Tikhonov Regularisation for (Large) Inverse Problems

Melina Freitag

Department of Mathematical Sciences
University of Bath

17th ILAS Conference
Braunschweig, Germany
23rd August 2011

joint work with C.J. Budd (Bath) and N.K. Nichols (Reading)
Inverse Problems

Data Assimilation as a Large Inverse Problem

Regularisation Parameter estimation in 4DVar
- Regularisation Parameter estimation
- Example

Application of $L_1$-norm regularisation in 4DVar
- Motivation: Results from image processing
- $L_1$-norm regularisation in 4DVar
- Examples
Outline

Inverse Problems

Data Assimilation as a Large Inverse Problem

Regularisation Parameter estimation in 4DVar
  Regularisation Parameter estimation
  Example

Application of $L_1$-norm regularisation in 4DVar
  Motivation: Results from image processing
  $L_1$-norm regularisation in 4DVar
  Examples
Ill-posed Problems

Given an operator $A$ we wish to solve

$$Af = g.$$ 

It is well-posed if
Ill-posed Problems

Given an operator $A$ we wish to solve

$$Af = g.$$ 

It is well-posed if

- solution exits
Ill-posed Problems

Given an operator $A$ we wish to solve

$$Af = g.$$ 

It is well-posed if

- solution exits
- solution is unique
Ill-posed Problems

Given an operator $A$ we wish to solve

$$Af = g.$$ 

It is well-posed if

- solution exits
- solution is unique
- is stable ($A^{-1}$ continuous)
Ill-posed Problems

Given an operator $A$ we wish to solve

$$Af = g.$$ 

It is well-posed if

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

but ..

In finite dimensions existence and uniqueness can be imposed, but

- discrete problem of underlying ill-posed problem becomes ill-conditioned
- singular values of $A$ decay to zero
An Illustrative Example

Fredholm first kind integral equation in 1D

\[ g(x) = \int_{0}^{1} k(x - x') f(x') dx' =: (Af)(x), \quad 0 < x < 1 \]

- \( f \) light source intensity as a function of \( x \)
- \( g \) image intensity
- \( k \) kernel representing blurring effects, e.g. \( k(x) = C \exp \left( -\frac{x^2}{2\gamma^2} \right) \), \( C, \gamma \) are positive parameters.
An Illustrative Example

Fredholm first kind integral equation in 1D

\[ g(x) = \int_0^1 k(x - x')f(x')dx' =: (Af)(x), \quad 0 < x < 1 \]

- \( f \) light source intensity as a function of \( x \)
- \( g \) image intensity
- \( k \) kernel representing blurring effects, e.g. \( k(x) = C \exp\left(-\frac{x^2}{2\gamma^2}\right) \), \( C, \gamma \) are positive parameters.

Discretisation

- use a piecewise smooth source \( f \)
- determine \( A \) using standard numerical quadrature;

\[
(A)_{ij} = hC \exp\left(-\frac{(i-j)h^2}{2\gamma^2}\right), \quad 1 \leq i, j \leq n, \quad h = \frac{1}{n}
\]

\[ \gamma = 0.05, \quad C = \frac{1}{\gamma \sqrt{2\pi}}. \]
Forward Problem

Given $f$ and the kernel $k$, determine the blurred image $g = Af$, or the discrete version

$$g = Af.$$
Forward Problem

Given $f$ and the kernel $k$, determine the blurred image $g = Af$, or the discrete version

$$g = Af.$$  

**Figure:** True solution $f$ and blurred image $g$
Inverse Problem

Given the kernel $k$, and the blurred image $g$, determine the source $f$ from $g = Af$, solve the discrete linear system

$$g = Af.$$ 

**Figure:** True solution and blurred image $g$
Inverse Problem

Given the kernel $k$, and the blurred image $g$, determine the source $f$ from $g = Af$, solve the discrete linear system

$$g = Af.$$
Inverse Problem

Problem: data $g$ are observed and contain noise and $A$ is ill-conditioned:

$$g_{\text{exact}} + e = Af,$$

$e$ is unknown white noise.

Figure: True solution and discrete noisy data
Inverse Problem

Problem: data $g$ are observed and contain noise and $A$ is ill-conditioned:

$$g_{\text{exact}} + e = Af,$$

$e$ is unknown white noise.

Singular Value Decomposition

Let

$$A = U\Sigma V^T = \sum_{i=1}^{r} \sigma_i u_i v_i^T$$

where

- $\Sigma = \text{diag}(\sigma_1, \ldots, \sigma_r)$ and $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_r > 0$
- $U^T U = I$ and $V^T V = I$
Inverse Problem - Regularisation needed

Least squares solution (with and without noise)

\[ f_{\text{exact}} = A^\dagger g_{\text{exact}} = \sum_{i=1}^{r} \frac{u_i^T g_{\text{exact}}}{\sigma_i} v_i \]
Inverse Problem - Regularisation needed

Least squares solution (with and without noise)

\[ f_{\text{exact}} = A^\dagger g_{\text{exact}} = \sum_{i=1}^{r} \frac{u_i^T g_{\text{exact}}}{\sigma_i} v_i \]

\[ f = A^\dagger g = A^\dagger(g_{\text{exact}} + e) = \sum_{i=1}^{r} \frac{u_i^T g_{\text{exact}}}{\sigma_i} v_i + \sum_{i=1}^{r} \frac{u_i^T e}{\sigma_i} v_i \]

\[ = f_{\text{exact}} + \sum_{i=1}^{r} \frac{u_i^T e}{\sigma_i} v_i \]
Inverse Problem - Regularisation needed

Least squares solution (with and without noise)

\[
\begin{align*}
\mathbf{f}_{\text{exact}} &= \mathbf{A}^\dagger \mathbf{g}_{\text{exact}} = \sum_{i=1}^{r} \frac{\mathbf{u}_i^T \mathbf{g}_{\text{exact}}}{\sigma_i} \mathbf{v}_i \\
\mathbf{f} &= \mathbf{A}^\dagger \mathbf{g} = \mathbf{A}^\dagger (\mathbf{g}_{\text{exact}} + \mathbf{e}) = \sum_{i=1}^{r} \frac{\mathbf{u}_i^T \mathbf{g}_{\text{exact}}}{\sigma_i} \mathbf{v}_i + \sum_{i=1}^{r} \frac{\mathbf{u}_i^T \mathbf{e}}{\sigma_i} \mathbf{v}_i \\
&= \mathbf{f}_{\text{exact}} + \sum_{i=1}^{r} \frac{\mathbf{u}_i^T \mathbf{e}}{\sigma_i} \mathbf{v}_i
\end{align*}
\]
Tikhonov Regularisation

Regularised solution of the form

\[ f_\alpha = \sum_{i=1}^{r} \frac{\sigma_i^2}{\sigma_i^2 + \alpha^2} \frac{u_i^T g}{\sigma_i} v_i \]

\(\alpha\) regularisation parameter.
Tikhonov Regularisation

Regularised solution of the form

\[ f_\alpha = \sum_{i=1}^{r} \frac{\sigma_i^2}{\sigma_i^2 + \alpha^2} u_i^T g \sigma_i v_i \]

\( \alpha \) regularisation parameter.
Solution \( f_\alpha \) to the minimisation problem

\[
\min_f \left\{ \|g - Af\|_2^2 + \alpha^2 \|f\|_2^2 \right\}.
\]
Tikhonov Regularisation

Regularised solution of the form

$$f_\alpha = \sum_{i=1}^{r} \frac{\sigma_i^2}{\sigma_i^2 + \alpha^2} \frac{u_i^T g}{\sigma_i} v_i$$

$\alpha$ regularisation parameter.

Solution $f_\alpha$ to the minimisation problem

$$\min_f \left\{ \|g - Af\|_2^2 + \alpha^2 \|f\|_2^2 \right\}.$$  

Least squares solution $f_\alpha$ to the linear system

$$\begin{bmatrix} A \\ \alpha I \end{bmatrix} f = \begin{bmatrix} g \\ 0 \end{bmatrix}.$$
Tikhonov Regularisation

Regularised solution of the form

\[ f_\alpha = \sum_{i=1}^{r} \frac{\sigma_i^2}{\sigma_i^2 + \alpha^2} u_i^T g v_i \]

\( \alpha \) regularisation parameter.

Solution \( f_\alpha \) to the minimisation problem

\[ \min_{f} \{ \| g - Af \|_2^2 + \alpha^2 \| f \|_2^2 \} . \]

Least squares solution \( f_\alpha \) to the linear system

\[ \begin{bmatrix} A \\ \alpha I \end{bmatrix} f = \begin{bmatrix} g \\ 0 \end{bmatrix} . \]

Normal equations

\[ (A^T A + \alpha^2 I) f_\alpha = A^T g . \]
Tikhonov Regularisation

Regularisation parameter $\alpha$

Regularised solution of the form

$$f_\alpha = \sum_{i=1}^{r} \frac{\sigma_i^2}{\sigma_i^2 + \alpha^2} \frac{u_i^T g}{\sigma_i} v_i$$

Filter factor $\frac{\sigma_i^2}{\sigma_i^2 + \alpha^2}$ as diagonal entries of the filter matrix $\Psi$
Tikhonov Regularisation

Regularisation parameter $\alpha$

Regularised solution of the form

$$f_\alpha = \sum_{i=1}^r \frac{\sigma_i^2}{\sigma_i^2 + \alpha^2} \frac{u_i^T g}{\sigma_i} v_i$$

Filter factor $\frac{\sigma_i^2}{\sigma_i^2 + \alpha^2}$ as diagonal entries of the filter matrix $\Psi$

Regularisation and perturbation error

$$f_\alpha = V\Psi U^T g, \quad g = g_{\text{exact}} + e$$

$$= V\Psi \Sigma^{-1} U^T g_{\text{exact}} + V\Psi \Sigma^{-1} U^T e$$

$$= V\Psi \Sigma^{-1} U^T U \Sigma V^T f_{\text{exact}} + V\Psi \Sigma^{-1} U^T e$$

$$= V\Psi V^T f_{\text{exact}} + V\Psi \Sigma^{-1} U^T e$$
Tikhonov Regularisation

Regularisation parameter $\alpha$

Regularised solution of the form

$$f_\alpha = \sum_{i=1}^{r} \frac{\sigma_i^2}{\sigma_i^2 + \alpha^2} \frac{u_i^T g}{\sigma_i} v_i$$

Filter factor $\frac{\sigma_i^2}{\sigma_i^2 + \alpha^2}$ as diagonal entries of the filter matrix $\Psi$

Regularisation and perturbation error

$$f_\alpha = V\Psi U^T g, \quad g = g_{\text{exact}} + e$$

$$= V\Psi \Sigma^{-1} U^T g_{\text{exact}} + V\Psi \Sigma^{-1} U^T e$$

$$= V\Psi \Sigma^{-1} U^T U \Sigma V^T f_{\text{exact}} + V\Psi \Sigma^{-1} U^T e$$

$$= V\Psi V^T f_{\text{exact}} + V\Psi \Sigma^{-1} U^T e$$

$$f_{\text{exact}} - f_\alpha =$$
Tikhonov Regularisation

Regularisation parameter $\alpha$

Regularised solution of the form

$$f_\alpha = \sum_{i=1}^{r} \frac{\sigma_i^2}{\sigma_i^2 + \alpha^2} u_i^T g v_i$$

Filter factor $\frac{\sigma_i^2}{\sigma_i^2 + \alpha^2}$ as diagonal entries of the filter matrix $\Psi$

Regularisation and perturbation error

$$f_\alpha = V\Psi U^T g, \quad g = g_{\text{exact}} + e$$
$$= V\Psi \Sigma^{-1} U^T g_{\text{exact}} + V\Psi \Sigma^{-1} U^T e$$
$$= V\Psi \Sigma^{-1} U^T U \Sigma V^T f_{\text{exact}} + V\Psi \Sigma^{-1} U^T e$$
$$= V\Psi V^T f_{\text{exact}} + V\Psi \Sigma^{-1} U^T e$$

$$f_{\text{exact}} - f_\alpha = (I - V\Psi V^T) f_{\text{exact}} -$$

Regularisation error
Tikhonov Regularisation

Regularisation parameter $\alpha$

Regularised solution of the form

$$f_\alpha = \sum_{i=1}^{r} \frac{\sigma_i^2}{\sigma_i^2 + \alpha^2} \frac{u_i^T g}{\sigma_i} v_i$$

Filter factor $\frac{\sigma_i^2}{\sigma_i^2 + \alpha^2}$ as diagonal entries of the filter matrix $\Psi$

Regularisation and perturbation error

$$f_\alpha = V\Psi U^T g, \quad g = g_{\text{exact}} + e$$
$$= V\Psi \Sigma^{-1} U^T g_{\text{exact}} + V\Psi \Sigma^{-1} U^T e$$
$$= V\Psi \Sigma^{-1} U^T U\Sigma V^T f_{\text{exact}} + V\Psi \Sigma^{-1} U^T e$$
$$= V\Psi V^T f_{\text{exact}} + V\Psi \Sigma^{-1} U^T e$$

$$f_{\text{exact}} - f_\alpha = (I - V\Psi V^T) f_{\text{exact}} - V\Psi \Sigma^{-1} U^T e$$

Regularisation error

Perturbation error
Tikhonov Regularisation

Regularisation and perturbation error

\[ f_{\text{exact}} - f_{\alpha} = (I - V\Psi V^T)f_{\text{exact}} - V\Psi \Sigma^{-1}U^Te \]

Figure: Regularisation and perturbation error
Tikhonov Regularisation

Illustrative example

Figure: $\alpha$ too small
Tikhonov Regularisation

Illustrative example

Figure: $\alpha$ too large
Tikhonov Regularisation

Illustrative example

Figure: Good Value for $\alpha$
Outline

Inverse Problems

Data Assimilation as a Large Inverse Problem

Regularisation Parameter estimation in 4DVar
  Regularisation Parameter estimation
  Example

Application of $L_1$-norm regularisation in 4DVar
  Motivation: Results from image processing
  $L_1$-norm regularisation in 4DVar
  Examples
Data Assimilation in NWP

Find an estimate $x_i$ at time $i$ for the true state of the atmosphere $x_i^{\text{Truth}}$.

Observations $y_i$

- Satellites
- Ships and buoys
- Surface stations
- Planes
Data Assimilation in NWP

Find an estimate $x_i$ at time $i$ for the true state of the atmosphere $x_i^{\text{Truth}}$.

A priori information $x_i^B$

- background state (previous forecast)

Observations $y_i$

- Satellites
- Ships and buoys
- Surface stations
- Planes
Data Assimilation in NWP

Find an estimate $\mathbf{x}_i$ at time $i$ for the true state of the atmosphere $\mathbf{x}_i^{\text{Truth}}$.

A priori information $\mathbf{x}_i^B$

- background state (previous forecast)

Models

- an operator linking state space and observation space (imperfect)

\[ \mathbf{y}_i = H_i(\mathbf{x}_i) \]

Observations $\mathbf{y}_i$

- Satellites
- Ships and buoys
- Surface stations
- Planes
Data Assimilation in NWP

Find an estimate $x_i$ at time $i$ for the true state of the atmosphere $x_{i}^{\text{Truth}}$.

A priori information $x_i^B$

- background state (previous forecast)

Models

- an operator linking state space and observation space (imperfect)
  
  \[ y_i = H_i(x_i) \]

- a model for the atmosphere (imperfect)
  
  \[ x_{i+1} = M_{i+1,i}(x_i) \]

Observations $y_i$

- Satellites
- Ships and buoys
- Surface stations
- Planes
Data Assimilation in NWP

Find an estimate $x_i$ at time $i$ for the true state of the atmosphere $x_i^{\text{Truth}}$.

A priori information $x_i^B$
- background state (previous forecast)

Models
- an operator linking state space and observation space (imperfect)
  \[ y_i = H_i(x_i) \]
- a model for the atmosphere (imperfect)
  \[ x_{i+1} = M_{i+1,i}(x_i) \]

Observations $y_i$
- Satellites
- Ships and buoys
- Surface stations
- Planes

Assimilation algorithms
- find an (approximate) state of the atmosphere $x_i$ at times $i$ (usually $i = 0$)
- $x_i^A$: Analysis (estimation of the true state after the DA)
- forecast future states of the atmosphere
Schematics of Data Assimilation

Figure: Background state $\mathbf{x}^B$
Schematics of Data Assimilation

Figure: Observations $y$
Schematics of Data Assimilation

Figure: Analysis $x^A$ (consistent with observations and model dynamics)
Observations

ECMWF Data Coverage (All obs DA) - SYNOP/SHIP
21/APR/2008; 00 UTC
Total number of obs = 26683

ECMWF Data Coverage (All obs DA) - BUOY
21/APR/2008; 00 UTC
Total number of obs = 7438

ECMWF Data Coverage (All obs DA) - AIRCRAFT
21/APR/2008; 00 UTC
Total number of obs = 51809

ECMWF Data Coverage (All obs DA) - ATOVS
21/APR/2008; 00 UTC
Total number of obs = 341239
Data Assimilation in NWP

Under-determinacy

- Size of the state vector \( \mathbf{x} \): \( 432 \times 320 \times 50 \times 7 = \mathcal{O}(10^7) \)
- Number of observations (size of \( \mathbf{y} \)): \( \mathcal{O}(10^5 - 10^6) \)
Data Assimilation in NWP

Under-determinacy

- Size of the state vector $\mathbf{x}$: $432 \times 320 \times 50 \times 7 = \mathcal{O}(10^7)$
- Number of observations (size of $\mathbf{y}$): $\mathcal{O}(10^5 - 10^6)$

Assumptions

- background error $\epsilon^B = \mathbf{x}^B - \mathbf{x}^{\text{Truth}}$ and covariance matrix $\mathbf{B} = (\epsilon^B - \bar{\epsilon}^B)(\epsilon^B - \bar{\epsilon}^B)^T$
- observation error $\epsilon^O = \mathbf{y} - H(\mathbf{x}^{\text{Truth}})$ and covariance matrix $\mathbf{R} = (\epsilon^O - \bar{\epsilon}^O)(\epsilon^O - \bar{\epsilon}^O)^T$
- Non-trivial errors: $\mathbf{B}$, $\mathbf{R}$ are positive definite
- Uncorrelated errors: $(\mathbf{x}^B - \mathbf{x}^{\text{Truth}})(\mathbf{y} - H(\mathbf{x}^{\text{Truth}}))^T = 0$
Optimal least-squares estimator

Cost function
Solution to the optimisation problem \( x^A = \underset{x}{\arg \min} J(x) \) where

\[
J(x) = \frac{1}{2} (x - x^B)^T B^{-1} (x - x^B) + \frac{1}{2} (y - H(x))^T R^{-1} (y - H(x))
\]

\[
= J_B(x) + J_O(x)
\]

\( \Rightarrow \) Three-dimensional variational data assimilation (3DVar)
Optimal least-squares estimator

Cost function
Solution to the optimisation problem \( \mathbf{x}^A = \text{arg min} \ J(\mathbf{x}) \) where

\[
J(\mathbf{x}) = \frac{1}{2} (\mathbf{x} - \mathbf{x}^B)^T \mathbf{B}^{-1} (\mathbf{x} - \mathbf{x}^B) + \frac{1}{2} (\mathbf{y} - H(\mathbf{x}))^T \mathbf{R}^{-1} (\mathbf{y} - H(\mathbf{x}))
\]

\[
= J_B(\mathbf{x}) + J_O(\mathbf{x})
\]

\( \Rightarrow \)Three-dimensional variational data assimilation (3DVar)

Interpolation equations

\[
\mathbf{x}^A = \mathbf{x}^B + \mathbf{K}(\mathbf{y} - H(\mathbf{x}^B)), \quad \text{where}
\]

\[
\mathbf{K} = \mathbf{B}\mathbf{H}^T (\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1} \quad \mathbf{K} \ldots \text{gain matrix}
\]

\( \Rightarrow \) Optimal interpolation
Bayesian interpretation

Non-Gaussian PDF’s (probability density function)

- $P(x)$ is a priori PDF (background)
- $P(y|x)$ is the observation PDF (likelihood of the observations given background $x$)
Bayesian interpretation

Non-Gaussian PDF’s (probability density function)

- $P(x)$ is a priori PDF (background)
- $P(y|x)$ is the observation PDF (likelihood of the observations given background $x$)
- $P(x|y)$ conditional probability of the model state given the observations, Bayes theorem:

$$\arg_x \max P(x|y) = \arg_x \max \frac{P(y|x)P(x)}{P(y)}$$
Bayesian interpretation

Non-Gaussian PDF’s (probability density function)

- $P(x)$ is a priori PDF (background)
- $P(y|x)$ is the observation PDF (likelihood of the observations given background $x$)
- $P(x|y)$ conditional probability of the model state given the observations, Bayes theorem:

  \[
  \arg_x \max P(x|y) = \arg_x \max \frac{P(y|x)P(x)}{P(y)}
  \]

Gaussian PDF’s

\[
P(x|y) = c_1 \exp \left( -(x - x^B)^T B^{-1} (x - x^B) \right) \cdot \nonumber \\
\quad c_2 \exp \left( -(y - H(x))^T R^{-1} (y - H(x)) \right)
\]

$x^A$ is the maximum a posteriori estimator of $x^{Truth}$. 

Bayesian interpretation

Non-Gaussian PDF’s (probability density function)

- \( P(x) \) is a priori PDF (background)
- \( P(y|x) \) is the observation PDF (likelihood of the observations given background \( x \))
- \( P(x|y) \) conditional probability of the model state given the observations, Bayes theorem:

\[
\arg_x \max P(x|y) = \arg_x \max \frac{P(y|x)P(x)}{P(y)}
\]

Gaussian PDF’s

\[
P(x|y) = c_1 \exp \left( -(x - x^B)^T B^{-1} (x - x^B) \right) \cdot c_2 \exp \left( -(y - H(x))^T R^{-1} (y - H(x)) \right)
\]

\( x^A \) is the maximum a posteriori estimator of \( x^{\text{Truth}} \).

Maximising \( P(x|y) \) equivalent to minimising \( J(x) \)
Four-dimensional variational assimilation (4DVar)

Minimise the cost function

\[ J(x_0) = \frac{1}{2} (x_0 - x_0^B)^T B^{-1} (x_0 - x_0^B) + \frac{1}{2} \sum_{i=0}^{n} (y_i - H_i(x_i))^T R_i^{-1} (y_i - H_i(x_i)) \]

subject to model dynamics \( x_i = M_{i,0} x_0 \).

Figure: Copyright: ECMWF
Four-dimensional variational assimilation (4DVar)

Minimise the cost function

\[ J(x_0) = \frac{1}{2} (x_0 - x_0^B)^T B^{-1} (x_0 - x_0^B) + \frac{1}{2} \sum_{i=0}^{n} (y_i - H_i(x_i))^T R_i^{-1} (y_i - H_i(x_i)) \]

subject to model dynamics \( x_i = M_{i,0} x_0 \).
Four-dimensional variational assimilation (4DVar)

Minimise the cost function

$$J(x_0) = \frac{1}{2}(x_0 - x_0^B)^T B^{-1}(x_0 - x_0^B) + \frac{1}{2} \sum_{i=0}^{n} (y_i - H_i(x_i))^T R_i^{-1}(y_i - H_i(x_i))$$

subject to model dynamics $x_i = M_{i,0}x_0$. 

Figure: Copyright: ECMWF
Minimisation of the 4DVar cost function

- Use Newton’s method in order to solve $\nabla J(x_0) = 0$, that is

$$\nabla^2 J(x_0^k) \Delta x_0^k = -\nabla J(x_0^k)$$

$$x_0^{k+1} = x_0^k + \Delta x_0^k$$

$k \geq 0$
Minimisation of the 4DVar cost function

- Use **Newton’s method** in order to solve $\nabla J(x_0) = 0$, that is
  \[
  \nabla \nabla J(x_0^k) \Delta x_0^k = -\nabla J(x_0^k)
  \]
  \[
  x_0^{k+1} = x_0^k + \Delta x_0^k
  \]
  \(k \geq 0\)

- Use approximate Hessian - **Gauß-Newton method**

  \[
  \nabla J(x_0) = B^{-1}(x_0 - x_0^B) - \sum_{i=1}^{n} M_{i,0}(x_0)^T H_i^T R_i^{-1}(y_i - H_i(x_i)),
  \]

  and

  \[
  \nabla \nabla J(x_0) = B^{-1} + \sum_{i=1}^{n} M_{i,0}(x_0)^T H_i^T R_i^{-1} H_i M_{i,0}(x_0).
  \]
Relation between 4DVar and Tikhonov regularisation

4DVar minimises

\[ J(x_0) = \frac{1}{2} (x_0 - x_0^B)^T B^{-1} (x_0 - x_0^B) + \frac{1}{2} \sum_{i=0}^{n} (y_i - H_i(x_i))^T R_i^{-1} (y_i - H_i(x_i)) \]

subject to model dynamics \( x_i = M_{0 \rightarrow i} x_0 \)

Melina Freitag

Tikhonov Regularisation for (Large) Inverse Problems
Relation between 4DVar and Tikhonov regularisation

4DVar minimises

\[ J(x_0) = \frac{1}{2}(x_0 - x_B^0)^T B^{-1}(x_0 - x_B^0) + \frac{1}{2} \sum_{i=0}^{n} (y_i - H_i(x_i))^T R_i^{-1}(y_i - H_i(x_i)) \]

subject to model dynamics \( x_i = M_{0 \rightarrow i} x_0 \)

or

\[ J(x_0) = \frac{1}{2}(x_0 - x_B^0)^T B^{-1}(x_0 - x_B^0) + \frac{1}{2}(\hat{y} - \hat{H}(x_0))^T \hat{R}^{-1}(\hat{y} - \hat{H}(x_0)) \]

where

\[ \hat{H} = [H_0^T, (H_1 M_{10}(t_1, t_0))^T, \ldots, (H_n M_{n0}(t_n, t_0))^T]^T \]

\[ \hat{y} = [y_0^T, \ldots, y_n^T]^T \]

and \( \hat{R} \) is block diagonal with \( R_i, i = 0, \ldots, n \) on the diagonal.
Relation between 4DVar and Tikhonov regularisation

Solution to the optimisation problem

Cost function

\[ J(x_0) = \frac{1}{2} (x_0 - x_0^B)^T B^{-1} (x_0 - x_0^B) + \frac{1}{2} (\hat{y} - \hat{H}(x_0))^T \hat{R}^{-1} (\hat{y} - \hat{H}(x_0)) \]
Relation between 4DVar and Tikhonov regularisation

Solution to the optimisation problem

Cost function

\[ J(x_0) = \frac{1}{2} (x_0 - x_0^B)^T B^{-1} (x_0 - x_0^B) + \frac{1}{2} (\hat{y} - \hat{H}(x_0))^T \hat{R}^{-1} (\hat{y} - \hat{H}(x_0)) \]

Gauß-Newton method

\[ \nabla \nabla J(x_0^k) \Delta x_0^k = - \nabla J(x_0^k) \]
\[ x_0^{k+1} = x_0^k + \Delta x_0^k \]
Relation between 4DVar and Tikhonov regularisation

Solution to the optimisation problem

Cost function

\[ J(x_0) = \frac{1}{2} (x_0 - x_0^B)^T B^{-1} (x_0 - x_0^B) + \frac{1}{2} (\hat{y} - \hat{H}(x_0))^T \hat{R}^{-1} (\hat{y} - \hat{H}(x_0)) \]

Gauß-Newton method

\[
(B^{-1} + \hat{H}^T \hat{R}^{-1} \hat{H}) \Delta x_0^k = -B^{-1} (x_0^k - x_0^B) + \hat{H}^T \hat{R}^{-1} (\hat{y} - \hat{H}(x_0)) \\
x_0^{k+1} = x_0^k + \Delta x_0^k
\]
Relation between 4DVar and Tikhonov regularisation

Variable transform
Set

\[
(B^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H}) \Delta x_0^k = -B^{-1}(x_0^k - x_0^B) + \mathbf{H}^T \mathbf{R}^{-1}(\hat{y} - \hat{H}(x_0))
\]

\[
x_0^{k+1} = x_0^k + \Delta x_0^k
\]

Gauß-Newton method
Relation between 4DVar and Tikhonov regularisation

Variable transform

Set

\[ B = \sigma_B^2 C_B \]
\[ \hat{R} = \sigma_R^2 C_R \]

Gauß-Newton method

\[
(B^{-1} + \hat{H}^T \hat{R}^{-1} \hat{H}) \Delta x_0^k = -B^{-1}(x_0^k - x_0^B) + \hat{H}^T \hat{R}^{-1}(\hat{y} - \hat{H}(x_0))
\]
\[
x_0^{k+1} = x_0^k + \Delta x_0^k
\]
Relation between 4DVar and Tikhonov regularisation

Variable transform

Set

\[ B = \sigma_B^2 C_B \]
\[ \hat{R} = \sigma_R^2 C_R \]
\[ b = C_R^{-\frac{1}{2}} (\hat{y} - \hat{H}(x_0)) \]

Gauß-Newton method

\[
\begin{align*}
(B^{-1} + \hat{H}^T \hat{R}^{-1} \hat{H}) \Delta x_0^k &= -B^{-1}(x_0^k - x_0^B) + \hat{H}^T \hat{R}^{-1}(\hat{y} - \hat{H}(x_0)) \\
x_0^{k+1} &= x_0^k + \Delta x_0^k
\end{align*}
\]
Relation between 4DVar and Tikhonov regularisation

Variable transform

Set

\[
\begin{align*}
B &= \sigma_B^2 C_B \\
\hat{R} &= \sigma_R^2 C_R \\
b &= C_R^{-\frac{1}{2}} (\hat{y} - \hat{H}(x_0)) \\
A &= C_R^{-\frac{1}{2}} \hat{H} C_B^{\frac{1}{2}}
\end{align*}
\]

Gauß-Newton method

\[
\begin{align*}
(B^{-1} + \hat{H}^T \hat{R}^{-1} \hat{H}) \Delta x_0^k &= -B^{-1}(x_0^k - x_0^B) + \hat{H}^T \hat{R}^{-1}(\hat{y} - \hat{H}(x_0)) \\
x_0^{k+1} &= x_0^k + \Delta x_0^k
\end{align*}
\]
Relation between 4DVar and Tikhonov regularisation

Variable transform
Set

\[
\begin{align*}
B &= \sigma_B^2 C_B \\
\hat{R} &= \sigma_R^2 C_R \\
b &= C_R^{-\frac{1}{2}} (\hat{y} - \hat{H}(x_0)) \\
A &= C_R^{-\frac{1}{2}} \hat{H} C_B^{\frac{1}{2}} \\
\alpha^2 &= \frac{\sigma_R^2}{\sigma_B^2}
\end{align*}
\]

Gauß-Newton method

\[
\begin{align*}
(B^{-1} + \hat{H}^T \hat{R}^{-1} \hat{H}) \Delta x_0^k &= -B^{-1}(x_0^k - x_0^B) + \hat{H}^T \hat{R}^{-1}(\hat{y} - \hat{H}(x_0)) \\
x_0^{k+1} &= x_0^k + \Delta x_0^k
\end{align*}
\]
Relation between 4DVar and Tikhonov regularisation

Variable transform

Set

\[
\begin{align*}
B &= \sigma_B^2 C_B \\
\hat{R} &= \sigma_R^2 C_R \\
b &= C_R^{-\frac{1}{2}} (\hat{y} - \hat{H}(x_0)) \\
A &= C_R^{-\frac{1}{2}} \hat{H} C_B^{\frac{1}{2}} \\
\alpha^2 &= \frac{\sigma_R^2}{\sigma_B^2}
\end{align*}
\]

Gauß-Newton method

\[
(\alpha^2 I + A^T A) C_B^{-\frac{1}{2}} \Delta x_0^k = -\alpha^2 C_B^{-\frac{1}{2}} (x_0^k - x_0^B) + A^T b \\
x_0^{k+1} = x_0^k + \Delta x_0^k
\]
Relation between 4DVar and Tikhonov regularisation

Variable transform

\[ z^k = C_B^{-\frac{1}{2}} (x_0^k - x_0^B) \]

Gauß-Newton method

\[ (\alpha^2 I + A^T A)C_B^{-\frac{1}{2}} \Delta x_0^k = -\alpha^2 C_B^{-\frac{1}{2}} (x_0^k - x_0^B) + A^T b \]

\[ x_0^{k+1} = x_0^k + \Delta x_0^k \]
Relation between 4DVar and Tikhonov regularisation

Variable transform
Set

\[ z^k = C_B^{-\frac{1}{2}} (x_0^k - x_0^B) \]

Gauß-Newton method

\[ (\alpha^2 I + A^T A)(z^{k+1} - z^k) = -\alpha^2 z^k + A^T b \]
Relation between 4DVar and Tikhonov regularisation

Variable transform
Set
\[ z^k = C^{-\frac{1}{2}}_B (x^k_0 - x^B_0) \]

Gauß-Newton method

\[ (\alpha^2 I + A^T A)(z^{k+1} - z^k) = -\alpha^2 z^k + A^T b \]

Normal equations
Relation between 4DVar and Tikhonov regularisation

Variable transform
Set

\[ z^k = C_B^{-\frac{1}{2}} (x_0^k - x_0^B) \]

Gauß-Newton method

\[
(\alpha^2 I + A^T A)(z^{k+1} - z^k) = -\alpha^2 z^k + A^T b
\]

Normal equations

Least squares solution

\[
\left\| \begin{bmatrix} A \\ \alpha I \end{bmatrix} (z^{k+1} - z^k) + \begin{bmatrix} -b \\ \alpha z^k \end{bmatrix} \right\|_2^2 \rightarrow \min
\]

at each Gauß-Newton method step
Relation between 4DVar and Tikhonov regularisation

Variable transform
Set
\[ z^k = C^{-\frac{1}{2}}_B (x_0^k - x_0^B) \]

Gauß-Newton method

\[
(\alpha^2 I + A^T A)(z^{k+1} - z^k) = -\alpha^2 z^k + A^T b
\]

Normal equations

Least squares solution

\[
\left\| \begin{bmatrix} A & \alpha I \\ \alpha I & \alpha I \end{bmatrix} \right\| (z^{k+1} - z^k) + \left\| \begin{bmatrix} -b \\ -\alpha z^k \end{bmatrix} \right\|^2 \rightarrow \text{min}
\]

at each Gauß-Newton method step or

\[
\|Az^{k+1} - (Az^k + b)\|_2^2 + \alpha^2 \|z^{k+1}\|_2^2
\]

Tikhonov regularisation
Relation between 4DVar and Tikhonov regularisation

Variable transform

Set

\[ z^k = C_B^{-\frac{1}{2}}(x_0^k - x_0^B) \]

Gauß-Newton method

\[
(\alpha^2 I + A^T A)(z^{k+1} - z^k) = -\alpha^2 z^k + A^T b
\]

Normal equations

Least squares solution

\[
\left\| \begin{bmatrix} A & \alpha I \\ \alpha I & -\alpha z^k \end{bmatrix} (z^{k+1} - z^k) + \begin{bmatrix} -b \\ \alpha z^k \end{bmatrix} \right\|_2^2 \rightarrow \min
\]

at each Gauß-Newton method step or

\[
\left\| A z^{k+1} - g \right\|_2^2 + \alpha^2 \left\| z^{k+1} \right\|_2^2
\]

Tikhonov regularisation
Summary

Minimising the cost function

\[ J(x_0) = \frac{1}{2}(x_0 - x_0^B)^T B^{-1}(x_0 - x_0^B) + \frac{1}{2}(\hat{y} - \hat{H}(x_0))^T \hat{R}^{-1}(\hat{y} - \hat{H}(x_0)) \]

amounts to solving a Tikhonov regularised least squares problem at every step

\[ \|Az^{k+1} - g\|^2_2 + \alpha^2\|z^{k+1}\|^2_2 \]
Data Assimilation and Tikhonov regularisation

**Issues**

- Dynamic vs static
Data Assimilation and Tikhonov regularisation

Issues

- Dynamic vs static
- Nonlinear Dynamics (and chaotic)
Data Assimilation and Tikhonov regularisation

Issues

- Dynamic vs static
- Nonlinear Dynamics (and chaotic)
- Multiscale and Large Scale
Data Assimilation and Tikhonov regularisation

Issues

- Dynamic vs static
- Nonlinear Dynamics (and chaotic)
- Multiscale and Large Scale
- Many unknown parameters ($B$ ...)

$L_1$-norm regularisation

Outline Inverse Problems Data Assimilation Regularisation Parameter
Data Assimilation and Tikhonov regularisation

Issues

- Dynamic vs static
- Nonlinear Dynamics (and chaotic)
- Multiscale and Large Scale
- Many unknown parameters ($B \ldots$)
- Model error
Data Assimilation and Tikhonov regularisation

Issues

- Dynamic vs static
- Nonlinear Dynamics (and chaotic)
- Multiscale and Large Scale
- Many unknown parameters ($B$ ...)
- Model error

Previous/current work
Data Assimilation and Tikhonov regularisation

Issues

- Dynamic vs static
- Nonlinear Dynamics (and chaotic)
- Multiscale and Large Scale
- Many unknown parameters ($B ...$)
- Model error

Previous/current work

- Low Rank Kalman Filters [Houtekamer, Mitchell 1998]
Data Assimilation and Tikhonov regularisation

Issues

• Dynamic vs static
• Nonlinear Dynamics (and chaotic)
• Multiscale and Large Scale
• Many unknown parameters ($B \ldots$)
• Model error

Previous/current work

• Low Rank Kalman Filters [Houtekamer, Mitchell 1998]
Data Assimilation and Tikhonov regularisation

Issues

- Dynamic vs static
- Nonlinear Dynamics (and chaotic)
- Multiscale and Large Scale
- Many unknown parameters ($B$ ...)
- Model error

Previous/current work

- Low Rank Kalman Filters [Houtekamer, Mitchell 1998]
- Preconditioners [Haben, Lawless, Nichols 2010]
**Data Assimilation and Tikhonov regularisation**

**Issues**

- Dynamic vs static
- Nonlinear Dynamics (and chaotic)
- Multiscale and Large Scale
- Many unknown parameters \((B \ldots)\)
- Model error

**Previous/current work**

- Low Rank Kalman Filters [Houtekamer, Mitchell 1998]
- Preconditioners [Haben, Lawless, Nichols 2010]
- *Regularisation Parameter Estimation*
Data Assimilation and Tikhonov regularisation

Issues

- Dynamic vs static
- Nonlinear Dynamics (and chaotic)
- Multiscale and Large Scale
- Many unknown parameters ($B \ldots$)
- Model error

Previous/current work

- Low Rank Kalman Filters [Houtekamer, Mitchell 1998]
- Preconditioners [Haben, Lawless, Nichols 2010]
- Regularisation Parameter Estimation
- $L_1$-Norm Regularisation
Outline

Inverse Problems

Data Assimilation as a Large Inverse Problem

Regularisation Parameter estimation in 4DVar
  Regularisation Parameter estimation
  Example

Application of $L_1$-norm regularisation in 4DVar
  Motivation: Results from image processing
  $L_1$-norm regularisation in 4DVar
  Examples
Choosing the regularisation parameter $\alpha$

$$\|Af - g\|_2^2 + \alpha^2 \|f\|_2^2$$
Generalised Cross-Validation (Golub, Heath, Wahba 1979) GCV

- If a data value is omitted, then a good choice of the reconstruction should be able to predict the missing data value as well
Generalised Cross-Validation (Golub, Heath, Wahba 1979) GCV

- If a data value is omitted, then a good choice of the reconstruction should be able to predict the missing data value as well.
- Minimise the GCV functional

\[
G(\alpha) = \frac{\| (I - AV\Psi\Sigma^{-1}U^T)g \|^2}{(\text{trace}(I - AV\Psi\Sigma^{-1}U^T))^2}
\]
Generalised Cross-Validation (Golub, Heath, Wahba 1979) GCV

- If a data value is omitted, then a good choice of the reconstruction should be able to predict the missing data value as well
- Minimise the GCV functional

\[
G(\alpha) = \frac{\| (I - AV\Psi\Sigma^{-1}U^T)g \|^2}{\text{trace}(I - AV\Psi\Sigma^{-1}U^T))^2}
\]

\[
G(\alpha) = \frac{\sum_{i=1}^{N} \left( \frac{u_i^T g}{\sigma_i^2 + \alpha^2} \right)^2}{\left( \sum_{i=1}^{N} \frac{1}{\sigma_i^2 + \alpha^2} \right)^2}
\]

Melina Freitag
Generalised Cross-Validation (Golub, Heath, Wahba 1979) GCV

Figure: Parameter estimation using GCV
L-Curve Criterion (Hansen 1992)

- Log-log plot of the norm of the regularised solution $\|f\|$ versus the corresponding residual norm $\|Af - g\|$
L-Curve Criterion (Hansen 1992)

- Log-log plot of the norm of the regularised solution $\|f\|$ versus the corresponding residual norm $\|Af - g\|
- Best value of $\alpha$ determined by maximum curvature

$$R(\alpha) = \log \|Af_\alpha - g\|^2 \quad S(\alpha) = \log \|f_\alpha\|^2$$

$$k(\alpha) = \frac{R''(\alpha)S'(\alpha) - R'(\alpha)S''(\alpha)}{(R'(\alpha)^2 + S'(\alpha))^{3/2}}$$
L-Curve Criterion (Hansen 1992)

Figure: Parameter estimation using L-Curve
Discrepancy Principle (Morozov 1966)

- Choose a regularised solution such that

\[ \|g - A f_\alpha\|_2 = \tau \delta \]

where \(2 \leq \tau \leq 5\)
Discrepancy Principle (Morozov 1966)

- Choose a regularised solution such that
  \[ \| g - Af_{\alpha} \|_2 = \tau \delta \]
  where \( 2 \leq \tau \leq 5 \)
- \( \delta \) is the expected value of the error \( \| e \| \)
Discrepancy Principle (Morozov 1966)

- Choose a regularised solution such that
  
  \[ \| g - Af_\alpha \|_2 = \tau \delta \]

  where \( 2 \leq \tau \leq 5 \)

- \( \delta \) is the expected value of the error \( \| e \| \)

- Apply iterative method to \( g = Af \)
Discrepancy Principle (Morozov 1966)

• Choose a regularised solution such that

\[ \| g - Af_\alpha \|_2 = \tau \delta \]

where \( 2 \leq \tau \leq 5 \)

• \( \delta \) is the expected value of the error \( \| e \| \)

• Apply iterative method to \( g = Af \)

• First steps: reduce the residual error in the singular direction associated with larger singular values
Discrepancy Principle (Morozov 1966)

- Choose a regularised solution such that
  \[ \| \mathbf{g} - A\mathbf{f}_\alpha \|_2 = \tau\delta \]

  where \( 2 \leq \tau \leq 5 \)

- \( \delta \) is the expected value of the error \( \| \mathbf{e} \| \)

- Apply iterative method to \( \mathbf{g} = A\mathbf{f} \)

- First steps: reduce the residual error in the singular direction associated with larger singular values

- Latter steps: singular direction associated to smaller singular values are fitted - truncate the iteration before the amplified noise takes over
Discrepancy Principle (Morozov 1966)

Figure: Parameter estimation using Discrepancy Principle
The system is given by

\[ \frac{dX_i}{dt} = -X_{i-2}X_{i-1} + X_{i-1}X_{i+1} - X_i + F, \quad i = 1, \ldots, N, \]

cyclic boundary conditions \( X_0 = X_N, X_{-1} = X_{N+1}, X_{N+1} = X_1. \)

- \( F = 8, \ N = 40 \) (13 positive Lyapunov exponents).
- solver: Runge-Kutta method with time step \( h = 0.01 \)
N-dimensional (chaotic) Lorenz system (Lorenz-95)

The system is given by

$$\frac{dX_i}{dt} = -X_{i-2}X_{i-1} + X_{i-1}X_{i+1} - X_i + F, \quad i = 1, \ldots, N,$$

cyclic boundary conditions $X_0 = X_N, X_{-1} = X_{N+1}, X_{N+1} = X_1$.

- $F = 8, \ N = 40$ (13 positive Lyapunov exponents).
- solver: Runge-Kutta method with time step $h = 0.01$
- a unit time $T = 1$ is associated with 5 days
- assimilation window: 5 time steps (associated with 6 hours)
- subsequent forecast: 95 time steps (associated with 5 day forecast)
N-dimensional (chaotic) Lorenz system (Lorenz-95)

The system is given by

\[ \frac{dX_i}{dt} = -X_{i-2}X_{i-1} + X_{i-1}X_{i+1} - X_i + F, \quad i = 1, \ldots, N, \]

cyclic boundary conditions \( X_0 = X_N, \ X_{-1} = X_{N+1}, \ X_{N+1} = X_1. \)

- \( F = 8, \ N = 40 \) (13 positive Lyapunov exponents).
- solver: Runge-Kutta method with time step \( h = 0.01 \)
- a unit time \( T = 1 \) is associated with 5 days
- assimilation window: 5 time steps (associated with 6 hours)
- subsequent forecast: 95 time steps (associated with 5 day forecast)
- observations are taken as noise from the truth trajectory
N-dimensional (chaotic) Lorenz system (Lorenz-95)

The system is given by

\[ \frac{dX_i}{dt} = -X_{i-2}X_{i-1} + X_{i-1}X_{i+1} - X_i + F, \quad i = 1, \ldots, N, \]

cyclic boundary conditions \( X_0 = X_N, X_{-1} = X_{N+1}, X_{N+1} = X_1. \)

- \( F = 8, \) \( N = 40 \) (13 positive Lyapunov exponents).
- solver: Runge-Kutta method with time step \( h = 0.01 \)
- a unit time \( T = 1 \) is associated with 5 days
- assimilation window: 5 time steps (associated with 6 hours)
- subsequent forecast: 95 time steps (associated with 5 day forecast)
- observations are taken as noise from the truth trajectory
- Model error introduced by parameter change \( F_{mod} = 12. \)
Lorenz-95 dynamics

The system is given by

$$\frac{dX_i}{dt} = -X_{i-2}X_{i-1} + X_{i-1}X_{i+1} - X_i + F, \quad i = 1, \ldots, N,$$

cyclic boundary conditions $X_0 = X_N$, $X_{-1} = X_{N+1}$, $X_{N+1} = X_1$. 

![Graph of Lorenz-95 dynamics](image)
Lorenz-95 dynamics

The system is given by

\[
\frac{dX_i}{dt} = -X_{i-2}X_{i-1} + X_{i-1}X_{i+1} - X_i + F, \quad i = 1, \ldots, N,
\]

cyclic boundary conditions \( X_0 = X_N, X_{-1} = X_{N+1}, X_{N+1} = X_1. \)
Initial condition error

<table>
<thead>
<tr>
<th>Observation frequency</th>
<th>4DVAR</th>
<th>Discrepancy Principle</th>
<th>GCV</th>
<th>L-Curve</th>
</tr>
</thead>
<tbody>
<tr>
<td>every 10 points</td>
<td>0.7729</td>
<td>0.7608</td>
<td>0.7394</td>
<td>0.8101</td>
</tr>
<tr>
<td>every 5 points</td>
<td>0.8043</td>
<td>0.6725</td>
<td>0.6510</td>
<td>0.7727</td>
</tr>
<tr>
<td>every 2 points</td>
<td>0.5492</td>
<td>0.3309</td>
<td>0.2812</td>
<td>0.4469</td>
</tr>
</tbody>
</table>

**Table:** Comparison RMS error - no model error in the Lorenz system
Comparison - no model error in the Lorenz system

Figure: 4DVAR

Figure: Generalised Cross-Validation
Initial condition error

<table>
<thead>
<tr>
<th>Observation frequency</th>
<th>4DVAR</th>
<th>Discrepancy Principle</th>
<th>GCV</th>
<th>L-Curve</th>
</tr>
</thead>
<tbody>
<tr>
<td>every 10 points</td>
<td>3.4641</td>
<td>3.4156</td>
<td>6.1941</td>
<td>0.8579</td>
</tr>
<tr>
<td>every 5 points</td>
<td>5.3430</td>
<td>4.4666</td>
<td>6.0010</td>
<td>0.8651</td>
</tr>
<tr>
<td>every 2 points</td>
<td>26.5536</td>
<td>5.8955</td>
<td>11.0836</td>
<td>0.7630</td>
</tr>
</tbody>
</table>

**Table**: Comparison RMS error - with model error in the Lorenz system
Comparison - with model error in the Lorenz system

Figure: 4DVAR

Figure: L-Curve Criterion
Outline

Inverse Problems

Data Assimilation as a Large Inverse Problem

Regularisation Parameter estimation in 4DVar

Regularisation Parameter estimation Example

Application of $L_1$-norm regularisation in 4DVar

Motivation: Results from image processing

$L_1$-norm regularisation in 4DVar

Examples
Results from image deblurring: $L_1$ regularisation

Figure: Blurred picture
Results from image deblurring: $L_1$ regularisation

Figure: Tikhonov regularisation min \( \{ \|A\mathbf{r} - \mathbf{b}\|_2^2 + \alpha\|\mathbf{r}\|_2^2 \} \)
Results from image deblurring: $L_1$ regularisation

Figure: $L_1$-norm regularisation $\min \{ \|Ax - b\|_2^2 + \alpha \|x\|_1 \}$
\( L_1 \) regularisation

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

- 4DVar smears out sharp fronts
In image processing, $L_1$-norm regularisation provides edge preserving image deblurring!

- 4DVar smears out sharp fronts
- $L_1$-norm regularisation has the potential to overcome this problem!
2 Regularisation Methods

4DVar

\[ \min_{z^{k+1}} \| Az^{k+1} - c \|^2_2 + \alpha^2 \| z^{k+1} \|^2_2 \]
2 Regularisation Methods

4DVar

\[
\min_{z^{k+1}} \| Az^{k+1} - c \|_2^2 + \alpha^2 \| z^{k+1} \|_2^2
\]

Total Variation regularisation

\[
\min_{z^{k+1}} \| Az^{k+1} - c \|_2^2 + \alpha^2 \| z^{k+1} \|_2^2 + \beta \| Dx_0^{k+1} \|_1
\]

where \( x_0^{k+1} = C \frac{1}{B} z^{k+1} + x_0^B \) and \( D \) is a matrix approximating the derivative of the solution.
Least mixed norm solutions

Solve

$$\min_{z^{k+1}} \|Az^{k+1} - c\|_2^2 + \alpha^2 \|z^{k+1}\|_2^2$$

using **Least squares** and

$$\min_{z^{k+1}} \|Az^{k+1} - c\|_2^2 + \alpha^2 \|z^{k+1}\|_2^2 + \beta \|Dx_0^{k+1}\|_1$$

using **quadratic programming** (see Fu/Ng/Nikolova/Barlow 2006).
Least mixed norm solutions

Consider

$$\min_{z^{k+1}} \| A z^{k+1} - c \|_2^2 + \beta \| D x_0^{k+1} \|_1$$

where $$x_0^{k+1} = C_B^{\frac{1}{2}} z^{k+1} + x_0^B$$
Least mixed norm solutions

Consider

$$\min_{z^{k+1}} \|Az^{k+1} - c\|_2^2 + \beta \|Dx_0^{k+1}\|_1$$

where $x_0^{k+1} = C_B \frac{1}{2} z^{k+1} + x_B$

$$\min_{z^{k+1}} \|Az^{k+1} - c\|_2^2 + \beta \|DC_B^{\frac{1}{2}} z^{k+1} + Dx_B\|_1$$
Least mixed norm solutions

Consider

$$
\min_{z^{k+1}} \|Az^{k+1} - c\|^2 + \beta \|Dx_0^{k+1}\|_1
$$

where $x_0^{k+1} = C_{B}^{\frac{1}{2}}z^{k+1} + x_0^B$

$$
\min_{z^{k+1}} \|Az^{k+1} - c\|^2 + \beta \|DC_{B}^{\frac{1}{2}}z^{k+1} + Dx_0^B\|_1
$$

Set

$$
v = \beta DC_{B}^{\frac{1}{2}}z^{k+1} + \beta Dx_0^B.
$$

and split $v$ into its positive and negative part:

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

where

$$
v^+ = \max(v, 0)
$$

$$
v^- = \max(-v, 0)
$$
Least mixed norm solutions

With

\[ v = \beta DC_B^{\frac{1}{2}} z^{k+1} + \beta Dx_0^B \]

and

\[ v = v^+ - v^- \]

the solution to

\[
\min_{z^{k+1}} \|Az^{k+1} - c\|_2^2 + \beta \|DC_B^{\frac{1}{2}} z^{k+1} + Dx_0^B\|_1
\]

is equivalent to
Least mixed norm solutions

With

\[ \mathbf{v} = \beta \mathbf{D} \mathbf{C} \frac{1}{B} \mathbf{z}^{k+1} + \beta \mathbf{D} \mathbf{x}_0^B \]

and

\[ \mathbf{v} = \mathbf{v}^+ - \mathbf{v}^- \]

the solution to

\[
\min_{\mathbf{z}^{k+1}, \mathbf{v}^+, \mathbf{v}^-} \left\{ \mathbf{1}^T \mathbf{v}^+ + \mathbf{1}^T \mathbf{v}^- + \| \mathbf{A} \mathbf{z}^{k+1} - \mathbf{c} \|_2^2 \right\}
\]

is equivalent to

subject to

\[
\beta \mathbf{D} \mathbf{C} \frac{1}{B} \mathbf{z}^{k+1} + \beta \mathbf{D} \mathbf{x}_0^B = \mathbf{v}^+ - \mathbf{v}^-
\]

\[ \mathbf{v}^+, \mathbf{v}^- \geq 0. \]
Least mixed norm solutions

\[
\min_{z^{k+1}, v^+, v^-} \left\{ 1^T v^+ + 1^T v^- + \|Az^{k+1} - c\|_2^2 \right\}
\]

subject to

\[
\beta DC_B^{\frac{1}{2}} z^{k+1} + \beta D x_0^B = v^+ - v^-
\]

\[
v^+, v^- \geq 0.
\]
Least mixed norm solutions

\[
\min_{z^{k+1}, v^+, v^-} \left\{ 1^T v^+ + 1^T v^- + \|Az^{k+1} - c\|_2^2 \right\}
\]
subject to

\[
\beta DC_B^{1/2} z^{k+1} + \beta D x_0^B = v^+ - v^-
\]

\[
v^+, v^- \geq 0.
\]

or

\[
\min_w \left\{ \frac{1}{2} w^T G w + l^T w \right\}
\]
subject to

\[
E w = k \quad \text{and} \quad F w \geq 0.
\]

where

\[
G = \begin{bmatrix}
2A^T A & 0 \\
0 & 0
\end{bmatrix}, \quad l = \begin{bmatrix}
-2A^T b \\
1 \\
1
\end{bmatrix}, \quad F = \begin{bmatrix}
0 & -I \\
-I
\end{bmatrix}
\]

\[
E = \begin{bmatrix}
\beta DC_B^{1/2} & -I & I
\end{bmatrix}, \quad w = [z^{k+1}, v^+, v^-]^T, \quad k = -\beta D x_0^B
\]
Example 1 - Linear advection equation

\[ u_t + u_z = 0, \]

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

\[
    u(z, 0) = \begin{cases} 
    0.5 & 0.25 < z < 0.5 \\
    -0.5 & z < 0.25 \text{ or } z > 0.5.
    \end{cases}
\]

This wave moves through the time interval, the model equations are defined by the upwind scheme

\[
    U_{j}^{n+1} = U_{j}^{n} - \frac{\Delta t}{\Delta z} (U_{j}^{n} - U_{j-1}^{n}),
\]

\[
    U_{0}^{n+1} = U_{N}^{n+1},
\]

where \( j = 1, \ldots, N \), \( \Delta z = \frac{1}{N} \) and \( n \) is the number of time steps. We take \( N = 100 \), \( \Delta t = 0.005 \).
length of the assimilation window: 40 time steps
perfect observations, noisy and sparse observations
\( \mathbf{R} = 0.01 \).
\( \mathbf{B} = \mathbf{I} \) and \( \mathbf{B} = 0.1e^{-\frac{|i-j|}{2L^2}} \), where \( L = 5 \)
4DVar - perfect and full observations, $B = I$

Figure: $t = 0$

Figure: $t = 20$

Figure: $t = 40$

Figure: $t = 80$
L1 - perfect and full observations, $B = I$

Figure: $t = 0$

Figure: $t = 20$

Figure: $t = 40$

Figure: $t = 80$
4DVar - noisy and sparse observations, $B = I$
L1 - noisy and sparse observations, $B = I$

Figure: $t = 0$

Figure: $t = 20$

Figure: $t = 40$

Figure: $t = 80$
4DVar - perfect and full observations, $B = 0.1e^{-\frac{|i-j|}{2L^2}}$

Figure: $t = 0$

Figure: $t = 20$

Figure: $t = 40$

Figure: $t = 80$
L1 - perfect and full observations, $B = 0.1e^{-\frac{|i-j|}{2L^2}}$

Figure: $t = 0$

Figure: $t = 20$

Figure: $t = 40$

Figure: $t = 80$
4DVar - noisy and sparse observations, $B = 0.1 e^{-\frac{|i-j|}{2L^2}}$

Figure: $t = 0$

Figure: $t = 20$

Figure: $t = 40$

Figure: $t = 80$
L1 - noisy and sparse observations, $B = 0.1e^{-\frac{|i-j|}{2L^2}}$

Figure: $t = 0$

Figure: $t = 20$

Figure: $t = 40$

Figure: $t = 80$
Example 2 - Burgers’ 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

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

where \( j = 1, \ldots, N \), \( \Delta x = \frac{1}{N} \) and \( n \) is the number of time steps. We take \( N = 100, \Delta t = 0.001 \).
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-Friedrichs method (smearing out the shock)

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

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

\[ U_{j+1}^n = U_{j}^n - \frac{\Delta t}{2 \Delta x} (f(U_{j+1}^n) - f(U_{j-1}^n)) + \frac{\Delta t^2}{2 \Delta x^2} \left( A_{j+\frac{1}{2}} (f(U_{j+1}^n) - f(U_{j}^n)) - A_{j-\frac{1}{2}} (f(U_{j}^n) - f(U_{j-1}^n)) \right) \]
Visualisation - Truth trajectory and numerical solution

Lax-Friedrichs method

Figure: \( t = 0 \)

Lax-Wendroff method

Figure: \( t = 0 \)
Visualisation - Truth trajectory and numerical solution

Lax-Friedrichs method

![Graph showing comparison between truth and numerical solution using Lax-Friedrichs method at t = 25]

Lax-Wendroff method

![Graph showing comparison between truth and numerical solution using Lax-Wendroff method at t = 25]
Visualisation - Truth trajectory and numerical solution

Lax-Friedrichs method

Figure: $t = 50$

Lax-Wendroff method

Figure: $t = 50$
Visualisation - Truth trajectory and numerical solution

Lax-Friedrichs method

Lax-Wendroff method

Figure: $t = 100$
Visualisation - Truth trajectory and numerical solution

Lax-Friedrichs method

Lax-Wendroff method

Figure: $t = 200$

Figure: $t = 200$
Setup

- length of the assimilation window: 100 time steps
- noisy and sparse observations
- $R = 0.01$.
- $B = 0.1e^{-\frac{|i-j|}{2L^2}}$, where $L = 5$
Lax-Friedrichs method
4DVar - noisy and sparse observations, $B = 0.1e^{-|i-j|/2L^2}$

Figure: $t = 0$

Figure: $t = 50$

Figure: $t = 100$

Figure: $t = 200$
L1 - noisy and sparse observations, $B = 0.1e^{-\frac{|i-j|}{2L^2}}$

Figure: $t = 0$

Figure: $t = 50$

Figure: $t = 100$

Figure: $t = 200$
Lax-Wendroff method
4DVar - noisy and sparse observations, \( B = 0.1e^{-\frac{|i-j|}{2L^2}} \)

\begin{figure}[h]
\centering
\includegraphics[width=0.45\textwidth]{figure_t_0.png}
\caption{Figure: \( t = 0 \)}
\end{figure}

\begin{figure}[h]
\centering
\includegraphics[width=0.45\textwidth]{figure_t_50.png}
\caption{Figure: \( t = 50 \)}
\end{figure}

\begin{figure}[h]
\centering
\includegraphics[width=0.45\textwidth]{figure_t_100.png}
\caption{Figure: \( t = 100 \)}
\end{figure}

\begin{figure}[h]
\centering
\includegraphics[width=0.45\textwidth]{figure_t_200.png}
\caption{Figure: \( t = 200 \)}
\end{figure}
L1 - noisy and sparse observations, $B = 0.1e^{-\frac{|i-j|}{2L^2}}$

Figure: $t = 0$

Figure: $t = 50$

Figure: $t = 100$

Figure: $t = 200$
Conclusions and further work

Conclusions

- regularisation parameter estimation methods improve 4DVar analysis
- $L_1$-norm regularisation recovers discontinuity better than 4DVar
Conclusions and further work

Conclusions
- regularisation parameter estimation methods improve 4DVar analysis
- $L_1$-norm regularisation recovers discontinuity better than 4DVar

Future work
- Further work: analysis of methods; convergence
Conclusions and further work

Conclusions

- regularisation parameter estimation methods improve 4DVar analysis
- $L_1$-norm regularisation recovers discontinuity better than 4DVar

Future work

- Further work: analysis of methods; convergence
- Extension to 2D, 3D
Conclusions and further work

Conclusions

- regularisation parameter estimation methods improve 4DVar analysis
- $L_1$-norm regularisation recovers discontinuity better than 4DVar

Future work

- Further work: analysis of methods; convergence
- Extension to 2D, 3D
- Multiscale methods
Weather forecast

*Figure*: Weather forecast for Europe for Wednesday lunchtime


Thank you.
Thank you.

Workshop 2: October 24-28, 2011
Large-Scale Inverse Problems and Applications in the Earth Sciences