

# Joint route guidance and demand management using generalized MFDs<sup>\*</sup>

Charalambos Menelaou<sup>\*</sup> Stelios Timotheou<sup>\*</sup>  
Panayiotis Kolios<sup>\*</sup> Christos G. Panayiotou<sup>\*</sup>

<sup>\*</sup> *KIOS Research and Innovation Center of Excellence, and the Department of Electrical and Computer Engineering, University of Cyprus, {cmene102, pkolios, timotheou.stelios, christosp}@ucy.ac.cy*

---

## Abstract:

In this work we propose a Model Predictive Control (MPC) framework that combines multi-region route guidance with demand management at a macroscopic level. While route guidance is employed to control all vehicular routes, demand management is introduced to control the flows' departure times. In effect a portion of the demand may be instructed to wait at their origin before commencing their journey (i.e., delayed departure) and thus ensure that, when vehicles do enter the network, they will travel at free-flow conditions. We show that the resulting problem is a nonlinear optimization problem that is solved by a novel convex relaxation with tight lower bounds on the optimal solution. Extensive simulations are conducted to evaluate the performance of the proposed MPC convex optimization problem indicating the substantial performance improvements in the network utilization.

*Keywords:* Demand Management; Route Guidance; Modelling, Control and Optimization of Transportation Systems;

---

## 1. INTRODUCTION

Traffic congestion is increasingly becoming the number one issue in cities with many socio-economic adverse effects. Interestingly, traffic congestion does not necessary occur due to the infrastructure's limited capacity but is a result of the absence of effective management strategies. One of the most popular approaches to resolve this issue is route guidance that aims to reduce the imbalances across the urban network Papageorgiou (1990).

Most of the scientific work on route guidance consider detail microscopic models, with complete information of the underlying network conditions (e.g., the speed and position of all vehicles) assumed to be known Papageorgiou (1990). However, microscopic models tend to be highly complex, especially in large-scale networks Daganzo (2007). To reduce this complexity, large network areas are split into smaller homogenous regions Mazlounian et al. (2010) in which their traffic dynamics defined according to the Macroscopic Fundamental Diagram (MFD) Daganzo (2007). The MFD captures the macroscopic relations between the three mobility parameters, i.e., speed, flow, and density which can then be used for region-level route guidance.

The new connected capabilities of vehicles have strengthen the applicability of such route guidance methods. Cur-

rently, cars are already equipped with onboard units that inform drivers about real-time trip information while also suggesting alternative less congested routes to follow. Despite these great features, congestion remains because most of route guidance approaches aim only to optimize the *benefit of individuals*, as opposed to the *benefit of the whole* Macfarlane (2019). Nonetheless, literature indicates that routing solutions that intend to improve the *system optimum* can substantially reduce travel times and congestion, with a slight decrease in traffic demand Macfarlane (2019). To achieve this, drivers' route and departure time choices should be manipulated appropriately (i.e., distributing traffic demand in space-time) through the integration of route guidance with a demand management method for Commuter Transportation (2004).

In this work we employ a Model Predictive Control (MPC) framework to optimize the control decisions for route guidance and demand management taking into account future implications that are estimated utilizing the MFD model. MPC solutions utilizing the MFD characteristics have been considered in the past including Geroliminis et al. (2013) that employ a nonlinear MPC framework to control a free-way system and a two-region urban network, respectively.

The most of the aforementioned MPC approaches rely on nonlinear MFD models, resulting in nonconvex optimization problems with solutions that yield suboptimal results since nonconvex solvers may get stuck in local optima points. The work and Menelaou et al. (2019) tries to address this issue by approximating the nonconvex MPC problem with a Mixed Integer Linear Program (MILP). The work in Kouvelas et al. (2019) proposes, instead, an

---

<sup>\*</sup> This work has been supported by the European Union's Horizon 2020 research and innovation programme under grant agreement No 739551 (KIOS CoE), the Government of the Republic of Cyprus through the Directorate General for European Programmes, Coordination and Development and through the Research Promotion Foundation (Project: CULTURE/BR-NE/0517/14).

Extended Kalman Filter framework to provide traffic state estimates to the MPC scheme, transforming the nonlinear problem into a linear parameter-varying model.

To summarize, this work develops an MPC framework to jointly solve the multi-regional joint route guidance and demand management problem utilizing a generalized MFD. Demand management is used to regulate flow departure times (i.e., a portion of requesting flows may be delayed at their origin) while route guidance seeks to find the appropriate paths for the flows to follow in order to increase the network's efficiency. Given the origin and destination pair of requesting flows, an MPC framework is employed to find the best departure time and the best alternative route to follow by each requesting flow to minimize for all vehicles the total time spent in the network. Note that the entire time spend accounts for both the waiting time outside the network (i.e., delayed departure) and the travel time in the network. Hereafter, our main contributions are as follows:

- The formulation of an MPC framework that jointly optimizes route guidance and demand management assuming macroscopic traffic dynamics that follow a generalized MFD.
- Derivation of an efficient lower bound solution to the problem that can be achieved by relaxing all the nonconvex constraints. The proposed relaxation provides a tight lower bound on the optimal solution.
- For those cases that the lower bound solution is infeasible, we proposed two algorithmic methods demonstrating how a good quality feasible solution can be achieved by properly manipulating of control inputs.

The rest of this paper is organized as follows. Section 2 presents the multi-regional system model for joint route guidance and demand management and Section 3 derives the mathematical formulation of the MPC framework. Section 4.1 relaxes the problem to a Quadratically Constrained Program (QCP), and then utilizes the derived control inputs to propose two methods that guarantee feasibility of the computed solutions. Furthermore, Section 5 includes simulation results demonstrating that by introducing demand management the ordinary route guidance method is improved significantly. Finally, Section 6 concludes this work and discusses future research directions.

## 2. SYSTEM MODEL

### 2.1 Traffic Flow Model

An urban area is partitioned into  $|\mathcal{R}|$  homogeneous<sup>1</sup> regions denoted by parameter  $r \in \mathcal{R} = \{1, \dots, |\mathcal{R}|\}$ . In this work we assume that the traffic dynamics within each partitioned region are modelled according to macroscopic traffic relations, Immers and Logghe (2003), that follow a generalized MFD shape that can be approximated by an asymmetric unimodal curve, Knoop and Hoogendoorn (2013). Using the MFD we can outline all important parameters of region  $r$  such as the *jam density*  $\rho_r^J$ , as well as the critical density  $\rho_r^C$  and *capacity*  $q_r^C$  where the region operates at its maximum outflow, Immers and Logghe (2003). The flow-density MFD relationship is complemented by the fundamental relationship that the

*intended outflow* of region  $k$ ,  $q_r(\rho_r(k))$  (veh/h), is equal to the product of density  $\rho_r(k)$  (veh/km) and speed  $v_r(\rho_r(k))$  (km/h) at discrete time-step  $k$ , i.e.,

$$q_r(\rho_r(k)) = \rho_r(k)v_r(\rho_r(k)). \quad (1)$$

The term “intended outflow”,  $q_r(\rho_r(k))$ , indicates the total flow that  $r$  is ready to transfer to its neighbours and/or to the outside world when there are no flow/storage capacity restrictions from neighbouring regions. It has been empirically observed that  $\rho_r^C \leq \rho_r^J/2$  and that  $q_r(\rho_r(k))$  is well-approximated using a third-order polynomial of density, Geroliminis and Daganzo (2008)

$$q_r(\rho_r(k)) = a_{r_1}\rho_r(k)^3 + a_{r_2}\rho_r(k)^2 + a_{r_3}\rho_r(k) \quad (2)$$

where,  $a_{r_1}$ ,  $a_{r_2}$  and  $a_{r_3}$  are constant calibration parameters. It follows from (1) that the regional speed,  $v_r(\rho_r(k))$ , is related to density through a second order polynomial, i.e.,

$$v_r(\rho_r(k)) = a_{r_1}\rho_r(k)^2 + a_{r_2}\rho_r(k) + a_{r_3}. \quad (3)$$

Let  $\mathcal{O} \subseteq \mathcal{R}$  and  $\mathcal{D} \subseteq \mathcal{R}$  denote the sets of origin and destination regions of different flows, respectively. Also, let  $\mathcal{J}_r^- \subseteq \mathcal{R}$  denote the set of neighbouring regions directly accessible from region  $r \in \mathcal{R}$ , and similarly let  $\mathcal{J}_r^+ = \mathcal{J}_r^- \cup \{r\}$ , such that

$$\mathcal{J}_r = \begin{cases} \mathcal{J}_r^+, & \text{if } r \in \mathcal{D} \\ \mathcal{J}_r^-, & \text{otherwise.} \end{cases} \quad (4)$$

Furthermore, let  $d_{od}(k)$  (veh) denote the *new external demand* which determines the number of new vehicles requesting to enter into the network from  $o \in \mathcal{O}$  to  $d \in \mathcal{D}$  at time-step  $k$ . Similarly, let  $\tilde{d}_{od}(k)$  (veh) denote the *admitted external demand* that determines the number of vehicles that actually enter into the network from region  $o \in \mathcal{O}$  towards  $d \in \mathcal{D}$  at time-step  $k$ . Variable  $\tilde{d}_{od}(k)$  is restricted by three factors:

- (1) The physical ability of the region to accommodate more vehicles.
- (2) The maximum possible demand that can physically enter region  $o$  during time-step  $k$  denoted by  $D_{od}^{MAX}$ .
- (3) The proposed demand management approach that may only allow a portion of the requesting demand to enter into the network; in this case the remaining vehicles will wait at their origins (outside the network) until their admission time.

The *cumulative external demand*,  $D_{od}(k)$  (veh), is used to keep track of the remaining demand that remains to be served at time-step  $k$ , such that

$$D_{od}(k+1) = D_{od}(k) - \tilde{d}_{od}(k) + d_{od}(k), k = 1, 2, \dots, \quad (5)$$

where,  $D_{od}(1) = 0$ .

To keep track of the portion of traffic destined for different regions, we further introduce variables  $\rho_{rd}(k)$  which denote the density in region  $r \in \mathcal{R}$  that is destined to  $d \in \mathcal{D}$ . Clearly, it is true that

$$\rho_r(k) = \sum_{d \in \mathcal{D}} \rho_{rd}(k). \quad (6)$$

Similarly, variables  $q_{rd}(k)$  and  $q_{rjd}(k)$  denote the *intended transfer flow* from  $r \in \mathcal{R}$  to  $d \in \mathcal{D}$ , and from  $r \in \mathcal{R}$  to  $d \in \mathcal{D}$  through neighbouring region  $j \in \mathcal{J}_r$  respectively, defined as

<sup>1</sup> Homogeneity is defined according to the homogeneous distribution of accumulated traffic as proposed in Mazlounian et al. (2010).

$$q_{rd}(k) = v_r(\rho_r(k))\rho_{rd}(k), \quad (7)$$

$$q_r(\rho_r(k)) = \sum_{d \in \mathcal{D}} \sum_{j \in \mathcal{J}_r} q_{rjd}(k). \quad (8)$$

Note that in the case that  $\{r = j = d\} \in \mathcal{D}$ , the variables  $q_{ddd}(k)$  determine the number of vehicles reach their destination and hence exit the network.

As mentioned above, the intended transfer flow of a region  $r \in \mathcal{R}$  is restricted by the flow/storage capacity of its neighbouring regions  $j \in \mathcal{J}_r^-$  with variable  $C_{rj}(\rho_j(k))$ , denotes the inter-boundary capacity of region  $r$  with region  $j$ . The inter-boundary capacity specifies the maximum flow that can be exchanged between the two neighbouring regions, for a specific value of  $\rho_j(k)$ . According to Sirmatel and Geroliminis (2019),  $C_{rj}(\rho_j(k))$  can be modelled as

$$C_{rj}(\rho_j(k)) = \begin{cases} C_{rj}^{\text{MAX}}, & \text{if } \rho_j(k) \leq \alpha \rho_j^J, \\ \frac{C_{rj}^{\text{MAX}}}{1 - \alpha} \left(1 - \frac{\rho_j(k)}{\rho_j^J}\right), & \text{otherwise,} \end{cases} \quad (9)$$

where  $C_{rj}^{\text{MAX}}$  is the maximum inter-boundary capacity and  $\alpha \rho_j^J$  is the point where the inter-boundary capacity starts to decrease with  $0 < \alpha < 1$ . Therefore, the intended transfer flow is restricted from the volume of density in region  $r \in \mathcal{R}$ , while the transfer flow of neighboring region  $j$  is analogous to its remaining storage capacity which also depends on the transfer flows from other regions  $s \in \{\mathcal{J}_j - r\}$ . Hence, the *actual transfer flow* from  $r \in \mathcal{R}$  to  $j \in \mathcal{J}_r$ , denoted by variable  $\tilde{q}_{rjd}(k)$ , is defined as

$$\tilde{q}_{rjd}(k) = \min \left( q_{rjd}(k), C_{rj}(\rho_j(k)) \frac{q_{rjd}(k)}{\sum_{y \in \mathcal{D}} q_{rjy}(k)} \right). \quad (10)$$

Taking the above into account, the dynamics of traffic density aiming to move from region  $r \in \mathcal{R}$  to region  $d \in \mathcal{D}$  are defined as

$$\rho_{rd}(k+1) = \rho_{rd}(k) + \frac{1}{L_r} \tilde{d}_{rd}(k) + \frac{T_s}{L_r} \sum_{j \in \mathcal{J}_r} (\tilde{q}_{jrd}(k) - \tilde{q}_{rjd}(k)) \quad (11)$$

where  $L_r$  and  $T_s$  denote the average distance travelled (km) by each vehicle inside region  $r$  and the duration of the discrete simulation time -step, respectively. Similar to Sirmatel and Geroliminis (2019), in this work we have assumed that  $L_r$  is independent of the origin-destination flow pair and the route choices of drivers.

### 3. PROBLEM DESCRIPTION

In this section, we employ the traffic dynamics described in Section 2 to formulate an optimization problem that aims to optimize the performance of the traffic network under joint route guidance and demand management. *Route guidance* splits each origin-destination traffic flow into distinct paths, while *demand management* regulates the entry of different flows into the network to optimize performance. The optimal performance is defined in terms of the minimization of the *Total Time Spent* (TTS), that includes both the *Total Travel Time* (TTT) of all vehicles inside the network and the *Total Waiting Time* (TWT) outside the network from the request to the actual time of entry of vehicles.

#### 3.1 Objective function

In an effort to define our objective function, we introduce variables  $S^a(k)$  and  $S^b(k)$  representing the cumulative number of vehicles that request to enter the network and successfully arrive at their destination, respectively, as follows

$$S^a(k+1) = S^a(k) + \sum_{o \in \mathcal{O}} \sum_{d \in \mathcal{D}} d_{od}(k), \quad (12)$$

$$S^b(k+1) = S^b(k) + T_s \sum_{d \in \mathcal{D}} q_{ddd}(k), \quad (13)$$

for  $k = 1, 2, \dots$ , where  $S^a(1) = 0$  and  $S^b(1) = 0$ .

Summing over all time-steps, yields the TTS in the network of all vehicles  $J_{TTS}$  (veh-h)

$$J_{TTS} = T_s \sum_k (S^a(k) - S^b(k)). \quad (14)$$

Note that the TTS includes both the TWT and the TTT (TTS=TTT+TWT).

#### 3.2 Model Predictive Control Formulation

To formulate the considered optimization problem we follow a Model Predictive Control approach with the control time-step equal to the simulation time-step such that any control action can be taken every discrete time-step  $T_s$ . The control and prediction horizons are both equal to  $N_p$ , while a new MPC problem is solved every  $m$  time-steps. Hence, we solve the  $l$ -th MPC problem,  $l = 1, 2, \dots$ , for the time horizon  $\mathcal{K}_l = \{m(l-1) + 1, \dots, m(l-1) + N_p\}$  and apply to the traffic network the control actions corresponding to time-steps  $\{m(l-1) + 1, \dots, ml\}$ . The control variables of the MPC problem are the intended transfer flows,  $q_{rjd}(k)$  (route guidance), and admitted external flows  $\tilde{d}_{od}(k)$  (demand management) to minimize the *TTS* as follows:

$$(P_1) \quad \min J_{TTS}^{MPC}(l) = T_s \sum_{k \in \mathcal{K}_l} (S^a(k) - S^b(k)) \quad (15a)$$

s.t. Traffic dynamics: (2) – (13),

$$\tilde{d}_{od}(k) \leq D_{od}^{\text{MAX}}, \quad k \in \mathcal{K}_l, o \in \mathcal{O}, d \in \mathcal{D}, \quad (15b)$$

$$\tilde{d}_{od}(k) \leq D_{od}(k), \quad k \in \mathcal{K}_l, o \in \mathcal{O}, d \in \mathcal{D}, \quad (15c)$$

$$0 \leq \rho_r(k) \leq \rho_r^J, \quad k \in \mathcal{K}_l, r \in \mathcal{R}, \quad (15d)$$

Variables:  $\rho_r(k), \rho_{rd}(k), \tilde{d}_{od}(k), D_{rd}(k), q_r(k), q_{rd}(k), q_{rjd}(k), \tilde{q}_{rjd}(k), v_r(k), S^a(k), S^b(k)$ .

Problem  $P_1$  is an MPC optimization problem where constraints (2) – (13) define the traffic dynamics modelled according to a generalized shape MFD. Constraints (15b) and (15c) define the physical constraints of the external demand inflows ensuring that it is always smaller than the maximum possible external inflow,  $D_{od}^{\text{MAX}}$ , and the total external demand,  $D_{od}(k)$ . Similarly, constraint (15d) ensures that the density of each region is within physical limits. Problem  $P_1$  is a nonconvex Nonlinear Program (NLP) due to presence of the third order polynomial of Eq. (2), the bilinear term in Eq. (7) and nonlinear functions in Eq. (9) and Eq. (10).

#### 4. SOLUTION APPROACH

In this section, we first reformulate Problem  $P_1$  to obtain a lower bound solution (Section 4.1) and then utilize the derived control inputs for the development of two solution approaches. The *Flow Normalization Route Guidance and Demand Management* (FN-RGDM) normalizes the intended transfer flows from the lower bound solution to derive flow ratios that should be transferred from one region to its neighboring regions. The *Iterative Bounding Box Route Guidance and Demand Management* (IBB-RGDM) is an iterative method that produces bounding box constraints for the traffic densities and speeds and solves a convex optimization problem with tighter relaxations in each iteration.

##### 4.1 Lower Bound Solution to Problem $P_1$

In this section we present how problem  $P_1$  can be relaxed into a Quadratically Constrained Program, and hence solved to optimality using standard mathematical programming solvers. The developed formulation relaxes all nonconvex constraints with constraints whose domains are supersets of the corresponding nonconvex constraint domains. As a result, the obtained solution from this formulation yields lower bounds to the optimal objective value and hence, the particular formulation can be used to derive the optimality gap of any developed solution approach. Next, we derive superset convex constraints for the four nonconvex constraints of Problem  $P_1$ , namely, (2), (7), (9) and (10).

To convexify Eq. (2), we construct a convex envelop of the generalized MFD diagram using piecewise linear segments. Towards this direction, we define a set of affine functions of the form,  $a_n \rho_r(k) + b_n$ ,  $r \in \mathcal{R}$ ,  $n \in \mathcal{N}_r$  such that

$$q_r(\rho_r(k)) \leq a_n \rho_r(k) + b_n, \forall n \in \mathcal{N}_r, \quad (16)$$

where  $\mathcal{N}_r = \{1, \dots, N_r\}$  is the set of piecewise linear segments defined for the approximation of Eq. (2) in region  $r$ . Because we are interested in obtaining a lower bound solution, all affine functions should be selected to lie above the curvature of  $q_r(\rho_r(k))$ .

To convexify Eq. (3), we substitute the equality sign “=” with an inequality sign, either “ $\geq$ ” when  $a_{r_1} > 0$  or “ $\leq$ ” when  $a_{r_1} < 0$ . Usually  $a_{r_1} > 0$ , yielding the convex relaxation

$$v_r(\rho_r(k)) \geq a_{r_1} \rho_r(k)^2 + a_{r_2} \rho_r(k) + a_{r_3}. \quad (17)$$

Let us now consider constraint (7) which involves the product of two variables  $\rho_{rd}(k)$  and  $v_r(\rho_r(k))$ . As this constraint has a bilinear term, we use the McCormick method, McCormick (1976), to derive convex envelopes to these constraints using the lower and upper bounds of the two variables. In this case, we can use the physical limits of the variables, i.e.,  $0 \leq \rho_{rd}(k) \leq \rho_r^J$  and  $0 \leq v_r(\rho_r(k)) \leq v_r^f$ , where  $v_r^f$  denotes the free-flow speed  $v_r^f = v_r(0) = a_{r_3}$ , to obtain the following set of inequalities:

$$q_{rd}(k) \geq 0, \quad (18)$$

$$q_{rd}(k) \geq v_r(\rho_r(k)) \rho_r^J + \rho_r(k) v_r^f - v_r^f \rho_r^J, \quad (19)$$

$$q_{rd}(k) \leq v_r(\rho_r(k)) \rho_r^J, \quad (20)$$

$$q_{rd}(k) \leq \rho_r(k) v_r^f. \quad (21)$$

Eqs. (18) and (19) are referred as *underestimators*, while Eqs. (20) and (21) as *overestimators* of Eq. (7). The above linear inequalities construct a convex envelop to the original equality  $q_{rd}(k) = \rho_{rd}(k) v_r(\rho_r(k))$ , called McCormick envelope, that is a superset of the nonconvex feasibility domain of (7).

Finally, constraints (9) and (10) are handled together. It is true that constraint (10) is the minimum of two functions and thus it can be relaxed by substituting the equality sign “=” with the inequality sign “ $\leq$ ” yielding:

$$\tilde{q}_{rjd}(k) \leq q_{rjd}(k), \quad (22)$$

$$\tilde{q}_{rjd}(k) \leq C_{rj}(\rho_j(k)) \frac{q_{rjd}(k)}{\sum_{y \in \mathcal{D}} q_{rjy}(k)}. \quad (23)$$

Observing the two new constraints (22) and (23), it can be observed that constraint (22) is linear, while constraint (23) is nonconvex; hence, further relaxation is required. Taking the sum over all  $\tilde{q}_{rjd}(k)$  for  $d \in \mathcal{D}$  in constraint (23) yields

$$\sum_{d \in \mathcal{D}} \tilde{q}_{rjd}(k) \leq C_{rj}(\rho_j(k)), \quad (24)$$

which is a relaxed version of (23). This stems from the fact that individual constraints are always at least as tight as the sum of the associated constraints. Although convex, constraint (24) is still nonlinear due to presence of the min operator in  $C_{rj}(\rho_j(k))$ . It can be easily verified that constraint (24) is equivalent to

$$\sum_{d \in \mathcal{D}} \tilde{q}_{rjd}(k) \leq C_{rj}^{MAX}, \quad (25)$$

$$\sum_{d \in \mathcal{D}} \tilde{q}_{rjd}(k) \leq \frac{C_{rj}^{MAX}}{1 - \alpha} \left(1 - \frac{\rho_j(k)}{\rho_j^J}\right), \quad (26)$$

for all  $k \in \mathcal{K}_l$ ,  $r \in \mathcal{R}$ ,  $j \in \mathcal{J}_r$ .

Therefore, eqs. (9) and (10) are relaxed into the linear constraints (22), (25) and (26).

In sum, Problem  $P_1$  can be relaxed into:

$$\min J_{TTS}^{MPC}(l) = T_s \sum_{k \in \mathcal{K}_l} (S^a(k) - S^b(k)) \quad (27)$$

s.t. Constraints: (1), (4) – (6), (8), (11) – (13), (16) – (21), (22), (25) – (26).

Variables:  $\rho_r(k)$ ,  $\rho_{rd}(k)$ ,  $\tilde{d}_{od}(k)$ ,  $D_{rd}(k)$ ,  $q_r(k)$ ,  $q_{rd}(k)$ ,  $q_{rjd}(k)$ ,  $\tilde{q}_{rjd}(k)$ ,  $v_r(k)$ ,  $S^a(k)$ ,  $S^b(k)$ .

Formulation (27) is a Quadratically Constrained Program (QCP) that provides a lower bound to the optimal objective value which can be used to assess the optimality gap of any solution approach for Problem  $P_1$ . Although, formulation (27) may lead to infeasible solutions due to possible non-satisfaction of the relaxed constraints, feasible solutions can be obtained through the two developed approaches (FN-RGDM and IBB-RGDM) implemented within the MPC framework that follows.

##### 4.2 General MPC framework

In this work the considered physical plant is described according to the non-linear problem ( $P_1$ ) whereas, every  $m$  time-steps the external demands for the prediction horizon and the current state of the network are input into the

MPC controller which computes the Lower Bound values (according to problem 27) for the control variables (i.e., the admitted demands and intended transfer flows) for the entire prediction horizon. It should be emphasized here that the intended transfer flows  $q_{rjd}$  are not used directly to the physical plant, but are converted into *split ratios* (i.e.,  $a_{rjd}(k) = q_{rjd}(k) / \sum_{d \in \mathcal{D}} q_{rjd}(k)$ ) which denote the percentage of the actual transfer flow to each of the destination regions. Unfortunately even if ratios are used in the physical plant, the solution still may not be feasible due to the modelling differences of the actual and relaxed model. More specifically, the actual capacity accounted for different destinations in each region of the network may not be equal to what is assumed from the relaxed models in the lower bound solution. To address this, either FN-RGDM or IBB-RGDM methods are employed to ensure feasible control inputs. In doing so, flow rations ( $a_{rjd}^{nor}(k) q_{rjd}(k)$ ) are used to normalize the intended transfer flows of each region and are used as control inputs to the network. The control inputs are then used to update the state of the the physical plant for the next  $m$  time-steps using the nonlinear multi-regional model dynamics described in Section 2. The procedure is then repeated until the end of the simulation.

#### 4.3 Flow Normalization Route Guidance and Demand Management

The FN-RGDM method is responsible to find the normalized split ratios that will result in feasible transfer flows across neighbouring regions. In this method, split ratios are normalized as follows:

$$a_{rjd}^{nor}(k) = a_{rjd}(k) \frac{q_{rd}^{LB}(k)}{q_{rd}^{PS}(k)} \quad (28)$$

where  $q_{rd}^{LB}(k)$  and  $q_{rd}^{PS}(k)$  denote the portion of intended outflows obtained from the lower bound solution and the actual ones that are measured from the network, respectively. In this way, split ratios are changed proportionally to the actual available capacity of each particular region.

#### 4.4 Iterative Bounding Box Route Guidance and Demand Management

The IBB-RGDM is an iterative method that is based on three repetitive steps. On each iteration, the method temporarily simulates the physical plant to get future estimates about each region's density and speed, that in turn are used to create tight box constraints for traffic. In the sequel, problem (27) is linearised based on the derived box constraints. More specifically, at the first step the solution of (27) is simulated (using the nonlinear MFD models) for the whole prediction horizon. The estimated values of speed and density (i.e.,  $u_r^S(\rho_r(k))$  and  $\rho_r^S(k)$ ) are then used to derive tighter bounds in the linearised program (27) by replacing constraint Eq. (17) with following inequalities:

$$C_1 \rho_r^S(k) \leq \rho_r(k) \leq C_2 \rho_r^S(k), \quad (29)$$

$$C_3 u_r^S(\rho_r(k)) \leq v_r(\rho_r(k)) \leq C_4 u_r^S(\rho_r(k)), \quad (30)$$

where,  $C_i$  are constants such that  $C_1, C_3 \in [0, 1]$ ,  $C_2 \geq 1$  and  $C_4 \geq 1$ . By doing so, program (27) is replaced by program (31) where in each algorithm iteration tighter bounds are provided. The third step updates  $a_{rjd}(k)$  before

the procedure is repeated until the desired convergence is achieved. Note that this iterative algorithm can terminate by applying a gradient based methodology on the convergence or by setting a predefined number of iterations, as is done in the simulation results of this work. Finally the updated  $a_{rjd}(k)$  are normalized using (28) before control inputs are returned to the plant.

$$\min J_{TTS}^{MPC}(l) = T_s \sum_{k \in \mathcal{K}_l} (S^a(k) - S^b(k)) \quad (31)$$

s.t. Constraints: (1), (4) – (6), (8), (11) – (13), (16), (18) – (21), (22), (25) – (26) and (29) – (30).

Variables:  $\rho_r(k)$ ,  $\rho_{rd}(k)$ ,  $\tilde{d}_{od}(k)$ ,  $D_{rd}(k)$ ,  $q_r(k)$ ,  $q_{rd}(k)$ ,  $q_{rjd}(k)$ ,  $\tilde{q}_{rjd}(k)$ ,  $v_r(k)$ ,  $S^a(k)$ ,  $S^b(k)$ .

## 5. SIMULATION RESULTS

Hereafter, an extensive performance evaluation is conducted for the proposed methodologies considering a Manhattan-style network in which 4 regions are considered as origins and 4 as destinations. The network consists of 16 regions each one assumed to have identical MFDs with the following parameters:  $a_{r_1} = 8/1225$ ,  $a_{r_2} = -1992/735$ ,  $a_{r_3} = 14768/147$ ,  $\rho_r^C = 43$  (veh/km),  $\rho_r^J = 118$  (veh/km),  $L_r = 1$  (km) and  $q_r^C = 1850$  (veh/h).

Furthermore, the inter-boundary capacity constraint presented in Eq. (9) and (10) is considered, with  $C_{rj}^{MAX} = 2000$  (veh/h) and  $a = 0.25$ . The prediction and control horizons are set to  $mN_p = 20$  and  $m = 1$ , respectively, and a simulation sample time  $T_s = 30s$  is used. For comparison, the performance of the following schemes is examined:

- **SP**: In this scheme all vehicles follow the shortest distance path from their origin to their destination,
- **FN-RGDM**: The Flow Normalization Route Guidance and Demand Management as presented in Section 4.3 is employed,
- **IBB-RGDM**: The Iterative Bounding Box Route Guidance and Demand Management as presented in Section 4.4 is employed.

where the Gurobi solver Gurobi Optimization Inc. (2016) is used to solve all related optimization problems. Finally, all three schemes are evaluated for the following 3 demand scenarios: (i) *light* with average demand around 1000 veh/h, (ii) *moderate* with average demand around 3000 veh/h and (iii) *heavy* with average demand around 6000 veh/h. The demand loading lasts for one hour and varies for different O-D pairs. In this work we assume that the compliance rate of drivers is equal to 100%.

### 5.1 Performance evaluation

On Table 5.1 we present the performance results of the three schemes in terms of the Average Time Spent (ATS), the Average Travel Time (ATT), and the Average Waiting Time (AWT) at the origin (waiting occurs outside the network). Looking at these results, it is clear that the SP scheme leads to very large travel times for the high demand scenarios while the proposed schemes yield good performance for all loading scenarios, with their performance only slightly affected with increasing demand. Note that although waiting at the origin is not explicitly imposed in the SP scheme, waiting occurs implicitly for vehicles that want to enter a region that is in a gridlock.

		Demand Level		
		Light	Moderate	Heavy
ATS (min)	SP	2.73	198.38	670.01
	FN-RGDM	2.68	3.69	5.66
	IBB-RGDM	2.64	3.38	4.78
ATT (min)	SP	2.73	68.5	93.56
	FN-RGDM	2.68	3.49	4.33
	IBB-RGDM	2.64	3.23	3.69
AWT (min)	SP	0	127.3	573.26
	FN-RGDM	0	0.2	1.32
	IBB-RGDM	0	0.14	1.09

Table 1. Performance evaluation of different solution approaches for different demand levels.

Scenario Number	Average Demand	Optimality Gap	
		FN-RGDM	IBB-RGDM
1	1000 veh/h	4.25%	2.12 %
2	3000 veh/h	9.81 %	3.65%
3	4000 veh/h	14.94%	4.91%
4	5100 veh/h	18.22%	4.96%
5	6000 veh/h	25.80 %	5.71 %

Table 2. The optimality gap of FN-RGDM and IBB-RGDM compared to the lower bound.

### 5.2 Optimality Gap

To examine the optimality gap between the Lower Bound (LB) and both upper bound methods (i.e., FN-RGDM and IBB-RGDM) we have evaluated the optimal objective value of the LB obtained using formulation (27) with the problem solved once for the entire time horizon, i.e.,  $\mathcal{K} = \{1, \dots, T + N_p\}$ . For the FN-RGDM and IBB-RGDM, we consider  $m = 1$  and  $N_p = 20$  time-steps, similar to the results presented in the previous experiments. The optimality criterion of choice is the *optimality gap* defined as follows:

$$\text{Optimality Gap} = \frac{J_{\text{TTS}}^{\text{Alg}} - J_{\text{TTS}}^{\text{LB}}}{J_{\text{TTS}}^{\text{LB}}} \times 100\%,$$

where  $J_{\text{TTS}}^{\text{LB}}$  and  $J_{\text{TTS}}^{\text{Alg}}$ ,  $\text{Alg} = \{\text{FN-RGDM}, \text{IBB-RGDM}\}$ , denote the TTS values, according to Eq. (14), obtained from the LB solution and the FN-RGDM and IBB-RGDM solutions, respectively. Table 2 illustrates the optimality gap of the FN-RGDM and IBB-RGDM schemes for six demand scenarios of increasing average value for the same simulation time-step and duration (i.e.,  $T_s = 30$  s and  $T = 120$  min). From the results, we can observe that the IBB-RGDM method can provide results closer to optimality in all the cases considered. This is a very important result which highlights the fact that IBB-RGDM obtains tight bounds near optimality. Furthermore, again it is clear that IBB-RGDM outperforms FN-LB as the performance of the latter method tends to drop for increasing congestion levels.

## 6. CONCLUSIONS

This work investigate a novel multi-region route guidance approach with demand management using a generalized MFD model. The resulting framework is a nonlinear non-convex problem that is solved efficiently by implementing a convex relaxation solved using QCP optimization. The proposed methodology can successfully eliminate congested conditions by managing the departure times of

vehicle flows (i.e., vehicular flows may held at their origin before instructed to commence their journeys).

Future research will include a detailed comparison of the proposed methodology with other state-of-the-art approaches (e.g., ordinary route guidance and perimeter control) while we will also examining adaptations of state-of-the-art solutions using our proposed convex relaxation.

## REFERENCES

- Daganzo, C.F. (2007). Urban gridlock: macroscopic modeling and mitigation approaches. *Transportation Research Part B: Methodological*, 41(1), 49–62.
- for Commuter Transportation, T.A. (2004). Mitigating Traffic Congestion: The role of demand-side strategies. Technical report, U.S. Department of Transport, Federal Highway Administration.
- Geroliminis, N. and Daganzo, C.F. (2008). Existence of urban-scale macroscopic fundamental diagrams: Some experimental findings. *Transportation Research Part B: Methodological*, 42(9), 759–770.
- Geroliminis, N., Haddad, J., and 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(1), 348–359.
- Gurobi Optimization Inc. (2016). Gurobi Optimizer Reference Manual. URL <http://www.gurobi.com>.
- Immers, L. and Logghe, S. (2003). *Traffic flow theory*. Department of Civil engineering Section Traffic and Infrastructure, Belgium. Course H 111.
- Knoop, V.L. and Hoogendoorn, S.P. (2013). Empirics of a generalized macroscopic fundamental diagram for urban freeways. *Transportation Research Record*, 2391(1), 133–141. doi:10.3141/2391-13.
- Kouvelas, A., Saeedmanesh, M., and Geroliminis, N. (2019). Linear-parameter-varying approximation of nonlinear dynamics for model predictive flow control of urban multi-region systems. In *58th IEEE Conference on Decision and Control (CDC 2019)*.
- Macfarlane, J. (2019). When apps rule the road: The proliferation of navigation apps is causing traffic chaos. it’s time to restore order. *IEEE Spectrum*, 56(10), 22–27.
- Mazlounian, A., Geroliminis, N., and Helbing, D. (2010). The spatial variability of vehicle densities as determinant of urban network capacity. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 368(1928), 4627–4647.
- McCormick, G.P. (1976). Computability of global solutions to factorable nonconvex programs: Part i—convex underestimating problems. *Mathematical programming*, 10(1), 147–175.
- Menelaou, C., Timotheou, S., Kolios, P., and Panayiotou, C.G. (2019). Joint route guidance and demand management for multi-region traffic networks. In *2019 18th European Control Conference (ECC)*, 2183–2188.
- Papageorgiou, M. (1990). Dynamic modeling, assignment, and route guidance in traffic networks. *Transportation Research Part B: Methodological*, 24(6), 471 – 495.
- Sirmatel, I.I. and Geroliminis, N. (2019). Nonlinear moving horizon estimation for large-scale urban road networks. *IEEE Transactions on Intelligent Transportation Systems*, 1–12. doi:10.1109/TITS.2019.2946324.