Polyhedral hybrid systems and applications to the Cell Transmission Model

Jerome Thai

I. Introduction

Numerous traffic estimation techniques developed in the literature rely on density based traffic models such as the Lighthill-Whitham-Richards (LWR) partial differential equation (PDE) [18], [20] and its discretization using the Godunov scheme [16], [17], [21] (also known as the Cell Transmission Model (CTM) [6], [7] in the transportation literature). These highway traffic monitoring systems rely on large amounts of data from different sources. These include *inductive loop detectors* (ILD) used in the PeMS system [4] and *in-vehicle transponders* (IVTs) such as FasTrak. Recently, the available data on traffic has increased tremendously since the development of cellular phone based highway traffic monitoring. With the cellular phone communication infrastructure in place and privacy aware smartphone sensing technology in full expansion [15], a large volume of data from mobile devices is now available [14]. Large scale Applications include traffic flow estimation to assimilate velocity measurements [22], [23], which is a rapidly expanding field at the heart of mobile internet services. This points out on the necessity of powerful statistical filters and algorithms to efficiently assimilate the measurements.

In [22], [23] the *Ensemble Kalman Filter* is used to assimilate velocity measurements. In [19], a switching-mode model (SMM) has been derived from the CTM, which is a nonlinear discrete time dynamical system. This consists in switching among different sets of linear difference equations, defined as linear state-space model (SSM) or modes, combined with a hidden Markov model to describe the transitions from one mode to another. The Mixture Kalman filter algorithm [5] is employed to assimilate data in a switching state-space model. In this paper, we show that for a Daganzo-Newell fundamental diagram, the Godunov scheme applied to the LWR model (described in [7]) is a piecewise affine (PWA) dynamic system, where each affine component is a linear mode. Contrary to the SMM, where an additional statistical model, namely the hidden Markov model, is introduced, we unravel the PWA character of the original CTM.

II. Traffic flow modelling

A. The LWR Model

Lighthill and Whitham in 1955 [18] introduce a macroscopic dynamic model of traffic based on conservation of cars (II.1), using Greenshields' hypothesis [11] of a static flow/density relationship (II.2), known as the *fundamental diagram*. The model consists of the following two equations:

$$\frac{\partial \rho(x,t)}{\partial t} + \frac{\partial q(x,t)}{\partial x} = 0 \tag{II.1}$$

$$q(x,t) = Q(\rho(x,t)) \tag{II.2}$$

where $\rho(x,t)$ and $q(x,t)$ denote the density and the flow of vehicles at location x and time t respectively, and Q is the flux function which is assumed to be a function of the density only.

Equation (II.1) is the principle of conservation of mass, or in this case conservation of vehicles, from fluid dynamics. These equations can be written more compactly as:

$$\frac{\partial \rho(x,t)}{\partial t} + Q'(\rho(x,t))\frac{\partial \rho(x,t)}{\partial x} = 0 \tag{II.3}$$

This equation is commonly known as the *Lighthill-Whitham-Richards*, or LWR, model. Different fundamental diagrams have been suggested. Greenshields [11] found that freeway speed and density could be reasonably well approximated by a straight line. The expression of the velocity and the flux are then:

$$v = V_G(\rho) = v_f(1 - \frac{\rho}{\rho_{\text{jam}}}) \tag{II.4}$$

$$Q_G(\rho) = \rho V_G(\rho) = v_f(\rho - \frac{\rho^2}{\rho_{\text{jam}}}) \tag{II.5}$$

where v_f is the free flow (or maximum) velocity, and ρ_{jam} is the jam (or maximum) density. In this case, the flow is a quadratic function of the density.

Many researchers have later suggested alternative shapes that provide a better fit to the measured data. They all share the same characteristics **LWR1-6**:

LWR1. Greenshields' hypothesis of a static flow/density relationship: $q = Q(\rho(x,t))$

LWR2. $Q(0) = Q(\rho_{\text{jam}})$=0

LWR3. The continuous portions of $Q(\rho)$ are concave.

LWR4. $V(0) = v_f$, and $V(\rho_{\text{jam}}) = 0$.

LWR5. A critical density ρ_c can be defined where the maximum flow q_c is attained. Then, $Q(\rho)$ is increasing for $\rho \leq \rho_c$ and decrasing for $\rho > \rho_c$.

LWR6. The critical density ρ_c separates the fundamental diagram into two regimes: *free flow* when $\rho \leq \rho_c$ and *congestion* when $\rho > \rho_c$

Many researchers have later suggested alternative shapes that provide a better fit to the measured data. For instance, the widely used Daganzo-Newell velocity function assumes a constant velocity in free-flow and a hyperbolic velocity in congestion as shown in Figure II.1:

$$v = V_{DN}(\rho) = \begin{cases} v_f & \text{if } \rho \leq \rho_c \\ -\omega_f\left(1 - \frac{\rho_{\text{jam}}}{\rho}\right) & \text{if } \rho > \rho_c \end{cases} \tag{II.6}$$

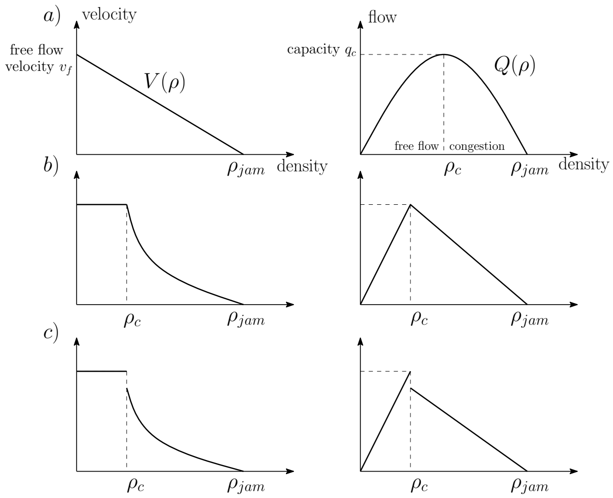

Fig. II.1: Speed and flow relationships (fundamental diagrams) for Greenshields (a), Daganzo-Newell (b), and discontinuous (c).

and the corresponding flux function is:

$$Q_{DN}(\rho) = \rho V_{DN}(\rho) = \begin{cases} v_f \rho & \text{if } \rho \leq \rho_c \\ -\omega_f \left(\rho - \rho_{\text{jam}}\right) & \text{if } \rho > \rho_c \end{cases} \tag{II.7}$$

where $\omega_f = v_f \rho_c / (\rho_{\text{jam}} - \rho_c)$ is the backwards propagation wave speed.

Measurements on the free-flow side are usually well represented by a straight line, whereas measurements in congestion tend to be more scattered. Some authors claim that there is a difference in the maximum measured flow $Q(\rho_c)$, depending on whether the freeway is in free-flow or congestion, and contend that a *discontinuity* exists at $\rho = \rho_c$ as in Figure II.1. This is described in [1], [3], [13] as a *capacity drop*, on the order of 4-10% in peak flow, as the freeway transitions into congestion.

B. Numerical Discretization

A good numerical method to solve the equations along roads is represented by the Godunov scheme, which is based on exact solutions to Riemann problems [9], [10]. This leads to the construction of a nonlinear discrete time dynamical system.

The Godunov discretization scheme is applied on the LWR PDE, where the discrete time step Δt is indexed by t, and the discrete space step Δx is indexed by i:

$$\rho_i^{t+1} = \rho_i^t - \frac{\Delta t}{\Delta x} \left(G(\rho_i^t, \rho_{i+1}^t) - G(\rho_{i-1}^t, \rho_i^t) \right) \tag{II.8}$$

In order to ensure numerical stability, the time and space steps are coupled by the CFL condition [17]: $c_{max} \frac{\Delta t}{\Delta x} \leq 1$ where c_{max} denotes the maximal characteristic speed.

For a family of flux functions $Q(\rho)$ that share the same characteristics **LWR1-6** listed above, the Godunov flux can be expressed as the minimum of the *sending flow* from the upstream cell and the *receiving flow* from the downstream cell through a boundary connecting two cells of a homogeneous road (i.e. the upstream and downstream cells have the same characteristics) [1]. The *sending flow* $S(\rho)$ is equal to the upstream flow if the upstream traffic is in free flow ($\rho \leq \rho_c$) or the capacity of the upstream section q_c if the upstream traffic is in congestion ($\rho > \rho_c$); on the other hand, the *receiving flow* $R(\rho)$ is equal to the capacity of the downstream section if the downstream traffic is in free flow or the downstream flow if the downstream traffic is in congestion. For this model, we note that for an heterogeneous segment (for instance due to a change in the number of lanes) a fundamental diagram with different parameters is defined at each cell i, consequently we add a subscript: $Q_i(\rho)$, $S_i(\rho)$, $R_i(\rho)$ and there is an implicit subscript for $G(\rho_i, \rho_{i+1})$.

$$G(\rho_1, \rho_2) = \min(S_1(\rho_1), R_2(\rho_2)) \tag{II.10}$$

$$S_1(\rho) = \begin{cases} Q_1(\rho) & \text{if } \rho \leq \rho_{c1} \\ q_{c1} & \text{if } \rho > \rho_{c1} \end{cases} \tag{II.11}$$

$$R_2(\rho) = \begin{cases} q_{c2} & \text{if } \rho \leq \rho_{c2} \\ Q_2(\rho) & \text{if } \rho > \rho_{c2} \end{cases} \tag{II.12}$$

where ρ_1 is the density of the cell upstream and ρ_2 is the density of the cell downstream. Then sending and receiving flows for the Daganzo-Newell fundamental diagram is:

$$S_1(\rho) = \begin{cases} v_{f1} \rho & \text{if } \rho \leq \rho_{c1} \\ q_{c1} & \text{if } \rho > \rho_{c1} \end{cases} \tag{II.13}$$

$$R_2(\rho) = \begin{cases} q_{c2} & \text{if } \rho \leq \rho_{c2} \\ -\omega_{f2} \left(\rho - \rho_{\text{jam2}}\right) & \text{if } \rho > \rho_{c2} \end{cases} \tag{II.14}$$

As shown in Figure II.2, the application of the Godunov scheme to the fundamental diagrams introduces intuitive concepts of *supply* and *demand* at the boundary connecting two cells. The upstream cell supplies the flow at the boundary up to capicity. We can note that in the discontinuous case, there is a drop in supply capacity when the upstream traffic is in congestion, as described in [1], [3], [13]. As a result, the flow through the boundary is smaller, even if the downstream cell can receive more flow. On the other hand, when the downstream traffic is congested, there is a decrease in demand from the downstream cell, limiting the flow through the boundary.

Important remark: For the rest of the paper, the widely-used *Cell Transmission Model* (CTM) described in [6] is

[1]There are various definitions of the Godunov flux $G(\rho_1, \rho_2)$ in the literature, notably in [8]:

$$G(\rho_1, \rho_2) = \begin{cases} \min_{\rho \in [\rho_1, \rho_2]} Q(\rho) & \text{if } \rho_1 \leq \rho_2 \\ \max_{\rho \in [\rho_2, \rho_1]} Q(\rho) & \text{if } \rho_2 \leq \rho_1 \end{cases} \tag{II.9}$$

This assumes that a fundamental diagram is defined at each boundary between two cells, which differs from the CTM.

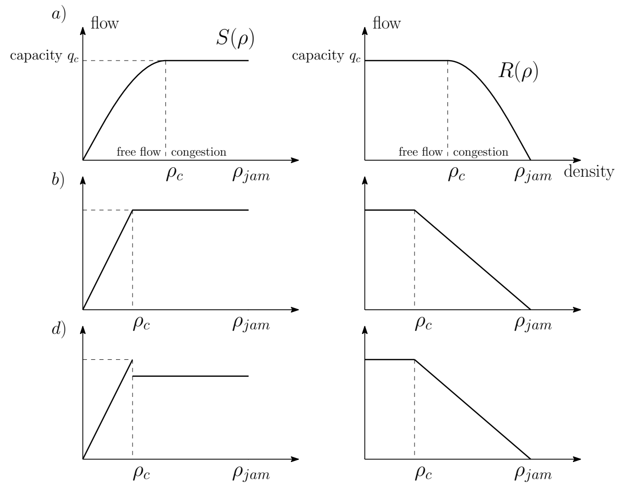

Fig. II.2: Sending and receiving flows for Greenshields (a), Daganzo-Newell (b), and discontinous (c) velocity functions.

chosen for our dynamic model and results are derived from it. We also suppose for simplicity and clarity that the segment of road we are modelling is homogeneous, i.e. the parameters of the fundamental diagram ω_f, v_f, ρ_{jam}, ρ_c, q_c are constant along the cells of the discretized road. And they are also time invariant because they are only related to the geometry of the highway, independently of the current traffic on it. All the results derived in the rest of the paper still remain for an heterogeneous road, in particular the piecewise affine character of the model and the tractibility of the Kalman filter algorithm, but the number of modes and the complexity increase. For more details on the heterogeneous case, see Appendix VIII-B.

Figure II.3 shows the explicit values taken by $G(\rho_1, \rho_2)$ for a partition of the space in different regions of the space (ρ_1, ρ_2). We will denote by **W**, **L**, and **D** the *white region*, *light region*, and *dark region* of the space (ρ_1, ρ_2) respectively.

$$G(\rho_1, \rho_2) = \begin{cases} R(\rho_2) & \text{if } (\rho_1, \rho_2) \in \mathbf{W} \\ q_c & \text{if } (\rho_1, \rho_2) \in \mathbf{L} \\ S(\rho_1) & \text{if } (\rho_1, \rho_2) \in \mathbf{D} \end{cases} \tag{II.15}$$

$$\begin{array}{ll} \mathbf{W} & = \{(\rho_1, \rho_2) \mid \rho_2 > F(\rho_1) \ , \ \rho_2 > \rho_c\} \\ \mathbf{L} & = \{(\rho_1, \rho_2) \mid \rho_1 > \rho_c \ , \ \rho_2 \leq \rho_c\} \\ \mathbf{D} & = \{(\rho_1, \rho_2) \mid \rho_2 \leq F(\rho_1) \ , \ \rho_1 \leq \rho_c\} \end{array} \tag{II.16}$$

where the boundary between the white and grey regions follows the $(\rho_1, \rho_2) = (\rho_1, F(\rho_1))$ trajectory with $F(\rho_1) = \bar{R}^{-1}(\bar{S}(\rho_1))$[2]for $\rho_1 \leq \rho_c$. $\bar{S}$ and $\bar{R}$ denote the restrictions of the sending and receiving flows to the sub-regions $[0, \rho_c)$ and $(\rho_c, \rho_{\text{jam}}]$ respectively, which also correspond to the left and right parts (w.r.t. ρ_c) of the fundamental diagram, as shown in the Figure II.3.

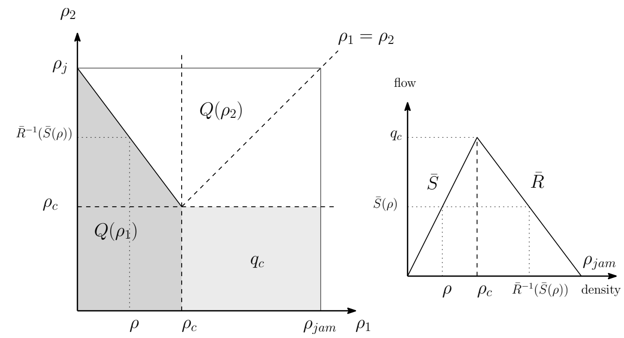

Fig. II.3: Values of $G(\rho_1, \rho_2)$ in the space (ρ_1, ρ_2).

When the velocity is the Daganzo-Newell function (II.6), the Godunov Flux becomes:

$$\begin{aligned} & G_{DN}(\rho_1, \rho_2) \\ & = \begin{cases} -\omega_f \left(\rho_2 - \rho_{\text{jam}}\right) & \text{if } (\rho_1, \rho_2) \in \mathbf{W} \\ q_c & \text{if } (\rho_1, \rho_2) \in \mathbf{L} \\ v_f \rho_1 & \text{if } (\rho_1, \rho_2) \in \mathbf{D} \end{cases} \end{aligned} \tag{II.17}$$

and the boundary between **W** and **D** regions is:

$$(\rho_1, \rho_2) = (\rho_1, -\frac{v_f}{\omega_f}\rho_1 + \rho_{\text{jam}}) \tag{II.18}$$

And **W**, **L**, **D** form a *polyhedral partition* of the space (ρ_1, ρ_2):

$$\begin{array}{ll} \mathbf{W} & = \{(\rho_1, \rho_2) \mid \rho_2 + \frac{v_f}{\omega_f}\rho_1 > \rho_{\text{jam}} \ , \ \rho_2 > \rho_c\} \\ \mathbf{L} & = \{(\rho_1, \rho_2) \mid \rho_1 > \rho_c \ , \ \rho_2 \leq \rho_c\} \\ \mathbf{D} & = \{(\rho_1, \rho_2) \mid \rho_2 + \frac{v_f}{\omega_f}\rho_1 \leq \rho_{\text{jam}} \ , \ \rho_1 \leq \rho_c\} \end{array} \tag{II.19}$$

III. A POLYHEDRAL PIECEWISE AFFINE MODEL

In the Godunov scheme (II.8), the update of the density ρ_i^{t+1} at cell i depends on the triplet $(\rho_{i-1}^t, \rho_i^t, \rho_{i+1}^t)$. In this section, we will refer to the fraction $\frac{\Delta t}{\Delta x}$ as α. Recall the Godunov scheme reads as:

$$\rho_i^{t+1} = \rho_i^t - \alpha \left(G(\rho_i^t, \rho_{i+1}^t) - G(\rho_{i-1}^t, \rho_i^t) \right) \tag{III.1}$$

A. Decomposition in different "modes"

ρ_i^{t+1} depends on whether both pairs (ρ_{i-1}^t, ρ_i^t) and (ρ_i^t, ρ_{i+1}^t) are in **W**, **L**, or **D** via $G(\rho_{i-1}^t, \rho_i^t)$ and $G(\rho_i^t, \rho_{i+1}^t)$. So there are nine possible combinations at cell i, which can be reduced to seven "modes" since the pairs (ρ_{i-1}^t, ρ_i^t) and (ρ_i^t, ρ_{i+1}^t) have ρ_i^t in common. Let's denote by $f(\rho_{i-1}^t, \rho_i^t, \rho_{i+1}^t)$ and $f_{DN}(\rho_{i-1}^t, \rho_i^t, \rho_{i+1}^t)$ the vector functions for ρ_i^{t+1} for the general and the Danganzo-Newell cases respectively, which variables are ρ_{i-1}^t, ρ_i^t, and ρ_{i+1}^t. Table III.1 list these seven possibilities, which can be easily derived from Figure II.3.

[2]Here, we formulate the more general case for equations (II.15, II.16) and we suppose that $\bar{R}$ is a strictly monotonic function on $(\rho_c, \rho_j]$, hence invertible, and $\bar{R}^{-1}$ denotes its inverse, which is the case for the Daganzo-Newell fundamental diagram.

Mode	(ρ_{i-1}^t, ρ_i^t)	(ρ_i^t, ρ_{i+1}^t)	$f(\rho_{i-1}^t, \rho_i^t, \rho_{i+1}^t)$
1	W	W	$\rho_i^t - \alpha(R(\rho_{i+1}^t) - R(\rho_i^t))$
2	W	L	$\rho_i^t - \alpha(q_c - R(\rho_i^t))$
3	L	W	$\rho_i^t - \alpha(R(\rho_{i+1}^t) - q_c)$
4	L	D	$\rho_i^t - \alpha(S(\rho_i^t) - q_c)$
5	D	W	$\rho_i^t - \alpha(R(\rho_{i+1}^t) - S(\rho_{i-1}^t))$
6	D	L	$\rho_i^t - \alpha(q_c - S(\rho_{i-1}^t))$
7	D	D	$\rho_i^t - \alpha(S(\rho_i^t) - S(\rho_{i-1}^t))$

TABLE III.1: Different values of $\rho_i^{t+1} = f(\rho_{i-1}^t, \rho_i^t, \rho_{i+1}^t)$ depending on the values of $G(\rho_{i-1}^t, \rho_i^t)$ and $G(\rho_i^t, \rho_{i+1}^t)$ in the space (ρ_1, ρ_2).

Mode	$f_{DN}(\rho_{i-1}^t, \rho_i^t, \rho_{i+1}^t)$
1	$(1-\alpha\omega_f)\rho_i^t + \alpha\omega_f\rho_{i+1}^t$
2	$(1-\alpha\omega_f)\rho_i^t + \alpha\omega_f\rho_c$
3	$\rho_i^t + \alpha\omega_f\rho_{i+1}^t - \alpha\omega_f\rho_c$
4	$(1-\alpha v_f)\rho_i^t + \alpha v_f\rho_c$
5	$\alpha v_f\rho_{i-1}^t + \rho_i^t + \alpha\omega_f\rho_{i+1}^t - \alpha\omega_f\rho_{\text{jam}}$
6	$\alpha v_f\rho_{i-1}^t + \rho_i^t - \alpha v_f\rho_c$
7	$\alpha v_f\rho_{i-1}^t + (1-\alpha v_f)\rho_i^t$

TABLE III.2: Different values of $\rho_i^{t+1} = f(\rho_{i-1}^t, \rho_i^t, \rho_{i+1}^t)$ depending on the values of $G(\rho_{i-1}^t, \rho_i^t)$ and $G(\rho_i^t, \rho_{i+1}^t)$ in the space (ρ_1, ρ_2).

For instance, for the first mode, (ρ_{i-1}^t, ρ_i^t) and (ρ_i^t, ρ_{i+1}^t) are both in **W** (see Figure II.3), thus $G(\rho_{i-1}^t, \rho_i^t) = R(\rho_i^t)$ and $G(\rho_i^t, \rho_{i+1}^t) = R(\rho_{i+1}^t)$, and then $\rho^{t+1} = f_1(\rho_{i-1}^t, \rho_i^t, \rho_{i+1}^t) = \rho_i^t - \alpha(R_+(\rho_{i+1}^t) - R(\rho_i^t))$, where f_1 is the first entry of f. By extending this result to an entire link with discrete state space indexed by $i = 1, \cdots, n$, where n is the number of space steps, we have a whole description of the space of "modes" along the link.

Remark: A priori, the number of modes in Table III.1 renders the approach of mode decomposition for estimation untractable: for n cells, the number of possible modes at any given time is technically 7^n. Since there is a correlation between two consecutive indices i and $i+1$, the number of modes for the entire link reduces from 3^n to an expression in the form of $a \cdot \beta^n + b \cdot \gamma^n + c \cdot \delta^n$ which lower and upper bounds are proved to be $3 \cdot 2^n$ and $3 \cdot (2.5)^n$ respectively (for full details, see Appendix VIII-A.4). And we will see later in section IV-C that the implementation of the Kalman filter in each mode has $O(n^2)$ time complexity and $O(n)$ space complexity.

We define J, the Jacobian matrix of f with respect to $(\rho_{i-1}^t, \rho_i^t, \rho_{i+1}^t)$ in each of the modes (which are all linear):

$$J = \left(\frac{\partial f_j}{\partial \rho_k}\right)_{j=1,\cdots,7, k=i-1,i,i+1} \tag{III.2}$$

Where f_j is the j-th entry of the vector function f defined in Table III.1. It is useful to make explicit the Jacobian matrix J_{DN} of the vector function f_{DN} with respect to $(\rho_{i-1}^t, \rho_i^t, \rho_{i+1}^t)$, and the constant term w:

$$J_{DN} = \begin{pmatrix} 0 & 1-\alpha\omega_f & \alpha\omega_f \\ 0 & 1-\alpha\omega_f & 0 \\ 0 & 1 & \alpha\omega_f \\ 0 & 1-\alpha v_f & 0 \\ \alpha v_f & 1 & \alpha\omega_f \\ \alpha v_f & 1 & 0 \\ \alpha v_f & 1-\alpha v_f & 0 \end{pmatrix} \tag{III.3}$$

$$w = \begin{pmatrix} 0 \\ \alpha\omega_f\rho_c \\ -\alpha\omega_f\rho_c \\ \alpha v_f\rho_c \\ -\alpha\omega_f\rho_{\text{jam}} \\ -\alpha v_f\rho_c \\ 0 \end{pmatrix} \tag{III.4}$$

Since f_{DN} is a *linear function* of $(\rho_{i-1}^t, \rho_i^t, \rho_{i+1}^t)$ as shown in Table III.1, we can notice that J_{DN} is constant. More notably, the vector function f_{DN} can be rewritten as:

$$\rho_i^{t+1} = f_{DN}(\rho_{i-1}^t, \rho_i^t, \rho_{i+1}^t) = J_{DN}\begin{pmatrix} \rho_{i-1}^t \\ \rho_i^t \\ \rho_{i+1}^t \end{pmatrix} + w \tag{III.5}$$

In the next section, we will see that the decomposition in "modes" as shown in Table III.1 leads to a piecewise affine formulation of the Godunov scheme in the case of the Daganzo-Newell fundamental diagram.

B. Polyhedral piecewise affine formulation

Let us consider a link with discrete time step indexed by $t \geq 0$ and discrete space step indexed by $i = 1, \cdots, n$, and let's denote $\boldsymbol{\rho}^t = (\rho_0^t, \rho_1^t, \cdots, \rho_n^t, \rho_{n+1}^t)$ a $n+2$ dimensional vector which describes the state of the link at time t in the space $\mathcal{S} = [0, \rho_{\text{jam}}]^{n+2}$. ρ_i^t is the density at time t and cell i. We can note that the ghost cells 0 and $n+1$ are included in the state of the link[3].

Definition of the space of modes: Let us denote by $\mathcal{M}_n$ the space of modes ($\mathcal{M}_n \subset \{1, \cdots, 7\}^n$). For $\boldsymbol{m} \in \mathcal{M}_n$, $\boldsymbol{m}$ is a vector of dimension n for which the i-th entry $m_i \in \{1, \cdots, 7\}$ is the mode at cell i. Equivalently, each element of $\mathcal{M}_n$ can be described as a sequence of regions in which the pair (ρ_i, ρ_{i+1}) is, for $i = 0, \cdots, n$. Hence, we define the equivalent space of modes $\tilde{\mathcal{M}}_n \subset \{\text{w,l,d}\}^{n+1}$, and for $\boldsymbol{s} \in \tilde{\mathcal{M}}_n$, $\boldsymbol{s}$ is a vector of dimension $n+1$ for which the i-th entry $s_i \in \{\text{w,l,d}\}$ is the region of the pair (ρ_i, ρ_{i+1}), for $i = 0, \cdots, n$. As we will see later, this second definition gives a description of the *partition of the space $\mathcal{S}$ into different polyhedra $\mathbf{P}_{\boldsymbol{m}}$ in which the mode is $\boldsymbol{m}$.*

[3]The values of ρ_0^t and ρ_{n+1}^t are given by the prescibed boundary conditions to be imposed on the in left and right side of the domain respectively. Note that these boundary values do not always affect the physical domain because of the nonlinear operator II.17, which causes the boundary conditions to be implemented in the weak sense. For more details, see [21].

The n-dimensional vector $\boldsymbol{m} \in \mathcal{M}_n$ describes the mode of the link at amy time, as defined in the previous section. At each time increment, the state of the link is updated through the following nonlinear dynamical system:

$$\boldsymbol{\rho}^{t+1} = F[\boldsymbol{\rho}^t] \tag{III.6}$$

with $F[\cdot]$ a n+2 dimensional function vector. When $\boldsymbol{\rho}^t \in \boldsymbol{P_m}$ (i.e. the mode at time t is $\boldsymbol{m}$), the i-th entry $\rho_i^{t+1} = F_i[\boldsymbol{\rho}^t]$ is:

$$\rho_i^{t+1} = \begin{cases} f_{m_i}(\rho_{i-1}^t, \rho_i^t, \rho_{i+1}^t) & \text{for} \quad i = 1, ..., n \\ u^{t+1} & \text{for} \quad i = 0 \\ d^{t+1} & \text{for} \quad i = n+1 \end{cases} \tag{III.7}$$

where m_i denotes the i-th entry of $\boldsymbol{m} \in \mathcal{M}_n$, i.e. the mode of cell i at time step t, u^{t+1} and d^{t+1} the boundary conditions upstream and downstream at time step $t+1$. $f_{m_i}(\rho_{i-1}^t, \rho_i^t, \rho_{i+1}^t)$ is the m_i-th entry of the function vector f evaluated in $(\rho_{i-1}^t, \rho_i^t, \rho_{i+1}^t)$. We note that $\rho_0^{t+1} = u^{t+1}$ and $\rho_{n+1}^{t+1} = d^{t+1}$, which means that the ghost cells are the boundary conditions of the CTM. For a Daganzo-Newell fundamental diagram the update operator of the dynamical system is:

$$\rho_i^{t+1} = \begin{cases} L_{m_i} \cdot \begin{pmatrix} \rho_{i-1}^t \\ \rho_i^t \\ \rho_{i+1}^t \end{pmatrix} + w_{m_i} & \text{for } i = 1, \cdots, n \\ u^{t+1} & \text{for } i = 0 \\ d^{t+1} & \text{for } i = n+1 \end{cases} \tag{III.8}$$

Where L_{m_i} is the m_i-th line of J_{DN} and w_{m_i} the m_i-th entry of w. With matrix notations:

$$\boldsymbol{\rho}^{t+1} = A_{\boldsymbol{m}} \boldsymbol{\rho}^t + b_{\boldsymbol{m}} + c_{t+1} \quad \text{if} \quad \boldsymbol{\rho}^t \in \mathbf{P}_{\boldsymbol{m}} \tag{III.9}$$

Where $A_{\boldsymbol{m}}$ is a tridiagonal matrix of size $(n+2) \times (n+2)$, such that the diagonal elements are $\{0, J_{m_1,2}, ..., J_{m_n,2}, 0\}$, the lower diagonal elements are $\{J_{m_1,1}, J_{m_2,1}, ..., J_{m_n,1}, 0\}$, and the upper diagonal elements are $\{0, J_{m_1,3}, J_{m_2,3}, ..., J_{m_n,3}\}$ where J (or J_{DN}) are defined in equations (III.2), (III.3). Equivalently:

$$A_{\boldsymbol{m}} = \begin{pmatrix} 0 & \cdots & 0 \\ L_{m_1} & & \\ & \ddots & \\ & & L_{m_n} \\ 0 & \cdots & 0 \end{pmatrix} \tag{III.10}$$

$b_{\boldsymbol{m}}$ and c_{t+1} are two vectors of dimension $(n+2)$ with entries $\{0, w_{m_1^t}, \cdots, w_{m_n^t}, 0\}$ and $\{u^{t+1}, 0, \cdots, 0, d^{t+1}\}$ repectively, and $\mathbf{P}_{\boldsymbol{m}}$ is the subset of space $\mathcal{S}$ where the mode is $\boldsymbol{m}$. We provide now a description of the partition of the space into the polyhedra $\mathbf{P}_{\boldsymbol{m}}$ in which the mode is $\boldsymbol{m}$. See appendix VIII-A.1 for details on polyhedra and their representation.

Polyhedral partition of the space: For a discretization into n cells, we chose to describe the ensemble of modes $\tilde{\mathcal{M}}_n$ in sequences $\boldsymbol{s} \in \{w, l, d\}^{n+1}$ and define $\mathbf{P}_{\boldsymbol{s}}$ the corresponding polyhedron for each sequence. Let us define 3^{n+1} polyhedra $\mathbf{W}_i, \mathbf{L}_i, \mathbf{D}_i$ for $i = 0, \cdots, n$ in the space $\mathcal{S}$:

$$\begin{aligned} \mathbf{W}_i &= \{(\rho_i, \rho_{i+1}) \mid \rho_{i+1} + \tfrac{v_f}{\omega_f}\rho_i > \rho_{\text{jam}} \ , \ \rho_{i+1} > \rho_c\} \\ \mathbf{L}_i &= \{(\rho_i, \rho_{i+1}) \mid \rho_i > \rho_c \ , \ \rho_{i+1} \leq \rho_c\} \\ \mathbf{D}_i &= \{(\rho_i, \rho_{i+1}) \mid \rho_{i+1} + \tfrac{v_f}{\omega_f}\rho_i \leq \rho_{\text{jam}} \ , \ \rho_i \leq \rho_c\} \end{aligned} \tag{III.11}$$

Each mode or possible sequence of regions $\boldsymbol{s} \in \mathcal{M}_n$ is valid in a polyhedron $\mathbf{P}_{\boldsymbol{s}}$ that is the intersection of $n+1$ polyhedra:

$$\mathbf{P}_{\boldsymbol{m}} = \bigcap_{i=0}^{n} \mathbf{Q}_i \tag{III.12}$$

where the polyhedra $\mathbf{Q}_i$ are

$$\mathbf{Q}_i = \begin{cases} \mathbf{W}_i & \text{if } s_i = w \\ \mathbf{L}_i & \text{if } s_i = l \\ \mathbf{D}_i & \text{if } s_i = d \end{cases} \tag{III.13}$$

Moreover, for two different modes $\boldsymbol{s}$ and $\boldsymbol{s}'$, and corresponding polyhedra $\mathbf{P}_{\boldsymbol{s}} = \bigcap_{i=0}^{n} \mathbf{Q}_i$ and $\mathbf{P}_{\boldsymbol{s}'} = \bigcap_{i=0}^{n} \mathbf{Q}'_i$, we can find an index i for which $\mathbf{Q}_i$ and $\mathbf{Q}'_i$ are disjoint. For instance, suppose without loss of generality that $\mathbf{Q}_i = \mathbf{W}_i$ and $\mathbf{Q}'_i = \mathbf{D}_i$, and we know that $\mathbf{W}_i$ and $\mathbf{D}_i$ are disjoint. Then in this case, the hyperplan $\{\boldsymbol{\rho} \mid \rho_{i+1} + \frac{v_f}{\omega_f}\rho_i = \rho_{\text{jam}}\}$ is a seperating hyperplan between $\mathbf{P}_{\boldsymbol{s}}$ and $\mathbf{P}_{\boldsymbol{s}'}$. Hence, $\mathbf{P}_{\boldsymbol{s}}$ and $\mathbf{P}_{\boldsymbol{s}'}$ are disjoint and the family $\{\mathbf{P}_{\boldsymbol{s}}\}_{\boldsymbol{s} \in \mathcal{M}_n}$ is a partition of $\mathcal{M}_n$.

IV. Extended Kalman filter

The Extended Kalman filter provides the state estimate $\hat{\boldsymbol{\rho}}^t$ as a gaussian distribution with mean $\boldsymbol{\mu}^t$ and covariance P_t given the sequence of observations $\boldsymbol{z}^{0:t}$, and sequence of control parameters $c^{0:t}$. We present here an algorithm for the implementation of the Extended Kalman filter to the Cell Transmission Model with n cells, which is a piecewise affine model we have seen earlier. Note that the state at time t is $\boldsymbol{\rho}^t = (\rho_0^t, \rho_1^t, ..., \rho_n^t, \rho_{n+1}^t)$, a vector of dimension $n+2$ that includes the two ghost cells 0 and $n+1$ which are the boundary conditions. When the state is in mode $\boldsymbol{m}^t$, this boils down to applying the coresponding Kalman filter with the update matrix $A_{\boldsymbol{m}}$ defined in III.10.

A. Mode detection

We present here a simple procedure to detect the mode of the state $\boldsymbol{\rho}$. It relies on the definition of polyhedra as a finite number of half-spaces (see Appendix VIII-A.1). For a state $\boldsymbol{\rho} = (\rho_0, \rho_1, ..., \rho_n, \rho_{n+1})$, a n+2 dimensional vector which describes the state of the link in the space $\mathcal{S} = [0, \rho_j]^{n+2}$, we present an algorithm that detects the mode of the state $\boldsymbol{m}$ (or $\boldsymbol{s}$) and provides a minimal H-representation of the polyhedron $P_{\boldsymbol{m}}$ (or $P_{\boldsymbol{s}}$). We first introduce the following indicator functions:

$$\begin{aligned}
&\alpha_i(\boldsymbol{\rho}) = 1_{\{\rho_{i+1}+\frac{v_f}{\omega_f}\rho_i>\rho_{\text{jam}}\}} \\
&\beta_i(\boldsymbol{\rho}) = 1_{\{\rho_i>\rho_c\}} \qquad \text{for } i=0,1,...,n \\
&\gamma_i(\boldsymbol{\rho}) = 1_{\{\rho_{i+1}>\rho_c\}}
\end{aligned} \tag{IV.1}$$

and we note $\mathbf{H}_{\alpha_i}$, $\mathbf{H}_{\beta_i}$, and $\mathbf{H}_{\gamma_i}$ the corresponding half-spaces. The dual half-spaces $\mathcal{S}\backslash\mathbf{H}$ are denoted by $\mathbf{H}^d_{\alpha_i}$, $\mathbf{H}^d_{\beta_i}$, and $H^d_{\gamma_i}$ and the corresponding indicator functions are $1-\alpha_i(\boldsymbol{\rho})$, $1-\beta_i(\boldsymbol{\rho})$, and $1-\gamma_i(\boldsymbol{\rho})$. We can notice that $\beta_{i+1}(\boldsymbol{\rho})=\gamma_i(\boldsymbol{\rho})$. Since we have:

$$\begin{aligned}
\mathbf{W}_i &= \mathbf{H}_{\alpha_i}\cap\mathbf{H}_{\gamma_i} \\
\mathbf{L}_i &= \mathbf{H}_{\beta_i}\cap\mathbf{H}^d_{\gamma_i} \\
\mathbf{D}_i &= \mathbf{H}^d_{\alpha_i}\cap\mathbf{H}^d_{\beta_i}
\end{aligned} \tag{IV.2}$$

for the polyhedra defined in (III.11)), their indicator functions are:

$$\begin{aligned}
&w_i(\boldsymbol{\rho}) = \alpha_i(\boldsymbol{\rho})\gamma_i(\boldsymbol{\rho}) \\
&l_i(\boldsymbol{\rho}) = \beta_i(\boldsymbol{\rho})(1-\gamma_i(\boldsymbol{\rho})) \qquad \text{for } i=0,1,...,n \\
&d_i(\boldsymbol{\rho}) = (1-\alpha_i(\boldsymbol{\rho}))(1-\beta_i(\boldsymbol{\rho}))
\end{aligned} \tag{IV.3}$$

Hence, evaluating the indicator functions $\alpha_i(\boldsymbol{\rho})$, $\beta_i(\boldsymbol{\rho})$, and $\gamma_i(\boldsymbol{\rho})$ for $i=0,...,n$ gives the mode $\boldsymbol{m}$ of state $\boldsymbol{\rho}$. Equations (III.12, III.13, IV.2) give an H-representation of $P_{\boldsymbol{s}}$ (see appendix VIII-A.1 for a formal definition of an H-representation).

B. Kalman filter algorithm

In order to use the *Kalman filter* to estimate the state of the link given a sequence of noisy observations, we model the process by adding a white noise to the underlying dynamic system model. The "true" state at time $t+1$, namely $\boldsymbol{\rho}^{t+1}$, is then:

$$\boldsymbol{\rho}^{t+1} = A_{\boldsymbol{m}}\boldsymbol{\rho}^t + b_{\boldsymbol{m}} + c_{t+1} + \boldsymbol{\eta}^{t+1} \quad \text{if} \quad \boldsymbol{\rho}^t\in\mathbf{P}_{\boldsymbol{m}} \tag{IV.4}$$

where $\boldsymbol{\eta}^t\sim N(0,Q_t)$ is the Gaussian zero-mean, white state noise with covariance Q_t. To apply the *control update* of the Kalman filter, it is then necessary to know the mode $\boldsymbol{m}$ of the state $\boldsymbol{\rho}^t$ (i.e. $\boldsymbol{m}$ such that $\mathbf{P}_{\boldsymbol{m}}$).

Additionally, the observation model for the link is given by:

$$\boldsymbol{y}^t = H_t\boldsymbol{\rho}^t + \boldsymbol{\chi}^t \tag{IV.5}$$

where $H_t\in\{0,1\}^{p_t\times n}$ is the linear observation observation matrix which encodes the p_t observations (each one of them being at a discrete cell on the highway) for which the density is observed during discrete time step t, and n is the number of cells along the link. The last term in (IV.5) is the white, zero mean observation noise $\boldsymbol{\chi}^t\sim N(0,R_t)$ with covariance matrix R_t.

Let $\boldsymbol{\mu}^t$ and P_t be the state estimate and the error covariance matrix at time t. Then the *predicted state estimate* $\boldsymbol{\mu}^{t+1:t}$ and *covariance estimate* $P_{t+1:t}$ of the *prediction step* are:

$$\begin{aligned}
&\boldsymbol{\mu}^{t+1:t} = A_{\boldsymbol{m}}\boldsymbol{\mu}^t + b_{\boldsymbol{m}} + c_{t+1} \text{ if } \quad \boldsymbol{\mu}^t\in\mathbf{P}_{\boldsymbol{m}} \\
&P_{t+1:t} = A_{\boldsymbol{m}}P_t(A_{\boldsymbol{m}})^T + Q_t
\end{aligned} \tag{IV.6}$$

The *measurement residual* $\boldsymbol{r}_{t+1}$, *residual covariance* S_{t+1}, *Kalman gain* K_{t+1}, *updated state estimate* $\boldsymbol{\mu}^{t+1}$, and *updated estimate covariance* P_{t+1} of the *update step* are:

$$\begin{aligned}
&\boldsymbol{r}_{t+1} = \boldsymbol{z}^{t+1} - H_{t+1}\boldsymbol{\mu}^{t+1:t} \\
&S_{t+1} = H_{t+1}P_{t+1:t}H^T_{t+1} + R_{t+1} \\
&K_{t+1} = P_{t+1:t}H^T_{t+1}S^{-1}_{t+1} \\
&\boldsymbol{\mu}^{t+1} = \boldsymbol{\mu}^{t+1:t} + K_{t+1}\boldsymbol{r}_{t+1} \\
&P_{t+1} = (I-K_{t+1}H_{t+1})P_{t+1:t}
\end{aligned} \tag{IV.7}$$

C. Implementation and complexity

Since the number of modes grows exponentially as the number of cells increase s (see Appendix VIII-A.4), it is computationally expensive to store a matrix $A_{\boldsymbol{m}}$ for each mode $\boldsymbol{m}$. Fortunately, it is possible to compute the *predicted state estimate* $\boldsymbol{\mu}^{t+1:t}$ and the *predicted covariance estimate* $P_{t+1:t}$ in linear time and quadratic time respectively, without forming any dense matrix $A_{\boldsymbol{m}}$. This relies on the tridiagonality of $A_{\boldsymbol{m}}$ and the homogeneity of the segment of road considered, which requires to store only the seven possible modes at each cell[4].

In particular, equation (III.8) gives a simple procedure to compute $\boldsymbol{\mu}^{t+1:t}$ in linear time from $\boldsymbol{\mu}^t$, J_{DN}, w, and c_{t+1}, by knowing $\boldsymbol{m}$ such that $\boldsymbol{\mu}^{t+1:t}\in\mathbf{P}_{\boldsymbol{m}}$:

$$\rho_i^{t+1} = \begin{cases} L_{m_i}\cdot\begin{pmatrix}\rho^t_{i-1}\\ \rho^t_i\\ \rho^t_{i+1}\end{pmatrix} + w_{m_i} & \text{for } i=1,...,n \\ u^{t+1} & \text{for } i=0 \\ d^{t+1} & \text{for } i=n+1\end{cases} \tag{IV.8}$$

Similarly, the double product $A_{\boldsymbol{m}}P_t(A_{\boldsymbol{m}})^T$ can be computed in quadratic time from P_t and J_{DN}. With the entries of $A_{\boldsymbol{m}}P_t$ indexed from 0 to $n+1$:

$$(A_{\boldsymbol{m}}P_t)_{i,j} = \begin{cases} L_{m_i}\cdot\begin{pmatrix}p_{i-1,j}\\ p_{i,j}\\ p_{i+1,j}\end{pmatrix} & \text{if } (i,j)\in\{1,...,n\}^2 \\ 0 & \text{otherwise}\end{cases} \tag{IV.9}$$

where L_{m_i} is the m_i-th line of J_{DN}, m_i the i-th entry of $\boldsymbol{m}$, and $p_{i,j}$ is the entry (i,j)-th entry of P_{t-1}. And the computation of the second matrix multiplication with entries indexed from 0 to $n+1$ is:

[4] In the case of a heterogeneous road (i.e. a different fundamental diagram for each cell), up to all nine possible local modes for each cell have to be stored, which is still bound by $9\times n$, where n is the number of cells.

$$(A_{\boldsymbol{m}} P_t (A_{\boldsymbol{m}})^T)_{i,j} = \begin{cases} (q_{i,j-1} \quad q_{i,j} \quad q_{i,j+1}).L_{m_j}^T & \text{if } (i,j) \in \{1,...,n\}^2 \\ 0 & \text{otherwise} \end{cases} \tag{IV.10}$$

where $q_{i,j}$ is the entry (i,j)-th entry of $A_{\boldsymbol{m}} P_t$. We can note that the first line and first column of P_t have only zero elements because the boundary condition $\rho_0^t = u^t$ is deterministic (i.e. $\text{cov}(u^t, \rho_i^t) = 0$ for $i = 1, \cdots, n$), and similarly the last line and last column of P_t are null since the boundary condition $\rho_{n+1}^t = d^t$ is deteministic.

The three equations (IV.8, IV.9, IV.10) show that both time complexity and space complexity of the *prediction step* are $O(n^2)$.

TO DO: complexity of the EKF vs the EnKF vs...

D. Analysis

TO DO: Not tuned because of the discontinuities in the derivative? But should perform well when the highway is in a mode for a long time.

V. Reduced Multiple Mode Kalman filter

A. Motivation and definitions

TO DO: write about intractibility and should select a good sample of modes. So we can choose the modes based on projections on the supporting hyperplanes of the minimal H-representation, or on the adjacent polyhedra. The estimate is the weighted sum of the state in each mode chosen, where the weight is a mix of likelihood with the residual mean and covariance of the kalman filters, and the distance divided by the covariance in one direction. And we can limit the number of modes selected by choosing the mode that are within a range, where the distance is the distance divided by the covariance in one direction. The heuristic in choosing the mode is that we choose a mode with only a few changes, because the state of the highway changes only locally for one time step (time step small from the CFL). And we suppose that the most likely modes are actually well represented by the adjacent ones, or the projected ones, and their linear combination. Adjacent polyhedra, minimal H-representation, and polyhedral partition. We have a minimal H-representation of a polyhedron. When changing a half-space to its dual, we have a new polyhedron that is non-empty (CHECK). How is it compared to the polyhedral partition? In how many polyhedra of the polyhedral partition is it inculded?

Faces of a polyhedron: A *hyperplane* can be written as a linear equality:

$$a_1 x_1 + a_2 x_2 + ... + a_n x_n = b \tag{V.1}$$

where n is the dimension of the space. It divides the space in two half-spaces. And a supportive hyperplane of a closed convex set C is a hyperplane ∂H such as $C \cap \partial H \neq \emptyset$ and $C \subseteq H$, where H is one of the two half-spaces. Given a polyhedron P, the inersection with any supportive hyperplane is a face of P. Moreover, a vertex is a zero-dimension face, an edge a one-dimension face, and a facet is a face of dimension $d-1$ if P is of dimension d. For a full-dimensional polyhedron, a facet is of dimension $n+1$ (recall that the space $\mathcal{S} = [0, \rho_j]^{n+2}$ is of dimension $n+1$).

Minimal H-representation: There exist infinitely many H-descriptions of a convex polytope. However, for a full-dimensional convex polytope, the minimal H-description is in fact unique and is given by the set of the facet-defining halfspaces [12]. Appendix VIII-A.2 gives an algorithm that finds the minimal H-representation of a polyhedron of the partition of $\mathcal{S}$ in our highway model.

The polyhedra are assumed closed in the following definitions.

Adjacent polyhedra: Two polyhedra P and P' in a polyhedral partition of the space $\mathcal{S}$ are said to be *k-adjacent* if they have a face of dimension k "in common". Formally, this is when there exists a hyperplane H in common between P and P' and the intersection of P, P' and H is of dimension k.

Exclusivity: In a polyhedral partition of the space $\mathcal{S}$ of dimension d, and given two $(d-1)$-adjacent polyhedra of the partition P and P', their $(d-1)$-adjacency is *exclusive* if they are the only two polyhedra sharing the facet. Formally, for H the common supportive hyperplane of P and P', this is when $P \cap P' \subset H$ is of dimension $d-1$ and $P \cap P'' \cap H$ and $P' \cap P'' \cap H$ are both of dimension $< d-1$ for any other polyhedron P'' of the partition. In our highway model (recall that the space $\mathcal{S} = [0, \rho_j]^{n+2}$ is of dimension $n+1$) we note that all (n+1)-adjacencies in the polyhedral partition of $\mathcal{S}$ are exclusive (see appendix VIII-A.3 for a proof). Therefore, for all the hyperplanes H in the minimal H-description of a polyhedron $P_{\boldsymbol{s}}$ of the partition of $\mathcal{S}$, $\partial H = \bar{H} \backslash H^0$ are supportive hyperplanes, where each corresponding facet seperates $P_{\boldsymbol{s}}$ from one and only one $(n+1)$-adjacent polyhedron $P_{\boldsymbol{s}'}$, and we say that s' is an adjacent mode of s. Such a property enables to find easily all the adjacent polyhedra of $P_{\boldsymbol{s}}$ from its minimal H-description (see appendix VIII-A.3 for more details). TO DO: make research and check the definition.

B. Neighbor modes search

TO DO: heuristics based on projection or adjacency and comparison with classic methods in statistical learning

Given a state ρ, and the state covariance P_t, let $\boldsymbol{s}$ be the mode, $\mathbf{P}_{\boldsymbol{s}}$ the associated polyhedron, $\bigcap_{i=0}^n \mathbf{Q}_i$ its H-representation given by (III.12, III.13, IV.2), and $\bigcap_{j=0}^k \mathbf{H}_j$ its minimal H-representation, where the closed half-spaces $\bar{\mathbf{H}}_j$ can be written:

$$\bar{\mathbf{H}}_j = \{\rho \mid a_j.\rho - b_j \leq 0\} \text{ for } j = 0, ..., k \tag{V.2}$$

The euclidian distance between ρ and each of the facet $\mathbf{F}_j = (\bigcap_{j'\neq j}\mathbf{H}_{j'}) \cap \partial\mathbf{H}_j$ of $\mathbf{P}_{\boldsymbol{s}}$ i.e.

$$d(\rho,\mathbf{F}_j) = \min_{\rho'\in\mathbf{F}_j}||\rho-\rho'|| \text{ for } j=0,...,k \tag{V.3}$$

is bounded on the left by the distance between ρ and the hyperplane $\partial\mathbf{H}_j$:

$$d(\rho,\mathbf{F}_j) \geq d(\rho,\partial\mathbf{H}_j) = \frac{|b_j - a_j.\rho|}{||a_j||} \text{ for } j=0,...,k \tag{V.4}$$

and we define the ratio:

$$d_j = \frac{d(\rho,\partial\mathbf{H}_j)}{a_j^T P_t a_j} \text{ for } j=0,...,k \tag{V.5}$$

for which we only look at the adjacent modes for which d_j is less than a given threshold d. Intuitively, when there is a high variance $a_j^T P_t a_j$ along the direction a_j orthogonal to $\partial\mathbf{H}_j$, there is a higher probability that the state at the next time step is in the half-space $\mathbf{H}_j^d$, which is the dual of $\mathbf{H}_j$, and therefore in the adjacent mode $\boldsymbol{s}_j$ with common supportive hyperplane $\partial\mathbf{H}_j$. We note that this is an approximation since the projection of ρ on $\partial\mathbf{H}_j$ is not always on a facet of the adjacent mode $\boldsymbol{s}_j$.

Given $j\in\{0,...,k\}$, there exists $i'\in\{0,...,n\}$ such that

$$\begin{array}{l}\mathbf{Q}_{i'} = \mathbf{H}_j \cap \mathbf{H}_{i'}\\ \mathbf{H}_j, \mathbf{H}_{i'} \in \{\mathbf{H}_{\alpha_{i'}}, \mathbf{H}^d_{\alpha_{i'}}, \mathbf{H}_{\beta_{i'}}, \mathbf{H}^d_{\beta_{i'}}, \mathbf{H}_{\gamma_{i'}}, \mathbf{H}^d_{\gamma_{i'}}\}\end{array} \tag{V.6}$$

and we can see from figure II.3 that there exists $\mathbf{R}_{i'} \in \{\mathbf{W}_{i'}, \mathbf{L}_{i'}, \mathbf{D}_{i'}\}$ defined in (III.11) different from $\mathbf{Q}_{i'}$ such that

$$\mathbf{H}^d_j \cap \mathbf{H}_{i'} \subset \mathbf{R}_{i'} \tag{V.7}$$

Let $\mathcal{I}_j$ be the set of all $i'\in\{0,...,n\}$ such that we have (V.6, V.7), then the only adjacent mode $\boldsymbol{s}_j$ with supportive hyperplane $\mathbf{H}_j$ is:

$$\mathbf{P}_{\boldsymbol{s}_j} = (\bigcap_{i\notin\mathcal{I}_j}\mathbf{Q}_i) \cap (\bigcap_{i'\in\mathcal{I}_j}\mathbf{R}_{i'}) \tag{V.8}$$

The adjacent mode vector $\boldsymbol{s}_j$ is obtained by re-evaluating all the indicator functions $\delta_{i'}(\boldsymbol{\rho})$ (defined in IV.1) associated to $\mathbf{H}_j$ for all $i'\in\mathcal{I}_j$.

Example: If $\mathbf{H}_j = \mathbf{H}_{\gamma_{i'}} = \mathbf{H}_{\beta_{i'+1}} = \{\boldsymbol{\rho} \mid \rho_{i'+1} > \rho_c\}$ and $d_j \leq d$ then $\gamma_{i'}(\boldsymbol{\rho}) = 1$ since $\boldsymbol{\rho}\in\mathbf{P}_{\boldsymbol{s}}\subset\mathbf{H}_j$. We have $\mathcal{I}_j = \{i', i'+1\}$ since $\mathbf{Q}_{i'}$, $\mathbf{Q}_{i'+1}$ are the *only affected* polyhedra in the H-representation of $\mathbf{P}_{\boldsymbol{s}}$ given by (III.12, III.13, IV.2), when $\mathbf{H}_j$ is changed to its dual $\mathbf{H}^d_j = \mathcal{S}\backslash\mathbf{H}_j$ in the definition of the adjacent polyhedron $\mathbf{P}_{\boldsymbol{s}_j}$. For $\mathbf{H}_{i'} \in \{\mathbf{H}_{\alpha_{i'}}, \mathbf{H}^d_{\alpha_{i'}}, \mathbf{H}_{\beta_{i'}}, \mathbf{H}^d_{\beta_{i'}}\}$, and $\mathbf{H}_{i'+1} \in \{\mathbf{H}_{\alpha_{i'+1}}, \mathbf{H}^d_{\alpha_{i'+1}}, \mathbf{H}_{\gamma_{i'+1}}, \mathbf{H}^d_{\gamma_{i'+1}}\}$ such that $\mathbf{Q}_{i'}, \mathbf{Q}_{i'+1}$ can be decomposed in this way (following definitions (III.12, III.13, IV.2)):

$$\begin{array}{l}\mathbf{Q}_{i'} = \mathbf{H}_{i'} \cap \mathbf{H}_j\\ \mathbf{Q}_{i'+1} = \mathbf{H}_{i'+1} \cap \mathbf{H}_j\end{array} \tag{V.9}$$

and the corresponding indicator functions take the values:

$$\begin{array}{l}w_{i'}(\boldsymbol{\rho}) = \alpha_{i'}(\boldsymbol{\rho})\gamma_{i'}(\boldsymbol{\rho}) = \alpha_{i'}(\boldsymbol{\rho})\\ l_{i'}(\boldsymbol{\rho}) = \beta_{i'}(\boldsymbol{\rho})(1-\gamma_{i'}(\boldsymbol{\rho})) = 0\\ l_{i'+1}(\boldsymbol{\rho}) = \beta_{i'+1}(\boldsymbol{\rho})(1-\gamma_{i'+1}(\boldsymbol{\rho})) = 1-\gamma_{i'+1}(\boldsymbol{\rho})\\ d_{i'+1}(\boldsymbol{\rho}) = (1-\alpha_{i'+1}(\boldsymbol{\rho}))(1-\beta_{i'+1}(\boldsymbol{\rho})) = 0\end{array} \tag{V.10}$$

There exist $\mathbf{R}_{i'} \in \{\mathbf{W}_{i'}, \mathbf{L}_{i'}, \mathbf{D}_{i'}\}$ and $\mathbf{R}_{i'+1} \in \{\mathbf{W}_{i'+1}, \mathbf{L}_{i'+1}, \mathbf{D}_{i'+1}\}$ defined in (III.11) such that

$$\begin{array}{l}\mathbf{H}^d_j \cap \mathbf{H}_{i'} \subset \mathbf{R}_{i'}\\ \mathbf{H}^d_j \cap \mathbf{H}_{i'+1} \subset \mathbf{R}_{i'+1}\end{array} \tag{V.11}$$

Then the adjacent polyhedron $\mathbf{P}_{\boldsymbol{s}_j}$ is given by:

$$\mathbf{P}_{\boldsymbol{s}_j} = (\bigcap_{i\notin\{i',i'+1\}}\mathbf{Q}_i) \cap (\mathbf{R}_{i'} \cap \mathbf{R}_{i'+1}) \tag{V.12}$$

and the indicator functions of $\mathbf{P}_{\boldsymbol{s}_j}$ have the same values as $\mathbf{P}_{\boldsymbol{s}}$ except for these ones, where $\gamma_{i'}(\boldsymbol{\rho}) = \beta_{i'+1}(\boldsymbol{\rho}) = 1$ is changed to $\gamma_{i'}(\boldsymbol{\rho}) = \beta_{i'+1}(\boldsymbol{\rho}) = 0$:

$$\begin{array}{l}w_{i'}(\boldsymbol{\rho}) = \alpha_{i'}(\boldsymbol{\rho})\gamma_{i'}(\boldsymbol{\rho}) = 0\\ l_{i'}(\boldsymbol{\rho}) = \beta_{i'}(\boldsymbol{\rho})(1-\gamma_{i'}(\boldsymbol{\rho})) = \beta_{i'}(\boldsymbol{\rho})\\ l_{i'+1}(\boldsymbol{\rho}) = \beta_{i'+1}(\boldsymbol{\rho})(1-\gamma_{i'+1}(\boldsymbol{\rho})) = 0\\ d_{i'+1}(\boldsymbol{\rho}) = (1-\alpha_{i'+1}(\boldsymbol{\rho}))(1-\beta_{i'+1}(\boldsymbol{\rho})) = 1-\alpha_{i'+1}(\boldsymbol{\rho})\end{array} \tag{V.13}$$

As the example shows, only the i'-th and (i'+1)-th entries of $\boldsymbol{s}_j$ differ from those of $\boldsymbol{s}$, and only two consecutive entries differ between two adjacent modes $\boldsymbol{s}$ and $\boldsymbol{s}'$ in the general case.

C. Prior gaussian distribution

TO DO: we get a mixture of gaussians that is approximates by a gaussian. comparison with the gaussian distribution given by the EnKF. Add references on the IMM, justify the formula of the weights cf. formula (3, 4, 5) in the state estimation for hybrid systems: applications to aircraft tracking.

Let $\boldsymbol{\mu}^{t-1}$ and P_{t-1} be the state estimate and the error covariance matrix at time $t-1$, $\boldsymbol{s}$ the mode of $\boldsymbol{\mu}^{t-1}$ (i.e. $\boldsymbol{\mu}^{t-1}\in\mathbf{P}_{\boldsymbol{s}}$), $\bigcap_{j=0}^k\mathbf{H}_j$ the minimal H-representation of $\mathbf{P}_{\boldsymbol{s}}$, $\boldsymbol{s}_j$ for $j\in\{0,...,k\}$ the adjacent modes defined by (V.6, V.7, V.8), and d_j the ratio defined in (V.5). We look at the adjacent modes such that $d_j\leq d$ where d is a threshold and apply the Kalman filter to each one of them. The *predicted state estimate* $\boldsymbol{\mu}_j(t:t-1)$ and *predicted covariance estimate* $P_j(t:t-1)$ in mode j are:

$$\begin{array}{l}\boldsymbol{\mu}_j(t:t-1) = A_{\boldsymbol{s}_j}\boldsymbol{\mu}^{t-1} + b_{\boldsymbol{s}_j} + c_t\\ P_j(t:t-1) = A_{\boldsymbol{s}_j}P_{t-1}(A_{\boldsymbol{s}_j})^T + Q_{t-1}\end{array} \tag{V.14}$$

The distribution of the predicted state is modeled as a mixture of Gaussian distributions, where each component $x_j = \mathcal{N}(\boldsymbol{\mu}_j(t:t-1),\ P_j(t:t-1))$ is the distribution of the predicted state in mode j:

$$x = \frac{1}{W_t}\sum_{j|d_j\leq d} w_j^t x_j \tag{V.15}$$

where $W_t = \sum_{j|d_j \le d} w_j^t$, and the weights w_j^t are the likelihood function fo mode j:

$$w_j^t = \mathcal{N}(\boldsymbol{r}_j(t);\ 0,\ S_j(t)) \tag{V.16}$$

and $\boldsymbol{r}_j(t)$ is the residual produced by the Kalman filter j, and $S_j(t)$ the corresponding residual covariance:

$$\begin{aligned} \boldsymbol{r}_j(t) &= \boldsymbol{z}^t - H_t \boldsymbol{\mu}_j(t:t-1) \\ S_j(t) &= H_t P_j(t:t-1) H_t^T + R_t \end{aligned} \tag{V.17}$$

It follows that the mean is:

$$\boldsymbol{\mu} = \frac{1}{W_t} \sum_{j|d_j \le d} w_j^t \boldsymbol{\mu}_j \tag{V.18}$$

and the second moment is:

$$\begin{aligned} \boldsymbol{\mu}^{(2)} &= E\left[xx^T\right] \\ &= \tfrac{1}{W_t} \textstyle\sum_{j|d_j \le d} w_j^t E\left[x_j x_j^T\right] \\ &= \tfrac{1}{W_t} \textstyle\sum_{j|d_j \le d} w_j^t \left(P_j + \boldsymbol{\mu}_j \boldsymbol{\mu}_j^T\right) \end{aligned} \tag{V.19}$$

Then the covariance Σ is

$$\begin{aligned} \Sigma &= \boldsymbol{\mu}^{(2)} - \boldsymbol{\mu}\boldsymbol{\mu}^T \\ &= \tfrac{1}{W_t} \textstyle\sum_{j|d_j \le d} w_j^t \left(P_j + \boldsymbol{\mu}_j \boldsymbol{\mu}_j^T\right) \\ &\quad - \tfrac{1}{W_t^2} \textstyle\sum_{j,j'|d_j,d_{j'} \le d} w_j^t w_{j'}^t \boldsymbol{\mu}_j \boldsymbol{\mu}_{j'}^T \end{aligned} \tag{V.20}$$

Finally, we assume that the distribution of the predicted state, which is the mixture of gaussians, is a single multivariate gaussian variable with mean $\boldsymbol{\mu}^{t:t-1}$ and covariance $P_{t:t-1}$ such that:

$$\begin{aligned} \boldsymbol{\mu}^{t:t-1} &= \tfrac{1}{W_t} \textstyle\sum_{j|d_j \le d} w_j^t \boldsymbol{\mu}_j \\ P_{t:t-1} &= \tfrac{1}{W_t} \textstyle\sum_{j|d_j \le d} w_j^t \left(P_j + \boldsymbol{\mu}_j \boldsymbol{\mu}_j^T\right) \\ &\quad - \tfrac{1}{W_t^2} \textstyle\sum_{j,j'|d_j,d_{j'} \le d} w_j^t w_{j'}^t \boldsymbol{\mu}_j \boldsymbol{\mu}_{j'}^T \end{aligned} \tag{V.21}$$

D. *Analysis*

time and space complexities and accuracy

VI. Conclusion and Future Work

VII. Acknowledgement

VIII. Appendix

A. *Polyhedral partition of modes*

For an entire link with discrete state space indexed by $i = 1, ..., n$, and state $\boldsymbol{\rho}^t = (\rho_0, \rho_1, ..., \rho_n, \rho_{n+1})$ a n+2 dimensional vector which describes the state of the link in the space $\mathcal{S} = [0, \rho_j]^{n+2}$, we present a whole description of the space of "modes" along it, partitioned in different polyhedra. We can note that the ghost cells are included. Since there is always the entry ρ_i in common for successive pairs (ρ_{i-1}, ρ_i) and (ρ_i, ρ_{i+1}), a correlation propagates along the link, reducing the number of modes to a quantity smaller than 3^n.

1) H-representation of a polyhedron: A convex polyhedron (or polytope) may be defined as an intersection of a finite number of half-spaces. Such definition is called a *half-space representation* or *H-representation*.
A closed half-space can be written as a linear inequality:

$$a_1 x_1 + a_2 x_2 + ... + a_n x_n \le b \tag{VIII.1}$$

where n is the dimension of the space containing the polytope under consideration. Hence, a *convex polytope* may be regarded as the set of solutions to the system of linear inequalities:

$$\begin{aligned} & a_{11}x_1 + a_{12}x_2 + ... + a_{1n}x_n \le b_1 \\ & a_{21}x_1 + a_{22}x_2 + ... + a_{2n}x_n \le b_2 \\ & ... \\ & a_{m1}x_1 + a_{m2}x_2 + ... + a_{mn}x_n \le b_m \end{aligned} \tag{VIII.2}$$

where m is the number of half-spaces defining the polytope (the dimension here is not related to the dimension of the space $\mathcal{S}$, and is used only for the definition). This can be concisely written as the matrix inequality:

$$Ax \le b \tag{VIII.3}$$

A is an $m \times n$ matrix, x is an $n \times 1$ column vector of variables, and b is an $m \times 1$ column vector of constants. Here, we don't differentiate *closed convex polyhedron* which are defined with large inequalities from *open convex polyhedron* which are defined with strict inequalities. And the polyhedra $\mathbf{W}_i, \mathbf{L}_i, \mathbf{D}_i$ defined in (III.11) for $i = 0, ..., n$ in the space $\mathcal{S}$ are represented with a mix of large and strict inequalities to have a proper partition of the space.

2) Algorithm to find the minimal H-representation: We use the indicator functions defined in (IV.1,IV.3). Starting from $i = 0$, we evaluate $\alpha_0(\boldsymbol{\rho})$, $\beta_0(\boldsymbol{\rho})$, and $\gamma_0(\boldsymbol{\rho})$ and get s_i, or equivalently Q_i, or the first two defining half-spaces of $\mathbf{P}_{\boldsymbol{s}}$. By construction, the intersection of these two half-spaces are the *minimal H-representation* of Q_0. Now suppose we have evalued the indicator functions and s_i sequentially for $i = 1, ..., k$ and constructed the *minimal H-representation* of $\bigcap_{i=0}^k \mathbf{Q}_i$.

Case $s_k = w$: We have:

$$\begin{aligned} & w_k(\boldsymbol{\rho}) = \alpha_k(\boldsymbol{\rho})\gamma_k(\boldsymbol{\rho}) = 1 \\ & \beta_{k+1}(\boldsymbol{\rho}) = \gamma_k(\boldsymbol{\rho}) = 1 \\ & w_{k+1}(\boldsymbol{\rho}) = \alpha_{k+1}(\boldsymbol{\rho})\gamma_{k+1}(\boldsymbol{\rho}) \\ & l_{k+1}(\boldsymbol{\rho}) = 1 - \gamma_{k+1}(\boldsymbol{\rho}) \\ & d_{k+1}(\boldsymbol{\rho}) = 0 \end{aligned} \tag{VIII.4}$$

This means that $s_{k+1} \in \{w, l\}$ and evaluating $\gamma_{k+1}(\boldsymbol{\rho})$ suffices to find s_{k+1}. It is also necessary to evaluate $\gamma_{k+1}(\boldsymbol{\rho})$ because its value is independent from the values of the indicator functions up to $i = k$ (given that $s_k = w$). Then s_{k+1} is exactly determined by one of the two half-spaces $H_{\gamma_{k+1}}$ and $H^d_{\gamma_{k+1}}$, and adding one of them to the *minimal H-representation* of $\bigcap_{i=0}^k \mathbf{Q}_i$ gives the *minimal H-representation* of $\bigcap_{i=0}^{k+1} \mathbf{Q}_i$.

Case $s_k = l$: We have:

$$
\begin{array}{l}
l_k(\boldsymbol{\rho}) = \beta_k(\boldsymbol{\rho})(1-\gamma_k(\boldsymbol{\rho})) = 1 \\
\beta_{k+1}(\boldsymbol{\rho}) = \gamma_k(\boldsymbol{\rho}) = 0 \\
w_{k+1}(\boldsymbol{\rho}) = \alpha_{k+1}(\boldsymbol{\rho})\gamma_{k+1}(\boldsymbol{\rho}) \\
l_{k+1}(\boldsymbol{\rho}) = 0 \\
d_{k+1}(\boldsymbol{\rho}) = 1 - \alpha_{k+1}(\boldsymbol{\rho})
\end{array} \tag{VIII.5}
$$

A similar analysis shows that adding $H_{\alpha_{k+1}}$ or $H^d_{\alpha_{k+1}}$ gives the *minimal H-representation* of $\bigcap_{i=0}^{k+1} \mathbf{Q}_i$.

Case $s_k = d$: We have no prior information on $w_{k+1}(\boldsymbol{\rho})$, $l_{k+1}(\boldsymbol{\rho})$, and $d_{k+1}(\boldsymbol{\rho})$ and have to evaluate each one of them to find the two defining half-spaces that give the *minimal H-representation* of $\bigcap_{i=0}^{k+1} \mathbf{Q}_i$, when added to the set of current defining half-spaces.

By recurrence, we have constructed a *minimal H-representation* of $\mathbf{P}_{\boldsymbol{s}} = \bigcap_{i=0}^{n} \mathbf{Q}_i$, namely

$$
\begin{array}{l}
\bigcap_{j=0}^{k} \mathbf{H}_j \\
\mathbf{H}_j \in \{\mathbf{H}_{\alpha_i}, \mathbf{H}^d_{\alpha_i}, \mathbf{H}_{\beta_i}, \mathbf{H}^d_{\beta_i}, \mathbf{H}_{\gamma_i}, \mathbf{H}^d_{\gamma_i}\}_{i=0,\ldots,n}
\end{array} \tag{VIII.6}
$$

Hence, each of the defining half-spaces is of dimension $n+1$ [12] (recall that the space $\mathcal{S} = [0, \rho_j]^{n+2}$ is of dimension $n+1$), and seperate by definition the polyhedron $\mathbf{P}_{\boldsymbol{s}}$ from $(n+1)$-adjacent polyhedra $\mathbf{P}_{\boldsymbol{s}'}$ of the polyhedral partition of $\mathcal{S}$.

3) Proof of the exclusivity of all (n+1)-adjcacencies: Given a polyhedron $\mathbf{P}_{\boldsymbol{s}}$ of the partition of $\mathcal{S}$, its minimal H-representation (VIII.6) $\bigcap_{j=0}^{k} \mathbf{H}_j$ given by the algorithm above, its H-representation $\bigcap_{i=0}^{n} \mathbf{Q}_i$ given by (III.12, III.13, IV.2), and an adjacent polyhedron $\mathbf{P}_{\boldsymbol{s}'}$, there exists a unique $j \in \{0, ..., k\}$ such that $\partial \mathbf{H}_j$ $(= \bar{\mathbf{H}}_j \backslash \mathbf{H}^0_j)$ is a common supportive hyperplane of $\mathbf{P}_{\boldsymbol{s}}$ and $\mathbf{P}_{\boldsymbol{s}'}$, by definition of the minimal H-representation. There exists $i' \in \{0, ..., n\}$ such that

$$
\begin{array}{l}
\mathbf{Q}_{i'} = \mathbf{H}_j \cap \mathbf{H}_{i'} \\
\mathbf{H}_j, \mathbf{H}_{i'} \in \{\mathbf{H}_{\alpha_{i'}}, \mathbf{H}^d_{\alpha_{i'}}, \mathbf{H}_{\beta_{i'}}, \mathbf{H}^d_{\beta_{i'}}, \mathbf{H}_{\gamma_{i'}}, \mathbf{H}^d_{\gamma_{i'}}\}
\end{array} \tag{VIII.7}
$$

and we can see from figure II.3 that there exists $\mathbf{R}_{i'}$ in $\{\mathbf{W}_{i'}, \mathbf{L}_{i'}, \mathbf{D}_{i'}\}$ defined in (III.11) different from $\mathbf{Q}_{i'}$ such that

$$
\mathbf{H}^d_j \cap \mathbf{H}_{i'} \subset \mathbf{R}_{i'} \tag{VIII.8}
$$

Let $\mathcal{I}$ be the set of all $i' \in \{0, ..., n\}$ such that we have (VIII.7, VIII.8), then $(\bigcap_{i \notin \mathcal{I}} \mathbf{Q}_i) \cap (\bigcap_{i' \in \mathcal{I}} \mathbf{R}_{i'})$ is a polyhedron of the partition. Let's denote its associated mode $\boldsymbol{s}''$.

$$
\bar{\mathbf{P}}_{\boldsymbol{s}} \cap \partial \mathbf{H}_j = (\bigcap_{i \notin \mathcal{I}} \bar{\mathbf{Q}}_i) \cap (\bigcap_{i' \in \mathcal{I}} \bar{\mathbf{H}}_{i'}) \cap \partial \mathbf{H}_j \subset (\bigcap_{i \notin \mathcal{I}} \bar{\mathbf{Q}}_i) \cap (\bigcap_{i' \in \mathcal{I}} \bar{\mathbf{R}}_{i'}) = \bar{\mathbf{P}}_{\boldsymbol{s}''} \tag{VIII.9}
$$

Moreover, $\mathbf{P}_{\boldsymbol{s}'}$ is seperated from $\mathbf{P}_{\boldsymbol{s}}$ by $\partial \mathbf{H}_j$, so:

$$
(\bigcap_{i \notin \mathcal{I}} \bar{\mathbf{Q}}_i) \cap (\bigcap_{i' \in \mathcal{I}} \bar{\mathbf{H}}_{i'}) \cap \bar{\mathbf{H}}^d_j \cap \bar{\mathbf{P}}_{\boldsymbol{s}'} \text{ is of full dimension and a subset of } \bar{\mathbf{P}}_{\boldsymbol{s}''} \tag{VIII.10}
$$

Since the closures of two polyhedra of the polyhedral partition are either equal or have an intersection at most a hyperplane, we have:

$$
\begin{array}{l}
\bar{\mathbf{P}}_{\boldsymbol{s}'} = \bar{\mathbf{P}}_{\boldsymbol{s}''} \\
\boldsymbol{s}' = \boldsymbol{s}''
\end{array} \tag{VIII.11}
$$

$$
\bar{\mathbf{P}}_{\boldsymbol{s}} \cap \partial \mathbf{H}_j \subset \bar{\mathbf{P}}_{\boldsymbol{s}'} \tag{VIII.12}
$$

so an adjacent polyhedra must be $\mathbf{P}_{\boldsymbol{s}'}$. This completes the proof.

4) Number of modes and intractibility: Suppose that the pair (ρ_0, ρ_1) is in the region $\mathbf{W}$, then the list of possible combinations in Table III.1 shows that (ρ_1, ρ_2) can be either in $\mathbf{W}$ or $\mathbf{L}$. Similarly, if (ρ_0, ρ_1) is in the region $\mathbf{L}$, (ρ_1, ρ_2) can be either in $\mathbf{W}$ or $\mathbf{L}$, and for (ρ_0, ρ_1) in $\mathbf{D}$, (ρ_1, ρ_2) can be either in $\mathbf{W}$, $\mathbf{L}$, or $\mathbf{D}$. As an example, Table VIII.1 describes all the possible sixteen combinations for the first three pairs (ρ_0, ρ_1), (ρ_1, ρ_2), and (ρ_2, ρ_3).

(ρ_0, ρ_1)	W				L					D				
(ρ_1, ρ_2)	W		L		W		D			W		L		
(ρ_2, ρ_3)	W	L	W	D	W	L	W	L	D	W	L	W	D	W

TABLE VIII.1: The sixteen possible modes for the first three pairs (ρ_0, ρ_1), (ρ_1, ρ_2), and (ρ_2, ρ_3).

We can recursively compute the number of "modes" M_k with respect to k, where k is the number of cells of the discretized link. Let's denote by w_k, l_k, and d_k the number of modes for which (ρ_k, ρ_{k+1}) is in $\mathbf{W}$, $\mathbf{L}$, and $\mathbf{D}$ respectively. Then we have these equations:

$$
w_0 = l_0 = d_0 = 1 \tag{VIII.13}
$$

$$
\begin{array}{lclclcl}
w_{k+1} & = & w_k & + & l_k & + & d_k \\
l_{k+1} & = & w_k & + & d_k & & \\
d_{k+1} & = & l_k & + & d_k & &
\end{array} \quad \text{for } k \geq 0 \tag{VIII.14}
$$

$$
n_k = w_k + l_k + d_k \quad \text{for } k \geq 0 \tag{VIII.15}
$$

Using matrix notations, equation (VIII.14) reads:

$$
\begin{bmatrix} w_{k+1} \\ l_{k+1} \\ d_{k+1} \end{bmatrix} = A \times \begin{bmatrix} w_k \\ l_k \\ d_k \end{bmatrix} \quad \text{where} \quad A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix} \tag{VIII.16}
$$

Then

$$
\begin{bmatrix} w_k \\ l_k \\ d_k \end{bmatrix} = A^k \times \begin{bmatrix} w_0 \\ l_0 \\ d_0 \end{bmatrix} \tag{VIII.17}
$$

It is possible to compute A^k explicitly by diagonalizing the matrix A, to obtain an explicit expression for w_k, l_k, and d_k in the form of $a.\beta^k + b.\gamma^k + c.\delta^k$. However, this analytical expression is unwieldy, so we will just derive lower and upper bounds to n_k. It is easy to see that $d_k \leq n_k/2$

number of cells	1	2	5	10	20
number of modes	7	16	182	10426	34206521
bound without analysis	7	49	16807	282475249	$8 \cdot 10^{16}$

TABLE VIII.2: Number of modes for a homogeneous road.

for $k \geq 0$, then we can prove recursively that $3 \cdot 2^k \leq n_k \leq 3 \cdot (2.5)^k$.

Even if we have found the minimal polyhedral partition of the space, the number of modes grows exponentially as the number of cells increases, so it is difficult to store all the possible modes. However, at any time step, the mode of each cell can be determined among the 7 possible modes and constructed sequentially building up the general mode of the segment of road.

B. The heterogeneous case

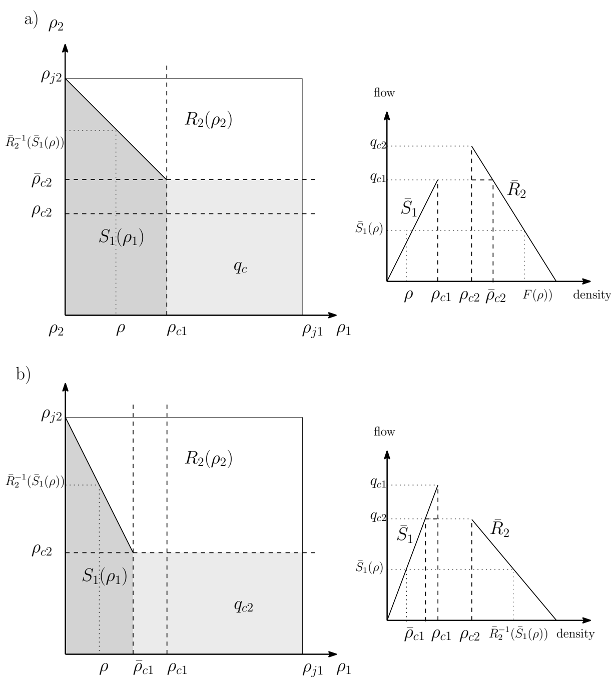

Fig. VIII.1: Values of $G(\rho_1, \rho_2)$ in the space (ρ_1, ρ_2) for the Daganzo-Newell fundamental diagram with different capacities $q_{c1} < q_{c2}$ for a) and $q_{c1} > q_{c2}$ for b). Note that for illustration purposes we suppose $\rho_{c1} \leq \rho_{c2}$.

In this section, we study the CTM for a heterogeneous road, i.e. ω_f, v_f, ρ_j, ρ_c, q_c can vary along the link and we add the subscript i: α_i, ω_{fi}, v_{fi}, ρ_{ji}, ρ_{ci}, q_{ci} for the parameters of the fundamental diagram Q_i at cell i, and the associated sending and receiving flows S_i, R_i. Figure VIII.1 shows the explicit values taken by $G(\rho_1, \rho_2)$ in different regions of the space (ρ_1, ρ_2) when $q_{c1} < q_{c2}$. We note that the critical density ρ_{c2} is increased to the *effective* critical value $\bar{\rho}_{c2}$ for the receiving flow R_2, with $\bar{\rho}_{c2} = \bar{R}_2^{-1}(q_{c1})$, and the *effective* capacity is $\bar{q}_c = q_{c1}$, which is the capacity of the sending flow S_1. And Figure VIII.1 also shows the explicit values taken by $G(\rho_1, \rho_2)$ in different regions of the space (ρ_1, ρ_2) when $q_{c1} > q_{c2}$. Similarly, the critical density ρ_{c1} is decreased to the *effective* value $\bar{\rho}_{c1}$ for the sending flow S_1 with $\bar{\rho}_{c1} = \bar{S}_1^{-1}(q_{c2})$, and the *effective* capacity is $\bar{q}_c = q_{c2}$, which is the capacity of the receiving flow R_2. The Godunov flux has a more general expression VIII.18:

$$G(\rho_1, \rho_2) = \begin{cases} R_2(\rho_2) & \text{if } (\rho_1, \rho_2) \in \mathbf{W} \\ \bar{q}_c & \text{if } (\rho_1, \rho_2) \in \mathbf{L} \\ S_1(\rho_1) & \text{if } (\rho_1, \rho_2) \in \mathbf{D} \end{cases} \tag{VIII.18}$$

$$\begin{array}{lll} \mathbf{W} & = \{(\rho_1, \rho_2) \mid & \rho_2 > F(\rho_1) \ , \ \rho_2 > \bar{\rho}_{c2}\} \\ \mathbf{L} & = \{(\rho_1, \rho_2) \mid & \rho_1 > \bar{\rho}_{c1} \ , \ \rho_2 \leq \bar{\rho}_{c2}\} \\ \mathbf{D} & = \{(\rho_1, \rho_2) \mid & \rho_2 \leq F(\rho_1) \ , \ \rho_1 \leq \bar{\rho}_{c1}\} \end{array} \tag{VIII.19}$$

where the boundary between the white and grey regions follows the $(\rho_1, \rho_2) = (\rho_1, F(\rho_1))$ trajectory with $F(\rho_1) = \bar{R}_2^{-1}(\bar{S}_1(\rho_1))$ for $\rho_1 \leq \rho_{c1}$. $\bar{S}$ and $\bar{R}$ denote the restrictions of the sending and receiving flows to the sub-regions $[0, \rho_c)$ and $(\rho_c, \rho_j]$ respectively, which also correspond to the left and right parts (w.r.t. ρ_c) of the fundamental diagram, as shown in the Figure VIII.1.

When the velocity is the Daganzo-Newell function (II.6), the Godunov Flux VIII.18 becomes II.17:

$$G_{DN}(\rho_1, \rho_2) = \begin{cases} -\omega_{f2}\,(\rho_2 - \rho_{j2}) & \text{if } (\rho_1, \rho_2) \in \mathbf{W} \\ \bar{q}_c & \text{if } (\rho_1, \rho_2) \in \mathbf{L} \\ v_{f1}\rho_1 & \text{if } (\rho_1, \rho_2) \in \mathbf{D} \end{cases} \tag{VIII.20}$$

and the boundary between the white and dark-grey regions is:

$$(\rho_1, \rho_2) = (\rho_1, -\frac{v_{f1}}{\omega_{f2}}\rho_1 + \rho_{j2}) \tag{VIII.21}$$

And **W**, **L**, **D** form a polyhedral partition of the space:

$$\begin{array}{lll} \mathbf{W} & = \{(\rho_1, \rho_2) \mid & \rho_2 + \frac{v_{f1}}{\omega_{f2}}\rho_1 > \rho_{j2} \ , \ \rho_2 > \bar{\rho}_{c2}\} \\ \mathbf{L} & = \{(\rho_1, \rho_2) \mid & \rho_1 > \bar{\rho}_{c1} \ , \ \rho_2 \leq \bar{\rho}_{c2}\} \\ \mathbf{D} & = \{(\rho_1, \rho_2) \mid & \rho_2 + \frac{v_{f1}}{\omega_{f2}}\rho_1 \leq \rho_{j2} \ , \ \rho_1 \leq \bar{\rho}_{c1}\} \end{array} \tag{VIII.22}$$

At cell i, this implies the effective density $\bar{\rho}^u_{ci}$ associated with the upstream boundary can be different from the effective density $\bar{\rho}^d_{ci}$ associated with the downstream boundary, depending of the capacity drops at these boundaries. Hence, using the notations introduced in section III-A all the combinations between (ρ_-, ρ) and (ρ, ρ_+) can be possible so we have nine modes. Consequently, for a discretization in n cells, the number of possible modes is 3^{n+1}.

number of cells	1	2	5	10	20
number of modes	9	27	729	177147	10^{10}

TABLE VIII.3: Number of modes for a heterogeneous road.

C. Kalman filtering-DRAFT

and $F_{LH}[.]$, have the same expression with $f_{DN}(.)$ and $f_{LH}(.)$ respectively (as defined in Table. For instance, when we have a Daganzo-Newell fundamental diagram with update operator for the dynamic system $F_{DN}[.]$ (i.e. $\boldsymbol{\rho}^t = F_{DN}[\boldsymbol{\rho}^{t-1}]$), and suppose that the mode at $x = i$ is three ($m_i = 3$) then[5]:

$$F_i[\boldsymbol{\rho}^{t-1}] = f_{DN,m_i}(\rho^{t-1}_{i_-,i,i_+}) = f_{DN,3}(\rho^{t-1}_{i_-,i,i_+}) = \rho_i^{t-1} + \alpha\omega_f \rho_{i+1}^{t-1} - \alpha\omega_f\rho_c \tag{VIII.23}$$

In order to use the *Extended Kalman filter* to estimate the state of the link given a sequence of noisy observations, we model the process in accordance with the framework of the *Extended Kalman filter* by adding a white noise to the underlying dynamic system model. The "true" state $\boldsymbol{\rho}^t$ is then:

$$\boldsymbol{\rho}^t = F[\boldsymbol{\rho}^{t-1}] + \boldsymbol{\eta}^t \tag{VIII.24}$$

where $\boldsymbol{\eta}^t \sim N(0, Q_t)$ is the Gaussian zero-mean, white state noise with covariance Q_t. The estimated state at time t is denoted by $\hat{\boldsymbol{\rho}}^t$ and the estimated covariance by P_t. The *prediction step* gives the *a priori* state estimate and covariance $\hat{\boldsymbol{\rho}}^{t:t-1}$ and $P_{t:t-1}$:

$$\begin{array}{ll} \text{Predicted state estimate:} & \hat{\boldsymbol{\rho}}^{t:t-1} = F[\hat{\boldsymbol{\rho}}^{t-1}] \\ \text{Predicted covariance estimate:} & P_{t:t-1} = F_{t-1}P_{t-1}(F_{t-1})^T + Q_{t-1} \end{array} \tag{VIII.25}$$

where F_t is the state transition defined to be the following Jacobian:

$$F_t = \left(\frac{\partial F_i[\rho^t]}{\partial \rho_j^t}\right)_{i,j} \tag{VIII.26}$$

The estimated mode $\hat{\boldsymbol{m}}^t$ associated to the estimated vector state $\hat{\rho}^t$ is defined from Table III.1. Specifically, in the context of our traffic model, the density at $x = i$ only depends on the densities at the neighbor points $x = i-1, i, i+1$. So F_t is a $(n+2) \times (n+2)$ tridiagonal matrix, such that the diagonal elements are $\{0, J_{\hat{m}_1,2}, ..., J_{\hat{m}_n,2}, 0\}$, the lower diagonal elements are $\{J_{\hat{m}_1,1}, J_{\hat{m}_2,1}, ..., J_{\hat{m}_n,1}, 0\}$, and the upper diagonal elements are $\{0, J_{\hat{m}_1,3}, J_{\hat{m}_2,3}, ..., J_{\hat{m}_n,3}\}$, where J is defined in Equation III.2.

In the case of the Daganzo-Newell fundamental diagram, the operator $F_{DN}[.]$ defined in (III.7) is *linear* (from the linearity of f_{DN}). Along with the assumption of a white state noise, the *prediction step* (IV.6) of the *Extended Kalman filter* is actually identical to the regular *Kalman filter*, and it is known from the theory that the *Kalman filter* is optimal [2].

[5] As we have seen earlier, α, ω_f, v_f, ρ_j, ρ_c, q_c can vary along the link and a proper notation would be α_i, ω_{fi}, v_{fi}, ρ_{ji}, ρ_{ci}, q_{ci} for the parameters of the fundamental diagram at $x = i$, so that:

$$F_i[\rho^{t-1}] = f_{DN,m_i}(\rho^{t-1}_{i_-,i,i_+}) = f_{DN,3}(\rho^{t-1}_{i_-,i,i_+}) = \rho_i^{t-1} + \alpha_i\omega_{fi}\rho_{i+1}^{t-1} - \alpha_i\omega_{fi}\rho_{ci}$$

In the case of a linear-hyperbolic velocity function, the operator $F_{LH}[.]$ is not linear if at some points of the discrete space the mode is between four and seven. Then the *Extended Kalman filter* is *near-optimal*.

We can note that the first line and first column of P_t have only zero elements because the boundary condition ρ_0^t is deterministic (i.e. $cov(\rho_0^t, \rho_i^t) = 0$ for $i = 1, ..., n$), and similarly the last line and last column of P_t are null since the boundary condition ρ_{n+1}^t is deteministic. Additionally, the observation model for the link is given by:

$$\boldsymbol{y}^t = H_t\boldsymbol{\rho}^t + \boldsymbol{\chi}^t \tag{VIII.27}$$

where $H_t \in \{0,1\}^{p_t \times n}$ is the linear observation observation matrix which encodes the p_t observations (each one of them being at a discrete cell on the highway) for which the density is observed during discrete time step t, and n is the number of cells along the link. The last term in (IV.5) is the white, zero mean observation noise $\boldsymbol{\chi}^t \sim N(0, R_t)$ with covariance matrix R_t. The *update step* is:

$$\begin{array}{ll} \text{Kalman gain:} & K_t = P_{t:t-1}H_t^T\left(H_tP_{t:t-1}H_t^T + R_t\right)^{-1} \\ \text{Updated state estimate:} & \hat{\boldsymbol{\rho}}^t = \hat{\boldsymbol{\rho}}^{t:t-1} + K_t(\boldsymbol{y}^t - H_t\hat{\boldsymbol{\rho}}^{t:t-1}) \\ \text{Updated estimate covariance:} & P_t = (I - K_tH_t)P_{t:t-1} \end{array} \tag{VIII.28}$$

The Extended Kalman filter provides the distribution of $\hat{\boldsymbol{\rho}}^t$ given the sequence of observations $\boldsymbol{y}^{0:t}$, sequence of modes $\hat{\boldsymbol{m}}^{0:t} = \{\hat{\boldsymbol{m}}^0, ..., \hat{\boldsymbol{m}}^t\}$, and sequence of control parameters $\boldsymbol{u}^{0:t}$, which is exactly equal to the distribution of $\hat{\boldsymbol{\rho}}^t$ since the sequence $\hat{\boldsymbol{m}}^{0:t}$ is deterministic. Concretely, the vector of control parameters $\boldsymbol{u}^t$ contain the vector of critical densities $\boldsymbol{\rho}_c$, the vector of jam densities $\boldsymbol{\rho}_j$, and the boundary conditions ρ_0^t and ρ_{n+1}^t. Theoretically, the result is obtained by marginalizing the joint distribution of $\hat{\boldsymbol{\rho}}^t$ and $\hat{\boldsymbol{m}}^{0:t}$ as follows:

$$\begin{aligned} p(\hat{\boldsymbol{\rho}}^t|\boldsymbol{y}^{0:t}, \boldsymbol{u}_{0:t}) &= \textstyle\int_{\mathcal{M}_n} p(\hat{\boldsymbol{\rho}}^t|\boldsymbol{y}^{0:t}, \boldsymbol{u}^{0:t}, \boldsymbol{m}^{0:t})p(\boldsymbol{m}^{0:t}|\boldsymbol{y}^{0:t}, \boldsymbol{u}^{0:t})d\boldsymbol{m}^{0:t} \\ &= \textstyle\int_{\mathcal{M}_n} p(\hat{\boldsymbol{\rho}}^t|\boldsymbol{y}^{0:t}, \boldsymbol{u}^{0:t}, \boldsymbol{m}^{0:t})\mathbf{1}_{\hat{\boldsymbol{m}}^{0:t}}d\boldsymbol{m}^{0:t} \\ &= p(\hat{\boldsymbol{\rho}}^t|\boldsymbol{y}^{0:t}, \boldsymbol{u}^{0:t}, \hat{\boldsymbol{m}}^{0:t}) \end{aligned} \tag{VIII.29}$$

D. Implementation-DRAFT

1) Algorithm for the prediction step: Algorithm using the structure of F_t and extension to a network

From Appendix VIII-A.4, In (IV.6), the *a priori* state estimate $\hat{\boldsymbol{\rho}}^{t:t-1}$ is derived through the algorithm in (III.7).

2) Accuracy: For a Daganzo-Newell fundamental diagram, we can note that the decomposition in different modes in (III.7) is in fact a special case of *Conditional Dynamic Linear Model* (CDLM) as described in [5] with a discrete latent indicator that is deterministic. In this case, there is only one estimated deterministic sequence of modes $\hat{\boldsymbol{m}}^{0:t}$ or latent indicators, hence there is no need to use a sequential Monte-Carlo method to sample an ensemble of trajectories of modes. The Mixture Kalman Filter [5] reduces to a simple Kalman Filter. For non-triangular fundamental diagrams which have the characteristics **LWR1-6**, a first order Taylor Series expansion is applied. Such a linearization is a good approximation. Then the Extended Kalman filter can be applied to each mode.

In our case, the estimated sequence of modes $\hat{\boldsymbol{m}}^{0:t}$ is readily infered through the Godunov scheme, and at all the cells along the link, including those where there is no observation. Contrary to some previous models as in [19], the modes are not directly sampled from density measurements along the highway. However, the mode is still indirectly infered from those measurements by assimilating the observations with the *update state* of the Kalman filter. Besides, we rely on the accuracy of the estimation of the mode provided by the Godunov scheme when we apply the Kalman filter for each mode. Such an assumption that favors the mode provided by the Godunov scheme is yet another reason why the sequential Monte-Carlo method (with the resampling step) is not used in the algorithm, because we then suppose that our sequence of modes is the one with the largest likelihoods. (to develop, do a test of likelihood)

3) Complexity: Comparison with the Extended Kalman filter and matlab simulations and results

REFERENCES

[1] K. Agyemang-Duah and F. Hall. Some issues regarding the numerical value of freeway capacity. *International Symposium on Highway Capacity*, 1991.

[2] B. D. O. Anderson and J. B. Moore. *Optimal Filtering*. Dover Publications, Inc., 2005.

[3] M. Cassidy and R. Bertini. Some traffic features at freeway bottlenecks. *Transportation Research*, 1999.

[4] C. Chen, P. Varaiya, and J. Kwon. An empirical assessment of traffic operations. *International Symposium on Transportation and Traffic Theory*, 2005.

[5] Rong Chen and Jun S. Liu. Mixture kalman filters. *Royal Statistical Society*, 2000.

[6] C. F. Daganzo. The cell transmission model: a dynamic representation of highway traffic consistent with the hydrodynamic theory. *Transportation Research Part B 28, no. 4*, 1994.

[7] C. F. Daganzo. The cell transmission model, part ii: network traffic. *Transportation Research Part B 29, no. 2*, 1995.

[8] Mauro Garavello and Benedetto Piccoli. *Traffic Flow on Networks*. American Institute of Mathematical Sciences, 2006.

[9] Edwige Godlewski and Pierre-Arnaud Raviart. *Numerical approximation of hyperbolic systems of conservation laws*. Applied Mathematical Sciences, 1996.

[10] S.K. Godunov. A finite difference method for the numerical computation of discontinuous solutions of the equations of fluid dynamics. 1959.

[11] B. D. Greenshields. A study of traffic capacity. *Proceedings of the 14th annual meeting of the Highway Research Board*, 1934.

[12] Branko Grnbaum. *Convex Polytopes*. Springer, 2003.

[13] F. Hall and K. Agyemang-Duah. Freeway capacity drop and the definition of capacity. *Transportation Research Record*, 1991.

[14] J.-C. Herrera, D. B. Work, R. Herring, J. Ban, Q. Jacobson, and A. M. Bayen. Evaluation of traffic data obtained via gps-enabled mobile phones: the mobile century experiment. *Transportation Research Part C*, 2009.

[15] B. Hoh, M. Gruteser, R. Herring, J. Ban, D. Work, J.-C. Herrera, A. M. Bayen, M. Annavaram, and Q. Jacobson. Virtual trip lines for distributed privacy-preserving traffic monitoring. *6th International Conference on Mobile Systems, Applications, and Services*, 2008.

[16] J. P. Lebacque. The godunov scheme and what it means for first order traffic flow models. *13th International Symposium on Transportation and Traffic Theory*, 1996.

[17] R. J. LeVeque. *Numerical Methods for Conservation Laws*. Birkhuser Basel, 1992.

[18] M. J. Lighthill and G. B. Whitham. On kinematic waves. ii. a theory of traffic flow on long crowded roads. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, pages 317–345, 1955.

[19] Laura Munoz, Xiaotian Sun, Roberto Horowitz, and Luis Alvarez. Traffic density estimation with the cell transmission model. *Proceedings of the American Control Conference*, 2003.

[20] P. I. Richards. Shock waves on the highway. *Operations Research*, 4:42–51, 1956.

[21] I. S. Strub and A. M. Bayen. Weak formulation of boundary conditions for scalar conservation laws: an application to highway traffic modeling. *Int. J. Robust Nonlinear Control*, 2006.

[22] D. B. Work and A. M. Bayen. Impacts of the mobile internet on transportation cyber-physical systems: Traffic monitoring using smartphones. *National Workshop for Research on High-Confidence Transportation Cyber-Physical Systems: Automotive, Aviation, & Rail*, 2008.

[23] D. B. Work, A. M. Bayen, and Q. Jacobson. Automotive cyber-physical systems in the context of human mobility. *National Workshop on High-Confidence Automotive Cyber-Physical Systems*, 2008.