L1-regularisation in 4DVar

Melina Freitag

Department of Mathematical Sciences
University of Bath

University of Reading
16th December 2008

Outline

- Introduction
- Tikhonov regularisation
- Link between 4DVar and Tikhonov regularisation
- Motivation: Results from image processing
- First results on $L1$-regularisation in 4DVar

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

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

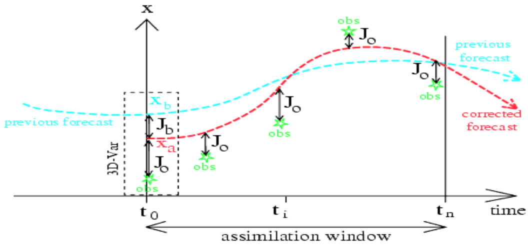


Figure: Copyright:ECMWF

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

Ill-posed problems

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

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

it is well-posed if

- solution exits
- solution is unique

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)

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.

Linear, finite dimensional case

Finite dimensions

- $\mathbf{A} : \mathbb{R}^n \rightarrow \mathbb{R}^n$, then $\mathbf{Ax} = \mathbf{b}$ is well-posed if $\mathbf{A}^{-1}$ exists.

Linear, finite dimensional case

Finite dimensions

- $\mathbf{A} : \mathbb{R}^n \rightarrow \mathbb{R}^n$, then $\mathbf{Ax} = \mathbf{b}$ is well-posed if $\mathbf{A}^{-1}$ exists.
- Existence imposed by considering least squares solutions

$$\mathbf{x}_{\mathrm{LS}} = \arg\min\{\|\mathbf{Ax} - \mathbf{b}\|\}$$

Linear, finite dimensional case

Finite dimensions

- $\mathbf{A} : \mathbb{R}^n \to \mathbb{R}^n$, then $\mathbf{A}\mathbf{x} = \mathbf{b}$ is well-posed if $\mathbf{A}^{-1}$ exists.
- Existence imposed by considering least squares solutions

$$\mathbf{x}_{\mathrm{LS}} = \arg\min\{\|\mathbf{A}\mathbf{x} - \mathbf{b}\|\}$$

- Uniqueness imposed by taking minimum norm least squares solution

$$\mathbf{x}_{\mathrm{MNLS}} = \arg\min\{\|\mathbf{x}_{LS}\|\} = \mathbf{A}^{\dagger}\mathbf{b}.$$

Linear, finite dimensional case

Finite dimensions

- $\mathbf{A} : \mathbb{R}^n \to \mathbb{R}^n$, then $\mathbf{Ax} = \mathbf{b}$ is well-posed if $\mathbf{A}^{-1}$ exists.
- Existence imposed by considering least squares solutions

$$\mathbf{x}_{\text{LS}} = \arg\min\{\|\mathbf{Ax} - \mathbf{b}\|\}$$

- Uniqueness imposed by taking minimum norm least squares solution

$$\mathbf{x}_{\text{MNLS}} = \arg\min\{\|\mathbf{x}_{LS}\|\} = \mathbf{A}^{\dagger}\mathbf{b}.$$

but ..

In the finite dimensional case one can replace $\mathbf{A}^{-1}$ by its pseudo-inverse $\mathbf{A}^{\dagger}$, 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

$$\mathbf{x}_\alpha = (\mathbf{A}^T\mathbf{A} + \alpha\mathbf{I})^{-1}\mathbf{A}^T\mathbf{b}$$

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

$$\begin{aligned} \mathbf{x}_\alpha &= (\mathbf{A}^T\mathbf{A} + \alpha\mathbf{I})^{-1}\mathbf{A}^T\mathbf{b} \\ &= (\mathbf{V}\boldsymbol{\Sigma}^T\mathbf{U}^T\mathbf{U}\Sigma\mathbf{V}^T + \alpha\mathbf{V}\mathbf{V}^T)^{-1}\mathbf{V}\boldsymbol{\Sigma}^T\mathbf{U}^T\mathbf{b} \end{aligned}$$

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

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

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

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

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

Bayesian Interpretation

Assuming X, B are random variables then

$$\pi(\mathbf{x}|\mathbf{b}) = \pi(\mathbf{b}|\mathbf{x})\pi(\mathbf{x})/\pi(\mathbf{b}),$$

Maximum a posteriori estimator is maximum of a posteriori pdf, hence minimise w.r.t. $\mathbf{x}$

$$-\log(\pi(\mathbf{x}|\mathbf{b})) = -\log(\pi(\mathbf{b}|\mathbf{x})) - \log(\pi(\mathbf{x}))$$

Example

If X and $\eta = B - AX$ are normally distributed then

$$\pi(\mathbf{x}) = C_1 \exp(-\frac{\|\mathbf{x}\|^2}{2\sigma_x}) \quad \text{and} \quad \pi(\mathbf{x}|\mathbf{b}) = C_2 \exp(-\frac{\|\mathbf{Ax}-\mathbf{b}\|^2}{2\sigma_\eta^2})$$

and Tikhonov cost functional is

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

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

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

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

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

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 $\mathbf{z} := \mathbf{F}_B^{-1/2}(\mathbf{x}_0 - \mathbf{x}_0^B)$

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.

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!

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

Example

Burger's equation

$$u_t + u\frac{\partial u}{\partial x} = \nu \frac{\partial^2 u}{\partial x^2}$$

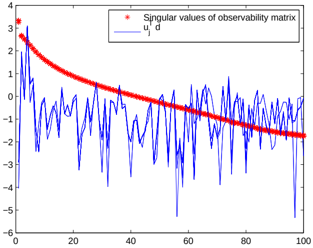
Optimal solution (4DVar)

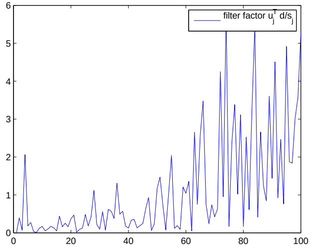
$$\mathbf{u}_0 = \mathbf{u}_0^B + \sum_j \frac{s_j^2}{\mu^2 + s_j^2} \frac{\mathbf{u}_j^T \hat{\mathbf{d}}}{s_j} \mathbf{v}_j, \quad \text{where} \quad \mu^2 = \frac{\sigma_O^2}{\sigma_B^2}.$$

Singular value analysis - observations everywhere

Optimal solution (4DVar)

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

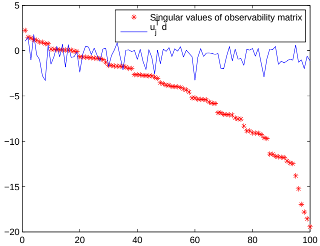


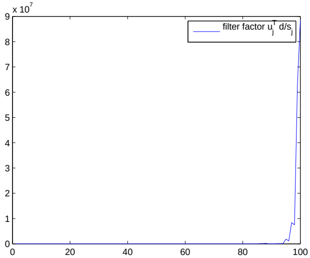


Singular value analysis - observations every 10 points

Optimal solution (4DVar)

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





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

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

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

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

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

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

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

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

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

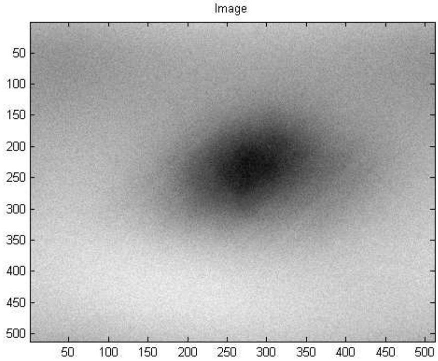


Figure: Blurred picture

Results from image deblurring: L1 regularisation

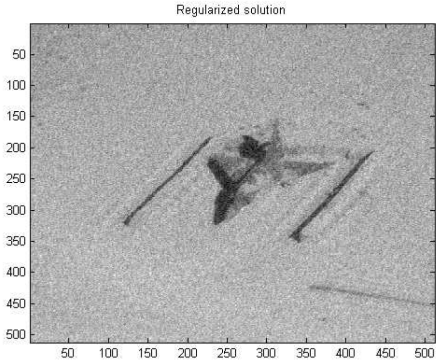


Figure: Tikhonov regularisation $\min\left\{\|\mathbf{Ax}-\mathbf{b}\|_2^2+\alpha\|\mathbf{x}\|_2^2\right\}$

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

Inviscid Burger's equation

Inviscid Burger's equation

$$u_t + uu_x = 0$$

Inviscid Burger's equation

Inviscid Burger's equation

$$u_t + uu_x = 0$$

Conservative first order upwind method

$$U_j^{n+1} = U_j^n - \frac{\Delta t}{\Delta x}(F^n_{j+\frac{1}{2}} - F^n_{j-\frac{1}{2}}),$$

where

$$F_{j+\frac{1}{2}} = \begin{cases} \frac{1}{2}U_j^2 & v_{j+\frac{1}{2}} > 0 \\ \frac{1}{2}U_{j+1}^2 & v_{j+\frac{1}{2}} < 0 \end{cases}$$

$$v_{j+\frac{1}{2}} = \begin{cases} \frac{1}{2}(U_j + U_{j+1}) & U_j \neq U_{j+1} \\ U_j & U_j = U_{j+1}. \end{cases}$$

Initial conditions

Initial conditions for the true solution are

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

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

Initial conditions

Initial conditions for the true solution are

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

with $\Delta x = \frac{1}{100}$ and $j = 1, \ldots, N$ and for the background solution they are

$$U_B^0(x(j)) = \begin{cases} 2.1 & 0 \leq x(j) < 3.5 \\ 0.6 & 3.5 \leq x(j) \leq 10. \end{cases}$$

Truth and Background trajectory

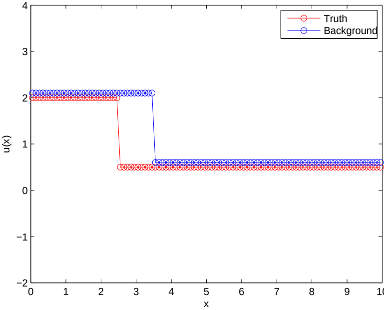


Figure: Initial conditions for Truth and Background, $t = 0$

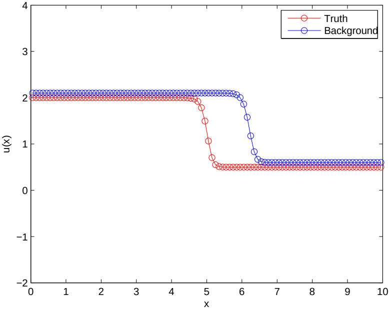


Figure: Truth and Background after 2000 time steps

Setup

- $\Delta t = 0.0001$

Setup

- $\Delta t = 0.0001$
- use "L1.1-regularisation" i.e.

$$\min \left\{ \|\mathbf{Ax} - \mathbf{b}\|_2^2 + \alpha \|\mathbf{x}\|_{1.1}^{1.1} \right\}$$

leading to a differentiable function and avoiding quadratic programming problem for the moment

Setup

- $\Delta t = 0.0001$
- use "L1.1-regularisation" i.e.

$$\min \left\{ \|\mathbf{Ax} - \mathbf{b}\|_2^2 + \alpha \|\mathbf{x}\|_{1.1}^{1.1} \right\}$$

  leading to a differentiable function and avoiding quadratic programming problem for the moment
- length of the assimilation window: 1000 time steps

Setup

- $\Delta t = 0.0001$
- use "L1.1-regularisation" i.e.

$$\min \left\{ \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2 + \alpha \|\mathbf{x}\|_{1.1}^{1.1} \right\}$$

  leading to a differentiable function and avoiding quadratic programming problem for the moment
- length of the assimilation window: 1000 time steps
- perfect and noisy observations

Setup

Perfect observations

4DVar

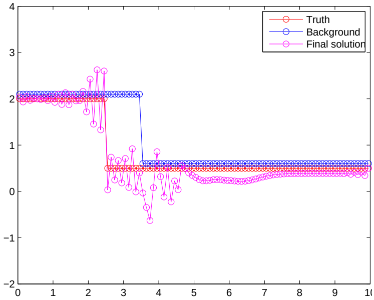


Figure: Truth and Background and final solution at time $t = 0$ (beginning of the assimilation window) using 4DVar

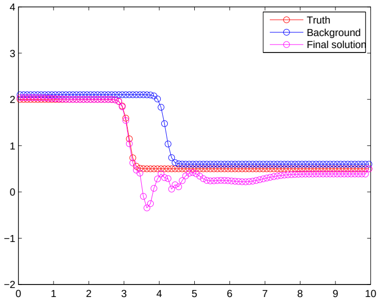


Figure: Truth and Background and final solution after 500 time steps (middle of the assimilation window) using 4DVar

4DVar

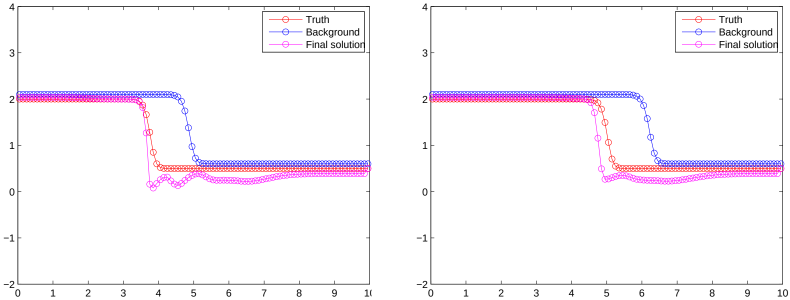


Figure: Truth and Background and final solution after 1000 time steps (which is the end of the assimilation window) using 4DVar

Figure: Truth and Background and final solution after 2000 time steps (assimilation window plus 1000 further steps forecast) using 4DVar

4DVar

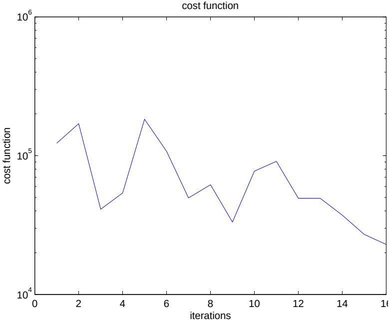


Figure: Cost function

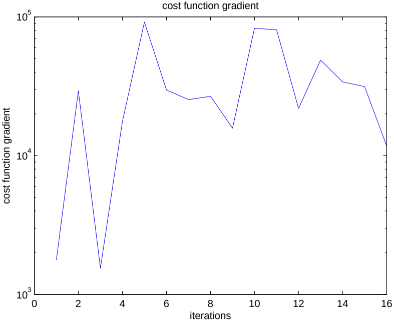


Figure: Cost function gradient

L1.1 Regularisation

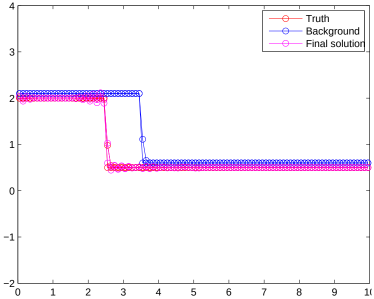


Figure: Truth and Background and final solution at time $t = 0$ (beginning of the assimilation window) using L1.1

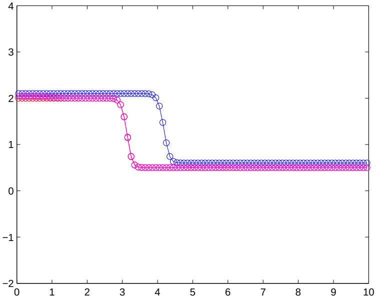


Figure: Truth and Background and final solution after 500 time steps (middle of the assimilation window) using L1.1

L1.1 Regularisation

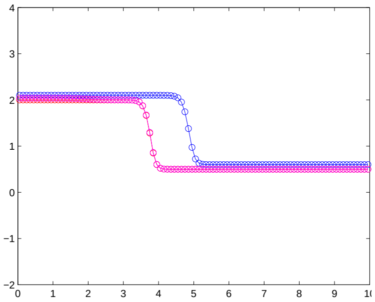


Figure: Truth and Background and final solution after 1000 time steps (which is the end of the assimilation window) using L1.1

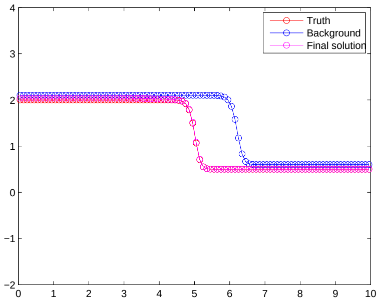


Figure: Truth and Background and final solution after 2000 time steps (assimilation window plus 1000 further steps forecast) using L1.1

L1.1 Regularisation

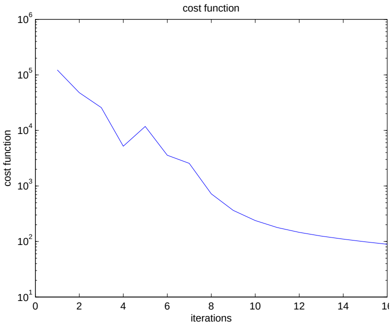


Figure: Cost function

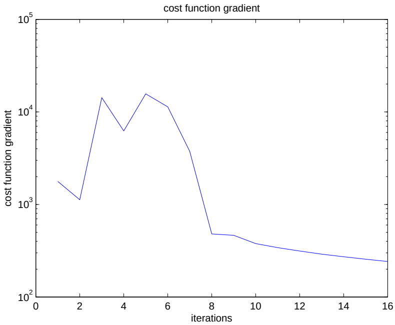


Figure: Cost function gradient

Comparison 4DVar and L1.1 Regularisation

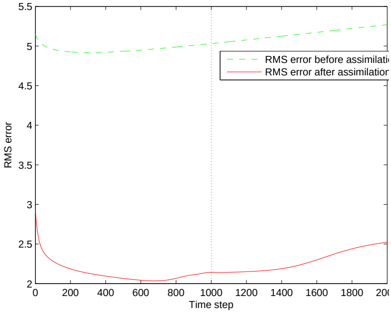


Figure: Root mean square error using 4DVar.

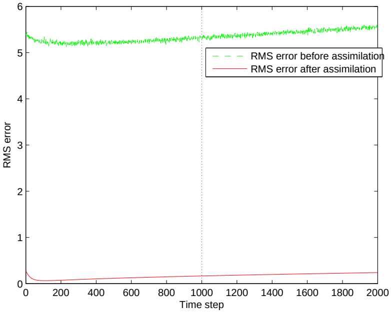


Figure: Root mean square error using L1.1.

Setup

Noisy observations

Comparison 4DVar and L1.1 Regularisation

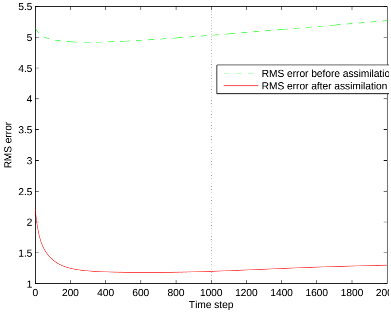


Figure: Root mean square error using 4DVar.

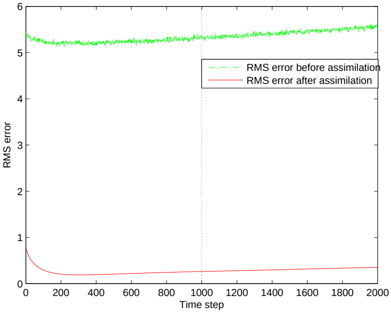


Figure: Root mean square error using L1.1.

4DVar - tridiagonal B

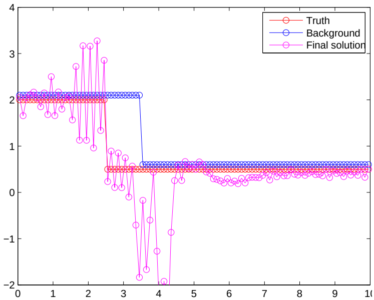


Figure: Truth and Background and final solution at time $t = 0$ (beginning of the assimilation window) using 4DVar

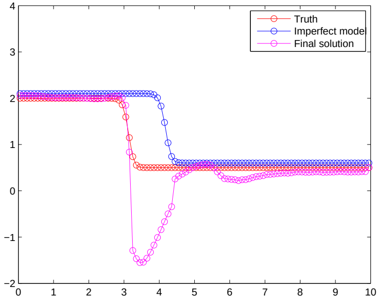


Figure: Truth and Background and final solution after 500 time steps (middle of the assimilation window) using 4DVar

4DVar - tridiagonal B

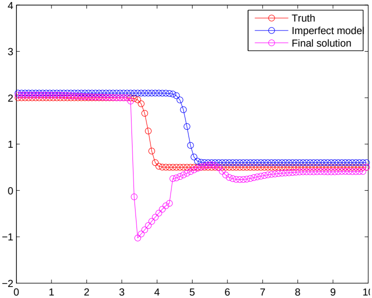


Figure: Truth and Background and final solution after 1000 time steps (which is the end of the assimilation window) using 4DVar

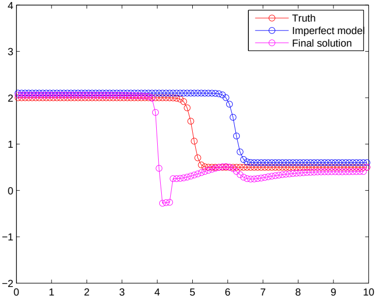


Figure: Truth and Background and final solution after 2000 time steps (assimilation window plus 1000 further steps forecast) using 4DVar

L1.1 Regularisation - tridiagonal B

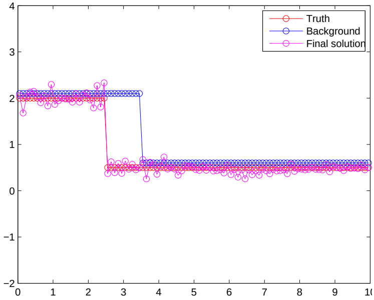


Figure: Truth and Background and final solution at time $t = 0$ (beginning of the assimilation window) using L1.1

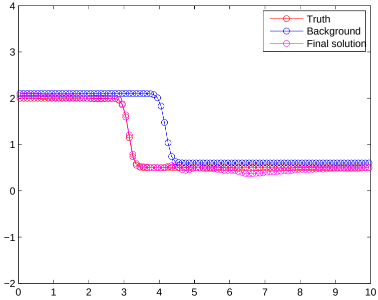


Figure: Truth and Background and final solution after 500 time steps (middle of the assimilation window) using L1.1

L1.1 Regularisation - tridiagonal B

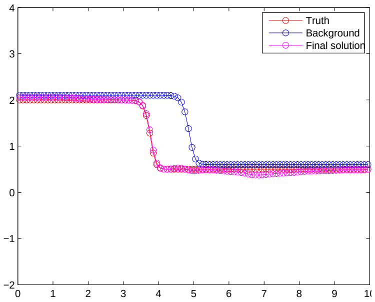


Figure: Truth and Background and final solution after 1000 time steps (which is the end of the assimilation window) using L1.1

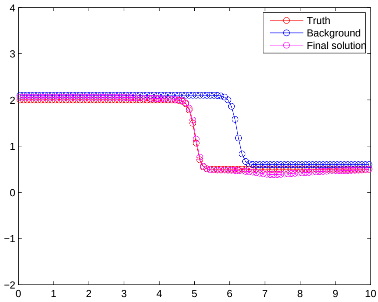


Figure: Truth and Background and final solution after 2000 time steps (assimilation window plus 1000 further steps forecast) using L1.1

4DVar - Gaussian exponential function in B, $b_{ij} = \sigma_b \exp(-r_{ij})$

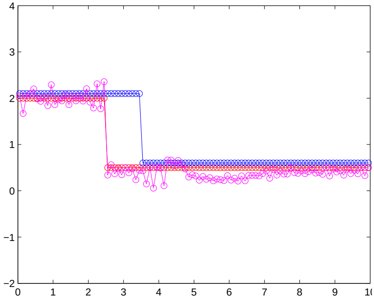


Figure: Truth and Background and final solution at time $t = 0$ (beginning of the assimilation window) using 4DVar

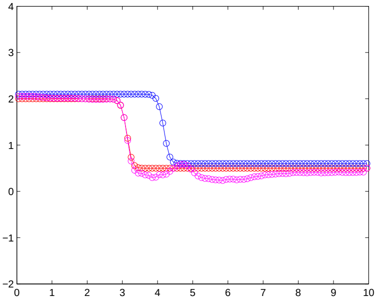


Figure: Truth and Background and final solution after 500 time steps (middle of the assimilation window) using 4DVar

4DVar - Gaussian exponential function in B, $b_{ij} = \sigma_b \exp(-r_{ij})$

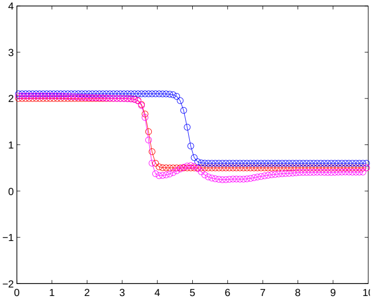


Figure: Truth and Background and final solution after 1000 time steps (which is the end of the assimilation window) using 4DVar

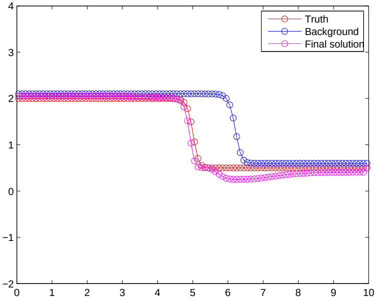


Figure: Truth and Background and final solution after 2000 time steps (assimilation window plus 1000 further steps forecast) using 4DVar

L1.1 Regularisation - Gaussian exponential function in B, $b_{ij} = \sigma_b \exp(-r_{ij})$

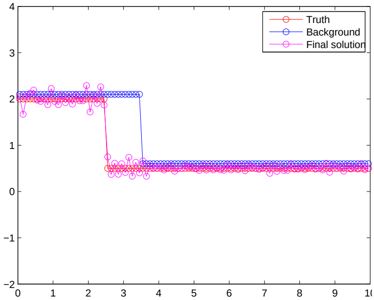


Figure: Truth and Background and final solution at time $t = 0$ (beginning of the assimilation window) using L1.1

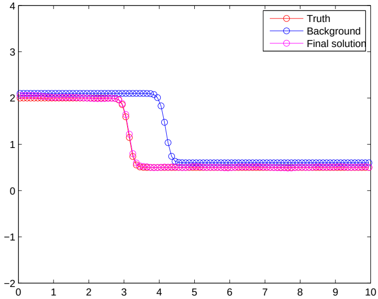


Figure: Truth and Background and final solution after 500 time steps (middle of the assimilation window) using L1.1

UNIVERSITY OF BATH

L1.1 Regularisation - Gaussian exponential function in B, $b_{ij} = \sigma_b \exp(-r_{ij})$

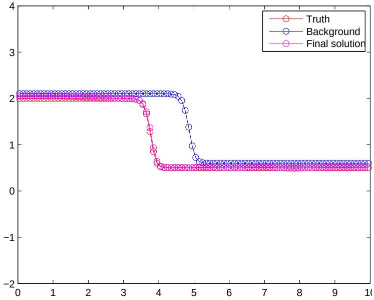


Figure: Truth and Background and final solution after 1000 time steps (which is the end of the assimilation window) using L1.1

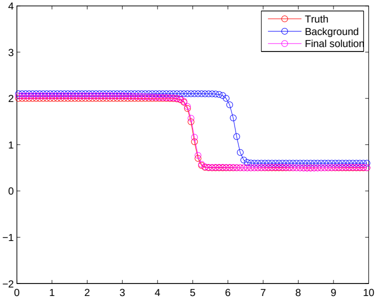


Figure: Truth and Background and final solution after 2000 time steps (assimilation window plus 1000 further steps forecast) using L1.1

Conclusions, questions and further work

- L1.1-norm regularisation recovers discontinuity better than 4DVar
- $L1.1$ minimisation gives smaller increments than 4DVar?
- Breakdown with fewer observations
- Implementation using quadratic/linear programming tools