School of Mathematical
and Physical Sciences

University of Reading

Department of Mathematics and Statistics

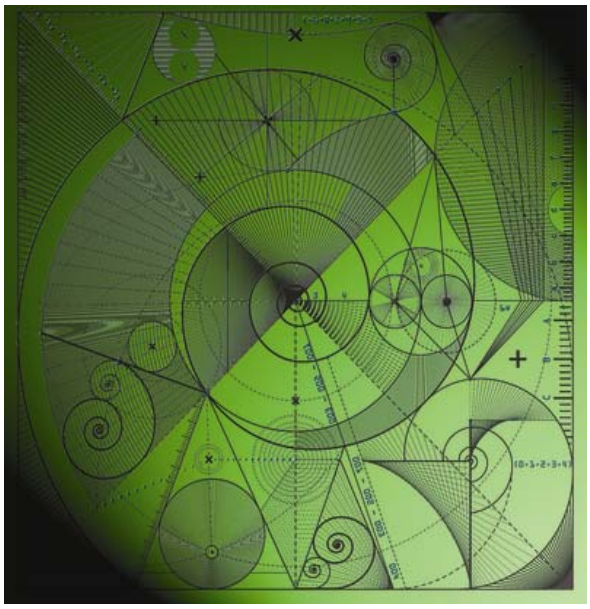
Preprint MPS-2011-13

21 October 2011

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



QUARTERLY JOURNAL OF THE ROYAL METEOROLOGICAL SOCIETY
Q. J. R. Meteorol. Soc. **00**: 1–14 (2010)
Published online in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/qj.000

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

M. A. Freitag[a*] , N. K. Nichols[b] and C. J. Budd[a]
[a]*Department of Mathematical Sciences, University of Bath, Claverton Down BA2 7AY, UK*
[b]*Department of Mathematics, The University of Reading, Box 220 Whiteknights, RG6 6AX, Reading, UK*

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 from stationary inverse problems to 4DVar a dynamical inverse problem and give examples for an L_1-norm penalty approach and a mixed Total Variation (TV) L_1-L_2-norm penalty approach. For problems with model error and where fronts are present the mixed TV L_1-L_2-norm penalty, which promotes sparsity, performs much better than the standard L_2-norm or L_1-norm regularisation in 4DVar.

KEY WORDS Data assimilation; Tikhonov regularisation; L_1 regularisation; TV regularisation, model error; linear advection equation

Received ; Accepted

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 initial conditions such that the forecast best fits both the observations and the background state (which is usually obtained from the previous forecast) within an interval called the assimilation window. Currently, in most operational 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 reviews 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 becomes edge-preserving as the process does not penalise the edges of the image. The L_1-norm penalty regularisation then recovers sharp edges in the image more precisely 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 that develop shocks, which is the case for moving weather fronts. In numerical weather prediction and ocean forecasting, it is recognized that the 4DVar assimilation method may not give a good analysis where there is a sharp gradient in the flow, such as a front (Bennett (2002); Lorenc (1981)). If the front is displaced in the background estimate, then the assimilation algorithm may smear the front and also underestimate the true amplitude of the shock (Johnson (2003)). In these cases the error covariances propagated implicitly by 4DVar are not representative of the correct error structures near the front. If model error is present, then there are systematic errors between the incorrect model trajectories and the observed data and therefore the strong constraint 4DVar, which assumes a perfect model, is not able to represent these errors correctly. Here we apply an L_1-norm penalty approach to several numerical examples containing sharp fronts for cases with model error. We show that the L_1-norm penalty approach applied to the gradient of the analysis vector (we call this mixed Total Variation (TV) L_1-L_2-norm penalty regularisation) performs better than the standard L_2-norm regularisation in 4DVar. With the use of the gradient operator and the L_1 norm, localisation of the gradient is enforced, which is important in tracking fronts. As an example we use the linear advection equation where sharp fronts and shocks are present. We use a numerical scheme that introduces some form of *model error* into the systems and find that, using an L_1-norm regularisation term, applied to the gradient of the

*Correspondence to: Department of Mathematical Sciences, University of Bath, Claverton Down BA2 7AY, UK. E-mail: m.freitag@maths.bath.ac.uk

solution, fronts are resolved more accurately than with the standard L_2-norm regularisation of 4DVar.

The aim of this paper is to examime the potential benefits of using L_1-norm regularisation in variational data assimilation. It presents a preliminary study showing that the method has potential to give improvement over existing approaches. Further investigation remains to be done in order to evaluate the technique in an operational setting.

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 explain how we solve the L_1-norm regularisation problem and the mixed TV L_1-L_2-norm regularisation problem. In Section 5 we state the model equations. Section 6 presents 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 6). We conclude with a section on future work.

2 4DVar and its relation to Tikhonov regularisation

In nonlinear 4DVar we aim to minimise the objective function

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

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 $x_i \in \mathbb{R}^m$, $i = 0, \ldots, N$ are the state vectors at time t_i. The function $\mathcal{M}_{i+1,i} : \mathbb{R}^m \to \mathbb{R}^m$ denotes the nonlinear model that evolves the state vector x_i at time t_i to the state vector x_{i+1} at time t_{i+1}. In weather forecasting the state vector $x_0^b \in \mathbb{R}^m$ is the best estimate from the previous assimilation cycle of the state of the system at the start of the window. The vectors $y_i \in \mathbb{R}^p$, $i = 1, \ldots, N$ contain the observations at times t_i and $H_i : \mathbb{R}^m \to \mathbb{R}^p$ is the observation operator that maps the model state space to the observation space.

Minimising (1) is a weighted nonlinear least-squares problem. By minimising $\mathcal{J}(x_0)$ we find an initial state $x_0 \in \mathbb{R}^m$, known as the *analysis*, such that the model trajectory is close to the background trajectory and to the observations in a suitable norm. The symmetric matrix $B \in \mathbb{R}^{m,m}$ and the symmetric matrices $R_i \in \mathbb{R}^{p,p}$, $i = 1, \ldots, N$ are assumed to represent the covariance matrices of the errors in the background and the observations respectively. The matrices R_i describe the combined effects of measurement errors, representativity errors (arising from the need to interpolate state vectors to the times and locations of the observations) and errors in the observation operator. Provided the background and observation errors have Gaussian distributions with mean zero, then minimising $\mathcal{J}(x_0)$ is equivalent to finding the *maximum a posteriori Bayesian estimate* of the true initial condition (Lorenc (1986)).

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

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

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

$$M_{i,0}(x_0) = M_{i,i-1}(x_{i-1})M_{i-1,i-2}(x_{i-2})\cdots M_{1,0}(x_0). \tag{5}$$

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

$$\widetilde{\nabla\nabla}\mathcal{J}(x_0) = B^{-1} + \sum_{i=1}^{N} M_{i,0}(x_0)^T H^T R^{-1} H M_{i,0}(x_0). \tag{6}$$

Both the summation terms in (4) and (6) can be obtained recursively using the adjoint equations

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

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

$$\nabla\mathcal{J}(x_0) = B^{-1}(x_0 - x_0^b) - \lambda_0, \tag{7}$$

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

$$\widetilde{\nabla\nabla}\mathcal{J}(x_0) = B^{-1} - \nabla\lambda_0. \tag{8}$$

Using these adjoint equations we avoid having to compute $M_{i,i-1}(x_{i-1})$ several times. We note that λ_i, $i = 0, \ldots, N$

DOI: 10.1002/qj

are vectors whereas $\nabla\lambda_i, i = 0, \ldots, N$ are square matrices of the dimension of the system state.

The approximate Hessian $\widetilde{\nabla}\widetilde{\nabla}\mathcal{J}(x_0)$ and $\nabla\mathcal{J}(x_0)$ are then used in (3), which is equivalent to a linearised least square problem. Here we solve this system directly. This approach is mathematically equivalent to the incremental 4DVar method as described in (Lawless *et al.* (2005a,b)); in the incremental method, however, the inner equations (3) are solved iteratively.

We may rewrite the objective function (1) in 4DVar as

$$\begin{aligned}\mathcal{J}(x_0) = &\frac{1}{2}(x_0 - x_0^b)^T B^{-1}(x_0 - x_0^b)\\ &+\frac{1}{2}(\hat{y} - \hat{\mathcal{H}}(x_0))^T \hat{R}^{-1}(\hat{y} - \hat{\mathcal{H}}(x_0)),\end{aligned} \tag{9}$$

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, $\hat{y} \in \mathbb{R}^{pN}$ is a vector and $\hat{R} \in \mathbb{R}^{pN,pN}$ is a block diagonal matrix with diagonal blocks equal to R. If we linearise $\mathcal{M}_{i,0}$ about x_0^b, then the Jacobian of the augmented matrix $\hat{\mathcal{H}}$ is given by

$$\hat{H} := \hat{H}(x_0^b) = \begin{bmatrix} HM_{1,0}(x_0^b) \\ HM_{2,0}(x_0^b) \\ \vdots \\ HM_{N,0}(x_0^b) \end{bmatrix}, \tag{10}$$

which 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 may write the linearised objective function that we aim to minimise as

$$\begin{aligned}\hat{J}(z) =&\|C_R^{-1/2}(\hat{y} - \hat{\mathcal{H}}(x_0^b)) - C_R^{-1/2}\hat{H}C_B^{1/2}z\|_2^2\\ &+ \mu^2\|z\|_2^2, \quad \mu^2 = \frac{\sigma_o^2}{\sigma_b^2}.\end{aligned} \tag{11}$$

This is equivalent to a linear least-squares problem with Tikhonov regularisation (Engl *et al.* (1996)), where μ^2 acts as the regularisation parameter. If we set

$$G := C_R^{-1/2}\hat{H}C_B^{1/2} \quad \text{and} \quad f := C_R^{-1/2}(\hat{y} - \hat{\mathcal{H}}(x_0^b)), \tag{12}$$

where $G \in \mathbb{R}^{pN,m}$ and $f \in \mathbb{R}^{pN}$, then equation (11) may be written as

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

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

$$\min_z\{\|f - Gz\|_2^2\} \tag{14}$$

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 observations are subject to errors, leading to errors in the vector f. Hence, we can see that the minimisation problem (14) 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 objective function $\hat{J}_2(z)$ in (13). The minimisation of the Tikhonov function (13) gives the regularised solution

$$z = (G^TG + \mu^2 I)^{-1}G^T f = \sum_{j=1}^{\min(pN,m)} \frac{\sigma_j^2}{\sigma_j^2 + \mu^2}\frac{u_j^T f}{\sigma_j}v_j, \tag{15}$$

(see, for example (Hansen *et al.*, 2006, Chapter 5) for details). 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 $U \in \mathbb{R}^{pN,pN}$ and $V \in \mathbb{R}^{m,m}$ orthonormal matrices and Σ is a $pN \times m$ matrix whith entries σ_j, $j = 1, \ldots, \min(pN, m) \geq 0$ on the leading diagonal and zeros elsewhere. Hence the factor $\sigma_j^2/(\sigma_j^2 + \mu^2)$ acts as a filter factor 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. The reason for the edge-preserving property of the L_1-norm is that the L_1-norm enforces a sparse solution (Donoho (2006)). 4DVar performs poorly for the recovery of fronts. For shocks and fronts the gradient of the solution is sparse and hence we introduce a mixed Total Variation L_1-L_2-norm approach which aims to recover fronts.

Hence we introduce and test two new approaches which are motivated by the L_1-norm regularisation and compare them to standard 4DVar: These are L_1-norm regularisation and a mixed Total Variation L_1-L_2-norm regularisation. Both are described in the next section.

3 L_1-norm and mixed L_1-L_2-norm regularisation

With the notation in (12), the minimisation problem in (11) can be written as (13) - known as standard Tikhonov regularisation - where the second term is a regularisation term and 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).

DOI: 10.1002/qj

Firstly, in this paper we consider the effects of L_1-norm regularisation for variational data assimilation by replacing the squared L_2-norm in the regularisation term $\mu^2\|z\|_2^2$ of (13) by the L_1-norm to obtain

$$\min_z \hat{J}_1(z) = \min_z\{\|f - Gz\|_2^2 + \mu^2\|z\|_1\},\ \mu^2 = \frac{\sigma_o^2}{\sigma_b^2}. \tag{16}$$

Equation (13) can be written as

$$\min_z \hat{J}_2(z) = \min_z \left\{ \left\| \begin{bmatrix} f \\ 0 \end{bmatrix} - \begin{bmatrix} G \\ \mu I \end{bmatrix} z \right\|_2^2 \right\},\ \mu^2 = \frac{\sigma_o^2}{\sigma_b^2}. \tag{17}$$

The minimisation problems (16) and (17) aim to produce a solution z and hence, with $z := C_B^{-1/2}(x_0 - x_0^b)$, an initial state $x_0 = C_B^{1/2}z + x_0^b$ such that the solution trajectory is both close to the background (the previous forecast) and the observations in some weighted norm. The solution to problems (16) promotes sparsity in the solution, hence it promotes a sparse vector z. We will see that this is not so useful for the computations.

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 Laplace (double-sided exponential) distribution. (For details, see the Appendix.)

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.

However, if it is known that fronts are present in the solution then the gradient of the solution will be sparse - hence the gradient of the initial state x_0 will be sparse. If we approximate the gradient by a matrix D given by

$$D = \begin{bmatrix} 1 & 0 & \dots & & \\ -1 & 1 & 0 & \dots & \\ 0 & -1 & 1 & 0 & \dots \\ & \ddots & \ddots & \ddots & \\ & \dots & 0 & -1 & 1 \end{bmatrix}, \tag{18}$$

then the minimisation problem for a sparse initial state and hence a sharp front becomes

$$\min_z \hat{J}_{TV}(z) = \min_z \left\{ \left\| \begin{bmatrix} f \\ 0 \end{bmatrix} - \begin{bmatrix} G \\ \mu I \end{bmatrix} z \right\|_2^2 + \delta\|Dx_0\|_1 \right\},\ \mu^2 = \frac{\sigma_o^2}{\sigma_b^2}, \tag{19}$$

where $x_0 = C_B^{1/2}z + x_0^b$, D is given by (18) and δ is another so-called regularisation parameter which needs to be chosen. The size of δ determines how much sparsity is enforced on the gradient of the solution (see Table I for examples with different choices of δ). We will see in Section 6 that minimising $\hat{J}_{TV}(z)$ in (19) gives a much better resolution of the fronts than minimising $\hat{J}_2(z)$ or $\hat{J}_1(z)$ in (17) or (16).

We find that, for fronts and shocks, regularisation with an added L_1-norm on the derivative of the initial condition in 4DVar gives much better results than the standard L_2-norm approach in the presence of model error. When an L_1-norm penalty term with a gradient as in (19) is added one often speaks of total variation (TV) regularisation (Strong and Chan (2003)). We call the problem in (19) mixed TV L_1-L_2-norm regularisation problem.

In the following section we explain how we solve the L_1-norm minimisation problem in (16) and the mixed TV L_1-L_2-norm minimisation problem in (19).

4 Least mixed norm solutions

Consider the minimisation problems (16) and (19). In order to solve these least mixed norm solutions we use an approach introduced by (Fu *et al.* (2006)). Both problems (16) and (19) are solved in a similar way. We explain the algorithm using the minimisation problem (19), the application of the algorithm to problem (16) is similar.

First, with $x_0 = C_B^{1/2}z + x_0^b$, problem (19) can be formulated as

$$\min_z \left\{ \left\| \begin{bmatrix} f \\ 0 \end{bmatrix} - \begin{bmatrix} G \\ \mu I \end{bmatrix} z \right\|_2^2 + \delta\|D(C_B^{1/2}z + x_0^b)\|_1 \right\}. \tag{20}$$

We let

$$v = \delta D(C_B^{1/2}z + x_0^b),$$

and split v into its non-negative and non-positive parts v^+ and v^-, that is

$$v = v^+ - v^-$$

and

$$v^+ = \max(v, 0), \qquad v^- = \max(-v, 0).$$

Problem (20) can then be written as

$$\min_{z,v^+,v^-} \left\{ \left\| \begin{bmatrix} f \\ 0 \end{bmatrix} - \begin{bmatrix} G \\ \mu I \end{bmatrix} z \right\|_2^2 + 1^T v^+ + 1^T v^- \right\}. \tag{21}$$

subject to the constraints

$$\delta D(C_B^{1/2}z + x_0^b) = v^+ - v^-, \tag{22}$$

$$v^+, v^- \geq 0. \tag{23}$$

Here 1 denotes the vector of all ones of appropriate size. This problem can then be written as

$$\min_w \left\{ \frac{1}{2} w^T H w + c^T w \right\} \tag{24}$$

subject to

$$Ew = g \quad \text{and} \quad Fw \geq 0, \tag{25}$$

where

$$w = \begin{bmatrix} z \\ v^+ \\ v^- \end{bmatrix}, \quad H = \begin{bmatrix} 2(G^TG + \mu^2 I) & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix},$$

$$c = \begin{bmatrix} -2G^T f \\ 1 \\ 1 \end{bmatrix}, \quad E = \begin{bmatrix} \delta D C_B^{1/2} & -I & I \end{bmatrix},$$

$$F = \begin{bmatrix} 0 & 0 & 0 \\ 0 & -I & 0 \\ 0 & 0 & I \end{bmatrix}, \quad g = -\delta D x_0^b,$$

and the block matrices I and 0 as well as the vectors 1 of all ones in the matrices H, E, F and c are of appropriate size. The objective function in (24) is convex as H is symmetric positive semi-definite. In order to solve the quadratic programming problem (24) with constraints (25) we use the MATLAB in-built function `quadprog.m`.

In the following section we consider a square wave advected using the linear advection equation as an example. We use a 'true' model (from which we take the observations) 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 observe that the new edge-preserving mixed TV L_1-L_2-norm regularisation indeed gives better results than the standard L_2-norm approach and the simple L_1-norm regularisation.

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 μ.

5 Models

In this section we consider the problem

$$u_t + [f(u)]_x = 0, \tag{26}$$

where $f(u)$ is given by

$$f(u) = u, \tag{27}$$

for the linear advection equation.

This general problem can be discretised using the upwind scheme

$$U_j^{n+1} = U_j^n - \frac{\Delta t}{\Delta x}\left(f(U_j^n) - f(U_{j-1}^n)\right). \tag{28}$$

All equations are valid for $j = 1, \ldots, N$, where f is given by (27). The CFL condition

$$\left|\frac{\max(f'(u))\Delta t}{\Delta x}\right| \leq 1 \tag{29}$$

needs to be satisfied for stability (Morton and Mayers (2005); LeVeque (1992)). For the linear advection equation (27) this condition just reduces to $\Delta t < \Delta x$. For more details on the above methods we refer to LeVeque (1992).

6 Linear advection equation

Consider the linear advection equation

$$u_t + u_x = 0, \tag{30}$$

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

$$u(x, 0) = \begin{cases} 0.5, & 0.25 < x < 0.5 \\ -0.5, & x < 0.25 \quad \text{or} \quad x > 0.5. \end{cases} \tag{31}$$

This wave moves through the time interval; the true solution is obtained by the method of characteristics (by advecting the inital condition at speed 1, that is $u(x, t) = u(x - t, 0)$) and the model equations are defined by the upwind scheme (28) with boundary conditions $U_0^n = U_N^n$, where $n = 1, \ldots, 80$, $\Delta x = \frac{1}{100}$ and n is the number of time steps. The same example is used in Griffith and Nichols (2000). For this example we take $\Delta t = 0.005$.

6.1 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 $B = 0.01I$ and $R = 0.01I$; hence we put equal emphasis on the observations and the background. Moreover, for the background we choose U_b^0 to be equal to the truth perturbed by Gaussian noise with mean zero and covariance B. The background thus contains errors with variance of order 0.01. 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

- standard 4DVar (minimisation problem (17)),
- L_1-norm regularisation (minimisation problem (16)), and
- mixed TV L_1-L_2-norm regularisation (minimisation problem (19)).

Figures 1 - 9 show the results for this example where the linear advection equation is used as a model.

In the plots the true solution is represented by a thick dot-dashed line (called 'Truth' in the legend). This true solution is unknown in practice. We take (noisy) observations from that true trajectory. The model solution (which is derived from the upwind method) is shown as a dashed line (called 'Imperfect model' in the legend). This solution represents the model solution, that is the solution that

DOI: 10.1002/qj

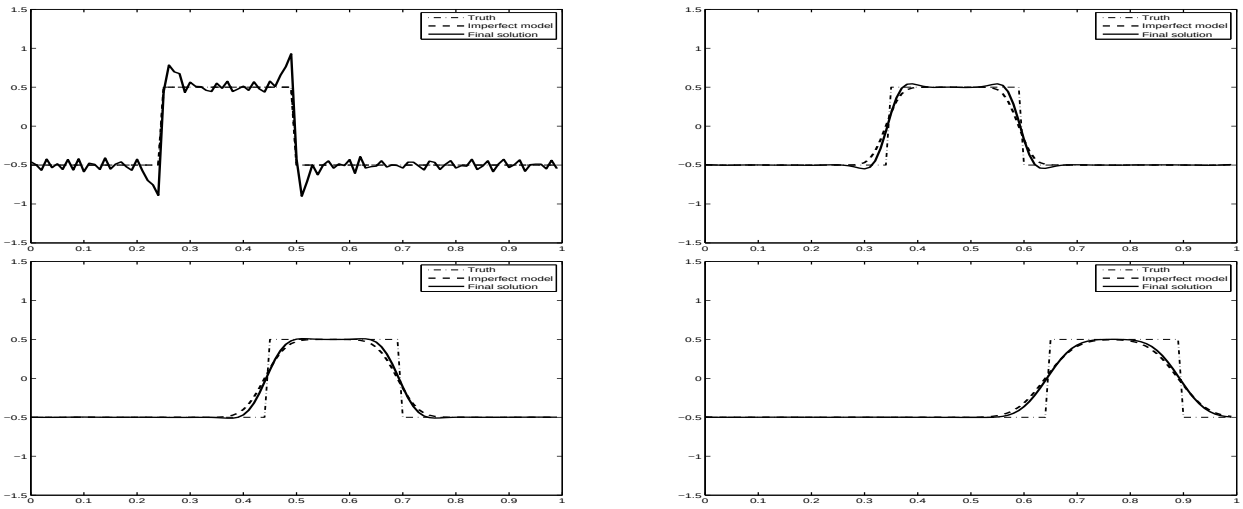


Figure 1. Results for **4DVar** applied to the linear advection equation where the initial condition is a square wave. We take **perfect observations at each point in time and space** over the assimilation interval which is 40 time steps. The four plots show the initial conditions at $t = 0$ and the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

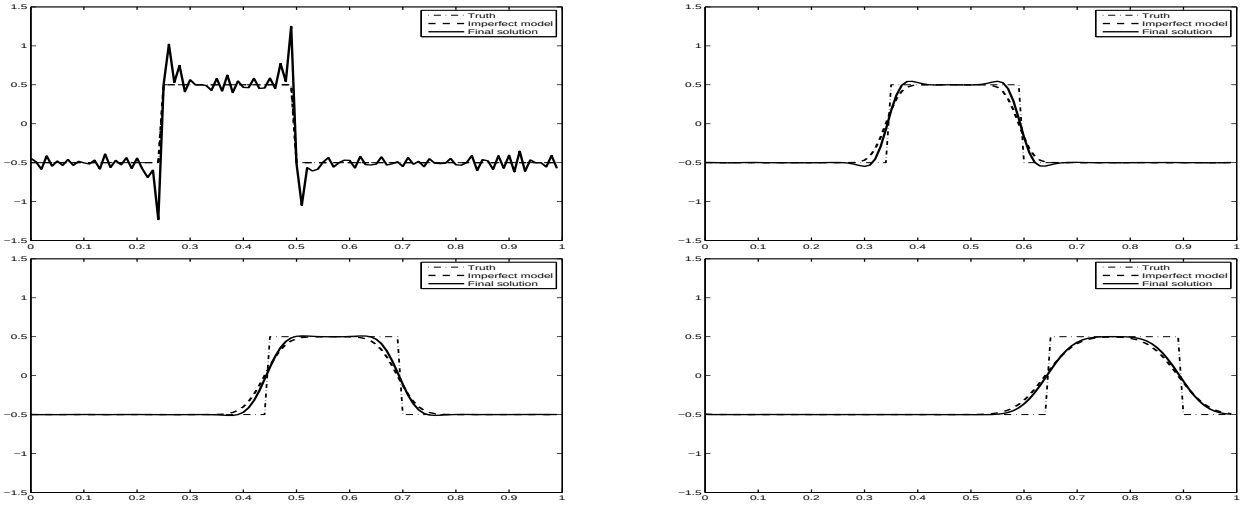


Figure 2. Results for L_1**-regularisation** for the same data as in Figure 1.

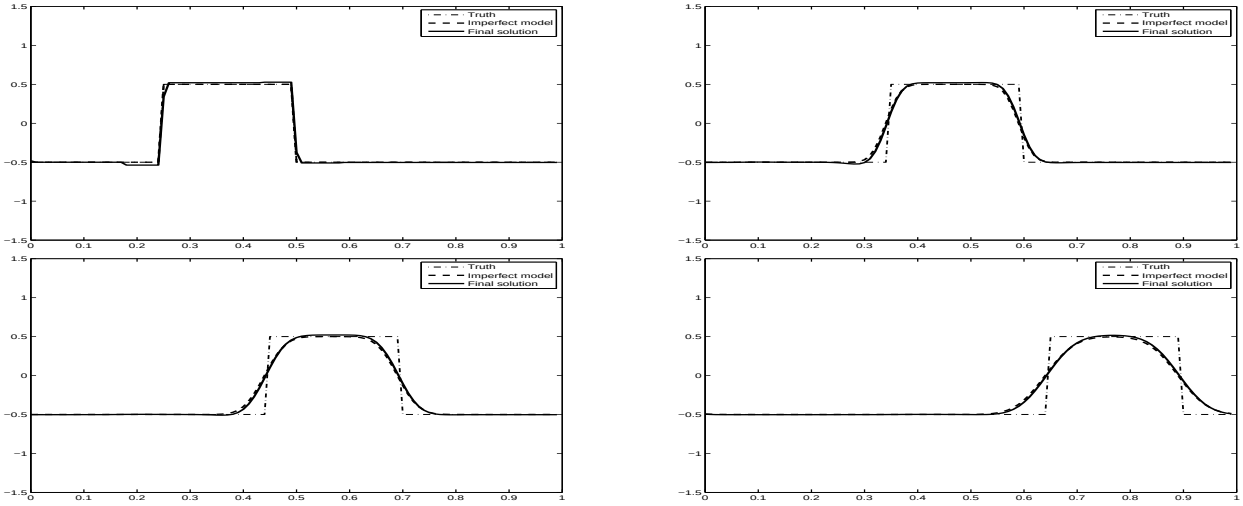


Figure 3. Results for **mixed TV** L_1-L_2**-norm regularisation** for the same data as in Figure 1. Mixed TV L_1-L_2-norm regularisation gives the best possible result for the initial condition.

DOI: 10.1002/qj

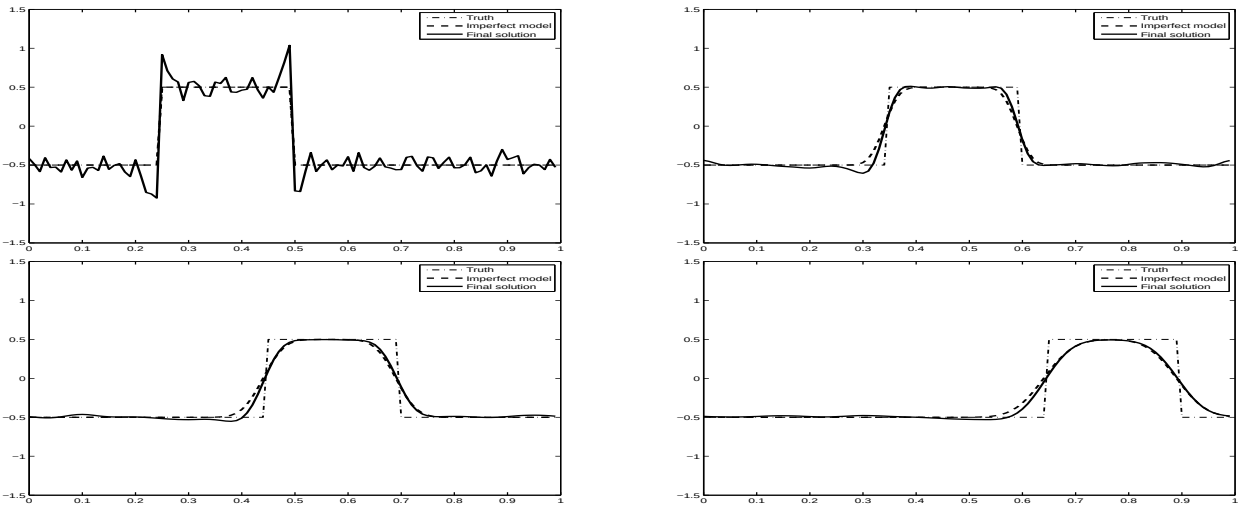


Figure 4. Results for **4DVar** for the same data as in Figure 1 but with **perfect observations every** 20 **points in space and every** 2 **time steps**. 4DVar leads to oscillations in the initial condition.

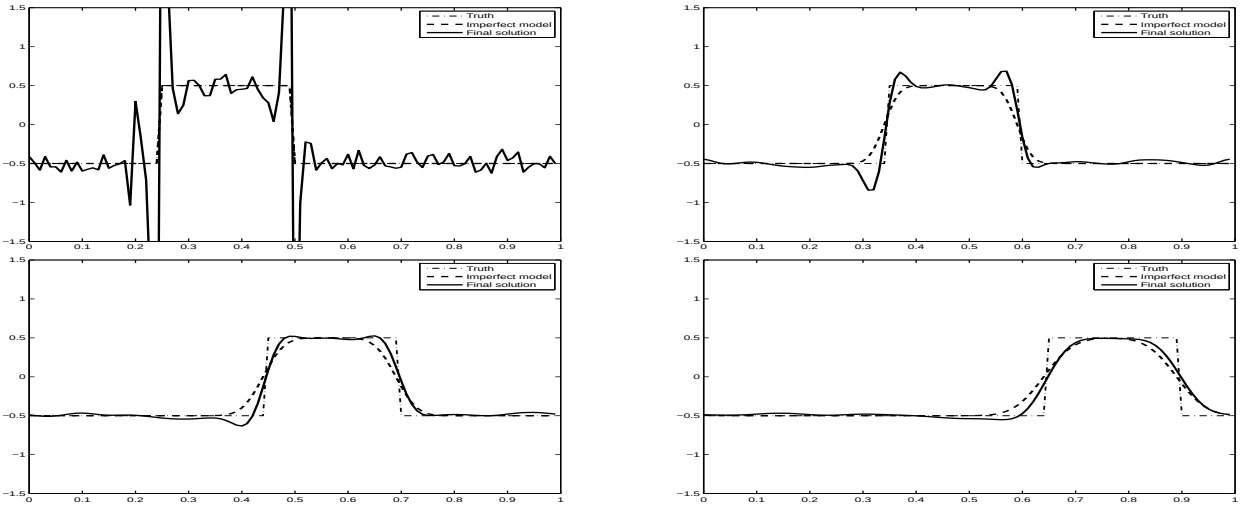


Figure 5. Results for L_1**-regularisation** for the same data as in Figure 4.

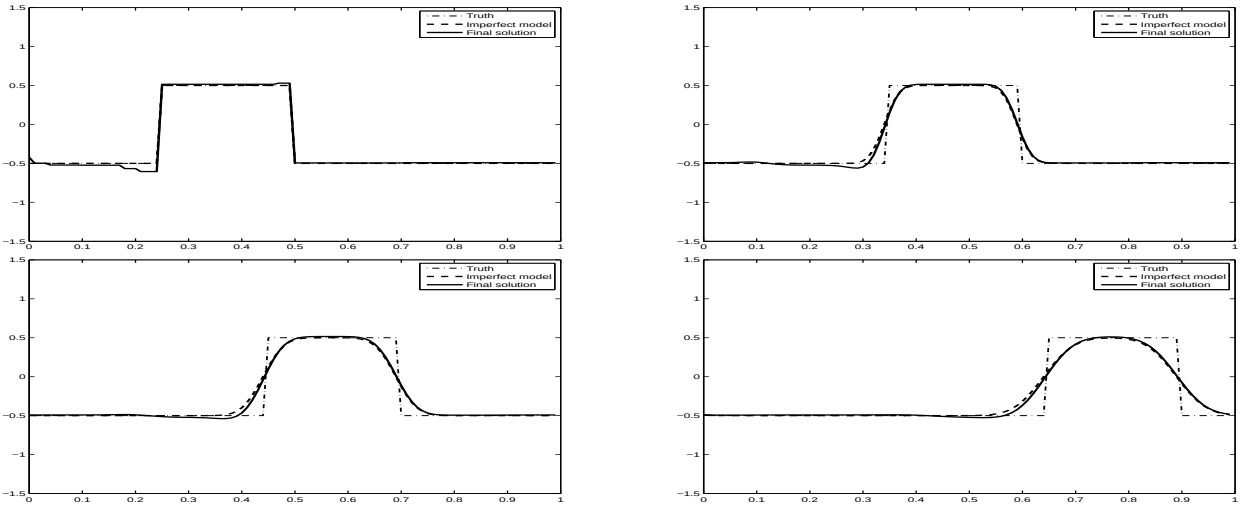


Figure 6. Results for **mixed TV** L_1-L_2**-norm regularisation** for the same data as in Figure 4. Mixed TV L_1-L_2-norm regularisation gives the best possible result for the initial condition.

DOI: 10.1002/qj

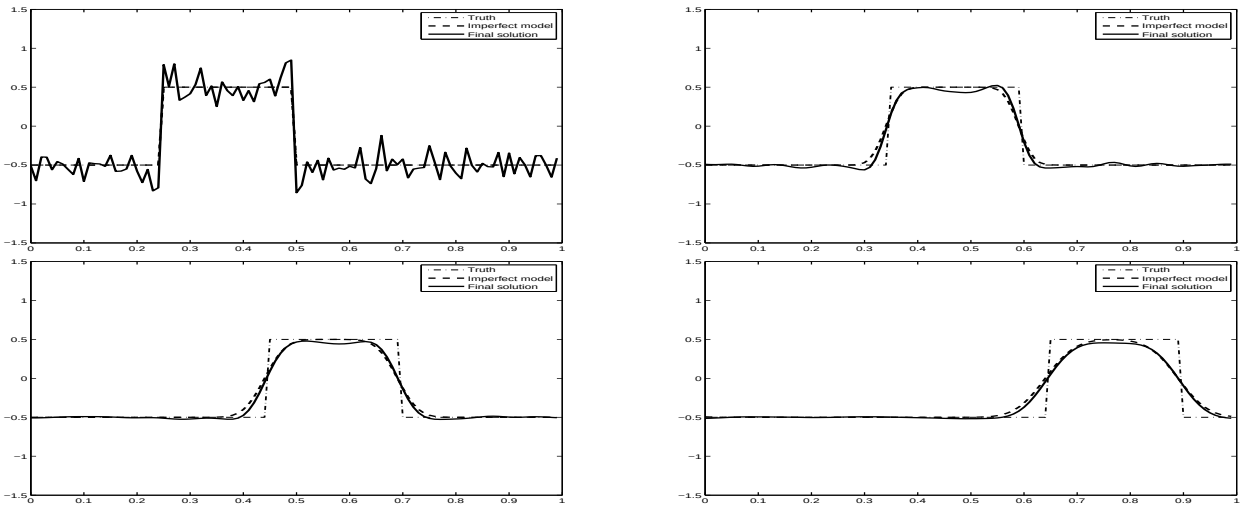


Figure 7. Results for **4DVar** for the same data as in Figure 1 but with **imperfect observations every** 20 **points in space and every** 2 **time steps**. 4DVar leads to bad oscillations in the initial condition and also to a misplaced discontinuity in the forecast.

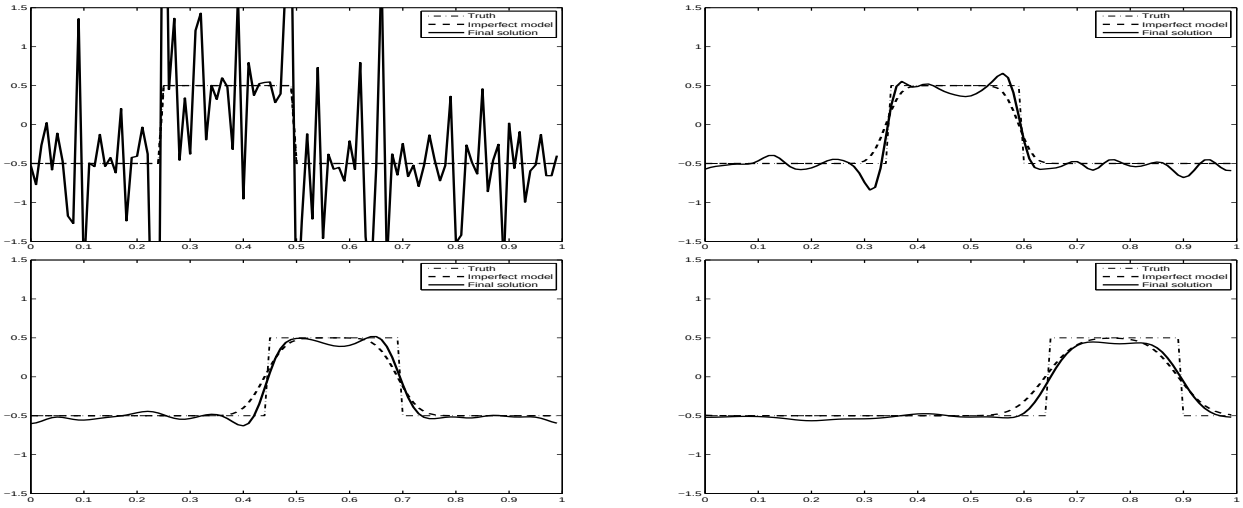


Figure 8. Results for L_1 **regularisation** for the same data as in Figure 7.

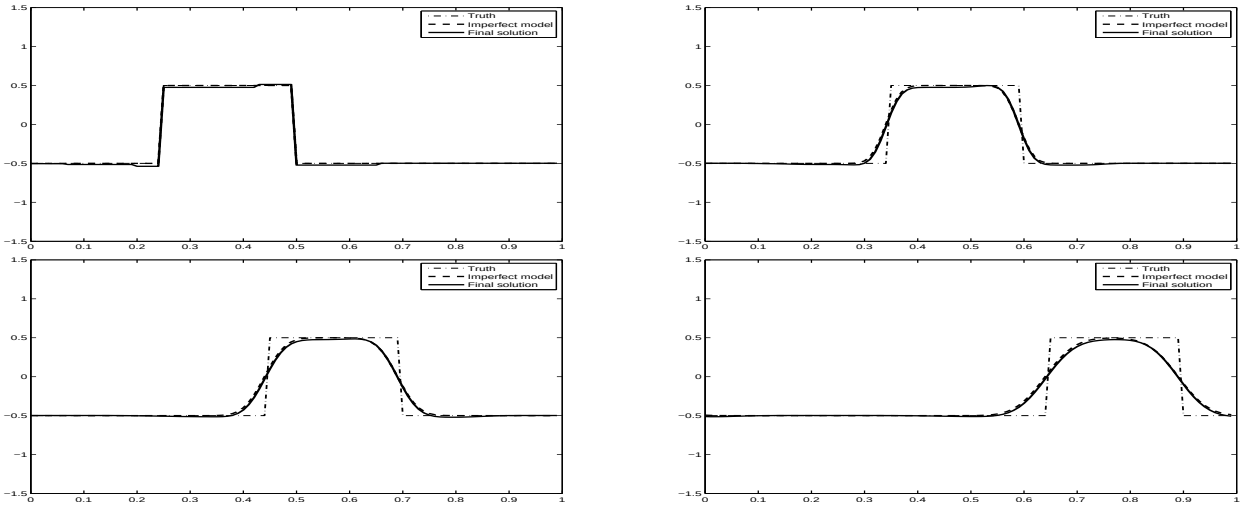


Figure 9. Results for **mixed TV** L_1**-**L_2**-norm regularisation** for the same data as in Figure 7. Mixed TV L_1-L_2-norm regularisation gives the best possible result for the initial condition.

DOI: 10.1002/qj

is obtained if we use the correct initial conditions and the (imperfect) model. It represents the best solution that we are able to achieve (if data assimilation gives us the perfect initial condition), as the model error is always present. The solution obtained from the assimilation process by incorporating the (perfect/partial/noisy) 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 (minimisation problem (17)), that for L_1-regularisation in Figure 2 (minimisation problem (16)) and that for mixed TV L_1-L_2-norm regularisation in Figure 3. The analysis obtained by 4DVar and L_1-regularisation is very inaccurate, with many oscillations and large over/undershoots near the discontinuities (first plots in Figures 1 and 2). When L_1-norm regularisation with the gradient is used, the initial condition more accurate (first plot in Figure 3). The same result is true for partial observations (Figures 4 and 5 for 4DVar and L_1-regularisation versus Figure 6 for mixed TV L_1-L_2-norm regularisation) and for imperfect partial observations (Figures 7 and 8 for 4DVar and L_1-norm regularisation versus Figure 9 for mixed TV L_1-L_2-norm regularisation). The second row $B = 0.01I$ of Table I quantifies the errors in the initial conditions for this situation for 4DVar, L_1-norm regularisation and the L_1-norm total variation approach. We see that for all types of observations we investigated (partial, full, perfect and noisy observations), L_1-norm TV regularisation gives the smallest initial condition error.

Traditional strong constraint 4DVar does not take model error into account. Hence 4DVar's attempts to compensate for the initial condition error are obstructed by the use of an imperfect forecast model and it therefore does not produce an accurate estimate of the truth at the initial time. The errors in the initial state estimated by 4DVar act to force the trajectory propagated by the incorrect model to match the observed data from the true model and hence act to compensate, on average, for the model error. From the final plots in Figures 4 and 7 for 4DVar we also see that the forecast is inaccurate due to the incorrect estimate produced at the end of the assimilation window. We also observe that the forecast in 4DVar leads to a slight phase shift and the wrong amplitude in the forecast, as well as overshooting and undershooting. If noisy observations are taken (see first plot in Figure 4 vs first plot in Figure 7), the oscillations in the initial condition are more frequent. For mixed TV L_1-L_2-norm regularisation (Figures 6 and 9) these problems do not occur. We see in the first plot of Figures 6 and 9 that the initial condition obtained from mixed TV L_1-L_2-norm regularisation is the most accurate and hence the best possible forecast (see final plots of Figures 6 and 9) is obtained (subject to model error). This behaviour is due to the property of mixed TV L_1-L_2-norm regularisation enforcing sparsity on the gradient of the solution.

In the next two subsections we change the experimental design of the problem slightly, in order to check the robustness of regularisations. We first check a more realistic background error covariance matrix in Section 6.2 and then investigate a change in the size of the assimilation window in Section 6.3.

6.2 Changing the background error covariance matrix

We take precisely the same experiment as in the previous Subsection 6.1; however, we change the background error covariance matrix from the identity matrix to a Gaussian covariance matrix B with entries

$$B_{ij} = \sigma_b^2 e^{-\frac{|i-j|}{2L^2}}, \quad \text{where} \quad L = 5, \tag{32}$$

and $\sigma_b^2 = 0.01$. Hence B is a symmetric matrix with diagonal entries equal to 0.01 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 0.01. Note that for this matrix the inverse is a tridiagonal matrix. For the background we choose Gaussian noise with covariance B and a mean value which is given by the truth. These errors are consistent with the choice of B.
We only present the results for imperfect and partial observations, as this represents the most realistic case; similar results are achieved in the cases of perfect observations and partial observations without noise. Further cases are summarised in Table I in Subsection 6.4. We also do not present the results for L_1 norm regularisation here as we have seen in Subsection 6.1 that this approach is not better than standard 4DVar. The more interesting case is the mixed TV L_1-L_2-norm regularisation.

Figures 10 and 11 show the results where the background error covariance matrix B is given by (32). For this choice of B, the results for 4DVar (Figure 10) are better than the results for the diagonal matrix B (Figure 7) because information is spread via the B matrix, and we see that the oscillations in the analysis are significantly reduced. However, mixed TV L_1-L_2-norm regularisation (Figure 11) still behaves consistently better than standard L_2-norm regularisation (Figure 10). In particular, the shape of the wave is distorted and there are small undershoots and overshoots in the 4DVar analysis (first plot in Figure 10), which lead to small errors and the wrong amplitude in the forecast (final plot in Figure 10). For the analysis using mixed TV L_1-L_2-norm regularisation, the inital condition (first plot in Figure 11) shows a smaller error than the initial condition in standard 4DVar (first plot in Figure 10) and the forecast is slightly better than the forecast in 4DVar (final plot in Figure 11). The quantities of the errors in the initial conditions for this particular case are summarised in the fifth row of Table I where we see that the errors using mixed TV L_1-L_2-norm regularisation are the smallest.

6.3 Changing the length of the assimilation window

Again, we take the same experimental data as in Subsection 6.1; 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

DOI: 10.1002/qj

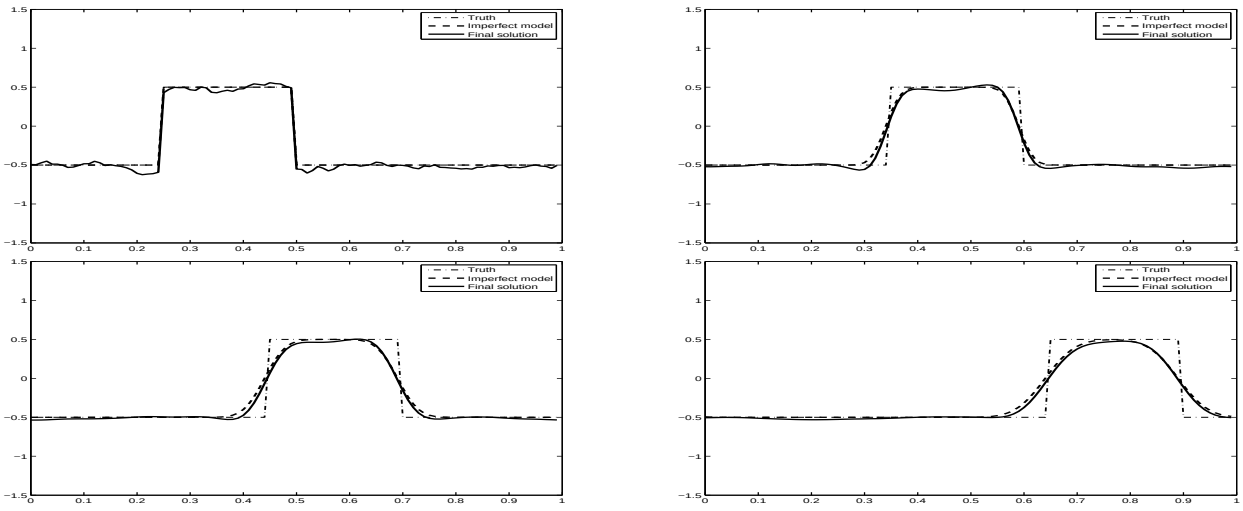


Figure 10. Results for **4DVar** for the same data as in Figure 7, but for B with $B_{ij} = 0.01\, e^{-\frac{|i-j|}{2L^2}}$, where $L = 5$.

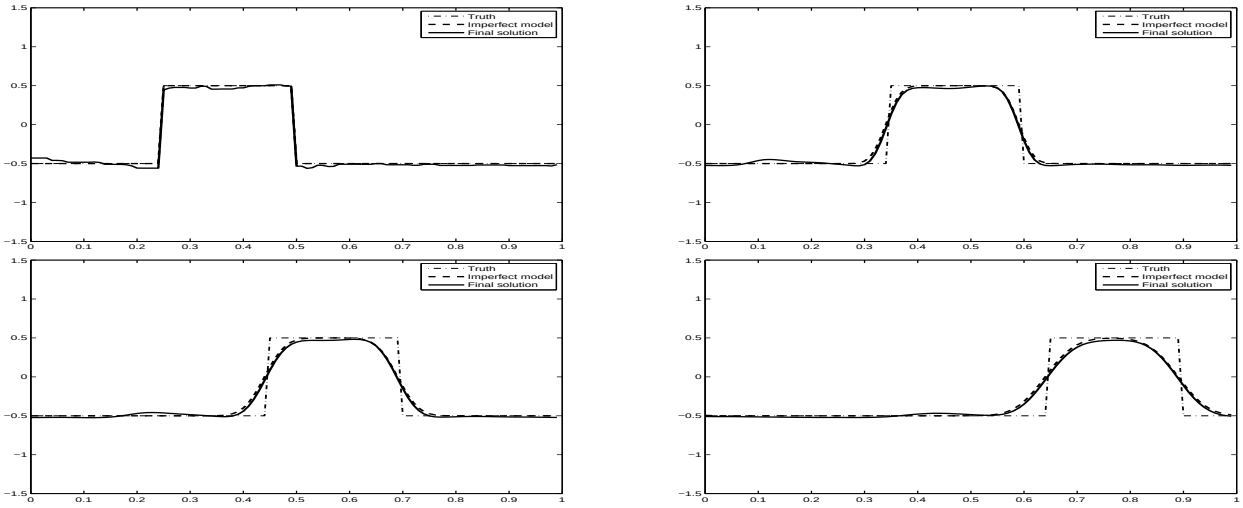


Figure 11. Results for **mixed TV L_1-L_2-norm regularisation** for the same data as in Figure 10.

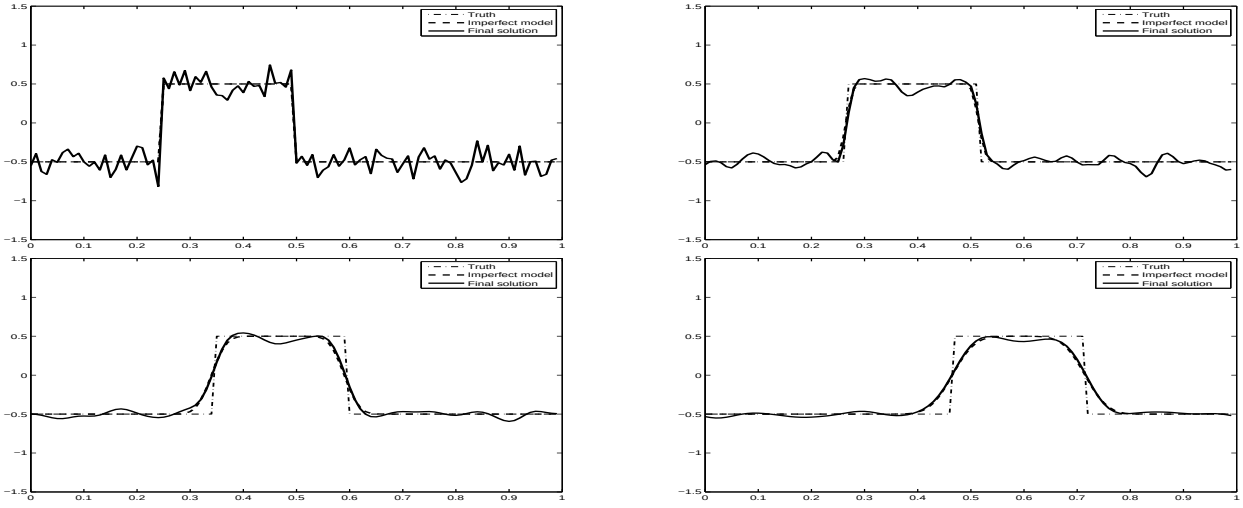


Figure 12. Results for **4DVar** applied to the linear advection equation where the initial condition is a square wave. We take **imperfect observations every** 5 **points in space and every** 2 **time steps** over the assimilation interval which is 5 time steps. The four plots show the initial conditions 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.

DOI: 10.1002/qj

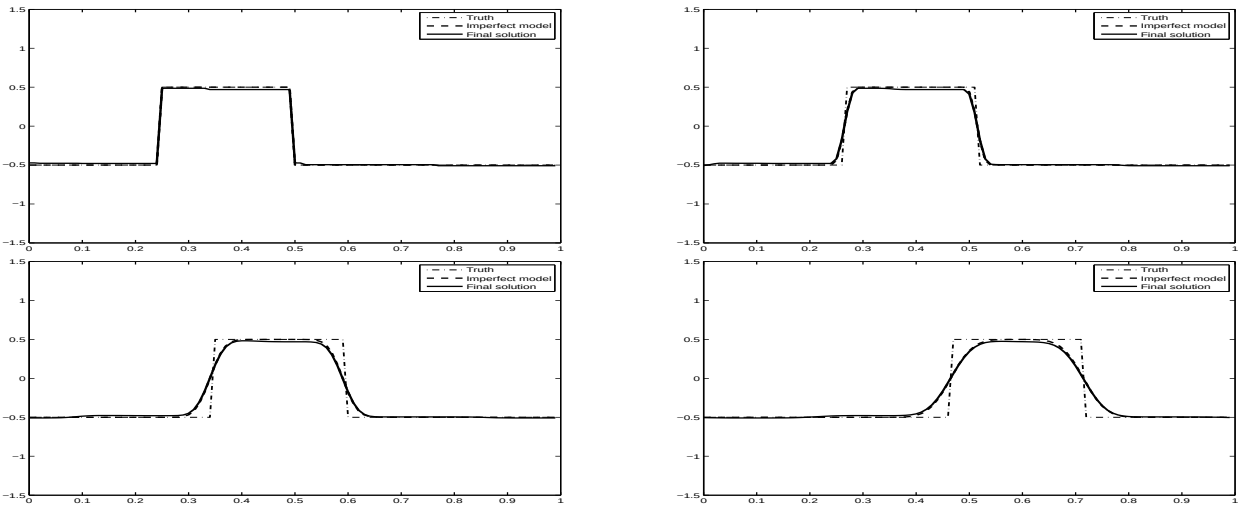


Figure 13. Results for **mixed TV L_1-L_2-norm regularisation** for the same data as in Figure 12. Mixed TV L_1-L_2-norm regularisation gives the best possible result for the initial condition.

Gaussian noise of mean zero and variance 0.01. For the background we again take the truth perturbed by Gaussian noise with covariance $B = 0.01I$.

Figures 12 and 13 show the results for a reduced size of the assimilation window. The first observation that we can make is that again the regularisation using the mixed TV L_1-L_2-norm (Figure 13) is consistently better than that using the L_2-norm (Figure 12). Standard 4DVar produces oscillations, in particular in the initial conditions, whereas the mixed TV L_1-L_2-norm regularisation does not show any oscillations. The 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 12). Again, for 4DVar, the forecast of the analysis does not keep the amplitude correctly (final plot in Figure 12), whereas L_1-norm regularisation provides a more accurate amplitude in the forecast (final plot in Figure 13), a property of the underlying imperfect model.

6.4 Summary of initial condition errors

In Table I we summarise the analysis errors (the errors between the analysis and the truth at $t = 0$), measured in the L_2 vector norm, for several scenarios. We choose observation errors with covariance $R = 0.01I$ and assimilation windows of length 40. The general experimental design is as in Section 6.1. We consider two types of covariance matrices for the background error, namely $B = \sigma_b^2 I$, and the double-sided exponential covariance matrix B given by (32). For both types of matrices we consider three different variances: $\sigma_b^2 = 1$, $\sigma_b^2 = 0.01$ and $\sigma_b^2 = 0.005$. For each matrix we use either perfect observations everywhere in time and space, perfect observations every 20 points in space and every 2 time steps or imperfect observations every 20 points in space and every 2 time steps, where the observations are taken as perturbations from the truth with Gaussian noise of mean zero and covariance B. Finally, the last three rows of Table I show the results for a smaller assimilation window of length 5.

We have also given results for different values of δ in (19). The emphasis on the sparsity of the gradient of the initial condition depends on this regularisation parameter. We have looked at three different values for δ and the best of all three results (that is the smallest error in the initial condition) is underlined in the table. The regularisation depends on the regularisation parameter but investigating the influence of this parameter and finding the optimal choice of δ is beyond the scope of this paper. We remark that for the plots in the previous subsections we used the value of δ which gave the smallest initial condition error.

We see from the entries in the table that the errors in the analysis at time $t = 0$ are consistently smaller for mixed TV L_1-L_2-norm regularisation than for standard 4DVar or L_1-norm regularisation. Mixed TV L_1-L_2-norm regularisation gives an error of about one magnitude smaller than for standard 4DVar. We also observe from the table that, for both standard 4DVar, L_1-norm regularisation and mixed TV L_1-L_2-norm regularization, the errors in the initial condition (analysis) decrease as the variance in the background error is reduced, that is, as the ratio of the background to observation variance decreases. This is consistent with the results of Haben *et al.* (2010), which show that the standard 4DVar assimilation problem becomes more well-conditioned (well-posed) as this ratio decreases. These examples demonstrate that, even where the noise in the background and observations is Gaussian with known covariances, the standard 4DVar approach does not produce as accurate an analysis as mixed TV L_1-L_2-norm regularisation in the presence of model error.

6.5 A shifted background

Finally, we consider the same problem as in Subsection 6.1 - with the same setup and error covariance matrices. However, here we shift the square wave in the background by 0.02 to the right, so that shock is displaced. The reason for this shift is a practical one; fronts are often resolved correctly in numerical weather forecasting, but

DOI: 10.1002/qj

Table I. Comparison between errors in the analysis in standard 4DVar, L_1-norm regularisation and mixed TV L_1-L_2-norm regularisation measured in the L_2-norm

		Standard 4DVar	L_1-norm regularisation	mixed TV L_1-L_2-norm regularisation		
				$\delta = 10$	$\delta = 100$	$\delta = 1000$
$B = I$	full perfect observations	2.3674	2.4392	1.1585	0.7674	0.2998
	partial perfect observations	12.8039	13.6598	9.3621	0.4643	2.7286
	partial imperfect observations	13.6182	14.4389	7.7128	0.4790	2.9110
$B = 0.01I$	full perfect observations	1.0609	1.4780	0.8963	0.6998	0.2531
	partial perfect observations	1.3791	10.0589	1.0935	0.2866	1.2440
	partial imperfect observations	1.4614	9.9083	1.0060	0.1719	1.3910
$B = 0.005I$	full perfect observations	0.9012	1.4567	07987	0.6417	0.2272
	partial perfect observations	0.8651	9.3547	0.6887	0.2260	0.8014
	partial imperfect observations	0.8979	8.5296	0.6566	0.1500	0.9141
B with entries $B_{ij} = e^{-\frac{\lvert i-j\rvert}{2L^2}}$ where $L = 5$	full perfect observations	1.1892	1.3703	0.9801	0.7391	0.2807
	partial perfect observations	2.7845	11.6647	2.2421	0.3832	2.7031
	partial imperfect observations	3.1041	11.1133	2.2780	0.5552	2.8524
B with entries $B_{ij} = 0.01e^{-\frac{\lvert i-j\rvert}{2L^2}}$ where $L = 5$	full perfect observations	0.4921	1.0184	0.4857	0.4346	0.1696
	partial perfect observations	0.3150	2.0667	0.2938	0.1633	0.9128
	partial imperfect observations	0.4161	1.5400	0.3997	0.3057	0.8456
B with entries $B_{ij} = 0.005e^{-\frac{\lvert i-j\rvert}{2L^2}}$ where $L = 5$	full perfect observations	0.4023	0.9396	0.3981	0.3636	0.1567
	partial perfect observations	0.2304	0.6327	0.2171	0.1455	0.6922
	partial imperfect observations	0.3225	0.5489	0.3139	0.2680	0.5686
$B = I$ and smaller length of assimilation window	full perfect observations	2.1595	2.1858	0.5812	0.3406	0.6591
	partial perfect observations	8.0773	8.2133	1.3201	0.5327	3.7108
	partial imperfect observations	11.2487	11.4258	1.6075	0.6121	3.6611
$B = 0.01I$ and smaller length of assimilation window	full perfect observations	0.6881	0.9963	0.4130	0.1996	0.4832
	partial perfect observations	0.9441	1.7047	0.6182	0.2129	1.6974
	partial imperfect observations	1.2017	2.5580	0.7971	0.1795	2.7750
$B = 0.005I$ and smaller length of assimilation window	full perfect observations	0.5463	0.8378	0.3677	0.1553	0.3939
	partial perfect observations	0.6809	1.4938	0.4903	0.1795	1.0246
	partial imperfect observations	0.8293	2.0489	0.6132	0.1510	1.1469

the front is often predicted to be in the wrong position. We simulate this situation in our simplified model by assuming a slightly shifted background. We add noise to this background, taken from a normal distribution with covariance matrix $B = 0.01I$ which is consistent with the error in the shifted background.

We only consider the case with partial noisy observations, since this is the most interesting and realistic one.

The results for this example are shown in the plots in Figures 14 and 15. The initial condition in 4DVar is clearly recovered very badly, with many oscillations (see first plot in Figure 14). Furthermore, at the end of the assimilation window the solution gives undershoots (see second plot in Figure 14). We also see that the amplitude of the front can is reduced in (see second and third plot in Figure 14).

However, the solution using the mixed TV L_1-L_2-norm regularisation provides a much better initial condition, with no oscillations present (see first plot in Figure 15). Moreover, there are no undershoots in the solution at the end of the assimilation window (see second plot in Figure 15). Therefore mixed TV L_1-L_2-norm regularisation gives a better initial condition for the forecast than standard 4DVar. Furthermore, mixed TV L_1-L_2-norm regularisation retains the amplitute of the front more accurately than 4DVar (see second and third plot in Figure 15).

7 Conclusions and future work

In this paper we have presented mixed TV L_1-L_2-norm regularisation, a new approach for variational data assimilation. We have given numerical examples where shock fronts are present in order to demonstrate that mixed TV L_1-L_2-norm regularisation gives 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 mixed TV L_1-L_2-norm regularisation approach in (19) is more involved than that for the standard approach in (13), practical implementations will also have to be investigated together with the efficiency of this new approach.

Appendix

The solution to the the data assimilation problem can be interpreted in statistical terms, where certain assumptions about the errors hold (Nichols (2010)). For the standard 4DVar problem, Gaussian errors are assumed for both the background and the observations, so the minimisation of the objective function (1) is equivalent to maximising the *a posteriori* likelihood estimate of the state, given the observations and the prior. A similar derivation can be

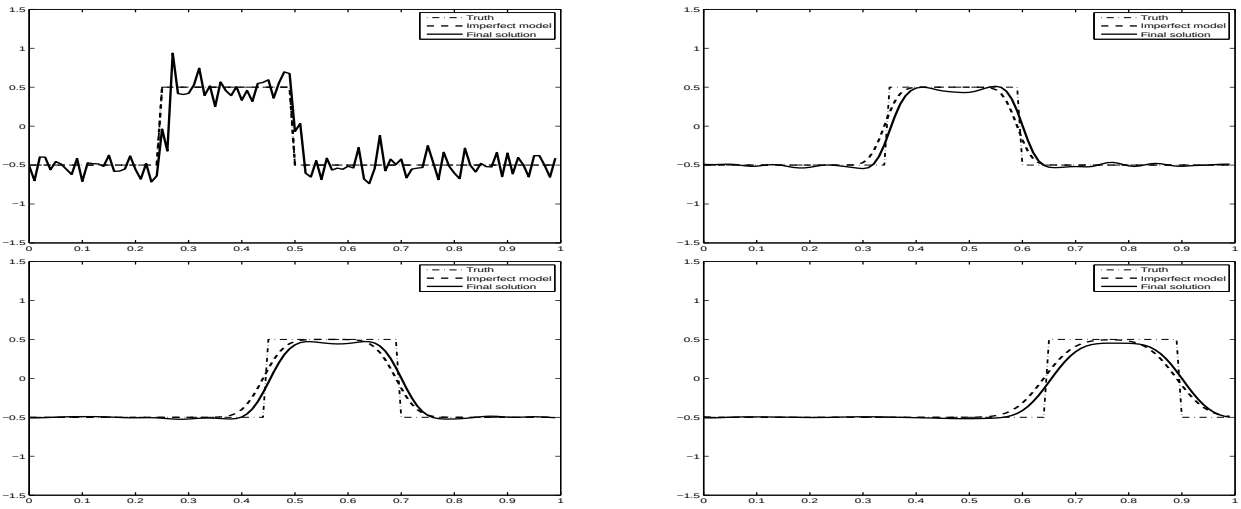


Figure 14. Results for **4DVar** for a shifted (and noisy) background.

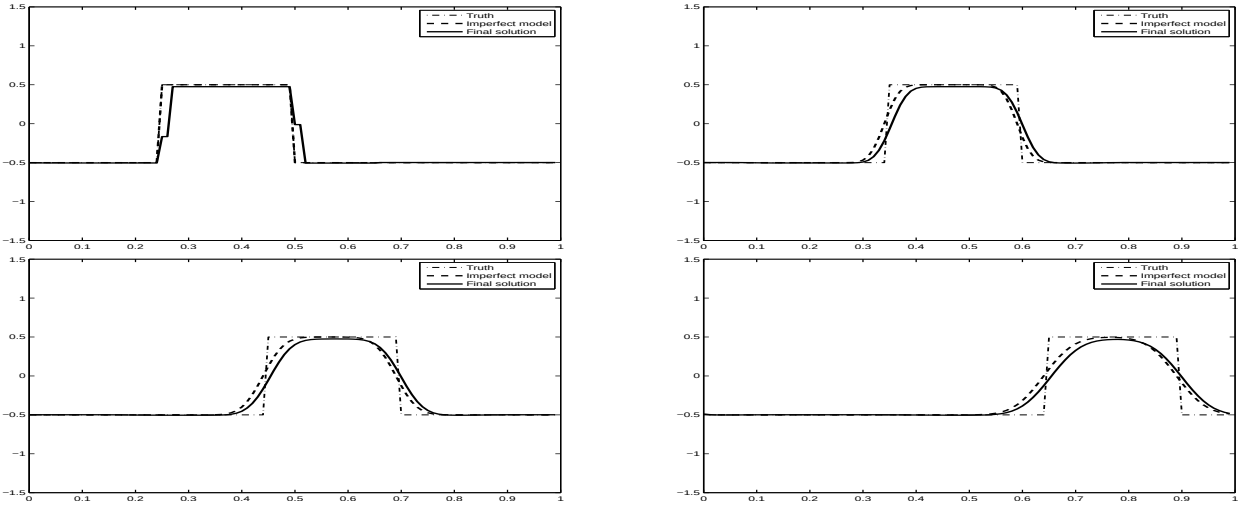


Figure 15. Results for **mixed TV L_1-L_2-norm regularisation** for the same data as in Figure 14.

made for L_1-norm regularisation (16) (which we are going to do here for the single variate case).

The addition of the penalty term $\mu^2\|z\|_1$ in (16) to the least squares term is sometimes also referred to as Lasso regression in statistics (Tibshirani (1996)). Now, $|z_i|$, where z_i is the ith entry of z, is proportional to the negative log-density of the Laplace (or double-sided exponential) distribution. Hence, the L_1-norm regularisation can be derived as a Bayesian posterior estimate, where the priors are independently distributed variables with Laplace probability density function

$$f(z_i) = \frac{1}{2\gamma} e^{-\frac{|z_i|}{\gamma}}, \tag{33}$$

where $\gamma = 1/\mu^2$. The in-depth mathematical investigation of L_1-norm regularisation is the subject of future research and beyond the scope of this paper.

In order to solve equation (3) at each step we use a direct method (backslash in Matlab).

We remark that the solution of the minimisation problem using the least mixed norm solution described in section 4, (see also(Fu *et al.* (2006))) is more expensive than standard 4DVar as the problem size is increased. More efficient methods need to be found for the minimisation; the details are beyond the scope of this paper.

We note that traditional 4DVar is not designed to deal with model error. Hence, for future work, a fairer comparison would be weak-constraint 4DVar (see, for example Trémolet (2006)) with L_1-regularisation.

Acknowledgement

The authors thank Nathan Smith (University of Bath) for helpful discussions on the subject of L_1-norm regularisation. 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.

Bennett A. F. 2002. *Inverse Modelling of the Ocean and Atmosphere (234 pp.)*. Cambridge: Cambridge University Press.

Daley R. 1991. *Atmospheric Data Analysis (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.

Donoho DL. 2006. Compressed Sensing. *Information Theory, IEEE Transactions on* **52**(4): 1289–1306.

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.

Haben SA, Lawless AS, Nichols NK. 2010. Conditioning and preconditioning of the variational data assimilation problem. *Computers & Fluids*, in press. (Published on line: 30 Nov 2010 doi:10.1016/j.compfluid.2010.11.025).

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. 2003. *Information Content of Observations in Variational Data Assimilation*. PhD Thesis. University of Reading.

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. An investigation of incremental 4D-Var using non-tangent linear models. *Q. J. R. Meteorol. Soc.* **131**: 459–476.

Lawless AS, Gratton S, Nichols NK. 2005. Approximate iterative methods for variational data assimilation. *Internat. J. Numer. Methods Fluids* **47**(10-11): 1129–1135.

LeVeque R. 1992. *Numerical methods for conservation laws*. Birkhäuser.

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

Lorenc AC. 1981. A global three-dimensional multivariate statistical interpolation scheme. *Mon. Wea. Rev.* **109**: 701–721.

Lorenc AC. 1986. Analysis methods for numerical weather prediction. *Q. J. R. Meteorol. Soc.* **112**: 1177–1194.

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. 13–39.

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.

Strong D, Chan T. 2003. Edge-preserving and scale-dependent properties of total variation regularization. *Inverse Problems* **19**(6): 165–187.

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

Tibshirani R. 1996. Regression shrinkage and selection via the lasso. *J. Roy. Statist. Soc. Ser. B* **58**(1): 267–288.

Trémolet Y. 2006. Accounting for an imperfect model in 4D-Var. *Q. J. R. Meteorol. Soc.* **132**: 2483–2504.