Nonlinear Model Predictive Control for Coordinated Traffic Flow Management in Highway Systems

Kimia Chavoshi and Anastasios Kouvelas

Abstract—The growing level of freeway traffic congestion comprises an everyday life issue with social, economic, and environmental implications for modern metropolitan areas. There is evidence that Variable Speed Limits (VSL) and Ramp Metering (RM) are two effective practical approaches to ameliorate traffic congestion. In this work we use the augmented METANET model, which is one of the most widely used macroscopic models for freeway traffic, to demonstrate the positive effects that these approaches can have on traffic flow and congestion. Since the modified METANET is a nonlinear model, nonlinear model predictive control (NLMPC) is a control method pathway for this system. It performs as a recursive on-line finite-horizon optimization of nonlinear problems, subject to the system dynamics and additional constraints, and has the privilege of prediction of future system states. We utilized the NLMPC method for the coordination of VSL and RM in highway networks. We simulate the implementation of the proposed control method on a freeway that contains a typical setting of on-ramps, off-ramps, as well as a lane drop that creates a physical bottleneck. The simulation results demonstrate significant improvement in the traffic flow conditions and provide useful insights about the way that VSL and RM manage to achieve this improvement. Understanding the special characteristics of capacity drop in highways, and how to ameliorate it, is crucial for future large-scale implementations.

I. INTRODUCTION

During the past decades freeway traffic has attracted a lot of attention from the literature, due to its crucial impact on safety, economy and the environment. Prior studies have proposed different approaches to improve traffic condition on freeway networks. Ramp Metering (RM) and Variable Speed Limits (VSL) have been widely considered as two effective methods to regulate the traffic flow on the freeway. RM methods provide better traffic conditions by controlling the on-ramp inflow on the freeway mainstream. RM strategies can be categorized as local and coordinated.

Local RM regulates the on-ramp inflow based on its neighbourhood traffic information, in order to ameliorate the local traffic conditions. ALINEA, presented in [1], is one of the most popular methods in this category, that employs a feedback control method. Although local ramp metering is well-known and widely used, especially due to its simplicity, it is easy to show that when looking at a bigger area it is sub-optimal and can be outperformed by coordinated RM strategies. Coordinated methods regulate the on-ramps inflows by utilizing system-wide traffic information, in order to enhance the overall network performance. Various methods have been presented in the literature for coordinated ramp metering; for instance [2] applied a reinforcement learning method to deal with equity issues (i.e. users from different on-ramps have equal access to the mainline). Moreover, [3] presented a multi-objective nonlinear optimization that includes two cost functions, for traffic and safety (based on a risk index model). A model predictive hierarchical control method is developed in [4], where the structure is composed from an estimation, an optimization, and a direct control layers, with focus on optimizing the total time spent. Note that various RM methods have been frequently implemented on motorways around the world. Nevertheless, the efficiency of these strategies deteriorates when they have to deal with high demands, or in cases of considering the equitable allocation of benefits among users (see [5] for more details).

VSL is another actuator that compared to RM can provide more direct and efficient control of traffic flow on the mainline of a freeway. Many different methods have been studied that apply VSL in order to improve freeway traffic conditions. For instance, [6] has proposed a model predictive control (MPC) approach. An optimal control strategy based on minimization of $L^2$ quadratic error to the desired outflow was developed by [7]. In another work, [8] studied VSL for a case study with connected vehicles, and designed a multi-objective optimization function to simultaneously optimize mobility, safety and environmental sustainability.

The integration of RM and VSL methods provides the opportunity to control mainline traffic flow conditions, while, at the same time, considering the equity and on-ramp queuing aspects. The study in [9] developed a feedback linearization VSL approach, coordinated with ALINEA/Q as a RM strategy, in order to maximize the flow rate and manage the on-ramp queues simultaneously. The combination of ALINEA and HERO as local and coordinated RM, respectively, together with a VSL algorithm is proposed by [10]. Furthermore, the work in [11] developed an MPC approach to compute the optimal coordination of RM and VSL. Although the proposed control methods for the combination of RM and VSL have presented promising results, this problem still requires more attention. The structure of the model that is considered in this MPC approach leads to a recursive Mixed Integer Non-Linear Problem (MINLP) optimization. By increasing the number of variables in MINLP, the computational time increases with an exponential rate. Consequently, the benchmark network that has been studied in the aforementioned paper is a simplified network that
consists only of one on-ramp.

In the current work, we aim at developing a Non-Linear Model Predictive Control (NLMPC) approach for the coordination of RM and VSL, and study the results of applying the developed control method on a more complicated case study with a more complex configuration of on-ramps, off-ramps and a mainline lane drop. Contrary to [11], we modify the model by using an approximation method that simplifies the optimization from MINLP to a Non-Linear Problem (NLP). This simplification accelerates the computational time and makes the NLMPC approach more practical for real world application.

The rest of the paper is organized as follows. The macroscopic model of the freeway traffic flow and the impact of implementing RM and VSL methods are described in Section 2. In Section 3, we explain our methodology and outline the NLMPC as a method to deal with nonlinear systems. The case study used for the simulation experiments is presented in Section 4. Afterwards, we investigate the simulation results for the uncontrolled case study and the controlled case study when coordinated RM and VSL are applied based on the NLMPC concept. Finally, Section 5 presents some useful insights, the conclusions and future work.

II. METANET MODEL

Many models have been developed to describe traffic flow in the macroscopic level. They model the traffic flow by formulating the dynamics of fundamental macroscopic traffic characteristics such as flow, density, and speed. According to the level of partial derivatives that the models cover, they are categorized as first order, second order and higher order models. METANET is one of the macroscopic second order traffic flow models, which is widely used for freeway networks. It represents the dynamics of traffic flow in terms of density and speed in the discretized time and space domains. Supposing a general segment \(i\) (discretized space that includes one on-ramp and one off-ramp) of a freeway with length \(\Delta_i\) and \(\lambda_i\) number of lanes at a discrete time step \(k\). We denote with \(q_{i-1}(k)\) the inflow that enters segment \(i\) from the upstream segment, and similarly, \(q_i(k)\) denotes the outflow of segment \(i\) that moves to the downstream segment. The variables \(r_i(k)\) and \(s_i(k)\) denote the on-ramp outflow and the off-ramp inflow, respectively. METANET express the dynamics of the traffic flow in segment \(i\) as follows [12]:

\[
\begin{align*}
\rho_i(k+1) &= \rho_i(k) + \frac{T}{\Delta_i \lambda_i} \left( q_{i-1}(k) - q_i(k) + r_i(k) - s_i(k) \right), \\
\frac{T}{\Delta_i} s_i(k) &= \beta_i(q_i(k)), \\
v_i(k+1) &= v_i(k) + \frac{T}{\tau} \left( V(\rho_i(k)) - v_i(k) \right) + \frac{T}{\Delta_i} v_i(k) (v_{i-1}(k) - v_i(k)) - \frac{vT}{\tau \Delta_i} \rho_{i+1}(k) - \rho_i(k) - \frac{\delta T}{\Delta_i \lambda_i} r_i(k) v_i(k) - \frac{\rho_i(k)}{\rho_{cr,i} + K}, \\
V(\rho_i(k)) &= v_{f,i} \exp \left( -\frac{1}{a_i} \left( \frac{\rho_i(k)}{\rho_{cr,i}} \right)^{a_i} \right),
\end{align*}
\]

where \(\rho_i(k)\) (veh/km/lane), \(q_i(k)\) (veh/h) and \(v_i(k)\) (km/h) denote the traffic density, flow, and speed, respectively, in segment \(i\) at time step \(k\). Based on (1), also known as conservation equation, the difference between total input flows (i.e. \(q_{i-1}(k)\) and \(r_i(k)\)) and total amount of output flows (i.e. \(q_i(k)\) and \(s_i(k)\)) in segment \(i\), results in the change of density \(\rho_i(k)\). The exiting rate \(\beta_i(q_i(k))\) denotes the ratio of \(s_i(k)\) to \(q_{i-1}(k)\). Equation (3) corresponds to the dynamics of speed and is composed by four different terms. The first is the so-called relaxation term, which demonstrates the tendency of vehicles to achieve the desired speed (i.e. the stationary speed \(V(\rho_i(k))\)). The second and third terms model the impact of spatial heterogeneity. The fourth term represents the speed drop caused by merging phenomena in case of including an on-ramp. The variable \(T\) is the time step (in sec) and \(\tau, v, K,\) and \(\delta\) denote model parameters that tune the weight of these four parts in the speed dynamics. The relationship between stationary speed and traffic flow is described by the well-known Fundamental Diagram (FD), demonstrated in equation (4), where \(v_{f,i}\) and \(\rho_{cr,i}\) represent the free flow speed and critical density, respectively. The original METANET model does not represent the effects of applying VSL and RM on freeway; therefore, we modify the METANET to describe impacts of VSL and RM utilization.

A. VSL modification on METANET model

Hegyi et al. in [11] proposed an extended version of equation (4) as follows:

\[
V(\rho_i(k)) = \min \left( (1 + \alpha) v_{VSL,i}(k), v_{f,i} \exp \left( -\frac{1}{a_i} \left( \frac{\rho_i(k)}{\rho_{cr,i}} \right)^{a_i} \right) \right)
\]

where \(v_{VSL,i}(k)\) is the speed limit on segment \(i\). The term \((1 + \alpha)\) denotes the non-compliance factor that models the disobedience of drivers towards speed limits. According to the above equation, the desired speed is the minimum between the limited speed caused by VSL and the desired speed derived from the FD without considering the VSL impact.

B. RM modification on METANET model

The inflow of the non-equipped on-ramps is derived by considering three terms. The first term is \(Q_0\) and denotes the on-ramp flow capacity due to the physical characteristics of the infrastructure (i.e. number of lanes). The second term, also called supply of space, represents the effect of the mainline congestion. Finally, the third term, also known as demand for space, represents the actual demand flow and is composed by the new arrivals \(d_i(k)\) (veh/h) and the vehicles already waiting in the ramp queue \(w_i(k)\) (veh). According to equation (7), for the RM equipped on-ramps, the inflow to the mainline is a portion of outflow in absence of RM. That is formulated as:

\[
r_i(k) = c_i(k) \times \min \left( Q_0, \frac{\rho_{max,i} - \rho_i(k)}{\rho_{max,i} - \rho_{cr,i}}, d_i(k) + w_i(k), \frac{Q_0}{T} \right),
\]
Finally, the dynamics of the on-ramp queue length $w_i(k)$ are considered limits and metering rates of the whole network. Note that the with the form of a dynamical system as follows:

$$w_i(k + 1) = w_i(k) + T(d_i(k) - r_i(k)), \quad (8)$$

where $c_i(k)$ is the metering rate that is bounded by $c_i(k) \in [c_{\text{min}}, 1]$, with $c_{\text{min}}$ denoting the minimum admissible value. Finally, the dynamics of the on-ramp queue length $w_i(k)$ are presented in equation (8), where $d_i(k)$ denotes the on-ramp demand flow (veh/h).

C. Dynamical System Definition

We summarize the METANET model described above with the form of a dynamical system as follows:

$$X(k + 1) = f(X(k), U(k), D(k)) \quad (9)$$

$$Y(k) = h(X(k)) \quad (10)$$

$$\forall i, 1 \leq i \leq N \quad (11)$$

$$X(k) = [\rho_i(k), v_i(k), w_i(k)]^T \quad (12)$$

$$U(k) = [v_{\text{VSL},i}(k), c_i(k)]^T \quad (13)$$

$$D(k) = [d_i(k), \beta_i(k)]^T \quad (14)$$

$$Y(k) = [\rho_i(k), v_i(k)]^T \quad (15)$$

where, $N$ is the number of segments; $X(k)$ is the system’s state vector which consists of the densities, mean speeds, and queue lengths of all freeway segments; $U(k)$ is the vector with the input (control) signals, which contains speed limits and metering rates of the whole network. Note that the on-ramp demands and off-ramp exiting rates are considered disturbances and are denoted with the vector $D(k)$. Finally, $Y(k)$ is the output vector that is assumed the same as the state vector. Based on the presented METANET model $f(\cdot)$ is a nonlinear vector function with a high order of complexity.

III. METHODOLOGY

In this work, the objective is to develop a coordinated control of VSL and RM in order to reduce congestion propagation and improve the traffic conditions on freeways. Non-Linear Model Predictive Control (NLMPC) is a straightforward control pathway to deal with nonlinear systems. It performs as a recursive on-line optimization of nonlinear problems subject to the system dynamics and additional constraints. In principle, the MPC type of controllers have the advantage of taking into account the impact of the predicted future behavior of the system into the current control signal designing.

In this study, we use METANET as the predictive model of macroscopic traffic flow on highway networks. In the modified version of METANET, equations (6)-(7), which describe the impacts of applying VSL and RM, contain minimum operators. In general, one can model the minimum operator by using an auxiliary binary variable in the optimization language. Therefore, in this problem the NLMPC appears as a recursive MINLP optimization. Either increasing the prediction horizon or the number of cells results in increasing the number of integer variables, and consequently the computational time grows exponentially. This drawback makes the designed NLMPC approach impractical for large-scale real networks. In order to tackle this problem we propose to approximate the minimum operators with a continuous function. The work in [13] has presented a log-sum-exp convex function $f(x) = \log(e^{x_1} + e^{x_2} + \cdots + e^{x_n})$ that is bounded as below:

$$\max(x_1, x_2, \ldots, x_n) \leq f(x) \leq \max(x_1, x_2, \ldots, x_n) + \log(n). \quad (16)$$

The above inequalities indicate that log-sum-exp function is an approximation of maximum operator, and thus the concept of a quasi-maximum function can be used. Quasi-maximum is a smooth approximation of the maximum operator:

$$\max(x_1, x_2, \ldots, x_n, \lambda) \simeq \frac{1}{\lambda} \log(e^{\lambda x_1} + e^{\lambda x_2} + \cdots + e^{\lambda x_n}), \quad \lambda > 0 \quad (17)$$

where $\lambda$ is a setting parameter; larger $\lambda$ increases the accuracy of the approximation. In this paper, we can prevent the usage of integer variables by applying the quasi-maximum to approximate the $\min\{\cdot\}$ operators in the predictive model. Therefore, the optimization transformes to a recursive NLP which requires less computational time compared with the recursive MINLP version that is applied in [11]. In order to justify this claim, we re-simulated the case study in [11] and applied the NLP method. The results indicate that the average computation time for one iteration of MPC is 1.13 s that is extremely faster than the 6 s reported by [11]. Subsequently, we define a nonlinear cost function form as follows:

$$\min_{U(k), 1 \leq k \leq k_c} \sum_{j=1}^{N_\phi} \left(\rho_i(j) - \rho_{cr,i}\right)^2 + \left(v_i(j) - v_{cr,i}\right)^2 + \omega \sum_{i=1}^{N_s} \left(v_i(k_p) - v_{cr,i}\right)^2 + \left(\rho_i(k_p) - \rho_{cr,i}\right)^2 \quad (18)$$

where, $N_\phi$ is the number of segments included in the freeway stretch under study. In fact, the optimization problem can be described as searching for the control signals $U(k)$ (RM rate and speed limits) that minimize the error of the density and speed with their corresponded desired values. In equation (18) $\rho_{cr,i}$ and $v_{cr,i}$ are the desired densities and speeds, respectively, that correspond to the critical point in the FD of the segment $i$. The critical point represents the optimal functioning point of the traffic flow in each segment. Therefore, this cost function results in trying to regulate the traffic flow close to the critical point in the FD. We denote $\omega$ as a weight to emphasise the importance of the final state of the system at the end of the prediction horizon in the cost function. In this study, we consider the control horizon $k_c$ equal to the prediction horizon $k_p$. By increasing the prediction horizon one can get more intuition about the future; therefore, the predicted trajectory is more accurate. However, increasing the prediction horizon will consequently increase the computational time and result in time-expensive solutions. In order to find an appropriate value we investigate the relationship between prediction horizon and performance of the controlled system. Vehicle Hours Traveled (VHT) is
a well-known criterion to evaluate performance of traffic networks, and is computed as follows:

\[ VHT = T \sum_{j=1}^{K} \sum_{i=1}^{N_j} \rho_i(j)v_i(j) \Delta(i,j) + T \sum_{j=1}^{K} \sum_{i=1}^{N_i} w_i(j) \]  

(19)

where \( K \) is the total simulation time. Figure 1 presents the effect of prediction horizon on VHT. By increasing the prediction horizon to more than 6 time steps, the improvement gained in terms of VHT is insignificant. In our experiments we choose to consider a prediction horizon of 9 time steps. Furthermore, in order to obtain reasonable solutions, the states and control signals need to be bounded by physical and operational constraints. The constraints on the state variables are derived from the fundamental diagram and are as follows:

\[
0 \leq \rho_i(k) \leq \rho_{i,\text{max}} \tag{20}
\]
\[
0 \leq v_i(k) \leq v_{f,i} \tag{21}
\]
\[
0 \leq w_i(k) \tag{22}
\]

We also define constraints on the control signals. Since the implementation of RM cannot block the on-ramp inflow completely, the RM rate is restricted by a lower bound (\( c_{\text{min}} = 0.2 \)), which is the minimum admissible value. On the other hand, the speed limit has also lower and upper bounds. Obviously, the speed limit cannot exceed the free flow speed; in addition, a lower bound (\( v_{\text{min}} = 50 \text{ km/h} \)) is considered due to safety reasons.

\[
C_{\text{min}} \leq c_i(k) \leq 1 \tag{23}
\]
\[
v_{\text{min}} \leq v_{\text{VSL},i} \leq v_{f,i} \tag{24}
\]

IV. CASE STUDY

The main concept that we have considered when structuring the case study is to demonstrate the performance of the control methodology in ameliorating the congestion under a relatively complex yet realistic network configuration and topology. The synthetic highway stretch that has been designed for the current case study is shown in Figure 2(a). The lane drop together with the on-ramp located at the ninth segment create a physical bottleneck that increases the risk of congestion that can spill-back towards the upstream segments of the freeway. Consequently, we expect to observe reductions in the amounts of off-ramps outflows that will aggravate congestion. Both on-ramps are equipped with RM infrastructure. Finally, we assume that VSL can be implemented on every segment along the freeway. Figure 2(b) depicts the designed network demand patterns (i.e. the mainstream inflow and the arriving flows at the on-ramps). Table I presents the calibrated values of the parameters of METANET model and the FDs.

V. SIMULATION RESULTS

In this section, we present the obtained results from the macroscopic traffic simulation. The simulation duration is 6000 sec, and we investigate two different scenarios: No Control (NC), and coordinated RM and VSL MPC-based control (RM&VSL). We compare the results of the simulations for the aforementioned scenarios and we study various aspects that can provide insights for other large-scale implementations. Figure 3 demonstrates the traffic flow characteristics based on flow, density and velocity for the two different scenarios. We can observe from the first column that for the uncontrolled freeway, a congestion is triggered in segment 8, which is downstream of the lane drop and upstream of the second on-ramp (see also Figure 2). The congestion then spills-back to the upstream segments of the freeway. By analysing the density and speed diagrams of NC scenario one can conclude that as moving towards the upstream of the freeway, the congestion levels become even higher. The densities reach values that are very close to jam density for the first three segments of the highway. In addition, the speed is close to zero and the freeway is close to become blocked. The diagrams of the second column depict the results of coordinated RM and VSL implementation. It is clear that the congestion disappears by applying the NLMPC control. Essentially, RM control improves the traffic flow conditions through managing the traffic inflows from the on-ramps to the mainline and VSL control can further enhance the performance of RM control; it provides further
TABLE I
THE SIMULATION PARAMETERS VALUES.

<p>| | | | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>$\rho_j$</td>
<td>$\rho_{cr}$</td>
<td>$Q$</td>
<td>$v_f$</td>
<td>$\tau$</td>
<td>$V$</td>
<td>$K$</td>
<td>$\delta$</td>
<td>$\alpha$</td>
</tr>
<tr>
<td>10 s</td>
<td>180 veh/km/lane</td>
<td>33.5 veh/km/lane</td>
<td>2000 veh/h</td>
<td>120 km/h</td>
<td>18 s</td>
<td>60 km$^2$/h</td>
<td>40 veh/km/lane</td>
<td>1.4</td>
<td>2</td>
</tr>
</tbody>
</table>

advantages for the traffic flow, as by inducing lower speeds on the freeway segments it enables the system to serve higher flows.

In addition, we evaluate the performance of the two scenarios based on VHT metric. For the NC scenario the VHT is equal to 1238.6 veh.h; implementing coordinated RM and VSL control reduces this value to 891.2 veh.h, which corresponds to 28.1%. Furthermore, Figure 4 presents the boundary, on-ramps, and aggregated queues diagrams. The comparison of the aggregated queue diagrams depicts that the coordinated control results in decreasing the aggregated queue lengths compared to the NC scenario. Hence, we can conclude from the integral of these diagrams (which express the contributions of different queues in VHT), that RM and VSL method increases the amount of queuing vehicles, and consequently the delays in the calculation of VHT; however, at the same time it increases the fluidity in the mainline of the freeway, which is beneficial for the total VHT of the system.

If we want to investigate the functional characteristics of the control methods we can observe Figure 5, which demonstrates the applied metering rates and variable speed limits. Moreover, this Figure depicts the coordination aspect of the control method. As we expected, the RMs are activated over long time periods; however, in correspondence with the

![Fig. 3. The flow, density and speed diagrams of NC and coordinated RM and VSL control scenarios applied on the METANET model.](image)

![Fig. 4. The boundary queue, on-ramps queue and aggregated queue diagrams for NC and coordinated RM and VSL control scenarios applied on the METANET model.](image)

![Fig. 5. The metering rates and dynamics speed limits resulted from the coordinated RM and VSL control scenario.](image)
VI. Conclusion and Future Work

In this paper, we developed an NLMPC method for the coordinated control of traffic flow on highway networks. In the methodology we proposed to utilize the quasi maximum function in order to approximate the cumbersome part of the predictive model. This approximation simplifies the optimization from an MINLP to an NLP. Afterwards, we simulated the macroscopic traffic flow model for a freeway with multiple on-ramps, off-ramps, and a lane drop. The results demonstrate that implementing the coordinated control scenario prevents congestion and enhances the traffic conditions. The proposed method decreases the VHT for 28.1% in comparison to the uncontrolled scenario. The MPC method can achieve optimal coordination, nevertheless, it results in a recursive NLP optimization which also has its drawbacks. As future work, we would like to focus on the linearization of METANET model to simplify the NLP and apply concepts from Linear Optimal Control.

REFERENCES