L1-regularisation in 4DVar

Melina Freitag

Department of Mathematical Sciences
University of Bath

UK MetOffice
3rd March 2009

Outline

Four-dimensional variational assimilation (4DVar)

Minimise the cost function

$$J(\mathbf{x}_0) = (\mathbf{x}_0 - \mathbf{x}_0^B)^T \mathbf{B}^{-1} (\mathbf{x}_0 - \mathbf{x}_0^B) + \sum_{i=0}^{n} (\mathbf{y}_i - H_i(\mathbf{x}_i))^T \mathbf{R}_i^{-1} (\mathbf{y}_i - H_i(\mathbf{x}_i))$$

subject to model dynamics $\mathbf{x}_i = M_{0 \to i} \mathbf{x}_0$

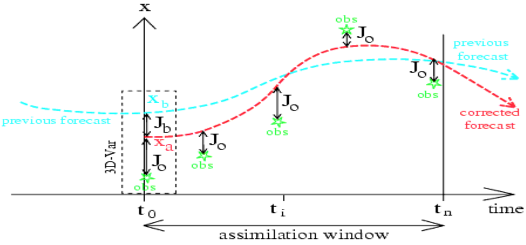


Figure: Copyright:ECMWF

Outline

Ill-posed problems

Given an operator $\mathbf{A}$ we wish to solve

$$\mathbf{A}\mathbf{x} = \mathbf{b}$$

it is well-posed if

- solution exits
- solution is unique
- is stable ($\mathbf{A}^{-1}$ continuous)

Equation is ill-posed if it is not well-posed.

Ill-posed problems

Given an operator $\mathbf{A}$ we wish to solve

$$\mathbf{A}\mathbf{x} = \mathbf{b}$$

it is well-posed if

- solution exits
- solution is unique
- is stable ($\mathbf{A}^{-1}$ continuous)

Equation is ill-posed if it is not well-posed.

but ..

In finite dimensions existence and uniqueness can be imposed, but

- discrete problem of underlying ill-posed problem becomes ill-conditioned
- singular values of $\mathbf{A}$ decay to zero
- $\mathbf{A}^{-1}$ is unstable!

A way out of this - Tikhonov regularisation

Solution to the minimisation problem

$$\begin{aligned} \mathbf{x}_\alpha &= \arg\min \left\{ \|\mathbf{A}\mathbf{x} - \mathbf{b}\|^2 + \alpha\|\mathbf{x}\|^2 \right\} \\ &= (\mathbf{A}^T\mathbf{A} + \alpha\mathbf{I})^{-1}\mathbf{A}^T\mathbf{b} \end{aligned}$$

where α is called the regularisation parameter.

Tikhonov regularisation using Singular Value Decomposition

Using the SVD of $\mathbf{A} = \mathbf{U}\Sigma\mathbf{V}^T$ the regularised solution in Tikhonov regularisation is given by

$$\begin{aligned}
\mathbf{x}_\alpha &= (\mathbf{A}^T\mathbf{A} + \alpha\mathbf{I})^{-1}\mathbf{A}^T\mathbf{b} \\
&= (\mathbf{V}\boldsymbol{\Sigma}^T\mathbf{U}^T\mathbf{U}\Sigma\mathbf{V}^T + \alpha\mathbf{V}\mathbf{V}^T)^{-1}\mathbf{V}\boldsymbol{\Sigma}^T\mathbf{U}^T\mathbf{b} \\
&= \mathbf{V}\mathrm{diag}\left(\frac{s_i^2}{s_i^2+\alpha}\frac{1}{s_i}\right)\mathbf{U}^T\mathbf{b}
\end{aligned}$$

$$\mathbf{x}_\alpha = \sum_{i=1}^{n} \frac{s_i^2}{s_i^2+\alpha}\frac{u_i^T\mathbf{b}}{s_i}\mathbf{v}_i$$

Outline

Relation between 4DVar and Tikhonov regularisation

4DVar minimises

$$J(\mathbf{x}_0) = (\mathbf{x}_0 - \mathbf{x}_0^B)^T \mathbf{B}^{-1} (\mathbf{x}_0 - \mathbf{x}_0^B) + \sum_{i=0}^{n} (\mathbf{y}_i - H_i(\mathbf{x}_i))^T \mathbf{R}_i^{-1} (\mathbf{y}_i - H_i(\mathbf{x}_i))$$

subject to model dynamics $\mathbf{x}_i = M_{0 \to i} \mathbf{x}_0$

Relation between 4DVar and Tikhonov regularisation

4DVar minimises

$$J(\mathbf{x}_0) = (\mathbf{x}_0 - \mathbf{x}_0^B)^T \mathbf{B}^{-1} (\mathbf{x}_0 - \mathbf{x}_0^B) + \sum_{i=0}^{n} (\mathbf{y}_i - H_i(\mathbf{x}_i))^T \mathbf{R}_i^{-1} (\mathbf{y}_i - H_i(\mathbf{x}_i))$$

subject to model dynamics $\mathbf{x}_i = M_{0\to i}\mathbf{x}_0$

or

$$J(\mathbf{x}_0) = (\mathbf{x}_0 - \mathbf{x}_0^B)^T \mathbf{B}^{-1} (\mathbf{x}_0 - \mathbf{x}_0^B) + (\hat{\mathbf{y}} - \hat{\mathbf{H}}(\mathbf{x}_0))^T \hat{\mathbf{R}}^{-1} (\hat{\mathbf{y}} - \hat{\mathbf{H}}(\mathbf{x}_0))$$

where

$$\hat{\mathbf{H}} = [H_0^T, (H_1 M(t_1, t_0))^T, \dots (H_n M(t_n, t_0))^T]^T$$

$$\hat{\mathbf{y}} = [\mathbf{y}_0^T, \dots, \mathbf{y}_n^T]^T$$

and $\hat{\mathbf{R}}$ is block diagonal with $\mathbf{R}_i$ on diagonal.

Relation between 4DVar and Tikhonov regularisation

Solution to the optimisation problem

Linearise about $\mathbf{x}_0$ then the solution to the optimisation problem

$$J(\mathbf{x}_0) = (\mathbf{x}_0 - \mathbf{x}_0^B)^T \mathbf{B}^{-1} (\mathbf{x}_0 - \mathbf{x}_0^B) + (\hat{\mathbf{y}} - \hat{\mathbf{H}}(\mathbf{x}_0))^T \hat{\mathbf{R}}^{-1} (\hat{\mathbf{y}} - \hat{\mathbf{H}}(\mathbf{x}_0))$$

is given by

$$\mathbf{x}_0 = \mathbf{x}_0^B + (\mathbf{B}^{-1} + \hat{\mathbf{H}}^T \hat{\mathbf{R}}^{-1} \hat{\mathbf{H}})^{-1} \hat{\mathbf{H}}^T \hat{\mathbf{R}}^{-1} \hat{\mathbf{d}}, \quad \hat{\mathbf{d}} = \hat{\mathbf{H}}(\mathbf{x}_0^B - \hat{\mathbf{y}})$$

Relation between 4DVar and Tikhonov regularisation

Solution to the optimisation problem

Linearise about $\mathbf{x}_0$ then the solution to the optimisation problem

$$J(\mathbf{x}_0) = (\mathbf{x}_0 - \mathbf{x}_0^B)^T \mathbf{B}^{-1} (\mathbf{x}_0 - \mathbf{x}_0^B) + (\hat{\mathbf{y}} - \hat{\mathbf{H}}(\mathbf{x}_0))^T \hat{\mathbf{R}}^{-1} (\hat{\mathbf{y}} - \hat{\mathbf{H}}(\mathbf{x}_0))$$

is given by

$$\mathbf{x}_0 = \mathbf{x}_0^B + (\mathbf{B}^{-1} + \hat{\mathbf{H}}^T \hat{\mathbf{R}}^{-1} \hat{\mathbf{H}})^{-1} \hat{\mathbf{H}}^T \hat{\mathbf{R}}^{-1} \hat{\mathbf{d}}, \quad \hat{\mathbf{d}} = \hat{\mathbf{H}}(\mathbf{x}_0^B - \hat{\mathbf{y}})$$

Singular value decomposition

Assume $\mathbf{B} = \sigma_B^2 \mathbf{I}$ and $\hat{\mathbf{R}} = \sigma_O^2 \mathbf{I}$ and define the SVD of the observability matrix $\hat{\mathbf{H}}$

$$\hat{\mathbf{H}} = \mathbf{U} \mathbf{\Sigma} \mathbf{V}^T$$

Then the optimal analysis can be written as

$$\mathbf{x}_0 = \mathbf{x}_0^B + \sum_j \frac{s_j^2}{\mu^2 + s_j^2} \frac{\mathbf{u}_j^T \hat{\mathbf{d}}}{s_j} \mathbf{v}_j, \quad \text{where} \quad \mu^2 = \frac{\sigma_O^2}{\sigma_B^2}.$$

Relation between 4DVar and Tikhonov regularisation

If $\mathbf{B}$ and $\hat{\mathbf{R}}$ are not multiples of the identity

$$J(\mathbf{x}_0) = (\mathbf{x}_0 - \mathbf{x}_0^B)^T \mathbf{B}^{-1} (\mathbf{x}_0 - \mathbf{x}_0^B) + (\hat{\mathbf{y}} - \hat{\mathbf{H}}(\mathbf{x}_0))^T \hat{\mathbf{R}}^{-1} (\hat{\mathbf{y}} - \hat{\mathbf{H}}(\mathbf{x}_0))$$

Relation between 4DVar and Tikhonov regularisation

If $\mathbf{B}$ and $\hat{\mathbf{R}}$ are not multiples of the identity

$$J(\mathbf{x}_0) = (\mathbf{x}_0 - \mathbf{x}_0^B)^T \mathbf{B}^{-1} (\mathbf{x}_0 - \mathbf{x}_0^B) + (\hat{\mathbf{y}} - \hat{\mathbf{H}}(\mathbf{x}_0))^T \hat{\mathbf{R}}^{-1} (\hat{\mathbf{y}} - \hat{\mathbf{H}}(\mathbf{x}_0))$$

Variable transformations

$\mathbf{B} = \sigma_B^2 \mathbf{F}_B$ and $\hat{\mathbf{R}} = \sigma_O^2 \mathbf{F}_R$ and define new variable $\mathbf{z} := \mathbf{F}_B^{-1/2}(\mathbf{x}_0 - \mathbf{x}_0^B)$

$$\hat{J}(\mathbf{z}) = \mu^2 \|\mathbf{z}\|_2^2 + \|\mathbf{F}_R^{-1/2} \hat{\mathbf{d}} - \mathbf{F}_R^{-1/2} \hat{\mathbf{H}} \mathbf{F}_B^{-1/2} \mathbf{z}\|_2^2$$

μ^2 can be interpreted as a regularisation parameter.
This is Tikhonov regularisation!

$$J(\mathbf{x}) = \|\mathbf{A}\mathbf{x} - \mathbf{b}\|^2 + \alpha \|\mathbf{x}\|_2^2$$

Outline

Blurred and exact images

The blurring process as a linear model

- Let $\mathbf{X}$ be the exact image
- Let $\mathbf{B}$ be the blurred image

$$\mathbf{x} = \text{vec}(\mathbf{X}) = \begin{bmatrix} \mathbf{x}_1 \\ \vdots \\ \mathbf{x}_N \end{bmatrix} \in \mathbb{R}^N, \quad \mathbf{b} = \text{vec}(\mathbf{B}) = \begin{bmatrix} \mathbf{b}_1 \\ \vdots \\ \mathbf{b}_N \end{bmatrix} \in \mathbb{R}^N$$

$N = m * n$ are related by the linear model

$$\mathbf{Ax} = \mathbf{b}$$

where $\mathbf{A}$ is the discretisation of a point spread function.

Blurred and exact images

The blurring process as a linear model

- Let $\mathbf{X}$ be the exact image
- Let $\mathbf{B}$ be the blurred image

$$\mathbf{x} = \mathrm{vec}(\mathbf{X}) = \begin{bmatrix} \mathbf{x}_1 \\ \vdots \\ \mathbf{x}_N \end{bmatrix} \in \mathbb{R}^N, \quad \mathbf{b} = \mathrm{vec}(\mathbf{B}) = \begin{bmatrix} \mathbf{b}_1 \\ \vdots \\ \mathbf{b}_N \end{bmatrix} \in \mathbb{R}^N$$

$N = m * n$ are related by the linear model

$$\mathbf{A}\mathbf{x} = \mathbf{b}$$

where $\mathbf{A}$ is the discretisation of a point spread function.

Noise $\mathbf{b} = \mathbf{b}_{\mathrm{exact}} + \mathbf{e}$

$$\mathbf{x}_{\mathrm{Naive}} = \mathbf{A}^{-1}\mathbf{b} = \mathbf{A}^{-1}\mathbf{b}_{\mathrm{exact}} + \mathbf{A}^{-1}\mathbf{e} = \mathbf{x} + \mathbf{A}^{-1}\mathbf{e}$$

Need regularisation techniques!

Standard technique: Tikhonov regularisation

$$\min\left\{\|\mathbf{Ax}-\mathbf{b}\|_2^2+\alpha\|\mathbf{x}\|_2^2\right\}$$

equivalent to

$$\mathbf{x}_\alpha=\sum_{i=1}^{n}\frac{s_i^2}{s_i^2+\alpha}\frac{u_i^T\mathbf{b}}{s_i}\mathbf{v}_i$$

Need regularisation techniques!

Standard technique: Tikhonov regularisation

$$\min \left\{ \|\mathbf{Ax} - \mathbf{b}\|_2^2 + \alpha \|\mathbf{x}\|_2^2 \right\}$$

equivalent to

$$\mathbf{x}_\alpha = \sum_{i=1}^{n} \frac{s_i^2}{s_i^2 + \alpha} \frac{u_i^T \mathbf{b}}{s_i} \mathbf{v}_i$$

L1 regularisation

In image processing, L_1-norm regularisation provides edge preserving image deblurring!

$$\min \left\{ \|\mathbf{Ax} - \mathbf{b}\|_2^2 + \alpha \|\mathbf{x}\|_1 \right\}$$

Results from image deblurring: L1 regularisation

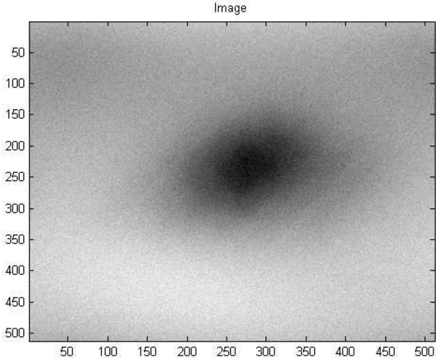


Figure: Blurred picture

Results from image deblurring: L1 regularisation

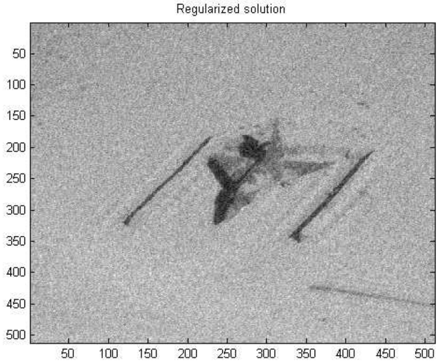


Figure: Tikhonov regularisation $\min\left\{\|\mathbf{A}\mathbf{x}-\mathbf{b}\|_2^2+\alpha\|\mathbf{x}\|_2^2\right\}$

Results from image deblurring: L1 regularisation

Figure: L1-norm regularisation $\min\left\{\|\mathbf{Ax}-\mathbf{b}\|_2^2+\alpha\|\mathbf{x}\|_1\right\}$

L1 regularisation

In image processing, L_1-norm regularisation provides edge preserving image deblurring!

- L1 regularisation beneficial in Data Assimilation?
- 4D Var smears out sharp fronts

L1 regularisation

In image processing, L_1-norm regularisation provides edge preserving image deblurring!

- L1 regularisation beneficial in Data Assimilation?
- 4D Var smears out sharp fronts
- L1 regularisation has the potential to overcome this problem

Outline

Example

Burger's equation

$$u_t + u\frac{\partial u}{\partial x} = u + f(u)_x = 0, \quad f(u) = \frac{1}{2}u^2$$

with initial conditions

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

Discretising

$$x(j) = 10(j - 1/2)\Delta x; \quad U^0(x(j)) = \begin{cases} 2 & 0 \leq x(j) < 2.5 \\ 0.5 & 2.5 \leq x(j) \leq 10. \end{cases}$$

with $\Delta x = \frac{1}{100}$ and $j = 1, \ldots, N$.

Exact solution and model error

Exact solution - method of characteristics

Riemann problem

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

where $s = 1.25$

Numerical solution - model error

- the Lax-Friedrich method (smearing out the shock)

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

- the Lax-Wendroff method (oscillations near the shock).

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

3 Regularisation Methods

4DVar

$$J(U^0) = \frac{1}{2}\|U_B^0 - U^0\|_{\alpha B}^2 + \frac{1}{2}\sum_{i=1}^{N}\|Y_i - H_i(U_i)\|_{R_i}^2$$

3 Regularisation Methods

4DVar

$$J(U^0) = \frac{1}{2}\|U_B^0 - U^0\|_{\alpha B}^2 + \frac{1}{2}\sum_{i=1}^{N}\|Y_i - H_i(U_i)\|_{R_i}^2$$

L1 regularisation

$$J(U^0) = \frac{1}{2}\|Z_B^0 - Z^0\|_p^p + \frac{1}{2}\sum_{i=1}^{N}\|Y_i - H_i(U_i)\|_{R_i}^2$$

where $p = 1$ (or $p = 1.0001$) and $Z = (\alpha B)^{-1/2}U$.

3 Regularisation Methods

4DVar

$$J(U^0) = \frac{1}{2}\|U_B^0 - U^0\|_{\alpha B}^2 + \frac{1}{2}\sum_{i=1}^{N}\|Y_i - H_i(U_i)\|_{R_i}^2$$

L1 regularisation

$$J(U^0) = \frac{1}{2}\|Z_B^0 - Z^0\|_p^p + \frac{1}{2}\sum_{i=1}^{N}\|Y_i - H_i(U_i)\|_{R_i}^2$$

where $p = 1$ (or $p = 1.0001$) and $Z = (\alpha B)^{-1/2}U$.

Total Variation regularisation

$$J(U^0) = \frac{1}{2}\|D(Z_B^0 - Z^0)\|_p^p + \frac{1}{2}\sum_{i=1}^{N}\|Y_i - H_i(U_i)\|_{R_i}^2$$

where D is a matrix approximating the derivative of the solution.

Setup

- $\Delta t = 0.001$
- length of the assimilation window: 100 time steps
- perfect observations
- Here: consider only Lax-Friedrich method

Truth trajectory and numerical solution

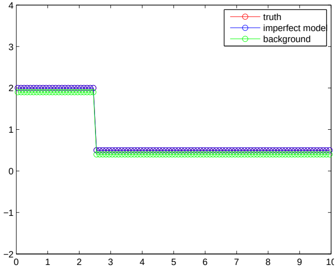


Figure: Initial conditions for Truth, Imperfect model and Background, $t = 0$

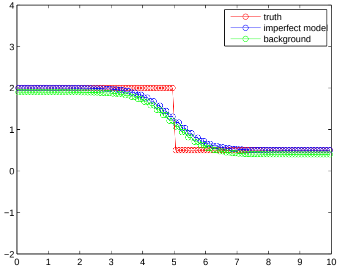
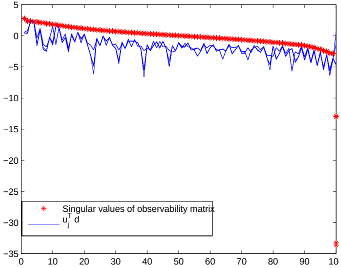


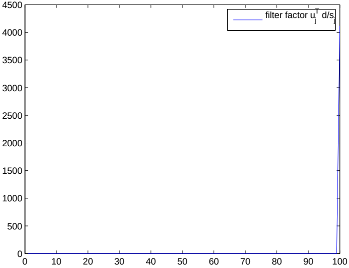
Figure: Truth, Imperfect model and Background after 200 time steps

Singular value analysis - observations everywhere

Optimal solution (4DVar)

$$\mathbf{u}_0 = \mathbf{u}_0^B + \sum_j \frac{s_j^2}{\mu^2 + s_j^2} \frac{\mathbf{u}_j^T \hat{\mathbf{d}}}{s_j} \mathbf{v}_j, \quad \text{where} \quad \mu^2 = \frac{\sigma_O^2}{\sigma_B^2}.$$

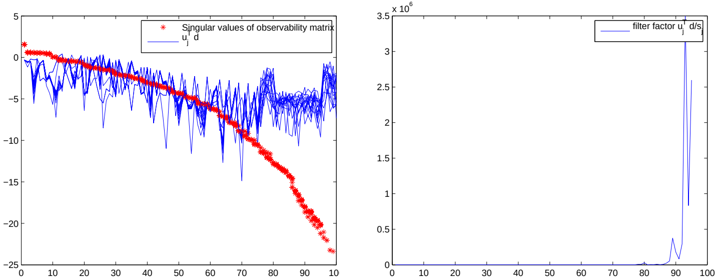




Singular value analysis - observations every 2 time steps and every 20 points in space

Optimal solution (4DVar)

$$\mathbf{u}_0 = \mathbf{u}_0^B + \sum_j \frac{s_j^2}{\mu^2 + s_j^2} \frac{\mathbf{u}_j^T \hat{\mathbf{d}}}{s_j} \mathbf{v}_j, \quad \text{where} \quad \mu^2 = \frac{\sigma_O^2}{\sigma_B^2}.$$



4DVar - observations everywhere

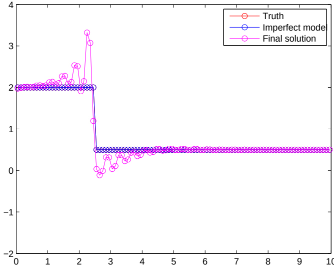


Figure: Truth and Background and final solution at time $t = 0$ (beginning of the assimilation window) using 4DVar

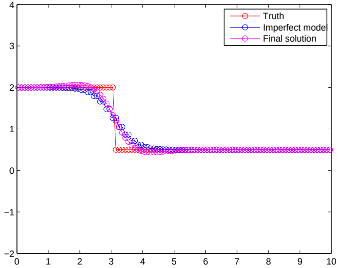


Figure: Truth and Background and final solution after 50 time steps (middle of the assimilation window) using 4DVar

4DVar - observations everywhere

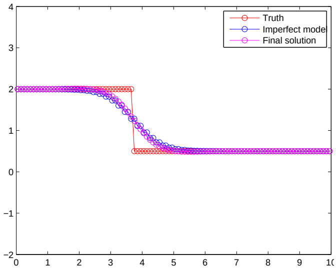


Figure: Truth and Background and final solution after 100 time steps (end of the assimilation window) using 4DVar

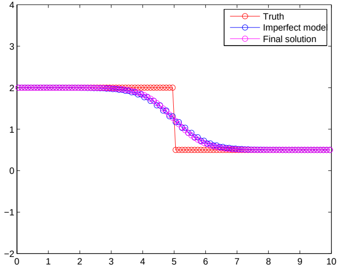


Figure: Truth and Background and final solution after 200 time steps using 4DVar

L1/TV regularisation - observations everywhere

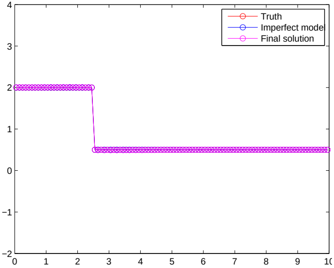


Figure: Truth and Background and final solution at time $t = 0$ (beginning of the assimilation window) using L1

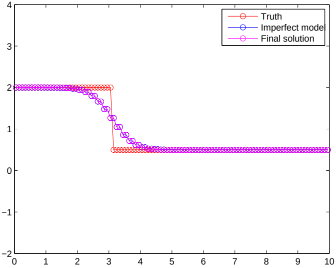


Figure: Truth and Background and final solution after 50 time steps (middle of the assimilation window) using L1

L1/TV regularisation - observations everywhere

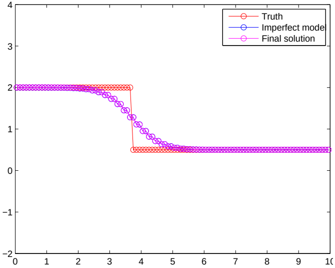


Figure: Truth and Background and final solution after 100 time steps (end of the assimilation window) using L1

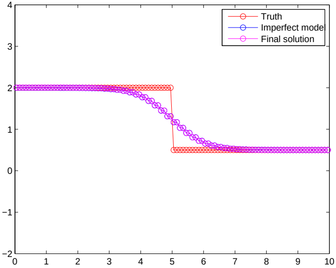


Figure: Truth and Background and final solution after 200 time steps using L1

4DVar - L1/TV regularisation - observations everywhere

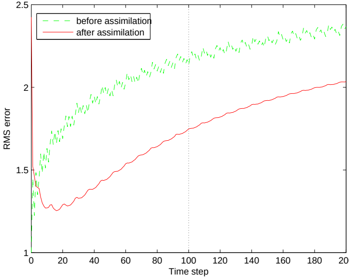


Figure: Root mean square error using 4D-Var.

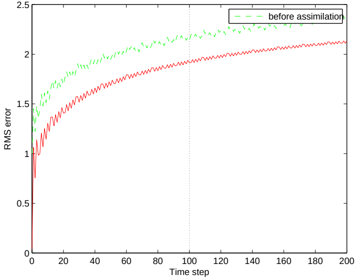


Figure: Root mean square error using L1/TV regularisation.

4DVar - observations every 2 time steps and every 20 points in space

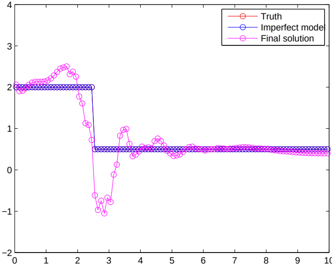


Figure: Truth and Background and final solution at time $t = 0$ (beginning of the assimilation window) using 4DVar

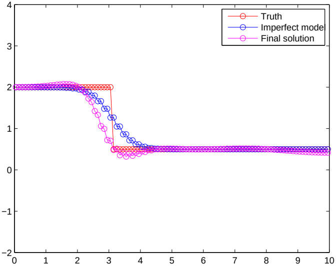


Figure: Truth and Background and final solution after 50 time steps (middle of the assimilation window) using 4DVar

4DVar - observations every 2 time steps and every 20 points in space

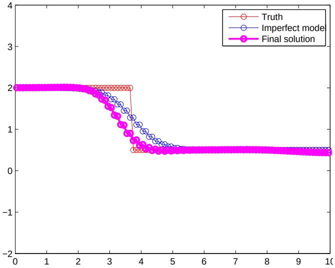


Figure: Truth and Background and final solution after 100 time steps (end of the assimilation window) using 4DVar

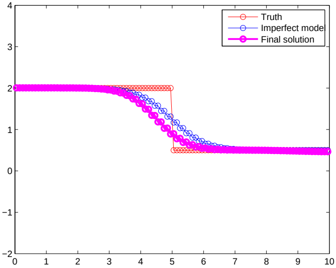


Figure: Truth and Background and final solution after 200 time steps using 4DVar

L1/TV regularisation - observations every 2 time steps and every 20 points in space

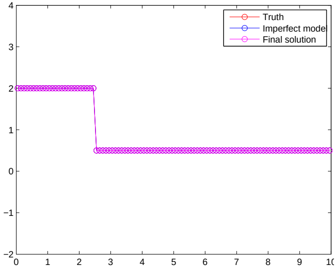


Figure: Truth and Background and final solution at time $t = 0$ (beginning of the assimilation window) using L1

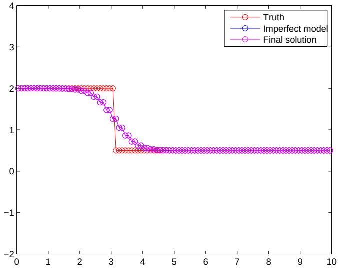


Figure: Truth and Background and final solution after 50 time steps (middle of the assimilation window) using L1

L1/TV regularisation - observations every 2 time steps and every 20 points in space

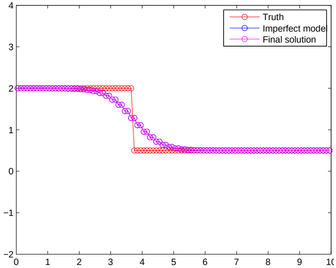


Figure: Truth and Background and final solution after 100 time steps (end of the assimilation window) using L1

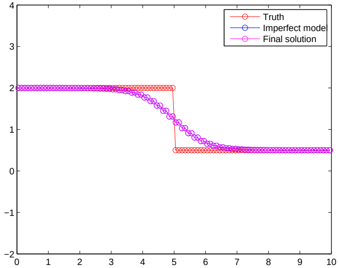


Figure: Truth and Background and final solution after 200 time steps using L1

4DVar - L1/TV regularisation - observations everywhere

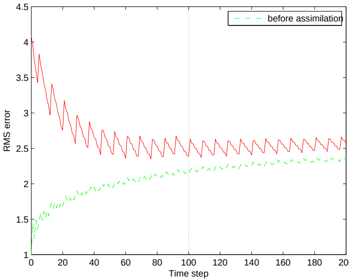


Figure: Root mean square error using 4D-Var.

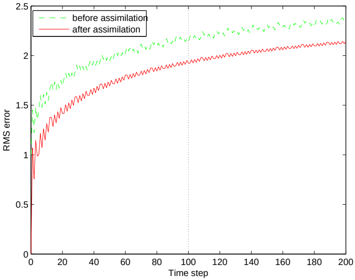


Figure: Root mean square error using L1/TV regularisation.

Conclusions, questions and further work

- L1-norm and TV regularisation recovers discontinuity better than 4DVar
- experiments with Lax-Wendroff similar
- experiments with noisy observations/different B matrices similar
- TV regularisation seems in general faster than L1 regularisation, but giving same results
- Implementation using quadratic programming tools