School of Mathematics, Meteorology
and Physics

University of Reading

Department of Mathematics

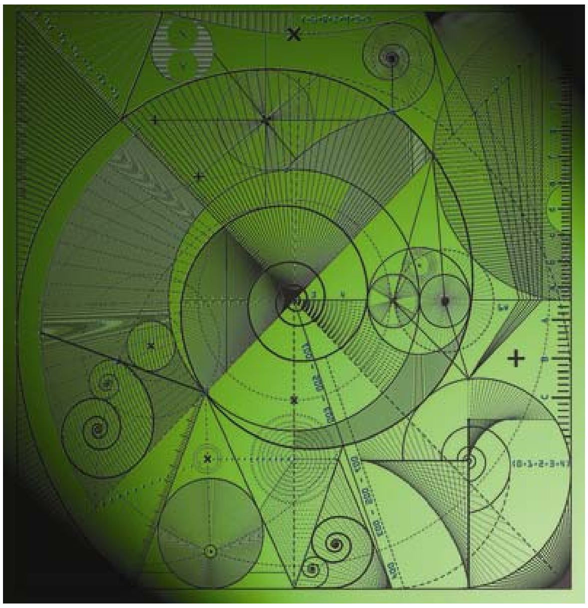
Preprint MPS_2010_14

21 April 2010

Regularization Techniques for Ill-posed Inverse Problems in Data Assimilation

by

M.A. Freitag, N.K. Nichols, C.J. Budd



Regularization Techniques for Ill-posed Inverse Problems in Data Assimilation

M.A. Freitag[a], N.K. Nichols[b,*], C.J. Budd[a]

[a]*Department of Mathematical Sciences, University of Bath, Bath BA2 7AY, UK*
[b]*Department of Mathematics, University of Reading, PO Box 220 Whiteknights, Reading RG6 6AX, UK*

Abstract

Optimal state estimation from given observations of a dynamical system by data assimilation is generally an ill-posed inverse problem. In order to solve the problem, a standard Tikhonov, or L_2, regularization is used, based on certain statistical assumptions on the errors in the data. The regularization term constrains the estimate of the state to remain close to a prior estimate. In the presence of model error, this approach does not capture the initial state of the system accurately, as the initial state estimate is derived by minimizing the average error between the model predictions and the observations over a time window. Here we examine an alternative L_1 regularization technique that has proved valuable in image processing. We show that for examples of flow with sharp fronts and shocks, the L_1 regularization technique performs more accurately than standard L_2 regularization.

Keywords: Ill-posed inverse problems, Tikhonov and L_1 regularization, variational data assimilation, nonlinear least-squares optimization, model error, Burgers' equation.

1. Introduction

The estimation of the states of a fluid dynamical system from given observations by data assimilation is generally an ill-posed inverse problem. Even with excellent numerical models of the system, prediction of the future states of the system is not possible without an accurate estimate of the current state

*Corresponding author: n.k.nichols@reading.ac.uk

of the system with which to initialize a forecast. For the very large systems arising in geosciences and environmental sciences, where there are few and sporadic observations, the problem is particularly hard to tackle. In operational weather and ocean forecasting centres and in the reservoir engineering industry, the problem is commonly formulated as a variational data assimilation problem, resulting in an optimization problem constrained by the dynamical equations describing the system.

The lack of available data, coupled with errors in the observations and in the prior state estimate, as well as in the model, leads to a highly ill-posed inverse problem. To make the problem amenable to solution, a standard form of regularization is used, based on certain statistical assumptions about the errors in the data. Under these assumptions the variational problem provides the maximum posterior Bayesian estimate of the current state of the system [7]. The problem is written as an L_2–norm nonlinear least squares problem with an L_2–norm regularization term. The regularization term constrains the estimate of the initial state to remain close to the prior estimate. The solution is essentially derived by averaging the error between the model prediction and the observations over a time window. As a result, the estimated initial state often contains unsatisfactory oscillations, determined to compensate for errors over the time window.

In this paper we examine alternative methods for regularizing the problem. These methods have proved valuable in image processing applications - especially in image de-blurring and reconstruction algorithms. In these applications the images tend to be static with extensive observations available, but are badly affected by errors. These methods are based on using L_1–norm regularization terms that effectively remove outliers in the data and provide strong edge preservation in the images [1]. To implement these methods in the data assimilation problem, we reformulate the state estimation problem as a least mixed–norm problem, where we regularize the least squares optimization problem using an alternative L_1– norm constraint on the prior estimate.

We compare these different regularization techniques for a test case where there are sharp fronts propagating in a fluid. Numerical methods that preserve different properties of the systems, such as Lax-Friedrich and Lax-Wendroff schemes, are applied. The results show that in the presence of model error the L_1–norm regularization methods capture the discontinuities in the initial conditions more accurately than using the classical L_2–norm regularization technique.

In the next section we describe the variational assimilation problem. In Section 3 we formulate it as a Tikhonov regularized problem and introduce the alternative L_1 regularized problem. In Section 4 we present experimental results comparing the two regularization techniques. The conclusions are summarized in Section 5.

2. Variational Data Assimilation

In optimal state estimation by variational data assimilation (VAR), a weighted nonlinear least-squares measure of the error between the model forecast and the available observations is minimized over a time window together with a penalty term that ensures that the optimal state estimate remains close to a prior estimate (or previous forecast). The problem is formulated as follows.

Given a *prior* estimate $\mathbf{x}_0^b \in \mathbb{R}^m$ (the *background*) of the current state of the system at time t_0 and observations $\mathbf{y}_i \in \mathbb{R}^{p_i}$, at times t_i for $i = 0, \ldots, N$, the aim of variational data assimilation is to find the optimal estimate for the initial state of the system $\mathbf{x}_0 \in \mathbb{R}^m$ (the *analysis*) at time t_0 that minimizes the objective function

$$J(\mathbf{x}_0) = (\mathbf{x}_0 - \mathbf{x}_0^b)^T \mathbf{B}^{-1} (\mathbf{x}_0 - \mathbf{x}_0^b) + \sum_{i=0}^{N} (\mathcal{H}_i(\mathbf{x}_i) - \mathbf{y}_i)^T \mathbf{R}_i^{-1} (\mathcal{H}_i(\mathbf{x}_i) - \mathbf{y}_i), \quad (1)$$

subject to the nonlinear forecast model equations given by

$$\mathbf{x}_i = \mathcal{M}_{i,i-1}(\mathbf{x}_{i-1}), \quad i = 1, \ldots, N. \quad (2)$$

Here $\mathcal{M} : \mathbb{R}^m \to \mathbb{R}^m$ and $\mathcal{H}_i : \mathbb{R}^m \to \mathbb{R}^{p_i}$ denote the evolution and observation operators of the system. We assume that the errors $(\mathbf{x}_0 - \mathbf{x}_0^b)$ in the background and the errors $(\mathbf{y}_i - \mathcal{H}_i(\mathbf{x}_i))$ in the observations are unbiased random Gaussian errors, uncorrelated in time, with covariance matrices $\mathbf{B}$ and $\mathbf{R}_i$, respectively, and that the errors in the background and the observations are uncorrelated. Under these statistical assumptions the solution to the assimilation problem yields the *maximum a posteriori* Bayesian estimate of the state of the system [7].

In practice, for computational efficiency, an incremental version of VAR is implemented in many operational centres. This method solves a sequence of linear approximations to the nonlinear least-squares problem and is equivalent to an approximate Gauss-Newton method for determining the analysis [6]. At each step of the procedure the current estimate $\mathbf{x}_0$ of the analysis is up-dated by an increment $\delta\mathbf{x}_0$ that minimizes the linear least-squares objective function

$$\tilde{J}(\delta\mathbf{x}_0) = (\delta\mathbf{x}_0 - \delta\mathbf{x}_0^b)^T\mathbf{B}^{-1}(\delta\mathbf{x}_0 - \delta\mathbf{x}_0^b) + (\hat{\mathbf{H}}\delta\mathbf{x}_0 - \hat{\mathbf{d}})^T\hat{\mathbf{R}}^{-1}(\hat{\mathbf{H}}\delta\mathbf{x}_0 - \hat{\mathbf{d}}), \quad (3)$$

subject to the linearized model equations

$$\delta\mathbf{x}_i = \mathbf{M}_{i,i-1}\,\delta\mathbf{x}_{i-1}, \quad i = 1, \ldots, N \quad (4)$$

where

$$\begin{aligned} \hat{\mathbf{H}} &= \left[\mathbf{H}_0^T, (\mathbf{H}_1\mathbf{M}_{1,0})^T, \ \ldots, \ (\mathbf{H}_N\mathbf{M}_{N,0})^T\right]^T, \\ \hat{\mathbf{d}}^T &= \left[\mathbf{d}_0^T, \ \mathbf{d}_1^T, \ \ldots, \ \mathbf{d}_N^T\right], \quad \text{with} \quad \mathbf{d}_i = \mathbf{y}_i - \mathcal{H}_i(\mathbf{x}_i), \end{aligned}$$

and $\delta\mathbf{x}_0^b = (\mathbf{x}_0^b - \mathbf{x}_0)$. The matrices $\mathbf{M}_{i,0}$ and $\mathbf{H}_i$ are linearizations of the evolution and observation operators $\mathcal{M}_{i,0}(\mathbf{x}_0)$ and $\mathcal{H}_i(\mathbf{x}_i)$ about the current estimated state trajectory $\mathbf{x}_i$, $i = 0, \ldots, N$, and $\hat{\mathbf{R}}$ is a block diagonal matrix with diagonal blocks equal to $\mathbf{R}_i$.

The linearized assimilation problem is generally ill-posed or highly ill-conditioned. Without the penalty term on the background, the problem is likely to be underdetermined, due to the relatively small number of available observations and the correspondingly low rank of $\hat{\mathbf{H}}$. The penalty term, where the covariance matrix $\mathbf{B}$ is positive definite, then acts as a regularization term and guarantees the existence of a solution to the problem. The problem, however, may still be very ill-conditioned (see [4]) and therefore difficult to solve accurately. In order to improve the conditioning of the linearized assimilation system, a transformation of the incremental states is introduced that decorrelates the errors in the prior states. For a spatially-distributed single-state system, the transformed problem is then easily written as a classical Tikhonov regularization problem and the alternative L_1 regularization problem can be derived directly. Both formulations are described in the next section.

3. Tikhonov and L_1 Regularization

A well-known technique for improving the conditioning of a linear least-squares problem is to apply a linear transformation to 'precondition' the system [2]. To illustrate the technique we here consider a system with a single spatially-distributed state and let $\mathbf{B} = \sigma_b^2\mathbf{C}_B$ and $\hat{\mathbf{R}} = \sigma_o^2\mathbf{C}_{\hat{R}}$, where

$\mathbf{C}_B$ and $\mathbf{C}_{\hat{R}}$ denote the correlation structures of the background and observation errors and σ_b^2 and σ_o^2 are the corresponding error variances. The strategy used in many forecasting centres is to precondition the problem symmetrically using the square root of the background error correlation matrix $\mathbf{C}_B^{1/2}$. The preconditioning is implemented using a transformation to new variables $\mathbf{z} = \mathbf{C}_B^{-1/2}(\delta\mathbf{x}_0 - \delta\mathbf{x}_0^b)$, which are then uncorrelated. With different statistical assumptions on the background errors, different regularizations of the problem then arise.

3.1. Tikhonov regularized problem

In terms of the transformed variable the linearized objective function (3) may be written

$$\hat{J}(\mathbf{z}) = \mu^2 \left\|\mathbf{z}\right\|_2^2 + \left\|\mathbf{G}\mathbf{z} - \hat{\mathbf{f}}\right\|_2^2, \tag{5}$$

where $\mu^2 = \sigma_o^2/\sigma_b^2$, $\mathbf{G} = \mathbf{C}_{\hat{R}}^{-1/2}\hat{\mathbf{H}}\mathbf{C}_B^{1/2}$ and $\hat{\mathbf{f}} = \mathbf{C}_{\hat{R}}^{-1/2}(\hat{\mathbf{d}} - \hat{\mathbf{H}}\delta\mathbf{x}_0^b)$. We see that this is the form of a classical Tikhonov regularization problem, where μ^2 is the regularization parameter. The solution gives the maximum posterior Bayesian estimate for the transformed analysis, assuming the background and observational errors have Gaussian distributions.

If the matrix $\mathbf{G}$ is very ill-conditioned, that is, it has singular values that decay rapidly with many that are zero or near zero, then without the regularization term the assimilation problem (5) is ill-posed. Letting $\mathbf{G} = \mathbf{U}\mathbf{S}\mathbf{V}^T$ be the singular value decomposition of $\mathbf{G}$ [2], we find that the minimizer of (5) is given by

$$\mathbf{z} = \sum_{i=0}^{r} \frac{s_i^2}{\mu^2 + s_i^2} \frac{\mathbf{u}_i^T\hat{\mathbf{f}}}{s_i}\mathbf{v}_i\,, \tag{6}$$

where s_i, $\mathbf{u}_i$, $\mathbf{v}_i$ are the singular values and corresponding left and right singular vectors of the matrix $\mathbf{G}$ and r is its rank. The factor $|\mathbf{u}_i^T\hat{\mathbf{f}}|$ determines the observational information that is projected onto the i^{th} right singular vector in the analysis. For perfect observations this factor typically has a magnitude similar to s_i. If the singular values become very small relative to this factor, however, then observational noise contained in $\hat{\mathbf{f}}$ may corrupt the solution significantly. The effect of the regularization parameter is to ensure that this noise is filtered from the solution. Information is lost, but the accuracy of the result is improved [5].

3.2. L_1 regularized problem

An alternative form of regularization is obtained by using an L_1-norm penalty term in place of the Tikhonov, or L_2-norm, penalty term. The linear objective function in this case takes the form

$$\hat{J}(\mathbf{z}) = \mu^2 \|\mathbf{z}\|_1 + \left\|\mathbf{G}\mathbf{z} - \hat{\mathbf{f}}\right\|_2^2, \tag{7}$$

The solution gives the maximum posterior Bayesian estimate for the transformed analysis provided that the background errors now have an exponential distribution and the observational errors have a Gaussian distribution.

This form of regularization is used popularly for image deblurring and image restoration since it effectively removes outliers in the data and provides strong edge recovery. It might be expected therefore that this form of regularization would more accurately recover sharp fronts and shocks in data assimilation for fluid flows. Our aim here is to investigate this hypothesis and to compare the L_1 and L_2 regularization techniques in an application to computational fluid dynamics. We consider a 1D single-state system with a moving front approximated by two different finite-difference schemes. Further experiments are described in [3].

4. Numerical Experiments

In order to test the two different regularization techniques in a data assimilation scheme, we consider the inviscid Burgers' equation

$$u_t + [f(u)]_x = 0, \tag{8}$$

where $f(u) = \frac{1}{2}u^2$, with initial conditions

$$u(x,0) = \begin{cases} 2\,, & 0 \leq x < 2.5 \\ 0.5\,, & 2.5 \leq x \leq 10\,. \end{cases} \tag{9}$$

The solution is given by

$$u(x,t) = \begin{cases} 2\,, & 0 \leq x < 2.5 + st \\ 0.5\,, & 2.5 + \mathrm{st} \leq x \leq 10\,, \end{cases} \tag{10}$$

where $s = 1.25$ is the shock speed. We approximate this system by the Lax-Friedrich finite-difference scheme

$$U_j^{n+1} = \frac{1}{2}(U_{j-1}^n + U_{j+1}^n) - \frac{\Delta t}{2\Delta x}(f(U_{j+1}^n) - f(U_{j-1}^n)), \tag{11}$$

$$U_j^{n+1} = \frac{1}{2}(U_{j-1}^n + U_{j+1}^n) - \frac{\Delta t}{2\Delta x}(f(U_{j+1}^n) - f(U_{j-1}^n)), \tag{12}$$

or by the Lax-Wendroff scheme in conservative form

$$U_j^{n+1} = U_j^n - \frac{\Delta t}{2\Delta x}(f(U_{j+1}^n) - f(U_{j-1}^n)) \tag{13}$$

$$+ \frac{\Delta t^2}{2\Delta x^2}\left(A_{j+\frac{1}{2}}(f(U_{j+1}^n) - f(U_j^n)) - A_{j-\frac{1}{2}}(f(U_j^n) - f(U_{j-1}^n))\right), \tag{14}$$

where $A_{j\pm\frac{1}{2}}$ is the Jacobian matrix $A(u) = f'(u)$ evaluated at $\frac{1}{2}(U_j^n + U_{j\pm1}^n)$ (see, for example, [8]). Here $U_j^n \approx u(j\Delta x, n\Delta t)$, and $\Delta x = 0.01$, $\Delta t = 0.001$ and $j = 1, \ldots, 100$. The Lax-Friedrich method is known to smear out the shock over time. The Lax-Wendroff method recovers the shock speed, but leads to oscillations near the shock. Both methods therefore introduce model errors into the system.

Observations are taken from the exact solution (10) at every 20 points in space and at every 2 time steps over a window of 100 time steps. The background is taken to be ${U^b}_j^0 = U_j^0 - 0.1$, a shifted initial condition, and the model is initialized with the true initial data. We choose $\mathbf{C}_B = \mathbf{I}$ and $\mathbf{C}_{\hat{R}} = \mathbf{I}$ with $\sigma_o^2 = 0.01$. We choose $\sigma_b^2 = 1.0$ for the Lax-Friedrich model and for the Lax-Wendroff example, which is more sensitive, we choose $\sigma_b^2 = 0.04$. We examine cases with perfect data and with noisy data, where random errors from a Gaussian distribution with variance $\sigma_o^2 = 0.01$ are added to the observations.

The assimilation problem is ill-posed for both numerical models. In Figure 1 the singular values s_i of the matrix $\mathbf{G}$ are plotted (red *) for each model on a log scale. We can see that these decay rapidly, with a large number that are numerically zero (below machine precision). The factors $\mathbf{u}_i^T\hat{\mathbf{f}}$ are also shown (blue line) and it is easy to see that the ratios $\mathbf{u}_i^T\hat{\mathbf{f}}/s_i$ become extremely large and hence without regularization the assimilation problem would become impossible to solve with any accuracy.

The aim of the assimilation is to reconstruct the initial conditions that minimize the errors between the observations from the exact system and the forecast from the imperfect model over the assimilation time window. The best solution over the assimilation window that can be attained is the model solution initialized with the true initial conditions, since the assimilation cannot correct the numerical model here.

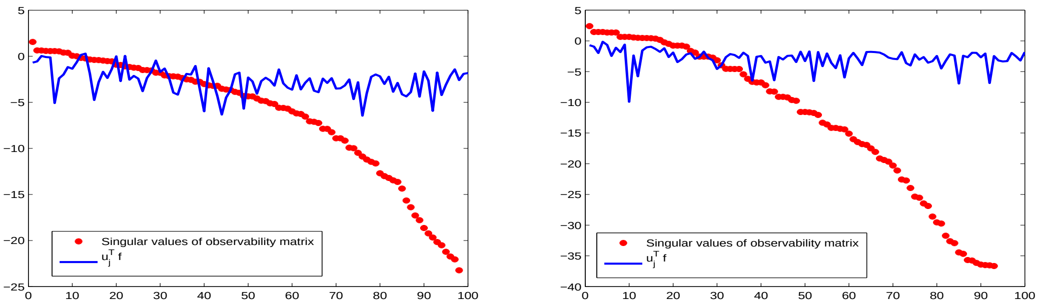


Figure 1: Singular values s_i (red *) of $\mathbf{G}$ for the Lax-Friedrich (left) and Lax-Wendroff (right) models and factors $\mathbf{u}_i^T \hat{\mathbf{f}}$ for the case of noisy data (blue line) plotted on a log scale.

4.1. Lax-Friedrich method

The results obtained by the L_2 and L_1 regularized assimilation techniques with perfect observations using the Lax-Friedrich numerical model are shown in Figure 2. From the model trajectory with the exact initial conditions (shown in blue), we see that the true shock front (shown in red) is smeared out over time, as expected for this numerical scheme. With the L_2 assimilation, the initial analysis is not recovered accurately, due to the model errors that are introduced over the window and also to the errors in the background. The initial analysis contains significant oscillations, leading to an inaccurate position for the shock front in the trajectory at the end of the assimilation window. For the L_1 assimilation, remarkably, the true initial condition is recovered exactly and the trajectory exactly matches the model solution over the time window.

With noisy observations, the results for the classical L_2 variational assimilation, shown in Figure 3, are considerably worse, with significant incorrect oscillations throughout the window. Again, the L_1 regularized assimilation technique captures the initial condition exactly - a surprising achievement! In this case we have chosen a simple background covariance that does not smooth the errors between the background and observations over the window and does not enforce a strong contraint on the analysis to remain close to the prior estimate. If instead, we use a Markov correlation matrix for the prior estimate, then the L_2 assimilation produces results very similar to those with perfect observations, whilst the L_1 regularization method continues to produce the initial analysis exactly [3].

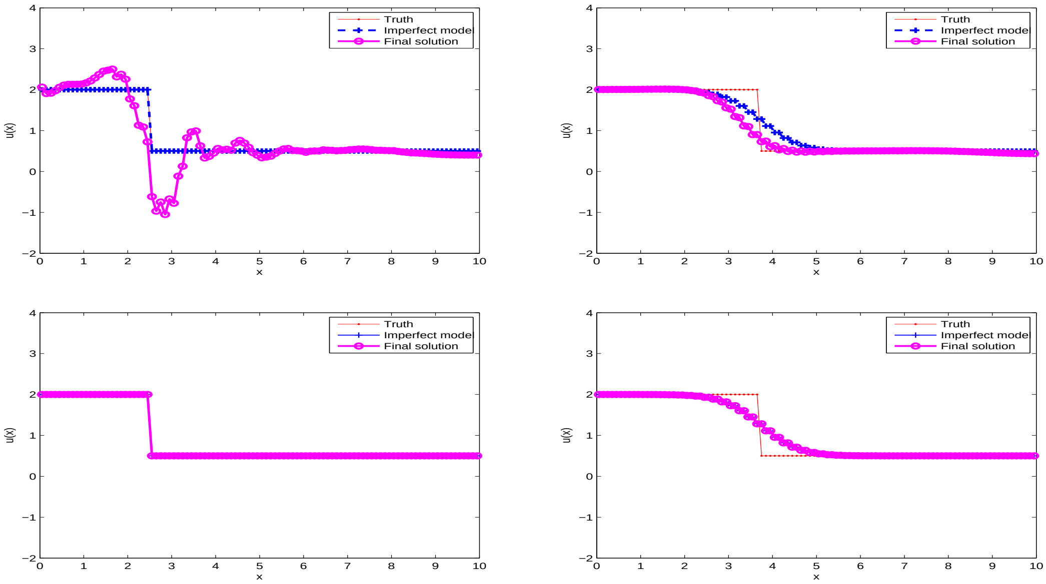


Figure 2: Results for Lax-Friedrich method for perfect (partial) observations. The left/right plots show the solution at the beginning ($t = 0$)/end ($t = 100$) of the assimilation window, respectively. The analysis trajectories from the L_2 and L_1 methods are shown in the upper/lower rows respectively. The true solution is shown in red, the exact model solution is shown in blue and the model solution after assimilation is shown in magenta.

4.2. Lax-Wendroff method

In Figure 4 the results from the L_2 and L_1 regularized assimilation schemes are shown for the Lax-Wendroff numerical model with noisy observational data. The trajectories are shown here at the initial time and at the end of a forecast period of 100 time steps from the end of the assimilation window. The model trajectory with the exact initial conditions (shown in blue) introduces oscillations around the true shock front (shown in red), but maintains the position of the shock accurately, as expected with this numerical scheme. We see again that with the classical L_2 assimilation, the initial analysis is not recovered accurately, due to the model errors that are introduced over the window and also to the errors in the background. The initial analysis contains large oscillations behind the shock. Over the time window, these are damped and the position of the shock front is maintained quite well. However, the solution still contains significant oscillations at the end of the assimilation window and hence the prediction for the shock at the

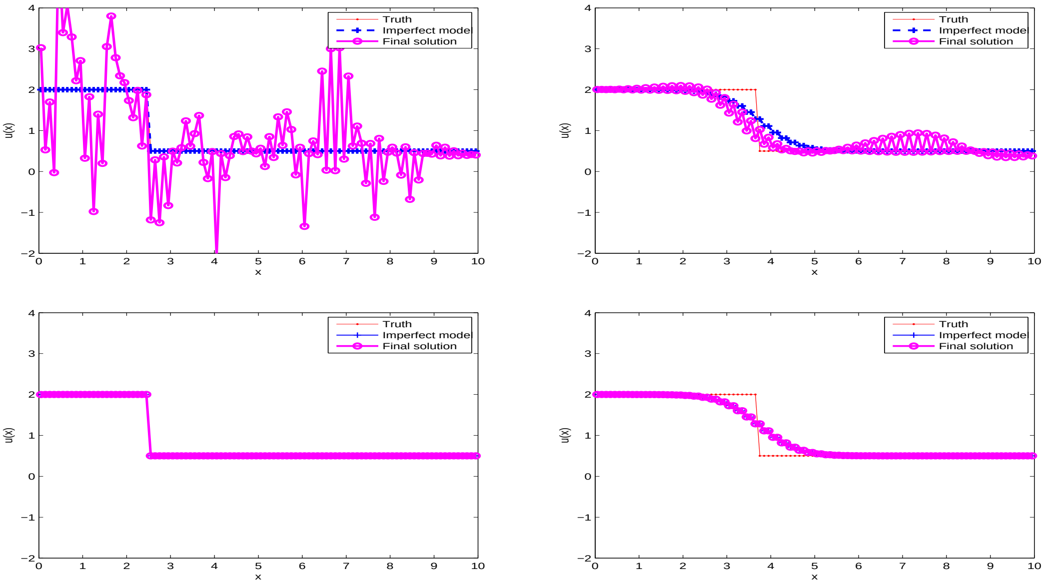


Figure 3: Results for Lax-Friedrich method with imperfect (partial) observations. The left/right plots show the solution at the beginning ($t = 0$)/end ($t = 100$) of the assimilation window, respectively. The analysis trajectories from the L_2 and L_1 methods are shown in the upper/lower rows respectively. The true solution is shown in red, the exact model solution is shown in blue and the model solution after assimilation is shown in magenta.

end of the forecast window is very inaccurate. With the L_1 regularization, in contrast, the true initial condition is recovered exactly, and the solution at the end of the forecast contains small oscillations but captures the shock position accurately, which is the best that is possible with this numerical model.

5. Conclusions

Variational data assimilation is popularly used in the geosciences to provide an optimal estimate of the state of a dynamical system by combining observations of the system with a model forecast over a time window. The problem is highly ill-posed due to lack of data coupled with errors in the observations and in the model. To make the problem amenable to solution, a standard L_2-norm regularization technique is used that constrains the state estimate to remain close to a specified prior estimate. Here we have proposed an alternative regularization strategy based on an L_1-norm penalty

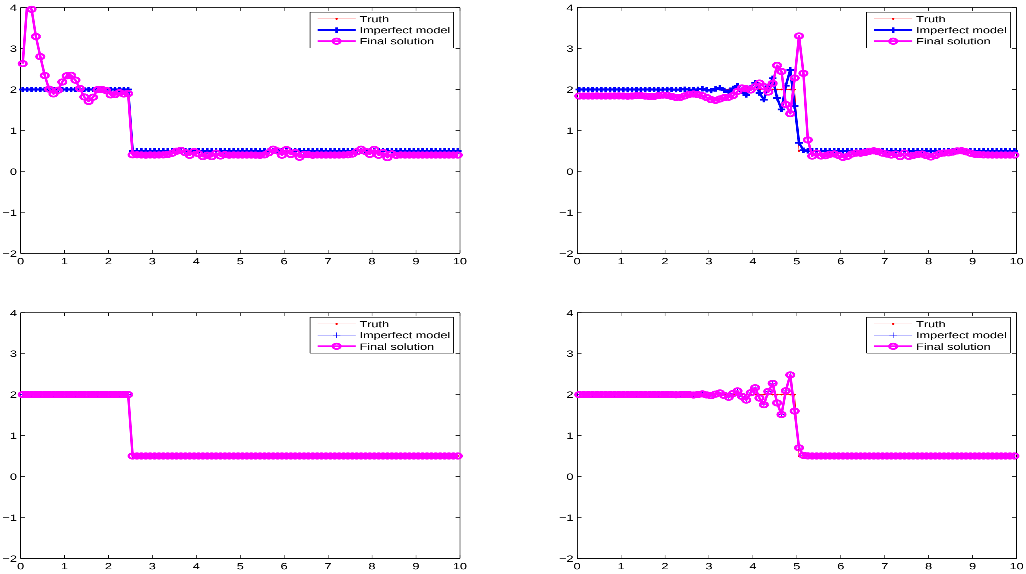


Figure 4: Results for Lax-Wendroff method with imperfect (partial) observations. The left/right plots show the solution at the beginning ($t = 0$)/end ($t = 200$) of the forecast window, respectively. The analysis trajectories from the L_2 and L_1 methods are shown in the upper/lower rows respectively. The true solution is shown in red, the exact model solution is shown in blue and the model solution after assimilation is shown in magenta.

approach, which is expected to remove outliers in the data and provide strong edge recovery. We have examined both the traditional L_2 and the proposed L_1 regularization techniques for a test case with sharp fronts propagating in a fluid, using two numerical models with different properties. The results show that in the presence of model error, the L_1 regularization method capture the discontinuities in the initial states much more accurately than the standard Tikhonov L_2 technique. Other test examples that support these conclusions can be found in [3]. The challenge now is to develop this approach for application to large multi-variable systems.

Acknowledgements

The work of the first and third authors was supported in part by a Great Western Research Grant and by an EPSRC Critical Mass Grant for the Bath Institute for Complex Systems. The work of the second author was supported in part by the NERC National Centre for Earth Observation.

References

[1] Engl, H., M. Hanke and A. Neubauer, 1996: Regularization of inverse problems. Kluwer Academic Pub.

[2] Golub, G.H. and C.F. Van Loan, 1996: *Matrix Computations*,(3rd ed.), Johns Hopkins University Press.

[3] Freitag, M.A., N.K. Nichols and C.J. Budd, 2010: Resolution of sharp fronts in the presence of model error in variational data assimilation, University of Bath, Bath Institute for Complex Systems, Preprint 1/10 (2010).

[4] Haben, S.A., A.S. Lawless and N.K. Nichols, 2009: Conditioning of the 3DVar data assimilation problem, University of Reading, Dept. of Mathematics, Math. Rpt. Series.

[5] Johnson, C., B.J. Hoskins and N.K. Nichols, 2005: A singular vector perspective of 4D-Var:Filtering and interpolation, *Quart. J. Roy. Met Soc.* **131**, 1–19.

[6] Lawless, A.S., S. Gratton and N.K. Nichols, 2005: An investigation of incremental 4D-Var using non-tangent linear models, *Quart. J. Roy. Met. Soc.*, **131**, 459–476.

[7] Lorenc, A., 1988: Optimal nonlinear objective analysis, *Quart. J. Roy. Met. Soc.*, **114**, 205–240.

[8] LeVeque, R., 1992: *Numerical methods for conservation laws.* Birkh*äuser*.