Ensemble Kalman filtering with residual nudging

– some recent results

Xiaodong Luo (IRIS), and Ibrahim Hoteit (KAUST)

29 May 2013

Outline

Problem statement (1/2)

Systems

Consider the state / parameter estimation problem in the following system

$$\begin{aligned} \mathbf{x}_k &= \mathcal{M}_{k,k-1}(\mathbf{x}_{k-1}) + \mathbf{u}_k \,, \\ \mathbf{y}_k &= \mathcal{H}_k(\mathbf{x}_k) + \mathbf{v}_k \,, \end{aligned} \tag{1}$$

with

Problem statement (1/2)

Systems

Consider the state / parameter estimation problem in the following system

$$\begin{aligned} \mathbf{x}_k &= \mathcal{M}_{k,k-1}(\mathbf{x}_{k-1}) + \mathbf{u}_k \,, \\ \mathbf{y}_k &= \mathcal{H}_k(\mathbf{x}_k) + \mathbf{v}_k \,, \end{aligned} \tag{1}$$

with

- $\mathbf{x} \in \mathbb{R}^m \rightarrow$ m-dimensional model state / parameter; $\mathbf{y} \in \mathbb{R}^p \rightarrow$ p-dimensional observation;

Systems

Consider the state / parameter estimation problem in the following system

$$\begin{aligned} \mathbf{x}_k &= \mathcal{M}_{k,k-1}(\mathbf{x}_{k-1}) + \mathbf{u}_k \,, \\ \mathbf{y}_k &= \mathcal{H}_k(\mathbf{x}_k) + \mathbf{v}_k \,, \end{aligned} \quad (1)$$

with

- $\mathcal{M} \rightarrow$ state / parameter transition operator;
  $\mathcal{H} \rightarrow$ observation operator;

Problem statement (1/2)

Systems

Consider the state / parameter estimation problem in the following system

$$\begin{aligned} \mathbf{x}_k &= \mathcal{M}_{k,k-1}(\mathbf{x}_{k-1}) + \mathbf{u}_k \,, \\ \mathbf{y}_k &= \mathcal{H}_k(\mathbf{x}_k) + \mathbf{v}_k \,, \end{aligned} \tag{1}$$

with

- $\mathbf{u} \rightarrow$ model error;
  $\mathbf{v} \rightarrow$ observational error;
  $\mathbf{u}$ and $\mathbf{v}$ are independent;

Problem statement (2/2)

The time indices k are often dropped later on since we confine ourselves to the update (filtering) step of a filter.

Assumptions

- $\mathbb{E}(\mathbf{v}) = \mathbf{0}$ and $\mathbb{E}(\mathbf{v}\mathbf{v}^T) = \mathbf{R}$;
- $\mathbf{R}$ is non-singular and there exists a non-singular matrix $\mathbf{R}^{1/2}$ such that $\mathbf{R} = \mathbf{R}^{1/2}\,\mathbf{R}^{T/2}$;

Additional notations

- The measured observation (data) $\mathbf{y}^{tr}$ is decomposed as $\mathbf{y}^{tr} = \mathcal{H}(\mathbf{x}^{tr}) + \mathbf{v}^{tr}$, with $\mathbf{x}^{tr}$ being the truth, and $\mathbf{v}^{tr}$ the corresponding realization of the observational error;

Residual nudging (1/3)

Additional notations

- The measured observation (data) $\mathbf{y}^{tr}$ is decomposed as $\mathbf{y}^{tr} = \mathcal{H}(\mathbf{x}^{tr}) + \mathbf{v}^{tr}$, with $\mathbf{x}^{tr}$ being the truth, and $\mathbf{v}^{tr}$ the corresponding realization of the observational error;
- The residual $\mathbf{r}$ with respect to $\mathbf{x}$ is given by $\mathbf{r} \equiv \mathcal{H}(\mathbf{x}) - \mathbf{y}^{tr}$;

Residual nudging (1/3)

Additional notations

- The measured observation (data) $\mathbf{y}^{tr}$ is decomposed as $\mathbf{y}^{tr} = \mathcal{H}(\mathbf{x}^{tr}) + \mathbf{v}^{tr}$, with $\mathbf{x}^{tr}$ being the truth, and $\mathbf{v}^{tr}$ the corresponding realization of the observational error;
- The residual $\mathbf{r}$ with respect to $\mathbf{x}$ is given by $\mathbf{r} \equiv \mathcal{H}(\mathbf{x}) - \mathbf{y}^{tr}$;
- The residual norm $\|\mathbf{r}\|_{\mathbf{R}} \equiv \sqrt{\mathbf{r}^T \, \mathbf{R}^{-1} \, \mathbf{r}}$;

Residual nudging (2/3)

An ad hoc criterion

Consider the residual $\hat{\mathbf{r}}$ with respect to an estimate $\hat{\mathbf{x}}$. By definition we have

$$\hat{\mathbf{r}} = \mathcal{H}(\hat{\mathbf{x}}) - \mathbf{y}^{tr} = (\mathcal{H}(\hat{\mathbf{x}}) - \mathcal{H}(\mathbf{x}^{tr})) - \mathbf{v}^{tr} \, .$$

Therefore by the triangle inequality

$$\|\hat{\mathbf{r}}\|_{\mathbf{R}} \le \|\mathcal{H}(\hat{\mathbf{x}}) - \mathcal{H}(\mathbf{x}^{tr})\|_{\mathbf{R}} + \|\mathbf{v}^{tr}\|_{\mathbf{R}} \, .$$

Residual nudging (2/3)

An ad hoc criterion

Consider the residual $\hat{\mathbf{r}}$ with respect to an estimate $\hat{\mathbf{x}}$. By definition we have

$$\hat{\mathbf{r}} = \mathcal{H}(\hat{\mathbf{x}}) - \mathbf{y}^{tr} = (\mathcal{H}(\hat{\mathbf{x}}) - \mathcal{H}(\mathbf{x}^{tr})) - \mathbf{v}^{tr} \, .$$

Therefore by the triangle inequality

$$\|\hat{\mathbf{r}}\|_{\mathbf{R}} \leq \|\mathcal{H}(\hat{\mathbf{x}}) - \mathcal{H}(\mathbf{x}^{tr})\|_{\mathbf{R}} + \|\mathbf{v}^{tr}\|_{\mathbf{R}} \, .$$

- If the filter performs reasonably well, we expect that $\|\mathcal{H}(\hat{\mathbf{x}}) - \mathcal{H}(\mathbf{x}^{tr})\|_{\mathbf{R}}$ may not be significantly larger than $\|\mathbf{v}^{tr}\|_{\mathbf{R}}$.

Residual nudging (2/3)

An ad hoc criterion

Consider the residual $\hat{\mathbf{r}}$ with respect to an estimate $\hat{\mathbf{x}}$. By definition we have

$$\hat{\mathbf{r}} = \mathcal{H}(\hat{\mathbf{x}}) - \mathbf{y}^{tr} = (\mathcal{H}(\hat{\mathbf{x}}) - \mathcal{H}(\mathbf{x}^{tr})) - \mathbf{v}^{tr} \, .$$

Therefore by the triangle inequality

$$\|\hat{\mathbf{r}}\|_{\mathbf{R}} \leq \|\mathcal{H}(\hat{\mathbf{x}}) - \mathcal{H}(\mathbf{x}^{tr})\|_{\mathbf{R}} + \|\mathbf{v}^{tr}\|_{\mathbf{R}} \, .$$

- On the other hand,
$$\begin{aligned}(\mathbb{E}(\|\mathbf{v}\|_{\mathbf{R}}))^2 &\leq \mathbb{E}(\|\mathbf{v}\|_{\mathbf{R}}^2) \\ &= \text{trace}\left(\mathbf{R}^{-1}\mathbb{E}(\mathbf{v}\mathbf{v}^T)\right) \\ &= \text{trace}(\mathbf{I}_p) = p \, ,\end{aligned}$$
where $\mathbf{I}_p$ is the p-dimensional identity matrix (p is the observation size).

Residual nudging (3/3)

Therefore our criterion is to require

$$\|\hat{\mathbf{r}}\|_{\mathbf{R}} \leq \beta \sqrt{p}$$

for some positive scalar β. This is used as our objective when we "design" the filter.

Remark: Later we may also consider a modified criterion $\beta_l \sqrt{p} \leq \|\hat{\mathbf{r}}\|_{\mathbf{R}} \leq \beta_u \sqrt{p}$ for some coefficients $0 \leq \beta_l \leq \beta_u$.

Outline

Method A (1/3)

The objective

The residual norm $\|\hat{\mathbf{r}}^a\|_{\mathbf{R}}$ of the analysis $\hat{\mathbf{x}}^a$ (i.e., the posterior estimate) be no larger than $\beta\sqrt{p}$ for a pre-chosen $\beta > 0$.

IRIS
International Research Institute of Stavanger

Method A (2/3)

Procedures (Luo and Hoteit, 2012, 2013b)

Given an analysis $\hat{\mathbf{x}}^a$ from a data assimilation algorithm,

1. calculate the residual norm $\|\hat{\mathbf{r}}^a\|_{\mathbf{R}}$;

Method A (2/3)

Procedures (Luo and Hoteit, 2012, 2013b)

Given an analysis $\hat{\mathbf{x}}^a$ from a data assimilation algorithm,

1. calculate the residual norm $\|\hat{\mathbf{r}}^a\|_{\mathbf{R}}$;
2. if $\|\hat{\mathbf{r}}^a\|_{\mathbf{R}} \leq \beta\sqrt{p}$, accept $\hat{\mathbf{x}}^a$ as the final estimate without any change (stop);

Method A (2/3)

Procedures (Luo and Hoteit, 2012, 2013b)

Given an analysis $\hat{\mathbf{x}}^a$ from a data assimilation algorithm,

1. calculate the residual norm $\|\hat{\mathbf{r}}^a\|_{\mathbf{R}}$;
2. if $\|\hat{\mathbf{r}}^a\|_{\mathbf{R}} \leq \beta\sqrt{p}$, accept $\hat{\mathbf{x}}^a$ as the final estimate without any change (stop);
3. otherwise, calculate a fraction coefficient $c = \beta\sqrt{p}/\|\hat{\mathbf{r}}^a\|_{\mathbf{R}}$, and let the modified estimate $\tilde{\mathbf{x}}^a$ be given by

$$\tilde{\mathbf{x}}^a = c\,\hat{\mathbf{x}}^a + (1-c)\,\mathbf{x}^o (\text{ called observation inversion }) \,,$$

   where $\mathbf{x}^o$ is a solution of the equation $\mathbf{H}\mathbf{x} = \mathbf{y}^{tr}$;

NB: Step 2 can be incorporated into Step 3 by letting

$$c = \min(1, \beta\sqrt{p}/\|\hat{\mathbf{r}}^a\|_{\mathbf{R}})$$

Method A (3/3)

An analysis of the procedures

With the above procedures, the (potentially) modified residual is given by

$$\begin{aligned}
\tilde{\mathbf{r}}^a &= \mathbf{H}\tilde{\mathbf{x}}^a - \mathbf{y}^{tr} \\
&= \mathbf{H}\tilde{\mathbf{x}}^a - \mathbf{H}\mathbf{x}^o \quad (\text{by } \mathbf{H}\mathbf{x}^o = \mathbf{y}^{tr}) \\
&= c\,(\mathbf{H}\hat{\mathbf{x}}^a - \mathbf{H}\mathbf{x}^o) \quad (\text{by } \tilde{\mathbf{x}}^a = c\,\hat{\mathbf{x}}^a + (1-c)\,\mathbf{x}^o) \\
&= c\,\hat{\mathbf{r}}^a \quad (\text{by } \mathbf{H}\mathbf{x}^o = \mathbf{y}^{tr} \text{ again}).
\end{aligned}$$

Since $c = \min(1, \beta\sqrt{p}/\|\hat{\mathbf{r}}^a\|_{\mathbf{R}}) \leq \beta\sqrt{p}/\|\hat{\mathbf{r}}^a\|_{\mathbf{R}}$,

$$\begin{aligned}
\|\tilde{\mathbf{r}}^a\|_{\mathbf{R}} &= \|c\,\hat{\mathbf{r}}^a\|_{\mathbf{R}} \\
&\leq \|\hat{\mathbf{r}}^a\|_{\mathbf{R}} \times \beta\sqrt{p}/\|\hat{\mathbf{r}}^a\|_{\mathbf{R}} \\
&= \beta\sqrt{p}\,, \text{ as desired.}
\end{aligned}$$

Method B (1/4)

Motivation

- The previous method is applicable to various data assimilation methods (e.g., the particle filter, see Luo and Hoteit, 2013b);

Method B (1/4)

Motivation

- The previous method is applicable to various data assimilation methods (e.g., the particle filter, see Luo and Hoteit, 2013b);
  On the other hand, though, it requires an observation inversion and thus incurs extra computational cost;

Method B (1/4)

Xiaodong Luo (IRIS), and Ibrahim Hoteit (KAUST)

Motivation

- In the second method, observation inversion is avoided;

Method B (1/4)

Motivation

- In the second method, observation inversion is avoided;
  The associated analytic results, however, may be filter-dependent.

The objective (method B)

The residual norm $\|\hat{\mathbf{r}}^a\|_{\mathbf{R}}$ of the analysis $\hat{\mathbf{x}}^a$ is bounded by $\beta_l\sqrt{p} \leq \|\hat{\mathbf{r}}^a\|_{\mathbf{R}} \leq \beta_u\sqrt{p}$ for some pre-chosen β_l and β_u.

An analysis in the framework of ensemble transform Kalman filter (Luo and Hoteit, 2013a)

Consider a family of mean update formulae in the framework of ensemble transform Kalman filter (ETKF), in terms of

$$\hat{\mathbf{x}}^a = \hat{\mathbf{x}}^b + \mathbf{G}\left(\mathbf{y}^{tr} - \mathbf{H}\hat{\mathbf{x}}^b\right) ,$$

$$\mathbf{G} = \hat{\mathbf{C}}^b\mathbf{H}^T\left(\delta\,\mathbf{H}\hat{\mathbf{C}}^b\mathbf{H}^T + \gamma\,\mathbf{R}\right)^{-1} ,$$

where

Method B (2/4)

An analysis in the framework of ensemble transform Kalman filter (Luo and Hoteit, 2013a)

Consider a family of mean update formulae in the framework of ensemble transform Kalman filter (ETKF), in terms of

$$\hat{\mathbf{x}}^a = \hat{\mathbf{x}}^b + \mathbf{G}\left(\mathbf{y}^{tr} - \mathbf{H}\hat{\mathbf{x}}^b\right) ,$$

$$\mathbf{G} = \hat{\mathbf{C}}^b\mathbf{H}^T\left(\delta\,\mathbf{H}\hat{\mathbf{C}}^b\mathbf{H}^T + \gamma\,\mathbf{R}\right)^{-1} ,$$

where

- δ and γ are some positive scalars;

Method B (2/4)

An analysis in the framework of ensemble transform Kalman filter (Luo and Hoteit, 2013a)

Consider a family of mean update formulae in the framework of ensemble transform Kalman filter (ETKF), in terms of

$$\hat{\mathbf{x}}^a = \hat{\mathbf{x}}^b + \mathbf{G}\left(\mathbf{y}^{tr} - \mathbf{H}\hat{\mathbf{x}}^b\right) ,$$

$$\mathbf{G} = \hat{\mathbf{C}}^b \mathbf{H}^T \left(\delta\, \mathbf{H}\hat{\mathbf{C}}^b \mathbf{H}^T + \gamma\, \mathbf{R}\right)^{-1} ,$$

where

- $\hat{\mathbf{C}}^b$ is a symmetric, positive semi-definite matrix;

  NB: In general $\hat{\mathbf{C}}^b$ may be related, but not necessarily proportional, to the sample error covariance $\hat{\mathbf{P}}^b$ of the background ensemble.

Method B (2/4)

An analysis in the framework of ensemble transform Kalman filter (Luo and Hoteit, 2013a)

Consider a family of mean update formulae in the framework of ensemble transform Kalman filter (ETKF), in terms of

$$\hat{\mathbf{x}}^a = \hat{\mathbf{x}}^b + \mathbf{G}\left(\mathbf{y}^{tr} - \mathbf{H}\hat{\mathbf{x}}^b\right) ,$$

$$\mathbf{G} = \hat{\mathbf{C}}^b \mathbf{H}^T \left(\delta\, \mathbf{H}\hat{\mathbf{C}}^b\mathbf{H}^T + \gamma\, \mathbf{R}\right)^{-1} ,$$

where

- If $\delta = 1$, then the above update formula resembles that in the EnKF, with $1/\gamma$ being analogous to the multiplicative inflation factor.

Method B (3/4)

Let

$$\mathbf{A} \equiv \mathbf{R}^{-1/2}\mathbf{H}\hat{\mathbf{C}}^b\mathbf{H}^T\mathbf{R}^{-T/2},$$

then it can be shown that

$$(\mathbf{R}^{-1/2}\hat{\mathbf{r}}^a) = \Phi(\mathbf{R}^{-1/2}\hat{\mathbf{r}}^b),$$

$$\text{with } \Phi \equiv \mathbf{I}_p - \mathbf{A}\left(\delta\,\mathbf{A} + \gamma\mathbf{I}_p\right)^{-1},$$

$$\hat{\mathbf{r}}^{\bullet} = \mathbf{H}\hat{\mathbf{x}}^{\bullet} - \mathbf{y}^{tr}.$$

Therefore

$$\|\hat{\mathbf{r}}^a\|_{\mathbf{R}} = \|\mathbf{R}^{-1/2}\hat{\mathbf{r}}^a\|_2 = \|\Phi\,(\mathbf{R}^{-1/2}\hat{\mathbf{r}}^b)\|_2.$$

Applying to $\|\hat{\mathbf{r}}^a\|_{\mathbf{R}}$ the following inequalities

$$\|\mathbf{M}^{-1}\|_2^{-1}\,\|\mathbf{z}\|_2 \leq \|\mathbf{M}\,\mathbf{z}\|_2 \leq \|\mathbf{M}\|_2\,\|\mathbf{z}\|_2$$

Method B (4/4)

for the specific case $\delta = 1$ we have the following bounds for γ:

$$\frac{\xi_l}{1-\xi_l}\,\lambda_{max} \le \gamma \le \frac{\xi_u}{1-\xi_u}\,\lambda_{min}\,,$$

subject to $\beta_l \le \frac{\beta_u}{\kappa + (1-\kappa)\,\xi_u}$ (from $\frac{\xi_l}{1-\xi_l}\,\lambda_{max} \le \frac{\xi_u}{1-\xi_u}\,\lambda_{min}$) ,

with $\xi_l, \xi_u \in [0,1)$,

where

Method B (4/4)

for the specific case $\delta = 1$ we have the following bounds for γ:

$$\frac{\xi_l}{1-\xi_l}\lambda_{max} \leq \gamma \leq \frac{\xi_u}{1-\xi_u}\lambda_{min} \, ,$$

subject to $\beta_l \leq \dfrac{\beta_u}{\kappa + (1-\kappa)\,\xi_u}$ (from $\dfrac{\xi_l}{1-\xi_l}\lambda_{max} \leq \dfrac{\xi_u}{1-\xi_u}\lambda_{min}$) ,

with $\xi_l, \xi_u \in [0, 1)$,

where

- $\xi_{\bullet} \equiv \beta_{\bullet} \sqrt{p} / \|\hat{\mathbf{r}}^b\|_{\mathbf{R}}$;

Method B (4/4)

for the specific case $\delta = 1$ we have the following bounds for γ:

$$\frac{\xi_l}{1-\xi_l}\,\lambda_{max} \le \gamma \le \frac{\xi_u}{1-\xi_u}\,\lambda_{min}\,,$$

subject to $\beta_l \le \frac{\beta_u}{\kappa + (1-\kappa)\,\xi_u}$ (from $\frac{\xi_l}{1-\xi_l}\,\lambda_{max} \le \frac{\xi_u}{1-\xi_u}\,\lambda_{min}$),

with $\xi_l, \xi_u \in [0, 1)$,

where

- $\xi_\bullet \equiv \beta_\bullet \sqrt{p} / \|\hat{\mathbf{r}}^b\|_{\mathbf{R}}$;
- $\lambda_\bullet$ are the eigenvalues of $\mathbf{A} \equiv \mathbf{R}^{-1/2}\mathbf{H}\hat{\mathbf{C}}^b\mathbf{H}^T\mathbf{R}^{-T/2}$;

Method B (4/4)

for the specific case $\delta = 1$ we have the following bounds for γ:

$$\frac{\xi_l}{1-\xi_l}\lambda_{max} \le \gamma \le \frac{\xi_u}{1-\xi_u}\lambda_{min} \, ,$$

subject to $\beta_l \le \dfrac{\beta_u}{\kappa + (1-\kappa)\,\xi_u}$ (from $\dfrac{\xi_l}{1-\xi_l}\lambda_{max} \le \dfrac{\xi_u}{1-\xi_u}\lambda_{min}$) ,

with $\xi_l, \xi_u \in [0, 1)$,

where

- $\xi_\bullet \equiv \beta_\bullet \sqrt{p}/\|\hat{\mathbf{r}}^b\|_{\mathbf{R}}$;
- $\lambda_\bullet$ are the eigenvalues of $\mathbf{A} \equiv \mathbf{R}^{-1/2}\mathbf{H}\hat{\mathbf{C}}^b\mathbf{H}^T\mathbf{R}^{-T/2}$;
- $\kappa = \lambda_{max}/\lambda_{min}$, i.e., the condition number of $\mathbf{A}$;

Remarks: $\xi_l, \xi_u \in [1, +\infty)$ corresponds to trivial or infeasible cases (Luo and Hoteit, 2013a).

Outline

Motivation

It can be shown that, with linear observations,

Kalman update:

$$\hat{\mathbf{x}}^a = \hat{\mathbf{x}}^b + \mathbf{G}\left(\mathbf{y}^{tr} - \mathbf{H}\hat{\mathbf{x}}^b\right) ,$$

$$\mathbf{G} = \hat{\mathbf{C}}^b \mathbf{H}^T \left(\mathbf{H}\hat{\mathbf{C}}^b \mathbf{H}^T + \gamma\, \mathbf{R}\right)^{-1} .$$

Motivation

It can be shown that, with linear observations,

Kalman update:

$$\hat{\mathbf{x}}^a = \hat{\mathbf{x}}^b + \mathbf{G}\left(\mathbf{y}^{tr} - \mathbf{H}\hat{\mathbf{x}}^b\right),$$

$$\mathbf{G} = \hat{\mathbf{C}}^b\mathbf{H}^T\left(\mathbf{H}\hat{\mathbf{C}}^b\mathbf{H}^T + \gamma\,\mathbf{R}\right)^{-1}.$$

$$\Updownarrow$$

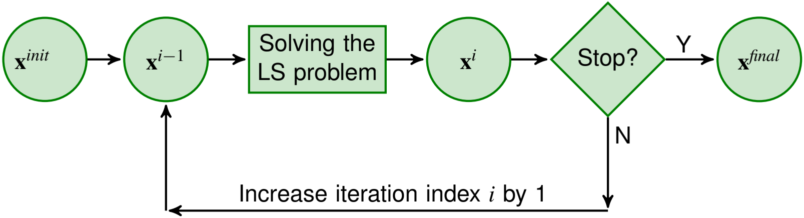
Least squares problem:

$$\hat{\mathbf{x}}^a = \underset{\mathbf{x}}{\operatorname{argmin}} \|\mathbf{y}^{tr} - \mathbf{H}\mathbf{x}\|^2_{\mathbf{R}} + \gamma\|\mathbf{x} - \hat{\mathbf{x}}^b\|^2_{\hat{\mathbf{C}}^b}.$$

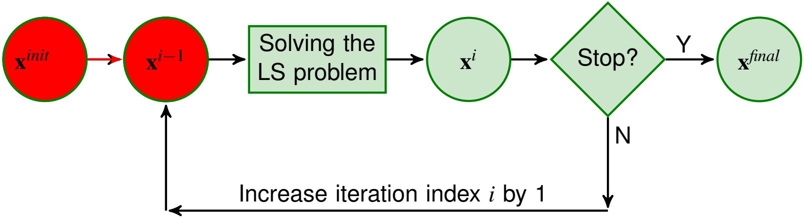
Linearization and iteration

With a nonlinear observation operator $\mathcal{H}$, one may first seek a "local" solution to the above least squares problem by linearizing $\mathcal{H}$, and then gradually expand the searching regime through iteration.

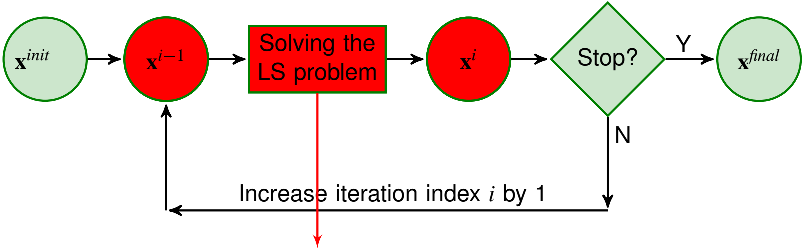
Linearization and iteration



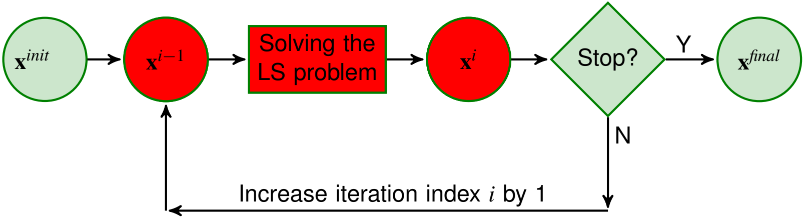
Linearization and iteration



Linearization and iteration



Linearization and iteration

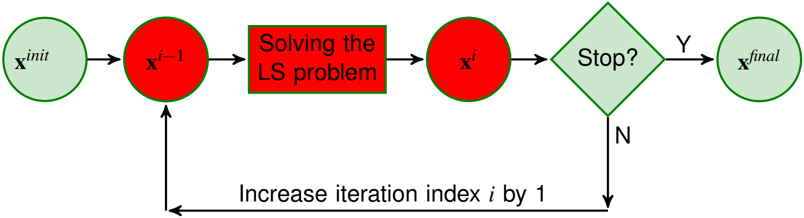


The least squares problem:

$$\underset{\mathbf{x}^i}{\operatorname{argmin}} \|\mathbf{y}^{tr} - \mathcal{H}(\mathbf{x}^{i-1}) - \mathbf{J}^{i-1}(\mathbf{x}^i - \mathbf{x}^{i-1})\|_{\mathbf{R}}^2 + \gamma^i \|\mathbf{x}^i - \mathbf{x}^{i-1}\|_{(\hat{\mathbf{C}}^b)^i}^2 \,,$$

where

Linearization and iteration



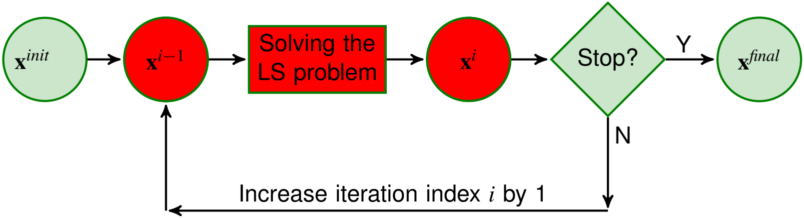
The least squares problem:

$$\underset{\mathbf{x}^i}{\text{argmin}} \, \|\mathbf{y}^{tr} - \mathcal{H}(\mathbf{x}^{i-1}) - \mathbf{J}^{i-1}(\mathbf{x}^i - \mathbf{x}^{i-1})\|_{\mathbf{R}}^2 + \gamma^i \|\mathbf{x}^i - \mathbf{x}^{i-1}\|_{(\hat{\mathbf{C}}^b)^i}^2 \, ,$$

where

- $\mathbf{J}^{i-1}$ is the Jacobian of $\mathcal{H}$ around the previous estimate $\mathbf{x}^{i-1}$;

Linearization and iteration



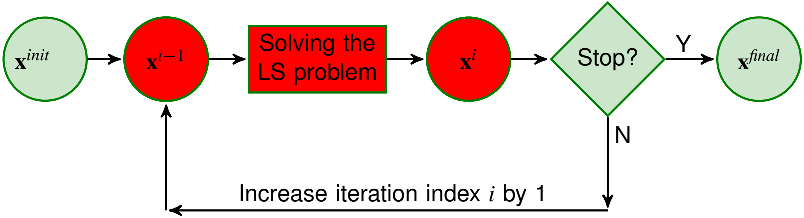
The least squares problem:

$$\underset{\mathbf{x}^i}{\text{argmin}} \, \|\mathbf{y}^{tr} - \mathcal{H}(\mathbf{x}^{i-1}) - \mathbf{J}^{i-1}(\mathbf{x}^i - \mathbf{x}^{i-1})\|^2_{\mathbf{R}} + \gamma^i \|\mathbf{x}^i - \mathbf{x}^{i-1}\|^2_{(\hat{\mathbf{C}}^b)^i} \, ,$$

where

- $\mathbf{J}^{i-1}$ is the Jacobian of $\mathcal{H}$ around the previous estimate $\mathbf{x}^{i-1}$;
- γ and $\hat{\mathbf{C}}^b$ (if necessary) may be adaptive with iteration;

Linearization and iteration



The least squares problem:

$$\underset{\mathbf{x}^i}{\text{argmin}} \, \| \mathbf{y}^{tr} - \mathcal{H}(\mathbf{x}^{i-1}) - \mathbf{J}^{i-1}(\mathbf{x}^i - \mathbf{x}^{i-1}) \|_{\mathbf{R}}^2 + \gamma^i \| \mathbf{x}^i - \mathbf{x}^{i-1} \|_{(\hat{\mathbf{C}}^b)^i}^2 \, ,$$

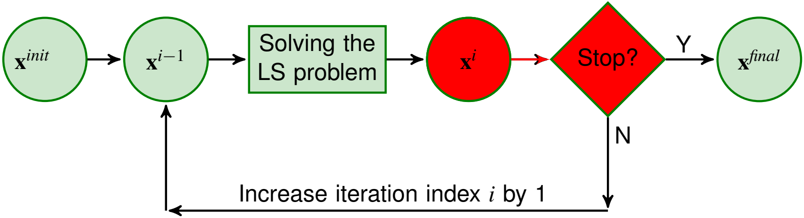
is solved by

$$\mathbf{x}^i = \mathbf{x}^{i-1} + \mathbf{G}^i \left(\mathbf{y}^{tr} - \mathcal{H}(\mathbf{x}^{i-1}) \right) \, ,$$

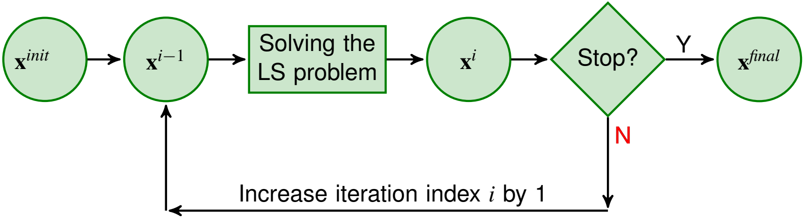
$$\mathbf{G}^i = (\hat{\mathbf{C}}^b)^i (\mathbf{J}^{i-1})^T \left(\mathbf{J}^{i-1} (\hat{\mathbf{C}}^b)^i (\mathbf{J}^{i-1})^T + \gamma^i \, \mathbf{R} \right)^{-1} \, ,$$

similar to the iteration formulae used in Chen and Oliver (2012); Emerick and Reynolds (2012).

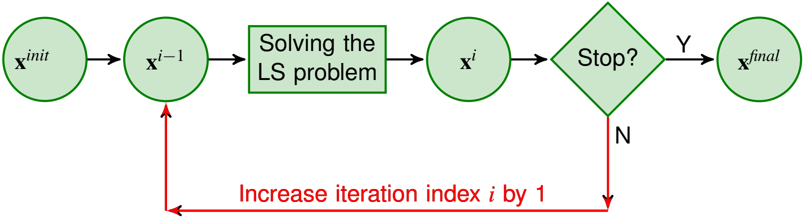
Linearization and iteration



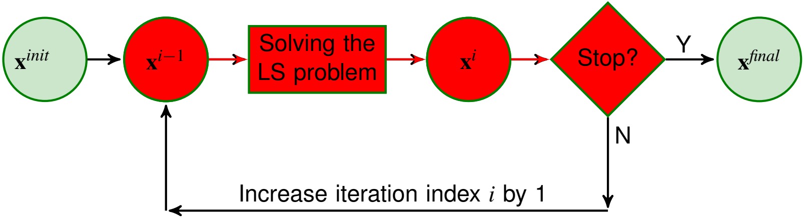
Linearization and iteration



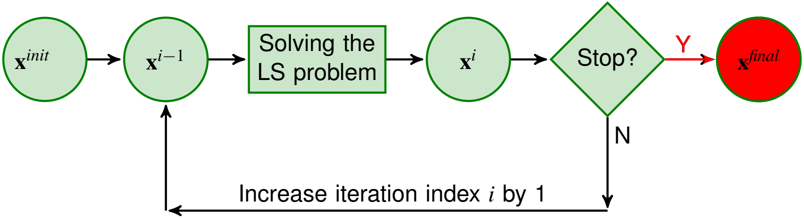
Linearization and iteration



Linearization and iteration



Linearization and iteration



Application to a parameter estimation problem (1/5)

Experiment settings

Application to a parameter estimation problem (1/5)

Experiment settings

- Dynamical model: a modified 40-dimensional Lorenz 96 model

$$\frac{dx_j}{dt} = \left(x_{j+1} - x_{j-2}\right) x_{j-1} - x_j + F_j, \; j = 1, \cdots, 40. \qquad (2)$$

with F_j being the parameters to be estimated;

NB: The model integration step = 0.05, and the assimilation time window = 1000 integration steps.

Application to a parameter estimation problem (1/5)

Experiment settings

- True parameters $F_j = 8 \; \forall \, j$;

  Initial ensembles of x_j and F_j are drawn at random from the normal distribution $N(0, 1)$;

Application to a parameter estimation problem (1/5)

Experiment settings

- Parameter time evolution model: $(F_j)_{k+1} = (F_j)_k$ with k being the time index;

  NB: Alternative model $(F_j)_{k+1} = (F_j)_k + (w_j)_k$ with noise $(w_j)_k$ is also possible;

Application to a parameter estimation problem (1/5)

Experiment settings

- Observations: 10 out of 40 state variables $(1, 6, 11, \cdots)$ are observed with Gaussian measurement noise (zero mean and variance 1);

  The observation operator is x itself for a state variable x that is to be observed;

Application to a parameter estimation problem (1/5)

Experiment settings

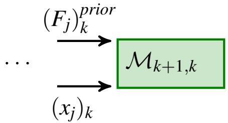
- The observation frequency is every 4 integration steps;

  The parameters F_j are estimated every S_a steps, for instance, S_a may be a multiple of 4 integration steps;

  NB: S_a is not necessarily equal to the length of the assimilation time window (1000 here). Instead, it may be shorter.

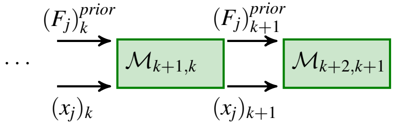
Application to a parameter estimation problem (1/5)

Experiment settings



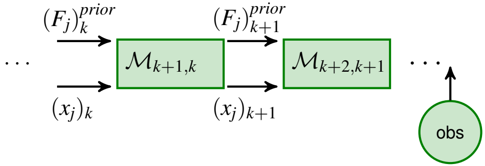
Application to a parameter estimation problem (1/5)

Experiment settings



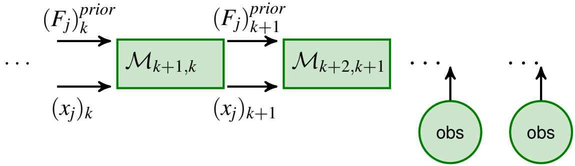
Application to a parameter estimation problem (1/5)

Experiment settings



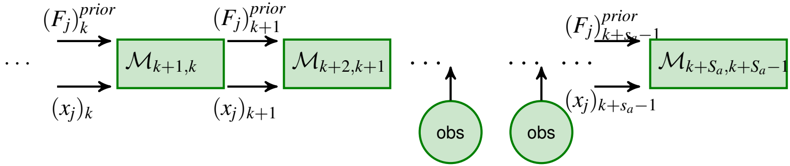
Application to a parameter estimation problem (1/5)

Experiment settings

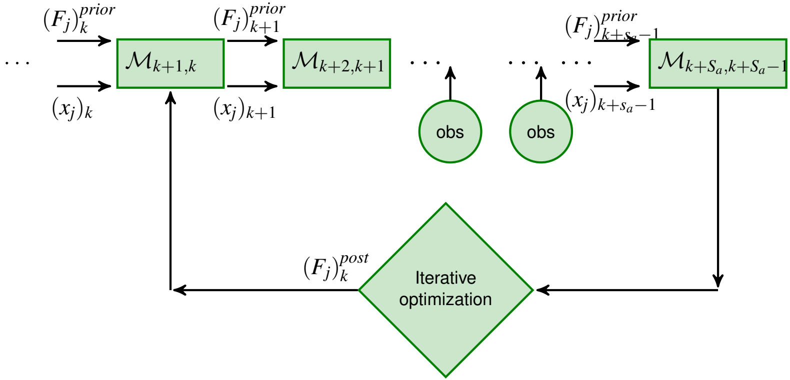


Application to a parameter estimation problem (1/5)

Experiment settings

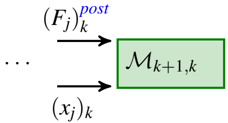


Experiment settings



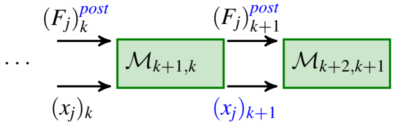
Application to a parameter estimation problem (1/5)

Experiment settings



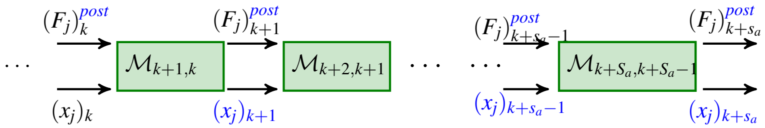
Application to a parameter estimation problem (1/5)

Experiment settings



Application to a parameter estimation problem (1/5)

Experiment settings



Application to a parameter estimation problem (2/5)

Results

Results

iterative optimization	$S_a = 100$	
	max iter no. $= 200$	max iter no. $= 1000$
RMSEs of initial / final ensemble means	7.9822 / 7.3963	7.9822 / 6.7884
CPU time	486.1082	3321.1528

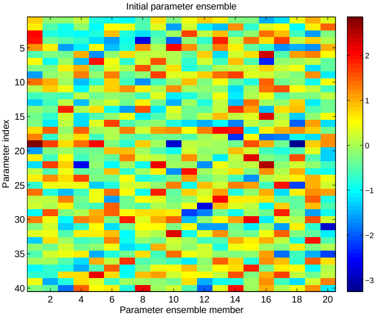
iterative optimization	$S_a = 1000$	
	max iter no. $= 200$	max iter no. $= 1000$
RMSEs of initial / final ensemble means	7.9822 / 7.9754	7.9822 / 7.9754
CPU time	777.8134	5440.5005

Application to a parameter estimation problem (3/5)

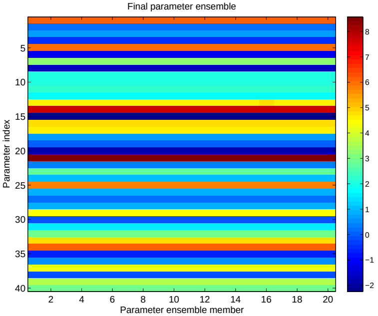
Results (cont'd)

Results of the iterative optimization with $S_a = 100$ and iter no. = 1000.

vertical: parameter variable index; horizontal: ensemble member index;



(a) Initial parameter ensemble

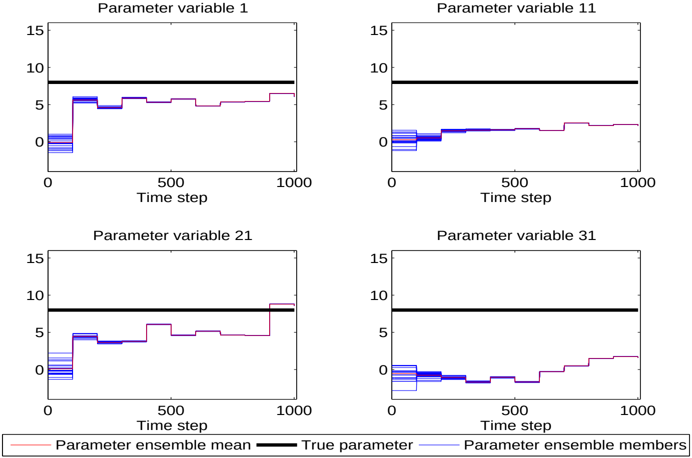


(b) Final parameter ensemble

Application to a parameter estimation problem (4/5)

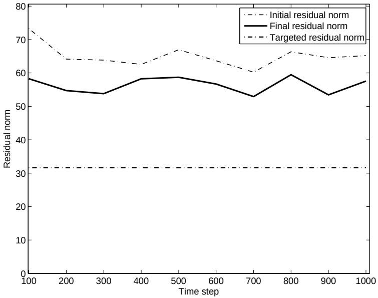
Results (cont'd)

Time series of the estimated parameter variables $F_1, F_{11}, F_{21}, F_{31}$

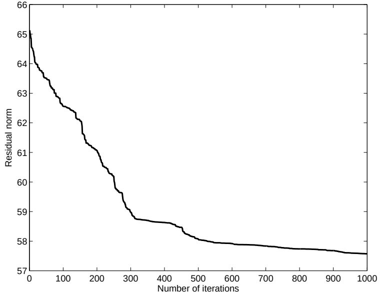


Results (cont'd)

Results of the iterative optimization with $S_a = 100$ and iter no. $= 1000$.



(a) Time series of the residual norms within the assimilation time window



(b) Residual norm reduction at time step 1000

Outline

Future plan

Possible issues in future investigations might include

- the choice of β;

Future plan

Possible issues in future investigations might include

- the choice of β;
- alternative ways to approximate the Jacobian matrix;

Future plan

Possible issues in future investigations might include

- the choice of β;
- alternative ways to approximate the Jacobian matrix;
- alternative optimization algorithms;

Future plan

Possible issues in future investigations might include

- the choice of β;
- alternative ways to approximate the Jacobian matrix;
- alternative optimization algorithms;
- applications to reservoir data assimilation and other related problems;

Future plan

Possible issues in future investigations might include

- the choice of β;
- alternative ways to approximate the Jacobian matrix;
- alternative optimization algorithms;
- applications to reservoir data assimilation and other related problems;
- others based on your feedback;

Special thanks to

financial supports from the projects

- **Rateallokering;**
- **RDA Geology;**
- **Integrated Workflow and Realistic Geology;**

References

Chen, Y. and D. Oliver, 2012: Levenberg-Marquardt forms of the iterative ensemble smoother for efficient history matching and uncertainty quantification, submitted.

Emerick, A. A. and A. C. Reynolds, 2012: Ensemble smoother with multiple data assimilation. *Computers & Geosciences*, **55**, 3–15.

Luo, X. and I. Hoteit, 2012: Ensemble Kalman filtering with residual nudging. *Tellus A*, **64**, 17 130, doi:10.3402/tellusa.v64i0.17130.

Luo, X. and I. Hoteit, 2013a: Covariance inflation in the ensemble Kalman filter: a residual nudging perspective and some implications. *Mon. Wea. Rev*, accepted.

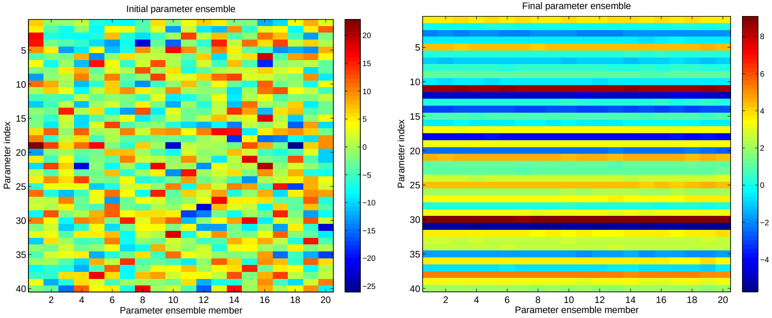
Luo, X. and I. Hoteit, 2013b: Efficient particle filtering through residual nudging. *Quart. J. Roy. Meteor. Soc.*, in press, doi:10.1002/qj.2152.

Extra results with the parameter model $(F_j)_{k+1} = (F_j)_k + (w_j)_k$ (1/3)

Results

Results of the iterative optimization with $S_a = 100$ and iter no. = 1000.

vertical: parameter variable index; horizontal: ensemble member index;

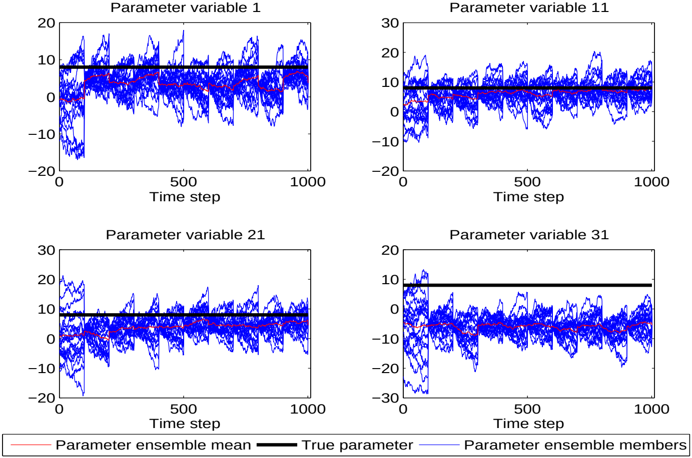


(a) Initial parameter ensemble **(b)** Final parameter ensemble

Extra results with the parameter model $(F_j)_{k+1} = (F_j)_k + (w_j)_k$ (2/3)

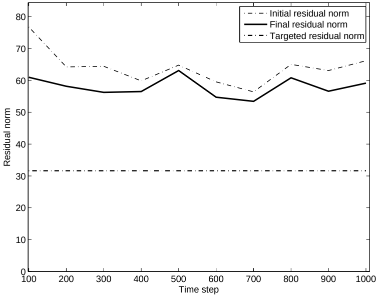
Results (cont'd)

Time series of the estimated parameter variables $F_1, F_{11}, F_{21}, F_{31}$

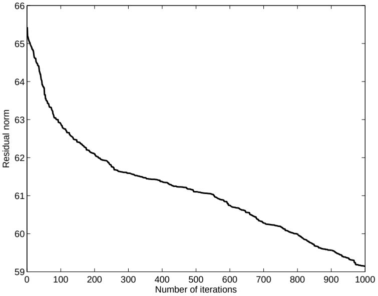


Results (cont'd)

Results of the iterative optimization with $S_a = 100$ and iter no. $= 1000$.



(a) Time series of the residual norms within the assimilation time window



(b) Residual norm reduction at time step 1000