A Macroscopic Traffic Data-Assimilation Framework Based on the Fourier–Galerkin Method and Minimax Estimation

Tigran T. Tchrakian and Sergiy Zhuk

***Abstract*—In this paper, we propose a new framework for macroscopic traffic state estimation. Our approach is a robust "discretize" then "optimize" strategy, based on the Fourier–Galerkin projection method and minimax state estimation. We assign a Fourier–Galerkin reduced model to a macroscopic model of traffic flow, described by a hyperbolic partial differential equation. Taking into account *a priori* estimates for the projection error, we apply the minimax method to construct the state estimate for the reduced model that gives us, in turn, the estimate of the Fourier–Galerkin coefficients associated with a solution of the original macroscopic model. We illustrate our approach with a numerical example that demonstrates its shock capturing capability using only sparse measurements and under high uncertainty in initial conditions. We present implementation details for our algorithm, as well as a comparison of our method against the ensemble Kalman filter applied to a "local" discretization of the same traffic flow model.**

***Index Terms*—Data assimilation, Fourier–Galerkin, macroscopic traffic flow models, minimax, shock waves, viscosity solutions.**

I. Introduction

MACROSCOPIC models of traffic flow can be used together with traffic data in order to estimate the state of traffic on a link or a set of connected links. The basic idea is to run the model, which outputs a solution in time and space, and to combine this solution with the available data over the spatiotemporal domain in order to obtain an estimate of the traffic state. Past approaches have involved linearizing [1] such models in order to apply Kalman filtering [2], [3] or extended Kalman filtering [4], or using the nonlinear model as it stands, and employing techniques such as particle filtering [5]. While each category of methods has its respective advantages, what they have in common is that they are applied to a reduced model that is obtained from the original macroscopic model by "local" methods. Local methods use limited information from the spatial domain in order to compute the solution at a given point in that domain. Specifically, all of the approaches listed above use finite differences as a solution method for the underlying partial differential equation (PDE), where the solution in a spatial "cell" is influenced by the solution in its neighboring cells. The resulting system is a set of difference equations, each describing the evolution of the state variable (typically density) at a given point in space.

1) Motivation: In contrast to local solution methods, "global" methods use information from the entire spatial domain and deliver continuous (in space) approximations for the solution of the macroscopic model. This allows the solution at any point in the domain to be computed. While local methods can capture shocks and provide an adequate solution of the model, they have certain drawbacks when it comes to data assimilation. Specifically, the discretization in space should take into account the locations of sensors in order to make it possible to relate the state of the reduced model and the measurements. This becomes a problem if sensors are moving or their locations are imprecisely given. On the other hand, the solution provided by a local method is accurate if the spatial discretization correctly captures local patterns of the PDE's initial data. However, since a data-assimilation method is supposed to operate without complete knowledge of the initial traffic density, it follows that the corresponding grid should capture correctly the entire class of initial data, which describe all admissible initial traffic densities. This, in general, would require one to work on very fine grids, leading to high computational costs.

To overcome the mentioned shortcomings of the local methods, we propose a global approach to traffic state estimation by using a global method, namely, the Fourier–Galerkin method, to transform the macroscopic traffic model to a system of ordinary differential equations (ODEs), and then perform data assimilation on the resulting reduced model by using the minimax approach.

2) Contribution and Related Work: Our main contribution is an iterative minimax state estimation algorithm for nonlinear hyperbolic equations, which is based on a global method and allows one to capture and track shocks for macroscopic traffic models with uncertain initial conditions given incomplete and noisy observations.

The Fourier–Galerkin method belongs to the class of so-called projection methods. The method works by projecting a solution of the macroscopic model onto a subspace generated by $\{e^{inx}\}_{|n|\leq N/2}$, and then studying the evolution of the projection coefficients only. We refer the reader to [6] for further details on projection methods. Notice that, in general, the Fourier–Galerkin method itself leads to an unstable reduced

Manuscript received January 14, 2014; revised June 12, 2014; accepted July 30, 2014. Date of publication September 17, 2014; date of current version January 30, 2015. This work was supported by the European Union Seventh Framework Programme FP7/2007-2013 under Grant 287867. The Associate Editor for this paper was A. Hegyi.

The authors are with IBM Research, Dublin, Ireland (e-mail: tigran@ie.ibm.com; sergiy.zhuk@ie.ibm.com).

Digital Object Identifier 10.1109/TITS.2014.2347415

model in that the reduced model is subject to unphysical oscillations. This instability is explained as follows: a shock solution is discontinuous and therefore, for the corresponding truncated Fourier series to be accurate, it should necessarily contain coefficients with large wavenumbers corresponding to highly oscillatory basis functions. Now, since the shock moves, the number, i.e., N, of basis functions should be sufficiently large to resolve the dynamics of the shock over time. In other words, in the exact Fourier–Galerkin model composed of infinitely many projection coefficients, the energy travels from low- to high-frequency modes, and if this phenomenon is not reflected by the truncated Fourier–Galerkin model, then the energy "travels back" and instability occurs. Therefore, in order to capture shock formation and correctly track these shocks, we apply a modification of the vanishing viscosity method that allows us to efficiently deal with Gibbs phenomenon (see [7]) by damping out the high-frequency modes, preventing the energy from traveling back.

The Fourier–Galerkin method provides estimates of the projection error in the L^2-norm. We stress that there is no statistical information available on projection errors, and therefore, the statistical uncertainty description of Kalman filter or particle filters does not apply. This motivates us to apply the minimax approach. It allows us to construct a guaranteed estimate of the state of a differential equation with uncertain but bounded initial conditions and model errors (in our case, these errors are associated with projection errors, which are bounded in L^2). The minimax approach is based upon the following idea: to minimize the worst case estimation error. In other words, one computes the estimation error for the worst case realization of the uncertain model error and initial conditions and then defines the minimax estimate as that which has minimal worst case error. Thus, by construction, the minimax estimate is robust to all possible realizations of uncertain parameters lying within the given bounding set. We refer the reader to [8]–[11] for basic information on the minimax framework. Discussion on robust projection methods by means of the minimax state estimation approach can be found in [12] and [13].

The iterative minimax state estimation algorithm is based on the following idea: we consider a given estimate of the state of the reduced model and use the bilinear structure of the reduced model in order to generate a new estimate. Each iteration of the process is based on the previous estimate and is obtained by applying the minimax approach to the linear state equation with a parameter. One stage of the proposed procedure can be seen as a data-driven fixed-point iteration, namely, given the estimate from the previous iteration and observations, we construct a linear system and estimate its state in order to produce a minimax state estimate that will be used in the next iteration as a starting point. To the best of our knowledge, the proposed iterative minimax state estimation approach is new and has not been presented yet in the literature. A similar algorithm was proposed for image motion estimation in [14]. In contrast to [14], we work with hyperbolic equations and perform a detailed numerical assessment. Data assimilation for systems of hyperbolic conservation laws based on the calculus of variations was proposed in [15], where the authors adopt the strategy: "optimize," then "discretize" so that the estimate of the initial density does not depend on a discretization method. In the present paper, we solve the filtering problem: that is, given observations in the past up until time instant T, to estimate traffic density at time T without observing the initial condition. We adopt a robust "discretize-optimize" strategy, that is, we take into account the projection error. Although we do not present a rigorous convergence proof, the numerical results look promising.

3) Outline of This Paper: The remainder of this paper is organized as follows: in Section II, we describe the flow model used in this paper, and in Section III, we explain briefly how finite-difference methods have previously been used to obtain a system for state estimation, and motivate the use of global methods. In Section IV, we describe the Fourier–Galerkin method. In Section V, we outline the data-assimilation procedure for the reduced model. The iterative minimax algorithm and numerical results are presented in Section VI. We conclude and describe future work in Section VII.

4) Notation: $\mathbb{R}^n$ $(\mathbb{C}^n)$ stands for the n-dimensional Euclidean space over the field of real (complex) numbers. $L^2(0,T)$ denotes the space of measurable square-integrable functions on $(0,T)$. In what follows, we will often write L^2 referring to $L^2(0,T)$ over a compact interval.

II. Flow Model

In this paper, we employ the Lighthill–Whitham–Richards [16], [17] (LWR) model. This is the standard equilibrium traffic flow model consisting of a scalar conservation law, i.e.,

$$u(x,t)_t + f\left(u(x,t)\right)_x = 0 \tag{1}$$

with initial data

$$u_0(x) = u(x,0) \tag{2}$$

where $u : \mathbb{R} \times \mathbb{R}_+ \to \mathbb{R}$ is the traffic density, $x \in \mathbb{R}$ and $t \in \mathbb{R}_+$ are the independent variables, space and time, respectively, and $f : \mathbb{R} \to \mathbb{R}$ is the flux function. A typical flux function is that of Greenshields [18], given by

$$f(u) = uV_m\left(1 - \frac{u}{u_m}\right) \tag{3}$$

where the constants V_m and u_m are the maximum speed and the maximum density, respectively.

III. Finite-Difference-Based State Estimation

A standard way of solving a model such as LWR is to use the finite-difference method. Using that approach, the density solution obtained is a grid function. That is, at each computational time step, i.e., Δt, the solution given by the finite-difference method is

$$u_i^j = u(i\Delta x, j\Delta t), \quad i = \ldots -1, 0, 1 \ldots; \quad j = 0, 1, 2, \ldots. \tag{4}$$

Defining the vector $U^j = u_i^j, i = \ldots -1, 0, 1 \ldots$, an n-level difference scheme is one for which the solution U^j is

constructed from $U^{j-1}, U^{j-2}, \ldots, U^{j-n+1}$. Difference schemes that have been used to solve (1), (2), are typically two-level methods, where the solution at a given time step is constructed only from the solution at the previous time step. In addition, the schemes applied to this problem usually require limited information from the spatial domain. For a first-order Godunov scheme applied to a scalar conservation law, solution u_i^j depends on either u_{i+1}^{j-1} or u_{i-1}^{j-1}, depending on the signal speed. The use of limited information is why such schemes are termed "local methods". For a two-level method, the scheme can be written as

$$U^{j+1} = \mathcal{H}U^j \tag{5}$$

where $\mathcal{H}$ is the difference operator. For traffic flow, the signal speed, i.e., $f'(u)$, will change in sign depending on the density, i.e., u; thus, the difference operator will need to change the direction of spatial differencing as appropriate. As a result, $\mathcal{H}$ will depend on the grid function U^j, meaning (5) is not a linear system. However, attempts have been made to linearize such operators. Muñoz *et al.* [1] took such an approach and used a piecewise linear flux function to obtain a switched-linear system. The number of linear systems they switch between depends on some assumptions made about the number of congested and free-flow regions, which can exist on a single link. This linearized model was used for estimation and control [2], [3] with the assumption that at most one of each region (congested/free-flow) can exist on a link. This resulted in a system that switched between five different linear systems, for which they used a Mixture Kalman Filter for the estimation/control. The measurements in that work are of the state (with some additive noise); thus, the transformation from observation space to state space is the identity matrix provided all components of the state vector, U^j can be measured. In general, although sensors are sparse, and the observation space is of lower dimension than the state space. The observations must be taken at grid points, meaning that such an approach works best for fixed sensors, and the spatial discretization should be chosen such that the grid points coincide with the sensor locations.

Another well-known work on traffic data assimilation is that by Herrera and Bayen [19], who used a "Newtonian relaxation method", which is a heuristic method that relaxes the dynamical model toward the observations. That scheme is suitable for mobile sensors through the use of an additional term in the flow model, which acts as either a source or a sink, in order to relax the conservation law toward the measurements from GPS-equipped vehicles. This method also falls into the "local" category because of its use of finite difference to discretize the model. Therefore, the incorporation of the measurements into the model is affected by the choice of spatial discretization. Other authors have used a Lagrangian reformulation of traffic flow models to do traffic state estimation [20], and this moving coordinate system has been extended to networks [21]. The approach there is also local, except rather than using local information from the spatial domain to influence a particular point in space, a particular point in the medium (i.e., motorcars) is influenced by information from the medium in its vicinity.

In the following section, we describe spectral (global) solution methods and will later show how using the resulting system for estimation allows us to take measurements from any point in space, unlike in finite-difference-based methods.

IV. Galerkin Method for LWR Model

Here, we derive the Fourier–Galerkin reduced model for LWR, and then get it into a form suitable for data assimilation.

A. Derivation of Reduced Model

We consider the initial value problem defined by (1) and (2), where $x \in [0, 2\pi]$ and $t \geq 0$. We also restrict our attention to periodic boundary conditions, i.e., $u(0,t) = u(2\pi, t)$. In the Fourier–Galerkin method, solutions are sought in the space, i.e., span$\{e^{inx}\}_{|n|\leq N/2}$ [6]. That is, u is approximated by u_N, which is the $N+1$-term truncated series

$$u_N(x,t) = \sum_{n=-N/2}^{N/2} a_n(t) e^{inx} \tag{6}$$

where $\{e^{inx}\}_{|n|\leq N/2}$ are trigonometric polynomials in complex form. Defining the residual

$$R_N(x,t) = \frac{\partial u_N}{\partial t} + \frac{\partial f(u_N)}{\partial x} \tag{7}$$

and requiring it to be orthogonal to span$\{e^{inx}\}_{|n|\leq N/2}$, we obtain the coefficients $a_n(t)$. However, solutions to (1) can develop shock discontinuities, even for smooth initial data. This will give rise to strong oscillations, which will spread to the entire spatial domain, causing the Fourier–Galerkin method to fail. Tadmor [7] showed that the addition of "spectral viscosity" to a scalar conservation law can help overcome this problem, and allow, moreover, the correct entropy solution to be obtained. The conservation law with viscosity is

$$u_t + f(u)_x = \varepsilon u_{xx} \tag{8}$$

where ε is a constant. Substituting (6) into (8), and using the flux function (3), where we set $V_m = u_m = 1$, leads to the residual

$$\begin{aligned} R_N(x,t) = &\sum_{n=-N/2}^{N/2} \dot{a}_n(t) e^{inx} + \sum_{n=-N/2}^{N/2} ina_n(t) e^{inx} \\ &- 2 \sum_{k=-N/2}^{N/2} \sum_{n=-N/2+k}^{N/2+k} ika_{n-k}(t) a_k(t) e^{inx} \\ &+ \varepsilon \sum_{n=-N/2}^{N/2} n^2 a_n(t) e^{inx}. \end{aligned} \tag{9}$$

Forcing the orthogonal projection of this residual onto span$\{e^{inx}\}_{|n|\leq N/2}$ to vanish, we obtain the following system

of ODEs

$$\frac{da_n}{dt} = \sum_{\substack{k=-N/2 \\ |n-k|\le N/2}}^{N/2} 2ika_{n-k}(t)a_k(t) - ina_n(t) - \varepsilon n^2 a_n(t),$$
$$n = -N/2 \dots N/2. \quad (10)$$

The last term on the right-hand side is due to the viscosity. Tadmor [7] showed that for scalar conservation laws, the addition of numerical viscosity on the higher modes leads to convergence to the unique entropy solution. Thus, the viscosity term is only activated for $|n/2| \ge m$, where m is some threshold wavenumber. This will be discussed in further detail in Section VI. The initial coefficients can be obtained from the data (2) by computing the integral

$$a_n(0) = \frac{1}{2\pi}\int_0^{2\pi} u_0(x)e^{-inx}dx, \quad n = -N/2 \dots N/2. \quad (11)$$

B. System and Observation Operators for Data Assimilation

In order to proceed with the data assimilation, we now express the system as a single differential equation and define an observation operator. We write (10) as

$$\frac{dX}{dt} = A(X)X \quad (12)$$

where $X : \mathbb{R} \to \mathbb{C}^{N+1}$ is a vector containing the $N+1$ expansion coefficients, and $A : \mathbb{C}^{N+1} \to \mathbb{C}^{N+1}$ is constructed using (10). The observation equation is given by

$$Y(t) = HX(t) + \eta(t) \quad (13)$$

where $Y : \mathbb{R} \to \mathbb{R}^M$ is a vector of density measurements, and $H : \mathbb{C}^{N+1} \to \mathbb{R}^M$ is the observation operator, which is a mapping from the complex state space of expansion coefficients to the real observation space of measured densities. This operator is constructed using (6), and the M density measurements can be taken anywhere on the interval, $x \in [0, 2\pi]$.

C. Mappings From Physical to Standard Intervals

The intervals for space and density, which we have chosen, are obviously not physically realistic. In spectral methods, computations are performed on standard intervals, over which the chosen basis functions are mutually orthogonal. Since orthogonality of the basis functions is the key ingredient of spectral methods, it is convenient to remain on standard intervals for computation. However, real systems are not necessarily defined on standard intervals; thus, it is necessary to define mappings from the physical intervals on which a real system is defined to the standard interval. For the Fourier basis, we use the standard interval $[0, 2\pi]$. We now define a general standard interval $I_s = [\xi_1, \xi_2]$ and a physical interval, $I_p = [a, b]$, where a and b could be a pair of 1-D coordinates along a roadway with $a < b$. We continue to use x to denote the space variable over I_s, and introducing z to denote the space variable over the physical interval, I_p, we define the mapping

$$\psi(z) : I_p \to I_s \quad \text{as} \quad x = \psi(z). \quad (14)$$

This mapping could take the form

$$x = \psi(z) = \frac{z-a}{b-a}(\xi_2 - \xi_1) + \xi_1. \quad (15)$$

Mappings for state variables can be defined analogously. In Section VI, we will discuss such mappings again before presenting the numerical results.

In the following section, we will use the system (12), (13) for data assimilation.

V. Data Assimilation

Here, we describe the experimental data-assimilation algorithm used in this paper and compare it qualitatively to some existing traffic state estimation works.

A. Minimax Framework

The algorithm is based on the minimax state estimation framework; we briefly recall it here. Assume that the state of the process of interest (a vector of projection coefficients in our case) at time instant t is described by vector $X(t)$ and the evolution of $X(t)$ over time is governed by a system of differential equations

$$\frac{dX}{dt} = A(p)X + e^m(t), \quad X(0) = X_0 + e^b \quad (16)$$

where X_0 is the initial condition [vector of initial coefficients obtained by using (11)], e^m represents model error, e^b describes the error in the initial condition, and p is a parameter, possibly depending on the state vector X. In our case, term e^m addresses the projection error introduced by the Galerkin method. In other words, e^m models the impact of higher order Fourier modes onto those described by the Galerkin model (12). Since

$$\int_0^T \int_0^{2\pi} u^2(x,t)dx\,dt < +\infty$$

we can assert that:

1) the vector of projection coefficients $X(t)$ is bounded in $L^2(0,T)$;
2) the error in the initial condition e^b is bounded in $\mathbb{R}^n$.

One can prove (we refer the reader to [6, p. 251] for further details on estimating the projection error) using properties of the conservation law (8) that the projection error e^m is bounded in L^2. Now, assuming that the vector of density observations, i.e., $Y(t)$, is related to the vector of projection coefficients, $X(t)$, through relation (13), and the observation error η is bounded, we propose the following uncertainty description:

$$\left\|Q_0^{\frac{1}{2}}e^b\right\|_{\mathbb{R}^n}^2 + \int_0^T \left\|Q^{\frac{1}{2}}e^m\right\|_{\mathbb{R}^n}^2 + \left\|R^{\frac{1}{2}}\eta\right\|_{\mathbb{R}^m}^2 dt \le 1 \quad (17)$$

where Q_0 and Q are symmetric semidefinite matrices representing worst case bounds on the energy of projection errors, namely, Q_0 accounts for the projection error introduced by projecting the initial condition [see (11)], and Q accounts for the projection error of the Galerkin method. The matrix, i.e., R, which is symmetric and positive-definite, describes the observation error, and $Q^{1/2}$ stands for a square root of the matrix Q. In fact, Q_0, Q, and R represent designer parameters and are assumed to be given *a priori*. We refer the reader to [6, p. 251] for further details on getting *a priori* error bounds for projection errors for the Fourier–Galerkin method.

Now, we are ready to introduce a formal definition for the minimax estimate, namely, $\widehat{X}(t)$ is called a minimax estimate of $X(t)$ iff the following inequality holds for all $\ell \in \mathbb{R}^{N+1}$:

$$\sup_{(e^b,e^m,\eta) \text{ verifying } (17)} \left(\ell^T X(t) - \ell^T \widehat{X}(t)\right)^2 \leq \sup_{(e^b,e^m,\eta) \text{ verifying } (17)} \left(\ell^T X(t) - \ell^T X^*(t)\right)^2$$

where $X^*(t)$ solves (16) for some e^b, e^m, and η verifying (17).

Let us give an intuitive geometric interpretation of the minimax estimator. It is not hard to see that any linear operator transforms an ellipsoid in n-dimensional Euclidean space into another ellipsoid which may be, in particular, stretched or rotated or squeezed. Now, one may recognize that (17) defines an ellipsoid in some functional space, and (16) may be considered as a linear transformation of this ellipsoid, namely, the tuple (e^b, e^m, η) is mapped through (16) to the unique vector $X(t)$ which corresponds to the solution of (16). It is not hard to see that all such $X(t)$ form an ellipsoid E, namely, X^* belongs to the ellipsoid, E, iff there is e^b, e^m and η that verify (17) and the trajectory of (16) corresponding to the tuple (e^b, e^m, η) verifies (13) and passes through X^* at time t. The minimax estimate $\widehat{X}(t)$ is then defined as a central point of the ellipsoid E (so-called Chebyshev center): that is for all $X^*(t)$ solving (16) for some e^b, e^m and η, the following inequality holds true:

$$\left|e_i^T X^*(t) - e_i^T \widehat{X}(t)\right| \leq e_i^T K e_i \tag{18}$$

where e_i is the ith vector of the canonical Euclidean basis, and $K(t)$ is a symmetric positive definite matrix representing the minimax gain. It turns out (see [11] for detailed derivations) that the minimax estimate $\widehat{X}(t)$ and minimax gain K solve the following system:

$$\begin{aligned}
\frac{dK}{dt} &= A(p)K + KA^T(p) + Q^{-1} - KH^T RHK \\
K(0) &= Q_0^+ \\
\frac{d\widehat{X}}{dt} &= A(p)\widehat{X} + K^{-1}H^T R\left(Y(t) - H\widehat{X}(t)\right) \\
\widehat{X}(0) &= X_0
\end{aligned} \tag{19}$$

where Q_0^+ denotes the pseudoinverse of the matrix Q_0 and the minimax gain K defines ellipsoid E. Existence of the unique matrix-valued function $t \mapsto K(t)$ with symmetric positive definite values $K(t)$ follows from classical linear-quadratic control results [22].

Now, we use the following algorithm to estimate the state of the nonlinear (12). Assume that $\widehat{X}^{(1)}(t)$ represents an estimate of the projection coefficients, X (obtained, for instance, solving the conservation law (1) for given initial and boundary conditions). We set $p = \widehat{X}^{(1)}(t)$ in (16). We then denote by $\widehat{X}^{(2)}(t)$ the minimax estimate obtained from (19) with $p = \widehat{X}^{(1)}(t)$. In such a way, we get a mapping of a given estimate, i.e., $\widehat{X}^{(1)}$, into the corresponding minimax estimate $\widehat{X}^{(2)}$. Now, thinking of this mapping as a contraction, we approximate its fixed point, that is, we assign to the minimax estimate $p = \widehat{X}^{(2)}$ a new minimax estimate $\widehat{X}^{(3)}$ and repeat this procedure. This gives us a sequence of estimates, $\{\widehat{X}^{(k)}\}$. Assuming that this sequence is convergent, we can define the estimate of the nonlinear (12) as a limiting point in $L^2(0,T)$ of $\{\widehat{X}^{(k)}\}$. In the sequel, we refer to the aforementioned algorithm as an iterative minimax algorithm. In the following section, we present numerical results on the convergence of the proposed algorithm. We note that the same idea is used in the numerical analysis for advection–diffusion PDEs with nonlinear advection and linear diffusion terms. Namely, time marching (performed after the spacial discretization) is done so that the nonlinear advection term is explicitly treated and is thus linear in the state variable, whereas the linear diffusion term is treated implicitly, which retains linearity and introduces stability [23, Chapter 3]. Now, combining this implicit–explicit scheme with data assimilation, we get a robust parameter estimation algorithm.

B. Qualitative Comparison With Existing Techniques

In Section III, we cited some existing works in traffic state estimation, which use local solution methods. Now that the Fourier–Galerkin reduced model has been derived, and the minimax data-assimilation framework has been described, it is useful to highlight some of the fundamental differences between the proposed approach and some of the existing approaches.

The global nature of our algorithm is a feature that sets it apart from existing techniques, which all use local methods to discretize the model in space. For example, Muñoz *et al.* [1] take a finite-difference discretization of a scalar conservation law with a piecewise linear flux function (termed the cell transmission model [24]) and use it to derive a switched-linear system. Assumptions on the number of congested/free-flow regions, which exist on a link, are required in order to arrive at such a system. In our case, we do not need to make such assumptions, allowing us to exploit the full features of the traffic flow model within the estimation framework, without compromising the dynamics described by the model. As mentioned previously, the continuous-in-space property of our algorithm is another feature of the "global" approach, which distinguishes it from state-of-the-art techniques. This gives us more flexibility in our observation locations. Of course, in finite-difference/volume-based discretization schemes, grids can be made fine enough to overcome accuracy issues arising from sensor location, but this will result in higher computational costs. For estimators where the spatial discretization is

of the order of tens of meters, the benefits brought about by the use of global methods may indeed be significant.

The estimation algorithm, which we present, is fully nonlinear. In this regard, the closest "local" state estimation works to this are those which make use of particle filter [5] or Ensemble Kalman Filter [25]; thus, a comparison with such a technique would be informative. We will return to this in Section VI-D.

VI. Numerical Implementation and Results

Here, we present some numerical results for the Fourier–Galerkin reduced model and minimax data assimilation, demonstrating the shock capturing capabilities of the iterative algorithm.

Before presenting the numerical results, we will briefly return to the subject of standard and physical intervals. In Section IV-C, we discussed the use of a mapping function, $\psi(\cdot)$, which can transform coordinates in the physical space onto a standard interval, on which computations are performed. Here, it should be understood that the interval, $[0, 2\pi]$, on which x is defined can be mapped to and from a physical interval whose limits define the start and endpoints of a section of roadway. In addition, we have previously defined $u \in [0, 1]$. The lower limit of this interval will be the same in physical space, as it corresponds to the empty road condition. The upper limit is mapped to the jam density, u_m, which is one of the parameters that must be defined for a physically meaningful flux function, (3). Typical values of u_m range from 0.2 to 0.3 veh/m. The other parameter that must be set is V_m, which is the maximum allowed speed on the road. The velocity is also defined on the standard interval, [0, 1]. Again, the lower limit of this interval coincides with the physical one, whereas the upper limit corresponds to V_m and can be mapped back and forth through the use of a suitable mapping function. Setting realistic parameters and defining the correct mapping functions is important for implementing the method in practice, but since we are concerned with the computational aspects of data assimilation in this paper, the following numerical results are displayed on standard intervals.

A. Solution of Fourier–Galerkin Reduced Model

We first get the initial coefficients by computing the integral (11). This can be done by quadratures or by using fast Fourier transform. We start with the initial data

$$u_0(x) = e^{-20(x-1)^2} \tag{20}$$

which is a Gaussian, centered at $x = 1$. The viscosity term in (10) is only activated for the higher modes [7], $|n/2| \geq m$, where m is the wavenumber beyond which the viscosity is added. This can be achieved by replacing the viscosity term with $-H[|n/2| - m]\varepsilon n^2 a_n(t)$, where $H[\cdot]$ is the Heaviside step function. Tadmor [7] solved Burgers' equation using $\varepsilon m \approx 0.25$. Having obtained the $N + 1$ initial coefficients, $a_n(0)$, for the initial data, (20), we solve the system, (10), using a fourth-order Runge–Kutta scheme (RK4) with $N = 100$ and $m = 30$.

Fig. 1 shows the solution of the initial value problem for the Fourier–Galerkin model with periodic boundary conditions at different time instants. In addition, shown is the solution of the same initial value problem using a first-order Godunov scheme. Fig. 1(a) shows the smooth initial profile, and Fig. 1(b) shows the profile steepening at $t = 0.1$, and a shock beginning to form. We see that oscillations start to form here due to the appearance of the shock discontinuity. In Fig. 1(c), we see that at $t = 1$, the solution is a shock wave and a rarefaction wave, and that oscillations are still present. However, the solution is still stable. Fig. 1(d) shows the solution at $t = 5.7$. The oscillations have now been damped out, except near the shock. However, since the Fourier–Galerkin solution still coincides with the Godunov solution, the former appears to capture shocks successfully. The inclusion of periodic boundary conditions is evident from the last figure, where the flow is observed to exit through the right boundary and reenter through the left boundary.

B. Iterative Minimax Algorithm: Numerical Results

To perform the data assimilation, we use the Godunov solution of (1) with prescribed initial data to mimic the actual traffic. We solve the Fourier–Galerkin model numerically (as was done in Section VI-A) using a perturbed initial condition, and store its solution over the computational time-window. We then sample the Godunov solution, and perturb the values artificially with additive noise in order to generate measurements. Once the Fourier–Galerkin and Godunov solutions have been precomputed over the same time-window, we launch the iterative minimax algorithm. In other words, we use the Godunov solution as "ground truth" to generate observations by sampling its values and assimilating them into the Fourier–Galerkin model. Although the Godunov and Fourier–Galerkin methods are approximating the same PDE, (1), the way in which each one computes spatial derivatives is different. As a result, the observed data come out of a different numerical scheme (Godunov) than the one used to construct the reduced model (Fourier–Galerkin). This makes the present experiment more realistic.

Before describing the iterative minimax algorithm (which was briefly introduced in Section V), we first define some quantities.

The ground truth (Godunov solution) is stored over the simulation period as $\boldsymbol{U}$. We use $\boldsymbol{U}_i$ to refer to the Godunov solution of the initial value problem at time step i (i.e., $\boldsymbol{U}_0$ is a grid function approximating $u_0(x)$). The solution, $\boldsymbol{U}_i$ is a grid function over a uniform spatial grid, Γ (used for the Godunov scheme). However, the observations are taken over a (possibly sparser) set of grid points, $\Pi(\Gamma) := \{x_1, x_2, \ldots, x_M\}$, where $0 < x_1 < x_2 < \cdots < x_M \leq 2\pi$. We use $\boldsymbol{U}_i^{obs}$ to denote the set of M observations at time step i, taken over $\Pi(\Gamma)$. Finally, we set $Y_i = \boldsymbol{U}_i^{obs} + \eta_i$, where η_i is additive noise.

The Fourier–Galerkin reduced model, (12), is solved with RK4, using the same computational time step as the Godunov scheme aforementioned. This is done in order to ensure coherence between the model (Fourier–Galerkin) and the "truth" (Godunov). We solve (12) with initial data, $X_0 + e^b$, where X_0 is the vector of coefficients obtained from the data, (2), using (11), and e^b is additive noise. The solution of this initial value problem stored over the full simulation period is denoted as $\widehat{X}^{(1)}$ and provides an initial estimate of the density over

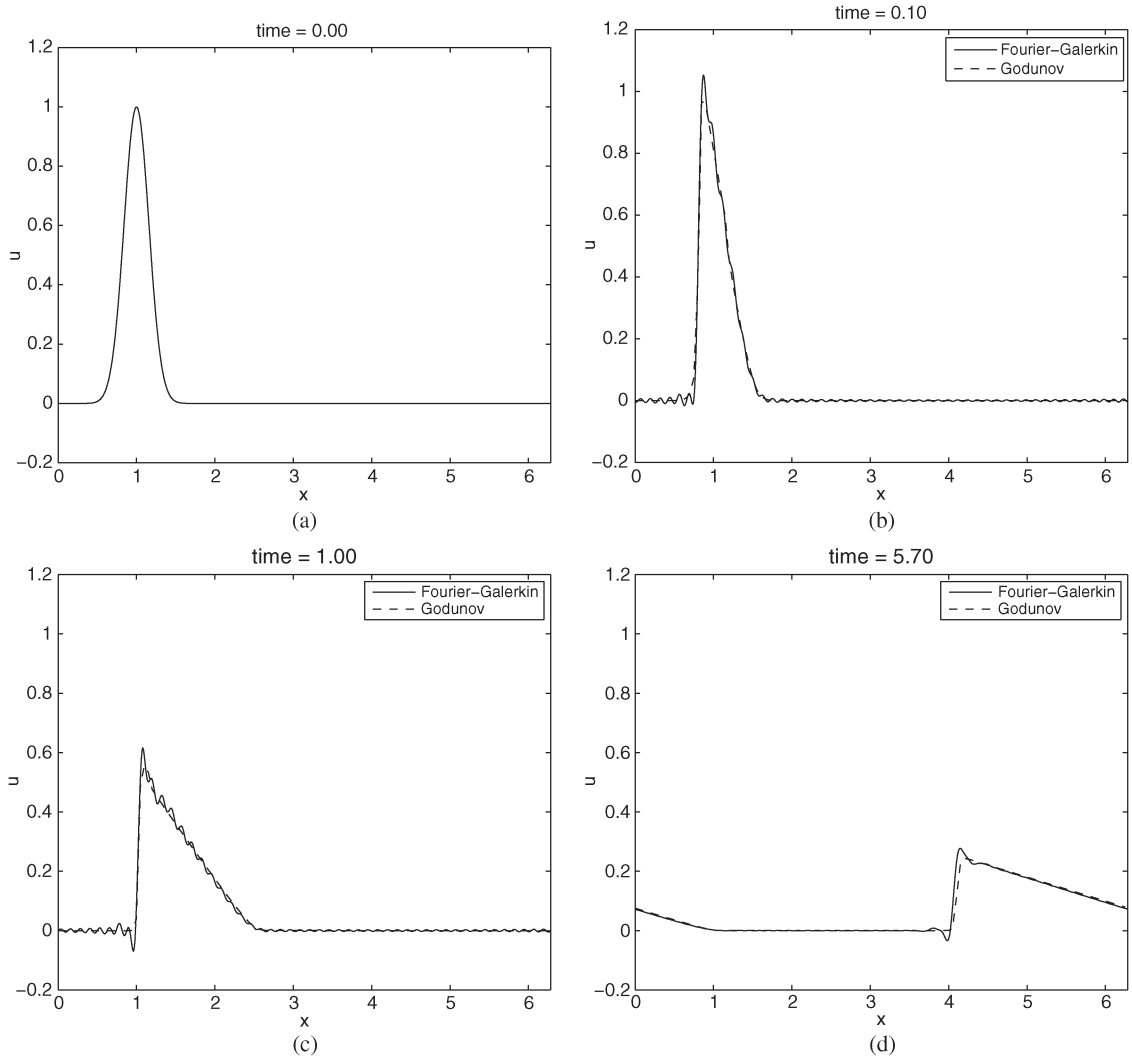

Fig. 1. Numerical solution to LWR with initial data (20). (a) Initial profile. (b) Shock forming at $t = 0.1$. (c) Shock-rarefaction solution at $t = 1$. (d) Solution at $t = 5.7$ showing periodic B.C.

the simulation time-window. At each of the T time steps of the Fourier–Galerkin solution computation, we construct the operators, A and H, using (10) and (6), respectively.

With the aforementioned quantities in mind, the iterative minimax algorithm is summarized in Algorithm 1. For details on the midpoint method used therein, see Appendix A. We initialize the algorithm using the data, (20), to generate the "truth", $\boldsymbol{U}$, and using the Fourier–Galerkin model with initial data, $u_0(x) + \sin^2(x)$ to generate $\widehat{X}^{(1)}$. The two initial profiles ("truth" and guess) are shown in Fig. 2(a). It should be noted that the perturbed initial density exceeds its maximum value, that is $u_m = 1$. Although this will initially result in an unphysical solution (as $u \in [0, u_m]$), we study this case nevertheless in order to demonstrate the efficacy of our approach. The Fourier–Galerkin solution and ground truth are shown at $t = 3$ and $t = 8$ in Fig. 2(b) and (c), respectively. We see that the two densities significantly diverge; the Fourier–Galerkin solution of the conservation law develops two shocks, whereas only one is present in the ground truth. We now implement Algorithm 1 with the following parameters:

- $R = 0.5I$, i.e., we allow for small noise in observations;
- $Q = 0.5I$, i.e., we assume that the projection error is small;
- $Q_0 = 0I$, i.e., we do not trust our initial condition at all.

We use I to denote the identity matrix of appropriate dimension. The number of observations, i.e., M, is taken to be 5 (due to the periodic boundary conditions, the observations at the left and right boundaries are the same), and these are uniformly distributed over the spatial domain. To numerically construct $\widehat{X}_k$ we apply the implicit midpoint method to (16)–(19) as was suggested in [12]. We refer the reader to Appendix A for the implementation details.

Algorithm 1 Iterative minimax algorithm

Compute $\widehat{X}^{(1)}$ and $\boldsymbol{U}$
for $i = 1$ to T **do**
  $\boldsymbol{U}_i^{obs} \leftarrow \boldsymbol{U}_i$ restricted to $\Pi(\Gamma)$
  $Y_i \leftarrow \boldsymbol{U}_i^{obs} + \eta_i$
end for

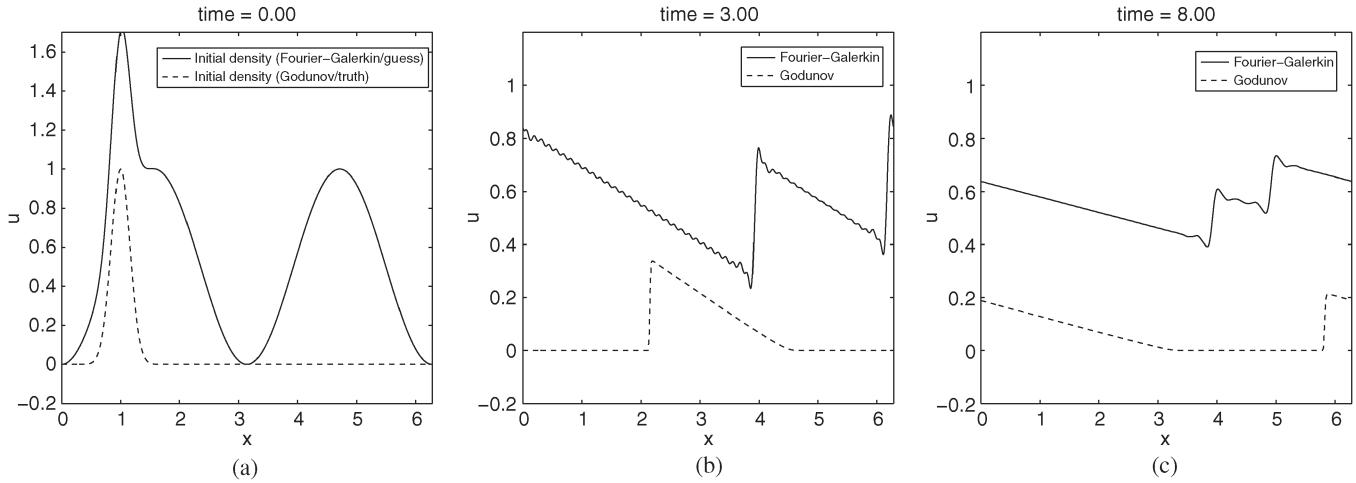

Fig. 2. Fourier–Galerkin (model) and Godunov ("truth") running over same time interval. (a) Initial data (truth and guess). (b) Solutions at $t = 3$. (c) Solutions $t = 8$.

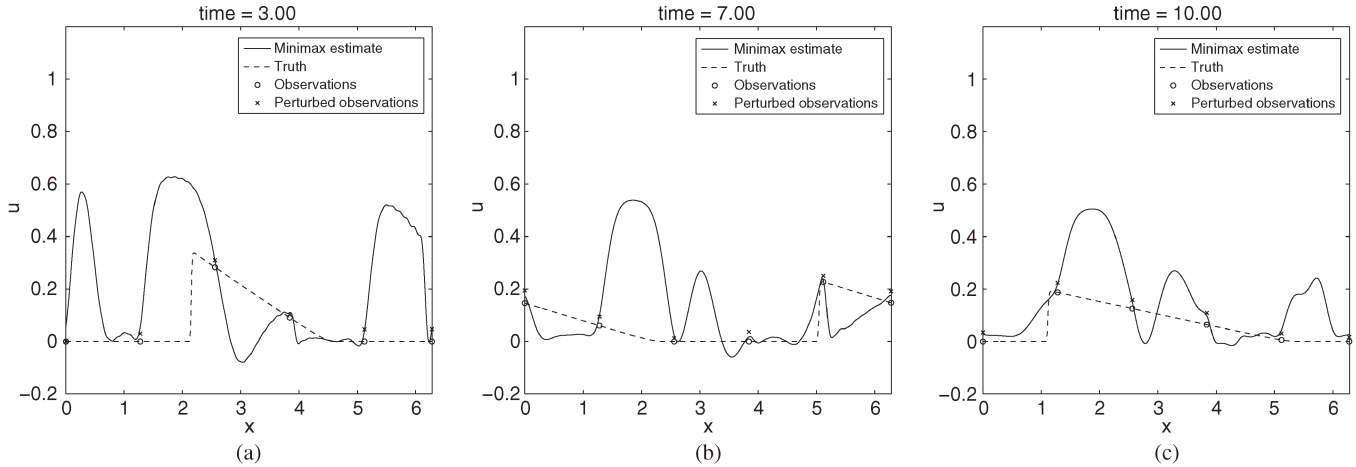

Fig. 3. First iteration of minimax estimate and ground truth. (a) Estimate and truth at $t = 3$. (b) Estimate and truth at $t = 7$. (c) Estimate and truth at $t = 10$.

$Y \leftarrow (Y_1 \dots Y_T)^T$
$i \leftarrow 1$
repeat
Get (discrete) minimax estimate, $\widehat{X}^{(i+1)}$, applying the implicit midpoint method to (16)–(19), with $p = \widehat{X}^{(i)}$ and Y.
Compute the distance between successive estimates, $\nu = \|\widehat{X}^{(i+1)} - \widehat{X}^{(i)}\|_{L^2}$
i++
until $\nu \ll \mu$ for a given threshold, $\mu > 0$.

The results at three different times for the first iteration are shown in Fig. 3. We see that for $k = 2$ (that is the first iteration), $H\widehat{X}_k$ does not converge to the "truth"; the relative L^2 error at the end-time (which is computed over a dense spatial grid and is thus not restricted to the observation points) is ≈ 5.7420. However, the estimate conforms to the truth in the vicinity of the observations, which is due to our choice of R. The results of the third iteration are shown in Fig. 4. We now observe much better convergence of the algorithm (the relative L^2 error at the end-time is ≈ 0.3841). At $t = 3$, the estimate contains a shock followed by a rarefaction; this rarefaction is close to that observed in the ground truth, but the location of the shock is incorrect. The estimate also contains a spike that does not appear in the ground truth. By $t = 7$, the spike in the density estimate has disappeared, and the shock rarefaction observed in the ground truth is captured in the estimate. The shock is then tracked for future time in the estimate;[1] we see that the estimate at $t = 10$ still matches the ground truth closely.

Note that, in this example, the observations are sparse over the spatial domain; this sparsity is such that the shock profile cannot be reconstructed from the observations alone using conventional statistical techniques. While the initial data for the Fourier–Galerkin model are very noisy (in fact, we do not even assume that we have reliable information about initial traffic density), the estimate, which combines its solution with

[1]For a video of the third minimax iteration, showing the tracking of the shock, see the first author's website.

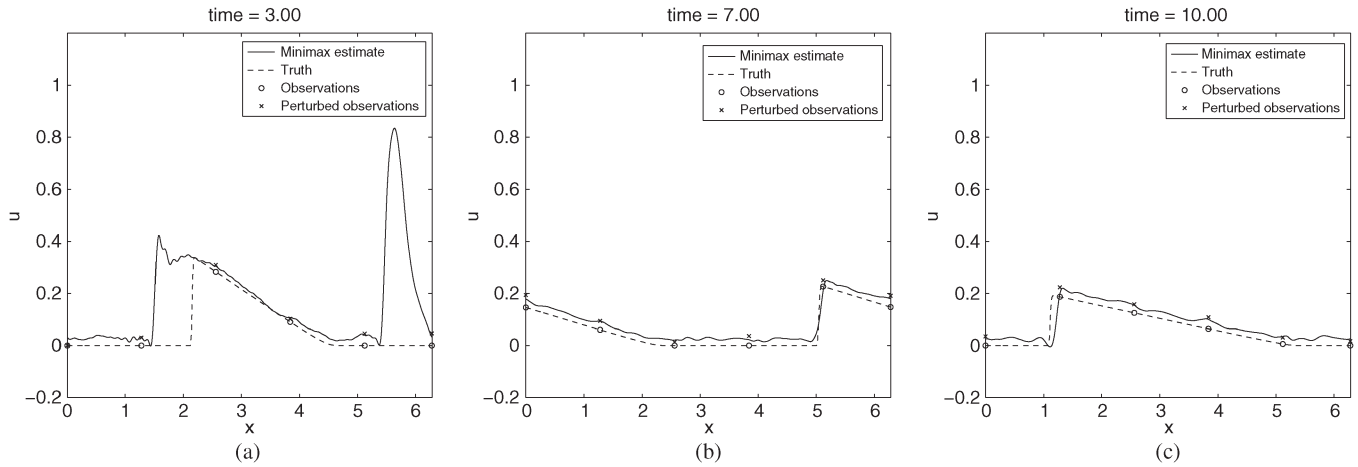

Fig. 4. Third iteration of minimax estimate and ground truth. (a) Estimate and truth at $t = 3$. (b) Estimate and truth at $t = 7$. (c) Estimate and truth at $t = 10$.

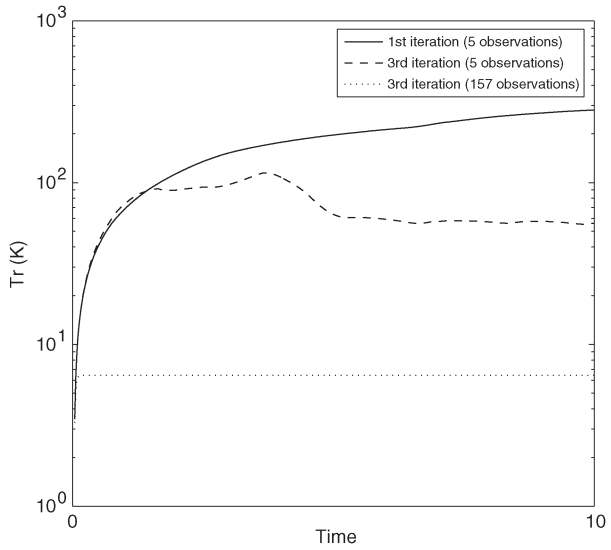

Fig. 5. $Tr(K)$ over time for different iterations.

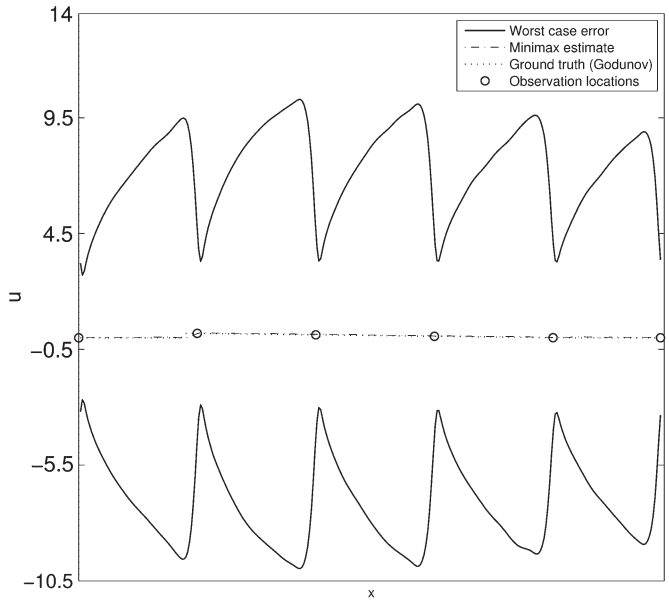

Fig. 6. Minimax uncertainty estimate for $u(x, T)$, $x \in [0, 2\pi]$.

observations, captures the shock; this is evident from Fig. 4(b) and (c) where we see that the observations do not give the precise location of the shock, yet the shock is captured well in the estimate. In addition, the noise in the initial condition is so large that spurious shocks develop in the Fourier–Galerkin solution (see Fig. 2), but these are all damped out in the estimate after only 3 iterations. This demonstrates the shock-capturing capability of the iterative minimax algorithm, with very sparse observations and a high degree of uncertainty in the initial conditions.

Fig. 5 shows the trace of the (discrete) minimax gain $(Tr(K))$ over the data-assimilation time-window for three different cases, plotted on a log scale in y. The first case is the first iteration of the minimax algorithm with five observations. The second and third cases are of the third iteration with 5 and 157 observations, respectively. Although it is difficult to see in the figures, each series starts at zero. We see that for the case with many observations, the gain stabilizes after a short time $(Tr(K) \approx 6)$, and the uncertainty level for this case [as described by (18)] is much lower than for the cases with fewer observations, as one would expect.

Fig. 6 shows the worst case error bounds as well as the estimate and truth for the final time step of the third iteration. Namely, for the fixed time instant, we plot the values of the density over space, the estimate of the density and uncertainty bounds (18): each uncertainty bound is centered around the value of the estimate. It is difficult to distinguish the estimate and the truth, as the scale of the figure is dominated by the worst case error. Note that the figure shows how the uncertainty in the parameters is carried over by the conservation law. In particular, near the observation locations, the worst case error is significantly smaller than elsewhere, as expected.

An important property of the observation operator, H, is that it can be constructed from (6) in such a way as to map from the state space to an observation space, in which the measurements are taken at any points on the interval $x \in [0, 2\pi]$. This is in contrast to finite-difference-based methods, which require

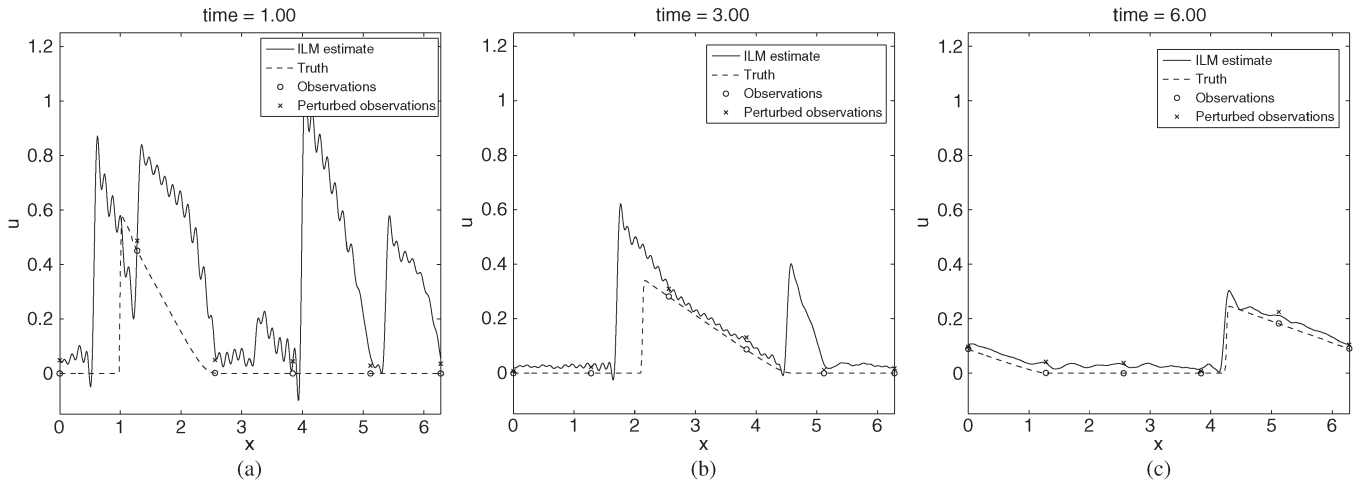

Fig. 7. Estimates at three points in time for ILM. (a) ILM estimate and truth at $t = 1$. (b) ILM estimate and truth at $t = 3$. (c) ILM estimate and truth at $t = 6$.

measurements from fixed points in space, which ideally should coincide with the grid points of the spatial discretization.

C. Implicitly Linearized Online Minimax Filter

Having observed the convergence of the iterative minimax algorithm, we now consider a potential online implementation. Since the algorithm presented is iterative in nature, it is not ideally suited to the real-time estimation problem, as it requires multiple computations to be performed over the same time-window. Clearly, for real-time data assimilation, where estimates must be instantaneously produced, the algorithm may not be directly applicable, although due to the fact that the Fourier–Galerkin model can operate much faster than real time, it should be possible to design a quasi-real-time implementation based on the iterative minimax algorithm. Here, we will apply the minimax filter in an "online" fashion by linearizing about each new state estimate and propagating the solution of the resulting system forward in time using the implicit midpoint method. Applied in this way, the minimax filter is now, in a sense, analogous to the extended Kalman filter (EKF). However, the linearization employed in our technique differs from that employed in the EKF, resulting in a stable scheme (which would not necessarily be true for standard EKF linearization), which is summarized in Algorithm 2, where $\widehat{X}_i$ is used to denote the solution of the second equation in (19) at the ith time step. This is not to be confused with $\widehat{X}^{(i)}$, which appears in the iterative minimax algorithm and was used to denote the solution over the simulation period for the ith iteration. See Appendix B for details on the implicit midpoint method employed in Algorithm 2. We apply Algorithm 2 on the same initial data and observation sparsity that was used in Section VI-B. The results for this are shown in Fig. 7. We see that at $t = 1$, the estimate is still far from the truth as it recovers from the very imprecise initial condition. At $t = 3$, we see that the shock present in the truth has still not been correctly approximated, and that a second spurious shock remains. By $t = 6$, the spurious shock has been damped out and the true shock has been well-approximated. The filter tracks the true state with good accuracy from then onward. This example demonstrates the potential of the Fourier–Galerkin and minimax approach for real-time application.

Algorithm 2 Implicitly linearized minimax

Define initial data, u_0
Generate observations over simulation time-window, $\boldsymbol{U}$
Project $u_0 + \sin^2(x)$ into Fourier basis to obtain X_0
$\widehat{X}_0 \leftarrow X_0$
for $i = 1$ to T **do**
  $\boldsymbol{U}_i^{obs} \leftarrow \boldsymbol{U}_i$ restricted to $\Pi(\Gamma)$
  $Y_i \leftarrow \boldsymbol{U}_i^{obs} + \eta_i$
  Assemble $A(\widehat{X}_{i-1})$
  Advance system (19) forward in time using implicit mid point method to obtain $\widehat{X}_i$
end for

D. Numerical Comparison With Ensemble Kalman Filter

What remains now is a comparison with state-estimation techniques based on local methods. In Sections III and V-B, we cited a number of state-of-the-art techniques and highlighted some differences/similarities to the proposed method. We noted that the ensemble Kalman filter (EnKF), which represents a particular case of a fully nonlinear particle filter, has been employed for the traffic state estimation problem [25]. Since for a wide class of Markov diffusion processes, it makes sense to assume local normality [26], the EnKF assumes that the state of the process at time instant $t + 1$ is a Gaussian random vector with mean given by the state predicted by the model (the state obtained by applying the model to the estimate of the state vector at time t) and second moments defined by the distribution of the model error, which is assumed to be Gaussian as well. Now, the observed vector is also a Gaussian with mean defined by the linear combinations of the actual state vector and covariance given by the noise characteristics. Given all these, the prediction of EnKF is just an ensemble mean conditioned on observed data, and with a covariance matrix that is a

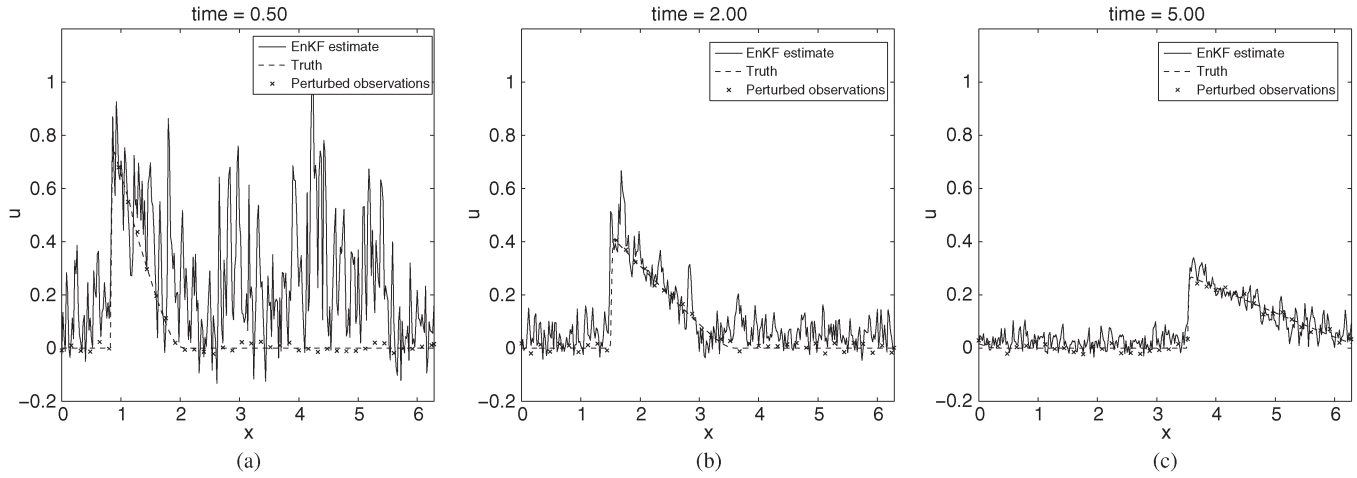

Fig. 8. Estimates at three points in time for EnKF with 40 observations. (a) EnKF estimate and truth at $t = 0.5$. (b) EnKF estimate and truth at $t = 2$. (c) EnKF estimate and truth at $t = 5$.

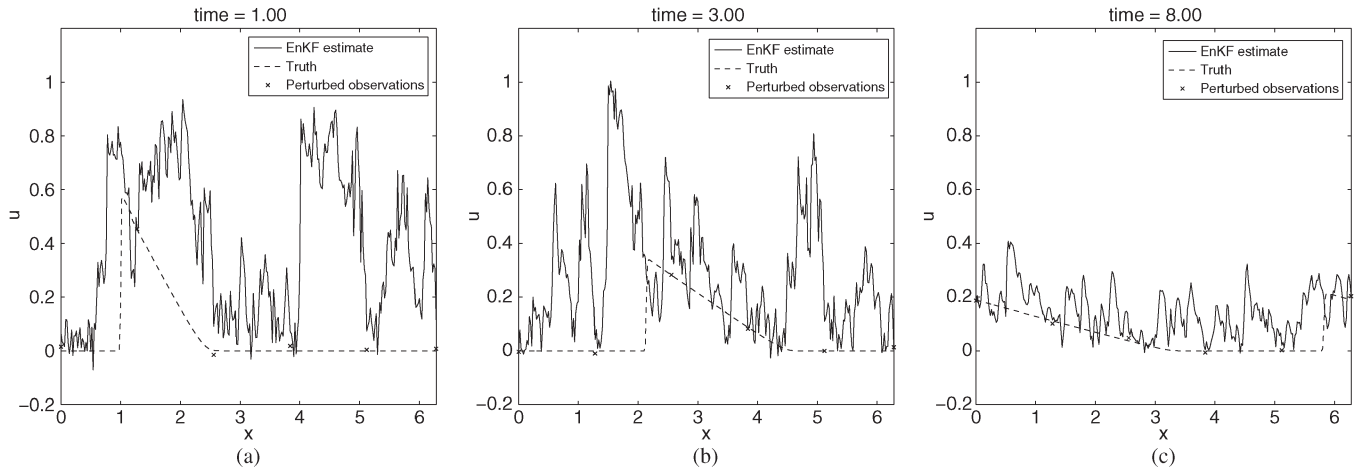

Fig. 9. Estimates at three points in time for EnKF with five observations. (a) EnKF estimate and truth at $t = 1$. (b) EnKF estimate and truth at $t = 3$. (c) EnKF estimate and truth at $t = 8$.

combination of the empirical covariance matrix arising from ensembles with the observation error covariance. We stress that assuming a Gaussian model error is equivalent to adding a "diffusion term" (the Laplacian differential operator in the most simple case) to the original conservation law. Moreover, the statistics of the discretization/projection error is usually not available, so that the setting proposed in this paper (that is just to assume that the error is bounded) is more realistic and therefore robust.

Here, we will apply EnKF to the same test case that we studied for both the iterative and implicitly linearized minimax (ILM) algorithms, and compare the results. The observation and error noise was taken to be the same as that used for the iterative and ILM algorithms. Since we wish to provide a comparison of our proposed algorithm with a *local* method, we apply EnKF to the Godunov discretization of the data. See Appendix C for details on this algorithm. The results for the algorithm applied to the test case with 40 observations over space are shown in Fig. 8. We see that the EnKF does appear to converge to the truth, and even capture the shock, using 40 observations. However, that number of observations is large compared with the number used for the iterative and ILM test cases. With 40 observations, the shock is well-localized by the data themselves, so the role of the model is not significant. We now drop the number of observations in space to 5 for a direct comparison with the minimax algorithms (see Fig. 9). We see that this time, the EnKF algorithm does not converge to the truth. While the estimate does recover somewhat from the very imprecise initial condition like before, it does not provide a good estimate of the solution. In addition, the EnKF estimate is quite noisy; there is no smoothness even in regions, where the density should be smooth. In the case with 40 observations where the shock is captured, observations are present at the location of the discontinuity; when the observations become sparser in space, the shock is not well-approximated at all. This is likely due to the local nature of the discretization scheme.

VII. Conclusion and Future Work

We distinguish between finite-difference-based traffic state estimation, and estimation based on global methods, such as the Fourier–Galerkin projection method presented here. The approach looks promising. The continuous-in-space reconstruction does not require the observation locations to coincide with

grid points, giving the method the ability to take measurements from any part of the spatial domain or even address moving sensors. This would appear to give the method a considerable advantage over finite-difference-based methods, in which measurements must be taken at grid points.

One striking feature of the method is its ability to track shocks with few observations, and high uncertainty in the initial traffic density. In the example studied, the initial guess of the density contained wild oscillations not present in the "true" initial density. In spite of the deviation between the true and guessed density profiles, the iterative minimax algorithm was capable of tracking the shock observed in the "true" density using very sparse observations. It should also be noted that the observations by themselves would not have been sufficient to reconstruct the density profile, indicating that the system dynamics provided by the Fourier–Galerkin model play a vital role. This highlights a considerable advantage that a model-based method such as this one possesses over purely data-driven estimators. In addition, any spurious shocks that were generated due to the noise in the guessed initial density were eradicated.

A point we must address is the choice of periodic boundary conditions. Clearly, for most practical applications of the proposed method, nonperiodic boundary conditions will be required; for instance, in the modeling of network traffic, the left (resp. right) boundary condition of a link should be informed by the state of its predecessor (resp. successor) link(s). The physical interpretation of periodic boundary conditions in this problem is that of a ring-road with no exits/entries. This test case has been studied in the analysis of the stability of microscopic models and has generated important results [27]. Taking periodic basis functions as a test bed, we observe convergence of the proposed iterative minimax algorithm for a particular case of quadratic flux. Based on this, we conjecture that the convergence results seen here still hold the same validity that they would under the incorporation of more general boundary conditions and strictly concave fluxes. We leave the rigorous study of this topic for further research.

Appendix A
Implicit Midpoint Method for Iterative Minimax Algorithm

Equation (19) is solved using the following midpoint method (a detailed derivation is presented in [12]):

$$Init:\ n=0,\ dt:=\frac{T}{M},\ M>0,\ \widehat{X}(0)=X_0$$

$$Update:\ n\to(n+1):$$

$$\widehat{X}\left((n+1)dt\right)=2\bar{X}-\widehat{X}(ndt)$$

$$\bar{X}=\widehat{X}(ndt)+\bar{A}(p)\bar{X}+\bar{V}\bar{U}^{-1}H^{T}R(\bar{Y}-H\bar{X})$$

where $\bar{Y}=(Y((n+1)dt)+Y(ndt)/2)$; thus, the approximation of the value of Y at the midpoint of the interval $(dtn, dt(n+1))$, and $\bar{A}(p)=(A(\widehat{X}^{(i)}(dt(n+1)))+A(\widehat{X}^{(i)}(dtn))/2)$ (here p stands for the minimax estimate which comes from the previous iteration so $p(t)=\widehat{X}^{(i)}(t)$ and the aforementioned algorithm computes $\widehat{X}^{(i+1)}(t)$), and $\bar{U}$, $\bar{V}$ are evaluated as

$$Init:\ U(0)=I,\ V(0)=Q_0^{+},\ K(0)=V(0)U^{-1}(0)$$
$$Update:\ n\to(n+1):$$
$$\begin{pmatrix}U\left(dt(n+1)\right)\\ V\left(dt(n+1)\right)\end{pmatrix}=2\begin{pmatrix}\bar{U}\\ \bar{V}\end{pmatrix}-\begin{pmatrix}I\\ K(n)\end{pmatrix}$$
$$\begin{pmatrix}\bar{U}\\ \bar{V}\end{pmatrix}=\begin{pmatrix}I\\ K(n)\end{pmatrix}+\begin{pmatrix}-\bar{A}(p)^{T} & H^{T}RH\\ Q^{-1} & \bar{A}(p)\end{pmatrix}\begin{pmatrix}\bar{U}\\ \bar{V}\end{pmatrix}$$
$$K(n):=V(dtn)U^{-1}(ndt).$$

Appendix B
Implicit Midpoint Method for ILM Algorithm

This numerical scheme is exactly the same as the one presented in Appendix A but $\bar{A}(p)$ is now defined differently, namely, $\bar{A}(p):=A(X(ndt))$. This is to say that the projection coefficients are assumed to be constant over small time intervals.

Appendix C
Ensemble Kalman Filter (EnKF) Algorithm

The EnKF algorithm is summarized as follows:

1) Approximate initial data, $u_0(x)+\sin^2(x)$, by grid function U_0^0.
2) Taking U_0^0 as the mean of a Gaussian prior distribution, draw K samples $(U_0^i,\ i=1\dots K)$ from that distribution.
3) Use Godunov scheme to predict all $K+1$ grid functions at the next time step $(U_1^i,\ i=0\dots K)$ using each initial grid function $(U_0^i,\ i=0\dots K)$
4) Compute the ensemble mean, $\bar{U}_1$, of $U_1^i,\ i=0\dots K$.
5) Compute empirical estimate of ensemble variance, P_1, from $\bar{U}_1$ and ensemble members.
6) Compute the Kalman gain

$$G_1=P_1H_1^{T}\left(H_1P_1H_1^{T}+R\right)^{-1}$$

where R is the observation error covariance as before, and H_1 is a suitable observation matrix for time step 1.

7) The $K+1$ grid functions are updated with new data as

$$\widehat{U}_1^i=U_1^i+\left(Y_1-H_1U_1^i+\varepsilon_1\right),\quad i=0\dots K$$

where ε_1 is the model error.

Steps 3)–8) are then repeated in order to obtain estimates are time steps 2, 3, ... Finally, the estimate of the density at time n is given by the ensemble mean.

References

[1] L. Muñoz, X. Sun, R. Horowitz, and L. Alvarez, "Traffic density estimation with the cell transmission model," in *Proc. Amer. Control Conf.*, 2003, vol. 5, pp. 3750–3755.

[2] X. Sun, L. Muñoz, and R. Horowitz, "Highway traffic state estimation using improved mixture Kalman filters for effective ramp metering control," in *Proc. 42nd IEEE Conf. Decis. Control*, 2003, vol. 6, pp. 6333–6338.

[3] X. Sun, L. Muñoz, and R. Horowitz, "Mixture Kalman filter based highway congestion mode and vehicle density estimator and its application," in *Proc. Amer. Control Conf.*, 2004, vol. 3, pp. 2098–2103.

[4] Y. Wang and M. Papageorgiou, "Real-time freeway traffic state estimation based on extended Kalman filter: A general approach," *Transp. Res. B, Methodol.*, vol. 39, no. 2, pp. 141–167, Feb. 2005.
[5] L. Mihaylova, R. Boel, and A. Hegyi, "Freeway traffic estimation within particle filtering framework," *Automatica*, vol. 43, no. 2, pp. 290–300, Feb. 2007.
[6] J. Hesthaven, D. Gottlieb, and S. Gottlieb, *Spectral Methods for Time-Dependent Problems*. Cambridge, U.K.: Cambridge Univ. Press, 2007.
[7] E. Tadmor, "Convergence of spectral methods for nonlinear conservation laws," *SIAM J. Numer. Anal.*, vol. 26, no. 1, pp. 30–44, Feb. 1989.
[8] M. Milanese and R. Tempo, "Optimal algorithms theory for robust estimation and prediction," *IEEE Trans. Autom. Control*, vol. 30, no. 8, pp. 730–738, Aug. 1985.
[9] F. L. Chernousko, *State Estimation for Dynamic Systems*. Boca Raton, FL, USA: CRC, 1994.
[10] A. Nakonechny, "A minimax estimate for functionals of the solutions of operator equations," *Arch. Math. (Brno)*, vol. 14, no. 1, pp. 55–59, 1978.
[11] A. Kurzhanski and I. Vályi, *Ellipsoidal Calculus for Estimation and Control Systems & Control: Foundations & Applications*. Boston, MA, USA: Birkhäuser, 1997.
[12] S. Zhuk, J. Frank, I. Herlin, and R. N. Shorten, *State Estimation for Linear Parabolic Equations: The Minimax Projection Method*, submitted for publication.
[13] S. Zhuk, "Minimax projection method for linear evolution equations," in *Proc. 52nd IEEE Conf. Decis. Control*, 2013, pp. 2556–2561.
[14] I. Herlin, O. Nakonechnyi, and S. Zhuk, "Minimax optical flow estimation from a sequence of 2D images," in *Proc. 20th PDMU*, Brno, Czech Republic, 2012, pp. 1–13.
[15] C. Bardos and O. Pironneau, "Data assimilation for conservation laws," *Methods Appl. Anal.*, vol. 12, no. 2, pp. 103–134, Jun. 2005.
[16] M. J. Lighthill and G. B. Whitham, "On kinematic waves II. A theory of traffic flow on long crowded roads," *Proc. R. Soc. London. Ser. A, Math. Phys. Sci.*, vol. 229, no. 1178, pp. 317–345, May 1955.
[17] P. I. Richards, "Shock waves on the highway," *Oper. Res.*, vol. 4, no. 1, pp. 42–51, Feb. 1956.
[18] B. D. Greenshields, "A study of traffic capacity," in *Proc. HRB*, 1935, vol. 14, pp. 448–477.
[19] J. C. Herrera and A. M. Bayen, "Incorporation of Lagrangian measurements in freeway traffic state estimation," *Transp. Res. B, Methodol.*, vol. 44, no. 4, pp. 460–481, May 2010.
[20] Y. Yuan, J. W. C. Van Lint, R. Wilson, F. van Wageningen-Kessels, and S. Hoogendoorn, "Real-time Lagrangian traffic state estimator for freeways," *IEEE Trans. Intell. Transp. Syst.*, vol. 13, no. 1, pp. 59–70, Mar. 2012.
[21] F. van Wageningen-Kessels, Y. Yuan, S. P. Hoogendoorn, H. van Lint, and K. Vuik, "Discontinuities in the Lagrangian formulation of the kinematic wave model," *Transp. Res. C, Emerg. Technol.*, vol. 34, pp. 148–161, Sep. 2013.
[22] T. Reid, *Riccati Differential Equations*. New York, NY, USA: Academic Press, 1972.
[23] R. Peyret, *Spectral Methods for Incompressible Viscous Flow*, vol. 148. New York, NY, USA: Springer-Verlag, 2002, ser. Applied Mathematical Sciences. [Online]. Available: http://books.google.co.uk/books?id=qaHBYoTlpBEC
[24] C. F. Daganzo, "The cell transmission model: A dynamic representation of highway traffic consistent with the hydrodynamic theory," *Transp. Res. B, Methodol.*, vol. 28, no. 4, pp. 269–287, Aug. 1994.
[25] D. Work *et al.*, "An ensemble Kalman filtering approach to highway traffic estimation using GPS enabled mobile devices," in *Proc. 47th IEEE Conf. CDC*, Dec. 2008, pp. 5062–5068.
[26] I. Gihman and A. Skorokhod, *Introduction to the Theory of Random Processes*. New York, NY, USA: Dover, 1997, ser. Dover Books on Mathematics.
[27] M. Bando, K. Hasebe, A. Nakayama, A. Shibata, and Y. Sugiyama, "Dynamical model of traffic congestion and numerical simulation," *Phys. Rev. E*, vol. 51, no. 2, pp. 1035–1042, Feb. 1995.

Tigran T. Tchrakian received the B.A.I., M.Sc., and Ph.D. degrees from Trinity College Dublin, Dublin, Ireland, in 2005, 2007, and 2010, respectively.

He is a Research Scientist with IBM Research, Ireland, where he works on applications of partial differential equations to problems in "smarter sities" domains.

Sergiy Zhuk received the B.A., M.Sc., and Ph.D. degrees in applied mathematics from Taras Shevchenko National University of Kyiv, Kiev, Ukraine, in 2000, 2002, and 2006, respectively.

He was an Alain Bensoussan (ERCIM) Fellow with INRIA Paris-Rocquencourt and Marie Curie (ERCIM) Fellow with CWI, Amsterdam, The Netherlands, in 2009 and 2011, respectively. He is currently a Research Scientist with IBM Research, Ireland, where he works on parameter estimation for partial differential equations with applications in "smarter cities" domains (water, energy, and transportation).