_c l} \mid w_c\left(\rho_m - \rho_N\right) - v_f \rho_c \mid,$$
$$f_{5,N} = \frac{1}{v_f l} \mid v_f \rho_N - v_f \rho_c \mid,$$
$$f_{6,N} = \frac{1}{v_f l} \mid v_f \rho_{N-1} - v_f \rho_c \mid,$$
$$f_{7,N} = \frac{T}{4l}\left(\frac{\bar{\beta}_{N-1}}{2} v_f \rho_c + \left(\frac{\bar{\beta}_{N-1}}{\beta_{N-1}} + 1\right) w_c \rho_m\right)$$

- $i \in \Omega \setminus \Omega_I \cup \Omega_O, i \neq \{1, N\}$

$$f_{1,i} = \frac{1}{v_f w_c l} \mid \delta_{i-1} - \sigma_i \mid, f_{2,i} = \frac{1}{v_f w_c l} \mid \delta_i - \sigma_{i+1} \mid,$$
$$f_{3,i} = \frac{1}{v_f w_c l} \mid \bar{\beta}_{i-1} \min(v_f \rho_{i-1}, v_f \rho_c)$$
$$- \frac{\bar{\beta}_{i-1}}{\beta_{i-1}} w_c(\rho_m - \check{\rho}_{i-1}) \mid,$$
$$f_{4,i} = \frac{1}{w_c l} \mid w_c(\rho_m - \rho_i) - v_f \rho_c \mid,$$
$$f_{5,i} = \frac{1}{v_f l} \mid v_f \rho_i - v_f \rho_c \mid,$$
$$f_{6,i} = \frac{1}{w_c l} \mid w_c(\rho_m - \rho_{i+1}) - v_f \rho_c \mid,$$
$$f_{7,i} = \frac{1}{v_f w_c l} \mid v_f \hat{\rho}_{i+1}$$
$$- \frac{\xi_{i+1}}{w_c} \min(w_c(\rho_m - \rho_{i+1}), v_f \rho_c) \mid,$$
$$f_{8,i} = \frac{1}{v_f l} \mid v_f \rho_{i-1} - v_f \rho_c \mid,$$
$$f_{9,i} = \frac{T}{4l}\left(\left(\frac{\bar{\beta}_{i-1}}{2} + \frac{\xi_{i+1}}{2 w_c} - 1\right) v_f \rho_c + \left(\frac{\bar{\beta}_{i-1}}{\beta_{i-1}} + \frac{\xi_{i+1}}{2 w_c}\right) w_c \rho_m\right)$$

- $i \in \Omega_I$

$$f_{1,i} = \frac{1}{v_f w_c l} \mid \delta_{i-1} - \sigma_i \mid, f_{2,i} = \frac{1}{v_f w_c l} \mid \delta_i - \sigma_{i+1} \mid,$$
$$f_{3,i} = \frac{1}{v_f l} \mid v_f \rho_{i-1} - v_f \rho_c \mid,$$
$$f_{4,i} = \frac{1}{w_c l} \mid w_c(\rho_m - \rho_i) - v_f \rho_c \mid,$$
$$f_{5,i} = \frac{1}{v_f l} \mid v_f \rho_i - v_f \rho_c \mid,$$
$$f_{6,i} = \frac{1}{w_c l} \mid w_c(\rho_m - \rho_{i+1}) - v_f \rho_c \mid,$$
$$f_{7,i} = \frac{1}{v_f w_c l} \mid v_f \hat{\rho}_i - \frac{\xi_i}{w_c} \min(w_c(\rho_m - \rho_i), v_f \rho_c) \mid,$$
$$f_{8,i} = \frac{1}{w_c l} \mid w_c(\rho_m - \rho_i) - v_f \rho_c \mid$$
$$f_{9,i} = \frac{T \xi_i}{8l}\left(\rho_m + \frac{v_f \rho_c}{w_c}\right)$$

- $i \in \Omega_O$

$$f_{1,i} = \frac{1}{v_f w_c l} \mid \delta_{i-1} - \sigma_i \mid, f_{2,i} = \frac{1}{v_f w_c l} \mid \delta_i - \sigma_{i+1} \mid,$$
$$f_{3,i} = \frac{1}{v_f l} \mid v_f \rho_{i-1} - v_f \rho_c \mid,$$
$$f_{4,i} = \frac{1}{w_c l} \mid w_c(\rho_m - \rho_i) - v_f \rho_c \mid,$$
$$f_{5,i} = \frac{1}{v_f w_c l} \mid \min(v_f \rho_i, v_f \rho_c) - \frac{1}{\beta_i} w_c(\rho_m - \check{\rho}_i) \mid,$$
$$f_{6,i} = \frac{1}{w_c l} \mid w_c(\rho_m - \rho_{i+1}) - v_f \rho_c \mid,$$
$$f_{7,i} = \frac{1}{v_f l} \mid v_f \rho_i - v_f \rho_c \mid,$$
$$f_{8,i} = \frac{T}{4l}\left(\left(\frac{3}{2} - \frac{1}{\bar{\beta}_i}\right) v_f \rho_c + \left(1 - \frac{1}{\beta_i} - \frac{1}{\bar{\beta}_i}\right) w_c \rho_m\right)$$

- $i = N + \bar{i}, \bar{i} \in \hat{\Omega}, j \in \Omega_I$

$$f_{1,i} = \frac{1}{w_c l} \mid \hat{\sigma}_j - \hat{f}_j \mid,$$
$$f_{2,i} = \frac{1}{v_f w_c l} \mid v_f \hat{\rho}_j - \frac{\xi_j}{w_c} \min(w_c(\rho_m - \rho_j), v_f \rho_c) \mid,$$
$$f_{3,i} = \frac{1}{w_c l} \mid w_c(\rho_m - \hat{\rho}_j) - v_f \rho_c \mid,$$
$$f_{4,i} = \frac{1}{w_c l} \mid w_c(\rho_m - \rho_j) - v_f \rho_c \mid,$$
$$f_{5,i} = \frac{T}{2l}\left(\left(\frac{1}{2} - \frac{\xi_j}{2 w_c}\right) v_f \rho_c + \left(\frac{1}{2} - \frac{\xi_j}{2 w_c}\right) w_c \rho_m\right)$$

- $i = N + N_I + \bar{i}, \bar{i} \in \check{\Omega}, j \in \Omega_O$

$$f_{1,i} = \frac{1}{v_f w_c l} \mid \delta_j - \sigma_{j+1} \mid, f_{2,i} = \frac{1}{v_f l} \mid \check{\delta}_j - \check{f}_j \mid,$$
$$f_{3,i} = \frac{1}{v_f w_c l} \mid \min(v_f \rho_j, v_f \rho_c) - \frac{1}{\beta_j} \check{\sigma}_j \mid,$$
$$f_{4,i} = \frac{1}{w_c l} \mid w_c(\rho_m - \rho_{j+1}) - v_f \rho_c \mid,$$
$$f_{5,i} = \frac{1}{v_f l} \mid v_f \check{\rho}_j - v_f \rho_c \mid, f_{6,i} = \frac{1}{v_f l} \mid v_f \rho_j - v_f \rho_c \mid,$$

$$f_{7,i} = \frac{T}{2l}\left(\left(\frac{\beta_j}{4} + \frac{\beta_j}{2\bar{\beta}_j} - \frac{1}{2}\right) v_f \rho_c + \left(\frac{1}{2} + \frac{\beta_j}{2\bar{\beta}_j}\right) w_c \rho_m\right)$$

$\boldsymbol{G} \in \mathbb{R}^{n\times g}$ is of the form

$$\begin{bmatrix} G_1 & 0 & \cdots & \cdots & 0 \\ 0 & G_2 & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & G_{n-1} & 0 \\ 0 & \cdots & \cdots & 0 & G_n \end{bmatrix}, \text{ where } G_i, i \in [1, n]$$

are row vectors with the same number of columns as the number of rows in f_i. The elements of each row vector G_i can be denoted by $G_{i,j}$, where j is in the range of the length of G_i, and are defined as follows:

- $i = 1$

$$G_i = \frac{Tv_f w_c}{4}\left[-\frac{2}{v_f} \quad 2 \quad -\frac{1}{v_f} \quad \frac{1}{w_c} \quad \left(\frac{1}{v_f} - \frac{\xi_2}{2v_f w_c}\right) \quad -1 \quad \frac{4}{Tv_f w_c}\right]$$

- $i = N$

$$G_i = \frac{Tv_f w_c}{4}\left[-2 \quad \frac{2}{w_c} \quad -1 \quad -\frac{1}{v_f} \quad \frac{1}{w_c} \quad -\frac{\bar{\beta}_{N-1}}{2w_c} \quad \frac{4}{Tv_f w_c}\right]$$

- $i \in \Omega \setminus \Omega_I \cup \Omega_O, i \neq \{1, N\}$

$$G_i = \frac{Tv_f w_c}{4}\left[-2 \quad 2 \quad -1 \quad -\frac{1}{v_f} \quad \frac{1}{w_c} \quad \left(\frac{1}{v_f} - \frac{\xi_{i+1}}{2v_f w_c}\right) \quad -1 \quad -\frac{\bar{\beta}_{i-1}}{2w_c} \quad \frac{4}{Tv_f w_c}\right]$$

- $i \in \Omega_I$

$$G_i = \frac{Tv_f w_c}{4}\left[-2 \quad 2 \quad -\frac{1}{w_c} \quad -\frac{1}{v_f} \quad \frac{1}{w_c} \quad \frac{1}{v_f} \quad -1 \quad -\frac{\xi_i}{2v_f w_c} \quad \frac{4}{Tv_f w_c}\right]$$

- $i \in \Omega_O$

$$G_i = \frac{Tv_f w_c}{4}\left[-2 \quad \frac{2}{\bar{\beta}_i} \quad -\frac{1}{w_c} \quad -\frac{1}{v_f} \quad 1 \quad \frac{1}{v_f \bar{\beta}_i} \quad \frac{1}{2w_c} \quad \frac{4}{Tv_f w_c}\right]$$

- $i = N + \bar{i}, \bar{i} \in \hat{\Omega}, j \in \Omega_I$

$$G_i = \frac{Tw_c}{4}\left[-2 \quad 2v_f \quad -1 \quad \frac{\xi_j}{w_c} \quad \frac{4}{Tw_c}\right]$$

- $i = N + N_I + \bar{i}, \bar{i} \in \check{\Omega}, j \in \Omega_O$

$$G_i = \frac{Tv_f}{4}\left[-\frac{2w_c\beta_j}{\bar{\beta}_j} \quad 2 \quad -w_c\beta_j \quad -\frac{w_c\beta_j}{v_f\bar{\beta}_j} \quad 1 \quad -\frac{\beta_j}{2} \quad \frac{T}{4v_f}\right]$$

$\boldsymbol{B_u} \in \mathbb{R}^{n\times(2+N_I+N_O)}$. The elements of $\boldsymbol{B_u}$ can be denoted by $\boldsymbol{B}_{i,j}$ and are given by the following function

$$\boldsymbol{B}_{i,j} = \begin{cases} \frac{T}{2l} f_{in}, & \text{if } i = j = 1 \\ -\frac{T}{2l} f_{out}, & \text{if } i = N, j = 2 \\ \frac{T}{2l}\hat{f}_k, & \text{if } i = N + \bar{i}, j = 2 + \bar{i}, \bar{i} \in \hat{\Omega}, k \in \Omega_I \\ -\frac{T}{2l}\check{f}_k, & \text{if } i = N + N_I + \bar{i}, j = 2 + N_I + \bar{i}, \\ & \bar{i} \in \check{\Omega}, k \in \Omega_I \\ 0, & \text{otherwise.} \end{cases}$$

Appendix B
Calculation of Jacobian Matrix

To calculate the Jacobian matrix for the non-linear vector valued function $\boldsymbol{f}(\boldsymbol{x}, \boldsymbol{u})$ containing $|\cdot|$ and $\min(\cdot,\cdot)$ functions, we first replace all the $\min(\cdot,\cdot)$ functions by their analytical form (27) and then calculate the derivative using the definition of the derivative for $|\cdot|$ function given as

$$\frac{\partial \mid x \mid}{\partial x} = \frac{x}{\mid x \mid}, x \neq 0 \tag{28}$$

At $x = 0$, we simply take the derivative to be 0. The elements of the Jacobian matrix can be calculated using (28) along with the chain rule of differentiation. For example, the element $f_{2,i}$ in $f_i, i \in \Omega_I, j = N + \bar{j}, \bar{j} \in \hat{\Omega}$ can be differentiated with respect to the state variable x_i, for $\delta_{i-1} \neq \sigma_i$, as follows:

$$\frac{\partial f_{1,i}}{\partial x_i} = \frac{1}{v_f w_c l} \cdot \frac{\partial \mid \delta_{i-1} - \sigma_i \mid}{\partial x_i} = -\frac{1}{v_f w_c l} \cdot \frac{\delta_{i-1} - \sigma_i}{\mid \delta_{i-1} - \sigma_i \mid} \cdot \frac{\partial \sigma_i}{\partial x_i},$$

where

$$\frac{\partial \sigma_i}{\partial x_i} = \frac{\partial \min(w_c(\rho_m - x_i), v_f\rho_c)}{\partial x_i} - \frac{\partial r_i}{\partial x_i}.$$

Here, for $w_c(\rho_m - x_i) \neq v_f\rho_c$,

$$\frac{\partial \min(w_c(\rho_m - x_i), v_f\rho_c)}{\partial x_i} = -\frac{w_c}{2}\left(1 - \frac{w_c(\rho_m - x_i) - v_f\rho_c}{\mid w_c(\rho_m - x_i) - v_f\rho_c \mid}\right),$$

and for $v_f x_j \neq \frac{\xi_i}{w_c}\min(w_c(\rho_m - x_i), v_f\rho_c)$,

$$\frac{\partial r_i}{\partial x_i} = -\frac{\xi_i}{2w_c} \cdot \frac{w_c}{2}\left(1 - \frac{w_c(\rho_m - x_i) - v_f\rho_c}{\mid w_c(\rho_m - x_i) - v_f\rho_c \mid}\right) \cdot \left(1 + \frac{v_f x_j - \frac{\xi_i}{w_c}\min(w_c(\rho_m - x_i), v_f\rho_c)}{\mid v_f x_j - \frac{\xi_i}{w_c}\min(w_c(\rho_m - x_i), v_f\rho_c) \mid}\right).$$

Therefore, the derivative can be written as

$$\frac{\partial f_{1,i}}{\partial x_i} = \frac{1}{2v_f l} \cdot \frac{\delta_{i-1} - \sigma_i}{\mid \delta_{i-1} - \sigma_i \mid}\left(1 - \frac{w_c(\rho_m - x_i) - v_f\rho_c}{\mid w_c(\rho_m - x_i) - v_f\rho_c \mid}\right) \cdot \left(1 - \frac{\xi_i}{2w_c}\left(1 + \frac{v_f x_j - \frac{\xi_i}{w_c}\min(w_c(\rho_m - x_i), v_f\rho_c)}{\mid v_f x_j - \frac{\xi_i}{w_c}\min(w_c(\rho_m - x_i), v_f\rho_c) \mid}\right)\right).$$

Note that the above expression is valid when $\delta_{i-1} \neq \sigma_i$, $w_c(\rho_m - x_i) \neq v_f\rho_c$ and $v_f x_j \neq \frac{\xi_i}{w_c}\min(w_c(\rho_m - x_i), v_f\rho_c)$.

If, on the other hand, $w_c(\rho_m - x_i) = v_f\rho_c$, then

$$\frac{\partial f_{1,i}}{\partial x_i} = \frac{1}{2v_f l} \cdot \frac{\delta_{i-1} - \sigma_i}{\mid \delta_{i-1} - \sigma_i \mid} \cdot \left(1 - \frac{\xi_i}{2w_c}\left(1 + \frac{v_f x_j - \frac{\xi_i}{w_c}\min(w_c(\rho_m - x_i), v_f\rho_c)}{\mid v_f x_j - \frac{\xi_i}{w_c}\min(w_c(\rho_m - x_i), v_f\rho_c) \mid}\right)\right),$$

if $v_f x_j = \frac{\xi_i}{w_c}\min(w_c(\rho_m - x_i), v_f\rho_c)$, then

$$\frac{\partial f_{1,i}}{\partial x_i} = \frac{1}{2v_f l} \cdot \frac{\delta_{i-1} - \sigma_i}{\mid \delta_{i-1} - \sigma_i \mid}\left(1 - \frac{\xi_i}{2w_c}\right) \cdot \left(1 - \frac{w_c(\rho_m - x_i) - v_f\rho_c}{\mid w_c(\rho_m - x_i) - v_f\rho_c \mid}\right).$$

if both $w_c(\rho_m - x_i) = v_f\rho_c$, and $v_f x_j = \frac{\xi_i}{w_c}\min(w_c(\rho_m - x_i), v_f\rho_c)$, then

$$\frac{\partial f_{1,i}}{\partial x_i} = \frac{1}{2v_f l} \cdot \frac{\delta_{i-1} - \sigma_i}{\mid \delta_{i-1} - \sigma_i \mid}\left(1 - \frac{\xi_i}{2w_c}\right).$$

and if $\delta_{i-1} = \sigma_i$, then $\frac{\partial f_{1,i}}{\partial x_i} = 0$. Similarly, other elements of the Jacobian matrix can also be derived.

Appendix C
Calculation of Lipschitz constant

In this section we present the methodology for calculating a Lipschitz constant γ that satisfies

$$||\boldsymbol{f}(\boldsymbol{x}, u) - \boldsymbol{f}(\hat{\boldsymbol{x}}, u)||_2 \leq \gamma||\boldsymbol{x} - \hat{\boldsymbol{x}}||_2$$

To calculate γ, we use the following inequalities related to the $|\cdot|$ function:

$$|| a \mid - \mid b || \;\leq\; \mid a - b \mid \tag{29a}$$

$$\mid a \pm b \mid \;\leq\; \mid a \mid + \mid b \mid \tag{29b}$$

$$\mid a_1 \mid + \mid a_2 \mid \;\leq\; \sqrt{2}(a_1^2 + a_2^2)^{1/2} = \sqrt{2}||\boldsymbol{a}||_2 \tag{29c}$$

By using one or more of these inequalities we can calculate the Lipschitz constant for each component of the vector function $\boldsymbol{f}(\cdot)$, and combine them to obtain the final γ.

For example, let us calculate the the Lipschitz constant for the expression $f_{1,i}, i \in \Omega_I$, which, for $j = N + \bar{j}, \bar{j} \in \hat{\Omega}$, is given as

$$f_{1,i} = \frac{1}{v_f w_c l} \mid \min(v_f x_{i-1}, v_f\rho_c) - \min(w_c(\rho_m - x_i), v_f\rho_c) - \min\left(v_f x_j, \frac{\xi_i}{w_c}\min(w_c(\rho_m - x_i), v_f\rho_c)\right) \mid .$$

Using (27), (29a), and (29b), and simplifying, we can write the following inequality for $\mid f_{1,i} - \hat{f}_{1,i} \mid$:

$$\mid f_{1,i} - \hat{f}_{1,i} \mid \leq \frac{1}{2v_f w_c l}(2v_f \mid x_{i-1} - \hat{x}_{i-1} \mid + 2\Big(1 - \frac{\xi_i}{2w_c}\Big)w_c \mid x_i - \hat{x}_i \mid + 2v_f \mid x_j - \hat{x}_j \mid + \xi_i \mid x_i - \hat{x}_i \mid).$$

By applying (29c) to combine the terms with the same coefficients and then adding up all the coefficients, we can write the above inequality as

$$\mid f_{1,i} - \hat{f}_{1,i} \mid \leq \frac{1}{l}\Big(\frac{\sqrt{2}}{w_c} + \frac{1}{v_f}\Big)||\boldsymbol{x} - \hat{\boldsymbol{x}}||_2.$$

Let us call the coefficient of $||\boldsymbol{x} - \hat{\boldsymbol{x}}||_2$ in the above equation as $\gamma_{1,i}$. We can obtain the γ_i corresponding to f_i by adding these coefficient values for all the components of the vector f_i.

The values of γ_i for different cases of mainline sections and ramps are given below:

- $i = 1$
$$\frac{1}{l}\Big(\frac{1+\sqrt{2}}{w_c} + \frac{1}{v_f} + \frac{\xi_2}{v_f w_c} + 4\Big)$$
- $i = N$
$$\frac{1}{l}\Big(\frac{2\bar{\beta}_{N-1}}{w_c} + \Big(1 + 2\frac{\bar{\beta}_{N-1}}{\beta_{N-1}}\Big)\frac{1}{v_f} + 4\Big)$$
- $i \in \Omega \setminus \Omega_I \cup \Omega_O, i \neq \{1, N\}$
$$\frac{1}{l}\Big(\frac{1+\sqrt{2}+2\bar{\beta}_{i-1}}{w_c} + \Big(1 + \frac{\bar{\beta}_{i-1}}{\beta_{i-1}}\Big)\frac{2}{v_f} + \frac{\xi_{i+1}}{v_f w_c} + 4\Big)$$
- $i \in \Omega_I$
$$\frac{1}{l}\Big(\frac{2+\sqrt{2}}{w_c} + \frac{2}{v_f} + \frac{\xi_i}{v_f w_c} + 5\Big)$$
- $i \in \Omega_O$
$$\frac{1}{l}\Big(\frac{2+\bar{\beta}_i}{w_c} + \Big(2 + \frac{\bar{\beta}_i}{\beta_i} + \frac{1}{\beta_i}\Big)\frac{1}{v_f} + 4\Big)$$
- $i = N + \bar{i}, \bar{i} \in \hat{\Omega}, j \in \Omega_I$
$$\frac{1}{l}\Big(\frac{1}{w_c} + \frac{\xi_j}{v_f w_c} + 3\Big)$$
- $i = N + N_I + \bar{i}, \bar{i} \in \check{\Omega}, j \in \Omega_O$
$$\frac{1}{l}\Big(\frac{1+\bar{\beta}_j}{w_c} + \Big(1 + \frac{\bar{\beta}_j}{\beta_j} + \frac{1}{\beta_j}\Big)\frac{1}{v_f} + 4\Big)$$

So, for every $f_i, i \in [1, n]$, there exists a γ_i such that$\mid f_i - \hat{f}_i \mid \leq \gamma_i||\boldsymbol{x} - \hat{\boldsymbol{x}}||_2, \gamma_i \geq 0$. Since we know that

$$||\boldsymbol{f}(x, u) - \boldsymbol{f}(\hat{x}, u)||_2^2 = \sum_{i=1}^{n} \mid f_i - \hat{f}_i \mid^2 \leq \sum_{i=1}^{n} \gamma_i^2||\boldsymbol{x} - \hat{\boldsymbol{x}}||_2^2$$

then γ for $\boldsymbol{f}(x, u)$ can be defined using function $h(\cdot)$ which is given by (30).

$$\begin{aligned}
&h\left(v_f, \omega_c, l, \xi_1, \xi_2, \ldots, \xi_{N_I}, \beta_1, \beta_2, \ldots, \beta_{N_O}\right) \\
&= \frac{1}{l}\Bigg(\Big(\frac{1+\sqrt{2}}{w_c} + \frac{1}{v_f} + \frac{\xi_2}{v_f w_c} + 4\Big)^2 \\
&+ \Big(\frac{2\bar{\beta}_{N-1}}{w_c} + \Big(1 + 2\frac{\bar{\beta}_{N-1}}{\beta_{N-1}}\Big)\frac{1}{v_f} + 4\Big)^2 \\
&+ \sum_{\substack{i \in \Omega \setminus \Omega_I \cup \Omega_O, \\ i \neq \{1,N\}}} \Big(\frac{1+2\bar{\beta}_{i-1}+\sqrt{2}}{w_c} + \Big(1 + \frac{\bar{\beta}_{i-1}}{\beta_{i-1}}\Big)\frac{2}{v_f} \\
&+ \frac{\xi_{i+1}}{v_f w_c} + 4\Big)^2 + \sum_{i \in \Omega_I}\Big(\frac{2+\sqrt{2}}{w_c} + \frac{2}{v_f} + \frac{\xi_i}{v_f w_c} + 5\Big)^2 \\
&+ \sum_{j \in \Omega_I}\Big(\frac{1}{w_c} + \frac{\xi_j}{v_f w_c} + 3\Big)^2
\end{aligned}$$

$$+\sum_{i\in\Omega_O}\Big(\frac{2+\bar{\beta}_i}{w_c}+\big(2+\frac{\bar{\beta}_i}{\beta_i}+\frac{1}{\beta_i}\big)\frac{1}{v_f}+4\Big)^2$$
$$+\sum_{j\in\Omega_O}\Big(\frac{1+\bar{\beta}_j}{w_c}+\Big(1+\frac{\bar{\beta}_j}{\beta_j}+\frac{1}{\beta_j}\Big)\frac{1}{v_f}+4\Big)^2\Bigg)^{1/2} \tag{30}$$

Appendix D
Proof of Theorem 3

Let $\boldsymbol{z}=\boldsymbol{Z}\boldsymbol{e}$ be the performance output for estimation error $\boldsymbol{e}$ and $\boldsymbol{w}$ be an unknown bounded disturbance. Let $V(\boldsymbol{e})=\boldsymbol{e}^\top\boldsymbol{P}\boldsymbol{e}$ be a Lyapunov function candidate where $\boldsymbol{P}\succ 0$. It can be shown that—for example, see [58]—the estimation error dynamics (23) is $\mathcal{L}_\infty$ stable with performance level $\mu=\sqrt{\mu_0\mu_1+\mu_2}$ if there exist constants $\mu_0,\mu_1,\mu_2\in\mathbb{R}_+$ such that

$$\mu_0\|\boldsymbol{w}\|_2^2\le V(\boldsymbol{e})\ \Rightarrow \Delta V(\boldsymbol{e})\le 0 \tag{31a}$$
$$\|\boldsymbol{z}\|_2^2\le \mu_1 V(\boldsymbol{e})+\mu_2\|\boldsymbol{w}\|_2^2, \tag{31b}$$

for all $t\ge 0$ where $\Delta V(\boldsymbol{e})\triangleq V(\boldsymbol{e}[k+1])-V(\boldsymbol{e}[k])$. First, realize that (31a) holds if there exists $\alpha>0$ such that

$$\begin{aligned}&V(\boldsymbol{e}[k+1])-V(\boldsymbol{e}[k])+\alpha\Big(V(\boldsymbol{e}[k])-\mu_0\|\boldsymbol{w}[k]\|_2^2\Big)\le 0\\ &\Leftrightarrow \boldsymbol{e}^\top[k+1]\boldsymbol{P}\boldsymbol{e}[k+1]-(1-\alpha)\boldsymbol{e}^\top[k]\boldsymbol{P}\boldsymbol{e}[k]\\ &\qquad -\alpha\mu_0\boldsymbol{w}^\top[k]\boldsymbol{w}[k]\le 0\\ &\Leftrightarrow \mathrm{Blkdiag}\left([(\alpha-1)\boldsymbol{P},\boldsymbol{O},-\alpha\mu_0\boldsymbol{I}]\right)+\boldsymbol{\Psi}^\top\boldsymbol{P}^{-1}\boldsymbol{\Psi}\preceq 0,\end{aligned} \tag{32}$$

where matrix $\boldsymbol{\Psi}$ is defined as

$$\boldsymbol{\Psi}\triangleq\begin{bmatrix}\boldsymbol{P}\boldsymbol{A}-\boldsymbol{P}\boldsymbol{L}\boldsymbol{C} & \boldsymbol{P}\boldsymbol{G} & \boldsymbol{P}\boldsymbol{B}_{\mathrm{w}}-\boldsymbol{P}\boldsymbol{L}\boldsymbol{D}_{\mathrm{w}}\end{bmatrix}.$$

Since the nonlinear function $\boldsymbol{f}(\cdot)$ is locally Lipschitz in $\boldsymbol{\Omega}$, then it holds that

$$\|\Delta\boldsymbol{f}[k]\|_2^2\le\gamma_l^2\|\boldsymbol{e}[k]\|_2^2\Leftrightarrow \mathrm{Blkdiag}\left(\left[-\gamma_l^2\boldsymbol{I},\boldsymbol{I},\boldsymbol{O}\right]\right)\preceq 0. \tag{33}$$

By applying the S-procedure to (32) from (33) for $\epsilon\ge 0$, we get

$$\mathrm{Blkdiag}\left(\left[(\alpha-1)\boldsymbol{P}+\epsilon\gamma_l^2\boldsymbol{I},-\epsilon\boldsymbol{I},-\alpha\mu_0\boldsymbol{I}\right]\right)+\boldsymbol{\Psi}^\top\boldsymbol{P}^{-1}\boldsymbol{\Psi}\preceq 0,$$

which, thanks to the Schur Complement, is equivalent to

$$\left[\begin{array}{c|c}\mathrm{Blkdiag}\left(\left[(\alpha-1)\boldsymbol{P}+\epsilon\gamma_l^2\boldsymbol{I},-\epsilon\boldsymbol{I},-\alpha\mu_0\boldsymbol{I}\right]\right) & *\\ \hline \boldsymbol{\Psi} & -\boldsymbol{P}\end{array}\right]\preceq 0.$$

Notice that the above is equivalent to (24b) provided that $\boldsymbol{Y}\triangleq\boldsymbol{P}\boldsymbol{L}$. Next, substituting $\boldsymbol{z}=\boldsymbol{Z}\boldsymbol{e}$ to (31b) yields

$$\begin{aligned}&\|\boldsymbol{Z}\boldsymbol{e}\|_2^2-\mu_1V(\boldsymbol{e})-\mu_2\|\boldsymbol{w}\|_2^2\le 0\\ &\Leftrightarrow \boldsymbol{e}^\top\boldsymbol{Z}^\top\boldsymbol{Z}\boldsymbol{e}-\mu_1\boldsymbol{e}^\top\boldsymbol{P}\boldsymbol{e}-\mu_2\boldsymbol{w}^\top\boldsymbol{w}\le 0\\ &\Leftrightarrow \mathrm{Blkdiag}\left(\left[-\mu_1\boldsymbol{P}+\boldsymbol{Z}^\top\boldsymbol{Z},-\mu_2\boldsymbol{I}\right]\right)\preceq 0.\end{aligned}$$

By employing congruence transformation (provided that $\mu_1>0$) and applying the Schur Complement, it is not difficult to show that the above is equivalent to (24c). Finally, the objective function (24a) tries to minimize the performance index, which dictates how much the disturbance $\boldsymbol{w}$ will impact the performance vector $\boldsymbol{z}$. This implies that if optimization problem (24) is solved, then the estimation error dynamics given in (23) is $\mathcal{L}_\infty$ stable with performance level $\mu=\sqrt{\mu_0\mu_1+\mu_2}$ and observer gain given as $\boldsymbol{L}=\boldsymbol{P}^{-1}\boldsymbol{Y}$, thus completing the proof. ■