Conditioning and Preconditioning of the Optimal State Estimation Problem

Nancy Nichols, Stephen Haben and Amos Lawless

University of Reading
Outline

• Optimal State Estimation
• Conditioning
• Preconditioning
• Ill-posedness and Regularization
• Conclusions
1. Optimal State Estimation
State Estimation / Data Assimilation

Aim: Find the optimal estimate (analysis) of the expected states of a system, consistent with both observations and the system dynamics given:

- Numerical prediction model (PDE)
- Observations of the system (over time)
- Background state (prior estimate)
- Estimates of the error statistics
Optimal Bayesian Estimate

\[
\begin{align*}
\min \ J(x_0) &= \frac{1}{2} (x_0 - x_0^b)^T B^{-1} (x_0 - x_0^b) \\
&+ \sum_{i=0}^{n} (\mathcal{H}_i[x_i] - y_i)^T R_i^{-1} (\mathcal{H}_i[x_i] - y_i) \\
\text{subject to} \quad x_i &= \mathcal{M}(t_i, t_0, x_0) \\
x_0^b &\quad \text{- Background state (prior estimate)} \\
y_i &\quad \text{- Observations} \\
\mathcal{H}_i &\quad \text{- Observation operator} \\
B &\quad \text{- Background error covariance matrix} \\
R_i &\quad \text{- Observation error covariance matrix}
\end{align*}
\]
Significant Properties:

- Very large number of unknowns \((10^7 - 10^8)\)
- Few observations \((10^5 - 10^6)\)
- System nonlinear unstable/chaotic
- Multi-scale dynamics
Solution of the Problem

• Solve iteratively by an approximate Gauss-Newton (incremental variational) method

• Solves a sequence of linear least-squares problems

• Use conjugate gradient/quasi-Newton iteration method to solve inner linear least-squares problem.

• Must solve in real time
Conditioning of the Problem

Accuracy/rate of convergence depend on the condition number \( \frac{\lambda_{\text{max}}}{\lambda_{\text{min}}} \) of the Hessian:

\[
B^{-1} + \hat{H}^T \hat{R}^{-1} \hat{H}
\]

where

\[
\hat{H} = \begin{pmatrix}
    H_0 \\
    H_1 \mathcal{M}_{0,1} \\
    \vdots \\
    H_n \mathcal{M}_{0,n}
\end{pmatrix}
\]

\[
\hat{R} = \begin{pmatrix}
    R_0 & 0 & \cdots & 0 \\
    0 & R_1 & \cdots & 0 \\
    \vdots & \vdots & \ddots & \vdots \\
    0 & 0 & \cdots & R_n
\end{pmatrix}
\]

\[
\mathcal{M}_{0,k} = \left. \frac{\partial \mathcal{M}_{0,k}}{\partial x} \right|_{x_0}
\]

\[
H_k = \left. \frac{\partial H_k}{\partial x} \right|_{\mathcal{M}_{0,k}(x_0)}
\]
Three Questions

• What is the effect of the covariance structure on the conditioning of the Hessian

• What is the effect of preconditioning on the problem

• How does the observational system affect the conditioning of the problem

Reference: (HLN)
Haben, Lawless and Nichols, *UoR Maths Rpt 3/2009*
2. Conditioning of the Hessian
Optimal Bayesian Estimate – 3DVar

Minimizes with respect to initial state $x_0$:

$$J(x_0) = \frac{1}{2} (x_0 - x_0^b)^T B^{-1} (x_0 - x_0^b) + \frac{1}{2} (H(x_0) - y)^T R^{-1} (H(x_0) - y),$$

- Hessian is:
  $$A = B^{-1} + (H)^T R^{-1} H$$

- $B$ and $R$ are covariance matrices that depend on the correlation length-scales of the errors; $H$ is the Jacobian of $H$. 
Minimizes with respect to initial state $x_0$:

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

- Hessian is:

$$A = B^{-1} + (H)^T R^{-1} H$$

- $B$ and $R$ are covariance matrices that depend on the correlation length scales of the errors; $H$ is the Jacobian of $\mathcal{H}$. 
Background Covariance Matrix

Define $\mathbf{B} = \sigma_b^2 \mathbf{C}$ on a 1D periodic domain by a Gaussian correlation structure:

$$(\mathbf{B})_{i,j} = \sigma_b^2 \exp \left( \frac{-r_{i,j}^2}{2L^2} \right)$$

- $r_{i,j}$ distance between points at position $i$ and $j$
- $L$ correlation lengthscale
- $\sigma_b^2$ background error variance
- Circulant
Circulant Matrices

\[ \mathbf{C} = \begin{pmatrix}
    c_0 & c_1 & c_2 & c_3 & \cdots & c_{N-2} & c_{N-1} \\
    c_{N-1} & c_0 & c_1 & c_2 & \cdots & c_{N-3} & c_{N-2} \\
    c_{N-2} & c_{N-1} & c_0 & & \cdots & \vdots & \vdots \\
    \vdots & \vdots & \ddots & \ddots & & \vdots & \vdots \\
    c_2 & & & \cdots & \ddots & c_0 & c_1 \\
    c_1 & c_2 & \cdots & c_{N-2} & c_{N-1} & c_0 
\end{pmatrix} \]

Eigenvalues

\[ \lambda_m = \sum_{k=0}^{N-1} c_k e^{-2\pi imk/N}. \]
Condition of B - Gaussian
Condition of Hessian

\[ \mathbf{A} = \mathbf{B}^{-1} + (\mathbf{H})^T \mathbf{R}^{-1} \mathbf{H} \]

- \( \mathbf{B} = \sigma_b^2 \mathbf{C} \), where \( \mathbf{C} \) is a correlation matrix.
- Assume observations are at grid points.
- Assume observation errors uncorrelated.
- \( \sigma_0^2 \) variance observation errors.
- \( \mathbf{R} = \sigma_0^2 \mathbf{I} \)
- \( \mathbf{H} \) linear.
- \( \mathbf{H}^T \mathbf{H} \) diagonal.
Condition Number of Hessian

Assume $\mathbf{C}$ has a **circulant** covariance structure.

**Bounds** on the conditioning of the Hessian are:

$$\alpha \kappa(\mathbf{C}) \leq \kappa(\mathbf{B}^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H}) \leq \left(1 + \left(\frac{\sigma_b^2}{\sigma_o^2}\right) \lambda_{\min}(\mathbf{C})\right) \kappa(\mathbf{C})$$

where

$$\alpha = \left(\frac{1 + \frac{p}{N} \frac{\sigma_b^2}{\sigma_o^2} \lambda_{\min}(\mathbf{C})}{1 + \frac{p}{N} \frac{\sigma_b^2}{\sigma_o^2} \lambda_{\max}(\mathbf{C})}\right)$$

and

$p = \text{number of observations}$
Conditioning of Hessian

Condition Number of $(B^{-1} + HR^{-1}H^T)$ vs Length Scale

Periodic Gaussian Exponential

$$B_{ij} = \sigma_b^2 \exp \left( \frac{-r_{i,j}^2}{2L^2} \right)$$

Blue = condition number  Red = bounds
Conditioning of Hessian

Condition Number of $(B^{-1} + HR^{-1}H^T)$ vs Length Scale

Periodic Gaussian Exponential

$$B_{ij} = \sigma_p^2 \exp \left( \frac{-r_{ij}^2}{2L^2} \right)$$

Blue = condition number

Laplacian 2nd Derivative

$$B^{-1} = \gamma^{-1} \left( I + \frac{l^4}{2\Delta x^4} (L)^2 \right)$$

Red = bounds
3. Preconditioning of the Hessian
Control Variable Transform

To improve conditioning transform to new variable:

- $z = B^{1/2} (x_0 - x_0^b)$
- Uncorrelated variable
- Equivalent to preconditioning by $B^{1/2}$
- Hessian of transformed problem is

$$I + B^{1/2} \hat{H}^T \hat{R}^{-1} \hat{H}B^{1/2}$$
Preconditioned Hessian

Bounds on the conditioning of the preconditioned Hessian are:

\[ 1 + \frac{\sigma_b^2}{\sigma_0^2} \gamma \leq \kappa (I + \sigma_0^{-2}(B^{1/2}H^T H B^{1/2})) \leq 1 + \frac{\sigma_b^2}{\sigma_0^2} \nu_0 \]

where

\[ \nu_0 = \| HCH^T \|_\infty, \quad \gamma = \frac{1}{p} \sum_{i,j \in J} C_{i,j} \cdot \]

\( \nu_0 \) changes slowly as a function of length scale.
Preconditioned Hessian - Gaussian

Condition number as a function of length scale

Preconditioned (blue)
Bounds (red)
Preconditioned Hessian - Gaussian

Assume two observations at $k^{th}$ and $m^{th}$ grid points

$$\kappa \left( I + \sigma_o^{-2} (B^{1/2} H^T H B^{1/2}) \right) = 1 + \frac{\sigma_b^2}{\sigma_o^2} \left| c_{k,m} \right|$$

Condition number decreases as the separation of the observations increases
Preconditioned Hessian - Gaussian

Condition number as a function of observation spacing
Extension to 4DVar

Convergence depends on condition number of

\[ I + B^{1/2} \hat{H}^T \hat{R}^{-1} \hat{H}B^{1/2} \]

where

\[ \hat{H} = \begin{pmatrix} H_0 \\ H_1 M_{0,1} \\ \vdots \\ H_n M_{0,n} \end{pmatrix} \quad \hat{R} = \begin{pmatrix} R_0 & 0 & \cdots & 0 \\ 0 & R_1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & R_n \end{pmatrix} \]

\[ M_{0,k} = \frac{\partial M_{0,k}}{\partial x} \bigg|_{x_0} \]

\[ H_k = \frac{\partial H_k}{\partial x} \bigg|_{M_{0,k}(x_0)} \]
Preconditioned Hessian

Bounds on the conditioning of the preconditioned Hessian are:

\[
1 + \frac{1}{p(n+1)} \frac{\sigma_b^2}{\sigma_o^2} \sum_{i,j=1}^{p(n+1)} (\hat{H}C\hat{H}^T)_{i,j} \leq \kappa(A_p) \leq 1 + \frac{\sigma_b^2}{\sigma_o^2} \|\hat{H}C\hat{H}^T\|_\infty
\]

where

- \( B = \sigma_b^2 \mathbf{C} \), \( \mathbf{C} \) is correlation matrix
- \( R_k = \sigma_0^2 \mathbf{I} \) for \( k = 0, \ldots, n \)
- Advection model discretized using upwind scheme
4D Background Correlation Matrix

Each block describes a covariance matrix between the background errors at different times.
Condition of Preconditioned 4DVar – using SOAR Correlation Matrix
Convergence Rates of CG in 4DVar – using SOAR Correlation Matrix

<table>
<thead>
<tr>
<th>Lengthscale</th>
<th>Unprecond</th>
<th>Precond</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.01</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>0.1</td>
<td>54</td>
<td>11</td>
</tr>
<tr>
<td>0.2</td>
<td>187</td>
<td>12</td>
</tr>
<tr>
<td>0.3</td>
<td>361</td>
<td>12</td>
</tr>
</tbody>
</table>
4.

Regularization
Preconditioned Hessian – 4DVar

Let \( B = \sigma_b^2 C_B \) \( \hat{R} = \sigma_o^2 I \) \( \mu^2 = \frac{\sigma_o^2}{\sigma_b^2} \)

Then the preconditioned optimal state estimation problem may be written in the form of a classical Tikhonov regularized problem:

\[
\hat{J} = \mu^2 \| z \|_2^2 + \| \hat{H} C_{1/2} B z - \hat{d} \|_2^2
\]

\[
\hat{H} = \begin{pmatrix}
H_0 \\
H_1 M_{0,1} \\
\vdots \\
H_n M_{0,n}
\end{pmatrix} \quad \hat{d} = \begin{pmatrix}
y_0 - \mathcal{H}_0(x_0^b) \\
y_1 - \mathcal{H}_1(x_1^b) \\
\vdots \\
y_n - \mathcal{H}_n(x_n^b)
\end{pmatrix}
\]
Rewrite the 4D Var solution as:

\[ Z = \sum_{i} \frac{\lambda_i^2}{\mu^2 + \lambda_i^2} \frac{u_i^T \hat{d}}{\lambda_i} v_i \]

where \( \mu^2 = \frac{\sigma_0^2}{\sigma_b^2} \) and \( \lambda_i, v_i, u_i \) are the singular values and right and left singular vectors of \( \hat{H}C_B^{1/2} = U\Lambda V^T \).

Johnson et al, 2005a,b
Department of Mathematics
SVD

2 pairs of RSVs contribute towards the analysis increment. The second pair lie in the near-null space.

Johnson et al, 2005a,b
Department of Mathematics
Effect of Observational Noise

Values of $\frac{\mathbf{u}_j^T \hat{\mathbf{d}}}{\lambda_j}$ for perfect and noisy observations.

Johnson et al, 2005a,b
Department of Mathematics
5. Conclusions
Conclusions

• The condition number of the Hessian is sensitive to lengthscale.

• **Preconditioning with** $B^{1/2}$ **generally reduces** the condition number and **improves the convergence rates**.

• The condition number of the Hessian **increases** as the observations become more accurate and as their separation decreases.
References


Future

Many more challenges left!