

23rd International Symposium on Transportation and Traffic Theory, ISTTT 23, 24-26 July 2019,
Lausanne, Switzerland

City-wide traffic control: modeling impacts of cordon queues

Wei Ni^a, Michael Cassidy^{a,*}

^a*Department of Civil and Environmental Engineering, University of California, Berkeley*

Abstract

Optimal cordon-metering rates are obtained using Macroscopic Fundamental Diagrams in combination with flow conservation laws. A model-predictive control algorithm is also used so that time-varying metering rates are generated based on their forecasted impacts. Our scalable algorithm can do this for an arbitrary number of cordoned neighborhoods within a city. Unlike its predecessors, the proposed model accounts for the time-varying constraining effects that cordon queues impose on a neighborhood's circulating traffic, as those queues expand and recede over time. The model does so at every time step by approximating a neighborhood's street space occupied by cordon queues, and re-scaling the MFD to describe the state of circulating traffic that results. The model also differentiates between saturated and under-saturated cordon-metering operations. Computer simulations of an idealized network show that these enhancements can substantially improve the predictions of both, the trip completion rates in a neighborhood and the rates that vehicles cross metered cordons. Optimal metering policies generated as a result are similarly shown to do a better job in reducing the Vehicle Hours Traveled on the network. The VHT reductions stemming from the proposed model and from its predecessors differed by as much as 14%.

© 2017 The Authors. Published by Elsevier B.V.

Peer-review under responsibility of the scientific committee of the 23rd International Symposium on Transportation and Traffic Theory.

Keywords: Macroscopic Fundamental Diagram; Traffic control; Cordon metering

1. Introduction

A sizable literature exists on the re-timing of traffic signals to meter inflows to cordoned neighborhoods; e.g., see [1, 2, 4, 5, 7–13, 15–17, 19]. In some of those efforts, metering rates were optimized using Macroscopic Fundamental Diagrams (MFDs) in combination with flow conservation laws [5, 7, 15]. The works have produced what we will call Neighborhood Transmission Models.

These NTMs separately estimate a neighborhood's time-varying accumulations of vehicles that are bound for destinations residing in that same neighborhood, and in each of the other neighborhoods within a city. The original NTM in [5, 15] has a problem regarding trips of the latter (i.e. inter-neighborhood) kind. When cordon metering substantially delays vehicles from crossing from one neighborhood to another, those vehicles are returned to their

* Corresponding author. Tel.: +1-510-642-7702 ; fax: +1-510-643-8919.

E-mail address: cassidy@ce.berkeley.edu

neighborhood's circulating traffic for another go at the boundary. They start from scratch as if previous attempts at boundary crossings had not just occurred.

Recent work in [7] has addressed this deficiency by differentiating a neighborhood's accumulations of circulating vehicles from those queued at its metered cordons; see the example in Fig.1(a). Though a notable advancement, this more discriminating NTM comes with its own deficiency in terms of accounting for the constraining effects that a neighborhood's outbound cordon queues can impose on its circulating traffic. To its credit, the discriminating model in [7] recognizes that the street space occupied by (possibly lengthy) cordon queues can diminish the neighborhood's capacity to serve the rest of its traffic. It further recognizes that this constraining effect can be accounted for by re-scaling the neighborhood's MFD downward to account for the street space lost in storing cordon queues. The problem, however, is that users of the discriminating NTM are advised to: (i) estimate the maximum length to which cordon queues are likely to expand during the rush; (ii) down-size the MFD to account for the street space lost in storing those maximum-lengthed queues; and (iii) use this down-sized MFD for the entire analysis, as if the maximum-lengthed queues were always present in the neighborhood.

Readers of [7] are not advised how to estimate the cordon queues' maximum length. (This is said in [7] to be a subject of future research.)¹ Yet, even with a good estimate of the queues' maximum length, use of a single re-scaled MFD can be problematic during transition periods when those queues may be growing gradually toward their maximum extension (e.g. early in a rush); and then gradually receding (late in the rush).

The present paper offers an enhancement. It captures the time-varying impacts of cordon queues on network capacity by re-scaling a neighborhood's MFD in dynamic fashion. The MFD is incrementally down-sized each period that cordon queues expand,² and is re-scaled upward during periods when those queues are diminishing. The MFD's re-scaling can thus occur in any and every time step of an analysis, and seems to us a simple, but physically-realistic way of addressing the heterogeneous traffic conditions and hysteresis due to cordon queues.

The proposed NTM and its predecessors are tested using the AIMSUN micro-level traffic simulation model [3]. It furnished the closest approximations to ground truth that are available to us. The simulations show that our proposed NTM does as good or better a job in predicting the rates that vehicles both, cross metered cordons and complete their trips, as compared against predictions from the earlier models in [5, 7, 15]. These improved predictions produce, in turn, optimal cordon-metering rates that do a better job of reducing the Vehicle Hours Traveled in a city. Improvements can be sizable, as will be shown in due course.

Simulations also show that inputs needed by the proposed NTM can be collected from the information systems onboard connected vehicles, even when the penetration rate of those special vehicles is small. As a final aside, simulations show that optimal metering rates tend to alternate over time between one that is highly relaxed, and another that is highly restrictive. This finding is important because it lends credibility to a key assumption in our model, as will be explained.

2. Background

The NTMs in [5, 7, 15] are now reviewed and critiqued in more depth. For each model in that family, the vehicle accumulation in a neighborhood i at time t , $n_i(t)$, is assumed to be a measurable input. It is used with an MFD to obtain i 's total flow, $f_i(n_i)$, defined as the flow integrated over neighborhood i with units of $veh \cdot km/min$; see Fig.1(b). Recall from sec.1 that the $n_i(t)$ is composed of i 's accumulations of vehicles bound for neighborhood j , $n_{ij}(t)$, and those bound for destinations within i itself, $n_{ii}(t)$. How one might disentangle these more disaggregate accumulations from the $n_i(t)$ is not clear from the literature. We will assume that in the near future, the information available from connected vehicles in the traffic mix can be used to sample the $\frac{n_{ij}(t)}{n_i(t)}$ and $\frac{n_{ii}(t)}{n_i(t)}$, and to estimate other inputs to be described in due course.

¹ We add for good measure that readers are not advised how to re-scale an MFD to account for maximum-lengthed cordon queues.

² Cordon queues could grow so long that the resulting MFD is too small to be well-defined. One could prevent such queue expansion by imposing a penalty term in the objective function. However, imposing control rules for the sake of maintaining the integrity of the traffic model runs the risk of adding to a neighborhood's vehicle-hours traveled, and seems to us like letting the tail wag the dog.

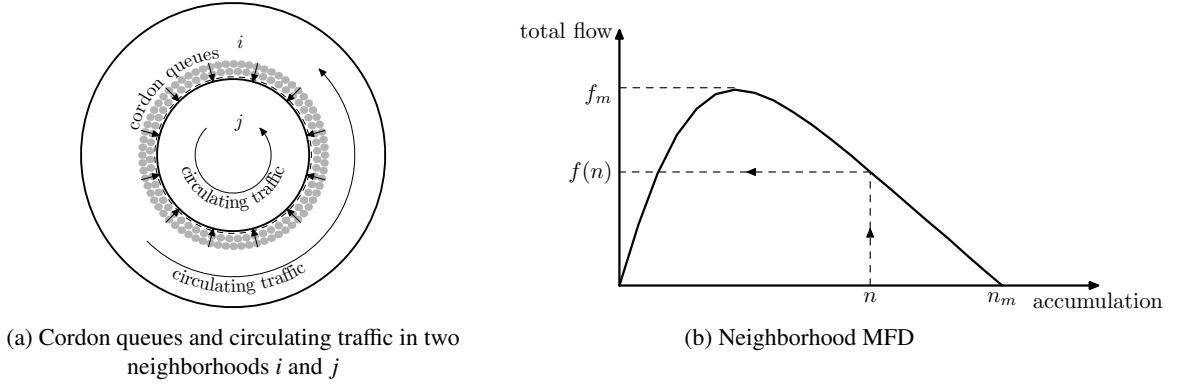


Fig. 1: Cordon queues and MFD

2.1. Original Model

To see how the earliest NTM in [5, 15] incorporated flow conservation laws, denote L_{ij} as the average distance that vehicles in i must travel to reach i 's boundary with j .³ Thus, $\frac{f_i(n_i(t))}{L_{ij}} \cdot \frac{n_{ij}(t)}{n_i(t)}$ is the rate that i 's vehicles bound for j reach that boundary; with units of *veh/min*. The minimum of that rate and the cordon metering rate, $u_{ij}(t) \cdot C_{ij}$, gives the rate that j -bound vehicles exit i for $j \neq i$, where: C_{ij} is the unmetered capacity at the cordon separating one neighborhood from another (i and j in the present context); and $u_{ij} \in [u_{min}, u_{max}]$ is the fraction of the capacity that is allowed to flow through the cordon.⁴ This cordon-crossing rate is subtracted from $\lambda_{ij}(t)$, the inter-neighborhood demand generated in i , with units of *veh/min*. The subtraction enables the updating of the n_{ij} at each time step of duration τ . Although different from what appears in [5, 15],⁵ the update can be expressed as

$$n_{ij}(t + \tau) - n_{ij}(t) = \tau \cdot \left[-\min\left\{\frac{f_i(n_i(t))}{L_{ij}} \cdot \frac{n_{ij}(t)}{n_i(t)}, C_{ij} \cdot u_{ij}(t)\right\} + \lambda_{ij}(t) \right]. \quad (1)$$

The n_{ii} are similarly updated by considering the intra-neighborhood demand generated in i , $\lambda_{ii}(t)$, the trip completion rate in i , and the inflows that were allowed to cross into i from j ; i.e.

$$n_{ii}(t + \tau) - n_{ii}(t) = \tau \cdot \left[-\frac{f_i(n_i(t))}{L_{ii}} \cdot \frac{n_{ii}(t)}{n_i(t)} + \lambda_{ii}(t) + \sum_{j \neq i} \min\left\{\frac{f_i(n_j)}{L_{ji}} \cdot \frac{n_{ji}(t)}{n_j(t)}, C_{ji} \cdot u_{ji}(t)\right\} \right], \quad (2)$$

where L_{ii} is the average distance that intra-neighborhood vehicles travel in i .

The problem arises when metering delays beyond duration τ those outbound vehicles that are queued and ready to cross the cordon. Note that (1) and (2) have no mechanism to carry-over those queued vehicles into the next time step. Those vehicles return instead to the n_{ij} , and are then assumed to be confronted (again!) by the distance L_{ij} that separates them from the cordon that they had just failed to cross.

These artificial additions to a neighborhood's circulating traffic can constrain its trip completions. And sending back queued vehicles can artificially starve a cordon line of flow if the metering there is suddenly relaxed.

2.2. More Discriminating Model

The problem was addressed in [7] by decomposing the $n_{ij}(t)$ into its accumulations of circulating vehicles, $n_{c,ij}(t)$, and of vehicles queued at its cordon, $n_{q,ij}(t)$, such that $n_{ij}(t) = n_{c,ij}(t) + n_{q,ij}(t)$. For convenience, denote i 's total

³ The L_{ij} can be estimated knowing the physical size of neighborhood i .

⁴ One might reasonably designate $u_{max} = 1$, to describe the flow that can enter the cordon when the traffic signals on the periphery are not re-timed to function as meters. One might impose a lower bound $u_{min} > 0$, to avoid the ire of j -bound drivers.

⁵ References [5, 15] omit the capacity of the metered cordon in the formulation of the system dynamics.

accumulations of circulating and cordon-queued vehicles as $n_{c,i}(t) = \sum_j n_{c,ij}(t)$ and $n_{q,i}(t) = \sum_j n_{q,ij}(t)$, respectively. This more discriminating NTM can therefore be formulated as

$$n_{c,ii}(t + \tau) - n_{c,ii}(t) = \tau \cdot \left[-\frac{\tilde{f}_i(n_{c,i}(t))}{L_{ii}} \cdot \frac{n_{c,ii}(t)}{n_{c,i}(t)} + \lambda_{ii}(t) + \sum_{j \neq i} C_{ji} \cdot u_{ji}(t) \right], \quad (3)$$

$$n_{c,ij}(t + \tau) - n_{c,ij}(t) = \tau \cdot \left[-\frac{\tilde{f}_i(n_{c,i}(t))}{L_{ij}} \cdot \frac{n_{c,ij}(t)}{n_{c,i}(t)} + \lambda_{ij}(t) \right], \quad (4)$$

$$n_{q,ij}(t + \tau) - n_{q,ij}(t) = \tau \cdot \left[-C_{ij} \cdot u_{ij}(t) + \frac{\tilde{f}_i(n_{c,i}(t))}{L_{ij}} \cdot \frac{n_{q,ij}(t)}{n_{c,i}(t)} \right]. \quad (5)$$

Equation (5) means that a cordon's residual queues are carried-over from one time step to the next. Moreover, the inclusion of C in (3) and (5) remedies any underestimates of cordon-crossing rates that might otherwise occur should metering abruptly relax.

New concerns emerge, however. Note from (3) and (4) how vehicle flow in i , $\tilde{f}_i(n_{c,i}(t))$, is solely a function of its circulating accumulation. As per our earlier remarks, the discriminating NTM captures the constraining effects of cordon queues in a static way. Recall that readers of [7] are advised to assume that the accumulation of cordon-queued vehicles always corresponds to maximum-lengthed queues, $n_{q,i,max}$. Once these accumulations are estimated in some (unspecified) fashion, the neighborhood MFD is downsized, such that $\tilde{f} \leq f$. The re-scaled MFD is used for the entire analysis period.⁶

Of further interest, the solution furnished in [7] was an analytical one. Though an impressive achievement, the approach limited the NTM's application to scenarios involving only two neighborhoods in a city. Moreover, the analytical solutions were tested only against numerical ones from that same (more discriminating) model. Those tests showed the analytical solutions to be consistent with numerical ones, but say nothing about the model's physical realism.

3. Proposed NTM

The model now proposed recognizes that a neighborhood's circulating flows are functions of its time-varying accumulations of both, the circulating and the cordon-queued vehicles, $n_c(t)$ and $n_q(t)$. We model the time-varying influence of n_q in a simple way as will be explained momentarily. To support our simple approach, we will show that an optimal cordon-control scheme is of bang-singular type, but entails bang-bang control almost everywhere in the system state space. This means that optimal metering along a cordon tends to alternate over time between fully relaxed and highly restrictive rates.⁷

The implication is that cordon queues that grow long enough to constrain circulating traffic will tend to be dense. This should be the case whether bang-bang control switches frequently or infrequently between extreme metering rates for reasons offered below.

Consider, for example, the cordon queues that persist during prolonged periods of very restrictive metering. Those queues will be dense in their own right. They can only grow denser by spilling-back and coalescing with upstream queues created by other traffic signals nearby. Moreover, if the queues that formed during extended periods of fully-relaxed metering were to grow long, their coalescence with other queues could make them dense as well. Or, if control switches frequently between relaxed and restrictive rates, recovery waves can be halted if and when they arrive at upstream signals during their red phases. Queues on the upstream links would remain jammed for the remainders of those reds, and possibly beyond if turning vehicles create queues downstream; see [6].

We exaggerate the above considerations and assume that cordon queues exhibit jam densities at all times. The assumption enables us to approximate the time-varying accumulation of cordon-queued vehicles and the street space lost

⁶ Note as an aside how the inclusion of C in (3) and (5) assumes that metered green times are always saturated. This ignores the possibility of under-saturated metering owing to low demand for crossing a cordon. Our numerical tests in sec.5 will relax this problematic assumption, as per a suggestion offered in a footnote of [7].

⁷ Theoretical derivations are presented in Appendix A, and experimental support is furnished in sec.5.

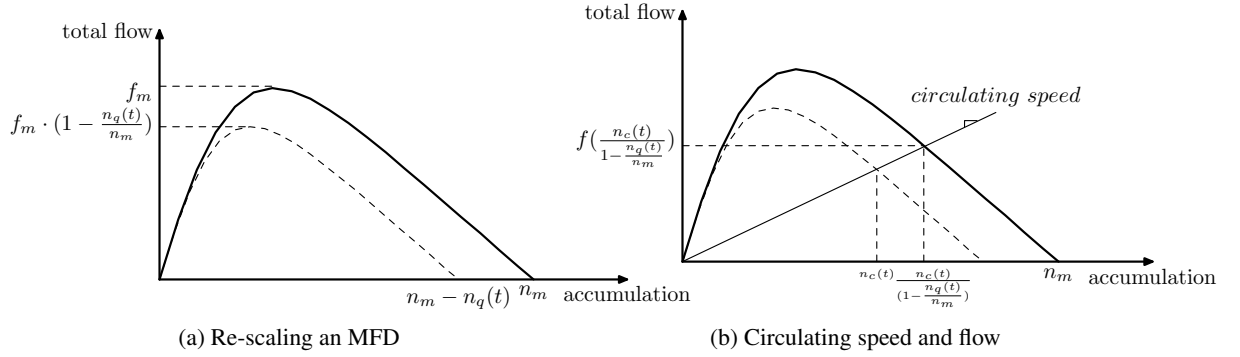


Fig. 2: Re-scaled MFD and implication on circulating speed

in storing those queues. The approximations are based on lower-bound estimates; i.e. vehicles in less-than-jammed queues exhibit larger-than-jammed spacings, and thus collectively occupy greater street space. Still, our lower-bound estimates seem to capture a lion's share of the constraining effects that cordon queues impose on neighborhood circulation. This point will be illustrated in sec.5. The point for now is that our approximation enables a rescaling downward of a neighborhood's MFD. Details of the re-scaling are furnished below.

3.1. Modeling Cordon-Queue Impacts

With the assumption of jammed cordon queues, the maximum accumulation of circulating vehicles that a neighborhood can store becomes $n_m - n_q$; i.e. the neighborhood's vehicle-storage capacity, n_m , is diminished by the factor $(1 - \frac{n_q}{n_m})$, as shown in Fig.2(a). That same multiplicative factor applied to f_m , the neighborhood's maximum possible total flow in the absence of cordon queues, gives the capacity left available for circulating flows. The assumption here is only that the neighborhood's capacity to circulate traffic diminishes in proportion to its street space given to cordon queues.

Capturing the constraining effect that those queues exert on circulating speeds would require adjustments using the same factor. Hence, the neighborhood's average circulating speed is given by $f(\frac{n_c}{1 - \frac{n_q}{n_m}}) \cdot (\frac{n_c}{1 - \frac{n_q}{n_m}})^{-1}$, as exemplified by the slope of the chord in Fig.2(b). The product of this speed and n_c is the neighborhood's total circulating flow. That flow is therefore

$$\bar{f}(n_c, n_q) = f\left(\frac{n_c}{1 - \frac{n_q}{n_m}}\right) \cdot \left(1 - \frac{n_q}{n_m}\right), \quad (6)$$

as given by a re-scaled MFD, like the dashed ones in Figs.2(a) and 2(b). The proposed re-scaling can occur in every time step.

Recall that (6) represents a departure from the logic behind the original NTM [5, 15], which assumes that circulating flows $f(n) = f(n_q + n_c)$, as if vehicles in the n_q were themselves circulating. The proposed logic is likewise different from that of the discriminating NTM [7], which assumes that street space is always lost to maximum-lengthed cordon queues.

3.2. Saturated vs Under-Saturated Metering and Model Formulation

Denote as $d_{ij}(t)$ the number of vehicles to cross a cordon from i to j during time window $[t, t + \tau]$. We do not assume that $d_{ij}(t)$ always equals $C_{ij} \cdot u_{ij}(t) \cdot \tau$, as if the meters were always saturated. When cordon queues are sufficiently short, $d_{ij}(t)$ is instead the sum of the $n_{q,ij}(t)$, the cordon's queued vehicles present at the start of the time window, and the inter-neighborhood vehicles to arrive at the cordon during duration τ . Hence,

$$d_{ij}(t) = \min\left\{C_{ij} \cdot u_{ij}(t) \cdot \tau, n_{q,ij}(t) + \tau \cdot \frac{\bar{f}_i(n_{c,i}(t), n_{q,i}(t))}{L_{ij}} \cdot \frac{n_{c,ij}(t)}{n_{c,i}(t)}\right\}. \quad (7)$$

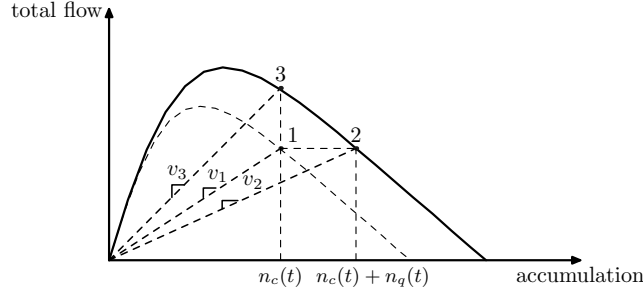


Fig. 3: Circulating speeds predicted by proposed NTM and predecessors

The formulation for the proposed NTM is thus

$$n_{c,ii}(t + \tau) - n_{c,ii}(t) = \tau \cdot \left[-\frac{\bar{f}_i(n_{c,i}(t), n_{q,i}(t))}{L_{ii}} \cdot \frac{n_{c,ii}(t)}{n_{c,i}(t)} + \lambda_{ii}(t) \right] + \sum_{j \neq i} d_{ji}(t), \quad (8)$$

$$n_{c,ij}(t + \tau) - n_{c,ij}(t) = \tau \cdot \left[-\frac{\bar{f}_i(n_{c,i}(t), n_{q,i}(t))}{L_{ij}} \cdot \frac{n_{c,ij}(t)}{n_{c,i}(t)} + \lambda_{ij}(t) \right], \quad (9)$$

$$n_{q,ij}(t + \tau) - n_{q,ij}(t) = -d_{ij}(t) + \tau \cdot \frac{\bar{f}_i(n_{c,i}(t), n_{q,i}(t))}{L_{ij}} \cdot \frac{n_{c,ij}(t)}{n_{c,i}(t)}. \quad (10)$$

3.3. Theoretical Insights

Comparisons between the proposed NTM and its predecessors offer further insights regarding model predictions of trip-completion rate. We start by comparing the proposed NTM with the original in [5, 15]. To that end, we note that

$$\frac{\bar{f}_i(n_{c,i}(t), n_{q,i}(t))}{n_{c,i}(t)} \geq \frac{f_i(n_{c,i}(t) + n_{q,i}(t))}{n_{c,i}(t) + n_{q,i}(t)}, \quad (11)$$

which indicates that an average circulating speed predicted for a certain scenario by the proposed model tends to exceed that predicted by the original one. Example speeds labeled v_1 and v_2 in Fig.3 illustrate a case in point.

General verification of (11) comes by expanding its left-side as

$$\frac{\bar{f}_i(n_{c,i}(t), n_{q,i}(t))}{n_{c,i}(t)} = f\left(\frac{n_{c,i}(t)}{1 - n_{q,i}(t)/n_m}\right) \cdot \frac{1 - n_{q,i}(t)/n_m}{n_{c,i}(t)}, \quad (12)$$

so that the expansion has the form $g(x) = f(x)/x$. The function $g(x)$ is non-increasing because

$$g'(x) = \frac{f'(x) \cdot x - f(x)}{x^2} = \frac{1}{x} \left(f'(x) - \frac{f(x) - f(0)}{x - 0} \right) \leq 0, \quad (13)$$

where the inequality in (13) holds because f is concave and $f(0) = 0$. With function g , equation (11) becomes

$$g\left(\frac{n_{c,i}(t)}{1 - n_{q,i}(t)/n_m}\right) \geq g(n_{c,i}(t) + n_{q,i}(t)). \quad (14)$$

As regards the left-side of (14), it can be shown that

$$\frac{n_{c,i}(t)}{1 - n_{q,i}(t)/n_m} = n_{c,i}(t) + n_{q,i}(t) \cdot \frac{n_{c,i}(t)}{n_m - n_{q,i}(t)}, \quad (15)$$

which cannot exceed the right-side argument of (14), $n_{c,i}(t) + n_{q,i}(t)$. Hence, (14) holds knowing that g is non-increasing. Equation (11) therefore holds as well.

Having thus verified (11), we can expand both sides of that inequality as

$$\frac{\tilde{f}_i(n_{c,i}(t), n_{q,i}(t))}{L_{ii}} \cdot \frac{n_{c,ii}(t)}{n_{c,i}(t)} \geq \frac{f_i(n_{c,i}(t) + n_{q,i}(t))}{L_{ii}} \cdot \frac{n_{c,ii}(t)}{n_{c,i}(t) + n_{q,i}(t)}, \quad (16)$$

which indicates that trip-completion rates predicted by the proposed NTM tend to exceed those predicted by the original model. As previously noted, the latter's artificial inclusion of $n_{q,i}(t)$ in circulating traffic tends to constrain trip completions.

Comparing predictions between the proposed NTM and the discriminating one in [7] tells a different story. As we discussed, [7] statically down-sizes the MFD by assuming an accumulation $n_{q_i,max}$, which corresponds to maximum-lengthed cordon queues.

It suffices to consider the case in which $n_{q_i,max} \approx 0$, $\tilde{f} = f$. We note that

$$\frac{\tilde{f}_i(n_{c,i}(t), n_{q,i}(t))}{f_i(n_{c,i}(t))} \leq 1, \quad (17)$$

as is clear from the example data points labeled “1” and “3” in Fig.3. Thus,

$$\frac{\tilde{f}_i(n_{c,i}(t), n_{q,i}(t))}{L_{ii}} \cdot \frac{n_{c,ii}(t)}{n_{c,i}(t)} \leq \frac{f_i(n_{c,i}(t))}{L_{ii}} \cdot \frac{n_{c,ii}(t)}{n_{c,i}(t)} = \frac{\tilde{f}_i(n_{c,i}(t))}{L_{ii}} \cdot \frac{n_{c,ii}(t)}{n_{c,i}(t)}, \quad (18)$$

which indicates that trip-completion rates predicted by the proposed NTM tend to be smaller than those predicted by the discriminating model if cordon queues are assumed to be short in length; e.g. if $n_{q_i,max} \approx 0$. The opposite outcomes can be expected when $n_{q_i,max}$ is assumed to be large. These points will be illustrated in sec.5.

4. Control

As in [5, 7, 15], we seek $u_{ij}^*(t)$, the optimal metering action to manage cordon crossings from i to j so as to minimize vehicle-hours traveled (VHT) on a network. And like those previous works, we do so assuming that trip-making demand, $\lambda_{ij}(t)$, is given. We discretize the entire control period into small time steps of duration τ , and obtain numerical solutions using model-predictive control with a rolling horizon of H steps. Denoting the start of each planning horizon as $t = t_0$, our objective function for each horizon takes the form

$$\min_u \sum_{1 \leq h \leq H} \sum_{i,j} n_{q,ij}(t_0 + h \cdot \tau) + n_{c,ij}(t_0 + h \cdot \tau), \quad (19)$$

where system dynamics are constrained by (8)-(10), and accumulations at t_0 are estimated as described below.

4.1. Initial System States

Denote as α the penetration of connected vehicles in the traffic mix, since their onboard systems can provide needed inputs. At the start of every planning horizon,

$$n_{c,ij}(t_0) + n_{q,ij}(t_0) = \frac{K_{ij}(t_0)}{\alpha}, \quad (20)$$

if at that t_0 there are $K_{ij}(t_0)$ connected vehicles in i observed heading for j . Assume for simplicity that $n_{q,ij}(t_0)$ need travel no distance to reach i 's boundary with j , and since the $n_{c,ij}(t_0)$ travel $L_{ij}/2$ on average, the total distance that needs to be covered in crossing that boundary is

$$n_{c,ij}(t_0) \cdot L_{ij}/2 + n_{q,ij}(t_0) \cdot 0 = \frac{\sum_{1 \leq k \leq K_{ij}(t_0)} l_k(t_0)}{\alpha}. \quad (21)$$

It is easy to show that

$$n_{c,ij}(t_0) = \frac{\sum_{1 \leq k \leq K_{ij}(t_0)} l_k(t_0)}{\alpha \cdot L_{ij}/2}, \quad (22)$$

$$n_{q,ij}(t_0) = \frac{K_{ij}(t_0)}{\alpha} - \frac{\sum_{1 \leq k \leq K_{ij}(t_0)} l_k(t_0)}{\alpha \cdot L_{ij}/2}, \quad (23)$$

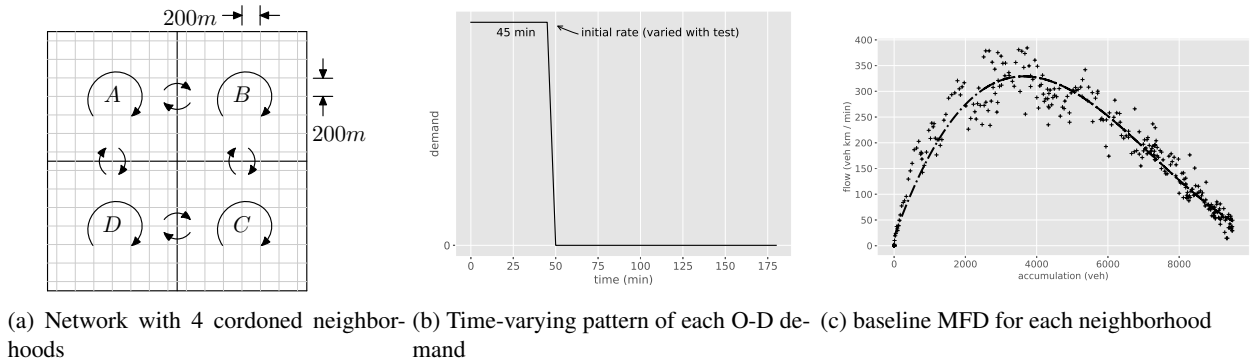


Fig. 4: Test inputs

where $l_k(t_0)$ is the distance that connected vehicle k must travel at time t_0 to reach the boundary, as measured by that vehicle.

4.2. Numerical Solution

The dimensions of an NTM system increase quadratically with the number of cordoned neighborhoods. Analytical solutions for an arbitrary number of neighborhoods therefore seem out of reach.⁸

We turned to a numerical method instead. Rather than directly solving the non-linear problem using a single-shooting method as in [5, 7, 15], we sought a more efficient approach that is scalable with problem size. Hence we chose the iterative Linear Quadratic Regulator (iLQR) method [14], along with several improvements to the regularization and line-search aspects of the algorithm developed in [18]. Details of this algorithm and brief discussion of why it suits our problem so well are furnished in Appendix B.

5. Numerical Analysis

Tests of the proposed NTM and its predecessors entailed use of the AIMSUN software [3] to simulate traffic on a signalized network of 14x14 two-way streets. The set-up was idealized to minimize details regarding inputs. Each street segment was 200m long, with 2 lanes in each travel direction. The network was cordoned into four equally-sized neighborhoods labeled A-D in Fig.4(a).

Demands for the O-Ds (shown with arrows in Fig.4(a)) each took the time-varying pattern shown in Fig.4(b).⁹ Rates varied across tests as will be described in due course, but were always set so that all trips were completed within each test's 3-h period. All (simulated) drivers received traffic updates at 5-min intervals, and responded by altering their routes as per the model in [3]. The penetration of connected vehicles, α , was set at just 5%. Traffic in each of our four identically-sized neighborhoods was described by an identical MFD for baseline conditions; i.e. conditions in the absence of cordon control. This MFD was estimated from AIMSUN simulations under a full range of spatially-uniform O-D demands, and free of cordon metering. The data points produced from those simulations are shown in Fig.4(c). A best-fit curve defining each neighborhood's baseline MFD is shown in the figure as well.

5.1. Predictive Strengths

The following illustrates how the proposed NTM can produce improved estimates of neighborhood traffic conditions. Initial demand for each O-D was set at 8,000 vph; see again Fig 4(b). To test the predictive strength of all

⁸ Recall that the analytical solution furnished in [6] was for two neighborhoods (only).

⁹ Each trip crossed no more than one cordon line to avoid the complications of modeling the driver route-choice behavior involved in touring three or more neighborhoods. The dramatic reduction in demand starting at $t = 50\text{min}$ created non-steady-state conditions, but only for a brief period.

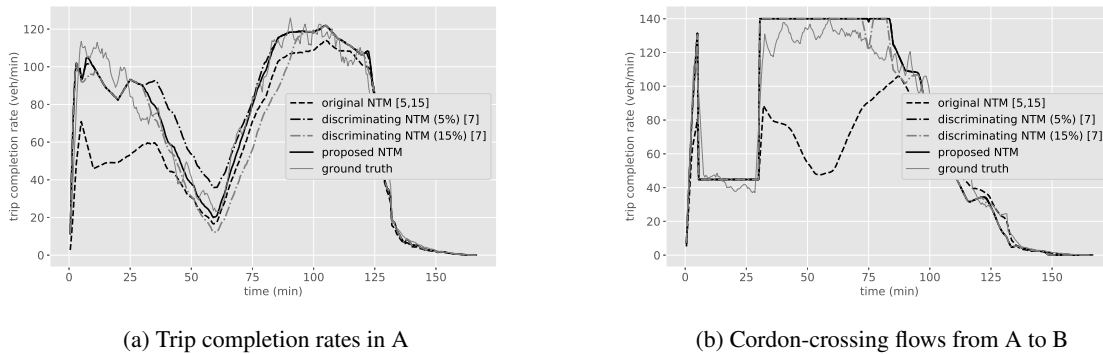


Fig. 5: Predictive strengths of different NTMs

NTMs, cordon queues were created deliberately in the early going by restrictively metering the collective crossings from neighborhood *A* to *B* at 1,500 vph. The metering persisted for 30 mins, starting at $t = 5\text{min}$ to allow for a warm-up period. No other cordon-control actions were taken.

Figure 5(a) presents trip-completion rates in *A* as predicted by: (i) the original NTM in [5, 15]; (ii) the discriminating NTM in [7] with the MFD modestly down-sized by 5% to reflect one guess for the impacts of maximum-lengthed cordon queues; (iii) the discriminating model with the MFD more dramatically down-sized by 15% to reflect another guess;¹⁰ and (iv) the proposed model.

“Ground-truth” rates generated by AIMSUN are shown as well. Visual inspection of Fig.5(a) clearly shows that predictions from the proposed NTM fit the ground-truth data best.¹¹

In contrast, the original NTM [5, 15] underestimated rates in the early going, when *A*’s cordon queues were lengthiest. The queued vehicles that were sent back into *A*’s circulating traffic seem indeed to have artificially constrained trip completions.

The discriminating NTM [7] with a 5% reduction in capacity overestimated rates toward the middle of the test, which is when neighborhood *A*’s circulating traffic entered the congested regime. The overestimates occurred because the model underestimated *A*’s capacity lost in storing cordon queues when those queues grew longest. The model did a better job of predicting trip-completion rates later in the period, when cordon queues receded.

The opposite occurred when the MFD was diminished by 15%. In that case, the discriminating NTM matched ground-truth data quite well in the early going, when cordon queues were long. When those queues later receded, the model overestimated the constraining effects on circulating traffic. It under-predicted trip-completion rates as a result.

The outcomes from the discriminating model underscore the problem with accounting for cordon-queue impacts in static fashion. Attempts to strike a balance by reducing the MFD by some intermediate amount did not improve predictions over the duration of each test; see again footnote 10.

We turn now to the time-varying rates at which vehicles crossed the cordon when traveling from *A* to *B*. Four of the curves in Fig 5(b) present predictions made (for each time step) by the proposed NTM and by its predecessors. Simulated ground-truth rates are shown by a fifth curve. Note from the latter how cross-cordon flows: fell during the 30-mins of restrictive metering; rebounded when metering was discontinued; and gradually dropped thereafter when cordon queues (and demand) diminished.

The figure shows that the ground-truth data are best matched by the proposed NTM and the discriminating model, irrespective of the MFD-reduction.¹² The virtual tie among these models is not surprising, since the discriminating and

¹⁰ The decision to re-scale the MFD down by 5% and 15% represents our best guesses at bracketing neighborhood *A*’s capacity losses due to maximum-lengthed cordon queues. Different guesses for those losses did not produce appreciable improvements in model predictions.

¹¹ Comparing predictions from the proposed model against ground truth produces a root-mean-square error of 7.7 vehicles/min. This is smaller than the RMSEs of 20.5 vehicles/min and 10.5–12.2 vehicles/min produced by the original and the discriminating NTMs, respectively.

¹² Predictions of the proposed model produced an RMSE of 11.7 vehicles/min, which is virtually the same as those from the discriminating NTM. The RMSE for the original NTM was 35.1 vehicles/min.

the proposed models were equipped with the same formulations for distinguishing under- and over-saturated metering operations; see again sec.3 and footnote 6 in particular.

The original NTM, in contrast, underestimated cross-cordon flows during the hour that followed metering's deactivation. Because the original model sent back vehicles that were queued at the cordon, they were not on the scene when opportunities came to saturate the cordon.

5.2. Control Outcomes

Our NTM's enhanced predictions lead to improved cordon-control policies. To demonstrate, demand was set so that neighborhood *A* was rich in destinations, and therefore in special need of protection via cordon metering [4]. Initial demand was therefore increased to 20,000 vph for each O-D pair bound for *A*, and was lowered to 5,000 vph for each remaining O-D; see again Fig.4(b). All inter-neighborhood trips were subject to cordon metering. As per the reasoning in footnote 4, the control variable, u_{ij} , was constrained in $[0.33, 1.0]$, which are expressed as ratios of the metering rate to the cordon's unmetered capacity. Optimal values, u_{ij}^* , were separately obtained from the proposed NTM and from its predecessors. To facilitate fair comparisons: the $u_{ij}^*(t)$ were in all cases selected using the iLQR method with a rolling horizon of $H = 20$ steps, each of duration $\tau = 5$ mins;¹³ and the resulting time-varying accumulations over the network were always generated using AIMSUN. Outcomes for the entire network are shown in Fig.6, along with a fifth curve for the do-nothing case, in which the traffic signals along the cordon lines were not retimed to function as meters. Each curve is the average of five simulations with distinct random seeds.

The proposed NTM wins. It reduced VHT on the network (the area under a curve) by nearly 15% relative to the do-nothing strategy. The original NTM reduced VHT by only 3%. The more discriminating model reduced network VHT by 6.5% when the MFD was only modestly diminished, and by a mere 1% when the MFD was down-sized more dramatically. The proposed model thus saved 12% more VHT than did the original one, and saved as much as 14% more VHT than did the discriminating NTM.

To explain these outcomes, we present the optimal metering actions imposed on vehicles traveling from *B* to *A*, $u_{BA}^*(t)$. Recall that metering was set with protecting neighborhood *A* in mind. We note too that the symmetries in our idealized set-up mean that metering actions imposed on vehicles traveling from *B* to *A* were identical to those imposed on vehicles traveling from *D* to *A*. The values selected by the proposed NTM and its predecessors are shown in Fig.7. Note that in all cases, the controller alternated between the minimum and maximum values allowed, $u_{min} = 0.33$ and $u_{max} = 1.0$. The outcome supports the theoretical analysis in Appendix A, indicating that bang-bang metering tends to be dominant over a control period.

Note by comparing Fig.7(a) with its counterparts how the original NTM [5, 15] selected u_{min} rather sparingly; i.e. u_{min} was deployed under the original model only twice, and only for brief periods during the 3-h test. By underestimating cordon-crossing rates (as we saw in Fig.5(b)), the original NTM overestimated metering's negative impacts. It therefore selected a metering policy that was quite relaxed. This is why its resulting network accumulations so closely resemble the do-nothing strategy, as evident in Fig.6.

The control policy generated by the more discriminating NTM [7] is once again a different story. This is because the model, in effect, sets aside a buffer in each neighborhood to store cordon queues at their estimated maximum lengths. This gives the discriminating NTM carte blanche to meter restrictively. Nothing is seen lost as a consequence, as long as the queueing buffer is not over-flowed. Indeed, Figures 7(b) and 7(c) show that the discriminating model generated relatively restrictive metering plans. Both figures display extended periods in which the cordon (separating neighborhood *B* from *A* in this case) was metered at u_{min} .

Note in particular how both figures display u_{min} for an extended duration toward the end of the analysis period. Cordon queues had diminished by that time. Yet the buffer remained in each neighborhood. The model thus failed to recognize that higher capacities to circulate vehicles had returned to neighborhood *A* (and elsewhere). In response, the model generated an unduly-restrictive plan to protect neighborhood *A* from congestion late in the day.

¹³ Values of H and τ were selected based on findings reported in [5]. At each control step, the iLQR method always converged within 10 iterations. Computation time for each step was about 5 seconds on an Intel i7-4770 CPU @ 3.4GHz. The algorithm was implemented in Python with Numpy and Tensorflow.

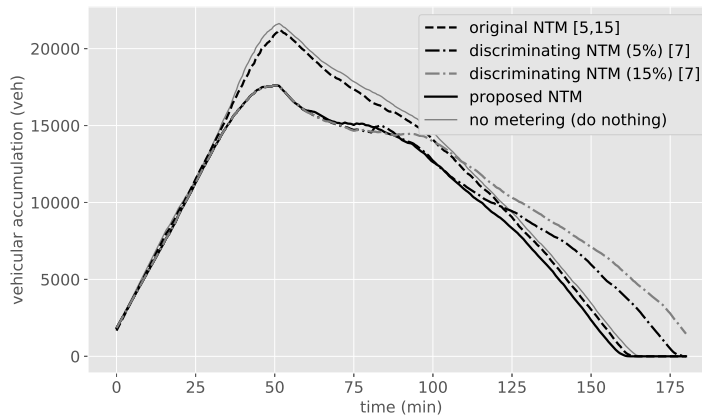
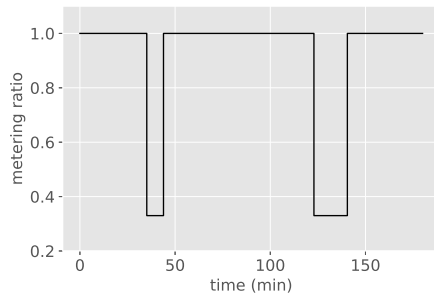
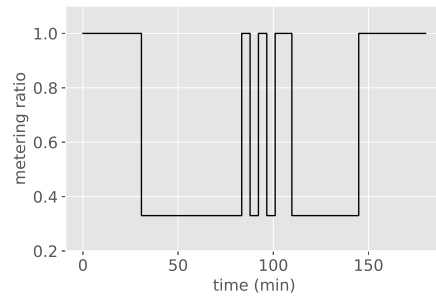


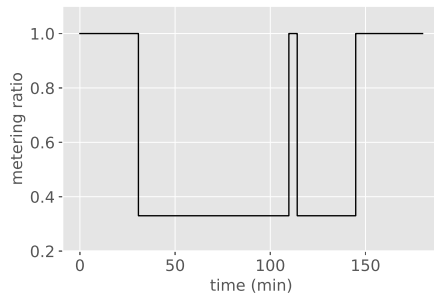
Fig. 6: Network-wide vehicle accumulation over time



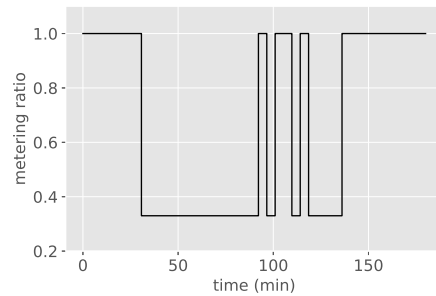
(a) Original NTM



(b) Discriminating NTM (5% MFD-reduction)



(c) Discriminating NTM (15% MFD-reduction)



(d) Proposed NTM

Fig. 7: Control actions for vehicles crossing from B to A

This was the discriminating model's downfall: congestion recovery was retarded by the late and lengthy occurrence of u_{min} . To see evidence of this, note how both dot-dashed curves in Fig.6 lie above their thin-solid (do-nothing) counterpart during the test's final hour or so.

The proposed NTM generated the metering pattern in Fig.7(d). This more effective policy is a kind-of compromise between those of its predecessors. Note, for example, how the final occurrence of u_{min} in Fig.7(d) persisted for a duration longer than what was generated by the original NTM, but shorter than what were produced by the discriminating model.

6. Summary and Future Research

Like the others of its kind, the Neighborhood Transmission Model presently proposed addresses the cordon-control problem on large geographic scales. Like its most recent predecessor [7], our NTM treats a neighborhood's circulating vehicles differently from those queued at its cordon. Unlike other NTMs, however, ours approximately accounts for the time-varying street space occupied by cordon queues, and models the attendant effects on circulating traffic in dynamic fashion. This is done by re-scaling the neighborhood's MFD with the assumption that cordon queues always exhibit jam densities.

The assumption is supported to some degree by our theoretical finding that an optimal metering strategy entails bang-singular control, which includes very restrictive metering; see Appendix A. Experimental support for this finding is furnished from simulations; see again Fig.7. The simulations further indicate that the proposed NTM does as good or better a job of predicting neighborhood traffic dynamics than do its predecessors, and that this leads to more effective cordon-metering policies. Policies were obtained by discretizing system dynamics into small time steps and applying the iLQR method to solve a rolling-horizon optimization. The method scales linearly with the number of variables (see Appendix B), and always converged in relatively few iterations (see again footnote 13).

Inputs needed at the start of each planning horizon may in the future come from connected vehicles. Our estimation methods that stemmed from this were a bit coarse, particularly the assumption that vehicles in a cordon queue need travel zero distance to reach the cordon line. Since the assumption will be subject to greater error as the queue lengths expand at cordons, future research might seek to refine the present estimates. Perhaps refinements will come by further coupling estimates with real-time measurements.

It is further worth noting that the proposed NTM implicitly assumes that inter-neighborhood-bound vehicles evenly distribute themselves across a cordon. This may be reasonable when drivers can readily adjust their routes in response to cordon queues. Still, our ongoing research on the subject indicates that network performance can often be improved by varying metering rates along a cordon line, in part to balance queue lengths there.

Finally, the proposed NTM assumes that the partitioning of neighborhoods occurs in static and *a priori* fashion. Our ongoing work in this realm suggests that there is merit in adapting cordon sizes and locations in real time, to accommodate a city's evolving congestion patterns.

Acknowledgements

This work was partially supported with funds from UCCONNECT, the University Transportation Center for federal region 9; and by the National Science Foundation.

Appendix A. “Bang-Singular” control

We show that the optimal control action, u^* , is “bang-singular” type by re-writing the discrete optimization problem of (19) in continuous fashion.¹⁴ Thus,

$$\begin{aligned} & \int_{t_0}^{t_0+H\tau} 1^T \mathbf{n} dt \\ \text{s.t. } & \dot{\mathbf{n}} = \lambda + f(\mathbf{n}) + g(\mathbf{n}, \mathbf{u}) \\ & \mathbf{u} \in [0, 1], \end{aligned}$$

where $\mathbf{n}$ is the vector of $[...n_{c,ij}...n_{q,ij}...]$; $\mathbf{u}$ is the vector of $[...u_{ij}...]$; λ is traffic demand; function f is the uncontrolled part of the system dynamics; and function g represents the cordon-crossing flows.

¹⁴ The continuous formulation is, of course, roughly equivalent to the discrete version when time-step duration, τ , is small. We note that [7] proved its optimal control solution to be bang-singular as well. The control results of [5, 15] were more continuous because extra regularization terms were added in their objective function.

Thanks to the special form of $g(\mathbf{n}, \mathbf{u})$, we can rewrite the problem as follows:

$$\begin{aligned} & \int_{t_0}^{t_0+H\tau} 1^T \mathbf{n} dt \\ \text{s.t. } & \dot{\mathbf{n}} = \lambda + f(\mathbf{n}) + A \cdot \mathbf{u} \\ & \mathbf{u} \in [u_{\min}(\mathbf{n}), u_{\max}(\mathbf{n})], \end{aligned}$$

where we transform function $g(\mathbf{n}, \mathbf{u})$ into $A \cdot \mathbf{u}$ and state-dependent control bounds $u_{\max}(n)$ and $u_{\min}(n)$.

The Hamiltonian of the system is:

$$\begin{aligned} H(\mathbf{n}, \mathbf{u}, \mathbf{p}, t) &= \mathbf{p}^T (\lambda + f(\mathbf{n}) + A \cdot \mathbf{u}) + 1^T \mathbf{n} \\ \nabla_{\mathbf{u}} H(\mathbf{n}, \mathbf{u}, \mathbf{p}, t) &= A^T \cdot \mathbf{p}, \end{aligned}$$

where $\mathbf{p}$ is the costate vector. Proof that the optimal control action, u^* , is “bang-singular” is now straightforward because our system is linear in the control variables and the objective function does not involve those variables. If the coefficient in $A^T \cdot \mathbf{p}$ is nonzero, use of Pontryagin’s maximum principle shows that the corresponding control is bang-bang. If any one of the coefficients in $A^T \cdot \mathbf{p}$ equals zero for a non-zero time interval, the corresponding control is of singular type, meaning that for a time control is not extreme. (We note for future reference that when undergoing control, the system trajectory is called a singular arc.) Now suppose $(A^T \cdot \mathbf{p})_i = 0$ for a non-zero time interval, i.e. the i -th coefficient equals zero and u_i is of singular type. Consider the costate equation

$$-\dot{\mathbf{p}}^T = \nabla_{\mathbf{n}} H(\mathbf{n}, \mathbf{u}, \mathbf{p}, t) = \mathbf{p}^T \cdot \nabla_{\mathbf{n}} f + 1^T,$$

and multiply both sides by A , such that

$$-\dot{\mathbf{p}}^T A = \mathbf{p}^T \cdot \nabla_{\mathbf{n}} f \cdot A + 1^T \cdot A.$$

Since $(\dot{\mathbf{p}}^T \cdot A)_i = 0$ and $1^T \cdot A = 0$, we have

$$(\mathbf{p}^T \cdot \nabla_{\mathbf{n}} f \cdot A)_i = 0,$$

which, together with $(\mathbf{p}^T \cdot A)_i = 0$, constrain the singular arc in a zero-measure subset of the system state space. Since our use of model-predictive control optimizes the system trajectory at each step but only applies the first control action, the observed initial system state at each time step has near-zero probability of lying exactly on these singular arcs. In the unlikely event that a state did lie on the singular arcs, measurement noise and prediction errors from the embedded dynamic model would soon cause the state to drift from singular control. Optimal control therefore tends almost always to be “bang-bang”, which explains why we did not see singular control in our numerical results.

Appendix B. Iterative LQR algorithm

The following algorithm is adopted from [14, 18]. Consider the trajectory optimization problem

$$\begin{aligned} \min_U & l_f(x_N) + \sum_{1 \leq i < N} l(x_i, u_i) \\ \text{s.t. } & x_{i+1} = f(x_i, u_i), \forall 1 \leq i < N, \end{aligned}$$

where: i is the index for time step; x is the vector of system state (vehicle accumulations in the present context); u is the control vector (metering actions); f is the system dynamic (a vector-valued traffic model); l is the cost function (VHT); and l_f is the terminal state cost. By defining $U_i = \{u_i, u_{i+1}, \dots, u_{N-1}\}$, the cost to go function J is defined as

$$J_i(x_i, U_i) = \sum_{i \leq j \leq N-1} l(x_j, u_j) + l_f(x_N).$$

The value function is defined by $V(x_i, i) = \min_{U_i} J_i(x_i, U_i)$, which has the recursive form

$$V(x_i, i) = \min_{u_i} [l(x_i, u_i) + V(f(x_i, u_i), i + 1)].$$

The Q function is defined as

$$Q(x_i, u_i) = l(x_i, u_i) + V(f(x_i, u_i), i + 1).$$

With all the above definitions, the iLQR algorithm first initializes a random control sequence U and then alternatively performs a “backward pass” and a “forward pass” to improve U until convergence.

The backward pass is done as follows:

$$\begin{aligned} Q_x &= l_x + f_x^T V'_x \\ Q_u &= l_u + f_u^T V'_x \\ Q_{xx} &= l_{xx} + f_x^T (V'_{xx} + \mu I_n) f_x \\ Q_{uu} &= l_{uu} + f_u^T (V'_{xx} + \mu I_n) f_u \\ Q_{ux} &= l_{ux} + f_u^T (V'_{xx} + \mu I_n) f_x \\ k &= -Q_{uu}^{-1} Q_u \\ K &= -Q_{uu}^{-1} Q_{ux} \\ \Delta V(i) &= \frac{1}{2} k^T Q_{uu} k + k^T Q_u \\ V_x(i) &= Q_x + K^T Q_{uu} k + K^T Q_u + Q_{ux}^T k \\ V_{xx}(i) &= Q_{xx} + K^T Q_{uu} K + K^T Q_{ux} + Q_{ux}^T K, \end{aligned}$$

where μ is a regularization factor. Note that in our application, $l_{xx}, l_{uu}, l_{ux}, l_u$ are all zeros, i.e. no direct hessian matrix computation is involved, which can save considerable computation time and makes the iLQR algorithm ideal for our application.

The forward pass to update the trajectory $(\hat{x}, \hat{y}, i)$ is:

$$\begin{aligned} \hat{u}(i) &= u(i) + \beta k(i) + K(i)(\hat{x}(i) - x(i)) \\ \hat{x}(i+1) &= f(\hat{x}(i), \hat{u}(i)) \\ \hat{x}(1) &= x(1); \end{aligned}$$

where β is the step size found by line-search. The iLQR algorithm repeats the backward and forward passes alternatively until the trajectory converges.

References

- [1] Aboudolas, K., Geroliminis, N., 2013. Perimeter and boundary flow control in multi-reservoir heterogeneous networks. *Transportation Research Part B: Methodological* 55, 265–281.
- [2] Ampountolas, K., Zheng, N., Geroliminis, N., 2017. Macroscopic modelling and robust control of bi-modal multi-region urban road networks. *Transportation Research Part B: Methodological* 104. doi:10.1016/j.trb.2017.05.007.
- [3] Casas, J., Ferrer, J.L., Garcia, D., Perarnau, J., Torday, A., 2010. Traffic simulation with aimsun, in: *Fundamentals of traffic simulation*. Springer, pp. 173–232.
- [4] Daganzo, C.F., 2007. Urban gridlock: macroscopic modeling and mitigation approaches. *Transportation Research Part B: Methodological* 41, 49–62.
- [5] Geroliminis, N., Haddad, J., Ramezani, M., 2013. Optimal perimeter control for two urban regions with macroscopic fundamental diagrams: A model predictive approach. *IEEE Transactions on Intelligent Transportation Systems* 14, 348–359.
- [6] Geroliminis, N., Skabardonis, A., 2011. Identification and analysis of queue spillovers in city street networks. *IEEE Transactions on Intelligent Transportation Systems* 12, 1107–1115.
- [7] Haddad, J., 2017. Optimal perimeter control synthesis for two urban regions with aggregate boundary queue dynamics. *Transportation Research Part B: Methodological* 96, 1–25.
- [8] Haddad, J., Geroliminis, N., 2012. On the stability of traffic perimeter control in two-region urban cities. *Transportation Research Part B: Methodological* 46, 1159–1176.
- [9] Haddad, J., Ramezani, M., Geroliminis, N., 2012. Model predictive perimeter control for urban areas with macroscopic fundamental diagrams, in: *American Control Conference (ACC)*, 2012, IEEE. pp. 5757–5762.

- [10] Jusoh, R.M., Ampountolas, K., 2017. Multi-gated perimeter flow control of transport networks, in: Control and Automation (MED), 2017 25th Mediterranean Conference on, IEEE. pp. 731–736.
- [11] Keyvan-Ekbatani, M., Carlson, R.C., Knoop, V.L., Hoogendoorn, S.P., Papageorgiou, M., 2016. Queuing under perimeter control: Analysis and control strategy, in: Intelligent Transportation Systems (ITSC), 2016 IEEE 19th International Conference on, IEEE. pp. 1502–1507.
- [12] Keyvan-Ekbatani, M., Papageorgiou, M., Knoop, V.L., 2015a. Controller design for gating traffic control in presence of time-delay in urban road networks. *Transportation Research Part C: Emerging Technologies* 59, 308–322.
- [13] Keyvan-Ekbatani, M., Yildirimoglu, M., Geroliminis, N., Papageorgiou, M., 2015b. Multiple concentric gating traffic control in large-scale urban networks. *IEEE Transactions on Intelligent Transportation Systems* 16, 2141–2154.
- [14] Li, W., Todorov, E., 2004. Iterative linear quadratic regulator design for nonlinear biological movement systems., in: ICINCO (1), pp. 222–229.
- [15] Ramezani, M., Haddad, J., Geroliminis, N., 2015a. Dynamics of heterogeneity in urban networks: aggregated traffic modeling and hierarchical control. *Transportation Research Part B: Methodological* 74, 1–19.
- [16] Ramezani, M., Haddad, J., Geroliminis, N., 2015b. Two-level hierarchical traffic control for heterogeneous urban networks, in: Control Conference (ECC), 2015 European, IEEE. pp. 3484–3489.
- [17] Sirmatel, I.I., Geroliminis, N., 2016. Model predictive control of large-scale urban networks via perimeter control and route guidance actuation, in: *IEEE Transactions on Intelligent Transportation Systems*, IEEE. pp. 6765–6770.
- [18] Tassa, Y., Erez, T., Todorov, E., 2017. Synthesis and stabilization of complex behaviors through online trajectory optimization, in: 2012 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2012), IEEE. pp. 4906–4913.
- [19] Yang, K., Zheng, N., Menendez, M., 2017. Multi-scale perimeter control approach in a connected-vehicle environment. *Transportation Research Part C: Emerging Technologies* 94. doi:10.1016/j.trc.2017.08.014.