Hybrid 4DVAR and nonlinear EnKS method without tangents and adjoints

El houcine Bergou,
Supervisor : Serge Gratton

INP-ENSEEIHT, and CERFACS Toulouse

September 2013
Outline

1. Problem statement
2. Globalisation methods
3. A LM-EnKS method
4. Computational results
5. Ongoing work
Problem statement

- Data assimilation is a daily life process.
- We try to balance the uncertainty in the data and in the forecast.

Application fields are in geosciences (weather forecasting, hydrology,...).
Problem statement

Let consider the following stochastic non necessary linear system:

\[
\begin{align*}
X_0 &= x_b + V_0, \quad V_0 \sim N(0, B) \\
X_i &= M_i(X_{i-1}) + V_i, \quad V_i \sim N(0, Q_i) \\
d_i &= H_i(X_i) + W_i, \quad W_i \sim N(0, R_i)
\end{align*}
\]

- \(X_i\) is the \(n\) dimensional state at time \(i\); it is random,
- \(d_i\) is the random observation vector at time \(i\),
- \(M_i\) is the (nonlinear) model propagator at time \(i\),
- \(H_i\) is the observation operator at time \(i\), it is not linear,
- \(x_b\) is background vector,
- \(B\) is the background error covariance matrix,
- \(Q_i\) and \(R_i\) are respectively the model, and observation, error covariance matrices at time \(i\),
Problem statement

Our goal is to find the best estimate of the state \( X_0, \ldots, X_k \) knowing the data set \( d_1, \ldots, d_k \).

4DVAR method solves this problem, in the sense of minimizing the sum of the squares of the errors, weighted by the error covariance matrices.
Our goal is to find the best estimate of the state $X_0, \ldots, X_k$ knowing the data set $d_1, \ldots, d_k$.

4DVAR method solves this problem, in the sense of minimizing the sum of the squares of the errors, weighted by the error covariance matrices.
Weak constraint 4DVar

⇒ nonlinear least-squares problem

\[ J(x_{0:k}) = \|x_0 - x_b\|^2_B - 1 + \sum_{i=1}^{k} \|x_i - M_i (x_{i-1})\|^2_{Q_i} - 1 + \sum_{i=1}^{k} \|d_i - H_i (x_i)\|^2_{R_i} - 1 \rightarrow \min_{x_{0:k}} \]
In the linear case, Kalman Filter and Kalman Smoother (KF and KS) and their Ensemble variants (EnKF and EnKS) (Evensen, 2009) give the pdf (mean and covariance) of the state knowing the data set.

In the non linear case, variants of the Kalman Filter were proposed such as Extended Kalman Filter (EKF). These methods may fail to find a minimum of 4DVar, especially for highly non linear case.

Iterated Kalman filter or 4DVar Incremental approach (Courtier et al., 1994) may also fail to converge.
Weak constraint 4DVar

- In the linear case Kalman Filter and Kalman Smoother (KF and KS) and their Ensemble variants (EnKF and EnKS) (Evensen, 2009) give the pdf (mean and covariance) of the state knowing the data set,

- In the non linear case, variants of the Kalman Filter were proposed such as Extended Kalman Filter (EKF). These methods may fail to find a minimum of 4DVar, especially for highly non linear case,

- Iterated Kalman filter or 4DVar Incremental approach (Courtier et al., 1994) may also fail to converge.
Weak constraint 4DVar

- In the linear case, Kalman Filter and Kalman Smoother (KF and KS) and their Ensemble variants (EnKF and EnKS) (Evensen, 2009) give the pdf (mean and covariance) of the state knowing the data set.

- In the non-linear case, variants of the Kalman Filter were proposed such as Extended Kalman Filter (EKF). These methods may fail to find a minimum of 4DVar, especially for highly non-linear case.

- Iterated Kalman filter or 4DVar Incremental approach (Courtier et al., 1994) may also fail to converge.
Incremental 4DVar

- **Incremental approach** (Courtier et al., 1994): linearization

\[ M_i (x_{i-1} + \delta x_{i-1}) \approx M_i (x_{i-1}) + M_i' (x_{i-1}) \delta x_{i-1}, \]
\[ H_i (x_i + \delta x_i) \approx H_i (x_i) + H_i' (x_i) \delta x_i, \]

- gives the **Gauss-Newton method**, (Bell, 1994), (Nichols et al., 2007) iterations
\[
x_0:0 \leftarrow x_0:0 + \delta x_0:0 \text{ with the linear least-squares problem for the increments}
\]
\[
\| x_0 + \delta x_0 - x_b \|^2_{B^{-1}} + \sum_{i=1}^{k} \| d_i - H_i (x_i) - H_i' (x_i) \delta x_i \|^2_{R_i^{-1}}
\]
\[
+ \sum_{i=1}^{k} \| x_i + \delta x_i - M_i (x_{i-1}) - M_i' (x_{i-1}) \delta x_{i-1} \|^2_{Q_i^{-1}}
\]
Incremental 4DVar

- Incremental approach (Courtier et al., 1994): linearization
  \[ \mathcal{M}_i (x_{i-1} + \delta x_{i-1}) \approx \mathcal{M}_i (x_{i-1}) + \mathcal{M}'_i (x_{i-1}) \delta x_{i-1}, \]
  \[ \mathcal{H}_i (x_i + \delta x_i) \approx \mathcal{H}_i (x_i) + \mathcal{H}'_i (x_i) \delta x_i, \]

- gives the Gauss-Newton method, (Bell, 1994), (Nichols et al., 2007) iterations \( x_{0:k} \leftarrow x_{0:k} + \delta x_{0:k} \) with the linear least-squares problem for the increments

\[
\| x_0 + \delta x_0 - x_b \|^2_{B^{-1}} + \sum_{i=1}^{k} \| d_i - \mathcal{H}_i (x_i) - \mathcal{H}'_i (x_i) \delta x_i \|^2_{R_i^{-1}}
\]

\[
+ \sum_{i=1}^{k} \| x_i + \delta x_i - \mathcal{M}_i (x_{i-1}) - \mathcal{M}'_i (x_{i-1}) \delta x_{i-1} \|^2_{Q_i^{-1}}
\]
Incremental 4DVar

\[
\|x_0 + \delta x_0 - x_b\|^2_{B^{-1}} + \sum_{i=1}^{k} \|d_i - H_i(x_i) - H'_i(x_i)\delta x_i\|^2_{R_i^{-1}}
\]

\[
+ \sum_{i=1}^{k} \|x_i + \delta x_i - M_i(x_{i-1}) - M'_i(x_{i-1})\delta x_{i-1}\|^2_{Q_i^{-1}}
\]

- **Tangent and adjoint** code needed,
- **Is difficult to parallelize,**
- **May fail to converge.**
Incremental 4DVar

\[
\|x_0 + \delta x_0 - x_b\|_B^{-1} + \sum_{i=1}^{k} \left\|d_i - \mathcal{H}_i (x_i) - \mathcal{H}'_i (x_i) \delta x_i\right\|_R^{-1}
\]

\[
+ \sum_{i=1}^{k} \left\|x_i + \delta x_i - \mathcal{M}_i (x_{i-1}) - \mathcal{M}'_i (x_{i-1}) \delta x_{i-1}\right\|_Q^{-1}
\]

- **Tangent and adjoint** code needed,
- **Is difficult** to parallelize,
- May fail to converge.
Incremental 4DVar

\[
\|x_0 + \delta x_0 - x_b\|^2_{B^{-1}} + \sum_{i=1}^{k} \|d_i - \mathcal{H}_i(x_i) - \mathcal{H}_i'(x_i) \delta x_i\|^2_{R_i^{-1}} \\
+ \sum_{i=1}^{k} \|x_i + \delta x_i - M_i(x_{i-1}) - M_i'(x_{i-1}) \delta x_{i-1}\|^2_{Q_i^{-1}}
\]

- **Tangent and adjoint** code needed,
- **Is difficult** to parallelize,
- **May fail** to converge.
Outline

1. Problem statement
2. Globalisation methods
3. A LM-EnKS method
4. Computational results
5. Ongoing work

Levenberg-Marquart Method
Linearized 4DVar as Kalman smoother
Derivative-free implementation of the EnKS - model
Globalisation methods

- **Convergence from any starting point** obtained with the globalization techniques based on the **control of the size of the increments**.
  - **Trust region method**: at each iteration a linearized problem is solved within a region where the linear approximation is trusted.
  - **Levenberg-Marquart method**: a penalized variant of the nonlinear least-squares problem is solved.
Globalisation methods

- **Convergence from any starting point** obtained with the globalization techniques based on the control of the size of the increments.
  - **Trust region method**: at each iteration a linearized problem is solved within a region where the linear approximation is trusted.
  - **Levenberg-Marquart method**: a penalized variant of the nonlinear least-squares problem is solved.
Globalisation methods

- Convergence from any starting point obtained with the globalization techniques based on the control of the size of the increments.
  - Trust region method: at each iteration a linearized problem is solved within a region where the linear approximation is trusted.
  - Levenberg-Marquart method: a penalized variant of the nonlinear least-squares problem is solved.
Levenberg-Marquart Method

- Add a penalty (Tikhonov regularization) to control the size of the increments,
- Let consider the following nonlinear least-squares:

\[
\arg \min_{x \in \mathbb{R}^n} F(x) = \| f(x) \|^2,
\]

where \( f \) from \( \mathbb{R}^n \rightarrow \mathbb{R}^m \) is a (possibly nonlinear) function.
- In the Levenberg-Marquart method, at each iteration we solve the linear least-squares problem:

\[
FL(x_j + \delta x) = \| f(x_j) + J_f(x_j)\delta x \|^2 + \gamma \| \delta x \|^2 \rightarrow \min_{\delta x},
\]

where \( x_j \) is the j-th iterate, \( J_f(x_j) \) is the Jacobian of \( f \) at \( x_j \) and \( \gamma \) is the regularization parameter,
Levenberg-Marquart Method

- Add a penalty (Tikhonov regularization) to control the size of the increments,
- Let consider the following nonlinear least-squares:

\[
\arg \min_{x \in \mathbb{R}^n} F(x) = \|f(x)\|^2,
\]

where \( f \) from \( \mathbb{R}^n \rightarrow \mathbb{R}^m \) is a (possibly nonlinear) function.

- In the Levenberg-Marquart method, at each iteration we solve the linear least-squares problem:

\[
FL(x_j + \delta x) = \|f(x_j) + J_f(x_j)\delta x\|^2 + \gamma \|\delta x\|^2 \rightarrow \min_{\delta x},
\]

where \( x_j \) is the j-th iterate, \( J_f(x_j) \) is the Jacobian of \( f \) at \( x_j \) and \( \gamma \) is the regularization parameter,
Levenberg-Marquart Method

- Add a penalty (Tikhonov regularization) to control the size of the increments,
- Let consider the following nonlinear least-squares:

\[
\arg \min_{x \in \mathbb{R}^n} F(x) = \|f(x)\|^2,
\]

where \( f \) from \( \mathbb{R}^n \to \mathbb{R}^m \) is a (possibly nonlinear) function.

- In the Levenberg-Marquart method, at each iteration we solve the linear least-squares problem:

\[
FL(x_j + \delta x) = \|f(x_j) + J_f(x_j)\delta x\|^2 + \gamma \|\delta x\|^2 \rightarrow \min_{\delta x},
\]

where \( x_j \) is the j-th iterate, \( J_f(x_j) \) is the Jacobian of \( f \) at \( x_j \) and \( \gamma \) is the regularization parameter,
The solution of this linear least-squares is a solution of the normal equation

$$(J_f(x_j)^T J_f(x_j) + \gamma I)\delta x = -J_f(x_j)f(x_j) = -\nabla F(x_j),$$

- When $\gamma = 0$, $\delta x = -(J_f(x_j)^T J_f(x_j))^{-1}\nabla F(x_j) = $ Incremental method (Gauss-Newton)(fast convergence).
- When $\gamma \to \infty$, $\delta x \to 0$ and it is positively proportional to $-\nabla F(x_j)$ (steepest descent),
- When $0 < \gamma < \infty$ there is a balance between the Gauss-Newton direction and steepest descent direction,
  $\Rightarrow$ The term $\gamma \|\delta x\|^2$ controls the step size as well as rotates the step direction towards the steepest descent.
Levenberg-Marquart Method

The solution of this linear least-squares is a solution of the normal equation

\[(J_f(x_j)^T J_f(x_j) + \gamma I)\delta x = -J_f(x_j)f(x_j) = -\nabla F(x_j),\]

When \(\gamma = 0\), \(\delta x = -(J_f(x_j)^T J_f(x_j))^{-1} \nabla F(x_j)\) (Incremental method (Gauss-Newton) (fast convergence)).

When \(\gamma \to \infty\), \(\delta x \to 0\) and it is positively proportional to \(-\nabla F(x_j)\) (steepest descent),

When \(0 < \gamma < \infty\) there is a balance between the Gauss-Newton direction and steepest descent direction,

\[\Rightarrow \text{ The term } \gamma \|\delta x\|^2 \text{ controls the step size as well as rotates the step direction towards the steepest descent.}\]
Levenberg-Marquart Method

- The solution of this linear least-squares is a solution of the normal equation

\[
(J_f(x_j)^T J_f(x_j) + \gamma I)\delta x = -J_f(x_j)f(x_j) = -\nabla F(x_j),
\]

- When \( \gamma = 0 \), \( \delta x = -(J_f(x_j)^T J_f(x_j))^{-1}\nabla F(x_j) = \) Incremental method (Gauss-Newton) (fast convergence).

- When \( \gamma \rightarrow \infty \), \( \delta x \rightarrow 0 \) and it is positively proportional to \( -\nabla F(x_j) \) (steepest descent).

- When \( 0 < \gamma < \infty \) there is a balance between the Gauss-Newton direction and steepest descent direction,

\[ \Rightarrow \text{The term } \gamma \|\delta x\|^2 \text{ controls the step size as well as rotates the step direction towards the steepest descent.} \]
Levenberg-Marquart Method

- The solution of this linear least-squares is a solution of the normal equation

\[(J_f(x_j)^T J_f(x_j) + \gamma I) \delta x = -J_f(x_j) f(x_j) = -\nabla F(x_j),\]

- When \( \gamma = 0 \), \( \delta x = - (J_f(x_j)^T J_f(x_j))^{-1} \nabla F(x_j) \) = Incremental method (Gauss-Newton) (fast convergence).
- When \( \gamma \to \infty \), \( \delta x \to 0 \) and it is positively proportional to \( -\nabla F(x_j) \) (steepest descent),
- When \( 0 < \gamma < \infty \) there is a balance between the Gauss-Newton direction and steepest descent direction,

\[\Rightarrow \text{The term } \gamma \| \delta x \|^2 \text{ controls the step size as well as rotates the step direction towards the steepest descent.}\]
In LM method $\gamma$ may remain constant over the iterations, or adaptive:

1. $\gamma$ remains constant, it must be chosen large enough to ensure the convergence.
2. $\gamma$ adaptive: at each iteration we compute

$$\psi = \frac{F(x_j) - F(x_j + \delta x)}{F(x_j) - FL(x_j + \delta x)}$$

- If $\psi \geq \sigma > 0$, we decrease $\gamma$.
- Else, we increase $\gamma$.
In LM method $\gamma$ may remain constant over the iterations, or adaptive:

1. $\gamma$ remains constant, it must be chosen large enough to ensure the convergence.
2. $\gamma$ adaptive: at each iteration we compute

$$
\psi = \frac{F(x_j) - F(x_j + \delta x)}{F(x_j) - FL(x_j + \delta x)}
$$

- If $\psi \geq \sigma > 0$, we decrease $\gamma$,
- Else, we increase $\gamma$. 
In LM method $\gamma$ may remain constant over the iterations, or adaptive:

1. $\gamma$ remains constant, it must be chosen large enough to ensure the convergence.
2. $\gamma$ adaptive: at each iteration we compute

$$\psi = \frac{F(x_j) - F(x_j + \delta x)}{F(x_j) - FL(x_j + \delta x)}$$

- If $\psi \geq \sigma > 0$, we decrease $\gamma$,
- Else, we increase $\gamma$, 

Levenberg-Marquart Method
Linearized 4DVar as Kalman smoother
Derivative-free implementation of the EnKS - model
In LM method $\gamma$ may remain constant over the iterations, or adaptive:

1. $\gamma$ remains constant, it must be chosen large enough to ensure the convergence.
2. $\gamma$ adaptive: at each iteration we compute

$$\psi = \frac{F(x_j) - F(x_j + \delta x)}{F(x_j) - FL(x_j + \delta x)}$$

- If $\psi \geq \sigma > 0$, we decrease $\gamma$,
- Else, we increase $\gamma$, 

Hybrid 4DVAR and nonlinear EnKS method
In LM method $\gamma$ may remain constant over the iterations, or adaptive:

1. $\gamma$ remains constant, it must be chosen large enough to ensure the convergence.
2. $\gamma$ adaptive: at each iteration we compute

$$\psi = \frac{F(x_j) - F(x_j + \delta x)}{F(x_j) - FL(x_j + \delta x)}$$

- If $\psi \geq \sigma > 0$, we decrease $\gamma$,
- Else, we increase $\gamma$, 

Levenberg-Marquart Method
Linearized 4DVar as Kalman smoother
Derivative-free implementation of the EnKS - model
Levenberg-Marquart Method

In 4DVar linearized problem we add regularization as follows,

\[ \tilde{J}(x_0:k) = \|x_0 + \delta x_0 - x_b\|^2_{B^{-1}} \]

\[ + \sum_{i=1}^{k} \|x_i + \delta x_i - \mathcal{M}_i(x_{i-1}) - \mathcal{M}_i'(x_{i-1}) \delta x_{i-1}\|^2_{Q_i^{-1}} \]

\[ + \sum_{i=1}^{k} \|d_i - \mathcal{H}_i(x_i) - \mathcal{H}_i'(x_i) \delta x_i\|^2_{R_i^{-1}} + \gamma \sum_{i=0}^{k} \|\delta x_i\|^2_{S_i^{-1}} \]

Assume the model and observation operator are regular and that \(\gamma\) is larger than a problem dependent constant: The gradient of the iterates goes to 0 for any initial iterate (global convergence property).
Levenberg-Marquart Method

In 4DVar linearized problem we add regularization as follows,

$$
\tilde{J}(x_0;k) = \| x_0 + \delta x_0 - x_b \|^2_{B^{-1}} \\
+ \sum_{i=1}^{k} \| x_i + \delta x_i - M_i(x_{i-1}) - M'_i(x_{i-1})\delta x_{i-1} \|^2_{Q_i^{-1}} \\
+ \sum_{i=1}^{k} \| d_i - H_i(x_i) - H'_i(x_i)\delta x_i \|^2_{R_i^{-1}} + \gamma \sum_{i=0}^{k} \| \delta x_i \|^2_{S_i^{-1}}
$$

Assume the model and observation operator are regular and that $\gamma$ is larger than a problem dependent constant: The gradient of the iterates goes to 0 for any initial iterate (global convergence property).
Write the linear least-squares problem for the increments $z_{0:k} = \delta x_{0:k}$ as

$$\| z_0 - z_b \|^2_{B^{-1}} + \sum_{i=1}^{k} \| z_i - M_i z_{i-1} - m_i \|^2_{Q_i^{-1}} + \sum_{i=1}^{k} \| d_i - H_i z_i \|^2_{R_i^{-1}}$$

where

- $z_b = x_b - x_0$
- $m_i = M_i (x_{i-1}) - x_i$
- $d_i = d_i - H_i (x_i)$
- $M_i = M_i' (x_{i-1})$
- $H_i = H_i' (x_i)$

This is the same function as minimized in the Kalman smoother for the following linear and gaussian system (Rauch et al., 1965; Bell, 1994)

- $Z_0 = z_b + V_0$, $V_0 \sim N(0, B)$
- $Z_i = M_i Z_{i-1} + m_i + V_i$, $V_i \sim N(0, Q_i)$
- $d_i = H_i Z_i + W_i$, $W_i \sim N(0, R_i)$
Linearized 4DVar as Kalman smoother

Write the linear least-squares problem for the increments $z_{0:k} = \delta x_{0:k}$ as

$$
\| z_0 - z_b \|_B^{-1} + \sum_{i=1}^{k} \| z_i - M_i z_{i-1} - m_i \|_{Q_i}^{-1} + \sum_{i=1}^{k} \| d_i - H_i z_i \|_{R_i}^{-1}
$$

$$
\begin{align*}
z_b &= x_b - x_0, & m_i &= \mathcal{M}_i (x_{i-1}) - x_i, & d_i &= d_i - \mathcal{H}_i (x_i) , \\
M_i &= \mathcal{M}_i' (x_{i-1}) , & H_i &= \mathcal{H}_i' (x_i)
\end{align*}
$$

- This is the same function as minimized in the Kalman smoother for the following linear and gaussian system (Rauch et al., 1965; Bell, 1994)

$$
\begin{align*}
Z_0 &= z_b + V_0, & V_0 &\sim N(0, B) \\
Z_i &= M_i Z_{i-1} + m_i + V_i, & V_i &\sim N(0, Q_i) \\
d_i &= H_i Z_i + W_i, & W_i &\sim N(0, R_i)
\end{align*}
$$

Hybrid 4DVAR and nonlinear EnKS method
Problem statement
Globalisation methods
A LM-EnKS method
Computational results
Ongoing work
References

Levenberg-Marquart Method
Linearized 4DVar as Kalman smoother
Derivative-free implementation of the EnKS - model

Ensemble Kalman filter (EnKF) and smoother (EnKS)

\[ Z_i^N = [z_i^1, \ldots, z_i^N] \] is ensemble of states at time \( i \), conditioned on all data up to time \( k \).

**Algorithm (EnKF)**

1. **Initialize**
   \[ z_{0|0}^\ell \sim N (z_b, B), \quad \ell = 1, \ldots, N. \] (1)

2. **For** \( i = 1, \ldots, k \), **advance in time**
   \[ z_{i|i-1}^\ell = M_i z_{i-1|i-1}^\ell + m_i + v_i^\ell, \quad v_i^\ell \sim N (0, Q_i), \] (2)
   \[ z_{i|i}^\ell = z_{i|i-1}^\ell - P_i^N H_i^T (H_i P_i^N H_i^T + R_i)^{-1} \]
   \[ \quad \times (H_i z_{i|i-1}^\ell - d_i - w_i^\ell), \quad w_i^\ell \sim N (0, R_i), \] (3)
Ensemble Kalman filter (EnKF) and smoother (EnKS)

- The EnKS is obtained by applying the same analysis step (3) as in the EnKF to the composite state $Z_{0:i}|i-1$ from time 0 to $i$, conditioned on data up to time $i-1$,

$$Z_0^N = \begin{bmatrix} Z_0^N |_{i-1} \\ Z_{i-1}^N |_{i-1} \\
... \\
Z_i^N |_{i-1} \end{bmatrix}. $$

in the place of $Z_i|i-1$.
- The observation term $H_i Z_{i}^N |_{i-1} - d_i$ becomes

$$ [0, \ldots, H_i] Z_{0:i}^N |_{i-1} - d_i = H_i Z_{i}^N |_{i-1} - d_i. \quad (5) $$
Ensemble Kalman filter (EnKF) and smoother (EnKS)

- The EnKS is obtained by applying the same analysis step (3) as in the EnKF to the composite state \( Z_{0:i|i-1} \) from time 0 to \( i \), conditioned on data up to time \( i - 1 \),

\[
Z_{0:i|i-1}^N = \begin{bmatrix}
Z_{0|i-1}^N \\
\vdots \\
Z_{i|i-1}^N
\end{bmatrix}.
\]

in the place of \( Z_{i|i-1} \).

- The observation term \( H_i Z_{i|i-1}^N - d_i \) becomes

\[
[0, \ldots, H_i] Z_{0:i|i-1}^N - d_i = H_i Z_{i|i-1}^N - d_i. \tag{5}
\]
Ensemble Kalman filter (EnKF) and smoother (EnKS)

Algorithm (EnKS)

Given $z_b$,

1. Initialize $z^\ell_{0|0} \sim N(z_b, B), \ell = 1, \ldots, N$.

2. For $i = 1, \ldots, k$, advance in time,

$$z^\ell_{i|i-1} = M_i z^\ell_{i-1|i-1} + m_i + v_i, \quad v_i \sim N(0, Q_i), \quad (6)$$

$$Z^N_{0:i|i} = Z^N_{0:i|i-1} - P^N_{0:i,0:i-1} \tilde{H}_0:i T (\tilde{H}_0:i P_{0:i,0:i-1} \tilde{H}_0:i T + R_i)^{-1} (7)$$

$$\cdot (\tilde{H}_0:i Z^N_{0:i|i-1} - d_i - w_i), \quad w_i \sim N(0, R_i), \quad (8)$$

where $\tilde{H}_0:i = [0, \ldots, H_i]$, and $P^N_{0:i,0:i-1}$ is the sample covariance matrix of $Z^N_{0:i|i-1}$. 
Derivative-free implementation of the EnKS - model

The linearized model \( M_i = M'_i (x_{i-1}) \) occurs only in advancing the time as an action on the ensemble \( Z^N = [z^n] = [\delta x^n] \),

\[
M_i \delta x^n_{i-1} + m_i = M'_i (x_{i-1}) \delta x^n_{i-1} + M_i (x_{i-1}) - x_i,
\]

Approximating by finite differences with a parameter \( \tau > 0 \):

\[
M_i \delta x^n_{i-1} + m_i \approx \frac{M_i (x_{i-1} + \tau \delta x^n_{i-1}) - M_i (x_{i-1})}{\tau} + M_i (x_{i-1}) - x_i,
\]

Accurate in the limit \( \tau \to 0 \).
Derivative-free implementation of the EnKS - model

The linearized model $M_i = M'_i (x_{i-1})$ occurs only in advancing the time as an action on the ensemble $Z^N = [z^n] = [\delta x^n]$, 

$$M_i \delta x^n_{i-1} + m_i = M'_i (x_{i-1}) \delta x^n_{i-1} + M_i (x_{i-1}) - x_i,$$

Approximating by finite differences with a parameter $\tau > 0$:

$$M_i \delta x^n_{i-1} + m_i \approx \frac{M_i (x_{i-1} + \tau \delta x^n_{i-1}) - M_i (x_{i-1})}{\tau} + M_i (x_{i-1}) - x_i,$$

Accurate in the limit $\tau \to 0$. 

Hybrid 4DVAR and nonlinear EnKS method
The linearized model \( M_i = M'_i(x_{i-1}) \) occurs only in advancing the time as an action on the ensemble \( Z^N = [z^n] = [\delta x^n] \),

\[
M_i \delta x^n_{i-1} + m_i = M'_i(x_{i-1}) \delta x^n_{i-1} + M_i(x_{i-1}) - x_i,
\]

Approximating by finite differences with a parameter \( \tau > 0 \):

\[
M_i \delta x^n_{i-1} + m_i \approx \frac{M_i(x_{i-1} + \tau \delta x^n_{i-1}) - M_i(x_{i-1})}{\tau} + M_i(x_{i-1}) - x_i,
\]

Accurate in the limit \( \tau \to 0 \).
The linearized model $M_i = M'_i(x_{i-1})$ occurs only in advancing the time as an action on the ensemble $Z^N = [z^n] = [\delta x^n]$, 

$$M_i \delta x^n_{i-1} + m_i = M'_i(x_{i-1}) \delta x^n_{i-1} + M_i(x_{i-1}) - x_i,$$

Approximating by finite differences with a parameter $\tau > 0$:

$$M_i \delta x^n_{i-1} + m_i \approx \frac{M_i (x_{i-1} + \tau \delta x^n_{i-1}) - M_i (x_{i-1})}{\tau} + M_i (x_{i-1}) - x_i,$$

Accurate in the limit $\tau \to 0$. 

Hybrid 4DVAR and nonlinear EnKS method
The linearized model $M_i = M'_i(x_{i-1})$ occurs only in advancing the time as an action on the ensemble $Z^N = [z^n] = [\delta x^n]$, \[
M_i \delta x^n_{i-1} + m_i = M'_i(x_{i-1}) \delta x^n_{i-1} + M_i(x_{i-1}) - x_i,
\]
Approximating by finite differences with a parameter $\tau > 0$: \[
M_i \delta x^n_{i-1} + m_i \approx \frac{M_i(x_{i-1} + \tau \delta x^n_{i-1}) - M_i(x_{i-1})}{\tau} + M_i(x_{i-1}) - x_i,
\]
Accurate in the limit $\tau \to 0$. 

Hybrid 4DVAR and nonlinear EnKS method
The linearized model $M_i = M'_i(x_{i-1})$ occurs only in advancing the time as an action on the ensemble $Z^N = [z^n] = [\delta x^n]$, 

$$M_i \delta x^n_{i-1} + m_i = M'_i(x_{i-1}) \delta x^n_{i-1} + M_i(x_{i-1}) - x_i,$$

Approximating by finite differences with a parameter $\tau > 0$: 

$$M_i \delta x^n_{i-1} + m_i \approx \frac{M_i(x_{i-1} + \tau \delta x^n_{i-1}) - M_i(x_{i-1})}{\tau} + M_i(x_{i-1}) - x_i,$$

Accurate in the limit $\tau \rightarrow 0$. 

Hybrid 4DVAR and nonlinear EnKS method
Derivative-free implementation of the EnKS - model

The linearized model $M_i = M'_i (x_{i-1})$ occurs only in advancing the time as an action on the ensemble $Z^N = [z^n] = [\delta x^n],$

$$M_i \delta x^n_{i-1} + m_i = M'_i (x_{i-1}) \delta x^n_{i-1} + M_i (x_{i-1}) - x_i,$$

Approximating by finite differences with a parameter $\tau > 0$:

$$M_i \delta x^n_{i-1} + m_i \approx \frac{M_i (x_{i-1} + \tau \delta x^n_{i-1}) - M_i (x_{i-1})}{\tau} + M_i (x_{i-1}) - x_i,$$

Accurate in the limit $\tau \to 0.$
The observation matrix occurs only in the action on the ensemble,

\[ H_i Z^N = \left[ H_i \delta x^1, \ldots, H_i \delta x^N \right]. \]

Approximating by finite differences with a parameter \( \tau > 0 \):

\[ H_i \delta x_i^n \approx \frac{H_i (x_{i-1} + \tau \delta x_i^n) - H_i (x_{i-1})}{\tau}, \]

Accurate in the limit \( \tau \to 0 \).

In tests below, \( \tau = 0.1 \) seems to work well enough. But this is application dependent: smaller \( \tau \) may be needed.
Derivative-free implementation of the EnKS - observation

The observation matrix occurs only in the action on the ensemble,

\[ H_i Z^N = [H_i \delta x^1, \ldots, H_i \delta x^N]. \]

Approximating by finite differences with a parameter \( \tau > 0 \):

\[ H_i \delta x^n_i \approx \frac{H_i \left( x_{i-1} + \tau \delta x^n_{i-1} \right) - H_i \left( x_{i-1} \right)}{\tau}, \]

Accurate in the limit \( \tau \rightarrow 0 \).

In tests below, \( \tau = 0.1 \) seems to work well enough. But this is application dependent: smaller \( \tau \) may be needed.
Derivative-free implementation of the EnKS - observation

The observation matrix occurs only in the action on the ensemble,

\[ H_i Z^N = \left[ H_i \delta x^1, \ldots, H_i \delta x^N \right] . \]

Approximating by finite differences with a parameter \( \tau > 0 \):

\[ H_i \delta x^n_i \approx \frac{H_i \left( x_{i-1} + \tau \delta x^n_{i-1} \right) - H_i \left( x_{i-1} \right)}{\tau} , \]

Accurate in the limit \( \tau \to 0 \).

In tests below, \( \tau = 0.1 \) seems to work well enough. But this is application dependent: smaller \( \tau \) may be needed.
The observation matrix occurs only in the action on the ensemble,

\[ H_i Z^N = \left[ H_i \delta x^1, \ldots, H_i \delta x^N \right]. \]

Approximating by finite differences with a parameter \( \tau > 0 \):

\[ H_i \delta x^n_i \approx \frac{H_i \left( x_{i-1} + \tau \delta x^n_{i-1} \right) - H_i (x_{i-1})}{\tau}, \]

Accurate in the limit \( \tau \to 0 \).

In tests below, \( \tau = 0.1 \) seems to work well enough. But this is application dependent: smaller \( \tau \) may be needed.
The observation matrix occurs only in the action on the ensemble,

\[ H_i Z^N = [H_i \delta x^1, \ldots, H_i \delta x^N]. \]

Approximating by finite differences with a parameter \( \tau > 0 \):

\[ H_i \delta x^n_i \approx \frac{H_i (x_{i-1} + \tau \delta x^n_{i-1}) - H_i (x_{i-1})}{\tau}, \]

Accurate in the limit \( \tau \to 0 \).

In tests below, \( \tau = 0.1 \) seems to work well enough. But this is application dependent: smaller \( \tau \) may be needed.
Outline

1. Problem statement
2. Globalisation methods
3. A LM-EnKS method
4. Computational results
5. Ongoing work
A LM-EnKS method

Given $x_0, x_1, ..., x_k, \gamma, \lambda > 1, \tau \leq 1, \sigma < 1$.

For outer loop $= 1, 2, ...$

- Initialize $z^\ell_{0|0} \sim N(0, B)$, for $\ell = 1, \ldots, N$

- for $i = 1, \ldots, k$ advance $z^\ell$ in time following (2), with the linearized operator approximated by finite differences:

$$z^\ell_{i|i-1} = \frac{M_i \left(x_{i-1} + \tau z^\ell_{i-1|i-1}\right) - M_i \left(x_{i-1}\right)}{\tau}$$

$$+ M_i \left(x_{i-1}\right) - x_i + v^\ell_i, \quad v^\ell_i \sim N(0, Q_i)$$

followed by the smoother analysis step with matrix-vector products $H_i z_i$ approximated by finite differences.
A LM-EnKS method

Given \( x_0, x_1, \ldots, x_k, \gamma, \lambda > 1, \tau \leq 1, \sigma < 1. \)

For outer loop = 1,2,...

- Initialize \( z_{0|0}^{\ell} \sim N(0, B) \), for \( \ell = 1, \ldots, N \)

- for \( i = 1, \ldots, k \) advance \( z_{i}^{\ell} \) in time following (2), with the linearized operator approximated by finite differences:

\[
\begin{align*}
z_{i|i-1}^{\ell} & = \frac{\mathcal{M}_i \left( x_{i-1} + \tau z_{i-1|i-1}^{\ell} \right) - \mathcal{M}_i \left( x_{i-1} \right)}{\tau} \\
& + \mathcal{M}_i \left( x_{i-1} \right) - x_i + v_i^{\ell}, \quad v_i^{\ell} \sim N(0, Q_i)
\end{align*}
\]

followed by the smoother analysis step with matrix-vector products \( H_i z_i \) approximated by finite differences.
Tikhonov regularization is considered as a further observations

\[ \tilde{d}_i = 0 = z_i + \tilde{W}_i \quad \tilde{W}_i \sim N \left( 0, \frac{1}{\gamma} S_i \right), \]

simply run the analysis step the second time with observation operator equal to identity and observation error covariance equal to \( \frac{1}{\gamma} S_i \).

\[ x_i \leftarrow x_i + \frac{1}{N} \sum_{\ell=1}^{N} z_{i|k}^{\ell}, \quad i = 1, \ldots, k \]
Tikhonov regularization is considered as a further observations

\[ \tilde{d}_i = 0 = z_i + \tilde{W}_i \quad \tilde{W}_i \sim N \left( 0, \frac{1}{\gamma S_i} \right), \]

simply run the analysis step the second time with observation operator equal to identity and observation error covariance equal to \( \frac{1}{\gamma} S_i \).

\[ x_i \leftarrow x_i + \frac{1}{N} \sum_{\ell=1}^{N} z_i^\ell |_{k}, \quad i = 1, \ldots, k \]
Outline

1. Problem statement
2. Globalisation methods
3. A LM-EnKS method
4. Computational results
5. Ongoing work

LM-EnKS for Lorenz 63 model
An example where Gauss-Newton does not converge

Hybrid 4DVAR and nonlinear EnKS method
Computational results

- To evaluate the performance of the method, we use the twin experiment technique.
- Truth = an integration of the model over time,
- We obtain the background by adding a gaussian perturbation to the initial state,
- We obtain the data $d_i$ by applying the observation operator $H_i$ to the truth and then adding a gaussian perturbation,
- Try to recover the truth using LM-EnKS method.
Computational results

- To evaluate the performance of the method, we use the twin experiment technique.
- Truth = an integration of the model over time,
- We obtain the background by adding a gaussian perturbation to the initial state,
- We obtain the data $d_i$ by applying the observation operator $H_i$ to the truth and then adding a gaussian perturbation,
- Try to recover the truth using LM-EnKS method.
Computational results

- To evaluate the performance of the method, we use the twin experiment technique.
- Truth = an integration of the model over time,
- We obtain the background by adding a gaussian perturbation to the initial state,
- We obtain the data $d_i$ by applying the observation operator $\mathcal{H}_i$ to the truth and then adding a gaussian perturbation,
- Try to recover the truth using LM-EnKS method.
Computational results

- To evaluate the performance of the method, we use the twin experiment technique.
- Truth = an integration of the model over time,
- We obtain the background by adding a gaussian perturbation to the initial state,
- We obtain the data $d_i$ by applying the observation operator $H_i$ to the truth and then adding a gaussian perturbation,
- Try to recover the truth using LM-EnKS method.
To evaluate the performance of the method, we use the twin experiment technique.

Truth = an integration of the model over time,

We obtain the background by adding a gaussian perturbation to the initial state,

We obtain the data $d_i$ by applying the observation operator $H_i$ to the truth and then adding a gaussian perturbation,

Try to recover the truth using LM-EnKS method.
Computational results
Lorenz 63 model

\[
\begin{align*}
\frac{dx}{dt} &= -\sigma (x - y) \\
\frac{dy}{dt} &= \rho x - y - xz \\
\frac{dz}{dt} &= xy - \beta z
\end{align*}
\]

$\sigma$, $\rho$ and $\beta$ are chosen to have the values 10, 28 and $8/3$ respectively.
Computational results

Parameters of the experiment

- The system is discretized using the fourth-order Runge-Kutta method.

\[ B = \sigma_b^2 \text{diag} \left( 1, \frac{1}{4}, \frac{1}{9} \right), \quad R_i = \sigma_r I, \]

\[ H_i(x, y, z) = (x^2, y^2, z^2). \]

- \( Q_i = \varepsilon I, \sigma_b = 1, \sigma_r = 1, \) and \( \varepsilon = 0.0001. \)
Figure: The first component $x(t)$ of the truth and five iterations of LM-EnKS. The initial conditions for the truth are $x(0) = 1$, $y(0) = 1$, and $z(0) = 1$, time step $dt = 0.1$, observations are the full state at each time, ensemble size is 100. And Root mean square error of LM-EnKS iterations over 50 timesteps.
Computational results

Lorenz 63 model

Figure: The first component $x(t)$ of the truth and five iterations of LM-EnKS. The initial conditions for the truth are $x(0) = 1$, $y(0) = 1$, and $z(0) = 1$, time step $dt = 0.1$, observations are the full state at each time, ensemble size is 100. And Root mean square error of LM-EnKS iterations over 50 timesteps.
Computational results
Lorenz 63 model

Figure: The first component $x(t)$ of the truth and five iterations of LM-EnKS. The initial conditions for the truth are $x(0) = 1$, $y(0) = 1$, and $z(0) = 1$, time step $dt = 0.1$, observations are the full state at each time, ensemble size is 100. And Root mean square error of LM-EnKS iterations over 50 time steps.
Computational results

Lorenz 63 model

**Figure**: The first component $x(t)$ of the truth and five iterations of LM-EnKS. The initial conditions for the truth are $x(0) = 1$, $y(0) = 1$, and $z(0) = 1$, time step $dt = 0.1$, observations are the full state at each time, ensemble size is 100. And Root mean square error of LM-EnKS iterations over 50 time steps.
Computational results
Lorenz 63 model

**Figure**: The first component $x(t)$ of the truth and five iterations of LM-EnKS. The initial conditions for the truth are $x(0) = 1$, $y(0) = 1$, and $z(0) = 1$, time step $dt = 0.1$, observations are the full state at each time, ensemble size is 100. And Root mean square error of LM-EnKS iterations over 50 timesteps.
Root mean square error of LM-EnKS iterations over 50 time steps

<table>
<thead>
<tr>
<th>Iteration</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMSE</td>
<td>20.16</td>
<td>15.37</td>
<td>3.73</td>
<td>2.53</td>
<td>0.09</td>
<td>0.09</td>
</tr>
</tbody>
</table>
An example where Gauss-Newton does not converge

\[(x_0 - 2)^2 + (3 + x_1^3)^2 + 10^6(x_0 - x_1)^2 \rightarrow \min\]

Could be seen as 4DVar problem with \(x_b = 2, B = I, M_1 = I, H_1(x) = -x^3, d_1 = 3, Q_1 = 10^{-6}\)
Adaptive gamma is better than fix gamma.
Advantages of LM-EnKS

- Solve the linear least-squares from 4DVar by EnKS, naturally \textit{parallel} over the ensemble members.
- Linear algebra glue is \textit{cheap}.
- Finite differences $\Rightarrow$ no tangent and adjoint operators needed.
- Add Tikhonov regularization to the linear least-squares $\Rightarrow$ Levelberg-Marquardt method, guaranteed convergence.
- Cheap and simple implementation of Tikhonov regularization within EnKS as an additional observation.
Advantages of LM-EnKS

- Solve the linear least-squares from 4DVar by EnKS, naturally parallel over the ensemble members.
- Linear algebra glue is cheap.
- Finite differences ⇒ no tangent and adjoint operators needed.
- Add Tikhonov regularization to the linear least-squares ⇒ Levelberg-Marquardt method, guaranteed convergence.
- Cheap and simple implementation of Tikhonov regularization within EnKS as an additional observation.
Advantages of LM-EnKS

- Solve the linear least-squares from 4DVar by EnKS, naturally parallel over the ensemble members.
- Linear algebra glue is cheap.
- Finite differences $\Rightarrow$ no tangent and adjoint operators needed.
- Add Tikhonov regularization to the linear least-squares $\Rightarrow$ Levelberg-Marquardt method, guaranteed convergence.
- Cheap and simple implementation of Tikhonov regularization within EnKS as an additional observation.
Advantages of LM-EnKS

- Solve the linear least-squares from 4DVar by EnKS, naturally parallel over the ensemble members.
- Linear algebra glue is cheap.
- Finite differences $\Rightarrow$ no tangent and adjoint operators needed.
- Add Tikhonov regularization to the linear least-squares $\Rightarrow$ Levelberg-Marquardt method, guaranteed convergence.
- Cheap and simple implementation of Tikhonov regularization within EnKS as an additional observation.
Advantages of LM-EnKS

- Solve the linear least-squares from 4DVar by EnKS, naturally parallel over the ensemble members.
- Linear algebra glue is cheap.
- Finite differences $\Rightarrow$ no tangent and adjoint operators needed.
- Add Tikhonov regularization to the linear least-squares $\Rightarrow$ Levelberg-Marquardt method, guaranteed convergence.
- Cheap and simple implementation of Tikhonov regularization within EnKS as an additional observation.
Outline

1. Problem statement
2. Globalisation methods
3. A LM-EnKS method
4. Computational results
5. Ongoing work
Variant of LM-EnKS and Convergence theory

- A variant of the method LM-EnKS using the information about the spread between the exact gradient of the 4DVar problem and the gradient of the model.
- In linear case (Mandel et al., 2009), (Le Gland et al., 2011) show that when $N \to \infty$, $\forall 1 \leq p < \infty$ the sample mean and covariance computed by EnKF converge in $L^p$ to the exact mean and covariance,
- We show the same result for the Kalman Smoother,
- When $\tau \to 0$ we prove that the LM-EnKS method is asymptotically equivalent to the method with the derivatives,
- When $\tau \to 0$ and $N \to \infty$, $\forall p, 1 \leq p < \infty$ we prove that at each iteration of LM-EnKS method, the sample mean converges in $L^p$ to the exact solution of the linearized problem.
Variant of LM-EnKS and Convergence theory

- A variant of the method LM-EnKS using the information about the spread between the exact gradient of the 4DVar problem and the gradient of the model.
- In linear case (Mandel et al., 2009), (Le Gland et al., 2011) show that when $N \to \infty$, $\forall 1 \leq p < \infty$ the sample mean and covariance computed by EnKF converge in $L^p$ to the exact mean and covariance,
- We show the same result for the Kalman Smoother,
- When $\tau \to 0$ we prove that the LM-EnKS method is asymptotically equivalent to the method with the derivatives,
- When $\tau \to 0$ and $N \to \infty$, $\forall p, 1 \leq p < \infty$ we prove that at each iteration of LM-EnKS method, the sample mean converges in $L^p$ to the exact solution of the linearized problem.
Variant of LM-EnKS and Convergence theory

- A variant of the method LM-EnKS using the information about the spread between the exact gradient of the 4DVar problem and the gradient of the model.
- In linear case (Mandel et al., 2009), (Le Gland et al., 2011) show that when $N \to \infty$, $\forall 1 \leq p < \infty$ the sample mean and covariance computed by EnKF converge in $L^p$ to the exact mean and covariance.
- We show the same result for the Kalman Smoother,
- When $\tau \to 0$ we prove that the LM-EnKS method is asymptotically equivalent to the method with the derivatives,
- When $\tau \to 0$ and $N \to \infty$, $\forall p, 1 \leq p < \infty$ we prove that at each iteration of LM-EnKS method, the sample mean converges in $L^p$ to the exact solution of the linearized problem.
A variant of the method LM-EnKS using the information about the spread between the exact gradient of the 4DVar problem and the gradient of the model.

In linear case (Mandel et al., 2009), (Le Gland et al., 2011) show that when $N \to \infty$, $\forall 1 \leq p < \infty$ the sample mean and covariance computed by EnKF converge in $L^p$ to the exact mean and covariance,

We show the same result for the Kalman Smoother,

When $\tau \to 0$ we prove that the LM-EnKS method is asymptotically equivalent to the method with the derivatives,

When $\tau \to 0$ and $N \to \infty$, $\forall p, 1 \leq p < \infty$ we prove that at each iteration of LM-EnKS method, the sample mean converges in $L^p$ to the exact solution of the linearized problem.
Variant of LM-EnKS and Convergence theory

- A variant of the method LM-EnKS using the information about the spread between the exact gradient of the 4DVar problem and the gradient of the model.
- In linear case (Mandel et al., 2009), (Le Gland et al., 2011) show that when $N \to \infty$, $\forall 1 \leq p < \infty$ the sample mean and covariance computed by EnKF converge in $L^p$ to the exact mean and covariance,
- We show the same result for the Kalman Smoother,
- When $\tau \to 0$ we prove that the LM-EnKS method is asymptotically equivalent to the method with the derivatives,
- When $\tau \to 0$ and $N \to \infty$, $\forall p, 1 \leq p < \infty$ we prove that at each iteration of LM-EnKS method, the sample mean converges in $L^p$ to the exact solution of the linearized problem.
Variant of LM-EnKS and Convergence theory

- A variant of the method LM-EnKS using the information about the spread between the exact gradient of the 4DVar problem and the gradient of the model.
- In linear case (Mandel et al., 2009), (Le Gland et al., 2011) show that when $N \to \infty$, $\forall 1 \leq p < \infty$ the sample mean and covariance computed by EnKF converge in $L^p$ to the exact mean and covariance,
- We show the same result for the Kalman Smoother,
- When $\tau \to 0$ we prove that the LM-EnKS method is asymptotically equivalent to the method with the derivatives,
- When $\tau \to 0$ and $N \to \infty$, $\forall p, 1 \leq p < \infty$ we prove that at each iteration of LM-EnKS method, the sample mean converges in $L^p$ to the exact solution of the linearized problem.
Thank you for your attention!
Some related work

- The equivalence between weak constraint 4DVar and Kalman smoothing is approximate for nonlinear problems, but still useful (Fisher et al., 2005).
- (Hamill et al. 2000) estimated background covariance from ensemble for 4DVar.
- Gradient methods in the span of the ensemble for one analysis cycle (i.e., 3DVAR) (Sakov et al., 2012) (with square root EnKF as a linear solver in Newton method), and (Bocquet and Sakov, 2012), who added regularization and use LETKF-like approach to minimize the nonlinear cost function over linear combinations of the ensemble.
- (Liu et al. 2008), (Liu et al. 2009) combine ensembles with (strong constraint) 4DVar and minimize in the observation space.
References


