Traffic Control and Monitoring with a Macroscopic Model in the Presence of Strong Congestion Waves

Denis Jacquet, Carlos Canudas de Wit, Damien Koenig

Traffic Control and Monitoring with a Macroscopic Model in the Presence of Strong Congestion Waves

Denis Jacquet*, Carlos Canudas de Wit and Damien Koenig
Laboratoire d'Automatique de Grenoble
BP. 46, 38 402 St. Martin d'Hères, France

***Abstract*— This paper treats the problem of designing controllers and estimators for freeway management applications using macroscopic models and optimal control. The model is a nonlinear conservation law that may develop and propagate discontinuities known as shock waves and thus prevent the use of variational techniques requiring regularity. We show in this paper how a valid first variation of the model can be computed, give an explicit formula of its solution and use it to evaluate gradients of optimization problems using the adjoint method. The coordinated ramp metering problem and the state estimation problem are treated as illustrations.**

I. Introduction

Traffic control and monitoring is a mature and still active research field. Given the large size of freeway infrastructures, macroscopic traffic models are often considered, the simplest one being the Lighthill-Whitham-Richards (LWR) model [1] that relies on the vehicle conservation principle and the constitutive assumption that vehicles travel at an equilibrium velocity function of the local density only. This model takes the form of a nonlinear hyperbolic partial differential equation also called a conservation law, a class known to be difficult to analyse [2] and simulate [3] because of the possible presence of propagating discontinuities called shock waves. For practical applications, the authors proposed in [4] an extension of the LWR model that takes into account on/off ramps and on-ramp saturations for heavily congested main lanes by adding an inhomogeneous term with Dirac distributions. Thought the consequences of this irregular source term should be addressed more carefully, we assume this model to be well-posed when interpreted in the distributional sense, which is standard for nonlinear conservation laws [5], [2].

The distributed nature of the model led some authors [6] to discretize the state equation in order to apply finite dimensional control techniques. Nevertheless, valid discretizations of conservation laws [3] as the Godunov method cannot be put in a form suitable for control (i.e. continuous or discrete state equation). A notable exception is the discretization proposed in [7] which is used in [8] but it assumes a piecewise linear flow function and leads to rather complicated hybrid controllers. The alternative is to perform the analytical design with the distributed system directly, discretizations being used for practical implementation only. Nevertheless, if the field of partial differential equation control is rather mature for elliptic, parabolic and hyperbolic linear systems [9], very few literature is available for nonlinear conservation laws and shocks are generally disregarded [10].

*Corresponding author: denis.jacquet@lag.insieg.inpg.fr

The contribution of this paper is to provide a new methodology that takes into account the possible propagation of shocks when solving optimal control problems with conservation laws. The motivation of such work is that many traffic management applications take the form of optimization problems. Examples are the design of coordinated ramp metering strategies that maximise the vehicle-distance-travelled and the estimation of the distributed traffic density that best matches the observations at finite sensor locations.

Optimal control problems involving distributed systems are classically solved iteratively using gradient descent methods that evaluate gradients of the cost functional based on the linearized dynamics. The advantage of this approach is to use the unavoidable discretizations only to solve the infinite dimensional conditions of optimality, making this approach computationally efficient and usually more consistent than discretize-then-optimize methods. Unfortunately, the extended LWR model may develop shock waves and is thus not linearizable, at least not in the usual sense. Though some works [11] discard these shocks when linearizing, some recent results [12] proposed an elegant way to compute parametric linearization of conservation laws using a distributional interpretation. Using the weak form of the conservation law, this paper provides a rigorous framework to compute the first variation of the traffic model and gives a formula to explicitly compute its solution. Though not being exactly a linearization in the strict sense, this first variation is enough to solve iteratively optimal control problems where gradients are evaluated using the classical method of adjoint calculus. In addition, this setting provides a nice interpretation of the necessary conditions of optimality that is not available in discretize-then-optimize approaches.

This paper treats the coordinated ramp metering problem and the state estimation problem but the proposed methodology may be applied to any control or estimation problem as soon as an optimization formulation is adopted. Numerical examples using real field data are proposed in the last section.

Notations. C_0^1 is the set of continuously differentiable functions with compact support. For ρ discontinuous on Γ, $\rho_{|\Gamma}^-$ and $\rho_{|\Gamma}^+$ are its left and right values along Γ and $[\rho]_{|\Gamma}{=}\rho_{|\Gamma}^+{-}\rho_{|\Gamma}^-$. Variables of integration are omitted in integrals.

II. The extended LWR model (ELWR)

The authors proposed in [4] a model for freeway sections with on/off ramps suited for applications in ramp metering, state estimation and missing data reconstruction. It is based on the vehicle and flow conservation principles, as in the LWR formulation [1], where a source term is introduced to handle ramps. For illustration, we consider a section of the beltway of Grenoble (France) as depicted on Figure 1.

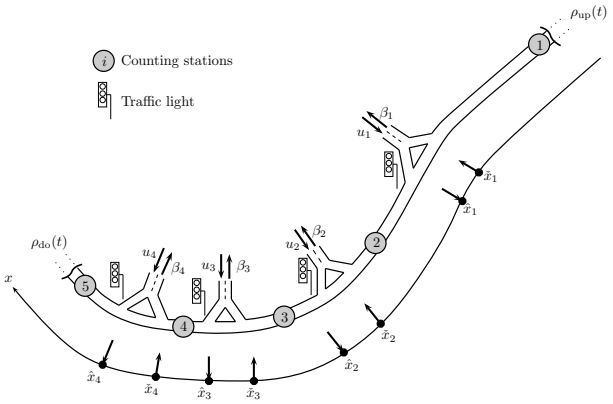

Fig. 1. Study case considered for illustration and numerical examples.

Let $(x,t) \in \Omega = (0,L) \times (0,T)$ be the space and time independent variables. The vehicle density, aggregated by lanes, is denoted $\rho(x,t)$ and plays the role of the state variable of the model. The vehicles are assumed to travel at an equilibrium speed $v(x,t) = V(\rho(x,t))$, leading to the flow function $\Phi(\rho) = \rho V(\rho)$ known as the fundamental diagram.We consider the Greenshield model [13] given by the concave quadratic form $\Phi(\rho) = \rho v_{\rm f} - \rho^2 v_{\rm f}/\rho_{\rm m}$ with $v_{\rm f}$ the free flow speed and $\rho_{\rm m}$ the maximal density. The N_u on-ramps are located at $\{\hat{x}_1, \ldots, \hat{x}_{N_u}\}$ and the N_β off-ramps at $\{\check{x}_1, \ldots, \check{x}_{N_\beta}\}$. The $i^{\rm th}$ on-ramp flow is set to $u_i(t)\Psi_i(\rho(\hat{x}_i,t))$ with $u(t)=(u_1(t),\ldots,u_{N_u}(t)) \in [0,1]^{N_u}$ the metering rates and $\Psi_i(\cdot)$ a smooth saturation function as depicted on Figure 2 that limits the inflow for large densities on the main lanes. The $i^{\rm th}$ off-ramp flow is set to be equal

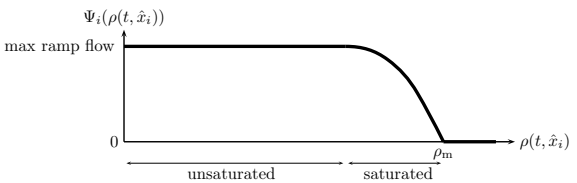

Fig. 2. Smooth saturation at on-ramp i.

to $\beta_i(t)\Phi(\rho(\check{x}_i,t))$ with $(\beta_1(t),\ldots,\beta_{N_\beta}(t)) \in [0,1]^{N_\beta}$ the split ratios and $\Phi(\cdot)$ the flow function. The functions $\Phi(\cdot)$ and $\Psi_i(\cdot)$ should be tuned (for instance by least square estimation if in parametric form) to reproduce the traffic dynamics in the region of interest.

With these notations, the traffic model in [4] writes

$$\begin{cases} \text{(DE)}\ \partial_t\rho + \partial_x\Phi(\rho) = \sum_i \delta_{\hat{x}_i} u_i \Psi_i(\rho) - \sum_j \delta_{\check{x}_j}\beta_j\Phi(\rho) \\ \text{(IC)}\ \ \rho(0,x) = \rho_I(x) \\ \text{(BC)}\ \begin{cases} \rho(t,0) \sim \rho_{\rm up}(t) \\ \rho(t,L) \sim \rho_{\rm do}(t) \end{cases} \end{cases} \tag{1}$$

where δ_ξ is the Dirac distribution centered at $x=\xi$. The symbol $\sim$ is used to point out that the Dirichlet boundary conditions are only active on a subset of $(0,T)$ determined by the BLN condition [14]. For illustration, at $x=0$, this condition tells that $\rho_{\rm up}$ applies only for entering characteristics, i.e. when $\Phi'(\rho(t,0)) > 0$ or if a discontinuous wave has positive speed, i.e. $\big(\Phi(\rho(t,0)) - \Phi(\rho_{\rm up}(t))\big)/\big(\rho(t,0) - \rho_{\rm up}(t)\big)$.

Being a conservation law, Equation (1) is valid only if it is interpreted in the distributional or weak sense [2] as ρ may develop discontinuities. A standard choice for the space of test functions is $\Lambda = C_0^1(]-\infty,T[\times]0,L[)$, making the weak sense interpretation of (DE) in (1) of the form

$$\sum_i \int_0^T u_i \Psi_i(\rho)\big|_{x=\hat{x}_i}\phi\big|_{x=\hat{x}_i} - \sum_j \int_0^T \beta_j\Phi(\rho)\big|_{x=\check{x}_j}\phi\big|_{x=\check{x}_j} + \int_0^T\int_0^L \rho\partial_t\phi + \Phi(\rho)\partial_x\phi + \int_0^L \rho_I\phi\big|_{t=0} = 0\ ,\ \forall\ \phi\in\Lambda \tag{2}$$

Equation (2) is a set of integral equations ensuring the conservation of vehicles and flows on any subset of Ω. Though its rigorous wellposedness requires additional analysis, it is assumed to have a unique piecewise-C^1 solution satisfying the entropy admissibility condition ([5], [15]). Curves of discontinuity, known as shock waves, are denoted $\{\Gamma_i\}_{i=1,\ldots,N_s}$ and correspond to strong congestion waves in traffic applications. Depending on the initial and boundary conditions as well as on the ramp flows, shocks may appear in finite time when characteristics intersect. Each curve of discontinuity Γ_i stating at time t_i^I is parameterized by $\Gamma_i = \{(s_i(t),t), t \in (t_i^I,T)\}$. The entropy admissibility condition ([5], [15]) tells that shocks are allowed to propagate with speed $\dot{s}_i(t) = [\Phi(\rho)]_{|\Gamma_i}/[\rho]_{|\Gamma_i}$ if $\rho_{|\Gamma_i}^- < \rho_{|\Gamma_i}^+$, i.e. if characteristic curves are oriented towards Γ_i near the shock.

As ρ is discontinuous at $x=\hat{x}_i$ and $x=\check{x}_j$, the products $\delta_{\hat{x}_i}u_i\Psi_i(\rho)$ and $\delta_{\check{x}_j}\beta_j\Phi(\rho)$ of a Dirac and a discontinuous function are ill-defined. This inconsistency can be removed with $\Psi_i(\rho)_{|\hat{x}_i} = \Psi_i\big(\max(\lim_{x\uparrow\hat{x}_i}\rho, \lim_{x\downarrow\hat{x}_i}\rho)\big)$ and $\Phi(\rho)_{|\check{x}_j} = \Phi(\lim_{x\uparrow\check{x}_j}\rho)$ from the traffic behavior.

To ease forthcoming calculus, Equation (DE) in (1) can be formulated with a divergence operator in the frame (e_x,e_t)

$$\nabla\cdot\begin{pmatrix}\Phi(\rho)\\ \rho\end{pmatrix} = b(\rho,u,\beta) \qquad \text{in } \Omega \tag{3}$$

with $b(\rho,u,\beta) = \sum_i \delta_{\hat{x}_i}u_i\Psi_i(\rho) - \sum_j \delta_{\check{x}_j}\beta_j\Phi(\rho)$. This formulation makes the method of characteristics [15] very explicit by showing that ρ is the integral of the inhomogeneous term $b(\rho,u,\beta)$ along the integral curves of the field

$$\eta(\rho) = \begin{pmatrix}\Phi'(\rho)\\ 1\end{pmatrix}$$

Contrary to linear hyperbolic equations, η is not a structural property of the model as it depends on the solution ρ.

III. First variation of the ELWR model

It is a known fact [16] that nonlinear hyperbolic equations cannot be linearized along a trajectory $\bar{\rho}$. The reason is that in the presence of shocks, small perturbations $\tilde{\rho}$ will propagate with increasing magnitude in time as they modify the shock speeds. Recently, a method was proposed in [12] to compute the linearized dynamics with respect to a finite dimensional parameter in the initial condition using distributional calculus. We present in what follows a new method based on the weak formulation to compute the first variation of the ELWR model. This operator will then be used to solve iteratively optimization problems.

Theorem (first variation of ELWR) *The first variation of the state equation along the trajectory $(\bar{\rho}, \bar{\rho}_I, \bar{u})$ with perturbation $(\tilde{\rho}, \tilde{\rho}_I, \tilde{u})$ is given by*

$$\left\{\begin{array}{l} \partial_t\tilde{\rho} + \partial_x\left(\Phi'(\bar{\rho})\tilde{\rho}\right) = \\ \qquad \sum_i \delta_{\hat{x}_i}\left(\bar{u}_i\Psi_i'(\bar{\rho})\tilde{\rho} + \tilde{u}_i\Psi_i(\bar{\rho})\right) - \sum_j \delta_{\check{x}_j}\beta_j\Phi'(\bar{\rho})\tilde{\rho} \\ \tilde{\rho}(0,x) = \tilde{\rho}_I \\ \left\{\begin{array}{ll} \tilde{\rho}(t,0) = 0 & \text{when } \rho_{\text{up}} \text{ applies.} \\ \tilde{\rho}(t,L) = 0 & \text{when } \rho_{\text{do}} \text{ applies.} \end{array}\right. \end{array}\right. \tag{4}$$

The dynamical part of (4) should be interpreted in the weak sense and can be put in the divergence form

$$\nabla\cdot\begin{pmatrix}\Phi'(\bar{\rho})\tilde{\rho} \\ \tilde{\rho}\end{pmatrix} = b_\rho(\bar{\rho},\bar{u},\beta)\tilde{\rho} + b_u(\bar{\rho})\tilde{u} \tag{5}$$

$$\text{with } \left\{\begin{array}{lcl} b_\rho(\bar{\rho},\bar{u},\beta) & = & \sum_i \delta_{\hat{x}_i}\bar{u}_i\Psi_i'(\bar{\rho}) - \sum_j \delta_{\check{x}_j}\beta_j\Phi'(\bar{\rho}) \\ b_u(\bar{\rho}) & = & \sum_i \delta_{\hat{x}_i}\Psi_i(\bar{\rho}) \end{array}\right.$$

Proof: The weak formulation (2) is a set of equations of the form $\mathcal{F} : (\rho, \rho_I, u) \in (L^1_{\text{loc}})^3 \to \mathbb{R}$. A Taylor expansion of $\mathcal{F}$ around $(\bar{\rho}, \bar{\rho}_I, \bar{u})$ with perturbation $(\tilde{\rho}, \tilde{\rho}_I, \tilde{u})$ can be performed by setting $\rho = \bar{\rho} + \tilde{\rho}$, $\rho_I = \bar{\rho}_I + \tilde{\rho}_I$ and $u = \bar{u} + \tilde{u}$. The function $\Phi(\cdot)$ being quadratic, we have the expansion $\Phi(\bar{\rho} + \tilde{\rho}) = \Phi(\bar{\rho}) + \Phi(\bar{\rho})'\tilde{\rho} + \frac{1}{2}\Phi''\tilde{\rho}^2$ with Φ'' a negative constant. The first variation of (2) is then obtained by removing the nonlinear terms that vanish as $(\tilde{\rho}, \tilde{\rho}_I, \tilde{u}) \to (0,0)$ and setting the boundary conditions of $\tilde{\rho}$ to 0 as there are not allowed to vary.

■

It makes sense that the first variation of a conservation law is itself a conservation law as the conservation principle should be fulfilled by the unperturbed and perturbed systems.

Theorem (Solution of the first variation of ELWR) *The solution of Equation (4) is*

$$\tilde{\rho} = \tilde{\rho}_c + \sum_{i=1}^{N_s}\delta_{\Gamma_i}\kappa_i \tag{6}$$

with $\tilde{\rho}_c$ the solution in $\Omega \setminus (\cup_i\Gamma_i)$ of

$$\left\{\begin{array}{l} \partial_t\tilde{\rho}_c + \Phi'(\bar{\rho})\partial_x\tilde{\rho}_c = b_\rho(\bar{\rho},\bar{u},\beta)\tilde{\rho}_c + b_u(\bar{\rho})\tilde{u} \\ \tilde{\rho}_c(0,x) = \tilde{\rho}_I(x) \end{array}\right. \tag{7}$$

and κ_i with $i = 1, ..., N_s$, the solutions of

$$\left\{\begin{array}{l} \frac{\mathrm{d}}{\mathrm{d}t}\kappa_i = b_\rho(\bar{\rho},\bar{u},\beta)_{|\Gamma_i}\kappa_i + \dot{\bar{s}}_i[\tilde{\rho}_c]_{|\Gamma_i} - [\Phi'(\bar{\rho})\tilde{\rho}_c]_{|\Gamma_i} \\ \kappa_i(t_i^I) = 0 \end{array}\right. \tag{8}$$

This result states that the solution of (4) is the sum of the piecewise-C^1 solution of (7) constructed from the characteristic field $\eta(\bar{\rho})$ and Dirac measures on the shock curves Γ_i. Equation (7) is well-posed at the shock locations as the entropy condition ([5], [15]) states that the characteristics are leaving the computational domain near these boundary curves. In (8), $\mathrm{d}/\mathrm{d}t$ is the full time derivative along a shock.

Proof: The proof relies on the integration by parts formula that states that for F a continuously differentiable vector field and g a continuously differentiable functional, we have

$$\int_\Omega (\nabla\cdot F)\, g = -\int_\Omega F\cdot\nabla g + \int_{\partial\Omega} F\cdot\nu g$$

with ν the unit outward normal vector to $\partial\Omega$. Let consider the special case of $\bar{\rho}$ with a single shock Γ on $x = \bar{s}(t)$. We can write $\bar{\rho} = \bar{\rho}^1 + (\bar{\rho}^2 - \bar{\rho}^1)H(x - \bar{s}(t))$ with $H(\cdot)$ the Heaviside distribution, $\bar{\rho}^1$ and $\bar{\rho}^2$ being the C^1 solutions at the left and right of the shock. A formal differentiation would give $\tilde{\rho} = \tilde{\rho}^1 + (\bar{\rho}^2 - \tilde{\rho}^1)H(x - \bar{s}(t)) - \tilde{s}(\bar{\rho}^2 - \bar{\rho}^1)\delta(x - \bar{s}(t))$, which suggests a solution of the form $\tilde{\rho} = \tilde{\rho}_c + \sum_i \delta_{\Gamma_i}\kappa_i$ as proposed in [12]. Interpreting (5) in the weak sense, we have the following set of equations for all $\phi \in C_0^1(]0,T[\times]0,L[)$

$$\begin{aligned} 0 &= \int_\Omega \begin{pmatrix}\Phi'(\bar{\rho})\tilde{\rho} \\ \tilde{\rho}\end{pmatrix}\cdot\nabla\phi + b_\rho(\bar{\rho},\bar{u},\beta)\tilde{\rho}\phi + b_u(\bar{\rho})\tilde{u}\phi \\ 0 &= \sum_k \int_{\Omega_k} \begin{pmatrix}\Phi'(\bar{\rho})\tilde{\rho}_c \\ \tilde{\rho}_c\end{pmatrix}\cdot\nabla\phi + b_\rho(\bar{\rho},\bar{u},\beta)\tilde{\rho}_c\phi + b_u(\bar{\rho})\tilde{u}\phi \\ &\quad + \sum_{i=1}^{N_s}\int_\Omega \delta_{\Gamma_i}\kappa_i\left(\begin{pmatrix}\Phi'(\bar{\rho}) \\ 1\end{pmatrix}\cdot\nabla\phi + b_\rho(\bar{\rho},\bar{u},\beta)\phi\right) \end{aligned}$$

with $\bigcup_k \Omega_k$ a decomposition of Ω where $\bar{\rho}$ is C^1. Note that $\delta_{\Gamma_i}\Phi'(\bar{\rho})$ is ill-defined as $\Phi'(\bar{\rho})$ is discontinuous along the curve Γ_i. Nevertheless, this difficulty can be overcame by using generalized characteristics [17] and Filippov's theory [18] of discontinuous fields. Without going into details, we set $\Phi'(\bar{\rho})_{|\Gamma_i} = [\Phi(\bar{\rho})]_{|\Gamma_i}/[\bar{\rho}]_{|\Gamma_i}$ which generalizes the way information propagates. An integration by parts gives

$$\begin{aligned} 0 &= \int_\Omega \left(-\nabla\cdot\begin{pmatrix}\Phi'(\bar{\rho})\tilde{\rho}_c \\ \tilde{\rho}_c\end{pmatrix} + b_\rho(\bar{\rho},\bar{u},\beta)\tilde{\rho}_c + b_u(\bar{\rho})\tilde{u}\right)\phi \\ &\quad + \sum_{i=1}^{N_s}\int_{t_i^I}^T \left(-[\Phi'(\bar{\rho})\tilde{\rho}_c]_{|\Gamma_i} + \dot{\bar{s}}_i(t)[\tilde{\rho}_c]_{|\Gamma_i}\right)\phi_{|\Gamma_i} \\ &\quad + \sum_{i=1}^{N_s}\int_{t_i^I}^T \kappa_i\left(\begin{pmatrix}\Phi'(\bar{\rho})_{|\Gamma_i} \\ 1\end{pmatrix}\cdot\nabla\phi_{|\Gamma_i} + b_\rho(\bar{\rho},\bar{u},\beta)_{|\Gamma_i}\phi_{|\Gamma_i}\right) \end{aligned}$$

The first term is equal to 0 with ρ_c the strong solution of (7), which can be computed by the method of characteristics [15] on each subset Ω_k. The second term is obtained using the fact that ϕ is continuous along the shocks as $\phi \in C_0^1$. Finally, given the value affected to $\Phi'(\bar{\rho})_{|\Gamma_i}$, we recognize the directional derivative of ϕ along the shock curve Γ_i

$$\frac{\mathrm{d}}{\mathrm{d}t}\phi_{|\Gamma_i} = \partial_t\phi_{|\Gamma_i} + \Phi'(\bar{\rho})_{|\Gamma_i}\partial_x\phi_{|\Gamma_i} = \partial_t\phi_{|\Gamma_i} + \dot{\bar{s}}_i(t)\partial_x\phi_{|\Gamma_i} \tag{9}$$

leading to

$$0 = \int_{t_i^I}^{T} \kappa_i \frac{\mathrm{d}}{\mathrm{d}t}\phi_{\mathbb{\Gamma}_i} + \Big(- [\Phi'(\bar\rho)\tilde\rho_c]_{|\Gamma_i} + \dot{\tilde s}_i(t)[\tilde\rho_c]_{|\Gamma_i} + b_\rho(\bar\rho,\bar u,\beta)_{\mathbb{\Gamma}_i}\kappa_i\Big)\phi_{\mathbb{\Gamma}_i}$$

An integration by parts in time concludes the proof. ∎

It should be pointed out that (4) is not rigorously a linearization of the state equation (1) as its solution (6) is a distribution whereas the solution of (1) is in L^1_{loc}. This mismatch of functional spaces does not allow the mapping $\mathcal{H} : u \mapsto \rho$ to be linearized as $\mathcal{H}(\bar u + \tilde u) \nsim \mathcal{H}(\bar u) + \mathcal{H}'(\bar u)\tilde u$. Nevertheless, all the information of a linearization is in (6) as the shock sensitivity is contained in its singular part. The same kind of inconsistency appears for the Heaviside distribution as $H(x - \bar a - \tilde a) \nsim H(x - \bar a) + \delta(x - \bar a)\tilde a$.

Note that the shock sensitivity is given by a time integration in Equation (8) as perturbations of the state near a shock modify the shock speed.

Optimization problems are often solved using iterative algorithms that evaluate gradients of the cost based on the first variation of the constraints. This result is thus a first step towards the development of methods able to solve optimal control problems involving nonlinear conservation laws.

IV. Application to optimization problems

A. Problem formulation

Many traffic management problems can be casted as optimal control problems with the traffic model as a dynamical constraint and possible constraints on the admissible states and controls. Using barrier techniques to move the last constraints to the cost function, these problems take the form

$$\begin{aligned} \underset{y}{\mathbf{Min}} \quad & \mathcal{J}(\rho,u) = \mathcal{J}_{\mathrm{obs}}(\rho) + \mathcal{J}_{\mathrm{bar}}(u) \\ & \qquad = \textstyle\int_0^T\int_0^L \mathcal{Q}(\rho) + \int_0^T \mathcal{R}(u) \\ \mathbf{Subj.\ to} \quad & (\mathrm{DE}) - (\mathrm{IC}) - (\mathrm{BC}) \text{ of Eq. } (1) \end{aligned} \tag{10}$$

where the decision variable y can be any variable of the problem depending on the application (generally the initial condition ρ_I or the metering rates u).

B. Gradient evaluation by the adjoint method

A simple variational analysis of Equation (10) by setting $\rho = \bar\rho + \tilde\rho$ and $u = \bar u + \tilde u$ gives the perturbed cost

$$\begin{aligned} \tilde{\mathcal{J}}(\tilde\rho,\tilde u) &= \tilde{\mathcal{J}}_{\mathrm{obs}}(\tilde\rho) + \tilde{\mathcal{J}}_{\mathrm{bar}}(\tilde u) \\ &= \int_0^T\int_0^L D_\rho\mathcal{Q}(\bar\rho)\tilde\rho + \sum_{i=1}^{N_u}\int_0^T D_{u_i}\mathcal{R}(\bar u)\tilde u_i \end{aligned}$$

But $\tilde\rho$ is not a decision variable and is defined through the traffic dynamics (1). A standard approach [9] is to use the adjoint equation of its first variation to evaluate $\tilde{\mathcal{J}}$. We have

$$\int_\Omega \{\partial_t\tilde\rho + \partial_x\left(\Phi'(\bar\rho)\tilde\rho\right) - \sum_i \delta_{\hat x_i}\bar u_i\Psi_i'(\bar\rho)\tilde\rho + \sum_j \delta_{\check x_j}\beta_j\Phi'(\bar\rho)\tilde\rho - \sum_i \delta_{\hat x_i}\tilde u_i\Psi_i(\bar\rho)\}\lambda =$$

$$\underbrace{\int_\Omega \{-\partial_t\lambda - \Phi'(\bar\rho)\partial_x\lambda - \sum_i \delta_{\hat x_i}\bar u_i\Psi_i'(\bar\rho)\lambda + \sum_j \delta_{\check x_j}\beta_j\Phi'(\bar\rho)\lambda\}\tilde\rho}_{\mathcal{A}(\lambda)}$$
$$+ \underbrace{\int_0^L \tilde\rho\lambda\big|_{t=T}}_{\mathcal{B}_T(\lambda,\tilde\rho)} - \underbrace{\int_0^L \tilde\rho\lambda\big|_{t=0}}_{\mathcal{B}_I(\lambda,\tilde\rho)} + \underbrace{\int_0^T \Phi'(\bar\rho)\tilde\rho\lambda\big|_{x=L}}_{\mathcal{B}_D(\lambda,\tilde\rho)} - \underbrace{\int_0^T \Phi'(\bar\rho)\tilde\rho\lambda\big|_{x=0}}_{\mathcal{B}_U(\lambda,\tilde\rho)}$$
$$- \sum_{i=1}^{N_u}\int_0^T \underbrace{\lambda_{|\hat x_i}\Psi_i(\bar\rho)_{|\hat x_i}}_{\mathcal{C}_i(\lambda)}\tilde u_i$$
$$+ \sum_{i=1}^{N_s}\int_{\Gamma_i} \underbrace{\frac{1}{\sqrt{1+\dot{\bar s}_i^2}}\begin{pmatrix}-1\\ \dot{\bar s}_i\end{pmatrix}\cdot\begin{pmatrix}[\Phi'(\bar\rho)\tilde\rho]_{|\Gamma_i}\\ [\tilde\rho]_{|\Gamma_i}\end{pmatrix}\lambda_{|\Gamma_i}}_{\mathcal{D}(\lambda,\tilde\rho)}$$

By setting

$$\begin{cases} \mathcal{A}(\lambda) &= D_\rho\mathcal{Q}(\bar\rho) \\ \mathcal{B}_T(\lambda,\tilde\rho) &= 0 \\ \mathcal{B}_D(\lambda,\tilde\rho) &= 0 \\ \mathcal{B}_U(\lambda,\tilde\rho) &= 0 \\ \mathcal{D}(\lambda,\tilde\rho) &= 0 \end{cases} \tag{11}$$

we obtain

$$\begin{aligned} \tilde{\mathcal{J}} &= \int_0^L \mathcal{B}_I(\lambda,\tilde\rho) + \sum_{i=1}^{N_u}\int_0^T \{\mathcal{C}_i(\lambda) + D_{u_i}\mathcal{R}(u)\}\tilde u_i \\ &= \int_0^L \lambda(0,\cdot)\tilde\rho_I + \sum_{i=1}^{N_u}\int_0^T \{\Psi_i(\bar\rho(\cdot,\hat x_i))\lambda(\cdot,\hat x_i) + D_{u_i}\mathcal{R}(u)\}\tilde u_i \end{aligned}$$

Theorem (Gradient evaluation) *The gradient of the abstract problem (10) may be evaluated by*

$$\nabla_{\rho_I}\mathcal{J} = \lambda(0,\cdot) \tag{12}$$
$$\nabla_{u_i}\mathcal{J} = \Psi_i(\bar\rho(\cdot,\hat x_i))\lambda(\cdot,\hat x_i) + D_{u_i}\mathcal{R}(u) \tag{13}$$

with the adjoint variable λ solving

$$\begin{cases} (\mathrm{DE})\ -\partial_t\lambda - \Phi'(\bar\rho)\partial_x\lambda = D_\rho\mathcal{Q}(\bar\rho) \\ \qquad\qquad + \sum_i \delta_{\hat x_i}\bar u_i\Psi_i'(\bar\rho)\lambda - \sum_j \delta_{\check x_j}\beta_j\Phi'(\bar\rho)\lambda \\ (\mathrm{FC})\ \lambda(T,x) = 0 \\ (\mathrm{BC}) \begin{cases} \lambda(t,0) = 0 \ \text{ when } \rho_{\mathrm{up}} \text{ does not apply.} \\ \lambda(t,L) = 0 \ \text{ when } \rho_{\mathrm{do}} \text{ does not apply.} \\ \lambda_{\mathbb{\Gamma}_i} = 0,\ i = 1,...,N_s \end{cases} \end{cases} \tag{14}$$

Proof: The adjoint equation (14) with final condition (FC) and boundary conditions (BC) is defined so that all equalities in (11) are satisfied. The boundary conditions of the adjoint equation (14) and of the first variation (4) apply on complementary domains that depends on the solution $\bar\rho$ of (1). Additional homogeneous boundary conditions $\lambda_{\mathbb{\Gamma}_i} = 0$ are provided at the shock locations to remove the singular part of (6). They make the reverse problem (14) well-posed as backward characteristics leave the boundaries Γ_i. ∎

This gradient evaluation method may be embedded in any descent algorithm to compute iteratively a solution to (10). Even with the analytic expressions (12,13,14), discretizations

are unavoidable as infinite dimensional calculus is needed. Numerical schemes solving (1) and (14) are available in [4].

The dual variable λ can be interpreted as a marginal cost. The backwards differential operator $-\partial_t\lambda - \Phi'(\bar\rho)\partial_x\lambda$ in (14) implies that λ is the integral of the inhomogeneous term along the backwards characteristics. The characteristic field being the same than for the first variation (4), the dual equation can be interpreted as the following

1) $D_\rho\mathcal{Q}$ is used to trig the adjoint variable where improvements are possible. It creates some marginal cost.
2) The marginal cost travels towards some regions where decision variables are available by following the backwards characteristics.

V. Traffic control and monitoring applications

Two traffic related problems are treated as illustrations. The first one is the optimization of a coordinated and traffic responsive ramp metering algorithm to reduce congestion on freeways. The second is the estimation of the traffic state where no sensor is available.

A. The ramp metering problem

Considering the objective of maximizing the vehicle-distance-travelled, [4] proposed the following formulation.

$$\begin{array}{ll}\underset{u}{\mathbf{Min}} & \mathcal{J}_{\rm r}(\rho) = -\int_0^T\int_0^L \Phi(\rho) \\ \mathbf{Subj.\ to} & \begin{cases} \text{(DE)} - \text{(IC)} - \text{(BC) of Eq. (1)} \\ u \in U_{\rm ad} = L^2\big((0,T),[0,1]\big)^{N_u} \end{cases}\end{array} \tag{15}$$

$U_{\rm ad}$ being a compact and convex subset, the barrier function

$$\mathcal{J}^M_{\rm bar}(u) = -\frac{1}{M}\sum_{i=1}^{N_u}\int_0^T \ln\big(u_i(1-u_i)\big)$$

can be used, leading to the augmented problem

$$\begin{array}{ll}\underset{u}{\mathbf{Min}} & \mathcal{J}_{\rm aug}(\rho,u) = \mathcal{J}_{\rm r}(\rho) + \mathcal{J}^M_{\rm bar}(u) \\ \mathbf{Subj.\ to} & \text{(DE)} - \text{(IC)} - \text{(BC) of Eq. (1)}\end{array} \tag{16}$$

The solution of (16) will converge to that of (15) as $M \to \infty$ while keeping solutions of (16) in the admissible set $U_{\rm ad}$. In this problem, the initial condition ρ_I is supposed to be known thus it is not a decision variable and $\tilde\rho_I = 0$. The first variation of the cost functional thus reduces to

$$\tilde{\mathcal{J}}_{\rm aug} = \sum_{i=1}^{N_u}\int_0^T \{\Psi_i(\rho(\cdot,\hat x_i))\lambda(\cdot,\hat x_i) + D_{u_i}\mathcal{J}^M_{\rm bar}(u)\}\tilde u_i$$

giving the following gradient in L^2

$$\nabla_{u_i}\mathcal{J}_{\rm aug} = \Psi_i(\rho(\cdot,\hat x_i))\lambda(\cdot,\hat x_i) + D_{u_i}\mathcal{J}^M_{\rm bar}(u) \tag{17}$$

and the dual equation

$$\begin{aligned}-\partial_t\lambda - \Phi'(\bar\rho)\partial_x\lambda &= \Phi'(\bar\rho) \\ &\quad + \sum_i \delta_{\hat x_i}\bar u_i\Psi_i'(\bar\rho)\lambda - \sum_j \delta_{\bar x_j}\beta_j\Phi'(\bar\rho)\lambda\end{aligned}$$

This result is in accordance with our previous results in [4] obtained with a slightly different method.

The following descent algorithm solves (15) iteratively.

Require: $u_i := u_i^{\rm init} \in (0,1)$, $M := M_{\rm init}$, ϵ_i, ϵ_o, ΔM
  while $\mathcal{J}^M_{\rm bar}(u)/\mathcal{J}_{\rm r}(\rho) > \epsilon_o$ **do**
    while $\|\nabla_u\mathcal{J}_{\rm aug}\| > \epsilon_i$ **do**
      Compute ρ from (1)
      Compute λ from (14)
      Compute $\nabla_{u_i}\mathcal{J}_{\rm aug}$ from (17)
      Update $u_i := u_i - t\nabla_{u_i}\mathcal{J}_{\rm aug}$, $t \in (0,1)$ s.t. $u \in (0,1)$
    end while
    M := M.ΔM
  end while

Figures 3 and 4 give the obtained results for the ramp metering optimization of the freeway section of Figure 1. A time horizon of 1.5 hour at the beginning of the afternoon rush hours is considered with real field initial and boundary data courtesy of DDE Isère. We observe that the metering rates decrease in Figure 4 to cope with the incoming congestion. The weakness of this method is that it requires the knowledge of the initial condition ρ_I and of the estimates of the boundary conditions $\rho_{\rm up}$ and $\rho_{\rm do}$. Receding horizon techniques may help to avoid propagation of errors in the estimates of ρ_I, $\rho_{\rm up}$ and $\rho_{\rm do}$.

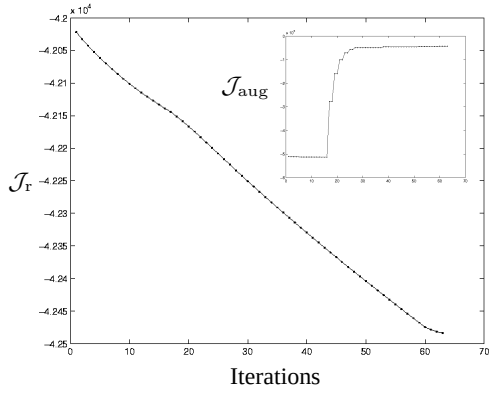

Fig. 3. Reduction of the costs $\mathcal{J}_{\rm r}$ and $\mathcal{J}_{\rm aug}$

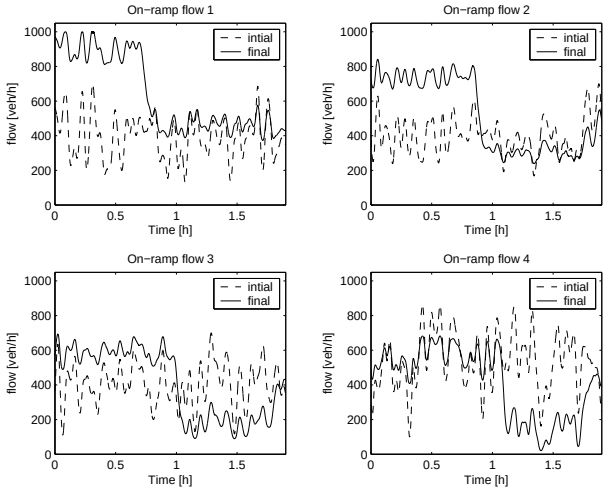

Fig. 4. Optimal on-ramp flows before and after optimization in the afternoon rush hours.

B. The state estimation problem

Consider the problem of estimating the current state $\rho(0,x)$ based on the density measurements $\xi_i(t)$ at a finite set of locations $\{\tilde{x}_i\}_{i=1}^{N_m}$ on the time horizon $(-T,0)$. As nonlinear conservation laws are not invertible (cannot be integrated backwards), iterations on the final condition would not be valid. The alternative is to search for the initial condition that minimizes the square error at the sensor locations, the final state being given by a one-to-one correspondance. The state estimation problem is formulated as following

$$\begin{aligned} \underset{\rho_I}{\textbf{Min}} \quad & \mathcal{J}_{\mathrm{o}}(\rho) = \textstyle\sum_i^{N_m} \frac{1}{2} \int_0^T (\rho(\cdot,\tilde{x}_i) - \xi_i)^2 \\ & = \textstyle\sum_i^{N_m} \frac{1}{2} \int_0^T \int_0^L \delta_{\tilde{x}_i} (\rho - \xi_i)^2 \\ \textbf{Subj. to} \quad & (\mathrm{DE}) - (\mathrm{IC}) - (\mathrm{BC}) \text{ of Eq. } (1) \end{aligned} \tag{18}$$

In this problem, the ramp metering rates u_i are supposed to be known leading to $\tilde{u}_i = 0$ for all i. We get

$$\tilde{\mathcal{J}}_{\mathrm{o}} = \int_0^L \lambda(0,\cdot)\tilde{\rho}_I$$

leading to the gradient

$$\nabla_{\rho_I} \mathcal{J}_{\mathrm{o}} = \lambda(0,\cdot) \tag{19}$$

and the adjoint equation

$$\begin{aligned} -\partial_t \lambda - \Phi'(\bar{\rho})\partial_x \lambda = {} & \delta_{\tilde{x}_i}(\rho - \xi_i) \\ & + \sum_i \delta_{\hat{x}_i} \bar{u}_i \Psi_i'(\bar{\rho})\lambda - \sum_j \delta_{\check{x}_j} \beta_j \Phi'(\bar{\rho})\lambda \end{aligned}$$

The marginal cost interpretation gives some insight on the limitations of the method. As characteristics linking the sensor locations to the initial condition in $(-T,0)$ are the only ones to provide information in the descent method, a lack of such characteristics would lead to a poor estimation. Nevertheless, this is a structural limitation of the system that cannot be overcome by other methods.

The following descent algorithm solve (18) iteratively and Figures 5 and 6 present the results for 5 sensors.

Require: $\rho_I := \rho_I^{\mathrm{init}} \in (0, \rho_{\mathrm{m}})$, ϵ
  while $\|\nabla_{\rho_I} \mathcal{J}_{\mathrm{o}}\| > \epsilon$ **do**
    Compute ρ from (1)
    Compute λ from (14)
    Compute $\nabla_{\rho_I} \mathcal{J}_{\mathrm{o}}$ from (19)
    Update $\rho_I := \rho_I - \nabla_{\rho_I} \mathcal{J}_{\mathrm{o}}$
  end while

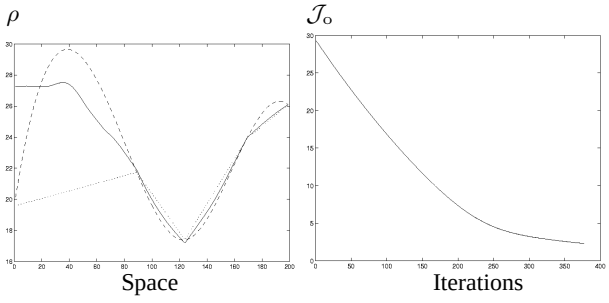

Fig. 5. Estimated initial condition (dashed: actual, plain: estimated, dot: linear interpolation of measurements) and cost evolution.

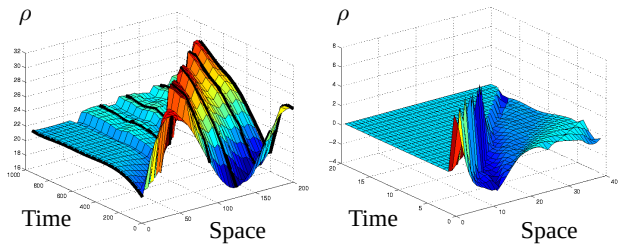

Fig. 6. Actual density distribution (left) and residual error after optimization (right). The 5 black lines show the sensor locations.

VI. Conclusion

This paper propose a rigorous methodology to compute the first variation of a scalar conservation law by using its weak interpretation. Though this first variation cannot be called rigorously a linearization, it can be used to solve optimization problems. Traffic control and monitoring applications were considered to show the practical usefulness of the approach.

References

[1] M. Lighthill and G. Whitham, "On kinematic waves II: a theory of traffic flow on long crowded roads," in *Proceedings Royal Society, London Series A*, 1955.

[2] A. Bressan, *Hyperbolic Systems of Conservation Laws - The One-dimensional Cauchy Problem*. Oxford University Press, 2000.

[3] R. J. LeVeque, *Numerical Methods for Conservation Laws*. Birkhäuser, 1992.

[4] D. Jacquet, C. Canudas de Wit, and D. Koenig, "Optimal ramp metering strategy with an extended LWR model: Analysis and computational methods," in *Proceedings of the 2005 IFAC World Congress*, 2005.

[5] R. Ansorge, "What does the entropy condition mean in traffic flow theory?" *Transportation Research Part B*, 1990.

[6] A. Kotsialos, M. Papageorgiou, M. Mangeas, and H. Haj-Salem, "Coordinated and integrated control of motorway networks via nonlinear optimal control," *Transportation Research*, 2002.

[7] C. Daganzo, "The cell transmission model: A dynamic representation of highway traffic consistent with the hydrodynamic theory," *Transportation Research Part B*, 1994.

[8] X. Sun, L. Muñoz, and R. Horowitz, "Mixture kalman filter based highway congestion mode and vehicle density estimator and its application," in *Proceedings of the 2004 American Control Conference*, 2004.

[9] J. Lions, *Optimal control of systems governed by partial differential equations*. Springer, 1971.

[10] J. de Halleux, C. Prieur, J.-M. Coron, B. d'Andra Novel, and G. Bastin, "Boundary feedback control in networks of open channels," *Automatica*, 2003.

[11] A. M. Bayen, R. Raffard, and C. J. Tomlin, "Network congestion alleviation using adjoint hybrid control: Application to highways," in *Hybrid Systems: Computation and Control*, 2004.

[12] C. Bardos and O. Pironneau, "Derivatives and control in the presence of shocks," *Computational Fluid Dynamics Journal*, 2003.

[13] L. Pipes, "Car following models and the fundamental diagram of road traffic," *Transportation Research*, 1967.

[14] C. Bardos, A. LeRoux, and J. Nedelec, "First order qusilinear equations with bounday conditions," *Communications in Partial Differential Equations*, 1979.

[15] L. Evans, *Partial Differential Equations*. American Mathematical Society, 1998.

[16] G. Whitham, *Linear and Nonlinear Waves*, ser. Pure and applied mathematics. John Wiley & Sons, 1974.

[17] C. Dafermos, "Generalized characteristics and the structure of solutions of hyperbolic conservation laws," *Indiana University Mathematics Journal*, 1977.

[18] A. Filippov, *Differential Equations with Discontinuous Right-hand Side*. Kluwer Academic Publisher, 1988.