School of Mathematics, Meteorology
and Physics

University of Reading

Department of Mathematics

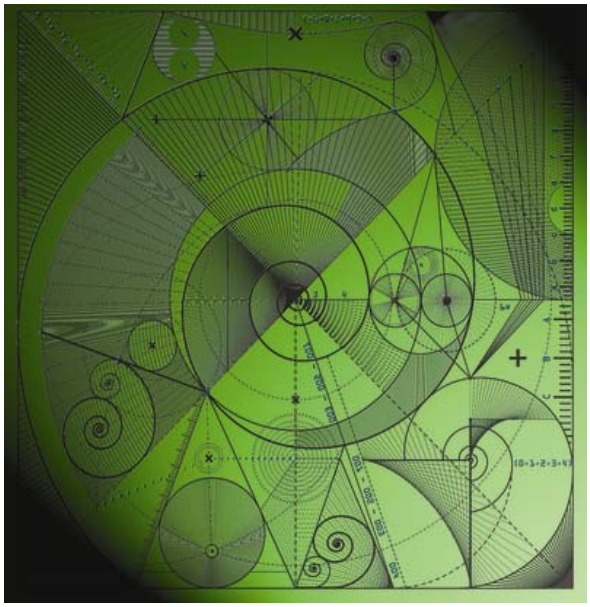
Preprint MPS_2010-04

12 March 2010

Resolution of sharp fronts in the presence of model error in variational data assimilation

by

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



Resolution of sharp fronts in the presence of model error in variational data assimilation

M. A. Freitag* N. K. Nichols† C. J. Budd‡

March 15, 2010

Abstract

We show that data assimilation using four-dimensional variation (4DVar) can be interpreted as a form of Tikhonov regularisation, a very familiar method for solving ill-posed inverse problems. It is known from image restoration problems that L_1-norm penalty regularisation recovers sharp edges in the image more accurately than Tikhonov, or L_2-norm, penalty regularisation. We apply this idea to 4DVar for problems where shocks are present and give some examples where the L_1-norm penalty approach performs much better than the standard L_2-norm regularisation in 4DVar.

1 Introduction

Data assimilation is a method for combining model forecast data with observational data in order to forecast more accurately the state of a system. One of the most popular data assimilation methods used in modern numerical weather prediction is four-dimensional data assimilation (4DVar) (Sasaki (1970); Talagrand (1981); Lewis *et al.* (2006)) which seeks the initial conditions such that the forecast best fits the observations within an interval called the assimilation window. Currently, in most weather centers systems and states of dimension $\mathcal{O}(10^7)$ or higher are considered, whereas there are considerably fewer observations, usually $\mathcal{O}(10^6)$ (see Daley (1991); Nichols (2010) for a review on data assimilation methods).

Linearised 4DVar can be shown to be equivalent to Tikhonov, or L_2-norm regularisation, a well-known method for solving ill-posed problems (Johnson *et al.* (2005)). Such problems appear in a wide range of applications (Engl *et al.* (1996)) such as geosciences and image restoration, the process of estimating an original image from a given blurred image. From the latter work it is known that by replacing the L_2-norm penalty term with an L_1-norm penalty function image restoration problems become edge-preserving as they do not penalise the edges of the image. The L_1-norm penalty regularisation then recovers sharp edges in the image better than the L_2-norm penalty regularisation (Hansen (1998); Hansen *et al.* (2006)). Edges in images lead to outliers in the regularisation term and hence, L_1-norms for the regularisation terms give a better result in image restoration. This is the motivation behind our approach for variational data assimilation.

The edge-preserving property of L_1-norm regularisation can be used for models which develop shocks, which is the case for moving weather fronts. We apply this idea to 4DVar for

*Department of Mathematical Sciences, University of Bath, Claverton Down BA2 7AY, UK
†Department of Mathematics, The University of Reading, Box 220 Whiteknights, RG6 6AX, Reading, UK
‡Department of Mathematical Sciences, University of Bath, Claverton Down BA2 7AY, UK

problems where shocks are present and give several numerical examples where the L_1-norm penalty approach performs better than the standard L_2-norm regularisation in 4DVar. Our examples include a linear advection equation and Burgers' equations where sharp fronts and shocks are present or develop. We will observe that with L_1-norm regularisation fronts as well as front speeds are resolved better than with the standard L_2-norm regularisation of 4DVar.

Section 2 gives an introduction to 4DVar and shows its relation to Tikhonov regularisation. In Section 3 we introduce the new algorithm and in Section 4 we state the model equations. Sections 5, 6, 7 and 8 present numerical examples, where the new L_1-norm regularisation is compared to standard 4DVar. In our examples we introduce several kinds of model error. Under these conditions it can be seen that L_1-norm regularisation outperforms 4DVar when sharp fronts are present (see Sections 5, 6 and 7) and both methods do equally well if no fronts are present (see Section 8). We conclude with a section on future work.

2 4DVar and its relation to Tikhonov regularisation

In nonlinear 4DVar we aim to minimise the cost functional

$$\begin{aligned}\mathcal{J}(x_0) =& \frac{1}{2}(x_0 - x_0^b)^T B^{-1}(x_0 - x_0^b) \\ &+ \frac{1}{2}\sum_{i=1}^{N}(y_i - \mathcal{H}_i(x_i))^T R_i^{-1}(y_i - \mathcal{H}_i(x_i))\end{aligned} \tag{1}$$

subject to the system equations

$$x_{i+1} = \mathcal{M}_{i+1,i}(x_i), \quad i = 0, \ldots, N-1. \tag{2}$$

This is a nonlinear constraint minimisation problem where the first term in (1) is called the background term, x_0^b is the background state at time $t = 0$ and x_i, $i = 1, \ldots, N$ are the state vectors at time t_i, which form the forecast trajectory. In weather forecasting the state vector x_0^b is the previous forecast at time $t = 0$. The vector y_i, $i = 1, \ldots, N$ contains the observations at times t_i. H_i is the observation operator which converts the model state into the observation state. Hence (1) is a nonlinear weighted least-squares problem. By minimising $J(x_0)$ we find an initial state x_0, known as the *analysis*, such that the forecast trajectory is close to the background trajectory and the observations in a suitable norm. The symmetric matrix B and the symmetric matrices R_i, $i = 1, \ldots, N$ are assumed to represent the covariance matrices of the errors in the background and the observations respectively. Provided the background and observation errors have Gaussian distributions with mean zero, then minimising $J(x_0)$ is equivalent to finding the *maximum a posteriori Bayesian estimator* for the true initial condition.

We apply a Gauß-Newton method (Dennis and Schnabel (1983)) in order to solve the minimisation problem (1). From a starting guess x_0^0, Newton's method for solving the gradient equation is

$$\nabla\nabla\mathcal{J}(x_0^k)\Delta x_0^k = -\nabla\mathcal{J}(x_0^k), \quad x_0^{k+1} = x_0^k + \Delta x_0^k, \tag{3}$$

for $k \geq 0$. In the Gauß-Newton method, the Hessian is replaced by an approximate Hessian $\widetilde{\nabla\nabla}\mathcal{J}(x_0^k)$ that neglects all the terms involving second derivatives of $\mathcal{M}_{i+1,i}$ and $\mathcal{H}_i$. We let $M_{i+1,i}$ be the Jacobian of $\mathcal{M}_{i+1,i}$. Here we only consider problems where the observation operator is linear, that is $\mathcal{H}_i(x_i) = H_i(x_i)$. Furthermore, both $R_i = R$ and $H_i = H$, are assumed to be unchanged over time.

The gradient of (1) is then given by

$$\begin{aligned}\nabla\mathcal{J}(x_0) =& B^{-1}(x_0 - x_0^b) \\ &- \sum_{i=1}^{N} M_{i,0}(x_0)^T H^T R^{-1}(y_i - H(x_i)),\end{aligned} \tag{4}$$

where $M_{i,0}(x_0)$ is the Jacobian of $\mathcal{M}_{i,0}(x_0)$. The chain rule gives

$$M_{i,0}(x_0) = M_{i,i-1}(x_{i-1})M_{i-1,i-2}(x_{i-2})\cdots M_{1,0}(x_0).$$

Taking the gradient of (4) and neglecting terms involving the gradient of $M_{i,0}(x_0)$ gives

$$\widetilde{\nabla\nabla}\mathcal{J}(x_0) = B^{-1} + \sum_{i=1}^{N} M_{i,0}(x_0)^T H^T R^{-1} H M_{i,0}(x_0). \tag{5}$$

Both the summation terms in (4) and (5) can be obtained recursively using the adjoint equations

$$\begin{aligned}\lambda_N &= 0, \\ \lambda_{i-1} &= M_{i,i-1}(x_{i-1})^T(\lambda_i + H^T R^{-1}(y_i - H(x_i))),\end{aligned}$$

for $i = N, \ldots, 1$, in order to find the gradient

$$\nabla\mathcal{J}(x_0) = B^{-1}(x_0 - x_0^b) - \lambda_0,$$

and similarly

$$\begin{aligned}\nabla\lambda_N &= 0 \\ \nabla\lambda_{i-1} &= M_{i,i-1}(x_{i-1})^T(\nabla\lambda_i - H^T R^{-1} H M_{i,0}(x_0)),\end{aligned}$$

for $i = N, \ldots, 1$, leads to

$$\widetilde{\nabla\nabla}\mathcal{J}(x_0) = B^{-1} - \nabla\lambda_0.$$

Using these adjoint equations we avoid having to compute $M_{i,i-1}(x_{i-1})$ several times. Note λ_i, $i = 0, \ldots, N$ are vectors whereas $\nabla\lambda_i$, $i = 0, \ldots, N$ are square matrices of the dimension of the system state.

The approximate Hessian $\widetilde{\nabla\nabla}\mathcal{J}(x_0)$ and $\nabla\mathcal{J}(x_0)$ are then used in (3), where the system is solved directly.

Note that this approach is mathematically equivalent to the incremental 4DVar method as described in (Lawless *et al.* (2005)); in the incremental method, however, the inner equations (3) are solved iteratively.

We can rewrite the cost function (1) in 4DVar as

$$\begin{aligned}\mathcal{J}(x_0) = &\frac{1}{2}(x_0 - x_0^b)^T B^{-1}(x_0 - x_0^b) \\ &+\frac{1}{2}(\hat{y} - \hat{\mathcal{H}}(x_0))^T \hat{R}^{-1}(\hat{y} - \hat{\mathcal{H}}(x_0)),\end{aligned}$$

where

$$\hat{\mathcal{H}}(x_0) = \begin{bmatrix} H\mathcal{M}_{1,0}(x_0) \\ H\mathcal{M}_{2,0}(x_0) \\ \vdots \\ H\mathcal{M}_{N,0}(x_0) \end{bmatrix}, \quad \text{and} \quad \hat{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}.$$

In general $\hat{\mathcal{H}}(x_0)$ is a nonlinear operator, $\hat{y}$ is a vector and $\hat{R}$ is a block diagonal matrix with diagonal blocks equal to R. If we linearise $\mathcal{M}_{i,0}$ about x_0^b and denote the Jacobian of $\mathcal{M}_{i,0}$ by $M_{i,0}$ then

$$\hat{H} := \hat{H}(x_0^b) = \begin{bmatrix} HM_{1,0}(x_0^b) \\ HM_{2,0}(x_0^b) \\ \vdots \\ HM_{N,0}(x_0^b) \end{bmatrix}, \tag{6}$$

is a matrix (it is essentially the observability matrix). Now writing $B = \sigma_b^2 C_B$ and $\hat{R} = \sigma_o^2 C_R$ and performing a variable transform $z := C_B^{-1/2}(x_0 - x_0^b)$ we can write the linearised objective function that needs to be minimised as

$$\begin{aligned} \hat{J}(z) =& \|C_R^{-1/2}(\hat{y} - \hat{H}(x_0^b)) - C_R^{-1/2}\hat{H}C_B^{1/2}z\|_2^2 \\ &+ \mu^2\|z\|_2^2, \quad \mu^2 = \frac{\sigma_o^2}{\sigma_b^2}. \end{aligned} \tag{7}$$

This is the same as the linear least-squares problem that needs to be solved for Tikhonov regularisation (Engl *et al.* (1996)), where μ^2 acts as the regularisation parameter here. If we set

$$G := C_R^{-1/2}\hat{H}C_B^{1/2} \quad \text{and} \quad f := C_R^{-1/2}(\hat{y} - \hat{\mathcal{H}}(x_0^b)), \tag{8}$$

then equation (7) may be written as

$$\hat{J}(z) = \|f - Gz\|_2^2 + \mu^2\|z\|_2^2, \quad \mu^2 = \frac{\sigma_o^2}{\sigma_b^2}. \tag{9}$$

If G is an ill-posed operator - or, in the discrete setting, an ill-conditioned matrix , then the minimisation problem

$$\min_z\{\|f - Gz\|_2^2\} \tag{10}$$

is hard to solve exactly, that is, the solution z does not continuously depend on the data. In data assimilation the matrix $G = C_R^{-1/2}\hat{H}C_B^{1/2}$ is generally ill-conditioned, which means it has singular values that decay rapidly and many are very small or even zero. This problem occurs if there are not enough observations in the system, which is typical for numerical weather prediction. Furthermore, the given observation data are subject to errors, leading to errors in the vector f. Hence, we can see that the minimisation problem (10) with an ill-conditioned system matrix G and an unreliable data vector f will lead to an unstable solution and some form of regularisation is required (for example preconditioning, Tikhonov regularisation, singular value filtering, etc.). We consider Tikhonov regularisation where a regularisation term $\mu^2\|z\|_2^2$ is introduced, which leads to the minimisation of $\hat{J}(z)$ in (9). The minimisation of the Tikhonov functional (9) gives the regularised solution

$$z = (G^TG + \mu^2 I)^{-1}G^T f = \sum_{j=1}^{q} \frac{\sigma_j^2}{\sigma_j^2 + \mu^2}\frac{u_j f}{\sigma_j}v_j,$$

where q is the length of the vector f (which is equivalent to N times the number of observations per time step). The vectors u_j and v_j are the singular vectors of G belonging to the singular values σ_j, where G has the singular value decomposition $G = U\Sigma V^T$, with U and V orthonormal matrices and Σ the diagonal matrix with entries $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_l \geq 0$. Hence the function $\frac{\sigma_j^2}{\sigma_j^2+\mu^2}$ acts as a filter function for small singular values σ_j.

It is known from image processing (Hansen *et al.* (2006)) that instead of taking the L_2-norm for the regularisation term $\mu^2\|z\|_2^2$ (that is the background term) the L_1-norm gives a better performance when sharp edges need to be recovered. We use this idea in order to try to improve the performance of 4DVar for the recovery of fronts.

3 L_1-norm regularisation

With the notation in (8), the minimisation problem in (7) can be written as

$$\min_z \hat{J}_2(z) = \min_z\{\|f - Gz\|_2^2 + \mu^2\|z\|_2^2\},\ \mu^2 = \frac{\sigma_o^2}{\sigma_b^2}, \tag{11}$$

where the second term is a regularisation term, μ^2 is the regularisation parameter. In the literature, there has been a growing interest in using L_1-norm regularisation for image restoration, see, for example Fu *et al.* (2006); Agarwal *et al.* (2007); Schmidt *et al.* (2007).

Hence, in this paper we consider the effects of L_1-norm regularisation for variational data assimilation by replacing the 2-norm in the regularisation term $\mu^2\|z\|_2^2$ of (11) by the 1-norm to obtain

$$\min_z \hat{J}_1(z) = \min_z\{\|f - Gz\|_2^2 + \mu^2\|z\|_1\},\ \mu^2 = \frac{\sigma_o^2}{\sigma_b^2}. \tag{12}$$

In fact, in general we can consider a p-norm, in the regularisation term, which leads to the minimisation problem

$$\min_z \hat{J}_p(z) = \min_z\{\|f - Gz\|_2^2 + \mu^2\|z\|_p^p\},\ \mu^2 = \frac{\sigma_o^2}{\sigma_b^2}, \tag{13}$$

The advantage of using the L_1-norm is that the solution is more robust to outliers. It has been observed that a small number of outliers have less influence on the solution, (Fu *et al.* (2006)). Edges in images lead to outliers in the regularisation term and hence, L_1-norms for the regularisation terms give a better result in image restoration. This is the motivation behind our approach for variational data assimilation. We will observe, that for fronts and shocks, L_1-norm regularisation in 4DVar gives much better results than the standard L_2-norm approach.

Note that both the L_2-norm and the L_1-norm minimisation can be interpreted from a Bayesian point of view. For the L_2-norm approach - which is equivalent to standard 4DVar - a Gaussian distribution is assumed for the error in the prior - that is, for the background error. For the L_1-norm the background error is assumed to have a Laplacian distribution (a modulus exponential distribution).

The following examples include a square wave advected using the linear advection equation, a moving front in Burgers' equation and a shock developing in Burger's equation. For all these examples we use a so-called true model (where we take the observation from) and another model, which is different from the truth and hence, introduces a model error. The different

models we use are introduced in the next section. In all examples we observed that the L_1-norm regularisation indeed gives better results than the L_2-norm approach.

Note that in all the examples we keep the regularisation parameter μ fixed, as we are only investigating the influence of the norm in the regularisation term, but not the size of the regularisation parameter.

4 Models

In this section we consider the problem

$$u_t + [f(u)]_x = 0, \tag{14}$$

where $f(u)$ is given either by

$$f(u) = u, \tag{15}$$

for the linear advection equation or by

$$f(u) = \frac{1}{2}u^2, \tag{16}$$

for the inviscid Burgers' equation.

This general problem can be discretised using the following numerical schemes. The upwind scheme is given by

$$U_j^{n+1} = U_j^n - \frac{\Delta t}{\Delta x}\left(f(U_j^n) - f(U_{j-1}^n)\right). \tag{17}$$

The model equations for the Lax-Friedrich method are

$$U_j^{n+1} = \frac{1}{2}(U_{j-1}^n + U_{j+1}^n) - \frac{\Delta t}{2\Delta x}(f(U_{j+1}^n) - f(U_{j-1}^n)), \tag{18}$$

and the model equations for the Lax-Wendroff method in conservative form are given by

$$\begin{aligned} U_j^{n+1} = U_j^n &- \frac{\Delta t}{2\Delta x}(f(U_{j+1}^n) - f(U_{j-1}^n)) \\ &+ \frac{\Delta t^2}{2\Delta x^2}\Big(A_{j+\frac{1}{2}}(f(U_{j+1}^n) - f(U_j^n)) \\ &\qquad - A_{j-\frac{1}{2}}(f(U_j^n) - f(U_{j-1}^n))\Big). \end{aligned} \tag{19}$$

All equations are valid for $j = 1, \ldots, N$, where f is given by (15) or (16). For the Lax-Wendroff method $A(u) = f'(u)$ denotes the Jacobian matrix and $A_{j\pm\frac{1}{2}}$ is the Jacobian matrix evaluated at $\frac{1}{2}(U_j^n + U_{j\pm1}^n)$. For more details on the the above methods we refer to LeVeque (1992).

5 Linear advection equation - Example 1

Consider the linear advection equation

$$u_t + u_x = 0,$$

on the interval $x \in [0,1]$, with periodic boundary conditions. The initial solution is a square wave defined by

$$u(x,0) = \begin{cases} 0.5, & 0.25 < x < 0.5 \\ -0.5, & x < 0.25 \quad \text{or} \quad x > 0.5. \end{cases}$$

This wave moves through the time interval; the true solution is obtained by the method of characteristics and the model equations are defined by the upwind scheme (17) with boundary conditions $U_0^{n+1} = U_N^{n+1}$, where $j = 1, \ldots, N$, $\Delta x = \frac{1}{N}$ and n is the number of time steps. The same example was used in Griffith and Nichols (2000). For this example we take $N = 100$, $\Delta t = 0.005$.

5.1 A standard experiment

We consider an assimilation window of length 40 time steps. After the assimilation period we compute the forecast for another 40 time steps, and hence, 80 time steps are considered in total. For the background and observation error covariance matrices we take $B = I$ and $R = 0.01I$, as we put more emphasis on the observations rather than on the background. Moreover, for the background we choose $U_b^0 = 0$. The background thus contains significant errors with variance of order unity. We test several cases.

1. Perfect observations are taken everywhere in time and space.

2. Perfect observations are taken every 20 points in space and every 2 time steps.

3. Imperfect observations are taken every 20 points in space and every 2 time steps; for the observations we introduce Gaussian noise with mean zero and variance 0.01.

For all cases we test 4DVar with both L_2-norm and L_1-norm regularisation. In the implementation we use $\|\cdot\|_p^p$ with $p = 1.00001$ instead of $p = 1$, in order to simplify the computations. Figures 1 - 6 show the results for Example 1, where the linear advection equation is used as a model. L_1-regularisation is advantageous if shocks and fronts are present.

In the plots the true solution is represented by a thick dot-dashed line (called 'Truth' in the legend). The model solution (which is derived from the upwind method) is shown as a dashed line (called 'Imperfect model' in the legend). This is the best that we can achieve in the data assimilation, as the model error is always present. The solution obtained from the assimilation process by incorporating the observations is given by the solid line (called 'Final solution' in the legend).

For perfect observations the result for 4DVar is shown in Figure 1, and that for L_1-regularisation in Figure 2. The analysis obtained by 4DVar is very inaccurate, with many oscillations near the discontinuity (first plot in Figure 1). When L_1-regularisation is used, the initial condition is correct (first plot in Figure 2). The same result is true for partial observations (Figure 3 for 4DVar versus Figure 4 for L_1-regularisation) and for imperfect partial observations (Figure 5 for 4DVar versus Figure 6 for L_1-regularisation). In 4DVar the initial conditions are chosen to compensate on average for the model errors over the entire time window, and hence 4DVar does not produce an accurate estimate of the truth at the initial time. From Figures 3 and 5 for 4DVar we also see that the forecast is inaccurate due to the incorrect estimate produced at the end of the assimilation window. For L_1-regularisation (Figures 4 and 6) these problems do not occur.

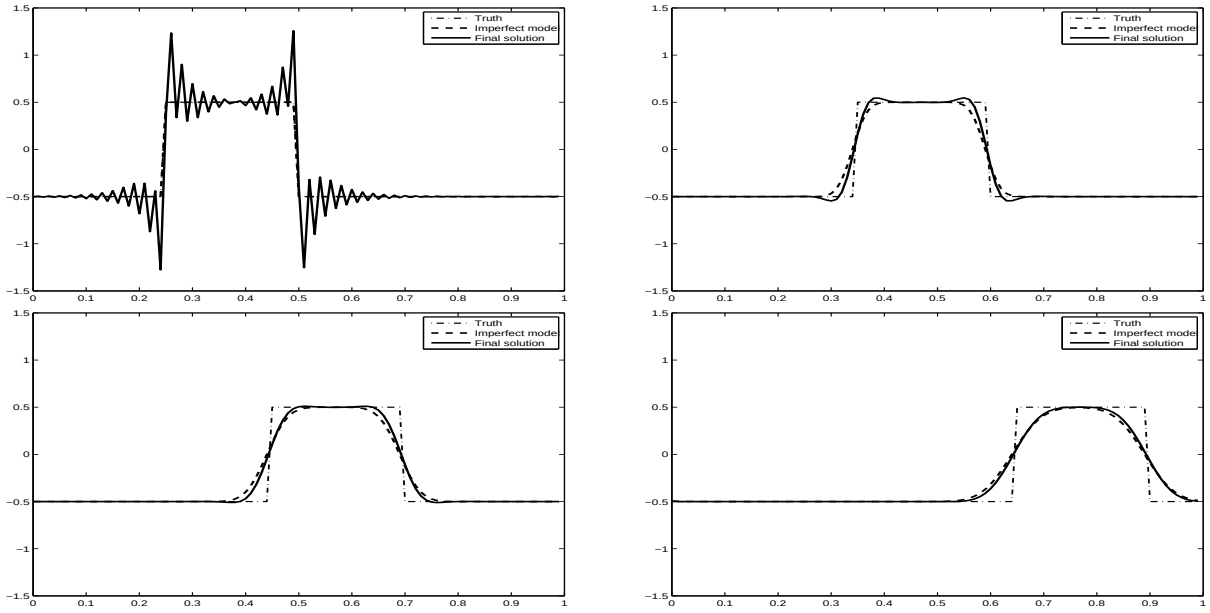


Figure 1: Results for **4DVar** applied to the linear advection equation where the initial condition is a square wave. We take **perfect observations at each point in time and space** over the assimilation interval which is 40 time steps. The four plots show the initial conditions at $t = 0$ and the result after 20, 40 and 80 time steps. 4DVar leads to oscillations in the initial condition.

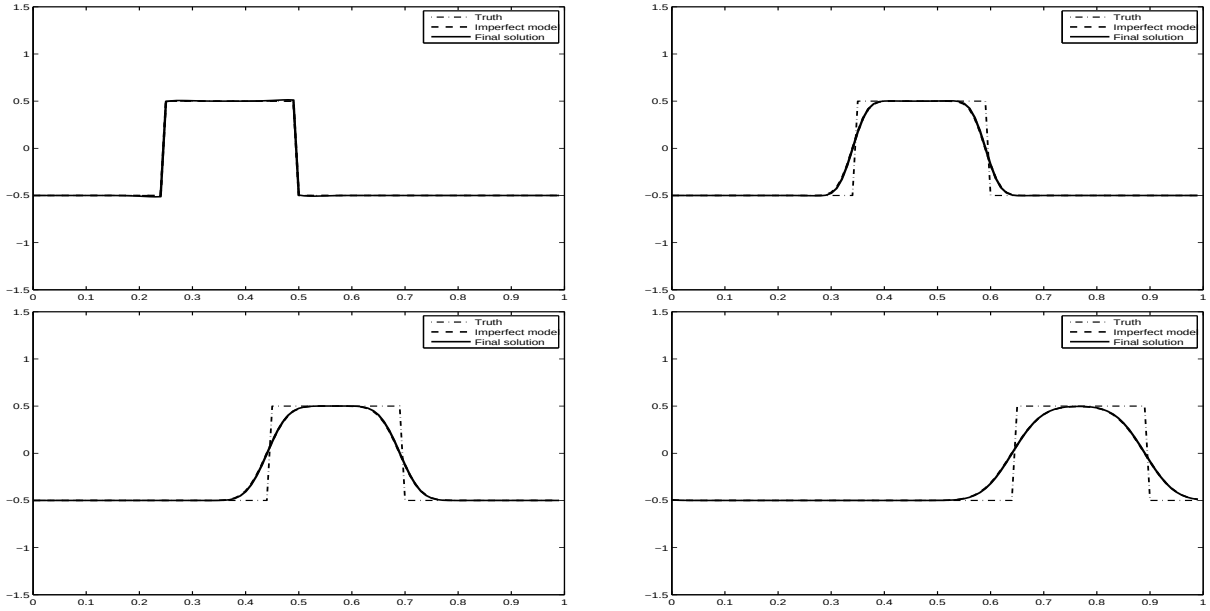


Figure 2: Results for L_1**-regularisation** for the same data as in Figure 1. L_1-regularisation gives the best possible result for the initial condition.

In the next two subsections we change the experimental design of the problem slightly, in order to check the robustness of L_1-norm regularisation.

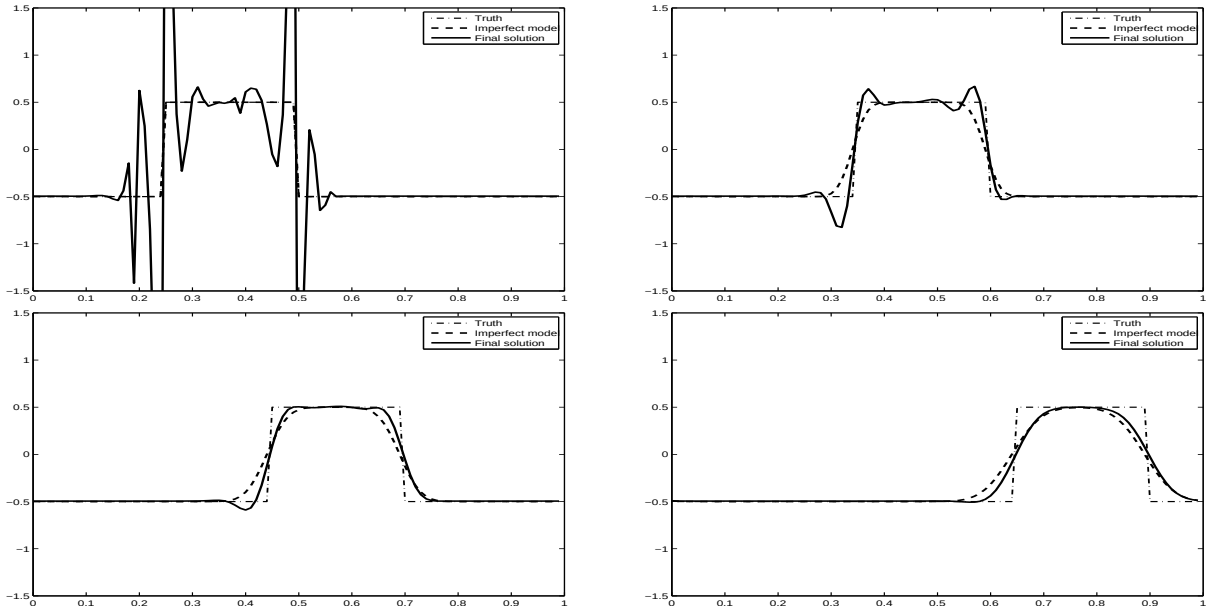


Figure 3: Results for **4DVar** for the same data as in Figure 1 but with **perfect observations every** 20 **points in space and every** 2 **time steps**. 4DVar leads to oscillations in the initial condition.

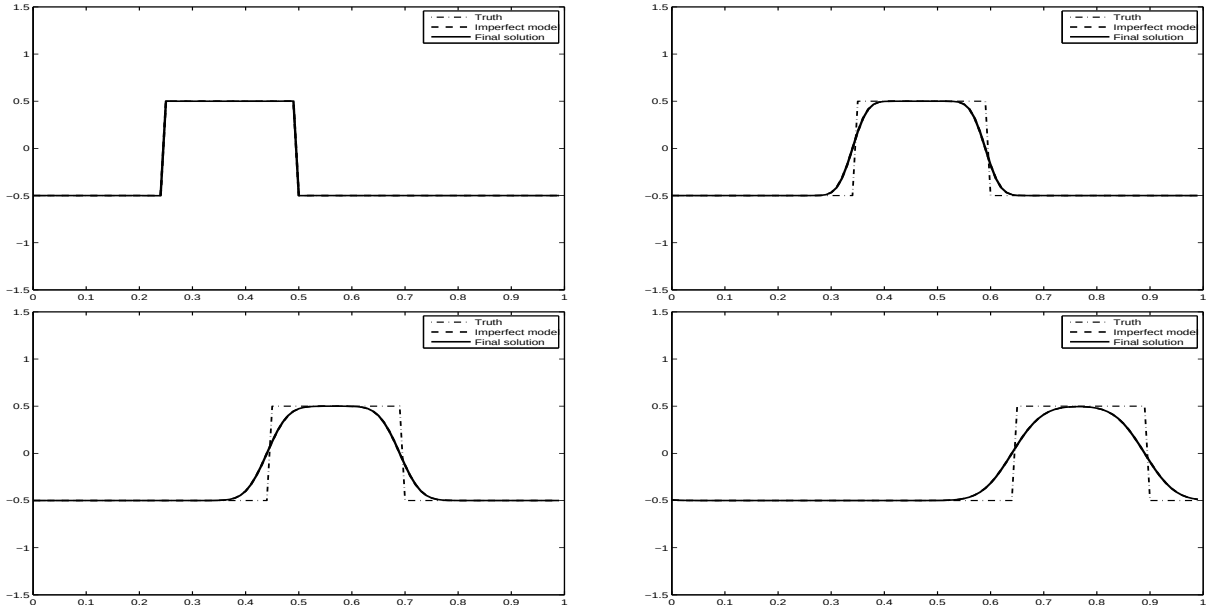


Figure 4: Results for L_1**-regularisation** for the same data as in Figure 3. L_1-regularisation gives the best possible result for the initial condition.

5.2 Changing the background error covariance matrix

We take precisely the same experiment as in the previous subsection 5.1, however, we change the background error covariance matrix from the identity matrix

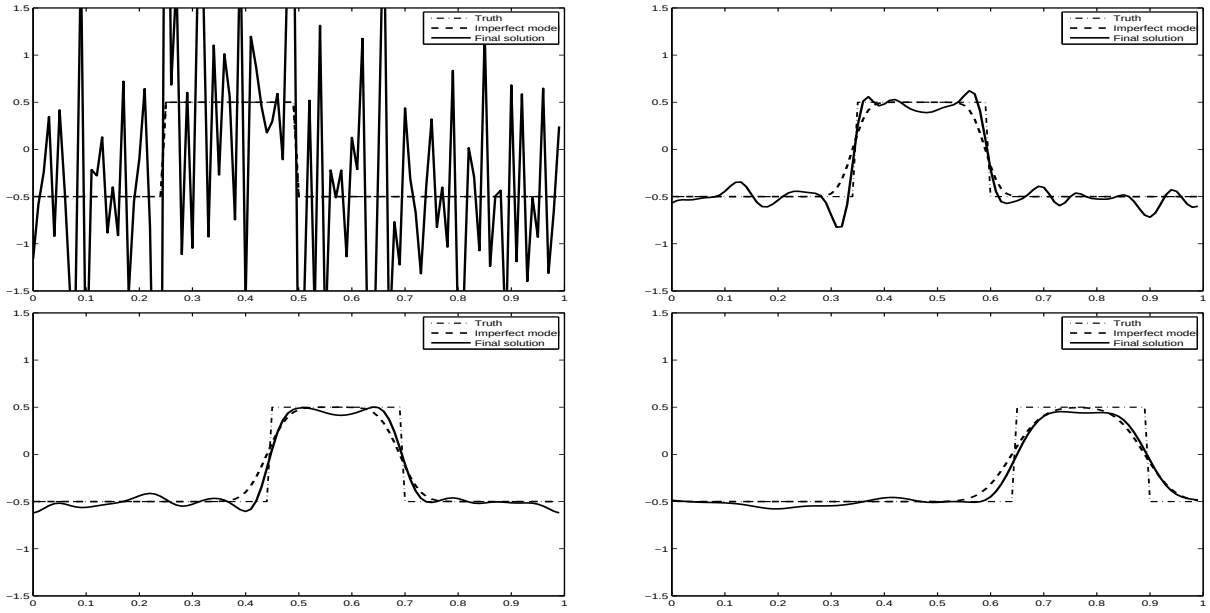


Figure 5: Results for **4DVar** for the same data as in Figure 1 but with **imperfect observations every** 20 **points in space and every** 2 **time steps**. 4DVar leads to bad oscillations in the initial condition and also to a misplaced discontinuity in the forecast.

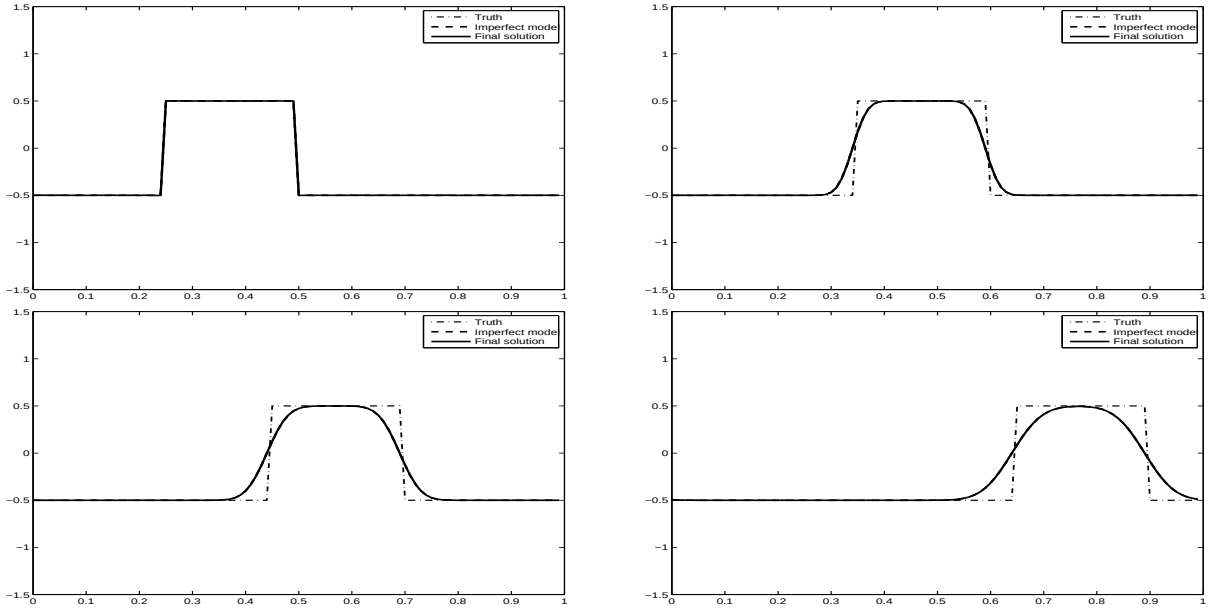


Figure 6: Results for L_1**-regularisation** for the same data as in Figure 5. L_1-regularisation gives the best possible result for the initial condition.

- to $B = 0.01I$ so that background and observation error covariances are equal and we put the same emphasis on background and observations. Furthermore, for this example, instead of taking a zero background we use $U_b^0(x(j)) = U^0(x(j)) - 0.1$, a shifted initial condition, which is closer to the truth and has an error within the specified variance.

- to a Gaussian covariance matrix B with entries

$$B_{ij} = e^{-\frac{|i-j|}{2L^2}}, \quad \text{where} \quad L = 5. \tag{20}$$

Hence B is a symmetric matrix with diagonal entries equal to unity and off-diagonal entries that decay exponentially. This background error covariance matrix spreads the information from the observations more adequately and the error variance is still unity. Note that for this matrix the inverse is a tridiagonal matrix. For the background we again choose $U_b^0 = 0$, which has errors consistent with the choice of B.

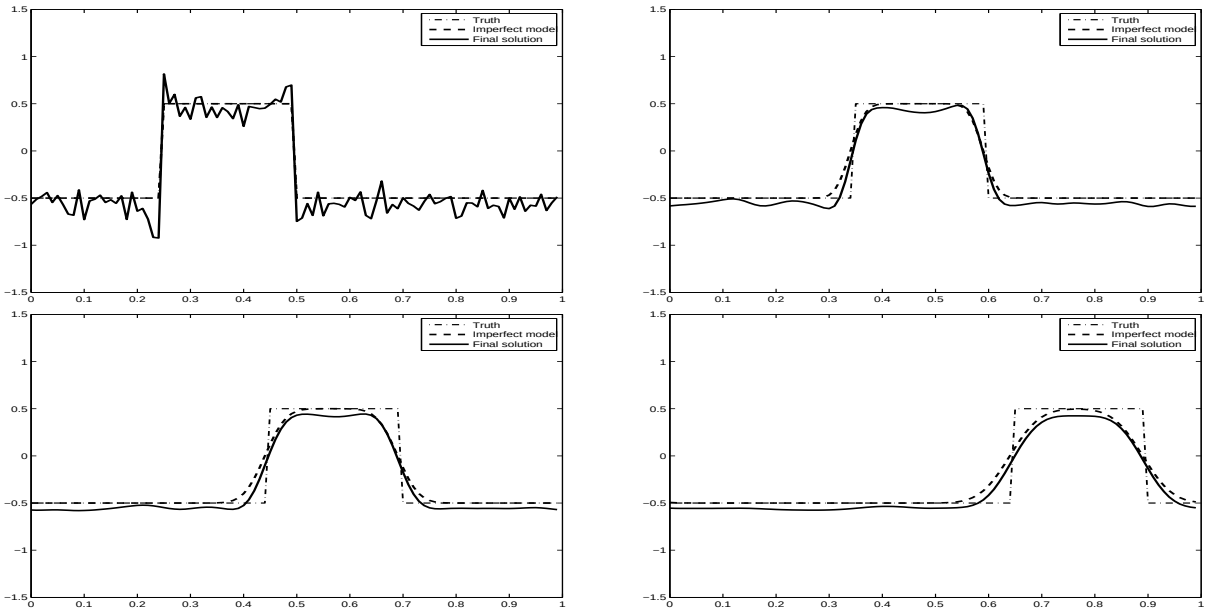


Figure 7: Results for **4DVar** for the same data as in Figure 5, but for $B = 0.01I$ and a different background estimate.

We only present the results for imperfect and partial observations, as this represents the most realistic case.

Figures 7 - 8 show the results for Example 1, when $B = 0.01I$. Clearly, L_1-regularisation gives the best possible result.

Figures 9 and 10 show the results where the matrix B from (20) is used as a background error covariance matrix. For this choice of B, the results for 4DVar (Figure 9) are better than the results for the diagonal matrix B (Figure 5) because information is spread via the B matrix. However, L_1-norm regularisation (Figure 10) still behaves consistently better than L_2-norm regularisation (Figure 9).

5.3 Changing the length of the assimilation window

Again, we take the same experimental data as in Subsection 5.1; this time, however, we reduce the size of the assimilation window from 40 time steps to 5 time steps and carry out the following test: we take imperfect observations every 5 points in space and every 2 time steps with Gaussian noise of mean zero and variance 0.01.

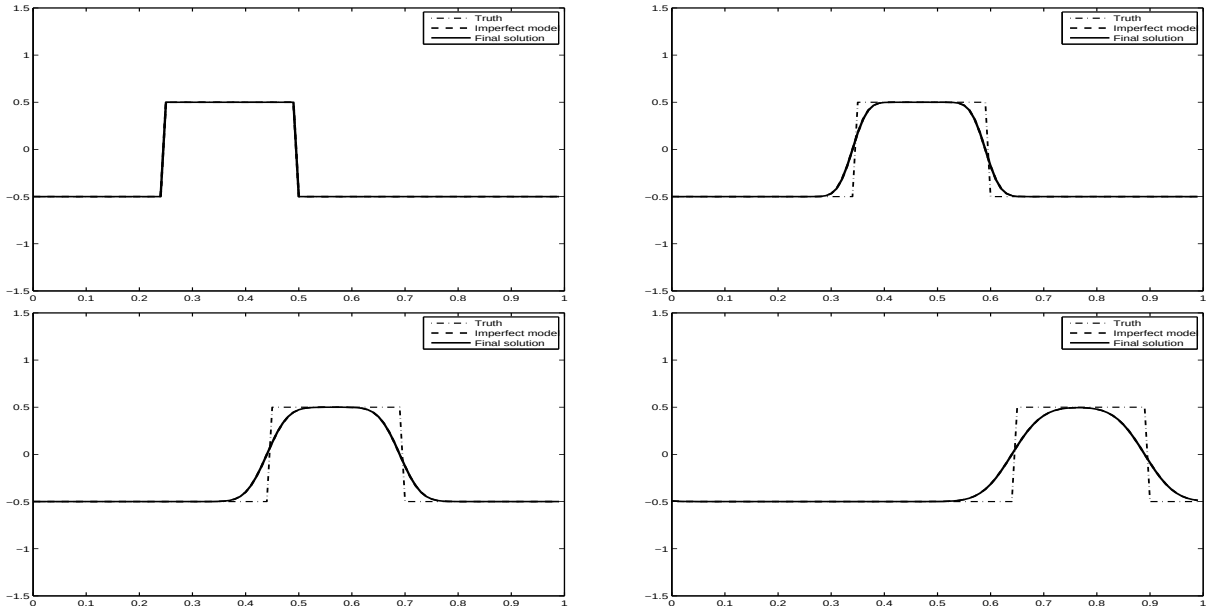


Figure 8: Results for L_1**-regularisation** for the same data as in Figure 7.

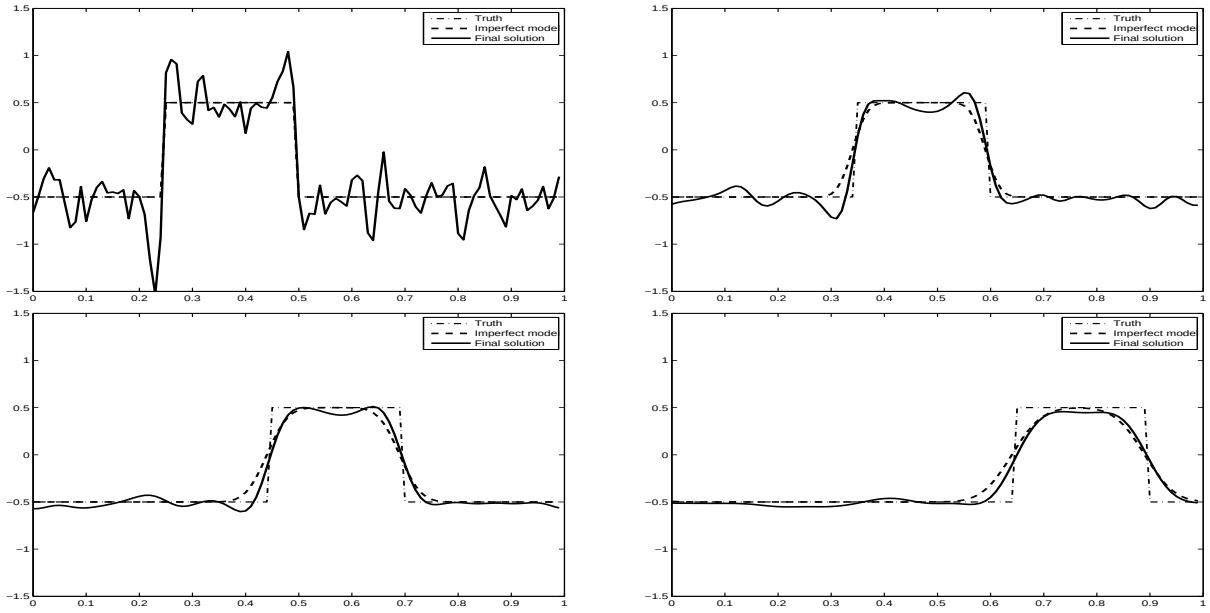


Figure 9: Results for **4DVar** for the same data as in Figure 5, but for B with $B_{ij} = e^{-\frac{|i-j|}{2L^2}}$, where $L = 5$.

Figures 11 and 12 show the results for Example 1, where the linear advection equation is used as a model and the size of the assimilation window is reduced. The first observation that we can make is that again the regularisation using the L_1-norm (Figure 12) is consistently better than that using the L_2-norm (Figure 11). Standard 4DVar produces oscillations, in particular in the initial conditions, whereas the L_1-norm regularisation does not show any oscillations. The

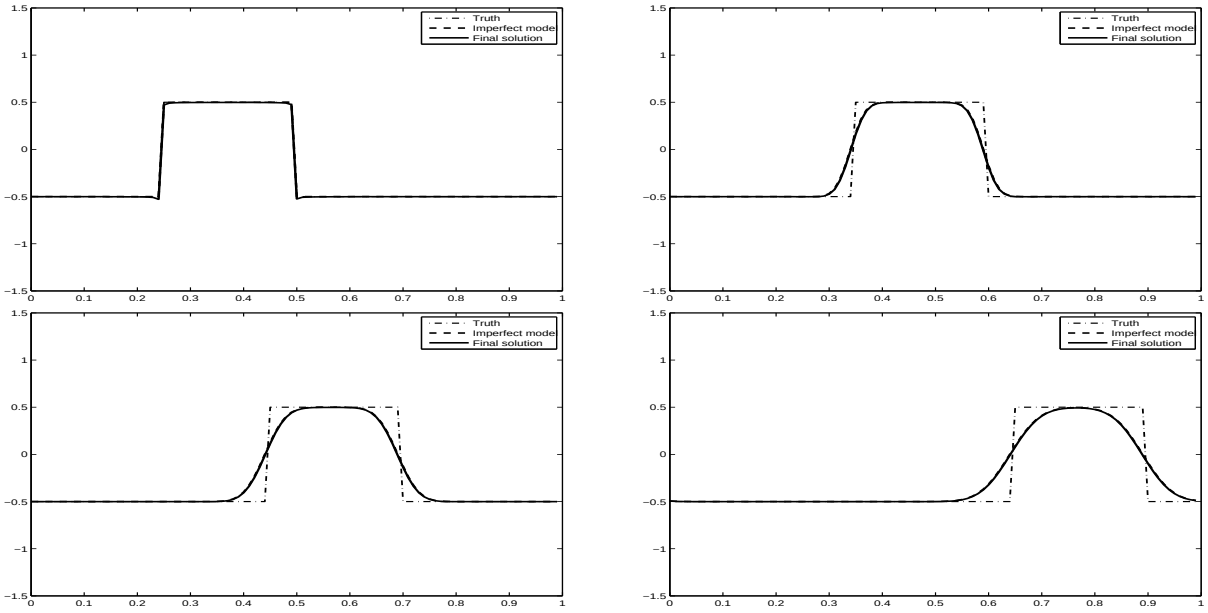


Figure 10: Results for L_1**-regularisation** for the same data as in Figure 9.

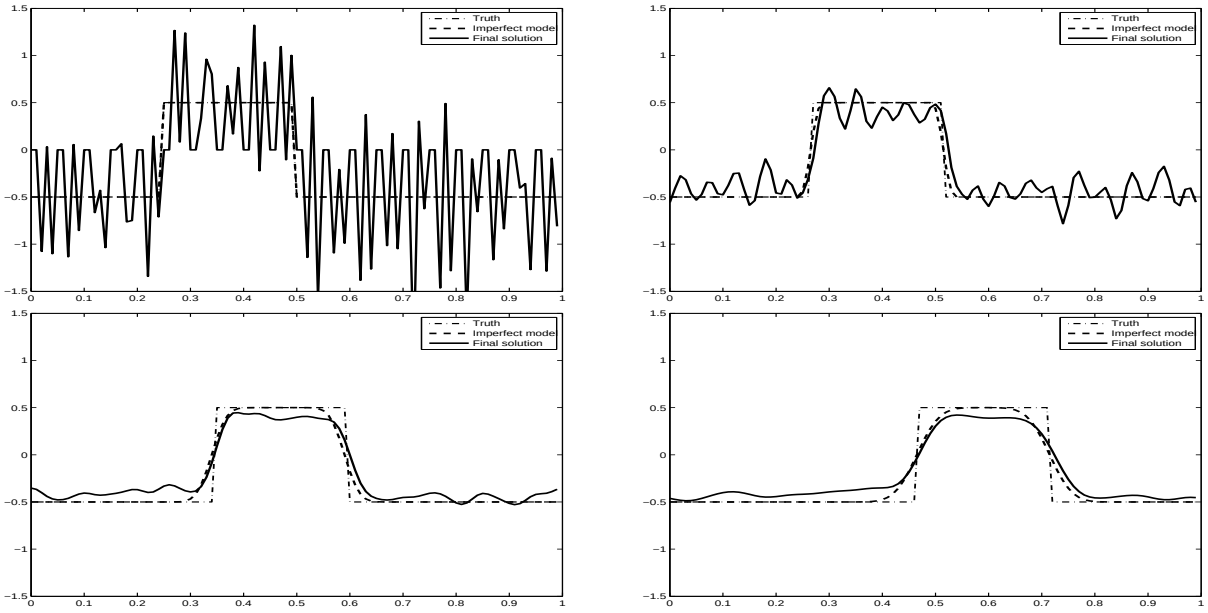


Figure 11: Results for **4DVar** applied to the linear advection equation where the initial condition is a square wave. We take **imperfect observations every** 5 **points in space and every** 2 **time steps** over the assimilation interval which is 5 time steps. The four plots show the initial conditions at $t = 0$ and the result after 5, 20 and 45 time steps. 4DVar leads to oscillations in the initial condition and a misplaced discontinuity in the forecast.

oscillations in the initial conditions in standard 4DVar then lead to errors in the forecast (see plots for $t = 5$, $t = 20$ and $t = 45$ in Figure 11). We observe also that for a smaller assimilation window, the oscillations in the 4DVar analysis are smaller than those in the larger assimilation

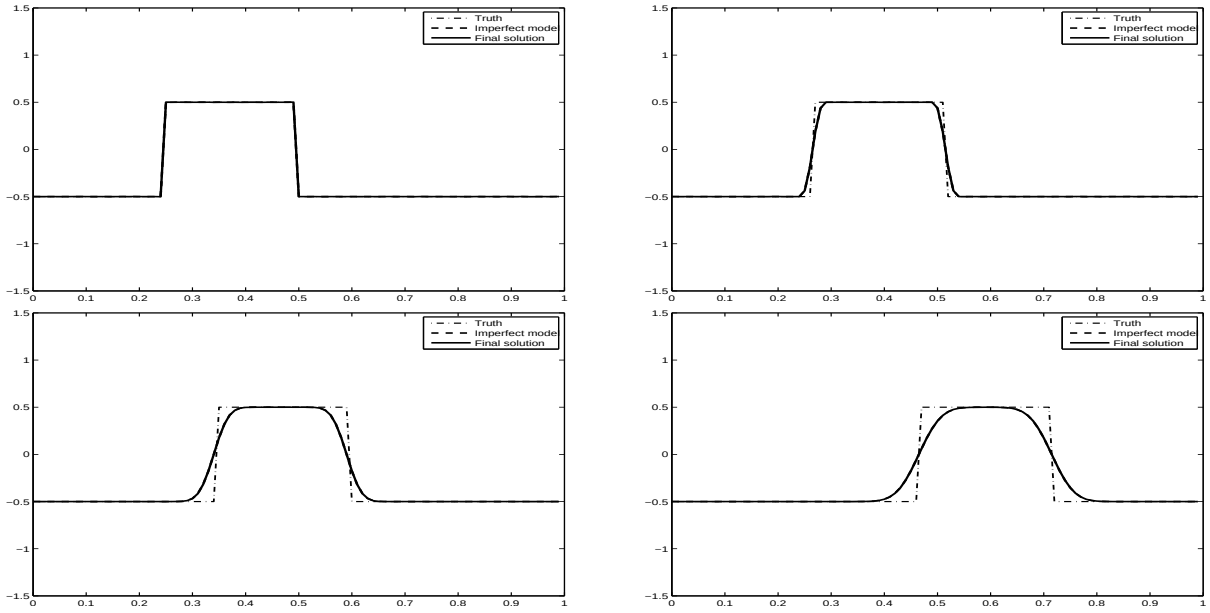


Figure 12: Results for L_1**-regularisation** for the same data in Figure 11. L_1-regularisation gives the best possible result for the initial condition.

window (compare the first plot in Figure 11 with the first plot in Figure 5). The primary reason is that there are fewer errors over the time interval for which the initial data must compensate on average.

6 The Burgers' equation - Example 2

Consider the inviscid Burgers' equation

$$u_t + uu_x = 0$$

that is equation (14), where the function f is given by (16), with initial conditions

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

Discretising these we obtain

$$x(j) = 10(j - 1/2)\Delta x; \tag{21}$$

$$U_j^0 = \begin{cases} 2, & 0 \le x(j) < 2.5 \\ 0.5, & 2.5 \le x(j) \le 10, \end{cases} \tag{22}$$

with $\Delta x = \frac{1}{N}$ and $j = 1, \ldots, N$. We are interested in this equation as the solution forms a shock front that moves forward in time. The method of characteristics results in a Riemann problem and its solution is given by

$$u(x,t) = \begin{cases} 2, & 0 \le x < 2.5 + st \\ 0.5, & 2.5 + st \le x \le 10, \end{cases} \tag{23}$$

where $s = 1.25$ is the shock speed. We use this result as the perfect model which generates the true solution to the problem. It is also used to generate the observations. For the imperfect model we choose one of the numerical solutions to the problem, namely the Lax-Friedrich method (18) or the Lax-Wendroff method (19).

It is well-known that the Lax-Friedrich method leads to smearing out the shock and hence introduces a model error there. The Lax-Wendroff method recovers the shock speed but leads to oscillations near the shock and hence a model error is introduced via this numerical scheme. We refer to LeVeque (1992) and Morton and Mayers (2005) for details. We implement both methods in 4DVar with L_2-regularisation and L_1-regularisation.

6.1 Lax-Friedrich method

The model equations for the Lax-Friedrich method are given by (18) for $j = 1, \ldots, N$, where f is given by (16), and hence a model error is introduced. The true states in this case are given by the exact solution to Burgers' equation (23). For the initial conditions $U^0(x(j))$ in the imperfect model we use (22) for comparison. For the background $U_b^0(x(j))$ we take $U_b^0(x(j)) = U^0(x(j)) - 0.1$, a shifted initial condition.

For all our tests we use the background error covariance matrix B from (20) in Section 5 and $R = 0.01I$, since we put more emphasis on the observations than on the background. As we have seen in Section 5, this choice for matrix B gives better results for 4DVar than a simple diagonal matrix. Furthermore, $\Delta t = 0.001$, $N = 100$ and the length of the assimilation window is 100 time steps. After the assimilation period we compute the forecast for another 100 time steps, hence, 200 time steps are plotted in total. We test the same cases as in Subsection 5.1, that is perfect full observations, perfect partial observations and imperfect partial observations.

For all cases we test 4DVar with both L_2-norm and L_1-norm regularisation. We only show the results for imperfect observations as this is the most realistic case. However, we observe that for perfect (full and partial) observations, 4DVar leads to oscillations near the discontinuity at the initial condition whereas L_1-norm regularisation gives the best possible result.

Specifically, if we take perfect observations everywhere in time and space, the final solution for 4DVar leads to oscillations in the initial condition, and hence the incorrect initial condition for the forecast. Over the assimilation and forecast period this error is decaying. The analysis for L_1-regularisation, however, coincides with the model solution, the best possible solution. This is true for both the initial condition and the forecast period.

If we consider perfect observations, but made only partially in space and time, again the final solution for 4DVar leads to oscillations in the initial condition and, therefore, the incorrect initial condition for the forecast. Since not enough observations are available this inaccurate initial condition leads to the wrong position of the shock front in the forecast. On the other hand the final solution for the L_1-regularisation coincides with the model solution.

The third case treats both partial and imperfect observations; the result for 4DVar is shown in Figure 13, the result for L_1-regularisation in Figure 14. The analysis for 4DVar does not resemble the true initial condition at all (first plot $t = 0$ in Figure 13). The RMS error and the oscillations are decaying over the assimilation period, but are still showing the wrong position for the shock in the forecast. For L_1-regularisation however (see Figure 14), we get the best possible solution.

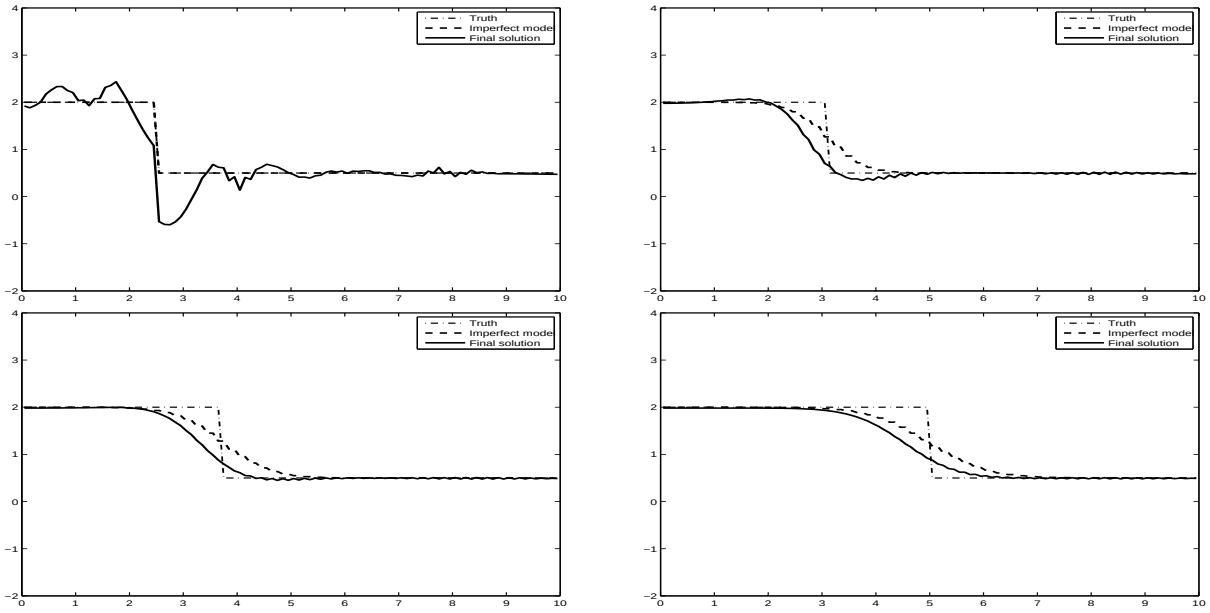


Figure 13: Results for **4DVar** applied to Burgers' equation where the initial condition is a step function and the Lax-Friedrich method is applied to solve the problem. We take **imperfect observations every** 20 **points in space and every** 2 **time steps** over the assimilation interval which is 100 time steps. The four plots show the initial condition at $t = 0$ and the result after 50, 100 and 200 time steps. 4DVar leads to oscillations in the initial condition and to the wrong position for the shock front in the forecast.

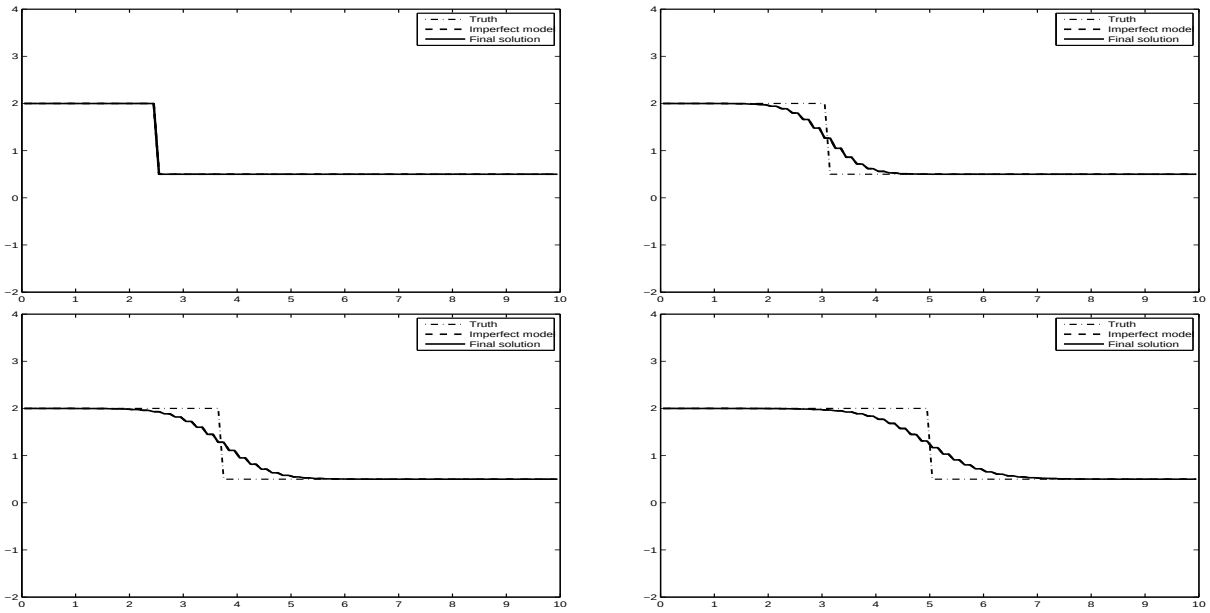


Figure 14: Results for L_1**-regularisation** applied to the same problem as in Figure 13. L_1-regularisation gives the best possible result for the initial condition.

6.2 Lax-Wendroff method

The model equations for the Lax-Wendroff method in conservative form are given by (19). It is well-known that by using the Lax-Wendroff method oscillations will appear near the shock; however, it does not smear out the shock and recovers the shock speed correctly. The true states in this case are again given by the exact solution to Burgers' equation (23). For the initial conditions $U^0(x(j))$ in the imperfect model we use (22) again for comparison. For the background $U_b^0(x(j))$ we use $U_b^0(x(j)) = U^0(x(j)) - 0.1$, a slightly shifted initial condition.

We use $B = 0.01I$, $R = 0.01I$ and otherwise the same data and experimental tests as for the Lax-Friedrich method. We only show the results for the imperfect observations here, as they represent the most realistic case.

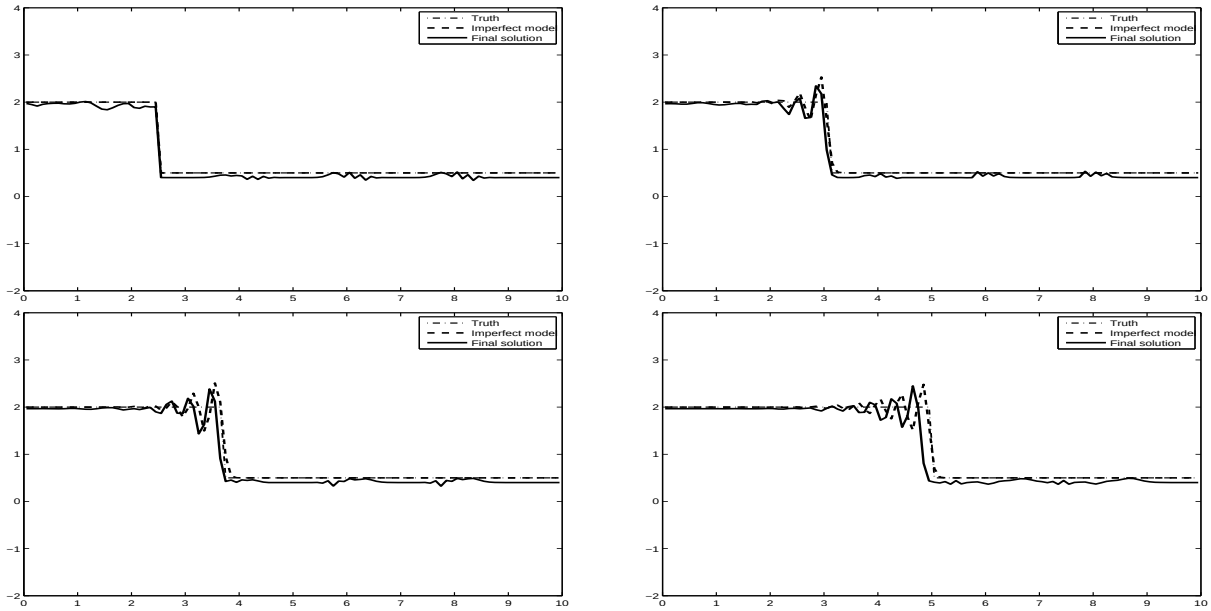


Figure 15: Results for **4DVar** applied to the Burgers' equation where the initial condition is a step function and the Lax-Wendroff method is applied to solve the problem. We take **imperfect observations every** 20 **points in space and every** 2 **time steps** over the assimilation interval which is 100 time steps. The four plots show the initial conditions at $t = 0$ and the result after 50, 100 and 100 time steps. 4DVar predicts the wrong position for the shock front in the forecast.

Figures 15 and 16 show the results for Example 2, where the Lax-Wendroff method is used for the model equations and imperfect partial observations are made. The result for L_1-norm regularisation is better than the result for 4DVar throughout the assimilation window and the forecast.

From Figure 15 we see that the assimilation of the observations using 4DVar leads to small errors in the initial conditions (see first plot in Figures 15). These errors lead to a slightly incorrect position for the shock in the forecast (see last plot in Figure 15). The solution for L_1 regularisation, shown in Figure 16, is equal, however, to the model solution, which is the best possible result.

Note that for Burgers' equation we also tested the change of the background error covariance matrix and the change of the size in the assimilation window. We found similar results as for

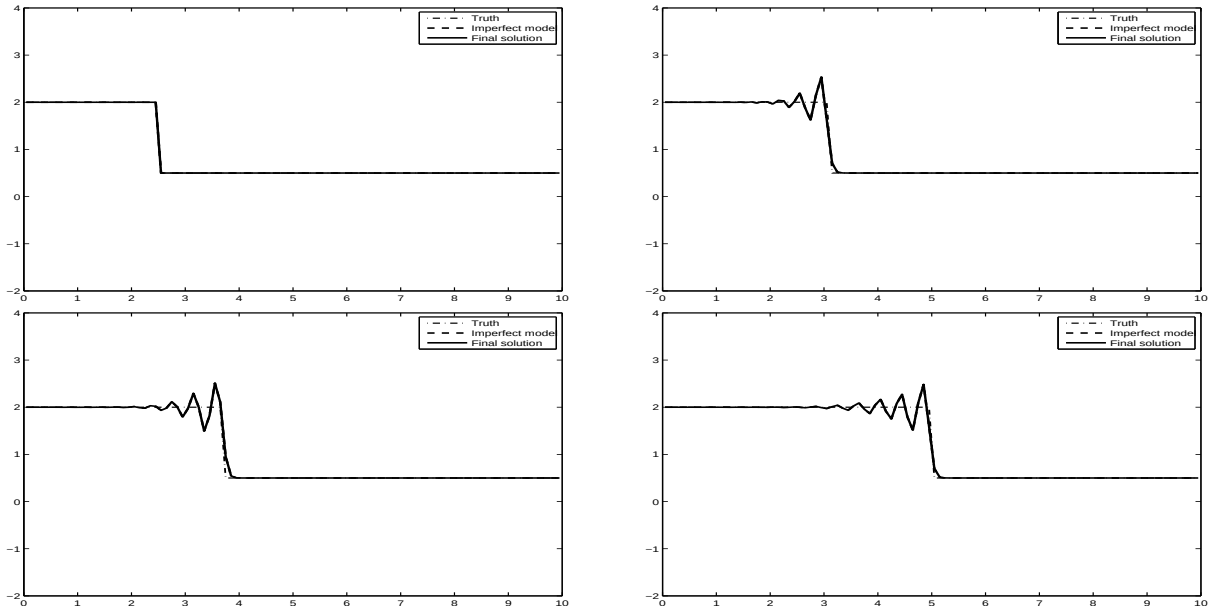


Figure 16: Results for L_1**-regularisation** with the same setup as in Figure 15. L_1-regularisation gives the best possible result for the initial condition.

the linear advection equation in Section 5.

We also note that we have tested several different values for p in equation (13). However, the best results were obtained when p was smaller than 1.01.

7 A shock formation in Burgers' equation

We next consider Burgers' equation (14) with f given by (16) again, where instead of (22) we have initial conditions

$$U^0(x(j)) = \sin(2\pi x(j)/10) + \frac{1}{2}\sin(\pi x(j)/10)$$

and $x(j)$ is given by (21). In this case, a shock develops in finite time leading to evolving high gradients in the forecast model.

For this problem we take the upwind scheme given by (17) with f given by (16) as the exact model (the 'Truth'), from which we take the observations. The output of the Lax-Friedrich method is taken as the model for the data assimilation process (and thereby, a model error is introduced). For our tests we use B from (20) and $R = 0.01I$. Furthermore, $\Delta t = 0.001$, $N = 100$ and the length of the assimilation window is 100 time steps. We compute a further 100 time steps for the forecast, so in total 200 time steps are calculated. For this problem we carry out the following test: We take imperfect observations at every 20th point in space and every second time step; for the observations we introduce Gaussian noise with mean zero and variance 0.01. We test 4DVar with both L_2-norm and L_1-norm regularisation.

Figures 17 - 19 show the results for this example. In Figure 17 the result for 4DVar is plotted, in Figure 18 the result for L_1-regularisation is shown. We immediately see that L_1-regularisation recovers the initial condition much more accurately than standard 4DVar with L_2-regularisation

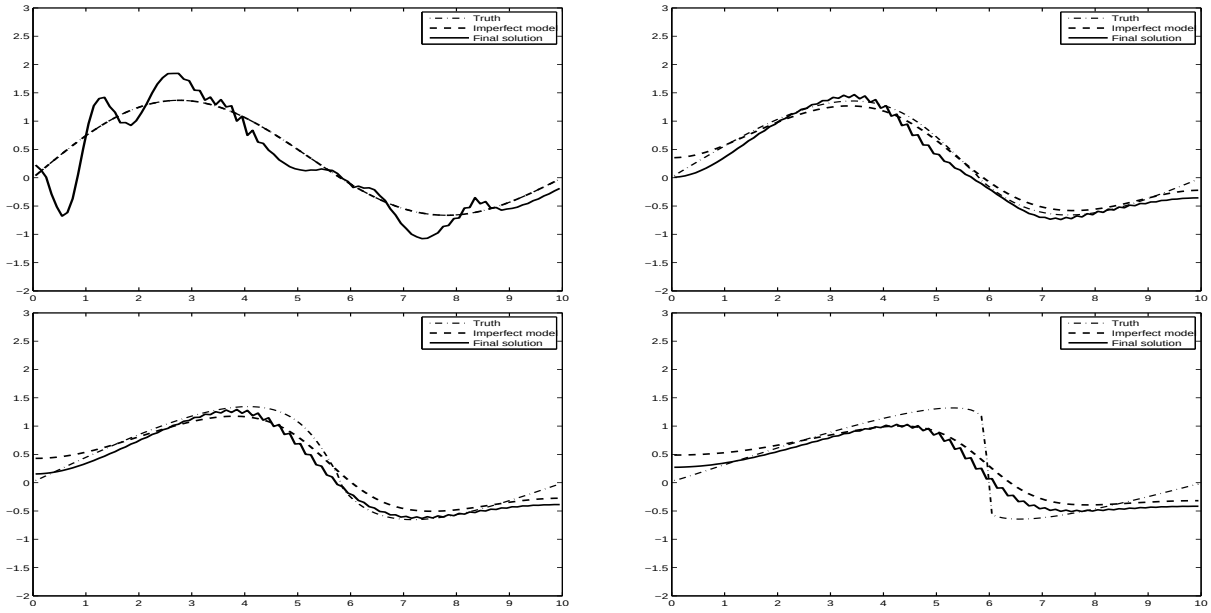


Figure 17: Results for **4DVar** applied to the Burgers' equation where the initial condition is $U^0(x(j)) = \sin(2\pi x(j)/10) + \frac{1}{2}\sin(\pi x(j)/10)$ and an upwind method is applied to solve the problem. We take **imperfect observations every** 20 **points in space and every** 2 **time steps** over the assimilation interval which is 100 time steps. The four plots show the initial conditions at $t = 0$ and the result after 50, 100 and 200 time steps. 4DVar leads to severe oscillations in the initial condition which are propagated into the forecast.

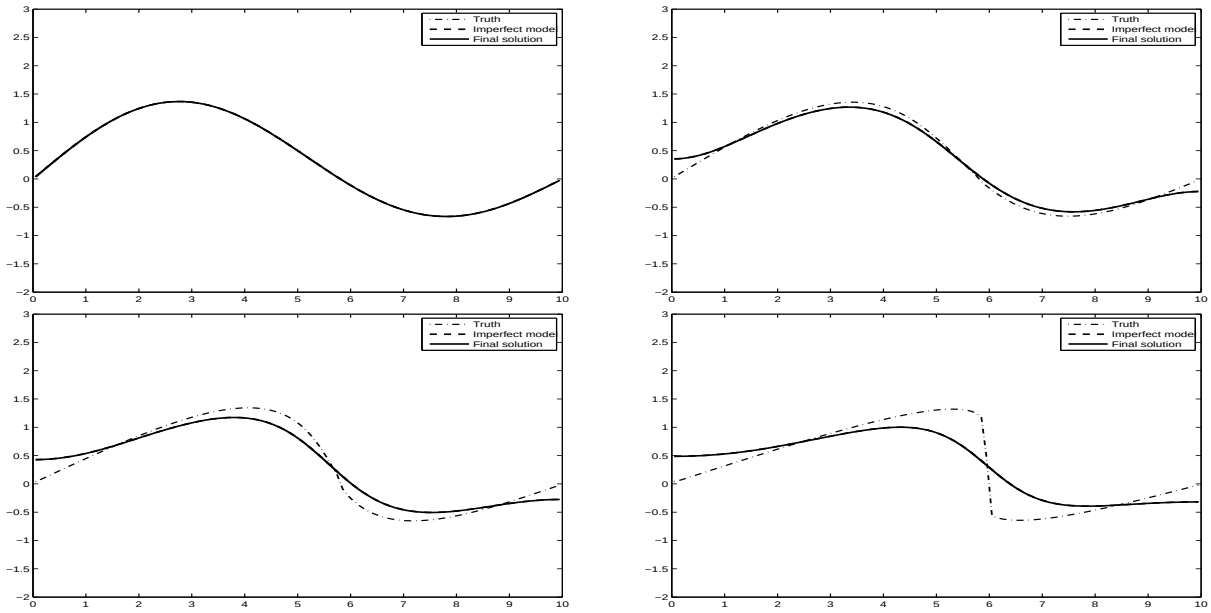


Figure 18: Results for L_1**-regularisation** applied to the same setup as in Figure 17. L_1-regularisation gives the best possible result for the initial condition.

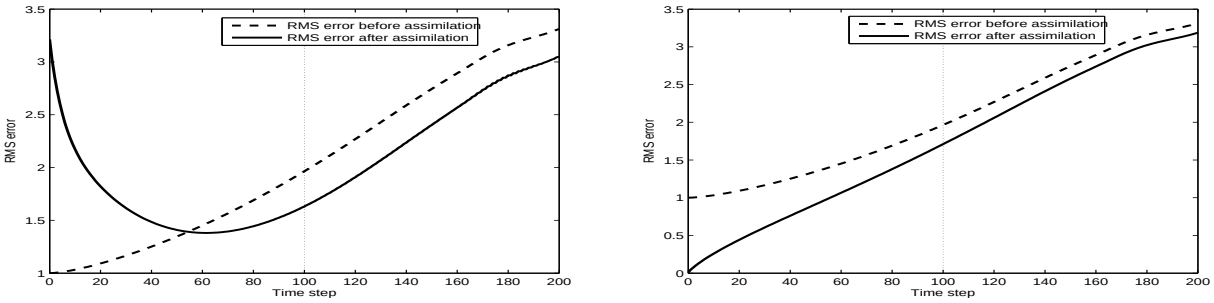


Figure 19: Root mean square error for data assimilation using Burgers' equation with initial conditions $U^0(x(j)) = \sin(2\pi x(j)/10) + \frac{1}{2}\sin(\pi x(j)/10)$ and an upwind scheme applied to solve the problem. The plot on the left shows the result for 4DVar. The plot on the right shows the result for L_1-norm regularisation.

(see the first plot in Figure 17, which shows severe oscillations versus the first plot in Figure 18 which gives a perfect match with the best possible solution). The oscillations that are produced with 4DVar can still be seen in the forecast after 200 time steps (see last plot in Figure 17 whilst L_1-norm regularisation produces the best possible forecast (see final plot in Figure 18).

Figure 19 shows the root mean square error for this example and therefore summarises the results in Figures 17 and 18. We observe that the dashed line, the error before the assimilation process, is the same for 4DVar and L_1-regularisation (note the different scales in the two graphs). However, for 4DVar the initial error is very large compared to the initial condition error using L_1-norm regularisation.

8 The Lorenz equations

In this last section we apply both 4DVar and L_1-regularisation to the Lorenz model (Lorenz (1963)), a system of three first order coupled, nonlinear ordinary differential equations for the variables x, y and z,

$$\begin{aligned}\dot{x} &= \sigma(y - x),\\ \dot{y} &= \rho x - y - xz,\\ \dot{z} &= xy - \beta z,\end{aligned}$$

with initial conditions $x(0) = x_0$, $y(0) = y_0$ and $z(0) = z_0$ and parameters $\sigma = 10$, $\rho = 28$ and $b = 8/3$. There are no discontinuities present in this problem. However, we use this test in order to check if L_1-regularisation also performs well for problems without a shock or sharp front. As initial conditions for this problem we use $x_0 = y_0 = z_0 = 1$. Furthermore $\Delta t = 0.01$, and the length of the assimilation window and the subsequent forecast is 200 time steps, so 400 time steps are considered in total. For the 'truth' we take the solution to the Lorenz equations using a 4th-order Runge-Kutta method. This perfect model is also used to generate the observations. Noisy observations are taken in all 3 variables at every time step with $R = 0.2I$. We use $B = I$. For the imperfect model we use the solution obtained using the explicit Euler method. Thereby we introduce a model error.

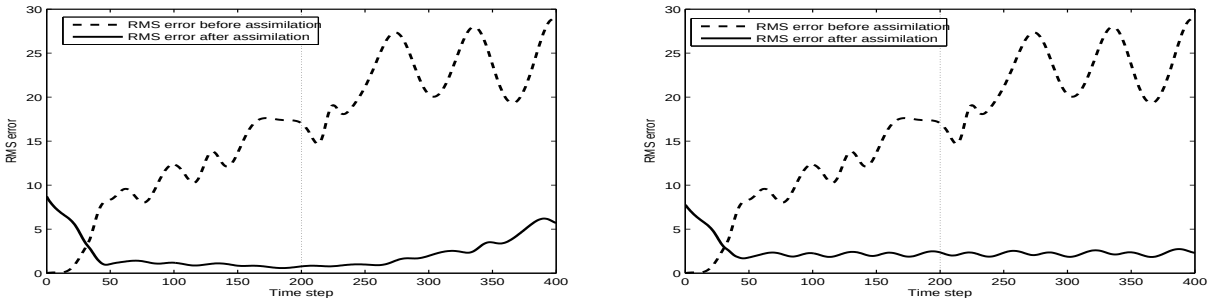


Figure 20: Root mean square error for data assimilation applied to the Lorenz equations. The plot on the left shows the result for 4DVar. The plot on the right shows the result for L_1-norm regularisation.

Figure 20 shows the results obtained where observations are assimilated into the Lorenz model using 4DVar or L_1-norm regularisation. The dashed line represents the RMS error before the observations are assimilated - the solid line shows the RMS error after the assimilation process. We see that both 4DVar and L_1-regularisation give similar results, establishing that L_1-regularisation does not do worse than 4DVar for problems without sharp fronts.

9 Conclusions and future work

In this paper we have presented L_1-regularisation, a new approach for variational data assimilation. We have given several numerical examples where shock fronts were present or developed in order to show that L_1-norm regularisation gives much better results than the standard 4DVar technique.

Future work will be to apply this technique to higher dimensional and possibly multi-scale problems. Because the minimisation process for the L_1-norm approach in (12) is more involved than that for the standard approach in (11), practical implementations for L_1-norm regularisation will also have to be examined together with the efficiency of this new approach.

Acknowledgements

The research of the first and third author is supported by Great Western Research (GWR) Grant "Numerical weather prediction: multi-scale methods and data assimilation" and by the Bath Institute for Complex Systems (BICS, EPSRC Critical Mass Grant). The research of the the second author is supported by the National Centre for Earth Observation (NCEO).

References

Agarwal V, Gribok AV, Abidi MA. 2007. Image restoration using L_1 norm penalty function. *Inverse Probl. Sci. Eng.* **15**(8): 785–809.

Daley R. 1991. *Atmospheric data assimilation (457 pp.)*. Cambridge: Cambridge University Press.

Dennis Jr JE, Schnabel RB. 1983. *Numerical methods for unconstrained optimization and nonlinear equations.* Prentice Hall Series in Computational Mathematics, Prentice Hall Inc.: Englewood Cliffs, NJ, ISBN 0-13-627216-9.

Engl H, Hanke M, Neubauer A. 1996. *Regularization of inverse problems*. Kluwer Academic Pub.

Fu H, Ng MK, Nikolova M, Barlow JL. 2006. Efficient minimization methods of mixed $l2$-$l1$ and $l1$-$l1$ norms for image restoration. *SIAM J. Sci. Comput.* **27**(6): 1881–1902 (electronic).

Griffith AK, Nichols NK. 2000. Adjoint methods in data assimilation for estimating model error. *Flow Turbul. Combust.* **65**(3-4): 469–488.

Hansen PC. 1998. *Rank-deficient and discrete ill-posed problems*. SIAM Monographs on Mathematical Modeling and Computation, Society for Industrial and Applied Mathematics (SIAM): Philadelphia, PA, ISBN 0-89871-403-6. Numerical aspects of linear inversion.

Hansen PC, Nagy JG, O'Leary DP. 2006. *Deblurring images*, *Fundamentals of Algorithms*, vol. 3. Society for Industrial and Applied Mathematics (SIAM): Philadelphia, PA, ISBN 978-0-898716-18-4; 0-89871-618-7. Matrices, spectra, and filtering.

Johnson C, Nichols NK, Hoskins BJ. 2005. Very large inverse problems in atmosphere and ocean modelling. *Internat. J. Numer. Methods Fluids* **47**(8-9): 759–771.

Lawless AS, Gratton S, Nichols NK. 2005. Approximate iterative methods for variational data assimilation. *Internat. J. Numer. Methods Fluids* **47**(10-11): 1129–1135, doi:10.1002/fld.851, URL http://dx.doi.org/10.1002/fld.851. 8th ICFD Conference on Numerical Methods for Fluid Dynamics. Part 2.

LeVeque R. 1992. *Numerical methods for conservation laws*. Birkh
"auser.

Lewis J, Lakshmivarahan S, Dhall S. 2006. *Dynamic data assimilation: a least squares approach*. Cambridge Univ Pr.

Lorenz EN. 1963. Deterministic nonperiodic flow. *Journal of the Atmospheric Sciences* **20**(2): 130–141.

Morton K, Mayers D. 2005. *Numerical solution of partial differential equations: an introduction*. Cambridge Univ Pr.

Nichols NK. 2010. Mathematical concepts of data assimilation. In: *Data Assimilation Making Sense of Observations*, Lahoz W, Khattatov B, Menard R (eds). Springer, pp. 1–20.

Sasaki Y. 1970. Some basic formalisms in numerical variational analysis. *Mon. Wea. Rev.* **98**: 875???883.

Schmidt M, Fung G, Rosales R. 2007. Fast optimization methods for l1 regularization: A comparative study and two new approaches. In: *ECML '07: Proceedings of the 18th European conference on Machine Learning*. Springer-Verlag: Berlin, Heidelberg, ISBN 978-3-540-74957-8, pp. 286–297.

Talagrand O. 1981. A study of the dynamics of four-dimensional data assimilation(initial conditions specification for numerical weather prediction). *Tellus* **33**: 43–60.