SECOND ORDER METHODS:
Computation of the analysis error covariance for non-linear dynamics

F.-X. Le Dimet, MOISE Project, LJK, University of Grenoble, France

Igor Gejadze
Department of Civil Engineering, University of Strathclyde, Glasgow, UK

V.P. Shutyaev, Institute of Numerical Mathematics, Russian Academy of Science, Moscow
Problem

• A prediction makes sense if it is associated to an evaluation of the error prediction

• Basic ingredients of prediction are
  – Models
  – Data
  – A priori information
All these sources of information have errors
How these errors propagates?
Three stages for predicting geophysical flows

- Data Assimilation Cost 10.
- Quality of prediction Cost 100.
Error Propagation : Generic Approach

\[ M(I, O) = 0 \]

\( M \) is the model \( O \) the output, \( I \) the input

\( \delta I \) is an error on the input, it will produce an error \( \delta O \) on the output.

\[ M(I + \delta I, O + \delta O) = 0 \]

First order approximation:

\[ M(I, O) + \frac{\partial M}{\partial I} \delta I + \frac{\partial M}{\partial O} \delta O = 0 \]

\[ \delta O = - \left[ \frac{\partial M}{\partial O} \right]^{-1} \frac{\partial M}{\partial I} \delta I \]

If \( \delta I \) is an unbiased random variable then:

\[ V_O = E(\delta O, \delta O^T) = - \left[ \frac{\partial M}{\partial O} \right]^{-1} \frac{\partial M}{\partial I} \delta I \delta I^T \left[ \frac{\partial M}{\partial O} \right]^T \left[ \frac{\partial M}{\partial O} \right]^{-T} = \left[ \frac{\partial M}{\partial O} \right]^{-1} \frac{\partial M}{\partial I} V_I \left[ \frac{\partial M}{\partial I} \right]^T \left[ \frac{\partial M}{\partial O} \right]^{-T} \]

Therefore we can estimate the covariance of the output from the covariance of the input.

What going on in the case of Variational Data Assimilation?
Variational Data Assimilation

\[
\frac{dX}{dt} = F(X)
\]
\[X(0) = U.\]
\[J(U) = \frac{1}{2} \int_{0}^{T} \left\| CX - X_{\text{obs}} \right\|^2 dt + \frac{1}{2} \left\| U - U_{b} \right\|^2\]

Determine \(U^*\) mimimizing \(J\)

Optimality System:

\[
\begin{aligned}
\frac{dX}{dt} &= F(X) \\
X(0) &= U. \\
\frac{dP}{dt} + \left[ \frac{\partial F}{\partial X} \right]^T P &= C^T (CX - X_{\text{obs}}) \\
P(T) &= 0 \\
\nabla J(U) &= -P(0) + U - U_{b} = 0
\end{aligned}
\]
Propagation of errors on observations toward errors on the initial condition

- The general approach can be applied.
- I is the observation, O is the initial condition.
- What should be considered as « model » M?
- Because initial condition and observation are linked only in the Optimality System it has to be considered as the « model » and the analysis must be carried out on it.
- Introducing a second order adjoint…
Problem statement

**Model of evolution process:**

\[
\frac{\partial \varphi}{\partial t} = F(\varphi) + f, \quad \varphi = \varphi(t, x), \quad t \in (0, T) \\
\varphi|_{t=0} = u
\]

**Objective function (for the initial value control):**

\[
2J(u) = \left(V_b^{-1}(u - u_b), u - u_b \right)_X + \left(V_o^{-1}(C\varphi - \varphi_o), C\varphi - \varphi_o \right)_o
\]

**Control problem:** \( J(u) = \inf_v J(v) \)

**Optimal solution (analysis) error:**

\[\delta u = u - \bar{u}, \quad E[\delta u] = 0\]

**Analysis Error covariance matrix:**

\[E[\delta u \delta u^T] = V_{\delta u}\]
Analysis covariance and the inverse Hessian - I

In the linear case $F(u) = F$ one has: $H = J''(u)$, $V_{\delta u} = H^{-1}$

This result can be found in any old handbook on statistics or regression.

In the nonlinear case the errors are related via nonlinear operator equation [1]:

$$H(\overline{u}, \delta u, \tau)\delta u = V_b^{-1} \xi_b + R(\overline{u}, \delta u) \cdot V_o^{-1} \xi_o$$

Analysis error Background error Observation error

With the following approximations $H(\overline{u}, \delta u, \tau) \approx H(\overline{u}, 0)$, $R(\overline{u}, \delta u) \approx R(\overline{u}, 0)$ one obtains

$$V_{\delta u} = E[\delta u \delta u^T] \approx H^{-1}(\overline{u}, 0)$$

‘Exact’ solution !!!

The sufficient condition for above approximations to be valid is:

$$F(\overline{\varphi} + \delta \varphi) - F(\overline{\varphi}) \approx F'(\overline{\varphi}) \delta \varphi$$

Tangent linear hypothesis

We call error due to above approximations – the linearization error. However, the ‘exact’ solution is not known – hence another error we call the origin error.
Operator R

\[
\left\{ \begin{array}{l}
-\frac{\partial \theta^*}{\partial t} - (F'(\varphi))^{*} \theta^* = C^* V g, \quad t \in (0, T) \\
\theta^*|_{t=T} = 0, \\
R g = \theta^*|_{t=0}, \quad g \in Y_{obs}.
\end{array} \right.
\]
Analysis covariance and the inverse Hessian - II

$$V_{\delta u} \approx H^{-1}(u)$$ How far this is valid?

1. if the TLH is valid;
2. if the TLH is not valid, but the truncation error is not cumulative.

$\frac{\partial \psi}{\partial t} - F'(\bar{\varphi} + \tau \delta \varphi) \psi = 0, \quad \psi(0,x) = \nu$

$- \frac{\partial \psi^*}{\partial t} - (F'(\bar{\varphi}) \psi^* = -CV_o^{-1}C^*\psi, \quad \psi^*(T,x) = 0$

$$H(u, \delta u, \tau) \nu = V_b^{-1} \nu - \psi^*(0,x)$$

$\forall t \in (0,T)$

$\frac{\partial \psi}{\partial t} - F'(\bar{\varphi}) \psi = 0, \quad \psi(0,x) = \nu$

$- \frac{\partial \psi^*}{\partial t} - (F'(\bar{\varphi})) \psi^* = -CV_o^{-1}C^*\psi, \quad \psi^*(T,x) = 0$

$$H(u) \nu = V_b^{-1} \nu - \psi^*(0,x)$$

Example of a non-cumulative error:

Solid line: $dF^{(1)} = F(\varphi) - F(\bar{\varphi})$

Dashed line: $dF^{(2)} = F'(\bar{\varphi})(\varphi - \bar{\varphi})$
Fully nonlinear ensemble method

1. Consider function $\overline{\varphi}$ as the exact solution to the problem
2. Start ensemble loop $l = 1, \ldots, L$
2.1 Generate using Monte-Carlo $\xi_{b,l}, \xi_{o,l}$
2.2 Compute $u_b = \overline{u} + \xi_{b,l}, \varphi_{obs} = C\overline{\varphi} + \xi_{o,l}$
2.3 Solve the original nonlinear DA problem with perturbed data and find $u_l$
2.4 Compute $\delta u_l = u_l - \overline{u}$
3. End ensemble loop.
4. Compute the statistics $\hat{\mathcal{V}}_{\delta u} = \frac{1}{L} \sum_{l=1}^{L} \delta u_l \delta u_l^T$

This method is used to compute the reference posterior covariance matrix. The sample size $L$ can be reduced with the sampling error compensation procedure.

Inverse Hessian by the quasi-Newton BFGS method

The BFGS forms the inverse Hessian in course of solving the auxiliary control problem:

$$
\begin{align*}
\frac{\partial \psi}{\partial t} - F'(\varphi)\psi &= 0, \quad t \in (0, t) \\
\psi(0, x) &= B\delta u \\
J_1(\delta u) &= \inf_{v} J_1(v)
\end{align*}
$$

$$
2J_1(\delta u) = (BV_b^{-1}(\delta u - \xi_b), B(\delta u - \xi_b))_X + (V_o^{-1}(C\psi - \xi_o), C\psi - \xi_o)_Y
$$

$\xi_b, \xi_o$ - synthetic data
$B$ - general preconditioner
$H^{-1}(\overline{u}) = B\widetilde{H}^{-1}B^*$
Example 1: Initialization problem

Model (non-linear convection-diffusion):

\[ \frac{\partial \varphi}{\partial t} + \frac{\partial (w \varphi)}{\partial x} - \frac{\partial}{\partial x} \left( k(\varphi) \frac{\partial \varphi}{\partial x} \right) = Q(\varphi) \]

\[ x \in (0,1), \quad t \in (0,T] \]

\[ \varphi(x,0) = \psi \]

\[ \frac{\partial \varphi(0,t)}{\partial x} = \frac{\partial \varphi(1,t)}{\partial x} = 0 \]

\[ \text{diag}(H^{-1}) \] and ensemble variance

\[ H^{-1} \] and ensemble covariance

[Diagram of field evolution and nonlinear diffusion coefficient]
When the main result is not valid \( V_{\delta u} \neq H^{-1}(\varphi) \)

In general nonlinear case one may not expect the inverse Hessian to be a satisfactory approximation to the analysis covariance (see below).

Model: 1D Burgers with strongly nonlinear dissipation term

\[
\frac{\partial \varphi}{\partial t} + \frac{1}{2} \frac{\partial (\varphi^2)}{\partial x} - \frac{\partial}{\partial x} \left( k \left( \frac{\partial \varphi}{\partial x} \right) \frac{\partial \varphi}{\partial x} \right) = 0
\]

\( x \in (0,1), \quad t \in (0,T] \)

\( \varphi(x,0) = u \)

\( \frac{\partial \varphi(0,t)}{\partial x} = 0, \quad \frac{\partial \varphi(1,t)}{\partial x} = 0 \)

\( k = k_0 + k_1 (\partial \varphi / \partial x)^2 \)

Field evolution: case A and case B

\( diag(H^{-1}(\bar{u},0)) \) and ensemble variance for initialization problem

Case A: sensors at \( x_k = (0.35, 0.4, 0.5, 0.6, 0.65) \)

Case B: sensors at \( x_k = (0.4, 0.45, 0.55, 0.6) \)

In Figures: inverse Hessian – solid line, ensemble estimate – dotted line, background variance – dashed line
Effective Inverse Hessian method (EIH): theory

Exact nonlinear operator equation
\[ H(\bar{u}, \delta u, \tau) \delta u = V_b^{-1} \xi_b + R(\bar{u}, \delta u) \cdot V_o^{-1} \xi_o \]

Exact analysis covariance (by definition)
\[ V_{\delta u} = E[\delta u \delta u^T] = E[H^{-1}(\bar{u}, \delta u, \tau)gg^T H^{-1}(\bar{u}, \delta u, \tau)] \]
\[ gg^T = V_b^{-1} \xi_b \xi_b^T V_b^{-1} + R(u, \delta u)V_o^{-1} \xi_o \xi_o^T V_o^{-1} R^T(u, \delta u) \]

Resulting from series of assumptions the equation above reduces to the form:

\[ V_{\delta u} \approx V = E[H^{-1}(\bar{u}, \delta u)] \]

Assumes nonlinear dynamics, but ‘close-to-linear’
statistical behavior (Ratkowski, 1983)

I. Computing the expectation
by Monte Carlo:
\[ V = \frac{1}{L} \sum_{l=1}^{L} H^{-1}(\bar{u}, \delta u_l) \]
\[ \delta u_l = u_l - \bar{u} \]

l-th optimal solution

II. Computing the expectation
by definition:
\[ V = \int_{-\infty}^{\infty} H^{-1}(\bar{u}, v)f_{\delta u}(v)dv \]

If we assume \( f_{\delta u} \) Gaussian, then
\[ V = \frac{1}{(2\pi)^{M/2}|V|^{1/2}} \int_{-\infty}^{\infty} H^{-1}(\bar{u}, v) \exp\left(-\frac{1}{2}v^T V^{-1} v\right)dv \]

\( v \) is dummy argument!
EIH method: implementation

**Preconditioning**

\[ \tilde{H} = B^* HB \quad \Leftrightarrow \quad H^{-1} = B \tilde{H}^{-1} B^* \]

\[ V = B \left( (2\pi)^{-M/2} |V|^{-1/2} \int_{-\infty}^{\infty} \tilde{H}^{-1}(\bar{u},v) \exp \left( -\frac{1}{2} v^T V^{-1} v \right) dv \right) B^* \]

1-level preconditioning: \( B = V_b^{1/2} \)

2-level preconditioning: \( B = V_b^{1/2} \tilde{H}^{-1/2}(\bar{u},0) \)

**Iterative process**

\[ V^{p+1} = B \left( (2\pi)^{-M/2} |V^p|^{-1/2} \cdots \int_{-\infty}^{\infty} \tilde{H}^{-1}(\bar{u},v) \exp \left( -\frac{1}{2} v^T (V^p)^{-1} v \right) dv \cdots \right) B^* \]

\[ V^0 = H^{-1}(\bar{u},0), \quad p = 0,1,... \]

*This integral is a matrix which can be presented in compact form*

**Monte Carlo (MC) implementation**

\[ V^{p+1} = B \left( \frac{1}{L} \sum_{l=1}^{L} \tilde{H}^{-1}(\bar{u},\delta u^p_l) \right) B^*, \quad p = 0,1,... \]

\[ E[\delta u^p] = 0, \quad E[\delta u^p \left( \delta u^p \right)^T] = V^p \]

For integration instead of MC one can use quasi-MC or multi-pole method for faster convergence (smaller L)
EIH method: example - 1

Relative error in the variance estimate $\epsilon$ by the ‘effective’ IH (asymptotic) and IH

Envelope for relative error in the sample variance estimate for $L=25$ (black) and $L=100$ (white)

$V_I$ - based on a set of optimal solutions

$V_{II}$ - does not require optimal solutions

$\epsilon_i = (\hat{V}_{i,i} - 1)/\widehat{V}_{i,i}^\circ, \quad \hat{\epsilon}_i = (\hat{V}_L)_{i,i}/\widehat{V}_{i,i}^\circ - 1$

$i = 1, \ldots, m$

$\widehat{V}^\circ$ - reference covariance
(same sample covariance with large $L$)

Can be improved using ‘localization’, but requires optimal solutions!
Relative error in the variance estimate $\varepsilon$ by the ‘effective’ IH (asymptotic) and IH

Envelope for $\varepsilon$ by the ‘effective IH’, $L=25$ (black) and $L=100$ (red)

Envelope for relative error in the sample variance estimate for $L=25$ (black) and $L=100$ (white)

$V_I$ - based on a set of optimal solutions

$V_{II}$ - does not require optimal solutions

$\varepsilon_i = \frac{(V_{i,i})}{\hat{\Sigma}_{i,i}} - 1$, $\hat{\varepsilon}_i = \frac{(\hat{V}_L)_{i,i}}{\hat{\Sigma}_{i,i}} - 1$

$i = 1, \ldots, m$

$\hat{\Sigma}$ - reference covariance (sample covariance with large $L$ and after ‘sampling error compensation’ procedure)

Can be improved using ‘localization’, but requires optimal solutions!
EIH method: examples 1-2, correlation matrix

Example 1

Reference correlation matrix

Example 2

Error in the correlation matrix by EIH method

Error in the correlation matrix by IH method
About the origin error

Left figure shows a set of variance vectors, which correspond to the set of ‘optimal solutions’ in the right figure.

‘Optimal solutions’ are generated around the ‘exact’ solution using statistically significant implementations of the background and observation errors.

Any ‘optimal solution’ from the set could be the origin (instead of the exact solution)! Therefore, the origin error could be of the same magnitude as the linearization error.

For reliable (robust) estimation one could use the upper bound of the set.
1. Covariance propagation using the TLM

a) Direct computation
\[ V_{t=T} = MB\tilde{V}_{t=0}B^*M^* + Q \]

\( \tilde{V}_{t=0} \) - compact analysis covariance (quasi-Newton or Lanczos), \( M \) - TLM propagator, \( Q \) – model error,
\( B = V_b^{1/2} \) - preconditioner

If \( \tilde{V}_{t=0} \) is known in the form of singular values/vectors, then
\[ V_{t=T} = MB(I + U(S - I)U^*)B^*M^* + Q = MBIB^*M^* + Q + MB(U(S - I)U^*)B^*M^* \]

Propagation of the analysis covariance: not feasible could be feasible!
However, could be done using ‘reduced’ order model

b) One can look for a compact representation of \( V(T) \) using the product \( V(T)v, v \)- vector.

Notice: the pdf (tabular frequency) is only locally non-Gaussian!
Propagation of the analysis covariance-2

2. Covariance propagation using the nonlinear model

**Unscented transformation** (Julier, Uhlman 1996)
- propagation of $2m$ **sigma points** (square-root vectors of the posterior covariance).
  This is a variant of quasi-Monte Carlo. Captures the posterior mean and covariance accurately to the $3^{rd}$ order (in terms of Taylor expansion series) for any nonlinearity.

**Monte Carlo method** (Evensen?)
- propagation of $L$ **random vectors**, which satisfy the posterior pdf. For large $L$ should probably provide the same level of accuracy ($3^{rd}$ order)

**Is the $3^{rd}$ order of accuracy better than the $1^{st}$ order?**
Good for propagation one time step (small in terms of characteristic time-length of processes supported by the model), with subsequent update step. However, in variational DA one has to propagate over the whole time window. $t \in (0, T), T >> \Delta t$ !
Unexpectedly, the TLM propagator ($1^{st}$ order) sometimes performs better than the nonlinear propagator ($3^{rd}$ order), as it is more robust.
Let us notice that the non-Gaussian pdf appears locally, i.e. the linearly propagated covariance can be taken as an basic estimate, which has to be improved locally.

**Is it a computational problem or fundamental problem?**
If the dynamics is highly nonlinear (chaotic), MC with the nonlinear propagator may converge more slowly (as compared to the MC with TLM propagator), or non-uniformly, or may not converge at all
Conclusions

In the linear case the analysis covariance is equal to the inverse Hessian

Reasonably nonlinear case

In the nonlinear case, one must distinguish the *linearization* error (results from linearization of operators around the ‘exact’ state) and the *origin* error (results from the difference between known and exact states)

For an exact origin:

the inverse Hessian is expected to approximate well the analysis covariance if the tangent linear hypothesis (TLH) is valid. In practice, this approximation can be sufficiently accurate even though the TLH is not valid

if the nonlinear DA problem exhibits a ‘close-to-linear’ statistical behavior, then the analysis covariance can be approximated by the ‘effective’ inverse Hessian.

For an approximate origin:

the likely magnitude of the origin error can be revealed by considering a set of variance vectors generated around an optimal solution

the upper bound of this set can be chosen to achieve reliable (robust) sequential state estimation

In an extremely nonlinear case the analysis covariance

Does not represent the pDF (though locally!)
References

Gejadze, I., Copeland, G.J.M., Le Dimet, F.-X., Shutyaev, V.P.
Computation of the posterior covariance in variational data assimilation problems with nonlinear dynamics.
*J. Comp. Physics.* (2010), in review

Gejadze, I., Le Dimet, F.-X., Shutyaev, V.P.
On optimal solution error covariances in variational data assimilation problems.
*J. Comp. Physics.* (2010), v.229, pp.2159-2178

Gejadze, I., Le Dimet, F.-X., Shutyaev, V.P.
On analysis error covariances in variational data assimilation.