A posteriori error estimates for the solution of inverse problems
UMD-VT Data Assimilation Day

Vishwas Rao and Adrian Sandu

Computational Science Laboratory (CSL)
Department of Computer Science
Virginia Tech.
visrao@cs.vt.edu
Outline

- Motivation
- Inverse problems
- Error estimation procedure
- Numerical experiments
- Conclusions
Motivation

- A posteriori error estimation is concerned with quantifying the error associated with the computed solution.

- Applications in Air-plane design, Aerodynamic shape optimization, Radar imaging, etc.

- Can be used as a tool for parameter tuning.

- Applications in optimal sensor configuration.
Inverse Problems

The forward problem

Estimate parameters → Mathematical model / Physical theory → Prediction of data

The inverse problem

Prediction of parameters → Mathematical model / Physical theory → Measured data

A posteriori error estimates for the solution of inverse problems. (http://csl.cs.vt.edu)
Inverse Problems - Mathematical Formulation I

▶ Discrete-time model:

\[ x_{k+1} = \mathcal{M}_{k,k+1}(x_k, \theta), \quad k = 0, \ldots, N-1, \quad x_0 = x_0(\theta), \quad (1) \]

\[ x \in \mathbb{R}^n, \quad \theta \in \mathbb{R}^m, \quad \mathcal{M} : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}^n. \]

▶ A general cost function with (1) as constraints

\[ J(x, \theta) = \sum_{k=0}^{N} r_k(x_k, \theta), \quad (2) \]

\[ J : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}, \quad r : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}. \]

▶ The Inverse problem:

\[ \theta^a = \arg \min_{\theta} J(x, \theta) \quad (3) \]

subject to \[ x_{k+1} - \mathcal{M}_{k,k+1}(x_k, \theta) = 0. \]
The 4D-Var cost function is

\[
J(x_0) = \frac{1}{2} \left( x_0 - x_0^b(\theta) \right)^T B_0^{-1}(\theta) \left( x_0 - x_0^b(\theta) \right) \\
+ \sum_{k=0}^{N} \frac{1}{2} \left( \mathcal{H}_k(x_k, \theta) - y_k \right)^T R_k^{-1}(\theta) \left( \mathcal{H}_k(x_k, \theta) - y_k \right),
\]

\[ J : \mathbb{R}^n \to \mathbb{R}, \quad \mathcal{H} : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^o, \quad y \in \mathbb{R}^o. \]
Lagrangian and the Jacobians

The Lagrangian function associated with the problem (3) is

\[ \mathcal{L} = \sum_{k=0}^{N-1} \left( r_k (x_k, \theta) - \lambda_{k+1}^T \cdot (x_{k+1} - M_{k,k+1}(x_k, \theta)) \right) \]

\[ + r_N (x_N, \theta) - \lambda_0^T \cdot (x_0 - x_0(\theta)) \]  

(5)

The model Jacobians are given by

\[ M_{k,k+1}(x, \theta) := (M_{k,k+1}(x, \theta))_x, \quad \nabla M_{k,k+1}(x, \theta) := (M_{k,k+1}(x, \theta))_\theta. \]  

(6)

The Jacobians of the cost function terms are given by

\[ (r_k)_x := (r_k(x, \theta))_x \bigg|_{x=x_k}, \quad (r_k)_\theta := (r_k(x, \theta))_\theta \bigg|_{x=x_k}. \]  

(7)
First order optimality conditions

Setting $\langle \nabla_\lambda L, \delta \lambda \rangle = 0$ gives the forward model ($\mathcal{F}$):

$$0 = x_{k+1} - M_{k,k+1}(x_k, \theta), \quad k = 0, \ldots, N - 1; \quad (8a)$$

Setting $\langle \nabla_x L, \delta x \rangle = 0$ gives the adjoint model ($\mathcal{A}$):

$$0 = \lambda_N - (r_N)^T_{x_N}, \quad (8b)$$

$$0 = \lambda_k - M^T_{k,k+1} \lambda_{k+1} - (r_k)^T_{x_k}, \quad k = N - 1, \ldots, 0;$$

Setting $\langle \nabla_\theta L, \delta \theta \rangle = 0$ gives the optimality condition ($\mathcal{O}$):

$$0 = (x_0)^T_{\theta} \lambda_0 + \sum_{k=0}^{N} (r_k)^T_{\theta} + \sum_{k=0}^{N-1} M^T_{k,k+1} \lambda_{k+1}. \quad (8c)$$
Sources of error

Sources of error:

1. Data errors.
2. Model errors.

A posteriori error estimates aims to quantify the effect of these errors on a particular aspect of the solution ($\theta^a$).
Perturbed inverse problem

Imperfect discrete model:

\[ \hat{x}_{k+1} = M_{k,k+1}(\hat{x}_k, \theta) + \Delta \hat{x}_{k+1}(\hat{x}_k, \theta), \quad k = 0, 1, \ldots, N - 1. \]  \hspace{1cm} (9)

Errors in the data lead to the following perturbed cost function:

\[ \hat{J}(\hat{x}, \theta) = \sum_{k=0}^{N} \left( r_k(\hat{x}_k, \theta) + \Delta r_k(\hat{x}_k, \theta) \right). \]  \hspace{1cm} (10)

The perturbed inverse problem solved in practice reads:

\[ \hat{\theta}^a = \arg \min_{\theta} \hat{J}(\hat{x}, \theta) \quad \text{subject to } (9). \]  \hspace{1cm} (11)
First order optimality conditions for the perturbed inverse problem I

The first order optimality conditions for the perturbed inverse problem (11) are:

Forward model ($\hat{F}$):

$$\Delta \hat{x}_{k+1} = \hat{x}_{k+1} - \mathcal{M}_{k+1}(\hat{x}_k, \theta), \quad k = 0, \ldots, N - 1; \quad (12a)$$

Adjoint model ($\hat{A}$):

$$\begin{align*}
(\Delta \hat{r}_N)^T_{\hat{x}_N} &= \hat{\lambda}_N - (\hat{r}_N)^T_{\hat{x}_N}, \\
(\Delta \hat{r}_k)^T_{\hat{x}_k} + (\Delta \hat{x}_{k+1})^T_{\hat{x}_k} \hat{\lambda}_{k+1} &= \hat{\lambda}_k - \hat{M}_{k+1}^T \hat{\lambda}_{k+1} - (\hat{r}_k)^T_{\hat{x}_k} \\
&= \hat{\lambda}_k - \hat{M}_{k+1}^T \hat{\lambda}_{k+1} - (\hat{r}_k)^T_{\hat{x}_k}, \\
k &= N - 1, \ldots, 0; \quad (12b)
\end{align*}$$
First order optimality conditions for the perturbed inverse problem II

Optimality condition $(\hat{O})$:

\[
\sum_{k=0}^{N} (\Delta \hat{r}_k)^T_{\theta} - \sum_{k=0}^{N-1} (\Delta \hat{x}_{k+1})^T_{\theta} \hat{\lambda}_{k+1} = (\hat{x}_0)^T_{\theta} \hat{\lambda}_0 + \sum_{k=0}^{N} (\hat{r}_k)^T_{\theta} + \sum_{k=0}^{N-1} \hat{M}_{k,k+1}^T \hat{\lambda}_{k+1}.
\]

Error estimation procedure [12/25]
Quantity of Interest

- Quantity of interest (QoI) is defined by a scalar functional $\mathcal{E} : \mathbb{R}^m \rightarrow \mathbb{R}$ that measures a certain aspect of the optimal parameter value
  \[ QoI = \mathcal{E}(\theta^a). \quad (13) \]

- An example of error functional (13) is the $\ell$-th component of the optimal parameter vector, $\mathcal{E}(\theta^a) = \theta^a_\ell$.

- The error in the QoI is
  \[ \Delta \mathcal{E} = \mathcal{E}(\hat{\theta}^a) - \mathcal{E}(\theta^a) \quad (14) \]
“Ideal” and “perturbed” super-Lagrangians

The super-Lagrangian associated with the scalar functional (13) with “ideal” first order optimality conditions (8) is defined as

\[ \mathcal{L}^E(\theta, x, \lambda, \mu, \nu, \zeta) = E(\theta) - \nu^T \cdot F - \mu^T \cdot A - \zeta^T \cdot O \]  

(15)

The super-Lagrangian associated with the scalar functional (13) with “perturbed” first order optimality conditions (12) is defined as

\[ \hat{\mathcal{L}}^E(\hat{\theta}^a, \hat{x}, \hat{\lambda}, \mu, \nu, \zeta) = E(\hat{\theta}^a) - \nu^T \cdot \hat{F} - \mu^T \cdot \hat{A} - \zeta^T \cdot \hat{O} \]  

(16)
Error estimate

- At \((\theta^a, x, \lambda, \mu, \nu, \zeta)\), super-Lagrangian is stationary. Hence we have:

\[
\Delta L^E = L^E(\hat{\theta}^a, \hat{x}, \hat{\lambda}, \mu, \nu, \zeta) - L^E(\theta^a, x, \lambda, \mu, \nu, \zeta) \approx 0. \tag{17}
\]

- Subtracting (15) from (16) and using the stationarity relation (17) leads to the following error estimate

\[
\Delta E \approx \Delta E^{est} = \Delta E_{fwd} + \Delta E_{adj} + \Delta E_{opt}, \tag{18a}
\]

\[
\Delta E_{fwd} = \nu^T \cdot (\hat{F} - F) = \sum_{k=0}^{N-1} \nu_{k+1}^T \cdot \Delta \hat{x}_{k+1}. \tag{18b}
\]

\[
\Delta E_{adj} = \mu^T \cdot (\hat{A} - A) = \sum_{k=0}^{N} \mu_k^T \cdot (\Delta \hat{r}_k)_{x_k}^T + \sum_{k=0}^{N-1} \mu_k^T \cdot (\Delta \hat{x}_{k+1})_{x_k}^T \hat{\lambda}_{k+1}. \tag{18c}
\]

\[
\Delta E_{opt} = \zeta^T \cdot (\hat{O} - O) = \zeta^T \cdot \left( \sum_{k=0}^{N} (\Delta \hat{r}_k)_\theta^T - \sum_{k=0}^{N-1} (\Delta \hat{x}_{k+1})_\theta^T \hat{\lambda}_{k+1} \right). \tag{18d}
\]
Procedure to compute the super-Lagrange parameters

▶ The variation of super-Lagrangian can be written as:

$$\delta \mathcal{L} = \delta \mathcal{E} - \sum_{k=0}^{N} \left\langle \nabla_{\lambda_k} \mathcal{L}, \delta \lambda_k \right\rangle - \sum_{k=0}^{N} \left\langle \nabla_{x_k} \mathcal{L}, \delta x_k \right\rangle - \left\langle \nabla_{\theta} \mathcal{L}, \delta \theta \right\rangle$$

▶ Setting \( \left\langle \nabla_{\lambda_k} \mathcal{L}, \delta \lambda_k \right\rangle = 0 \) for \( k = 0, \ldots, N \) leads to the TLM:

\[
\begin{align*}
\mu_0 &= - (x_0)_\theta \zeta; \\
\mu_k &= M_{k-1,k} \mu_{k-1} - M_{k-1,k} \zeta, \quad k = 1, \ldots, N.
\end{align*}
\]  

(19)

▶ Setting \( \left\langle \nabla_{x_k} \mathcal{L}, \delta x_k \right\rangle = 0 \) for \( k = N, \ldots, 0 \) leads to the following SOA model:

\[
\begin{align*}
\nu_N &= (r_N)_{x_N,x_N} \mu_N - (r_N)_{\theta,x_N} \zeta; \\
\nu_k &= M_{k,k+1}^T \nu_{k+1} + (M_{k,k+1}^T \lambda_{k+1})_{x_k} \mu_k + (r_k)_{x_k,x_k} \mu_k - (r_k)_{\theta,x_k} \zeta \\
&\quad - (M_{k,k+1}^T \lambda_{k+1})_{x_k} \zeta, \quad k = N - 1, \ldots, 0.
\end{align*}
\]  

(20)
Reduced-Hessian system

Setting $\langle \nabla_\theta \mathcal{L}^E , \delta \theta \rangle = 0$ gives:

$$0 = \mathcal{E}_\theta^T + (x_0)_\theta^T \nu_0 + \sum_{k=0}^{N-1} \mathcal{M}_{k,k+1}^T \nu_{k+1}$$

$$+ (r_N)_{x_N,\theta}^T \mu_N + \sum_{k=0}^{N-1} \left( \mathcal{M}_{k,k+1}^T \lambda_{k+1} + (r_k)_{x_k}^T \right)_\theta^T \mu_k$$

$$- \left( \lambda_0^T (x_0)_{\theta,\theta} \right)^T \zeta - \sum_{k=0}^{N} (r_k)_{\theta,\theta}^T \zeta - \sum_{k=0}^{N-1} \left( \mathcal{M}_{k,k+1}^T \lambda_{k+1} \right)_\theta^T \zeta.$$  

We can show that

$$\left( \nabla^2_{\theta,\theta} j \right)(\theta^a) \cdot \zeta = \mathcal{E}_\theta^T,$$  

where, $j$ is given by

$$j(\theta) = \mathcal{J} (x(\theta), \theta) = \sum_{k=0}^{N} r_k (x_k(\theta), \theta)$$
The final Algorithm to obtain the super-Lagrange parameters

- Solve the Hessian equation (22) for $\zeta$;
- Solve the TLM (19) forward in time for $\mu_k$, $k = 0, \ldots, N$;
- Solve the SOA model (20) backward in time for $\nu_k$, $k = N, \ldots, 0$. 
Shallow water model on a sphere - Experimental setup

- Shallow water models have been used to model the atmosphere.
- Synthetic observations are generated.
- The noise in the observations are normally distributed with a mean of 0 and a std. dev of 2%.
- Model errors are introduced artificially by changing the grid resolution.
Error estimates

<table>
<thead>
<tr>
<th></th>
<th>$\Delta \mathcal{E}_{\text{actual}}$</th>
<th>$\Delta \mathcal{E}_{\text{est}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Errors</td>
<td>$5.4701 \times 10^1$</td>
<td>$5.7268 \times 10^1$</td>
</tr>
<tr>
<td>Model Errors</td>
<td>1.9278</td>
<td>2.9683</td>
</tr>
</tbody>
</table>

Table: The comparison between actual error and the a posteriori error estimates for the shallow water model for hourly observations measured over a period of 24 hours.

$$\Delta \mathcal{E}_{\text{actual}} = \mathcal{E}(\hat{\theta}^a) - \mathcal{E}(\theta^{\text{actual}})$$
Forward model solution

(a) Zonal wind velocity
(b) Meridional wind velocity
(c) Height

Figure: Snapshots of the solution at the final time for the shallow water model.
Errors in the observations:
Mean = 0, Std. dev = 2%

Figure: Errors in the data collected for different variables at different grid points for the shallow water model.
Data error impact:

(a) Zonal wind velocity

(b) Meridional wind velocity

(c) Height

Figure: Data error contributions at different grid points to the error functional for the shallow water model for hourly observations measured over a period of 24 hours.
Model error impact (discrete):

(a) Zonal wind velocity

(b) Meridional wind velocity

(c) Height

**Figure:** Model error (discrete) contributions at different grid points to the error functional for the shallow water model for hourly observations measured over a period of 24 hours.
Conclusions and future work

- We have presented a framework to estimate the a posteriori error for general inverse problems.

- The estimates are within the accepted bounds for the model errors and very accurate for the data errors.

- Having an accurate adjoint machinery is very necessary for the methodology to be successful.

- The estimates can be used to improve the forecasts in weather prediction, place the sensors more effectively etc.

- This methodology has to be tested on larger problems.