# VARIATIONAL DATA ASSIMILATION FOR DISCRETE BURGERS EQUATION

AMIT APTE, DIDIER AUROUX, AND MYTHILY RAMASWAMY

Abstract. We present an optimal control formulation of the data assimilation problem for the Burgers' equation, with the initial condition as the control. First the convergence of the implicit Lax-Friedrichs numerical discretization scheme is presented. Then we study the dependence of the convergence of the associated minimization problem on different terms in the cost function, specifically, the weight for the regularization and the number of observations, as well as the *a priori* approximation of the initial condition. We present numerical evidence for multiple minima of the cost function without regularization, while only a single minimum is seen for the regularized problem.

## 1. Introduction

In recent years, there has been great interest in development of methods aimed at blending together sophisticated computational models of complex systems and the vast amounts of data about these systems that is now commonly available. E.g., Satellites provide detailed observations of the atmosphere and the oceans.[15] As a result, *data assimilation*, which refers to the process of combining data and the model output, has received a lot of attention not only from researchers in sciences and engineering, but also from the mathematical community in order to develop sound mathematical and statistical foundations for these methods.[2] There is a host of methods, such as the 4d-var (four dimensional variational) and EnKF (Ensemble Kalman Filter) which are prevalent in the atmospheric and oceanic sciences,[5, 16, 22, 13, 18, 12] while many new ones, such as Bayesian sampling,[1, 10] back and forth nudging,[3, 4] are under intense development. The main aim of this paper is to study some of these methods, in a model which is simple enough so as to be mathematically tractable but at the same time retains the essential features of relevance to applications in atmospheric sciences.

In particular, we will consider a dynamical model, in our example, the Burgers' equation, which is well-posed as an initial value problem. The main problem of data assimilation is that of estimating in an "optimal" manner the initial condition, given a set of noisy data which are obtained by observing the time evolution of a "true" initial condition of the physical dynamical systems which this model represents. In the context of geophysics, this is the problem of initialization of the numerical model.

We have chosen Burgers' equation as a dynamical model for several reasons. Firstly, it is a very well studied forward model – the existence and uniqueness of solutions is well known.[23] In fact, uniqueness of the solution of a specific variational formulation of the initialization problem is also known,[27] see sec. 2. Furthermore, data assimilation and related problems using Burgers' equation have been studied by several authors.[19, 9, 24, 7, 8] Unlike these previous studies, one of the aims of this study will be to understand the effect of "data density," i.e. increasing number of observations, on the optimal state estimates that we discuss. Lastly, our main aim is to use this model to study data assimilation methods based on Bayesian approaches and compare them with existing methods such as 4d-var and EnKF. This is still work in progress and we will present these results elsewhere in future. The major limitation of this study in particular, and Burgers' equation in general, is that the data we use will be "synthetic" or simulated data and not from any physical system. Thus we will not be dealing with the issues of errors in modelling.

This paper presents initial results from a larger study, of which overall scope is as follows. We would like to study the Bayesian formulation of the variational approach stated below (Sec. 2), where the cost function is seen as the logarithm of a posterior distribution function on appropriate function spaces. Such approach is being developed in other data assimilation problems as well.[10, 11] Further, we will formulate a version of the Kalman filter on these spaces, and then compare the two methods.

The paper is organized as follows. We will first discuss the data assimilation problem in the framework of optimal control theory, and present some of the available theoretical results. We will then describe the

analysis of the numerical methods we use for discretizing the continuous problem. The novel feature here is a discussion about the convergence of viscous Burgers' equation over a bounded domain. We will also present numerical results corroborating this analysis. We will then present the gradient descent methods and the numerical results for minimization of the cost functional whose minima represent the optimal estimate of the state based on the data.

We will end the paper with a discussion of our ongoing work and its relation to other contexts.

## 2. 4D-VAR METHOD FOR BURGERS' EQUATION

We will first formulate the data assimilation problem as an optimal control problem by constructing a cost function which measures "distance" between the given observations and our estimate of it, and whose minimum will give the optimal initial condition. This is precisely the commonly used 4d-var method.[18] Such a minimization problem is ill-posed.[24] We will use Tikhonov regularization by adding a "background" guess.[27, 17] Another regularization of this problem, given by the Landweber iteration method, is discussed in [24].

### 2.1. Cost function.

Let us consider a model given by viscous Burgers' equation

$$\frac{\partial z}{\partial t} + \frac{1}{2}\frac{\partial (z^2)}{\partial x} = \mu \frac{\partial^2 z}{\partial x^2} \tag{2.1}$$

for $(x,t) \in \Omega = (0,1) \times (0,T)$ and for a postive parameter $\mu$, with Dirichlet boundary conditions

$$z(0,t) = 0 = z(1,t)$$

and initial condition

$$z(x,0) = u(x).$$

For $u \in L^2(\Omega)$, it is known that there exists a unique solution $z \in L^2(0,T; H_0^1(\Omega)) \cap C([0,T]; L^2(\Omega))$ ( See, for example, [14], Chapter 2, section 2). For initial conditions with better regularity, the solution is also in a better space (see, for example, [27, 24]).

In parallel with the discussion in [27], we will assume that the observations $Z^d(t)$ at a fixed time $t$ are in a Hilbert space $\mathcal{Z}$ and that the state of the system $z(t;u)$ at time $t$ is related to the data by the *observation operator*, which is a linear mapping $C$, i.e. $Cz(t;u) \in mathcalZ$ as well. We will consider two distinct cases.

- *Continuous observations*: The observations will be available continuously in the domain $\Omega(0,T)$. The cost function can be written as

$$J(u) = \frac{1}{2}\int_0^T \left\| C(z(t;u)) - Z^d \right\|_{\mathcal{Z}}^2 \, dx \, dt + \frac{\alpha}{2}\int_0^1 |u - u^b|^2 \, dx \tag{2.2}$$

  where $\|\cdot\|_{\mathcal{Z}}$ is a norm on $\mathcal{Z}$. We will only consider the case $C = \mathrm{id}$ and the $L^2$ norm. We note that the above cost function is not the most natural one to consider when the observations are noisy, but continuous in time.
- *Discrete observations*: The observations are taken at a finite number $M$ of points in "space" $x$ and finitely many times. In particular,

$$C(z(t_i)) = \{z(x_1,t_i), \ldots, z(x_M,t_i)\} \qquad \text{for} \qquad 0 \le t_1 < t_2 \cdots < t_N = T$$

  The cost function in this case is

$$J(u) = \frac{1}{2}\sum_{i=1}^{N} \left| C(z(t_i,u)) - Z^d(t_i) \right|^2 + \frac{\alpha}{2}\int_0^1 |u - u^b|^2 \, dx\,, \tag{2.3}$$

  where the norm in the sum is simply the $L^2$ norm on $\mathcal{Z} \equiv \mathbb{R}^M$ which is finite-dimentional.

In both the above cases, the "optimal" initial condition will be given by the minimum of the cost function. Thus, we will look for

$$\overline{u} = \operatorname*{argmin}_{u \in \mathcal{U}} J(u)$$

for a reasonable class of controls, say $\mathcal{U} = H_0^1(0,1)$. Note that we have used Tikhonov regularization with $u^b(x)$ being an *a priori* approximation of the unknown initial condition $u$. We will later discuss the dependence of the minimum on both $\alpha$ as well as on $u^b(x)$.

2.2. **Adjoint equations.** In the continuous observation case, one can show that under reasonable assumptions, there exists at least one solution $\overline{u}$ to the minimization problem. Further, the first order optimality conditions verified by $\overline{u}$ can be derived using the fact $J'(\overline{u}) = 0$ and the existence of the co-state vector $P \in L^2(0,T;H_0^1(\Omega))$, satisfying

$$-\frac{\partial P}{\partial t} - \overline{z}\frac{\partial P}{\partial x} - \mu\frac{\partial^2 P}{\partial x^2} = C^*\left[C(\overline{z}(\overline{u})) - Z^d\right] \qquad \text{with} \qquad P(x,T) = 0. \tag{2.4}$$

Here $\overline{z}(\overline{u})$ is the solution of Eq. (2.1) with initial condition $\overline{u}$ and $C^*$ is the adjoint of $C$. In [27] sufficient conditions are derived, namely smallness of $T$, so that $J$ admits a unique minimum. The condition on smallness of $T$ depends on the observations as well as the viscosity $\mu$, and is difficult to verify in practice. Our main interest will be in studying numerical methods for finding a minimum and understanding whether this minimum is unique. We will also use the co-state or adjoint equation in the minimization algorithm in order to calculate the gradient of the cost function.

In the case of discrete observations, we are not aware of any results for existence of a minimum. But, even in that case, the above co-state equation with the right hand set to zero, and with jump conditions

$$P(x,t_i-) = P(x,t_i+) - \nabla_z \left|C(z(t_i,u)) - Z^d(t_i)\right|^2 \qquad \text{for} \qquad i = N, N-1, \ldots, 1$$

at observations times, can be used to calculate the gradient of the cost function. Adjoint methods have been discussed previously for various different optimal control problems, e.g., [25, 20].

## 3. Numerical Methods

Here we will consider two finite difference schemes for Burgers' equation and indicate their convergence and then derive the adjoint schemes and describe the 4d-var algorithm using steepest descent method. Throughout this section, we will use super- and sub-scripts for time and space discretization, respectively. In particular, for any function $U(x,t)$, let us denote

$$\begin{aligned} &U_j^m = U(x_j,t_m) \quad \text{for} \quad j = 0,\ldots,(n+1), \quad m = 0,\ldots,N, \quad \text{where} \\ &x_j = j\Delta x, \quad (n+1)\Delta x = 1, \quad t_m = m\Delta t, \quad N\Delta t = T. \end{aligned}$$

3.1. **Schemes for Burgers' equation.** We will consider two schemes here – the implicit Lax-Friedrichs scheme and the "centered difference" scheme. We will show that for the implicit Lax-Friedrichs scheme, the time step $\Delta t$ can be chosen to be much larger than that for the centered difference scheme and for the rest of the numerical work we will focus only on the implicit Lax-Friedrichs scheme.

3.1.1. *Implicit Lax-Friedrichs scheme.* Let us first consider the following descretization of the Burgers' equation.

$$\begin{aligned} \mathcal{L}^{\mathrm{LF}}U(x_j,t_m) &= \frac{U_j^{m+1} - \frac{U_{j+1}^m + U_{j-1}^m}{2}}{\Delta t} + \frac{1}{4\Delta x}((U_{j+1}^m)^2 - (U_{j-1}^m)^2) \\ &- \frac{\mu}{(\Delta x)^2}\left(U_{j+1}^{m+1} - 2U_j^{m+1} + U_{j-1}^{m+1}\right) = 0, \quad 1 \le j \le N-1. \end{aligned} \tag{3.1}$$

We first calculate the local truncation error $T_j^m = T(x_j,t_m)$, which is obtained by applying the scheme to the exact solution $z(x_j,t_m)$ :

$$T(x,t) = \mathcal{L}^{\mathrm{LF}}(z(x,t))$$

Assuming the solution to be smooth and using Taylor's theorem, we have for $k = \Delta t$ and $h = \Delta x$ ,

$$\frac{1}{k}\left\{z(x,t+k) - \frac{1}{2}\left[z(x+h,t) + z(x-h,t)\right]\right\} = z_t + \frac{1}{2}z_{tt}k - \frac{1}{2}z_{xx}\frac{h^2}{k} + o\left(k + \frac{h^2}{k}\right),$$

$$\begin{aligned} \frac{\mu}{h^2}\left[z(x+h,t+k) - 2z(x,t+k) + z(x-h,t+k)\right] &= \mu z_{xx} + \frac{\mu}{12}z_{xxxx}h^2 \\ &\quad + \mu z_{xxt}k + o\left(h^2 + k\right), \end{aligned}$$

$$\frac{1}{4h}\left[z^2(x+h,t) - z^2(x-h,t)\right] = \frac{1}{2}\left(z^2\right)_x + \left(zz_{xxx} + 3z_xz_{xx}\right)\frac{h^2}{6} + o\left(h^3\right),$$

where $z_x = z_x(x,t)$ etc. on the right hand side. Using the above expressions, we get

$$\mathcal{L}^{\mathrm{LF}} z(x_j,t_m) = (z_t + zz_x - \mu z_{xx}) + \left(\frac{1}{2}z_{tt} - \mu z_{xxt}\right)k - z_{xx}\frac{h^2}{2k} + \left(3z_x z_{xx} + zz_{xxx} - \frac{\mu}{2}z_{xxxx}\right)\frac{h^2}{6} + o\left(k + h^2 + \frac{h^2}{k}\right)$$

Choosing $h, k$ small such that $\frac{h}{k}$ is a positive constant,

$$|T(x,t)| \le \left(\frac{1}{2}|z_{tt}| + \mu|z_{xxt}|\right)k + |z_{xx}|\frac{h^2}{2k} + \left(3|z_x z_{xx}| + |zz_{xxx}| + \frac{\mu}{2}|z_{xxxx}|\right)\frac{h^2}{6} + o\left(k + h^2 + \frac{h^2}{k}\right) \tag{3.2}$$

and using the fact that the derivatives of $z$ are bounded in our domain, and for $h/k$ a constant,

$$|T(x,t)| \le Ck.$$

This shows that the local truncation error goes to zero as $k$ goes to zero with $\frac{h}{k}$ constant. Thus the scheme is consistent. [In fact, this is true for $h^p/k$ constant for any $0 < p < 2$ and the order of the scheme in this case is $\min(1, 2/p - 1)$. Thus, $p \le 1$ gives the scheme of highest order which is one.]

In order to prove convergence, we will extend $U_j^m$ as a piecewise constant function $U_k(x,t)$ for a time step $k$ for all $(x,t)$ in our domain and define the error to be

$$e_k(x,t) = U_k(x,t) - z(x,t)$$

where $z$ is the solution of Eq. (2.1). The scheme is said to be convergent if this error converges to zero in some norm as $k$ tends to zero. (See [21] for notations and definitions.) Henceforth, we will drop the subscript $k$ for the time-step $k$.

Let us multiply Eq. (3.1) by $k$ and write the scheme as

$$k\mathcal{L}^{\mathrm{LF}} U_j^m = \left(AU^{m+1}\right)_j - \left[H\left(U^m\right)\right]_j \tag{3.3}$$

where $H$ is a nonlinear operator defined by

$$\left[H\left(U^m\right)\right]_j := \frac{1}{2}\left(U_{j+1}^m + U_{j-1}^m\right) - \frac{k}{4h}\left[\left(U_{j+1}^m\right)^2 - \left(U_{j-1}^m\right)^2\right] \tag{3.4}$$

and the matrix $A$ is symmetric, tridiagonal with $(1 + 2\mu k/h^2)$ as diagonal and $(-\mu k/h^2)$ as sub- and super-diagonal entries. Noting that

$$\mathcal{L}^{\mathrm{LF}} U_j^m = 0, \qquad \text{and} \qquad \mathcal{L}^{\mathrm{LF}} z(x_j,t_k) = (T_k)_j^m,$$

we get

$$\left(Ae^{m+1}\right)_j = \left[H\left(U^m\right)\right]_j - \left[H\left(z^m\right)\right]_j - kT_j^m.$$

We need to estimate the first term on the RHS in order to get a recurrence relation for the error. Before doing that, we estimate the norm of the inverse of the matrix $A$. Let us denote for any vector $x = (x_1, x_2, \cdots, x_n) \in \mathbb{R}^n$,

$$\|x\|_\infty = \max_{1 \le j \le n} |x_j|$$

and for the bounded function $z$ defined on the domain $I \times [0,T]$,

$$\|z\|_\infty = \sup_{I \times [0,T]} |z(x,t)|$$

**Lemma 3.1.** $\left\|A^{-1}\right\|_\infty \le 1$

*Proof.* By definition,

$$\|A^{-1}\|_\infty = \max_{\|x\|_\infty = 1} \|A^{-1}x\|_\infty$$

There exists $x^* \in \mathbb{R}^n$ such that $\|x^*\|_\infty = 1$ and $\|A^{-1}\|_\infty = \|A^{-1}x^*\|_\infty$.

Let assume that $\|A^{-1}\|_\infty > 1$. Then $\|A^{-1}x^*\|_\infty > 1$. Let $y = A^{-1}x^*$, i.e. $x^* = Ay$. Then we have $\|y\|_\infty > 1$ and $\|Ay\|_\infty = 1$.

Let $i_0$ such that $y_{i_0} = \|y\|_\infty > 1$. (Replace $y$ by $-y$ if the maximum of the absolute values is reached with a negative value). Then the corresponding row of $Ay$ is

$$(Ay)_{i_0} = y_{i_0} + K(2y_{i_0} - y_{i_0-1} - y_{i_0+1})$$

where $K = \frac{\mu k}{h^2} > 0$ in the decomposition $A = I + KB$.

If $i_0 = 1$ or $n$, then there is only one "-1" in the corresponding row of $B$. As $y_{i_0} \geq y_i$ for all $i$, then $(Ay)_{i_0} \geq y_{i_0} > 1$. Thus $\|Ay\|_\infty > 1$. There is a contradiction, and thus $\left\|A^{-1}\right\|_\infty \leq 1$. □

**Remark 3.2.** *Note that this norm becomes very close to* 1 *if* $K$ *is close to* 0 *(but by definition of* $K$*, it should be a quite large real number), or if all the components of* $y$ *are close (*$y_{i_0} = y_{i_0-1} = y_{i_0+1} \Rightarrow (Ay)_{i_0} = y_{i_0}$*). It is possible to consider* $y = (1\ 1\ 1\ \ldots\ 1)^T$ *(as* $\|y\|_\infty = 1$*), and then,* $Ay = y$ *and then* $A^{-1}y = y$ *and* $\|A^{-1}\|_\infty \geq 1$ *(and this is the maximum).*

**Lemma 3.3.** *Let*

$$C_m = \max\{\, \|U^m\|_\infty, \|z\|_\infty\}$$

*for* $m$*,* $0 \leq m \leq N$*. If the CFL condition*

$$\frac{kC_0}{h} \;\leq\; 1 \tag{3.5}$$

*holds for* $k, h$ *with* $kN \leq T$ *and* $\frac{k}{h}$ *a positive constant, then we have*

$$C_m \;\leq\; C_0 \quad \forall\, 1 \leq m \leq N.$$

*Proof.* We can prove this by induction. From (3.3), we have for $m = 1$,

$$\begin{aligned}
\left(AU^1\right)_j &= \left[H\left(U^0\right)\right]_j = \frac{1}{2}\ \left(U^0_{j+1} + U^0_{j-1}\right) - \frac{k}{4h}\left[\left(U^0_{j+1}\right)^2 - \left(U^0_{j-1}\right)^2\right] \\
&= (\frac{1}{2} - \frac{k}{2h}U^0_{j+1})U^0_{j+1} + (\frac{1}{2} + \frac{k}{2h}U^0_{j-1})U^0_{j-1} \\
&= \frac{1}{2}(U^0_{j+1} + U^0_{j-1}) \leq \|U^0\|_\infty
\end{aligned}$$

by using the CFL condition. By the previous lemma and the definition of $C_0$, it follows that

$$\|U^1\|_\infty = \|A^{-1}AU^1\|_\infty \leq \|U^0\|_\infty \leq C_0$$

Hence the condition $\frac{kC_1}{h} \;\leq\; 1$ also holds. Now using the assumption $C_m \leq C_0$ and the condition $\frac{kC_m}{h} \;\leq\; 1$, proceeding as above, we can show that

$$C_{m+1} \leq C_0.$$

Thus the lemma follows. □

**Lemma 3.4.** *If the CFL condition Eq.* (3.5) *holds for* $k, h$ *with* $kN \leq T$ *and* $\frac{k}{h}$ *a positive constant, then we have for every* $m$*,* $1 \leq m \leq N$*,*

$$|(Ae^{m+1})_j| \;\leq\; \|e^m\|_\infty + k|(T_k)^m_j|.$$

*Proof.* From the previous lemma, it follows that the CFL condition

$$\frac{kC_m}{h} \;\leq\; 1$$

holds for all $m$, $1 \leq m \leq N$. Consider

$$\begin{aligned}
(H(U^m))_j - (H(z(x_j, t_m))_j &= \frac{1}{2}(e^m_{j+1} + e^m_{j-1}) - \frac{k}{4h}\{(U^m_{j+1})^2 - (z^m_{j+1})^2 - ((U^m_{j-1})^2) - (z^m_{j-1})^2)\} \\
&= (\frac{1}{2} - \frac{k}{2h}\theta^m_{j+1})e^m_{j+1} + (\frac{1}{2} + \frac{k}{2h}\theta^m_{j-1})e^m_{j-1}
\end{aligned}$$

using mean value theorem, for some $\theta^m_{j+1}$ between $U^m_{j+1}$ and $z(x_j, t_m)$. By the CFL condition, both coefficients are positive and hence

$$\begin{aligned} |(H(U^m))_j - (H(z(x_j,t_m)_j| &\leq (\frac{1}{2} - \frac{k}{2h}\theta^m_{j+1})|e^m_{j+1}| + (\frac{1}{2} + \frac{k}{2h}\theta^m_{j-1})|e^m_{j-1}| \\ &\leq \frac{1}{2}(|e^m_{j+1}| + |e^m_{j-1}|) \\ &\leq \|e^m\|_\infty \end{aligned}$$

Thus we get

$$|(Ae^{m+1})_j| \leq \|e^m\|_\infty + k|(T_k)^m_j|.$$

□

Using all these, now we can conclude the convergence of the scheme.

**Theorem 3.5.** *If the CFL condition Eq.* (3.5) *holds for $k, h$ with $kN \leq T$ and $\frac{k}{h}$ a positive constant, the implicit Lax-Friedrichs scheme is convergent.*

*Proof.* From the previous lemmas, we have

$$\begin{aligned} \|e^{m+1}\|_\infty &= \max_j |(A^{-1}Ae^{m+1})_j| \leq \|A^{-1}\|\|Ae^{m+1}\|_\infty \\ &\leq \|e^m\|_\infty + k|(T_k)^m_j|. \end{aligned}$$

Defining $E^{m+1} = \|e^{m+1}\|_\infty$, we have the recurrence relation

$$E^{m+1} \leq E^m + k \max_j |(T_k)^m_j|$$

Solving this iteratively, we get

$$E^{m+1} \leq E^0 + k\sum_{i=0}^{m} \max_j |(T_k)^i_j|$$

Recall that for a smooth solution $z$, the local truncation error, $|T(x,t)| \leq Ck$ for some constant depending on the bounds of the derivatives of $z$. Using $mk \leq T$ we get

$$E^{m+1} \leq E^0 + kCT.$$

This shows that the error goes to zero as $k$ goes to zero with $\frac{k}{h}$ fixed as a positive constant. □

**Remark 3.6.** *It is interesting to note that the viscous term treated implicitly makes the Lax-Friedrichs scheme stable, while the explicit treatment of the viscous term has been shown not to converge.*[8, 26] *Indeed, Lax-Friedrichs scheme (see equation 3.1) can be rewritten in the following way:*

$$\begin{aligned} U^{m+1}_j &= \frac{U^m_{j+1} + U^m_{j-1}}{2} - \frac{\Delta t}{4\Delta x}((U^m_{j+1})^2 - (U^m_{j-1})^2) + \frac{\mu\,\Delta t}{(\Delta x)^2}\left(U^m_{j+1} - 2U^m_j + U^m_{j-1}\right) \\ &= U^m_j + \frac{U^m_{j+1} - 2U^m_j + U^m_{j-1}}{2} - \frac{\Delta t}{4\Delta x}((U^m_{j+1})^2 - (U^m_{j-1})^2) + \frac{\mu\,\Delta t}{(\Delta x)^2}\left(U^m_{j+1} - 2U^m_j + U^m_{j-1}\right). \end{aligned} \tag{3.6}$$

*The viscosity of the numerical scheme is then $(1/2 + \mu\Delta t/\Delta x^2)$. The CFL condition*[8, 26] *for stability of conservative schemes requires that the numerical viscosity be no greater than $1/2$, which is not satisfied as long as $\mu > 0$.*

3.1.2. *Centered difference scheme.* The next scheme for Burgers' equation is the implicit centered scheme,

$$\begin{aligned} \mathcal{L}^{\mathrm{C}}U(x_j,t_m) &= \frac{U^{m+1}_j - U^m_j}{\Delta t} + \frac{1}{4\Delta x}\left((U^m_{j+1})^2 - (U^m_{j-1})^2\right) \\ &- \frac{\mu}{(\Delta x)^2}\left(U^{m+1}_{j+1} - 2U^{m+1}_j + U^{m+1}_{j-1}\right) \end{aligned} \tag{3.7}$$

Let us first calculate the local truncation error $T^m_j = T(x_j, t_m)$, in this case. As before assuming the solution to be smooth and using Taylor's theorem, we have for $k = \Delta t$ and $h = \Delta x$

$$\frac{z(x,t+k) - z(x,t)}{k} = z_t(x,t) + \frac{1}{2}z_{tt}(x,t)k + \circ(k),$$

Choosing $h$ and $k$ small as before, we get the local truncation error estimate

$$|T(x_j,t_m)| \leq \left(\frac{1}{2}|z_{tt}| + \mu|z_{xxt}|\right)k + \left(3|z_x z_{xx}| + |z z_{xxx}| + \frac{\mu}{2}|z_{xxxx}|\right)\frac{h^2}{6} + o\left(k+h^2\right) \tag{3.8}$$

Thus the local truncation error goes to zero as $k$ and $h$ go to zero and the scheme is consistent (for any path in the $(h,k)$ plane, unlike the implicit Lax-Friedrichs scheme).

In the earlier case, we could get a recurrence relation for the error and from there an error bound. In this case also, we need to check that the local errors do not amplify too much. Heuristically one can check that the amplification is not too large, at least in the linearized equation and then conclude for the nonlinear equation as the solution is smooth. We follow the approach as outlined in [14] for the linearized Burgers' equation with frozen coefficients

$$u_t + au_x = \mu u_{xx}$$

by taking Fourier transform and check Von Neumann's stability condition.

Extending the functions to whole of the real line by 0, we can define

$$U^m(\xi) = \frac{1}{\sqrt{2\pi}}\int_0^\infty e^{-\iota x\xi}U^m(x,t)dx.$$

Then by taking Fourier transform of the scheme with respect to $x$, we get

$$\left[1 + 4\mu\frac{k}{h^2}\sin^2\left(\frac{h\xi}{2}\right)\right]U^{m+1}(\xi) = \left[1 - \frac{\iota ka}{h}\sin\left(\frac{h\xi}{2}\right)\cos\left(\frac{h\xi}{2}\right)\right]U^m(\xi)$$

which gives the amplification factor

$$H(\xi) = \frac{1 - \frac{ka}{h}\left(\iota\ s\sqrt{1-s^2}\right)}{1 + \frac{4\mu k}{h^2}s^2}$$

where $s = \sin(h\xi/2)$. We have, $|H(\xi)| \leq 1$ if, for example,

$$\left(\frac{ka}{h}\right)^2 \leq 4\mu\frac{k}{h^2}$$

This gives a stability condition $k \leq \frac{4\mu}{a^2}$. If this holds, the scheme for the linearized equation converges as the global error goes to zero when we refine $k$. In the nonlinear case, instability may appear if the solution $z$ becomes unbouded ("$a$" plays the role of the norm of $u$).

3.2. **Adjoint Schemes.** We have already presented the "continuous adjoint" in eq. (2.4). In this section we discuss the adjoint of the Burgers' equation discretized using the implicit Lax-Friedrichs scheme. We have chosen to use this scheme in our numerical implementation, rather than discretizing the continuous adjoint equation.

The discretized cost function is

$$J(u) = \frac{1}{2}\sum_{m=0}^{N}\sum_{j=1}^{n-1}|U_j^m - \hat{U}_j^m|^2$$

where $u = (u_j)_{j=1}^{n-1} \in \mathbb{R}^{n-1}$ and $U^m = (U_j^m)_{j=1}^{n-1}$ is the solution of the numerical scheme with $u$ as the initial condition and $\hat{U} = (\hat{U}_j^m)_j$ is the given observation. More generally, let us take

$$J(u) = \sum_{m=0}^{N} g(U^m)$$

with $g$ a scalar function of $U^m$. Let us write the scheme as

$$\begin{aligned} AU^{m+1} &= H(U^m), \qquad 0 \leq m \leq N-1 \\ U^0 &= u \end{aligned}$$

for a nonlinear operator $H$ from $\mathbb{R}^{n-1}$ into $\mathbb{R}^{n-1}$. We define the augmented functional as

$$\tilde{J} = \sum_{m=0}^{N} g(U^m) + \sum_{m=0}^{N-1}\langle p^m, (AU^{m+1} - H(U^m))\rangle$$

Here $\langle \cdot, \cdot \rangle$ denotes the usual scalar product in $\mathbb{R}^{n-1}$. Taking first variations,

$$\begin{aligned}
\delta \tilde{J} &= \sum_{m=0}^{N} \langle g'(U^m), \delta U^m \rangle + \sum_{m=0}^{N-1} \langle p^m, A\delta U^{m+1}) \rangle - \langle\ p^m, H'(U^m)\delta U^m) \rangle \\
&= \sum_{m=0}^{N} \langle g'(U^m), \delta U^m \rangle + \sum_{m=1}^{N} \langle A^T p^{m-1},\ \delta U^m \rangle - \sum_{m=0}^{N-1} \langle H'(U^m)^T p^m, \delta U^m \rangle \\
&= \sum_{m=1}^{N-1} \langle g'(U^m) + A^T p^{m-1} - H'(U^m)^T p^m, \delta U^m \rangle + \langle g'(U^0), \delta U^0 \rangle - \langle p^0, H'(U^0)\delta U^0 \rangle \\
&+ \langle g'(U^N), \delta U^N \rangle + \langle A^T p^{N-1}, \delta U^N \rangle = 0
\end{aligned}$$

since at the optimal $u$, the first variation has to vanish. This gives the adjoint scheme as

$$\begin{aligned}
A^T p^{m-1} - H'(U^m)^T p^m &= -g'(U^m), \quad 1 \le m \le N-1 \\
g'(U^N) + A^T p^{N-1} &= 0.
\end{aligned}$$

The gradient of $J$ is given by

$$g'(U^0) = H'(U^0)^T p_0.$$

Thus for the implicit Lax-Friedrichs scheme, the adjoint scheme is

$$\frac{p_j^{m-1} - \frac{p_{j+1}^m + p_{j-1}^m}{2}}{\Delta t} + \frac{1}{2\Delta x} U_j^m (p_{j-1}^m - p_{j+1}^m) - \frac{\mu}{(\Delta x)^2} \left( p_{j+1}^{m-1} - 2p_j^{m-1} + p_{j-1}^{m-1} \right) = -(U_j^m - \hat{U}_j^m)$$

for $j$, $1 \le j \le n-1$. The gradient of $J$ is given by

$$\nabla J(u) = \left[ \frac{1}{2} \left( B^1 \right)^T - \frac{k}{2h} \left( B^0 \right)^T \right] p^0$$

where $B^1$ is the tridiagonal matrix with zero along the diagonal and 1 in off the diagonal while $B^0$ is the tridiagonal matrix with zero on the diagonal and $U^0 = (U_j^0)$ above the diagonal and $-U^0$ below the diagonal.

## 4. Numerical Results

In this section, we will first compare the numerical results for convergence of the implicit Lax-Friedrichs and the centered difference schemes, with eqs. (3.2) and (3.8), respectively. Next we discuss the various numerical experiments with the gradient descent method for finding a minimum of the cost function, for various choices of regularization and for varying number of observations.

4.1. **Implicit Lax-Friedrichs and centered difference schemes.** In order to compare the truncation error given by eqs. (3.2) and (3.8) with numerical truncation error, we will use the exact solutions of Burgers' equation obtained, e.g., by Cole-Hopf transformation.[6]

Fig. 1 shows the dependence on $\Delta x = h$, on $\mu$, and on $k/h$ of the $L^2$-norm (in space variable $x$) of the local truncation error $T(x,t)$ for fixed $t$. Using the variable $\lambda = k/h$, we write Eq. (3.2) as

$$T(\lambda; h, \mu) = a(\mu)h\lambda + b(\mu)h\frac{1}{\lambda} + c(\mu)h^2 \tag{4.1}$$

where $a, b, c$ are functions of $\mu$ through the derivatives of the solution as they appear in Eq. (3.2). The minimum of the graph of the error $T$ vs. $\lambda$ occurs at

$$\lambda_{\text{opt}} = \sqrt{\frac{b(\mu)}{a(\mu)}}, \tag{4.2}$$

which is independent of $h$ but depends on $\mu$. We also see that the minima depends on the solution under consideration and the time $t$. It is seen in Fig. 1(d) that the "optimal" $\lambda$ decreases with $\mu$, approximately as $1/\mu$. Numerically, we observed that for fixed $\mu$ and $h$, the $\lambda_{\text{opt}}$ for solution with an initial condition with fewer zeros is larger than $\lambda_{\text{opt}}$ for the solution with more zeros. Furthermore, at fixed $\lambda$ and $\mu$, the error increases with $h$. All these conclusions are clearly seen in Fig. 1. Based on this discussion, the choice of $(h, k)$ is made as follows. We first choose the smallest possible $h$, which is mainly limited by the memory

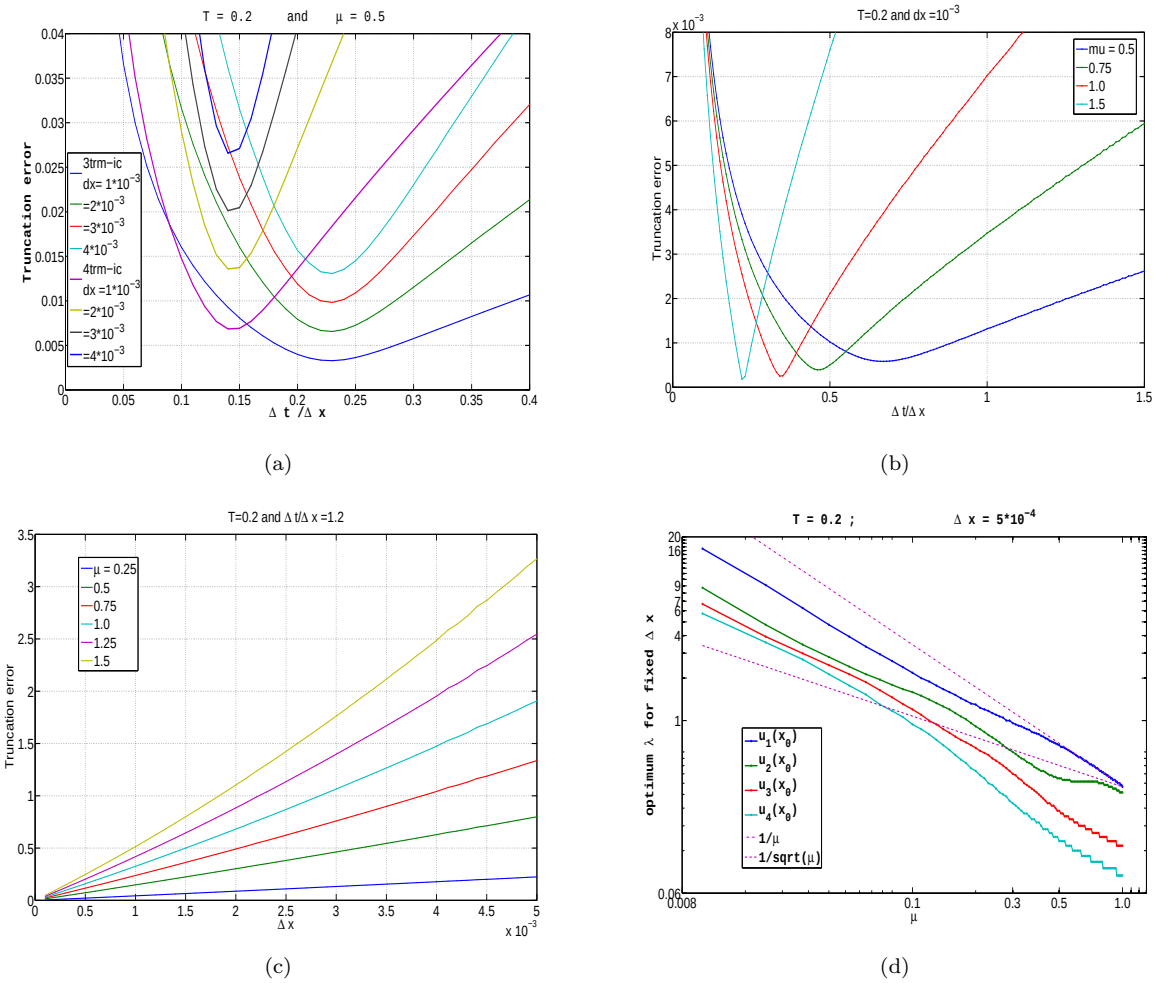
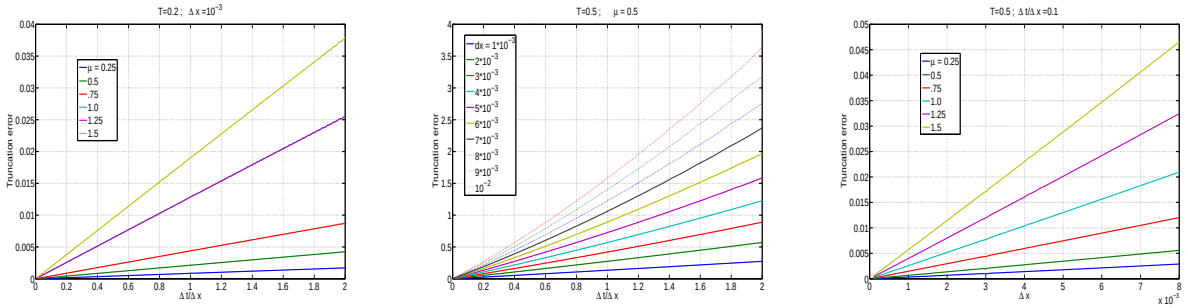



FIGURE 1. The $L^2$ norm, in "space" index $j$, of the truncation error $T_j^m$ Eq. (3.2) for the implicit Lax-Friedrichs scheme. (a) As a function of $\lambda = k/h$ for fixed $\mu = 0.5$. The thick lines are for one initial condition for various $h$, while thin ones for another initial condition for various $h$. (b) As a function of $\lambda$ for fixed $h = 10^{-3}$. Different lines are for different $\mu$. (c) As a function of $h$ for fixed $\lambda = 1.2$. Different lines are for different $\mu$. (d) The optimal $\lambda$ as a function of $\mu$ – it is seen to decrease approximately as $1/\mu$. The numerical results are in complete agreement with Eqs. (4.1)-(4.2).

available for computation. With this $h$, the $k$ is chosen to be greater than the largest $\lambda_{\rm opt}$, but smaller than $C_0/h$ in order to satisfy the CFL condition Eq. (3.5).

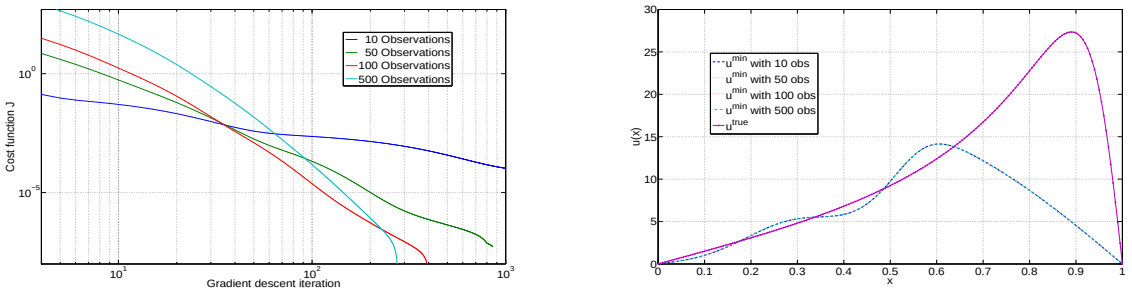
The behaviour of the truncation error for the centered implicit scheme is much simpler. As seen in Fig. 2, the smaller the $(h, k)$, the smaller is the truncation error. Thus, in this case, we must choose the smallest possible values of $(h, k)$, limited only by the memory and computation time.

4.2. **Gradient descent algorithm.** Now we will discuss the minima of the cost function $J$ found using the gradient descent method. We perform the "identical twin experiments" as follows. We choose an initial condition $u^{\rm true}$. We solve the Burgers' equation numerically and generate the data $Z^d$. This is then used to evaluate the cost function $J$. We would like to understand the relation between the minimum $u^{\rm min}$ found using numerical method and this "true" initial condition $u^{\rm true}$.




(a) (b) (c)

FIGURE 2. The $L^2$ norm, in "space" index $j$, of the truncation error $T_j^m$ Eq. (3.8) for the implicit centered scheme. (a) As a function of $\lambda$ for fixed $h = 10^{-3}$. The different lines are for different $\mu$. (b) As a function of $\lambda$ for fixed $\mu = 0.5$. Different lines are for different $h$. (c) As a function of $h$ for fixed $\lambda = 0.1$. Different lines are for different $\mu$. We see that in this case, the error increases with $\lambda$ and $h$, in agreement with Eq. (3.8).




(a) (b)

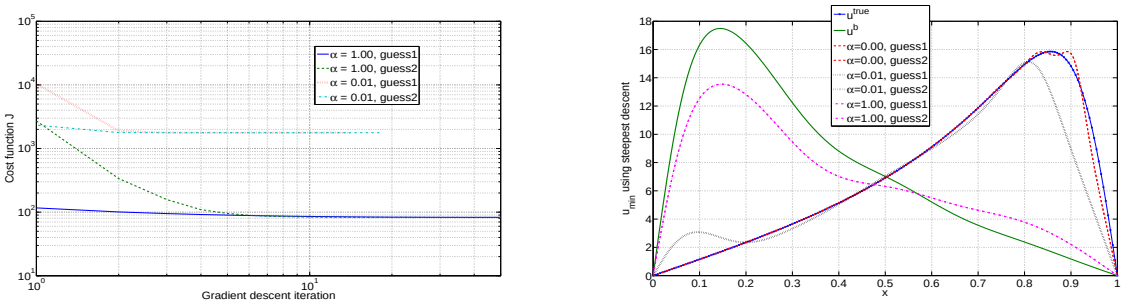
FIGURE 3. The behaviour of the steepest descent for $\alpha = 0$. (a) The cost function $J$ as a function of the gradient descent step, for different number of observations. The more the observations, the faster the gradient descent converges. (b) The $u^{\min}$ and the $u^{\text{true}}$. We see that $u^{\min}$ is almost independent of the number of observation (all four lines are almost identical), but is different from $u^{\text{true}}$.

4.2.1. *Non-regularized cost function with* $\alpha = 0$. In the case of discrete observations, we first look at the behaviour of the gradient descent as the number of observations is increased, when $\alpha = 0$ in the cost function, i.e., without the regularization term. We see from fig. 3(a) that the rate of decrease of $J$ with each descent step strongly depends on the number of observations.

We also see, from fig. 3(b), that irrespective of the number of observations, the minimum is different from the true initial condition. But in the case when $\alpha = 0$, one of the minima of the cost function is certainly $u^{\text{true}}$ since $J \geq 0$ and $J(u^{\text{true}}) = 0$. (Note that this is only true when the observations $Z^d$ are without any noise, which is the case in our numerical experiments. The case of noisy observations is of great interest but will be discussed elsewhere).

This seems to indicate that this cost function for $\alpha = 0$ could have multiple local minima. Indeed, starting with an initial guess which is close to $u^{\text{true}}$, we find that the gradient descent converges to a minimum which is very close, within numerical accuracy, to $u^{\text{true}}$.

4.2.2. *Regularized cost function with* $\alpha \neq 1$. First we discuss the minima of $J$ with $\alpha \neq 0$ and with $u_b = u^{\text{true}}$. In this case, we could numerically find only a single minimum which is close, within numerical accuracy, to




(a) (b)

FIGURE 4. The behaviour of the steepest descent for $\alpha \neq 0$. (a) The cost function $J$ as a function of the gradient descent step, for different values of $\alpha$. The larger the value of $\alpha$, the faster the gradient descent converges. (b) The $u^{\min}$ along with $u^{\text{true}}$ and $u^b$. We see that $u^{\min}$ is like an "interpolation" between $u^{\text{true}}$ and $u^b$, and is independent of the initial guess of the gradient descent.

$u^{\text{true}}$. This is because setting $u_b = u^{\text{true}}$ corresponds to "perfect observations" of the initial condition and thus the regularization term of the cost function dominates. Clearly, the unique minimum of $\|u - u^{\text{true}}\|^2$ is $u^{\text{true}}$. We also see that the gradient descent converges very fast [within a $O(10)$ iterations].

The most interesting case is when $\alpha \neq 0$ and $u_b \neq u^{\text{true}}$. This is of obvious interest, since in practice, the *a priori* guess $u_b$ would certainly not be the "true" state $u^{\text{true}}$. The behaviour of the gradient descent in this case is shown in fig. 4(a). It is clear that the larger the $\alpha$, the faster the convergence. We also see that the presence of the regularization leads to much faster convergence. We have also seen (but not shown in the figure) that the convergence does not depend so strongly on the number of observations as it does in the case $\alpha = 0$. We also note that the minimum in this case is certainly $J_{\min} \neq 0$, since both the "background term" and the "observational term" cannot be zero simultaneously.

Fig. 4(b) shows the minima obtained with varying $\alpha$. We see that irrespective of the initial guess, the gradient descent converges to a single minima. (This is true for various other guesses, but for clarity, only two minima for each value of $\alpha$ are shown in the figure.) Thus it seems that even for observations which are discrete in time and space, the cost function has a unique minima.

The figure also shows $u^{\text{true}}$ and $u^b$ for comparison. We clearly see that as $\alpha$ increases, the $u^{\min}$ comes closer to $u^b$, as expected. The minimum is approximately a linear combination between $u^{\text{true}}$ and $u^b$.

## 5. Conclusion

In summary, we have discussed the data assimilation problem in the optimal control setting, specifically for the Burgers' equation. This leads us to the numerical study of the discretization of Burgers' equation and the gradient descent algorithm for minimization of an appropriate cost function.

We first prove the convergence properties of the implicit Lax-Friedrichs discretization scheme under the CFL condition. We present numerical results that support the estimates for the truncation error and which clearly show that the implicit Lax-Friedrichs scheme allows much larger time steps than the centered difference scheme.

Next, we study the convergence of the gradient descent algorithm and its dependence on the various parameters of the problems, namely, the number of observations, the relative weight in the cost function of the regularization and the data, and the *a priori* approximation $u^b$ of the initial condition. We have presented numerical indications that the cost function without regularization has multiple minima, while the regularized cost function has unique minimum. The rate of convergence depends strongly on the number of observations in the former case, but not the latter case. The minimum obtained is seen to be a combination of the *a priori* background and the "true" state of the system as given by the observations. The interesting case of noisy observations, as well as probabilistic formulation of this model will be reported in the future.

## Acknowledgements

The authors would like to thank Sumanth Bharadwaj for his help with the computations. AA and MR would like to thank Praveen Chandrasekhar and Enrique Zuazua for illuminating discussions. DA and MR would like to thank the Indo-French Center for the Promotion of Advanced Research/ Centre Franco-Indien Pour la Promotion de la Recherche Avancee, for the financial support for this work through the project IFCPAR 3701-1.

## References

1. A. Apte, M. Hairer, A.M. Stuart, and J. Voss, *Sampling the posterior: An approach to non-gaussian data assimilation*, Physica D **230** (2007), 50–64.
2. A. Apte, C. K. R. T. Jones, A. M. Stuart, and J. Voss, *Ensemble data assimilation*, Int. J. Numerical Methods in Fluids **56** (2008), 1033–1046.
3. D. Auroux, *The back and forth nudging algorithm applied to a shallow water model, comparison and hybridization with the 4D-VAR*, Int. J. Numer. Methods Fluids **61** (2009), 911–929.
4. D. Auroux and J. Blum, *A nudging-based data assimilation method for oceanographic problems: the Back and Forth Nudging (BFN) algorithm*, Nonlin. Proc. Geophys. **15** (2008), 305–319.
5. Andrew F. Bennett, *Inverse modeling of the ocean and atmosphere*, Cambridge University Press, 2002.
6. E. R. Benton and G. W. Platzman, *A table of solutions of the one-dimensional Burgers' equation*, Quart. Appl. Math. **30** (1972), 195–212.
7. Carlos Castro, Francisco Palacios, and Enrique Zuazua, *An alternating descent method for the optimal control of inviscid Burgers' equation in the presence of shocks*, Math. Models and Methods in Appl.Sci. **18** (2008), 369–416.
8. ———, *Optimal control and vanishing viscosity for the Burgers' equation*, Integral Methods in Science and Engineering (C. Constanda and M. E. Pérez, eds.), vol. 2, Birkhauser, Boston, 2010, pp. 65–90.
9. Wanglung Chung, John M. Lewis, S. Lakshmivarahan, and S. K. Dhall, *Effect of data distribution in data assimilation using Burgers' equation*, Proceedings of the 1997 ACM symposium on Applied computing, 1997, pp. 521–526.
10. S.L. Cotter, M. Dashti, J.C. Robinson, and A.M. Stuart, *Bayesian inverse problems for functions and applications to fluid mechanics*, Inverse Problems **25** (2009), 115008.
11. S.L. Cotter, M. Dashti, and A.M. Stuart, *Approximation of bayesian inverse problems for pdes*, SIAM J. Numer. Anal **48** (2010), 322–345.
12. G. Evensen, *The ensemble Kalman filter: Theoretical formulation and practical implementation*, Ocean Dynamics **53** (2003), 343–367.
13. Geir Evensen, *Data assimilation: the ensemble Kalman filter*, Springer, 2007.
14. Edwige Godlewski and Pierre Arnaud Raviart, *Hyperbolic systems of conservation laws*, Mathematics and Applications, no. 3/4, Ellipses, 1991.
15. See e.g. `http://france.meteofrance.com/france/observations?62059.path=animationsatellite` or `http://www.mausam.gov.in/WEBIMD/satellitemainpage.jsp` as instances of the vast amount of weather data available.
16. Eugenia Kalnay, *Atmospheric modeling, data assimilation, and predictability*, Cambridge University Press, 2003.
17. A. Kirsch, *An introduction to the mathematical theory of inverse problems*, Springer-Verlag, New York, 1996.
18. F.-X. Le Dimet and O. Talagrand, *Variational algorithms for analysis and assimilation of meteorological observations: theoretical aspects*, Tellus **38A** (1986), 97–110.
19. J.-M. Lellouche, J.-L. Devenon, and I. Dekeyser, *Boundary control of Burgers' equation – a numerical approach*, Comput. Math. Applic. **28** (1994), 33–44.
20. ———, *Data assimilation by optimal control method in a 3D costal oceanic model: the problem of discretization*, J. Atmos. Ocean. Tech. **15** (1998), 470–481.
21. Randall J. LeVeque, *Numerical methods for conservation laws*, second ed., Lectures in Mathematics, Birkhauser Verlag, ETH Zurich, 1992.
22. J. M. Lewis, S. Lakshmivarahan, and S. K. Dhall, *Dynamic data assimilation: a least squares approach*, Encyclopedia of Mathematics and its Applications, no. 104, Cambridge University Press, 2006.
23. J.L. Lions, *Control of systems governed by partial differential equations*, Springer-Verlag, New York, 1972.
24. Johan Lundvall, Vladimir Kozlov, and Per Weinerfelt, *Iterative methods for data assimilation for Burgers's equation*, Journal of Inverse and Ill-Posed Problems **14** (2006), 505–535.
25. Antje Noack and Andrea Walther, *Adjoint concepts for the optimal control of Burgers' equation*, Comput. Optim. Applic. **36** (2007), 109–133.
26. E. Tadmor, *Numerical viscosity and the entropy condition for conservative difference schemes*, Math. Comp **43** (1984), 369–381.
27. Luther W. White, *A study of uniqueness for the initialization problem for Burgers' equation*, J. Math. Anal. Appl. **172** (1993), 412.

TIFR Center for Applicable mathematics, Post Bag No. 6503, Chikkabommasandra, Bangalore 560065 India
*E-mail address*: apte@math.tifrbng.res.in

Laboratoire J.-A. Dieudonne, Universite de Nice Sophia Antipolis, Parc Valrose, F-06108 Nice cedex 2, France
*E-mail address*: auroux@unice.fr

TIFR Center for Applicable mathematics, Post Bag No. 6503, Chikkabommasandra, Bangalore 560065 India
*E-mail address*: mythily@math.tifrbng.res.in