TECHNICAL REPORT DOCUMENTATION PAGE

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
PSR-19-06	N/A	N/A

4. Title and Subtitle Implications of Information Structure in Control of Urban Traffic Networks	**5. Report Date** September 2020
	6. Performing Organization Code N/A
7. Author(s) Ketan Savla, https://orcid.org/0000-0002-1668-6380	**8. Performing Organization Report No.** PSR-19-06-TO-017
9. Performing Organization Name and Address METRANS Transportation Center University of Southern California University Park Campus, RGL 216 Los Angeles, CA 90089-0626	**10. Work Unit No.** N/A **11. Contract or Grant No.** USDOT Grant 69A3551747109 Caltrans 65A0674 (TO-017)
12. Sponsoring Agency Name and Address California Department of Transportation 1227 O Street Sacramento, CA 95843	**13. Type of Report and Period Covered** Final report (8/1/2019–7/31/2020) **14. Sponsoring Agency Code** USDOT OST-R

15. Supplementary Notes
Link to project web page: https://www.metrans.org/research/implications-of-information-structure-in-control-of-urban-traffic-networks

16. Abstract
First, we consider optimal control of traffic flow over networks using a combination of variable speed limit, ramp meter, lane-changing, and routing control. While this problem has attracted significant attention, most of the prior work has been limited to centralized or open-loop control. We propose to develop the foundations for a framework to design closed-loop control under given information structures. Our emphasis will be on computational tractability and characterization of performance gap with respect to centralized control. Second, we propose to study optimal information design to influence route choice decisions of drivers in dynamic environments. Specifically, we adopt the framework of algorithmic persuasion, under which the system planner can exploit information asymmetry about the knowledge of the real-time state of the network to release noisy information or recommend routes to the drivers in order to optimize social objective. The study of algorithmic persuasion in the context of routing games is very recent, and more so, the existing work implicitly assumes the drivers to evaluate the incentive compliant nature of the recommendations from the system planner only asymptotically, they do not consider externality from drivers who do not participate in persuasion, and assume static traffic flow models. In this project, we propose to address these shortcomings to develop foundations for algorithmic persuasion in routing games. The methodological contributions will be supplemented with case studies using traffic data from the Los Angeles area, and with simulation case studies in VISSIM.

17. Key Words	18. Distribution Statement
TRT Terms: Information systems; Traffic flow; Traffic flow rate Identifier Terms: Urban Traffic Flow Guidance Systems Subject Areas: Transportation (General); Traffic and Transport Planning	No restrictions.

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	42	N/A

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

Implications of Information Structure in Control of Urban Traffic Networks

September 2020

A Report from the Pacific Southwest Region University Transportation Center

Ketan Savla, University of Southern California

Pacific Southwest Region UTC
University Transportation Center

USC University of Southern California

Implications of Information Structure in Control of Urban Traffic Networks

Final Report

CALTRANS # TO 65A0674

September 2020

Principal Investigator :

Ketan Savla
Sonny Astani Department of Civil & Environmental Engineering
3620 S. Vermont Avenue, KAP 254A, University of Southern California
Los Angeles, CA 90089-2531
Tel: 213-740-0670 Fax: 213-744-1426
ksavla@usc.edu

Disclaimer Statement

The contents of this report reflect the views of the authors, who are responsible for the accuracy of the data and information presented herein. This document is disseminated under the sponsorship of the Department of Transportation, University Transportation Centers Program, the California Department of Transportation and the METRANS Transportation Center in the interest of information exchange. The U.S. Government, the California Department of Transportation, and the University of Southern California assume no liability for the contents or use thereof. The contents do not necessarily reflect the official views or policies of the State of California, USC, or the Department of Transportation. This report does not constitute a standard, specification, or regulation.

Abstract

In the first part, finite-time optimal feedback control for traffic networks under information constraints is studied. By utilizing the framework of multi-parametric linear programming, it is demonstrated that when cost/constraints can be modeled or approximated by piecewise-affine functions, the optimal control has a closed-form state-feedback realization. The optimal feedback control law, however, has a centralized structure and requires instantaneous access to the state of the entire network that may lead to prohibitive communication requirements in large-scale networks. We subsequently examine the design of a decentralized (sub)-optimal feedback controller with a one-hop information structure, wherein the optimum outflow rate from each segment of the network depends only on the state of that segment and the state of the segments immediately downstream. The decentralization is based on the relaxation of constraints that depend on state variables that are unavailable according to the information structure. The resulting decentralized control scheme has a simple closed-form representation and is scalable to arbitrary large networks; moreover, we demonstrate that, with respect to certain meaningful performance indexes, the performance loss due to decentralization is zero; namely, the centralized optimal controller has a decentralized realization with a one-hop information structure and is obtained at no computational/communication cost.

In the second part, we consider a routing game among non-atomic agents where link latency functions are conditional on an uncertain state of the network. All the agents have the same prior belief about the state, but only a fixed fraction receive private route recommendations or a common message, which are generated by a known randomization, referred to as *private* or *public signal* respectively. The remaining non-receiving agents choose route according to Bayes Nash flow with respect to the prior. We develop a computational approach to solve the optimal information design problem, i.e., to minimize expected social latency cost over all public or *obedient* private signals. For a fixed flow induced by non-receiving agents, design of an optimal private signal is shown to be a generalized problem of moments for polynomial link latency functions, and to admit an atomic solution with a provable upper bound on the number of atoms. This implies that, for polynomial link latency functions, information design over private and public signals, when the non-receiving agents choose route according to Bayes Nash flow, can be equivalently cast as a polynomial optimization problem. This in turn can be arbitrarily lower bounded by a known hierarchy of semidefinite relaxations. The first level of this hierarchy is shown to be exact for the basic two link case with affine latency functions, and it relies on tightening the bound on the number of atoms in the support of optimal signal. We also identify a class of private signals over which the optimal social cost is non-increasing with increasing fraction of receiving agents. This does not require the link latency functions to be polynomial, and is in contrast to existing results where the cost of receiving agents under a *fixed* signal may increase with their increasing fraction.

Contents

List of Tables

1 Disclosure

K. Savla has financial interest in Xtelligent, Inc.

2 Acknowledgements

This research was performed in collaboration with Y. Zhu at the University of Southern California, and with S. Jafari formerly with the University of Southern California.

The high performance computing support provided by USC's Center for Advanced Research Computing in running some of the simulations during the project is gratefully acknowledged.

3 Introduction

Rapid advancements in technology have facilitated a tremendous increase in the number of control/decision and sensor points in urban traffic networks, ranging from an individual driver carrying smart phone to ramp meters to city-scale traffic control center. Due to the large volume of data generation, it is computationally, and arguably even technologically, infeasible to inter-connect all the points to each other for real-time applications. Therefore, it is of interest to study performance of traffic networks under various information structures, i.e., sparse interconnection of control/decision and sensor points. In this project, we propose to study such issues under two complementary topics, namely (i) feedback control of traffic networks using ramp metering, variable speed limit and routing control, and (ii) routing control through information design.

3.1 Feedback Control of Traffic Networks

In infrastructure flow networks such as traffic networks, the primary objective for control design is to regulate the flow while optimizing a certain performance index.

In the study of fluid dynamics at macroscopic scale, the fluid is treated as a continuum and its motion is described by the *mass conservation law* stating that "the rate of change of the mass of a fluid in a fixed region is equal to the difference between the rate of mass flow into and out of the region" [1]. Let $\rho(x, t)$ and $v(x, t)$ respectively denote the mass density and the velocity vector of a fluid at time t, at position $x = [x_1, x_2, x_3]^\top$ in the three-dimensional space. With the continuum representation of the fluid, the law of mass conservation is expressed as [1]:

$$\frac{\partial \rho}{\partial t} + \mathrm{div}(\rho v) = 0, \tag{1}$$

which is balancing the rate of change of the mass density ρ and the divergence of the mass flow rate ρv, where the *divergence* of a vector field $f = [f_1, f_2, f_3]^\top$ in Cartesian coordinates is defined as $\mathrm{div}(f) \triangleq \partial f_1/\partial x_1 + \partial f_2/\partial x_2 + \partial f_3/\partial x_3$. In order to simplify the analysis, fluid motion is often considered in one dimension reducing equation (1) to

$$\frac{\partial \rho}{\partial t} = -\frac{\partial u}{\partial x}, \tag{2}$$

where $u = \rho v$ is the mass flow rate of the fluid. Many real flows are essentially one-dimensional, and variations in parameters across streamlines can be ignored; or by averaging properties of the flow over an appropriate region, it can be analyzed in one dimension [1]. In general, however, there are situations for which the one-dimensional assumption leads to highly erroneous results [1].

Now, consider fluid motion in a region (cell) of length ℓ as shown in Fig. 1 with inflow rate of u_{in} into the cell and the outflow rate of u_{out}. A discretized version of (2), in both time and space,

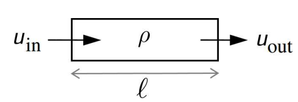


Figure 1: One-dimensional fluid motion in a region (cell) of length ℓ and internal (average) mass density ρ with inflow rate u_{in} and outflow rate u_{out}.

is given by

$$\rho^{k+1} = \rho^k + \frac{T_s}{\ell}\left(u_{in}^k - u_{out}^k\right), \tag{3}$$

where T_s is the sampling time period, ρ^k is the mass density of the fluid at time $t = kT_s$, and u^k_{in} and u^k_{out} are, respectively, the mass inflow and outflow rate into and from the cell at time step $t = kT_s$.

A widely-used approach for fluid flow control in a transport network is to partition the network into several segments, each of which is represented by a cell as shown in Fig. 1. Then, the following assumptions are made:

(i) The fluid dynamics in every cell is described by (3), that is, for cell i of length ℓ_i, mass density ρ^k_i, inflow rate y^k_i, and outflow rate u^k_i, we have

$$\rho^{k+1}_i = \rho^k_i + \frac{T_s}{\ell_i}(y^k_i - u^k_i). \tag{4}$$

(ii) The mass density in every cell ρ^k_i can be measured at each time step k.

(iii) The outflow rate from each cell can be controlled through a regulation mechanism. This can be done by placing an *active* network element (e.g. a control valve or a compressor) at interfaces between consecutive cells.

It should be noted that the inflow rate y^k_i to cell i is a known function of the outflow rates from the immediately upstream cells. If all immediate upstream cells of cell i are merged only into cell i, then y^k_i is equal to the sum of all flow rates leaving the upstream cells; otherwise, the inflow to cell i is determined according to *flow split ratios* of the network which are known *a priori*. Hence, if the outflow rate from every cell is known over a fixed period of time, then from (4), the state of the system (densities) is completely known over that period.

Fig. 2 shows a fluid transmission network with a line structure partitioned into n cells of possibly different length, where the cells are increasingly numbered from upstream to downstream. The outflow rate u_i from cell i can be controlled through a flow regulation mechanism.

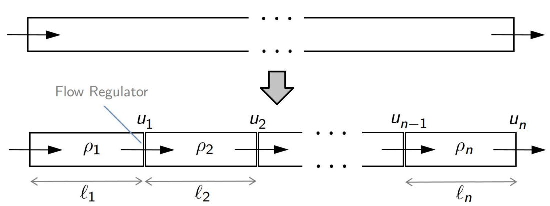


Figure 2: Partitioning a transmission network with a line structure into n cells. A flow regulator at the interface between any two consecutive cells controls the outflow rate u_i from each cell i. For this linear network, the inflow rate to cell i is $y^k_i = u^k_{i-1}$.

The control objective is to find time series of the outflow rates and the corresponding mass densities such as to optimize an integral performance index over a finite period of time, subject to dynamical and physical constraints of the network. In general, the optimization problem can be expressed as

$$\min_{u} \left[\Phi(\rho^N) + \sum_{k=0}^{N-1} L^k(\rho^k, u^k) \right], \quad \text{s.t. } (\rho, u) \in \Omega \tag{5}$$

where $\rho = [\rho^{0\top}, \ldots, \rho^{N\top}]^\top$, $u = [u^{0\top}, \ldots, u^{(N-1)\top}]^\top$, $\rho^k = [\rho^k_1, \ldots, \rho^k_n]^\top$, $u^k = [u^k_1, \ldots, u^k_n]^\top$, n is the number of cells, N is the final time step, Φ is the terminal cost functional, L is the running cost functional, and $\rightarrow\!|$ is the set of admissible state/control variables satisfying (4) and meeting supply-demand constraints. The complexity of an optimization problem depends mainly on the function forms of its objective function and constraint set. For the sake of tractability, we focus on *linear objective functions*. There are many meaningful cost functions of practical interest which can be expressed in a linear form [2–4].

The above framework has been widely used to formulate optimal flow control problems for complex transmission networks [5–9]. Of particular interest to this study is the work on highway traffic networks.

Traffic flow in highway transportation networks is often regulated by ramp metering and/or variable speed limit under the *Cell Transmission Model* (CTM) dynamics. The CTM is a simple macroscopic traffic model capturing most phenomena observed on highways including flow conservation, non-negativity, and congestion wave propagation [5,6]. Because of its analytical simplicity, the CTM is widely used for control design purposes, wherein a one-way road is partitioned into multiple cells as shown in Fig. 2, and the traffic flow in each cell is viewed as a homogeneous stream of vehicles with a dynamic described by (4). In this problem, the flow regulation is carried out by reducing the outflow rate from the cells, that is the flow regulation mechanism acts as a *control valve*.

Since the size and complexity of transportation networks are growing, design and implementation of an *efficient* control scheme providing an optimum operation has become more challenging and demanding. The existing results on finite-time optimal control of transport networks are mainly restricted to schemes with an open-loop feedforward control structure which are not robust in most actual applications. It is well known that the use feedback helps reducing the effects of modeling uncertainties and improving performance, especially when a simplified plant model is used to make the control design and analysis tractable.

One approach for optimal flow control is the *Model Predictive Control* (MPC) which is a model-based feedback control technique relying on real-time optimization [3, 4, 10–12]. Although the closed-loop operation of the MPC provides a certain degree of robustness with respect to modeling uncertainties, the primary challenge of implementing MPC in real-time is its computational complexity. The framework of *multi-parametric linear/quadratic programming* has been proposed to reduce on-line computation effort in MPC, and effective numerical algorithms have be developed to solve multi-parametric programming [13–15]. However, determination of the optimal control action at each time step involves centralized operations making its implementation costly or impractical for large-scale networks.

It is, therefore, desired to design an optimal, or at least suboptimal, feedback control law with a simple structure that requires access only to local information. Decentralized optimal control problems are often substantially more complex than the corresponding problems with centralized information. A trivial centralized optimal decision-making problem may become NP-hard under a decentralized information structure [16]. This is why most research has been focused on the design of meaningful suboptimal decentralized control policies and identification of tractable subclasses of problems [17,18]. Since no principled methodology exists for design and performance evaluation of decentralized optimal controllers, the problem is often attacked by applying suitable approximations and/or relaxations.

This work is an attempt to deal with decentralized feedback control design for some classes of flow networks. A new decentralization method is proposed for feedback flow control, which is based on the following logic: (i) Construct a centralized optimal state-feedback control scheme with respect to a global performance index generating the control input of the entire network at each sample time, given the state vector of the entire network. The resulting controller, in theory,

provides the ideal performance. In practice, however, such a controller may not be implementable.
(ii) Design a local version of the centralized optimal feedback control scheme for each portion of the network minimizing a local cost function. The performance metric associated with each local controller is a local version of the global (centralized) performance index, wherein only local state variables (specified by a given information structure) are used to generate the input command to the respective actuator. Due to the lack of analytical tools, performance evaluation of the decentralized scheme and comparison with centralized optimal control are done through numerical simulations.

3.2 Information Design for Traffic Networks

Route choice decision in traffic networks under uncertain and dynamic environments, such as the ones induced by recurring unpredictable incidents, can be a daunting task for agents. Private route recommendation or public information systems could therefore play an important role in such settings. While the agents have prior about the uncertain state, e.g., through experience or publicly available historic records, the informational advantage of such systems in knowing the actual realization gives the possibility of inducing a range of traffic flows through appropriate route recommendation or public information strategies.

A strategy of a recommendation system to map state realization to randomized private route recommendations for the agents is referred to as a *private signal* ; a strategy to map state realization to randomized public messages is referred to as a *public signal*. A private signal is feasible or *obedient*, if, to every agent, it recommends a route which is weakly better in expectation than the other routes. Under a public signal, the agents can be assumed to choose routes consistent with Bayes Nash flow with respect to the posterior. The problem of minimizing expected social latency cost over all obedient private signals or over all public signals is referred to as *information design* in this study. We are interested in these problems for *non-atomic* agents, when a fraction of agents do not participate in signaling and induce Bayes Nash flow with respect to the prior. The technical challenge is the joint consideration of optimal signal for receiving agents and the flow induced by non-receiving agents. The non-atomic setting is chosen in part to be consistent with the feedback control part of the project which uses macroscopic models.

Information design for *finite* agents has attracted considerable attention recently with applications in multiple domains, e.g., see [19] for an overview; the single agent case was studied in [20] as *Bayesian persuasion*. In the finite agent (and finite action) setting, the obedience condition on the signal can be expressed as finite linear constraints, one for each combination of actions by the agents. This allows to cast the information design problem as a tractable optimization problem. Techniques to further reduce computational cost of information design are presented in [21]. However, analogous computational approaches to solve information design for non-atomic agents, particularly for routing games, are lacking.

There has been a growing interest recently in understanding the impact of information in non-atomic routing games. For example, [22] demonstrates informational Braess paradox in which revealing information about all the links does not necessarily minimize social cost; [23, 24] illustrate that properly designed information structure could reduce price of anarchy; [25] demonstrates that information design only for a fraction of agents, while taking into account externality from flow induced by the rest, might be beneficial for social cost. Information *design* using private signals, as in this study, has also been pursued recently in [26]. Optimal public signals for some settings were characterized in [27]. While these existing works provide useful insights, the information *design* aspect of these works is restricted to stylized settings involving a network with just two parallel links, sub-optimal signals, and link latency functions which ensure non-zero flow on all links under all state realizations. It is not apparent to what extent can the methodologies underlying these studies, which typically rely on analytical solutions, be generalized. On the other hand, we develop a computational approach in this study, with focus on parallel networks for illustration. The approach however extends to general network setting.

Our key observation is that information design for polynomial link latency functions has strong connections with the *generalized problem of moments* (GPM) [28]. A GPM minimizes, over finite probability measures, a cost which is linear in moments with respect to these measures subject to constraints which are also linear in the moments. This connection allows to leverage computational tools developed for GPM, such as `GloptiPoly` [29], which utilizes a hierarchy of semidefinite relaxations to lower bound GPM arbitrarily closely by relaxation of sufficiently high order, at the expense of increasing computational cost. For a fixed flow induced by non-receiving agents, information design for receiving agents is indeed a GPM. Furthermore, since the cost and constraints involve moments up to a finite order, there exists an optimal signal which is atomic with provable upper bound on the number of atoms [30]. This is utilized to equivalently cast information design, when the non-receiving agents choose route according to Bayes Nash flow, as a polynomial optimization problem. This in turn can be arbitrarily approximated by known hierarchy of semidefinite relaxations [31], which can also be implemented in `GloptiPoly`. The first level of this hierarchy is shown to be exact for the basic two link case with affine latency functions, and it relies on using convexity of the cost function and the constraints to sharpen the bound from [30] for optimal solution.

The obedient constraints for the receiving agents in the information design setup of this study are reminiscent of characterization of (Bayes) correlated equilibrium. It is therefore natural to compare our approach with semidefinite programming based approaches for computation of correlated equilibria, e.g., in *continuous* polynomial games [32]. In [32], the action set is continuous and the agents are finite, and hence alternate formulations for correlated equilibrium are proposed which involve approximation through finite moments and discretization of the action set. On the other hand, in our setup, where the action set is finite and the agents are non-atomic, the constraints for the receiving agents are readily in computational form and involve moments up to a finite order without any approximation. This then allows us to consider an *equivalent* finite discretization, with known cardinality, of the agent population, to transform equivalently into a polynomial optimization problem. Thereafter, the use of semidefinite relaxation hierarchy for solution is standard.

The computational approach of this study can be utilized to complement the current studies on (paradoxical) effect of different fractions of receiving agents under specific public signals (primarily, full information). While existing work, e.g., [33, 34], studies the effect on population-specific (i.e., receiving and non-receiving) costs, we study the effect on the social cost, in the spirit of the social planner's perspective adopted in the study. We provide a class of private signals under which the optimal social cost is non-increasing with increasing fraction of receiving agents. The key idea is to use an optimal solution at a given fraction to synthesize signals which are feasible for all higher fractions and give the same cost. This monotonic result does not require the link latency functions to be polynomial. We also provide numerical examples to suggest that one should not however expect such a monotonic behavior in general for a public signal, even if it is optimal.

In summary, the main contributions of the information design part of the study are as follows. First, by making connection to GPM and associated semidefinite programming machinery, we point to a compelling computational framework to solve information design problems. Second, by establishing the existence of an atomic optimal solution, we provide credence to such a structural assumption often implicitly made in information design studies. The sharpening of the bound on the number of atoms that we illustrate in a simple case suggests the possibility of using the problem structure of information design to reduce the size of the optimization formulation, and hence the computation cost. Third, the result and underlying proof technique for the monotonic behavior of social cost under a reasonable class of private signals could be utilized to design private signals that guarantee performance which is robust to higher than anticipated agent participation rate. However, our results also suggest that this may be difficult to achieve through public signals. Overall, the contributions allow to considerably expand the scope of information design studies for non-atomic routing games, which has been limited so far to stylized settings.

The results for feedback control of traffic networks are presented in Sections 4-7, and the results for information design are presented in Sections 8-11. Due to space limitation, several technical details are not included here but can be found in extended versions [35, 36].

4 Setup for the Study of Feedback Control of Traffic Networks

Throughout the study of feedback control of flow networks, the set of integers $\{1, 2, \ldots, n\}$ is denoted by $\mathbb{N}_n$, and $\{(a_i)_{i \in \mathbb{N}_n}\} = \{a_1, a_2, \ldots, a_n\}$. A *convex polyhedron* is the intersection of finitely many half-spaces, i.e., $\{x \in \mathbb{R}^n \,|\, Ax \leq b\}$, for a matrix $A \in \mathbb{R}^{m \times n}$ and a vector $b \in \mathbb{R}^m$. A real-valued function $f(x)$ on $D \subseteq \mathbb{R}^n$ is said to be *increasing* (*decreasing*) if it is increasing (decreasing) in *every* coordinate.

Theorem 1. *[37] Consider the following multi-parametric linear program*

$$\begin{aligned} J^{*}(\theta) &= \min_{z} c^{\top} z \\ s.t.\ & Wz \leq G + S\theta, \quad \theta \in \Theta \subseteq \mathbb{R}^m, \end{aligned} \tag{6}$$

where $z \in \mathbb{R}^n$ is the decision variables vector and $\theta \in \mathbb{R}^m$ is a parameter vector, Θ is a closed polyhedral set, and c, W, G, S are constant matrices. Let Θ^{*} denote the region of parameters θ such that (6) is feasible. Then, there exists an optimizer $z^{*}(\theta) : \Theta^{*} \to \mathbb{R}^n$ which is a continuous and piecewise affine function of θ, that is

$$\begin{aligned} z^{*} &= pwa(\theta) \\ &= L_i\theta + l_i, \quad \text{if } \theta \in R_i, \quad i \in \mathbb{N}_p, \end{aligned} \tag{7}$$

where sets $R_i = \{\theta \in \Theta^{*} \,|\, \Pi_i\theta \leq \eta_i\}$ form a polyhedral partition of Θ^{*}, p is the number of polyhedral sets, L_i, l_i, Π_i, η_i are constant matrices, and $pwa(\cdot)$ is a generic symbol for piecewise affine functions on polyhedral sets. Moreover, the value function $J^{*}(\theta) : \Theta^{*} \to \mathbb{R}$ is a continuous, convex, and piecewise affine function of θ.

The Matlab-based Multi-Parametric Toolbox [14] together with YALMIP Toolbox [38] can be used to solve multi-parametric linear programs and compute the matrices L_i, l_i, Π_i, η_i in (7).

Consider the optimal control design problem (4)-(5). The objective is to design an optimal control with feedback architecture to benefit from the feedback properties such that the resulting control law is *suitable* for practical implementation. By 'suitable', we mean a controller meeting limitations in communication and computational power.

Fig. 3 shows a general network with a number of inflow/outflow rates, where λ_i^k and μ_i^k denote the i-th inflow and outflow at time k, respectively. The external inflow rates to the network act as exogenous inputs which cannot be manipulated by the controller. The controller can regulate only the outflow rate of each cell by monitoring the states of the network cells.

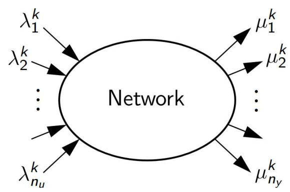


Figure 3: A general controlled network with n_u exogenous inflow rates λ_i^k and n_y outflow rates μ_i^k. The controller task is to regulate the outflow rate from each cell of the network to optimize a performance index.

Since the objective function and constraints in (5) depend on inflow rates to the network, then, in general, a complete knowledge of inflow rate signals over the control horizon is required to solve the optimization problem (5). The assumption that the external inflow rate over the control horizon is known *a priori* is, however, very restrictive in practice. The exogenous input to the network may not be known or predictable in all scenarios. When no knowledge on the inflow rate is available, a control law must be designed such that the feasibility of the solution (control/state variables) at any time for *any* admissible λ_i^k is guaranteed.

Disregarding communication and computational limitations, finding a globally optimal solution to (5) with a *feedback realization* is a difficult task in general. Hence, some assumptions and simplifications need to be made to make the control design tractable and its implementation feasible. It is desired to implement the solution to (4)-(5) in the form of a *static* state-feedback control as

$$u^k = \;{}^{k}(\rho^k), \tag{8}$$

where $u^k = [u_1^k, \ldots, u_n^k]^\top$ and $\rho^k = [\rho_1^k, \ldots, \rho_n^k]^\top$ are the vectors of cells' outflow rates and mass densities, respectively. A realization of the form (8) is possible when the performance index and constraints satisfy certain properties, or they are simplified through proper approximations to satisfy certain properties.

Remark 1. *The main reason for considering "static feedback" is the simplicity of control law. In a static state-feedback controller, the control action at each time k depends only on the current state vector at time k. One may consider a "dynamic feedback" controller, wherein the control action depends on the state variables in the current and previous sampling instants; this, however, makes design, analysis, and implementation of the controller more difficult.*

Although design of a centralized feedback optimal control (if it exists) provides the ideal performance, it may not be implementable for large-size networks, as it may require a significant computational resource and a fast and highly-reliable commutation system. It is, therefore, necessary to further simplify the control law to meet communication/computational constraints.

We present our results in a general setting which can then be specialized to highway traffic networks, e.g., CTM can be formulated as (4)-(5). We are interested in a finite-horizon decentralized control law with the following features:

- It consists of independent local controllers that have access only to information about their local neighborhood (i.e., local state variables and local network's parameters and architecture) providing a performance level (with respect to a certain performance index) sufficiently close to that of the optimal centralized controller.

- Each local controller can be implemented in a static state-feedback form, feeding back local state variables to generate the local control action. This requires that the constrained optimization problem associated with each local controller to be such that its optimal solution can be expressed in a feedback form.

We first study the centralized control design under certain assumptions such that the control law is optimal (w.r.t a performance index) and implementable in a state-feedback form. Subsequently, decentralization of the resulting centralized control scheme is investigated by considering a simple information structure. Communication constraints are often modeled by a fixed *information structure*; for example, in the network shown in Fig. 2, if only the mass density of cells i and $i+1$ are available to generate the outflow from every cell i, a desired decentralized realization of the i-th controller is

$$u_i^k = \mu_i^k(\rho_i^k, \rho_{i+1}^k), \tag{9}$$

expressing the current required outflow rate from cell i in terms of the current state of local cells i and $i+1$.

We argue that when the cost and constraint satisfy some separability condition, a local version of problem (4)-(5) can be constructed for each portion of the network. The i-th local controller (generating the outflow rate from cell i) has access only to the state of cells in a pre-specified neighborhood of cell i determined by a given information structure, then by expressing the solution to each local problem in a feedback form, a state-feedback decentralized control law is designed. It should be highlighted that for design of a local controller no information about the parameters, structure, and state of the rest of the network is used; only the feedback architecture of the control law can indirectly provide information about the status of the rest of the network. In other words, *feedback* is essential to keep a local controller from being completely blind about the rest of the network.

To further illustrate the proposed decentralization, let us consider the network shown in Fig. 2, and assume that only knowledge about cells i and $i+1$ are available to generate u_i. To design the i-th controller, we consider the sub-network consisting of only cells i and $i+1$, as shown in Fig. 4 and solve the centralized problem associated with the two-cell network. In the i-th local optimization problem, the decision variables are u_i^k, u_{i+1}^k, $k = 0, \ldots, N-1$, with zero inflow rate to cell i, but only u_i^k is used and implemented and the optimal values of u_{i+1}^k are unused. For this example, the i-th local optimization problem may be expressed as

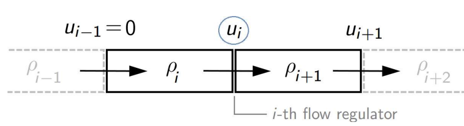


Figure 4: To design the i-th controller generating outflow rate u_i^k, a centralized optimal feedback control law for the sub-network consisting of only cells i and $i+1$ is designed. No knowledge about the rest of the network is available to the i-th controller.

$$\min_{u_i, u_{i+1}} \left\{ \Phi_i(\rho_i^N, \rho_{i+1}^N) + \sum_{k=0}^{N-1} L_i^k(\rho_i^k, \rho_{i+1}^k, u_i^k, u_{i+1}^k) \right\}$$
$$\text{s.t. } (u_i, u_{i+1}, \rho_i, \rho_{i+1}) \in \hat{\Omega}_i \tag{10}$$

where ℓ_i is the terminal cost functional, $\hat{g}_i$ is the running cost functional, and $\hat{\mathcal{U}}_i$ is the outflow constraint set associated with the i-th controller. The set $\hat{\mathcal{U}}_i$ is obtained by relaxing any constraint involving non-local variables and assuming zero inflow to the local network. From the solutions to (10), only u_i^k is kept for implementation and the remaining variables are discarded. As mentioned before, we would like to implement the solution to (10) in the form of a static state-feedback of local states as (9). The feedback realization of the solution is crucial as the values of $\rightarrow_i$ and $\rightarrow_{i+1}$ are affected by the action of the other controllers in the network.

The proposed decentralization scheme relies on the following properties:

- Existence of a global optimizer for the centralized problem with a state-feedback realization.
- Separability of the centralized problem such that for each sub-network a local optimization problem can be constructed, for which a global optimizer can be found in a feedback form.

The above points are clarified in the following sections.

5 Sub-Optimal Decentralized Feedback Control

The objective is to design a decentralized static feedback controller with a specific information structure e.g. a (unidirectional) one-hop information structure in a line network as $u_i^t = \kappa_i^t(x_i^t, x_{i+1}^t)$.

5.1 Approach 1: Truncation

Ignore any term in cost and constraint functions that depends on non-local variables, then solve the truncated optimization. As an example consider a 4-cell network with the following global optimization problem: The global cost function is given by

$$J = \sum_{t=0}^{N-1} \sum_{i=1}^{4} (\alpha_i^t x_i^t + \beta_i^t u_i^t), \tag{11}$$

subject to

$$\begin{aligned} x_i^{t+1} &= x_i^t + u_{i-1}^t - u_i^t, \quad i = 1, 2, 3, 4, \\ 0 \le u_i^t &\le d_i(x_i^t), \quad i = 1, 2, 3, 4, \\ u_i^t &\le s_{i+1}(x_{i+1}^t), \quad i = 1, 2, 3. \end{aligned} \tag{12}$$

Assume the objective is to express the second outflow rate u_2^t as a function of local states x_2^t and x_3^t. Let x_2^t, u_2^t, x_3^t, u_3^t be local variables for the local controller generating u_2^t as shown in Fig. 5. Then, the corresponding local optimization may be constructed by truncating the global optimization as

$$J_2 = \sum_{t=0}^{N-1} (\alpha_2^t x_2^t + \beta_2^t u_2^t) + (\alpha_3^t x_3^t + \beta_3^t u_3^t), \tag{13}$$

subject to

$$\begin{aligned} x_i^{t+1} &= x_i^t + u_{i-1}^t - u_i^t, \quad i = 2, 3, \\ 0 \le u_i^t &\le d_i(x_i^t), \quad i = 2, 3, \\ u_i^t &\le s_{i+1}(x_{i+1}^t), \quad i = 2. \end{aligned} \tag{14}$$

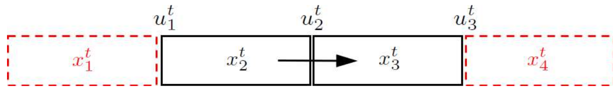


Figure 5: The local network associated with the second actuator local optimization problem generating u_2^t. Only the local states x_2^t, x_3^t are available to generate u_2^t.

Setting the inflow to the local network zero, i.e., $u_1^t = 0, \forall t$, makes the truncated problem a function of only the variables $x_2^t, u_2^t, x_3^t, u_3^t$. The solution to the truncated problem (13), (14) can be expressed as $u_2^t = (x_2^t, x_3^t)$. Clearly, this decentralized approach gives the centralized performance if every controller has access to the entire network state as no truncation will take place. We will show that this simple approach with a one-hop structure can give the centralized performance for some networks of any size.

5.2 Approach 2: Design of Local Optimization

Find a suitable parametrization to express each local optimization as a function of only local variables such that the minimization of local costs improves the global performance index. Consider the global optimization problem (13), (14). Assuming that $x_2^t, u_2^t, x_3^t, u_3^t$ are local variables for the local controller generating u_2^t, and given a priori knowledge on the entire network structure and parameters, determine the form/parameters of functionals f_2^t, g_2^t, h_2^t (independent of state variables), and formulate the local optimization problem as

$$\begin{aligned} J_2 &= \sum_{t=0}^{N-1} f_2^t(x_2^t, u_2^t, x_3^t, u_3^t) \\ g_2^t&(x_2^t, u_2^t, x_3^t, u_3^t) \leq 0 \\ h_2^t&(x_2^t, u_2^t, x_3^t, u_3^t) = 0 \end{aligned} \tag{15}$$

such that:

- The problem of determining 'suitable' functionals f_2^t, g_2^t, h_2^t is tractable (off-line computations).
- The solution to the local optimization is guaranteed to be feasible w.r.t. the original global problem.
- The local optimization problem is tractable for online implementation and its feasible global optimum can be found using the standard numerically efficient techniques.
- The decentralized approach gives the centralized performance if every controller has access to the entire network state.
- The local controllers implicitly collaborate with each other to improve the global performance.

6 Application to Traffic Networks

In recent years, due to the ever-increasing traffic demand, efficient control and management of transportation networks has received a great deal of attention. There has been a lot of research done on the optimal control of freeway networks based on various models for traffic systems, among which first-order models, such as the CTM, are widely used for control design. In a CTM-based traffic model, the network dynamics is described by (4), where ρ_i^k [veh/mi] is the traffic density, y_i^k [veh/hr] is the inflow rate, u_i^k [veh/hr] is the outflow rate, and ℓ_i [mi] is the length of cell i. The constraints are defined in terms of demand and supply functions, where the demand function $\bar{d}_i(\cdot)$ returns the maximum outflow from the cell as a function of its current traffic density, and the supply function $\bar{s}_i(\cdot)$ gives the maximum inflow into the cell as a function of its current traffic density [2]. The demand and supply functions are assumed to be of the form

$$\begin{aligned} \bar{d}_i(\rho_i) &= \min\{d_i(\rho_i), C_i\}, \\ \bar{s}_i(\rho_i) &= \min\{s_i(\rho_i), C_i\}, \end{aligned} \tag{16}$$

where d_i is continuous non-decreasing function of ρ_i with $d_i(0) = 0$ and s_i is continuous non-increasing function of ρ_i with $s_i(0) > 0$, and C_i [veh/hr] is maximum flow capacity of cell i. The *jam traffic density* of cell i is defined as $r_i = \inf\{\rho_i > 0 \,|\, s_i(\rho_i) = 0\}$. The functions $d_i(\cdot)$ and $s_i(\cdot)$ are often assumed to be affine of the form $d_i(\rho_i) = v_i\rho_i$ and $s_i(\rho_i) = w_i(r_i - \rho_i)$, where v_i [mi/hr] is the maximum traveling free-flow speed and w_i [mi/hr] is the backward congestion wave traveling speed of cell i. Then, in a *controlled network* via speed limit control, the feasible region for outflow and inflow rates are defined as [2]:

$$\begin{aligned} 0 &\leq u_i^k \leq \min\{v_i\rho_i^k, C_i\}, \\ 0 &\leq y_i^k \leq \min\{w_i(r_i - \rho_i^k), C_i\}. \end{aligned} \tag{17}$$

The flow regulation mechanism in a traffic network acts as a collection of *control valves*, each of which at each time can be open to the fullest extent possible, completely closed, or partially closed during the network operation.

Assumption 1. *The length of cells* ℓ_i *and the time interval* T_s *are chosen such that vehicles traveling at maximum speed* v_i *can not cross multiple cells in one time step, i.e.,* $v_i T_s \leq \ell_i$, $\forall i$. *Also, the backward congestion wave traveling speed* w_i *satisfies* $w_i T_s \leq \ell_i$, $\forall i$.

Assumption 1 is known as *Courant-Friedrichs-Lèvy* condition [2] which is a necessary condition for numerical stability in numerical computations. It can be easily verified that Assumption 1 together with constraints (17) ensure that at each time the density of each cell is non-negative and never exceeds the jam density.

A flow network can be represented by a directed graph, in which edges represent cells and vertices (or junctions) represent interface between consecutive cells which are the actuators' location. The junctions can be of either of the three types defined below.

Definition 1. *[2] A junction with a single incoming and a single outgoing cell is called* ordinary*; a junction with a single incoming cell and multiple outgoing cells is called* diverge*; and a junction with multiple incoming cells and a single outgoing cell is called* merge.

The following definitions and notations are used throughout this section.

Definition 2. *Consider a network whose topology is described by directed graph* G. *The set of edges of* G *corresponding to on-ramps is called the* source set *denoted by* E_{on}*, and the set of edges corresponding to off-ramps is called the* sink set *denoted by* E_{off}.

At any diverge junction, the traffic flow is distributed according to a given split percentage which are estimated from historical data [39].

Definition 3. *[2] The* split ratio *(or* turning ratio*)* $R_{ij} \in [0, 1]$ *is defined as the fraction of flow leaving cell* $i \notin E_{off}$ *that is directed towards cell* $j \neq i$*, where* $\sum_j R_{ij} = 1$*. If cells* i *and* j *are not adjacent or* $i = j$*,* R_{ij} *is defined to be zero.*

Definition 4. *Let cell* i *be an incoming cell to junction* $\sim_i$*, where* $\sim_i$ *denotes the* head *or the* downstream junction *of cell* i*. The set of all outgoing cells from junction* $\sim_i$ *is called the* out-neighborhood *of cell* i *and is denoted by* E_i^+*. If* $i \in E_{off}$*, then* E_i^+ *is the empty set. In other words,* E_i^+ *is the set of all direct successor of cell* i*. The elements of* E_i^+ *are referred to as the* out-neighbors *of cell* i*.*

Definition 5. *Let cell* i *be an outgoing cell from junction* $□_i$*, where* $□_i$ *denotes the* tail *or the* upstream junction *of cell* i*. The set of all incoming cells to junction* $□_i$ *is called the* in-neighborhood *of cell* i *and is denoted by* E_i^-*. If* $i \in E_{on}$*, then* E_i^- *is the empty set. In other words,* E_i^- *is the set of all direct predecessor of cell* i*. The elements of* E_i^- *are referred to as the* in-neighbors *of cell* i*.*

An example is shown in Fig. 6 clarifying the above definitions.

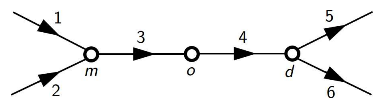


Figure 6: Directed graph of a six-cell network with source set (on-ramps) $E_{on} = \{1, 2\}$ and sink set (off-ramps) $E_{off} = \{5, 6\}$. The merge, ordinary, and diverge junctions are labeled m, o, and d, respectively. The split ratios of $R_{45} = 0.3$ and $R_{46} = 0.7$ imply that 30% of vehicles in cell 4 turn towards cell 5 and 70% of them turn toward cell 6. The in-neighborhood and out-neighborhood of cell 4 are $E_4^- = \{3\}$ and $E_4^+ = \{5, 6\}$, respectively.

It is often more convenient to express the dynamics and constraints in terms of the *traffic mass* of the cells. Let $x_i^k = \ell_i \rho_i^k$ [veh] denote the traffic mass of cell i at time k, then from (4) and (17), the dynamics and constraints of an n-cell network can be written as

$$x_i^{k+1} = x_i^k + T_s(y_i^k - u_i^k), \quad \forall i \in N_n \tag{18a}$$

$$y_i^k = >_i^k + \sum_{j=1}^n R_{ji} u_j^k, \tag{18b}$$

$$0 \le u_i^k \le \min\{(v_i/\ell_i)x_i^k, C_i\}, \tag{18c}$$

$$0 \le y_i^k \le \min\{w_i(r_i - (1/\ell_i)x_i^k), C_i\}, \tag{18d}$$

where $>_i^k$ is an exogenous inflow rate to cell $i \in E_{on}$ ($>_i^k = 0$, if $i \notin E_{on}$), and R_{ij}'s are split ratios. Then, for any $i \in E_{on}$, $y_i^k = >_i^k$, and for any $i \notin E_{on}$, $y_i^k = \sum_{j=1}^n R_{ji} u_j^k$.

Remark 2. *To ensure that* $>_i^k$ *is a feasible exogenous input to the network, it is typically assumed that the jam traffic density of any on-ramp is infinity,* $r_i = 1$*, and* $>_i^k \le C_i$*,* $\forall i \in E_{on}$*.*

Control Objective: Consider the network dynamics (18) and let $x^k = [x_1^k, \dots, x_n^k]^\top$ be the state vector and $u^k = [u_1^k, \dots, u_n^k]^\top$ be the control input vector of the network at time k. The control objective is to design a static feedback control law such that for *any* initial state x^0 and *any* exogenous inflow $>^k$, the feasibility of control actions is guaranteed and a performance index of the form (5), subject to (18) and a given information structure, over a fixed given control horizon [0,N] is optimized. In this study, we focus on linear objective functions, i.e. (5) with

$$\bullet(x^N) = \sum_{i=1}^n \alpha_i^N x_i^N,$$

$$^k(x^k, u^k) = \sum_{i=1}^n \left(\alpha_i^k x_i^k + {}^k_i u_i^k \right), \tag{19}$$

where N is a fixed final time, and $\alpha_i^k > 0$ and β_i^k are cost-weighting parameters.

Remark 3. *There are meaningful performance indexes which can be expressed in a linear form [2–4]; for example:*

(i) Minimization of the total travel time *of the network is equivalent to minimization of the total number of vehicles in the entire network, then the corresponding cost is* $J = \sum_{k=0}^{N} \sum_{i=1}^{n} x_i^k$.

(ii) Maximization of the total travel distance *is equivalent to maximization of the flows, then the following cost should be minimized* $J = -\sum_{k=0}^{N-1} \sum_{i=1}^{n} u_i^k$.

(iii) The total congestion delay *is defined as the time difference between actual travel time and the travel time in free-flow conditions whose minimization is equivalent to minimizing* $J = \sum_{k=0}^{N-1} \sum_{i=1}^{n} (x_i^k - (\ell_i/v_i) u_i^k)$.

For the centralized control, there is no information constraint and the control law is of the form $u^k = \mu^k(x^k)$. For the decentralized control, we consider controller with a *one-hope information structure* as defined below.

Definition 6. *A feedback controller is said to have a* uni-directional one-hop information structure*, if* u_i^k *depends only on* x_i^k *and the state of the cell(s) immediately downstream of cell* i*, i.e. those either entering or leaving the downstream junction of cell* i*. Similarly, in a* bi-directional one-hop information structure*,* u_i^k *depends only on* x_i^k *and the state of the cells immediately upstream a well as those of immediately downstream of cell* i.

Clearly, Definition 6 can be extended to uni/bi-directional p-hop structure. Throughout this study, however, we focus on uni-directional one-hop structure and call it "one-hop information structure", for short. As an example, for the network in Fig. 6, a decentralized static feedback controller with a one-hop information structure is of the form: $u_1^k = \mu_1^k(x_1^k, x_2^k, x_3^k)$, $u_2^k = \mu_2^k(x_2^k, x_1^k, x_3^k)$, $u_3^k = \mu_3^k(x_3^k, x_4^k)$, $u_4^k = \mu_4^k(x_4^k, x_5^k, x_6^k)$, $u_5^k = \mu_5^k(x_5^k)$, and $u_6^k = \mu_6^k(x_6^k)$.

6.1 Centralized Feedback Control

The external inflow rates δ_i^k, $i \in E_{on}$, to the network act as exogenous inputs which cannot be manipulated by the controller. In general, the solution to the optimization problem (5), (18), (19) depends on the values of δ_i^k; however, no *a priori* knowledge on δ_i^k is often available for control design. Analogous to the classical LQR problem where no uncontrolled exogenous input is considered for optimal control design [40], we design a centralized optimal controller under the assumption of $\delta_i^k = 0, \forall i \in E_{on}, k \in [0, N-1]$; and we refer to the resulting controller as *zero-inflow optimal control law*. Then, we show that in the presence of any non-zero inflow, the feasibility of the optimal solution is guaranteed.

Let us first suppose that the sequence of $\delta_i^k, \forall i \in E_{on}$ over the entire control horizon, is known beforehand; under this assumption the following theorem gives the *true optimal control law*.

Theorem 2. *The solution to (5), (18), (19) can be expressed in the form of a continuous piecewise affine static feedback law on polyhedra of the state vector as*

$$\begin{aligned}(u^k)^* &= pwa^k(x^k) \\ &= F_i^k x^k + f_i^k, \ \textit{if}\ x^k \in R_i^k,\end{aligned} \tag{20}$$

where $R_i^k = \{x \in \mathbb{R}^n \mid H_i^k x \le h_i^k\}$, $i \in \mathbb{N}_{p^k}$, *is the* i*-th polyhedral partition of the set of feasible states, and* p^k *is the number of polyhedral partitions at time* k*. The controller parameters* $F_i^k, f_i^k, H_i^k, h_i^k$ *can be computed offline; they are independent of* x^k, $\forall k$, *but may depend on the values of* λ_i^k.

Corollary 1. *Consider the optimization problem* (5), (18), (19)*. The* zero-inflow optimal control law *can be expressed as* (20)*, where matrices* $F_i^k, f_i^k, H_i^k, h_i^k$ *can be computed off-line. Moreover, the feasibility of the resulting control actions is guaranteed for any non-zero inflow rate* λ_i^k*, i.e., the constraints in* (18) *are always satisfied.*

A true optimal controller is not implementable as it needs unknown inflow rates over the control horizon; and a zero-inflow optimal controller may not be truly optimal. However, with respect to certain cost functions, the zero-inflow optimal feedback control law is truly optimal. We show that for problem (5), (18), (19), under certain assumptions on the network topology, if the cost functions satisfy certain properties, the zero-inflow centralized optimal feedback controller has a decentralized realization with a one-hop information structure which is truly optimal for any exogenous inflow.

Theorem 3. *Consider the problem* (5), (18), (19) *for a network with time-invariant split ratios and* no merge junction. *In addition, assume that cost-weighting parameters satisfy* $\alpha_i^k > \alpha_j^k > 0$, $\forall k, i$ *and* $\forall j \in E_i^+$, *and* $\beta_i^k \le \beta_i^{k+1} \le 0$, $\forall k, i$. *Then, the true optimal feedback control law (with centralized information) can be realized as*

$$
\begin{aligned}
(u_i^k)^\star &= pwa_i^k(x_i^k, (x_j^k)_{j\in E_i^+}) \\
&= \min\left\{\frac{v_i}{\ell_i}x_i^k, C_i, \left\{\frac{w_j}{R_{ij}}\left(r_j - \frac{1}{\ell_j}x_j^k\right), \frac{C_j}{R_{ij}}\right\}_{j\in E_i^+}\right\}.
\end{aligned}
\tag{21}
$$

The controller (21) has a one-hop information structure (see Definition 6); moreover, its parameters are obtain at *no computational cost* independent of the control horizon N. Indeed, the expression in the right-hand side of (21) is the upper limit of u_i^k which is known beforehand, that is $0 \le u_i^k \le \bar{u}_i^k$, where

$$
\bar{u}_i^k = \min\left\{\frac{v_i}{\ell_i}x_i^k, C_i, \left\{\frac{w_j}{R_{ij}}\left(r_j - \frac{1}{\ell_j}x_j^k\right), \frac{C_j}{R_{ij}}\right\}_{j\in E_i^+}\right\}. \tag{22}
$$

Hence, Theorem 3 states that, under the given assumptions, setting each outflow rate equal to its upper limit provides the true optimal performance. This is equivalent to opening every control valve to the fullest extent possible at each time. We refer to such scheme as *trivial control* or *uncontrolled scheme*.

Remark 4. *The conditions given in Theorem 3 are sufficient (not necessary) for a linear performance index with respect to which a centralized optimal control law has a realization with a specific one-hop information structure. It should be also noted that the optimal control is not necessarily unique.*

Remark 5. *The widely-used performance indexes in Remark 3 satisfy the properties given in Theorem 3.*

In general, however, the optimal controller needs access to the state of the entire network and depends on the control horizon. For a general network with a general linear cost functional, the closed-form of the control law (20) enables one to compute the controller parameters offline and stored in computer memory before the control actions are ever applied to the network. That is, there is no need to solve a large-size optimization problem at every time step for real-time implementation, unless there is a large variation in the network parameters. An optimal feedback controller of the form (20), however, suffers from two major drawbacks restricting its applicability to large-scale networks: (i) Even though the piecewise-affine form of the control law seems to be simple, when the number of cells and the control horizon increase, solving the corresponding multi-parametric linear programs may result in a very large number of polyhedral partitions, making the structure of the controller too complex. Although applying the merging algorithms [14, 15] may considerably reduce the number of polyhedral partitions, in general there may still be too many polyhedral sets. (ii) Determining the optimal control action at each time involves centralized operations, that is each local controller needs instantaneous access to the state of the entire network; this, however, may not be feasible for large-size networks, as implementation of a highly reliable and fast communication system may be impractical or too costly. It is, therefore, necessary to design an feedback control law with a simple structure that requires access only to local information, while providing a satisfactory performance level.

Remark 6. *Throughout the study, it is assumed that no information about the exogenous inflow to the network is available for control design (both centralized and decentralized), that is any controller is designed under the assumption of zero external inflow rate. The controller, however, can be applied to the network with any admissible inflow rates. For nonzero inflow rate, the feasibility of the control actions is guaranteed, but they are not necessarily truly optimal.*

6.2 Decentralized Feedback Control

In this subsection, the objective is to design a static state-feedback control law with a one-hop information structure for problem (5), (18), (19). Such a control law, for a general network, is of the form

$$u_i^k = {}_i^k(x_i^k, (x_j^k)_{j \in D_i}), \tag{23}$$

where D_i denotes the set of all cells, excluding cell i, leaving/entering the downstream junction (head) of cell i. From the definition of D_i, it follows that $E^+ \subseteq D_i$; also, for any $i \in E_{off}$, $D_i = \{\}$. For example, in Fig. 6, $D_1 = \{2, 3\}$ and $D_4 = \{5, 6\}$.

Design of a decentralized feedback controller can be viewed as solving an uncertain optimization problem, wherein non-local variables/parameters are unknown. The main challenges are how to ensure the feasibility of the solution and how to express or implement it in a feedback form.

Uncertain linear program has been the subject of a lot of research and several approaches have been proposed to deal with robust optimization problems [41] including: solving the problem for nominal values of the unknown parameters and then performing sensitivity analysis; formulating the problem as a stochastic optimization by incorporating the knowledge on the probability distribution of the uncertain parameters; and assigning a finite set of possible values to the uncertain parameters and determining a solution which is relatively good for all the scenarios [42]. Also, some research has focused on evaluating the impact of uncertainty on the cost by computing the worst and best optimum solutions [43]. In some other works, in order to ensure the feasibility of solution, a worst-case approach is considered which, in general, leads to extremely conservative solutions [41].

In this study, we follow the decentralized procedure proposed in Section 5 which lead to a simple decentralized control law with the desired information structure and provides a feasible solution that under certain conditions could provide the optimal centralized performance.

In order to design the i-th control law with a one-hope information structure, we design a centralized optimal static state-feedback controller for the sub-network consisting of cells i and any

$j \in D_i$, with zero inflow rate to cell i. Then, the i-th local optimization is

$$\min_{u_i,(u_j)_{j\in D_i}} \Big\{ \hat{\phi}_i(x_i^N,(x_j^N)_{j\in D_i}) + \sum_{k=0}^{N-1} \hat{\ell}_i^k(x_i^k,(x_j^k)_{j\in D_i},u_i^k,(u_j^k)_{j\in D_i}) \Big\} \quad \text{s.t. } (x_i,(x_j)_{j\in D_i},u_i,(u_j)_{j\in D_i}) \in \hat{\Omega}_i, \tag{24}$$

where

$$\hat{\phi}_i = \alpha_i^N x_i^N + \sum_{j\in D_i} \alpha_j^N x_j^N, \qquad \hat{\ell}_i^k = \alpha_i^k x_i^k + \beta_i^k u_i^k + \sum_{j\in D_i}(\alpha_j^k x_j^k + \beta_j^k u_j^k). \tag{25}$$

and the constraint set $\hat{\Omega}_i$ is defined by (18) with zero inflow rate $y_i^k = 0, \forall k$, wherein any constraint involving non-local variables is relaxed.

Theorem 4. *Consider the local optimization (24), for a sub-network of cells $i, j \in D_i$. The solution can be expressed as*

$$(\hat{u}_i^k)^* = pwa_i^k(x_i^k,(x_j^k)_{j\in D_i}). \tag{26}$$

which is a piecewise affine function on polyhedra of local state variables whose parameters can be computed off-line. Moreover, it satisfies all constraints in (18) for any k, i.

We refer to (26) as a "sub-optimal decentralized control law with a one-hope information structure". It should be highlighted that the separability property of the objective function and constraints has enabled us to simply construct a local version of the centralized optimization problem as (24).

The natural question that arises is how to evaluate the performance and sub-optimality level of the above decentralized control scheme. As mentioned earlier, in general, performance analysis of decentralized controllers is a very difficult task. Due to the lack of analytical tools, performance evaluation can be done through extensive numerical simulations. It should be noted that although the above decentralization procedure involves constraints/relaxations that may affect the conservativeness of the solution, under certain conditions, performance degradation due to decentralization is zero. It is easy to verify that if the conditions in Theorem 3 are satisfied, solving the local optimization (24) gives the true optimal controller.

Example 1. *In order to illustrate the decentralization process, let us consider a 3-cell network with*

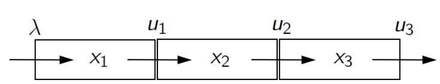


Figure 7: A 3-cell network with exogenous inflow rate λ^k.

the following cost function and constraints

$$\begin{aligned} \min\ & J = \sum_{k=0}^{N} (x_1^k + 4x_2^k + 2x_3^k) \\ \text{s.t. } & x_i^{k+1} = x_i^k + u_{i-1}^k - u_i^k, \quad u_0^k = \lambda^k, \\ & 0 \le u_i^k \le 0.9x_i^k, \quad i = 1,2,3, \\ & u_i^k \le 1 - 0.3x_{i+1}^k, \quad i = 1,2. \end{aligned} \tag{27}$$

Table 1 shows how the centralized controller can be implemented in a feedback form. In a decentralized control with a one-hop information structure, given local state at time t, the control action at each time $t = 0, 1, \ldots, N-1$, is obtained as shown in Table 2.

Table 1: Centralized control in feedback form.

Global optimization at time t	Control action at time t
Given x^t: $\min J^t = \sum_{k=t}^{N}(x_1^k + 4x_2^k + 2x_3^k)$ $x_1^{k+1} = x_1^k + 0 - u_1^k$, $x_2^{k+1} = x_2^k + u_1^k - u_2^k$, $x_3^{k+1} = x_3^k + u_2^k - u_3^k$, $0 \leq u_1^k \leq 0.9x_1^k$, $u_1^k \leq 1 - 0.3x_2^k$, $0 \leq u_2^k \leq 0.9x_2^k$, $u_2^k \leq 1 - 0.3x_3^k$, $0 \leq u_3^k \leq 0.9x_3^k$,	$u^t = {}^t(x^t)$, Discard u^k, $\forall k > t$.

Table 2: A decentralized control in feedback form.

Local optimization at time t	Control action at time t
Given x_1^t, x_2^t: $\min J_1^t = \sum_{k=t}^{N}(x_1^k + 4x_2^k)$ $x_1^{k+1} = x_1^k + 0 - u_1^k$, $x_2^{k+1} = x_2^k + u_1^k - u_2^k$, $0 \leq u_1^k \leq 0.9x_1^k$, $u_1^k \leq 1 - 0.3x_2^k$, $0 \leq u_2^k \leq 0.9x_2^k$,	$u_1^t = {}^t_1(x_1^t, x_2^t)$, Discard u_1^k, $\forall k > t$.
Given x_2^t, x_3^t: $\min J_2^t = \sum_{k=t}^{N}(4x_2^k + 2x_3^k)$ $x_2^{k+1} = x_2^k + 0 - u_2^k$, $x_3^{k+1} = x_3^k + u_2^k - u_3^k$, $0 \leq u_2^k \leq 0.9x_2^k$, $u_2^k \leq 1 - 0.3x_3^k$, $0 \leq u_3^k \leq 0.9x_3^k$,	$u_2^t = {}^t_2(x_2^t, x_3^t)$, Discard u_2^k, $\forall k > t$.
Given x_3^t: $\min J_3^t = \sum_{k=t}^{N} 2x_3^k$ $x_3^{k+1} = x_3^k + 0 - u_3^k$, $0 \leq u_3^k \leq 0.9x_3^k$,	$u_3^t = {}^t_3(x_3^t)$, Discard u_3^k, $\forall k > t$.

In a decentralized feedback control with a one-hop information structure (see Definition 6), by setting $y_i^k = 0$, $\forall k$, and removing any constraints involving state of non-local (not immediately down-stream) cells, the global optimization problem is decomposed into multiple lower-dimensional local problems. For any i, the i-th local optimization problem is such that, given local states at current time t, the feasibly of u_i^t is guaranteed (see Theorem 4).

7 Simulations: Feedback Control of Traffic Networks

As mentioned earlier, it is generally difficult to analytically evaluate performance of a decentralized control scheme in compared with that of an optimal centralized controller, hence comprehensive numerical simulations must be performed to demonstrate the effectiveness of a decentralization technique and numerically assess the level of sub-optimality.

Simulation 1 : Consider the 8-cell cyclic traffic network shown in Fig. 8, with cost function $J = \sum_{k=0}^{N} \sum_{i=1}^{8} \alpha_i x_i^k$, subject to (18), where $\alpha_i = 1$, for $i = 1, 2, 3, 5, 7, 8$, $\alpha_4 = 5$, and $\alpha_6 = 3$. Turning ratios at diverge junctions are $R_{38} = R_{34} = 0.5$ and $R_{56} = R_{57} = 0.5$. The external inflow

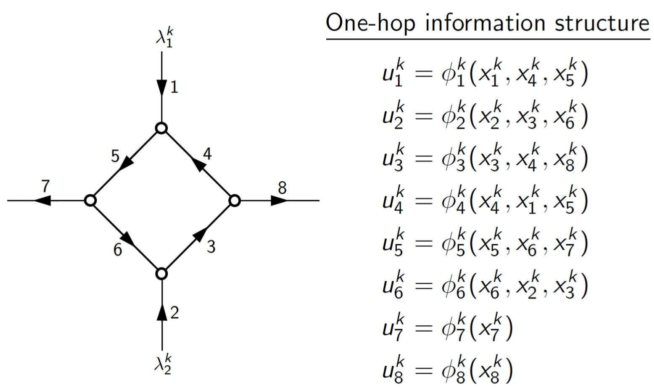


Figure 8: An 8-cell cyclic network with exogenous inflow rates λ_1^k and λ_2^k, where $E_{on} = \{1, 2\}$ and $E_{off} = \{7, 8\}$.

rate to the network are $\lambda_1^k = \lambda_2^k = 1, \forall k$. The other parameters are $v_i = 0.9$, $w_i = 0.3$, $C_i = 10$, $r_i = 10$, $\ell_i = 1$, $x_i(0) = 0$, $\forall i$, and $T_s = 1$. Let J_{cen}^* and J_{dec}^* be respectively the cost value of the centralized control and decentralized control with a one-hop information structure. Fig. 9 shows the relative percentage of the performance loss due to decentralization $\varepsilon = 100(J_{dec}^* - J_{cen}^*)/J_{cen}^*$ for different values of control horizon $N = 1, 2, \ldots, 30$.

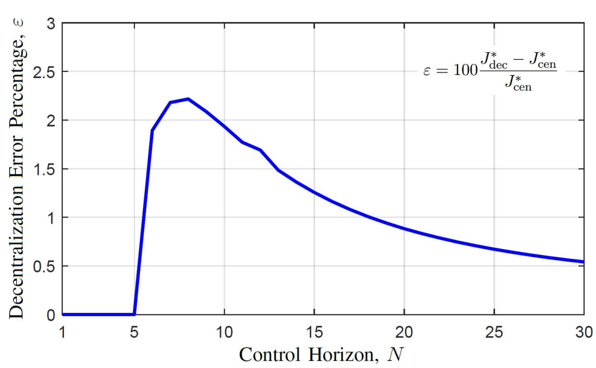


Figure 9: Relative performance loss due to decentralization as a function of control horizon. The maximum relative error is 2.22% at $N = 8$.

Simulation 2 : To evaluate the performance of the decentralized scheme, let us consider a larger network with more realistic architecture and parameters. We consider the freeway system of an area in the southern Los Angeles as shown in Fig. 10(a) modeled by the CTM. The directed graph of the network of the region of interest consisting of 32 cells is shown in Fig. 10(b). Consider

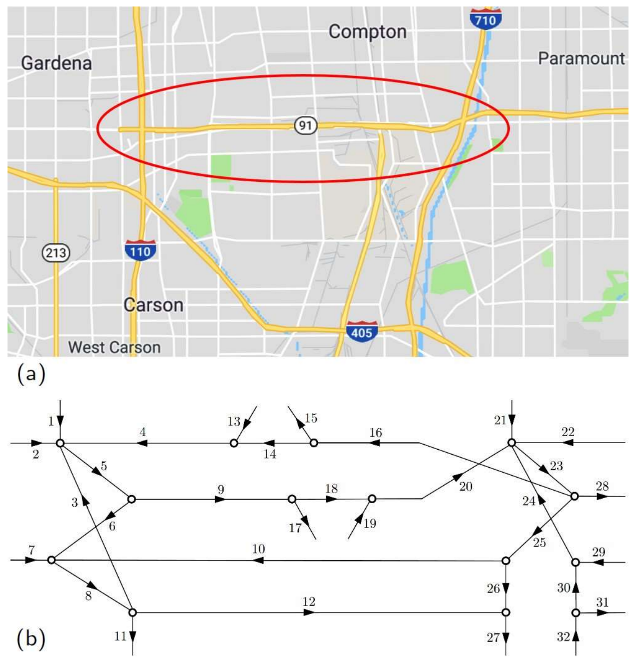


Figure 10: (a) The map of an area in the southern Los Angeles. The red ellipse shows the region used in our numerical simulation. (b) The directed graph of the transportation network of the region of interest with 32 cells, where $E_{on} = \{1, 2, 7, 13, 19, 21, 22, 29, 32\}$ and $E_{off} = \{11, 15, 17, 27, 28, 31\}$.

minimization of $J = \sum_{k=0}^{N} \sum_{i=1}^{n} x_i^k$, subject to (18), with the following parameters. The sampling time is $T_s = 1/360$ hr (or 10 sec). For on-ramp cells, the jam traffic density ρ_i is assumed to be infinity and for other cells is $\rho_i = 200$ veh/mi. For all cells, the backward congestion wave traveling speed is $w_i = 13$ mi/hr. For cells $3, 4, 9, 10, 12, 16, 20$, the cell's length is $\ell_i = 2$ mi, the free-flow speed is $v_i = 65$ mi/hr, and the maximum flow capacity is $C_i = 800$ veh/hr, and for other cells, $\ell_i = 0.5$ mi, $v_i = 25$ mi/hr, and $C_i = 400$ veh/hr. At any diverge junction, σ_i with incoming cell i, the turning ratios are time-invariant and are split uniformly between the outgoing cells, i.e., $R_{ij} = 1/n_{\sigma_i}$, where n_{σ_i} is the number of outgoing cells from junction σ_i; for example, $R_{8,3} = R_{8,11} = R_{8,12} = 1/3$. The exogenous inflow rate to on-ramp cells are $\lambda_i^k = 1, \forall k, i \in \{1, 2, 7, 13, 19, 21, 22, 29, 32\}$, and $x_i^0 = 0$, $\forall i$. Fig. 11 shows the relative percentage of the performance loss due to decentralization for different values of control horizon $N = 1, 2, \ldots, 10, 20, \ldots, 100$. For example, for $N = 60$, the optimal cost value of the centralized controller is $J^\star_{cen} = 33.7367$ and that of the decentralized one (with one-hop information structure) is $J^\star_{dec} = 33.7383$, then the relative decentralization performance loss $\varepsilon = 100(J^\star_{dec} - J^\star_{cen})/J^\star_{cen} = 0.0047\%$.

8 Setup for Information Design Study

We state a few key notations to be used for this study. $E_\nu[x]$ will denote the expected value of random variable x with respect to probability distribution ν. $\mathrm{int}(X)$ will denote the interior of set X and $\Delta(X)$ the set of all probability distributions on X. For an integer n, we let $[n] := \{1, 2, \ldots, n\}$.

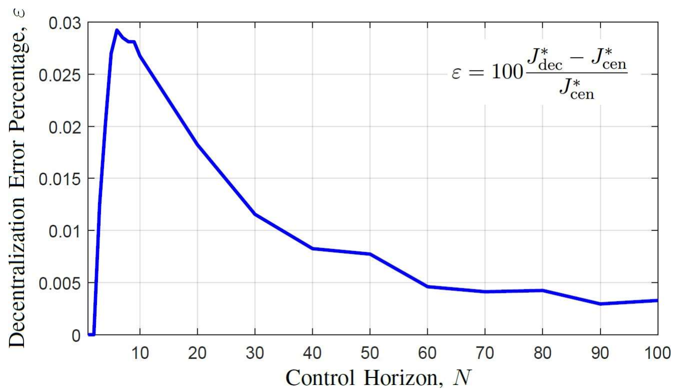


Figure 11: Relative performance loss due to decentralization as a function of control horizon. The maximum relative error is 0.0292% at $N = 6$.

For a vector $x \in \mathbb{R}^n$, let $\text{supp}(x) := \{i \in [n] \mid x_i \neq 0\}$ be the set of indices whose corresponding entries in x are not zero. For $\tau > 0$, let $\mathcal{P}_n(\tau) := \left\{x \in \mathbb{R}^n_{\geq 0} \mid \sum_{i \in [n]} x_i = \tau\right\}$ be the $(n-1)$-dimensional probability simplex of size τ. When $\tau = 1$, we shall simply denote the simplex as $\mathcal{P}_n$ for brevity in notation. $\mathbf{0}_{n \times m}$ and $\mathbf{1}_{n \times m}$ will denote $n \times m$ matrices all of whose entries are 0 and 1 respectively. In all these notations, the subscripts corresponding to size shall be omitted when clear from the context. For a matrix A, its transpose is denoted as A^T. For matrices A and B of the same size, their inner product is $A \cdot B = \sum_{i,j} A_{i,j} B_{i,j}$. $A \succeq 0$ for a symmetric matrix A will imply that it is positive semidefinite.

Consider a network consisting of n parallel links between a single source-destination pair. Without loss of generality, let the agent population generate a unit volume of traffic demand. The link latency functions $\ell_{\omega,i}(f_i)$, $i \in [n]$, give latency on link i as a function of flow f_i through them, conditional on the *state* of the network $\omega \in \Omega = \{\omega_1, \ldots, \omega_s\}$. Throughout the study, we shall make the following basic assumption on these functions.

Assumption 2. *For every $i \in [n]$, $\omega \in [s]$, $\ell_{\omega,i}$ is a non-negative, continuously differentiable and non-decreasing function.*

At times, we shall strengthen the assumption to $\ell_{\omega,i}$ being strictly increasing. A class of functions satisfying Assumption 2 which is attractive from a computational perspective is that of polynomial functions:

$$\ell_{\omega,i}(f_i) = \sum_{d=0}^{D} \alpha_{d,\omega,i} f_i^d, \qquad i \in [n],\ \omega \in [s] \tag{28}$$

with $\alpha_{0,\omega,i} > 0$ and $\alpha_{1,\omega,i} > 0$. We shall also let α_d refer to the $s \times n$ matrix whose entries are $\alpha_{d,\omega,i}$. Two instances of (28) commonly studied in the literature are affine and the Bureau of Public Roads (BPR) functions [44]. In the former case, $D = 1$ and in the latter case, $D = 4$ with $\alpha_1 = \alpha_2 = \alpha_3 = \mathbf{0}$. Furthermore, the BPR function has the following interpretation: $\alpha_{0,\omega,i}$ is the free flow time on link i when the state is ω, and $\left(0.15\,\frac{\alpha_{0,\omega,i}}{\alpha_{4,\omega}}\right)^{\frac{1}{4}}$ is the flow capacity of link i when the state is ω.

Let $\omega \sim \mu_0 \in \text{int}(\Delta(\Omega))$, for some prior μ_0 which is known to all the agents. The agents do not have access to the realization of ω, but a fixed fraction $\nu \in [0, 1]$ of the agents receives private route recommendations conditional on the realized state. These conditional recommendations are generated by a *signal* $\pi = \{\pi_\omega \in \Delta(P_n(\rho)) : \omega \in \Omega\}$ as follows. Given a realization $\omega \in \Omega$, sample a $x \in P_n(\rho)$ according to π_ω, and partition the agent population into $n+1$ parts with volumes $(x_1, \ldots, x_n, 1-\rho)$. All the agents are identical, and therefore in the non-atomic setting that we are considering here the partition can be formed by independently assigning every agent to a partition with probability equal to the volume of that partition. The agents in the $(n+1)$-th partition, with volume $1-\rho$, do not receive any recommendation, whereas all the agents in the i-th partition, $i \in [n]$, receive recommendation to take path i.

The signal π and the fraction ρ is publicly known to all the agents. Therefore, it is easy to see that the (joint) posterior on (x, ω), i.e., the proportion of agents getting different recommendations and the state of the network, formed by an agent who receives recommendation $i \in [n]$ is:

$$\mu^{\pi,i}(x,\omega) = \frac{x_i \, \pi_\omega(x) \, \mu_0(\omega)}{\sum_{\omega' \in \Omega} \int_{p \in P(\rho)} p_i \, \pi_{\omega'}(p) \, dp \, \mu_0(\omega')} \tag{29}$$

and the posterior formed by an agent who does not receive a recommendation is:

$$\mu^{\pi,\emptyset}(x,\omega) = \pi_\omega(x)\mu_0(\omega) \tag{30}$$

Remark 7. *One could consider an alternate setup where the set of agents who do not participate in the signaling scheme is pre-determined. These agents do not receive a recommendation and also do not have knowledge about π. In this case, (30) can be replaced with $\mu^{\emptyset}(x,\omega) = \frac{\mu_0(\omega)}{|P(\rho)|}$ obtained by replacing π_ω with the uniform distribution. The methodologies developed in this study also extend to this alternate setting.*

A signal is said to *obedient* if the recommendation received by every agent is weakly better, in expectation with respect to posterior in (29), than other routes, while the non-receiving agents induce a Bayes Nash flow with respect to their posterior in (30). Formally, a π is said to be obedient if there exists $y \in P_n(1-\rho)$ such that[1]:

$$\sum_\omega \int_x \ell_{\omega,i}(x_i+y_i)\mu^{\pi,i}(x,\omega)\,dx \le \sum_\omega \int_x \ell_{\omega,j}(x_j+y_j)\mu^{\pi,i}(x,\omega)\,dx, \quad i,j \in [n] \tag{31a}$$

$$\sum_\omega \int_x \ell_{\omega,i}(x_i+y_i)\mu^{\pi,\emptyset}(x,\omega)\,dx \le \sum_\omega \int_x \ell_{\omega,j}(x_j+y_j)\mu^{\pi,\emptyset}(x,\omega)\,dx, \quad i \in \mathrm{supp}(y),\ j \in [n] \tag{31b}$$

Plugging the expressions of beliefs from (29) and (30), noting that the denominators on both sides of the inequalities are the same in (31), and multiplying both sides of (31b) by y_i, one equivalently gets:

$$\sum_\omega \int_x \ell_{\omega,i}(x_i+y_i)\, x_i\, \pi_\omega(x)dx\, \mu_0(\omega) \le \sum_\omega \int_x \ell_{\omega,j}(x_j+y_j)\, x_i\, \pi_\omega(x)dx\, \mu_0(\omega), \quad i,j \in [n] \tag{32a}$$

$$\sum_\omega \int_x \ell_{\omega,i}(x_i+y_i)\, y_i\, \pi_\omega(x)dx\, \mu_0(\omega) \le \sum_\omega \int_x \ell_{\omega,j}(x_j+y_j)\, y_i\, \pi_\omega(x)dx\, \mu_0(\omega), \quad i,j \in [n] \tag{32b}$$

We emphasize that multiplying both sides by y_i allows to equivalently relax the restriction on i in terms of y in (31b) to get (32b).

[1] Throughout the study, unless noted otherwise, the summation over indices for degree, state and link, such as d, ω and i, respectively, are to be taken over the entire range, i.e., $\{0, \ldots, D\}$, $[s]$ and $[n]$, respectively.

The social cost is taken to be the expected total latency:

$$J(\pi, y) := \sum_{\omega, i} \int (x_i + y_i)\, \ell_{\omega,i}(x_i + y_i)\, \pi_\omega(x) dx\, \mu_0(\omega) \tag{33}$$

The information design problem can then be stated as

$$\min_{(\pi,y) \in \Pi \times P(1-\alpha)} J(\pi, y) \text{ s.t. } (32) \tag{34}$$

where Π is the concise notation for $\Delta(P(\alpha))^s$.

Remark 8. *(i) If there are multiple feasible* y *for a given* π*, then a solution* (π^*, y^*) *to* (34) *can be interpreted as implicitly requiring an additional action from the social planner to enforce* y^**. One could alternately consider a* robust *formulation by replacing* $\min_{(\pi,y)}$ *in* (34) *with* $\min_\pi \max_y$*. We leave such an extension for future consideration. Moreover, Lemma 1 below shows that, under a rather reasonable condition on the link latency functions, there exists a unique feasible* y *for every* π*, in which case the robust version is the same as* (34).

(ii) The revelation principle, e.g., see [19], implies that optimality in the class of obedient direct *private signals, i.e., signals which recommend routes, also ensures optimality within a broader class which includes* indirect *signals. An indirect signal provides noisy information about the state realization. The route choice is then determined by Bayes Nash flow with respect to the posterior beliefs induced by the signal. In Section 10, we consider a special case of indirect signals, known as* public *signals.*

(iii) The feasible set in (34) *is non-empty for all* $\alpha \in [0, 1]$*. A formal argument is postponed to Remark 13, after we have discussed public signals.*

Lemma 1. *For every* $\pi \in \Pi$*, a* $y \in P(1-\alpha)$ *satisfies* (32b) *if and only if it solves the following convex problem:*

$$\min_{y \in P(1-\alpha)} \sum_{\omega, i} \int_0^{y_i} \int_{x \in P(\alpha)} \ell_{\omega,i}(x_i + s)\, \pi_\omega(x) dx\, ds\, \mu_0(\omega) \tag{35}$$

Moreover, such a y *is unique if* $\{\ell_{\omega,i}\}_{\omega,i}$ *are strictly increasing over* $[0, 1]$.

Lemma 1 follows from a straightforward adaptation of the standard argument for Wardrop equilibrium in the deterministic case.

Remark 9. *Lemma 1 implies that, in order to ensure a unique feasible* y *for every* π*, it is sufficient to have* $\alpha_{1,\omega,i} > 0$ *for all* ω, i *for affine latency functions, and* $\alpha_{4,\omega,i} > 0$ *for all* ω, i *for BPR latency functions.*[2]

Following Lemma 1, minimizing $J(\pi, y)$ with respect to y for a fixed π is trivial for strictly increasing link latency functions. Joint optimization over π and y in (34) however is challenging, not the least because it involves optimizing over probability distributions. The next section presents finite dimensional formulations which are provably equivalent to (34) for polynomial link latency functions.

[2]Note that all the derivatives of the BPR latency function are zero at 0. However, one can easily show uniqueness in the special cases when, for a signal supported only on $x_i = 0$, (35) has a solution with $y_i = 0$.

9 Private Signals

In this section, unless stated otherwise, we assume that the link latency functions are polynomial, i.e., of the form in (28). Let us first consider minimizing $J(\pi, y)$ over π satisfying (32a), for a fixed y. Note that, for $y = \mathbf{0}$, this corresponds to the information design problem in the special case when $\lambda = 1$. Even in this special case, which has been studied previously in [24, 26], no comprehensive solution methodology exists.

We start by rewriting the information design problem in terms of moments of the signal π. Let z be the vector of all monomials in $x_1, \ldots, x_n$ up to degree $\lceil d/2 \rceil$, arranged in a lexicographical order. For example, for $D = 3$, $z = [1, x_1, \ldots, x_n, x_1^2, \ldots, x_1 x_n, x_2 x_1, \ldots, x_2 x_n, \ldots, x_n x_1, \ldots, x_n^2]^T$. For a fixed y, (34) can then be written as:

$$\min_{\pi \in \Pi} \sum_{\omega} \int C_{\omega}(y) \cdot zz^T \, \pi_{\omega}(x) dx \tag{36a}$$

$$\text{s.t.} \quad \sum_{\omega} \int A^{(i,j)}_{\omega}(y) \cdot zz^T \, \pi_{\omega}(x) dx \geq 0, \quad i, j \in [n] \tag{36b}$$

$$\sum_{\omega} \int B^{(i,j)}_{\omega}(y) \cdot zz^T \, \pi_{\omega}(x) dx \geq 0, \quad i, j \in [n] \tag{36c}$$

for appropriate symmetric matrices C_{ω}, $A^{(i,j)}_{\omega}$, and $B^{(i,j)}_{\omega}$; expressions for these matrices in the special case when $D = 1$ (i.e., affine link latency functions) are provided in [36]. The cost in (36a) is the same as the cost in (33), (36b) corresponds to the obedience constraint in (32a), and (36c) corresponds to (32b).

(36) is an instance of the *generalized problem of moments* (GPM) [28], which in turn can be solved numerically using `GloptiPoly` [29]. This software solves GPM by lower bounding it with semidefinite relaxations of increasing order. The stopping criterion on the order is however problem-dependent; approximations can be obtained by a user-specified order. In the special of $n = 2$, the first order relaxation is tight.

Proposition 1. *Let $n = 2$. For every $y \in P(1 - \lambda)$, (36) is equivalent to a semidefinite program.*

Remark 10. *Proposition 1 implies that, in the case of two links, when all the agents are receiving, i.e., $\lambda = 1$, computing optimal signal is tractable for arbitrary polynomial latency functions. This is to be contrasted with existing work, e.g., [24, 26], where an optimal signal is provided for such a setting only for certain affine link latency functions.*

Proposition 1 and the discussion before it suggests a natural alternating heuristic for solving (34): start with an arbitrary $y \in P(1 - \lambda)$, and alternate between solving (36) for a fixed y and finding a feasible y using Lemma 1. Under appropriate conditions on the latency functions, one can show that this heuristic results in a sequence of feasible (π, y) whose associated cost is monotonically decreasing, and hence convergent, though not necessarily to a global optimum of (34).

9.1 An Exact Polynomial Optimization Formulation via Atomic Signals

A natural approach to approximate the joint optimization in (34) is to discretize the support of π. A signal π is called m-*atomic*, $m \in \mathbb{N}$, if, for every $\omega \in \Omega$, π_{ω} is supported on m discrete points $x^{(k)} \in P(\lambda)$, $k \in [m]$. Let the set of such signals be denoted as $\Pi(m)$. It is easy to see that every signal in $\Pi(m)$ can be represented as a $s \times m$ row stochastic matrix. To emphasize the matrix notation, we let $\pi(k|\omega)$ denote the probability of recommending routes according to $x^{(k)}$ when the state realization is ω. Computing optimal signal in $\Pi(m)$ can be written as the following polynomial optimization problem[3]:

$$\min_{\substack{x^{(k)} \in \mathcal{P}(\Omega),\, k \in [m] \\ y \in \mathcal{P}(1-\Omega) \\ \pi \in \Pi(m)}} \sum_{k,\,\omega,\,i} \left(x_i^{(k)} + y_i \right) \ell_{\omega,i}(x_i^{(k)} + y_i)\, \pi(k|\omega)\, \mu_0(\omega) \tag{37a}$$

$$\text{s.t.} \quad \sum_{k,\,\omega} \ell_{\omega,i}(x_i^{(k)} + y_i)\, x_i^{(k)}\, \pi(k|\omega)\, \mu_0(\omega) \leq \sum_{k,\,\omega} \ell_{\omega,j}(x_j^{(k)} + y_j)\, x_i^{(k)}\, \pi(k|\omega)\, \mu_0(\omega), \quad i,j \in [n] \tag{37b}$$

$$\sum_{k,\,\omega} \ell_{\omega,i}(x_i^{(k)} + y_i)\, y_i\, \pi(k|\omega)\, \mu_0(\omega) \leq \sum_{k,\,\omega} \ell_{\omega,j}(x_j^{(k)} + y_j)\, y_i\, \pi(k|\omega)\, \mu_0(\omega), \quad i,j \in [n] \tag{37c}$$

In particular, for (28) with $D = 1$, i.e., affine link latency functions, the polynomials in the cost functions and the constraints are of degree 3. (37) can also be solved (approximately) using `GloptiPoly`. (37) gives an increasingly tighter upper bound to (34) with increasing $m \in \mathbb{N}$. While it is natural to expect the gap between (37) and (34) to go to zero as $m \to +\infty$, the gap in fact becomes zero for finite m.

Theorem 5. (34) *is equivalent to* (37) *for* $m > s\binom{D+n}{D+1}$

The upper bound in Theorem 5 on the number of atoms required to realize an optimal signal can be tightened in some cases, as we show in the next section.

9.2 Diagonal Atomic Signals

An atomic signal which has attracted particular attention is when π is the identity matrix of size s. We shall refer to such a signal as a *diagonal atomic signal*, and denote its finite support as x^{ω}, $\omega \in \Omega$. The polynomial optimization problem in (37) in this case simplifies to:

$$\min_{\substack{x^{\omega} \in \mathcal{P}(\Omega),\, \omega \in \Omega \\ y \in \mathcal{P}(1-\Omega)}} \sum_{\omega,\, i} (x_i^{\omega} + y_i)\, \ell_{\omega,i}(x_i^{\omega} + y_i)\, \mu_0(\omega) \tag{38a}$$

$$\text{s.t.} \quad \sum_{\omega} \ell_{\omega,i}(x_i^{\omega} + y_i)\, x_i^{\omega}\, \mu_0(\omega) \leq \sum_{\omega} \ell_{\omega,j}(x_j^{\omega} + y_j)\, x_i^{\omega}\, \mu_0(\omega), \quad i,j \in [n] \tag{38b}$$

$$\sum_{\omega} \ell_{\omega,i}(x_i^{\omega} + y_i)\, y_i\, \mu_0(\omega) \leq \sum_{\omega} \ell_{\omega,j}(x_j^{\omega} + y_j)\, y_i\, \mu_0(\omega), \quad i,j \in [n] \tag{38c}$$

In general, (38) gives an upper bound to (37) for $m > s$, and hence also for (34). The next result establishes the equivalence between the two formulations in a special case, and also establishes that

[3]Throughout the study, unless noted otherwise, the summation over index for discrete support, such as k, is to be taken over the entire range, i.e., m.

(38) is equivalent to the following semidefinite program:

$$\min_{M \succeq 0} \hat{J}(M) := C \cdot M \tag{39a}$$

$$\text{s.t.} \quad A^{(i,j)} \cdot M > 0, \quad i, j \in [n] \tag{39b}$$

$$B^{(i,j)} \cdot M > 0, \quad i, j \in [n] \tag{39c}$$

$$M(1,1) = 1 \tag{39d}$$

$$M(i,j) > 0, \quad i, j \in [(s+1)n+1] \tag{39e}$$

$$S_x^{(k)} \cdot M = 0, \quad S_y \cdot M = 0, \quad k \in [m] \tag{39f}$$

$$T_x^{(i,k)} \cdot M = 0, \quad T_y^{(i)} \cdot M = 0 \quad i \in [n], k \in [m] \tag{39g}$$

where the expressions for symmetric matrices C, $A^{(i,j)}$, $B^{(i,j)}$, $S_x^{(k)}$, S_y, $T_x^{(i,k)}$ and $T_y^{(i)}$ for the special case $D = 1$ are provided in [36].

Proposition 2. *If $n = 2$, then (38), (34) and (39) are all equivalent to each other for (28) with $D = 1$, i.e., for affine link latency functions.*

Remark 11. *(i) For $n = 2$ and $D = 1$, Proposition 2 implies that an optimal signal can be realized with s atoms, which is much less than the bound $\binom{3}{2} s = 3s$ given by Theorem 5.*

(ii) For $n = 2$ and $D = 1$, if $M^* = \begin{bmatrix} 1 & \xi^{*T} \\ \xi^* & M^{0,*} \end{bmatrix}$ is an optimal solution to (39), then $[x_1^{*1}, x_2^{*1}, \ldots, x_1^{*s}, x_2^{*s}, y_1, y_2]^T = \xi^*$ is an optimal solution for (38), and hence also for (34).

(iii) Proposition 2 and its proof approach (cf. [36]) might appear to be generalization of an observation in [26], which was made for $\lambda = 1$, and for a class of affine link latency functions. Not only do we remove these restrictions, but more importantly, our proof implicitly highlights that the obedience constraint needs more careful treatment than suggested in [26].

(iv) It is informative to contrast the different approaches of Proposition 1 and Proposition 2 for establishing tightness of the natural semidefinite relaxation of the corresponding variants of the information design problem. Proposition 5 simply relies on the ability to rewrite the problem in terms of univariate probability measures with compact support. On the other hand, Proposition 2 relies on the tightness of the GPM obtained by relaxation of the problem because it has optimal probability measures supported on single atoms.

9.3 Monotonicity of Optimal Cost Value under Diagonal Atomic Signals

Let $J^{\text{diag}}(x, y)$ denote the cost function in (38a), and let $J^{\text{diag},*}(\lambda)$ denote the optimal value for a given λ.

Theorem 6. *$J^{\text{diag},*}(\lambda)$ is continuous and monotonically non-increasing with respect to $\lambda \in [0, 1]$.*

Remark 12. *(i) Note that Theorem 6 does not require the link latency functions to be polynomial.*

(ii) In light of Proposition 2, Theorem 6 implies that, if $n = 2$ and if the link latency functions are affine, then the optimal cost value under all, i.e., not necessarily (diagonal) atomic, private signals is continuous and monotonically non-increasing in $\lambda \in [0, 1]$. However, this is not necessarily the case with public signals, as we illustrate in Section 11.

(iii) *The proof of Theorem 6 (cf. [36]) implies that for a (not necessarily optimal) atomic diagonal signal π^{diag} for some $\nu_1 \in [0,1]$, one can construct a simple ν-dependent atomic diagonal signal with the same social cost as π^{diag} for all $\nu \in [\nu_1, 1]$. In other words, one can construct a simple feedback (using ν) atomic diagonal signal around a* nominal *π^{diag} under which the social cost does not increase due to higher than nominal fraction of receiving agents for which π^{diag} is designed. This is to be contrasted with existing results according to which the cost of receiving agents may increase with their increasing fraction under a fixed (open-loop) signal, e.g., see [33, 34].*

10 Public Signals

A public signal is an indirect signal, under which, for every state realization, ν fraction of agents all receive the same message among $\{1, \ldots, m\} =: [m]$. Formally, a public signal is a map $\pi^{\mathrm{pub}} : \Omega \to \Delta([m])$, or can alternately be represented as a $s \times m$ row stochastic matrix. The posterior formed by agents when the message they receive is k is:

$$\mu^{\pi^{\mathrm{pub}},k}(\omega) = \frac{\pi^{\mathrm{pub}}(k|\omega)\mu_0(\omega)}{\sum_{\psi} \pi^{\mathrm{pub}}(k|\psi)\mu_0(\psi)}, \qquad \omega \in \Omega \tag{40}$$

The joint posterior formed by agents who do not receive message, but have knowledge of π^{pub}, is:

$$\mu^{\pi^{\mathrm{pub}},\emptyset}(k,\omega) = \pi^{\mathrm{pub}}(k|\omega)\mu_0(\omega), \qquad k \in [m],\ \omega \in \Omega \tag{41}$$

Public signals over m messages have strong parallel with, but are not equivalent to, m-atomic private signals considered in Section 9.1. We return to this connection in Proposition 3.

Let $x^{(k)} \in \mathcal{P}(\nu)$ be the link flow induced by receiving agents, when the message they receive is $k \in [m]$, and let $y \in \mathcal{P}(1-\nu)$ be the link flow induced by agents not receiving the message. $x^{(k)}$ is the Bayes Nash flow with respect to the posterior in (40) and y is the Bayes Nash flow with respect to the posterior in (41). That is, $x^{(k)}$ satisfies:

$$\sum_{\omega} \ell_{\omega,i}(x_i^{(k)} + y_i)\mu^{\pi^{\mathrm{pub}},k}(\omega) \le \sum_{\omega} \ell_{\omega,j}(x_j^{(k)} + y_j)\mu^{\pi^{\mathrm{pub}},k}(\omega), \qquad i \in \mathrm{supp}(x^{(k)}),\ j \in [n]$$

Substituting the expression from (40), the conditions on $\{x^{(1)}, \ldots, x^{(m)}\}$ can be collectively rewritten as

$$x_i^{(k)} \sum_{\omega} \Big(\ell_{\omega,i}(x_i^{(k)} + y_i) - \ell_{\omega,j}(x_j^{(k)} + y_j)\Big) \pi^{\mathrm{pub}}(k|\omega)\mu_0(\omega) \le 0, \qquad i,j \in [n], \quad k \in [m] \tag{42}$$

Similarly, the condition on y can be written as

$$y_i \sum_{k,\omega} \Big(\ell_{\omega,i}(x_i^{(k)} + y_i) - \ell_{\omega,j}(x_j^{(k)} + y_j)\Big) \pi^{\mathrm{pub}}(k|\omega)\mu_0(\omega) \le 0, \qquad i,j \in [n] \tag{43}$$

The social cost is:

$$J(\pi^{\mathrm{pub}}, x, y) := \sum_{k,i,\omega} (x_i^{(k)} + y_i)\,\ell_{\omega,i}(x_i^{(k)} + y_i)\,\pi^{\mathrm{pub}}(k|\omega)\mu_0(\omega) \tag{44}$$

Therefore, the problem of optimal public signal design can be written as:

$$\min_{\substack{x^{(k)} \in \mathcal{P}(\alpha),\, k \in [m] \\ y \in \mathcal{P}(1-\alpha) \\ \pi^{pub} \in \Pi(m)}} J(\pi^{pub}, x, y) \quad \text{s.t. (42) – (43)} \tag{45}$$

Similar to (37), (45) is a third degree polynomial optimization problem for affine link latency functions.

Example 2. *Two public signals which have attracted particular interest are* full information *and* no information*:*

$$\pi^{pub,\,full} = \begin{array}{c} \\ \omega_1 \\ \omega_2 \\ \vdots \\ \omega_s \end{array} \begin{bmatrix} k=1 & k=2 & \dots & k=m \\ 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix}, \qquad \pi^{pub,\,no} = \begin{array}{c} \\ \omega_1 \\ \omega_2 \\ \vdots \\ \omega_s \end{array} \begin{bmatrix} k=1 & k=2 & \dots & k=m \\ 1 & 0 & \dots & 0 \\ 1 & 0 & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ 1 & 0 & \dots & 0 \end{bmatrix} \tag{46}$$

where $m = s$ *for the full information signal, and* m *is arbitrary, e.g.,* $m = 1$*, for the no information signal. In general, any row-stochastic* $\pi^{pub,\,no}$ *with identical rows corresponds to a no information signal.*

It is sometimes of interest to evaluate the cost of a given public signal. The cost can be computed once the induced flows $x^{(k)}$, $k \in [m]$, and y are known, which in turn can be computed using the next result.

Lemma 2. *The link flows,* y *and* $x^{(k)}, k \in [m]$*, induced by a public signal* π^{pub} *are solutions to*

$$\min_{y \in \mathcal{P}(1-\alpha);\; x^{(k)} \in \mathcal{P}(\alpha),\, k \in [m]} \sum_{i,\omega,k} \int_0^{x_i^{(k)}+y_i} \ell_{\omega,i}(z)\, dz\, \pi^{pub}(k|\omega)\, \mu_0(\omega) \tag{47}$$

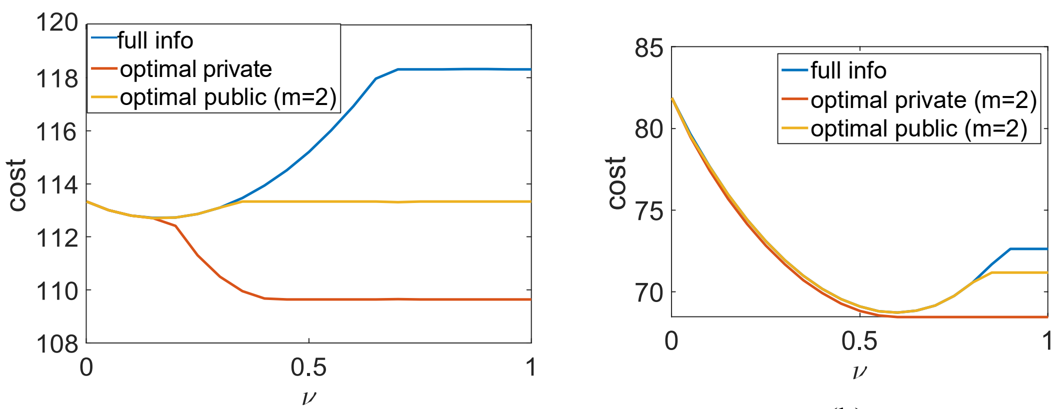
It is interesting to compare the formulations in (37) and (45) for m-atomic private signals and public signals with m messages respectively. While next result implies that every public signal with m messages can be equivalently realized by an m-atomic private signal, the converse is not true in general.

Proposition 3. *Given a* $\alpha \in [0,1]$*, for every public signal* π^{pub} *with* m *messages, there exists an* m*-atomic direct private signal with the same cost.*

Remark 13. *Proposition 3 implies that, for every* $\alpha \in [0,1]$*, there exists a feasible 1-atomic private signal corresponding to* $\pi^{pub,\,no}$ *in (46) with* $m = 1$*. Therefore, (34) is feasible for every* $\alpha \in [0,1]$*. Considering* s *duplicates of the same atom as for* $m = 1$ *case implies that (38) is feasible for all* $\alpha \in [0,1]$*. Feasibility of (37) can be established along similar lines.*

11 Simulations: Information Design for Traffic Networks

We compare the minimum cost achievable under private signals, public signals, and full information over two parallel links under affine (Section 11.1) and BPR latency functions (Section 11.2). The computations were performed using a combination of `GloptiPoly` and the `MultiStart` function (with `fmincon` solver) in MATLAB. In particular, the upper bound computed by `MultiStart` allows to certify optimality of the lower bound obtained from `GloptiPoly`, especially when the solution from `GloptiPoly` does not come with an explicit certificate of optimality. In all the instances, it was found sufficient to have 125 starting points for `MultiStart` and relaxation order of 3 for `GloptiPoly`. The no information signal corresponds to $\nu = 0$, when all the costs are expectedly equal. For both the scenarios, the total demand is set to be 5.



(a)

(b)

Figure 12: Comparison of minimum cost achievable under private signals, public signals and full information over two parallel links, under different ν for (a) affine latency functions and (b) BPR latency functions.

11.1 Affine Latency Functions

Figure 12(a) provides comparison between social costs for the following simulation parameters:

$$\alpha_0 = \begin{array}{c} \\ \omega_1 \\ \omega_2 \end{array} \begin{array}{c} \begin{array}{cc} i=1 & i=2 \end{array} \\ \begin{bmatrix} 5 & 25 \\ 20 & 15 \end{bmatrix} \end{array}, \qquad \alpha_1 = \begin{array}{c} \\ \omega_1 \\ \omega_2 \end{array} \begin{array}{c} \begin{array}{cc} i=1 & i=2 \end{array} \\ \begin{bmatrix} 4 & 2 \\ 1 & 2 \end{bmatrix} \end{array}, \qquad \mu_0 = \begin{array}{c} \omega_1 \\ \omega_2 \end{array} \begin{bmatrix} 0.6 \\ 0.4 \end{bmatrix}$$

The minimum social cost, i.e., the social cost when the planner can mandate which route every (receiving as well as non-receiving) agent takes for every realization of ω[4], for these parameters is 83.33. Following Proposition 2, optimal private signal is computed using (38). The approximation to optimal social cost under public signals using (45) was found to be identical for $m = 2, 3, 4$, and therefore these values are plotted under *optimal* public signal in Figure 12(a). The corresponding

[4]This is also referred to as the *first-best* strategy.

public signals for a few representative ⌫ are:

$$\text{⌫} = 0.25: \quad x = \begin{array}{c} \\ i=1 \\ i=2 \end{array}\begin{bmatrix} 1.25 & 0 \\ 0 & 1.25 \end{bmatrix}_{k=1,\,k=2}, \quad y = \begin{array}{c} i=1 \\ i=2 \end{array}\begin{bmatrix} 3.23 \\ 0.52 \end{bmatrix}, \quad \text{⇡}^{pub} = \begin{array}{c} !_1 \\ !_2 \end{array}\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}_{k=1,\,k=2}$$

$$\text{⌫} = 0.5: \quad x = \begin{array}{c} i=1 \\ i=2 \end{array}\begin{bmatrix} 2.06 & 2.06 \\ 0.44 & 0.44 \end{bmatrix}_{k=1,\,k=2}, \quad y = \begin{array}{c} i=1 \\ i=2 \end{array}\begin{bmatrix} 2.11 \\ 0.39 \end{bmatrix}, \quad \text{⇡}^{pub} = \begin{array}{c} !_1 \\ !_2 \end{array}\begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}_{k=1,\,k=2}$$

$$\text{⌫} = 0.75: \quad x = \begin{array}{c} i=1 \\ i=2 \end{array}\begin{bmatrix} 3.75 & 0 \\ 0 & 3.75 \end{bmatrix}_{k=1,\,k=2}, \quad y = \begin{array}{c} i=1 \\ i=2 \end{array}\begin{bmatrix} 0.42 \\ 0.83 \end{bmatrix}, \quad \text{⇡}^{pub} = \begin{array}{c} !_1 \\ !_2 \end{array}\begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}_{k=1,\,k=2}$$

$$\text{⌫} = 1: \quad x = \begin{array}{c} i=1 \\ i=2 \end{array}\begin{bmatrix} 4.17 & 0.2 \\ 0.83 & 4.8 \end{bmatrix}_{k=1,\,k=2}, \quad y = \begin{array}{c} i=1 \\ i=2 \end{array}\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad \text{⇡}^{pub} = \begin{array}{c} !_1 \\ !_2 \end{array}\begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}_{k=1,\,k=2}$$

and optimal private signals for the same representative ⌫ are:

$$\text{⌫} = 0.25: \quad x = \begin{array}{c} i=1 \\ i=2 \end{array}\begin{bmatrix} 0.32 & 0 \\ 0.93 & 1.25 \end{bmatrix}_{k=1,\,k=2}, \quad y = \begin{array}{c} i=1 \\ i=2 \end{array}\begin{bmatrix} 3.75 \\ 0 \end{bmatrix}, \quad \text{⇡} = \begin{array}{c} !_1 \\ !_2 \end{array}\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}_{k=1,\,k=2}$$

$$\text{⌫} = 0.5: \quad x = \begin{array}{c} i=1 \\ i=2 \end{array}\begin{bmatrix} 1.58 & 0.37 \\ 0.92 & 2.13 \end{bmatrix}_{k=1,\,k=2}, \quad y = \begin{array}{c} i=1 \\ i=2 \end{array}\begin{bmatrix} 2.5 \\ 0 \end{bmatrix}, \quad \text{⇡} = \begin{array}{c} !_1 \\ !_2 \end{array}\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}_{k=1,\,k=2}$$

$$\text{⌫} = 0.75: \quad x = \begin{array}{c} i=1 \\ i=2 \end{array}\begin{bmatrix} 2.83 & 1.62 \\ 0.92 & 2.13 \end{bmatrix}_{k=1,\,k=2}, \quad y = \begin{array}{c} i=1 \\ i=2 \end{array}\begin{bmatrix} 1.25 \\ 0 \end{bmatrix}, \quad \text{⇡} = \begin{array}{c} !_1 \\ !_2 \end{array}\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}_{k=1,\,k=2}$$

$$\text{⌫} = 1: \quad x = \begin{array}{c} i=1 \\ i=2 \end{array}\begin{bmatrix} 4.08 & 2.87 \\ 0.92 & 2.13 \end{bmatrix}_{k=1,\,k=2}, \quad y = \begin{array}{c} i=1 \\ i=2 \end{array}\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad \text{⇡} = \begin{array}{c} !_1 \\ !_2 \end{array}\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}_{k=1,\,k=2}$$

While the cost in Figure 12(a) shows non-monotonic behavior with respect to ⌫ in the full information case as well as under optimal public signal, the optimal cost is monotonically non-decreasing under private signals. Expectedly, the optimal cost under public signal is no greater than the cost under full information, and the optimal cost under private signal is no greater than under public signal. Interestingly, in this case, full information is an optimal public signal for small values of ⌫, and gives the same cost as an optimal private signal for even smaller values of ⌫.

11.2 BPR Latency Functions

Figure 12(b) provides comparison between social costs for the following simulation parameters:

$$\text{↲}_0 = \begin{array}{c} !_1 \\ !_2 \end{array}\begin{bmatrix} 5 & 25 \\ 20 & 15 \end{bmatrix}_{i=1,\,i=2}, \qquad \text{↲}_1 = \text{↲}_2 = \text{↲}_3 = \mathbf{0}, \qquad \text{↲}_4 = \begin{array}{c} !_1 \\ !_2 \end{array}\begin{bmatrix} 0.047 & 0.025 \\ 0.037 & 0.058 \end{bmatrix}_{i=1,\,i=2}, \qquad \mu_0 = \begin{array}{c} !_1 \\ !_2 \end{array}\begin{bmatrix} 0.6 \\ 0.4 \end{bmatrix}$$

These parameters correspond to free flow travel times and capacities being equal to ↲_0 and $\begin{array}{c} !_1 \\ !_2 \end{array}\begin{bmatrix} 2 & 3.5 \\ 3 & 2.5 \end{bmatrix}_{i=1,\,i=2}$ respectively. The minimum social cost for these parameters is 52.78.

The approximation to optimal social cost under private signals using (37) was found to be identical for $m = 2, 3, 4$, suggesting that $m = 2$ atoms are possibly sufficient to realize optimal private signal in this case. This is much less than the upper bound of $2\binom{6}{5} = 12$ atoms given by Theorem 5. Similarly, the approximation to optimal social cost under public signals using (45) was found to be identical for $m = 2, 3, 4$. Therefore, values for $m = 2$ are plotted under optimal private and optimal public, respectively, in Figure 12(b). Optimal public signals for a few representative ⌫ are:

$$⌫ = 0.25: \quad x = \begin{array}{c} \\ i=1 \\ i=2 \end{array} \begin{array}{cc} k=1 & k=2 \\ \left[\begin{array}{cc} 1.25 & 0 \\ 0 & 1.25 \end{array}\right. \end{array}, \quad y = \begin{array}{c} i=1 \\ i=2 \end{array} \left[\begin{array}{c} 3.75 \\ 0 \end{array}\right., \quad ⇡^{pub} = \begin{array}{c} \\ !_1 \\ !_2 \end{array} \begin{array}{cc} k=1 & k=2 \\ \left[\begin{array}{cc} 1 & 0 \\ 0 & 1 \end{array}\right. \end{array}$$

$$⌫ = 0.5: \quad x = \begin{array}{c} \\ i=1 \\ i=2 \end{array} \begin{array}{cc} k=1 & k=2 \\ \left[\begin{array}{cc} 2.5 & 0 \\ 0 & 2.5 \end{array}\right. \end{array}, \quad y = \begin{array}{c} i=1 \\ i=2 \end{array} \left[\begin{array}{c} 2.5 \\ 0 \end{array}\right., \quad ⇡^{pub} = \begin{array}{c} \\ !_1 \\ !_2 \end{array} \begin{array}{cc} k=1 & k=2 \\ \left[\begin{array}{cc} 1 & 0 \\ 0 & 1 \end{array}\right. \end{array}$$

$$⌫ = 0.75: \quad x = \begin{array}{c} \\ i=1 \\ i=2 \end{array} \begin{array}{cc} k=1 & k=2 \\ \left[\begin{array}{cc} 3.75 & 0 \\ 0 & 3.75 \end{array}\right. \end{array}, \quad y = \begin{array}{c} i=1 \\ i=2 \end{array} \left[\begin{array}{c} 1.25 \\ 0 \end{array}\right., \quad ⇡^{pub} = \begin{array}{c} \\ !_1 \\ !_2 \end{array} \begin{array}{cc} k=1 & k=2 \\ \left[\begin{array}{cc} 1 & 0 \\ 0 & 1 \end{array}\right. \end{array}$$

$$⌫ = 1: \quad x = \begin{array}{c} \\ i=1 \\ i=2 \end{array} \begin{array}{cc} k=1 & k=2 \\ \left[\begin{array}{cc} 5.0 & 2.08 \\ 0.0 & 2.92 \end{array}\right. \end{array}, \quad y = \begin{array}{c} i=1 \\ i=2 \end{array} \left[\begin{array}{c} 0 \\ 0 \end{array}\right., \quad ⇡^{pub} = \begin{array}{c} \\ !_1 \\ !_2 \end{array} \begin{array}{cc} k=1 & k=2 \\ \left[\begin{array}{cc} 1 & 0.13 \\ 0 & 0.87 \end{array}\right. \end{array}$$

and optimal private signals for the same representative ⌫ are:

$$⌫ = 0.25: \quad x = \begin{array}{c} \\ i=1 \\ i=2 \end{array} \begin{array}{cc} k=1 & k=2 \\ \left[\begin{array}{cc} 0.99 & 0 \\ 0.26 & 1.25 \end{array}\right. \end{array}, \quad y = \begin{array}{c} i=1 \\ i=2 \end{array} \left[\begin{array}{c} 3.75 \\ 0 \end{array}\right., \quad ⇡ = \begin{array}{c} \\ !_1 \\ !_2 \end{array} \begin{array}{cc} k=1 & k=2 \\ \left[\begin{array}{cc} 1 & 0 \\ 0 & 1 \end{array}\right. \end{array}$$

$$⌫ = 0.5: \quad x = \begin{array}{c} \\ i=1 \\ i=2 \end{array} \begin{array}{cc} k=1 & k=2 \\ \left[\begin{array}{cc} 2.24 & 0.0 \\ 0.26 & 2.5 \end{array}\right. \end{array}, \quad y = \begin{array}{c} i=1 \\ i=2 \end{array} \left[\begin{array}{c} 2.5 \\ 0 \end{array}\right., \quad ⇡ = \begin{array}{c} \\ !_1 \\ !_2 \end{array} \begin{array}{cc} k=1 & k=2 \\ \left[\begin{array}{cc} 1 & 0 \\ 0 & 1 \end{array}\right. \end{array}$$

$$⌫ = 0.75: \quad x = \begin{array}{c} \\ i=1 \\ i=2 \end{array} \begin{array}{cc} k=1 & k=2 \\ \left[\begin{array}{cc} 3.49 & 0.76 \\ 0.26 & 2.99 \end{array}\right. \end{array}, \quad y = \begin{array}{c} i=1 \\ i=2 \end{array} \left[\begin{array}{c} 1.25 \\ 0 \end{array}\right., \quad ⇡ = \begin{array}{c} \\ !_1 \\ !_2 \end{array} \begin{array}{cc} k=1 & k=2 \\ \left[\begin{array}{cc} 1 & 0 \\ 0 & 1 \end{array}\right. \end{array}$$

$$⌫ = 1: \quad x = \begin{array}{c} \\ i=1 \\ i=2 \end{array} \begin{array}{cc} k=1 & k=2 \\ \left[\begin{array}{cc} 4.74 & 2.01 \\ 0.26 & 2.99 \end{array}\right. \end{array}, \quad y = \begin{array}{c} i=1 \\ i=2 \end{array} \left[\begin{array}{c} 0 \\ 0 \end{array}\right., \quad ⇡ = \begin{array}{c} \\ !_1 \\ !_2 \end{array} \begin{array}{cc} k=1 & k=2 \\ \left[\begin{array}{cc} 1 & 0 \\ 0 & 1 \end{array}\right. \end{array}$$

The social cost profile in Figure 12(b) shows similar qualitative dependence on ⌫ as in Figure 12(a). Since diagonal atomic private signals are observed to be optimal (based on the sample values reported above), monotonicity of the corresponding cost is consistent with Theorem 6.

12 Conclusions and Future Work

12.1 Feedback Control of Traffic Networks

This project provides some structural insights into the finite-horizon optimal feedback control for flow networks. The enabling tool for the design of an optimal feedback control law is the multi-parametric linear program. It is well known that for large-size complex networks, the prohibitive computation and computation loads makes the design and implementation of a centralized controller too costly or impractical; moreover, the effect of noise, delay, or any type of error or failure in data transmission may substantially degrade the control quality. It is, therefore, necessary to develop decentralized feedback controllers with simple structure. A simple procedure is proposed to design a decentralized feedback control with a "one-hop" information structure. Moreover, it is shown that the optimal feedback controller with respect to certain linear performance indexes possesses a one-hop information structure, making the optimal controller suitable for practical implementations in large-scale networks. This suggests that if certain conditions are satisfies, the *trivial control* (with the least computational/communication cost) can provide the same (or very close) performance to that of the *centralized control* (with the most computational/communication cost).

For a given flow network of size n and control horizon N, it is invaluable to analytically determine when it is worth to implement uncontrolled scheme, or a decentralized control law with a p-hop information structure to achieve a satisfactory level of performance. We also plan to extend and evaluate the approach to higher order traffic dynamics, such as ARZ and PW models. Our ultimate objective is to develop a principled approach for distributed optimal control of physical infrastructure networks under given information constraints.

12.2 Information Design for Traffic Networks

Information design for non-atomic routing games is gaining increasing attention. While existing works provide useful insights through analysis of simple scenarios, the generality of these insights is not readily apparent. Relatedly, a computational approach to operationalize optimal information design for general settings does not exist to the best of our knowledge. By making connection to semidefinite programming (SDP), this project not only fills this gap, but also allows to leverage computational tools developed by the SDP community. The latter is particularly relevant for extending the approach to non-atomic games beyond routing.

There are several immediate directions for future work. The bound in Theorem 5 on the number of atoms required to realize optimal private signals may be computationally prohibitive for large networks. Proposition 2 and Section 11.2 on the other hand suggest the possibility of exploring problem structure to tighten the bound. A counterpart to Theorem 5 for public signals remains open. A relatively unexplored direction is to provide sub-optimality bounds for simple classes of signals such as diagonal atomic. Finally, it would be interesting to utilize the approach in this project to quantify the reduction in *price of anarchy* under information design. This will complement existing work, e.g., in [23], where such an analysis is provided under specific models for correlation between coefficients of affine latency functions across links, and under a specific class of signals.

13 Implementation

Not applicable.

14 References

References

[1] B. R. Munson, A. P. Rothmayer, T. H. Okiishi, and W. W. Huebsch. *Fundamentals of Fluid Mechanics*. John Wiley & Sons, Inc., 2012.

[2] G. Como, E. Lovisari, and K. Savla. Convexity and robustness of dynamic traffic assignment and freeway network control. *Transportation Research Part B: Methodological*, 91:446–465, 2016.

[3] Y. Han, A. Hegyi, Y. Yuan, S. Hoogendoorn, M. Papageorgiou, and C. Roncoli. Resolving freeway jam waves by discrete first-order model-based predictive control of variable speed limits. *Transportation Research Part C: Emerging Technologies*, 77:405–420, 2017.

[4] A. Muralidharan and R. Horowitz. Computationally efficient model predictive control of freeway networks. *Transportation Research Part C: Emerging Technologies*, 58:532–553, 2015.

[5] C. Daganzo. The cell transmission model: A dynamic representation of highway traffic consistent with the hydrodynamic theory. *Transportation Research Part B*, 28(4):269–287, 1994.

[6] L. Adacher and M. Tiriolo. A macroscopic model with the advantages of microscopic model: A review of cell transmission model's extensions for urban traffic networks. *Simulation Modelling Practice and Theory*, 86:102–119, 2018.

[7] P. Wong and R. Larson. Optimization of natural-gas pipeline systems via dynamic programming. *IEEE Transactions on Automatic Control*, 13(5):475–481, 1968.

[8] S. Misra, M. W. Fisher, S. Backhaus, R. Bent, M. Chertkov, and F. Pan. Optimal compression in natural gas networks: A geometric programming approach. *IEEE Transactions on Control of Network Systems*, 2(1):47–56, 2015.

[9] A. Martin, M. Möller, and S. Moritz. Mixed integer models for the stationary case of gas network optimization. *Mathematical Programming*, 105(2):563–582, 2006.

[10] A. Hegyi, B. D. Schutter, and H. Hellendoorn. Model predictive control for optimal coordination of ramp metering and variable speed limits. *Transportation Research Part C: Emerging Technologies*, 13(3):185–209, 2005.

[11] I. Papamichail, A. Kotsialos, I. Margonis, and M. Papageorgiou. Coordinated ramp metering for freeway networks– A model-predictive hierarchical control approach. *Transportation Research Part C: Emerging Technologies*, 18(3):311–331, 2010.

[12] M. Hadiuzzaman and T. Z. Qiu. Cell transmission model based variable speed limit control for freeways. *Canadian Journal of Civil Engineering*, 40(1):46–56, 2013.

[13] C. N. Jones, M. Baric, and M. Morari. Multiparametric linear programming with applications to control. *European Journal of Control*, 13(2).

[14] M. Herceg, M. Kvasnica, C. N. Jones, and M. Morari. Multi-Parametric Toolbox 3.0. In *Proc. of the European Control Conference*, pages 502–510, Zürich, Switzerland, July 17–19 2013.

[15] M. Baotic, F. J. Christophersen, and M. Morari. A new algorithm for constrained finite time optimal control of hybrid systems with a linear performance index. In *2003 European Control Conference (ECC)*, pages 3323–3328, 2003.

[16] J. Tsitsiklis and M. Athans. On the complexity of decentralized decision making and detection problems. *IEEE Transactions on Automatic Control*, 30(5):440–446, 1985.

[17] R. Cogill, M. Rotkowitz, B. V. Roy, and S. Lall. An approximate dynamic programming approach to decentralized control of stochastic systems. In *Control of Uncertain Systems: Modelling, Approximation, and Design*, pages 243–256. Springer Berlin Heidelberg, 2006.

[18] H. Lakshmanan and D. P. de Farias. Decentralized approximate dynamic programming for dynamic networks of agents. In *2006 American Control Conference (ACC)*, pages 1648–1653, June 2006.

[19] Dirk Bergemann and Stephen Morris. Information design: A unified perspective. *Journal of Economic Literature*, 57(1):44–95, 2019.

[20] Emir Kamenica and Matthew Gentzkow. Bayesian persuasion. *American Economic Review*, 101(6):2590–2615, 2011.

[21] Shaddin Dughmi and Haifeng Xu. Algorithmic Bayesian persuasion. In *ACM Symposium on the Theory of Computing*, 2016.

[22] Daron Acemoglu, Ali Makhdoumi, Azarakhsh Malekian, and Asuman Ozdaglar. Informational Braess' paradox: The effect of information on traffic congestion. *arXiv preprint arXiv:1601.02039*, 2016.

[23] Shoshana Vasserman, Michal Feldman, and Avinatan Hassidim. Implementing the wisdom of Waze. In *IJCAI*, pages 660–666, 2015.

[24] S. Das, E. Kamenica, and R. Mirka. Reducing congestion through information design. In *Allerton Conference on Communication, Control and Computing*, 2017.

[25] Manxi Wu and Saurabh Amin. Information design for regulating traffic flows under uncertain network state. In *2019 57th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pages 671–678. IEEE, 2019. Extended version at `https://arxiv.org/abs/1908.07105`.

[26] Hamidreza Tavafoghi and Demosthenis Teneketzis. Strategic information provision in routing games. Available at https://hamidtavaf.github.io/infodesign_routing.pdf.

[27] Olivier Massicot and Cedric Langbort. Public signals and persuasion for road network congestion games under vagaries. *IFAC-PapersOnLine*, 51(34):124–130, 2019.

[28] Jean B Lasserre. A semidefinite programming approach to the generalized problem of moments. *Mathematical Programming*, 112(1):65–92, 2008.

[29] Didier Henrion, Jean-Bernard Lasserre, and Johan Löfberg. Gloptipoly 3: moments, optimization and semidefinite programming. *Optimization Methods & Software*, 24(4-5):761–779, 2009.

[30] Christian Bayer and Josef Teichmann. The proof of Tchakaloff's theorem. *Proceedings of the American mathematical society*, 134(10):3035–3040, 2006.

[31] Jean B Lasserre. Global optimization with polynomials and the problem of moments. *SIAM Journal on optimization*, 11(3):796–817, 2001.

[32] Noah D Stein, Pablo A Parrilo, and Asuman Ozdaglar. Correlated equilibria in continuous games: Characterization and computation. *Games and Economic Behavior*, 71(2):436–455, 2011.

[33] Hani S Mahmassani and R Jayakrishnan. System performance and user response under real-time information in a congested traffic corridor. *Transportation Research Part A: General*, 25(5):293–307, 1991.

[34] Manxi Wu, Saurabh Amin, and Asuman E Ozdaglar. Value of information systems in routing games. *arXiv preprint arXiv:1808.10590*, 2018.

[35] S. Jafari and K. Savla. A decentralized optimal feedback flow control approach for transport networks. `https://arxiv.org/pdf/1805.11271.pdf`; preliminary version at ACC 2019.

[36] Y. Zhu and K. Savla. A computational approach for information design in non-atomic routing games. Available at `https://viterbi-web.usc.edu/~ksavla/papers/infodesign-sdp.pdf`, 2020.

[37] F. Borrelli. *Constrained Optimal Control of Linear and Hybrid Systems*. Springer, 2003.

[38] J. Lofberg. YALMIP: A toolbox for modeling and optimization in MATLAB.

[39] J. Krumm. Where will they turn: predicting turn proportions at intersections. *Personal and Ubiquitous Computing*, 14(7):591–599, 2010.

[40] A. K. Singh and B. C. Pal. An extended linear quadratic regulator for LTI systems with exogenous inputs. *Automatica*, 76:10–16, 2017.

[41] D. Bertsimas and A. Thiele. A robust optimization approach to inventory theory. *Operations Research*, 54(1):150–168, 2006.

[42] V. Gabrel, C. Murat, and N. Remli. Linear programming with interval right hand sides. *International Transactions in Operational Research*, 17(3):397–408, 2010.

[43] J. W. Chinneck and K. Ramadan. Linear programming with interval coefficients. *The Journal of the Operational Research Society*, 51(2):209–220, 2000.

[44] David Branston. Link capacity functions: A review. *Transportation research*, 10(4):223–236, 1976.