Hybrid 4DVAR and nonlinear EnKS method without tangents and adjoints

E. Bergou, S. Gratton, and J. Mandel

INP-ENSEEIHT, CERFACS Toulouse, and University of Colorado Denver

Probabilistic Approaches to Data Assimilation for Earth Systems Workshop, February 2013

Outline

Outline

Problem statement

Let consider the following stochastic non necessary linear system :

$$\begin{aligned} X_0 &= x_\mathrm{b} && +V_0, && V_0 \sim N\left(0, \boldsymbol{B}\right) \\ X_i &= \mathcal{M}_i(X_{i-1}) && +V_i, && V_i \sim N\left(0, \boldsymbol{Q}_i\right) \\ d_i &= \mathcal{H}_i(X_i) && +W_i, && W_i \sim N\left(0, \boldsymbol{R}_i\right) \end{aligned} \quad \text{where}$$

- X_i is the n dimensional state at time i ; it is random,
- d_i is the random observation vector at time i,
- $\mathcal{M}_i$ is the (nonlinear) model propagator at time i,
- $\mathcal{H}_i$ is the observation operator at time i, it is not linear,
- x_b is background vector,
- $\boldsymbol{B}$ is the background error covariance matrix,
- $\boldsymbol{Q}_i$ and $\boldsymbol{R}_i$ are respectively the model, and observation, error covariance matrices at time i,

Problem statement

$$
\begin{aligned}
X_0 &= x_\mathrm{b} & +V_0, \quad & V_0 \sim N\left(0, \boldsymbol{B}\right) \\
X_i &= \mathcal{M}_i(X_{i-1}) & +V_i, \quad & V_i \sim N\left(0, \boldsymbol{Q}_i\right) \\
d_i &= \mathcal{H}_i(X_i) & +W_i, \quad & W_i \sim N\left(0, \boldsymbol{R}_i\right)
\end{aligned}
$$

- 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.

Problem statement

$$
\begin{aligned}
X_0 &= x_{\mathrm{b}} && +V_0, && V_0 \sim N\left(0, \boldsymbol{B}\right) \\
X_i &= \mathcal{M}_i(X_{i-1}) && +V_i, && V_i \sim N\left(0, \boldsymbol{Q}_i\right) \\
d_i &= \mathcal{H}_i(X_i) && +W_i, && W_i \sim N\left(0, \boldsymbol{R}_i\right)
\end{aligned}
$$

- 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

- We want to determine $x_0, \ldots, x_k$ (x_i = state at time i) from background, model and observations (data)

$$\begin{array}{rcll} x_0 & \approx & x_\mathrm{b} & \text{state at time } 0 \approx \text{ the background,} \\ x_i & \approx & \mathcal{M}_i\left(x_{i-1}\right) & \text{state evolution } \approx \text{ by the model ,} \\ \mathcal{H}_i\left(x_i\right) & \approx & d_i & \text{value of observation operator } \approx \text{ data.} \end{array}$$

- $\Rightarrow$ nonlinear least-squares problem

$$J(x_{0:k}) = \|x_0 - x_\mathrm{b}\|^2_{\boldsymbol{B}^{-1}} + \sum_{i=1}^{k} \|x_i - \mathcal{M}_i\left(x_{i-1}\right)\|^2_{\boldsymbol{Q}_i^{-1}}$$

$$+ \sum_{i=1}^{k} \|d_i - \mathcal{H}_i\left(x_i\right)\|^2_{\boldsymbol{R}_i^{-1}} \to \min_{x_{0:k}}$$

Weak constraint 4DVar

- We want to determine $x_0, \ldots, x_k$ (x_i = state at time i) from background, model and observations (data)

$$\begin{array}{rcll} x_0 & \approx & x_\mathrm{b} & \text{state at time } 0 \approx \text{ the background,} \\ x_i & \approx & \mathcal{M}_i\left(x_{i-1}\right) & \text{state evolution } \approx \text{ by the model ,} \\ \mathcal{H}_i\left(x_i\right) & \approx & d_i & \text{value of observation operator } \approx \text{ data.} \end{array}$$

- $\Rightarrow$ nonlinear least-squares problem

$$J(x_{0:k}) = \|x_0 - x_\mathrm{b}\|^2_{\boldsymbol{B}^{-1}} + \sum_{i=1}^{k} \|x_i - \mathcal{M}_i\left(x_{i-1}\right)\|^2_{\boldsymbol{Q}_i^{-1}}$$

$$+ \sum_{i=1}^{k} \|d_i - \mathcal{H}_i\left(x_i\right)\|^2_{\boldsymbol{R}_i^{-1}} \rightarrow \min_{x_{0:k}}$$

Weak constraint 4DVar

- Originally in 4DVar (strong constraint), $x_i = \mathcal{M}_i(x_{i-1})$(perfect model). The weak constraint $x_i \approx \mathcal{M}_i(x_{i-1})$ accounts for model error (Trémolet, 2007).
- 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

- Originally in 4DVar (strong constraint), $x_i = \mathcal{M}_i\left(x_{i-1}\right)$(perfect model). The weak constraint $x_i \approx \mathcal{M}_i\left(x_{i-1}\right)$ accounts for model error (Trémolet, 2007).
- 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

- Originally in 4DVar (strong constraint), $x_i = \mathcal{M}_i\,(x_{i-1})$(perfect model). The weak constraint $x_i \approx \mathcal{M}_i\,(x_{i-1})$ accounts for model error (Trémolet, 2007).
- 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

- Originally in 4DVar (strong constraint), $x_i = \mathcal{M}_i\left(x_{i-1}\right)$(perfect model). The weak constraint $x_i \approx \mathcal{M}_i\left(x_{i-1}\right)$ accounts for model error (Trémolet, 2007).
- 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

$$\mathcal{M}_i\left(x_{i-1}+\delta x_{i-1}\right) \approx \mathcal{M}_i\left(x_{i-1}\right)+\mathcal{M}_i'\left(x_{i-1}\right)\delta x_{i-1},$$
$$\mathcal{H}_i\left(x_i+\delta x_i\right) \approx \mathcal{H}_i\left(x_i\right)+\mathcal{H}_i'\left(x_i\right)\delta x_i,$$

- 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

$$\left\|x_0+\delta x_0-x_{\mathrm{b}}\right\|^2_{\boldsymbol{B}^{-1}}+\sum_{i=1}^{k}\left\|d_i-\mathcal{H}_i\left(x_i\right)-\mathcal{H}_i'\left(x_i\right)\delta x_i\right\|^2_{\boldsymbol{R}_i^{-1}}$$
$$+\sum_{i=1}^{k}\left\|x_i+\delta x_i-\mathcal{M}_i\left(x_{i-1}\right)-\mathcal{M}_i'\left(x_{i-1}\right)\delta x_{i-1}\right\|^2_{\boldsymbol{Q}_i^{-1}}$$

Incremental 4DVar

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

$$\mathcal{M}_i\left(x_{i-1}+\delta x_{i-1}\right) \approx \mathcal{M}_i\left(x_{i-1}\right)+\mathcal{M}_i'\left(x_{i-1}\right)\delta x_{i-1},$$
$$\mathcal{H}_i\left(x_i+\delta x_i\right) \approx \mathcal{H}_i\left(x_i\right)+\mathcal{H}_i'\left(x_i\right)\delta x_i,$$

- 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

$$\left\|x_0+\delta x_0-x_{\mathrm{b}}\right\|_{\boldsymbol{B}^{-1}}^2+\sum_{i=1}^k\left\|d_i-\mathcal{H}_i\left(x_i\right)-\mathcal{H}_i'\left(x_i\right)\delta x_i\right\|_{\boldsymbol{R}_i^{-1}}^2$$
$$+\sum_{i=1}^k\left\|x_i+\delta x_i-\mathcal{M}_i\left(x_{i-1}\right)-\mathcal{M}_i'\left(x_{i-1}\right)\delta x_{i-1}\right\|_{\boldsymbol{Q}_i^{-1}}^2$$

Incremental 4DVar

$$\left\|x_0 + \delta x_0 - x_{\mathrm{b}}\right\|^2_{\boldsymbol{B}^{-1}} + \sum_{i=1}^{k} \left\|d_i - \mathcal{H}_i\left(x_i\right) - \mathcal{H}'_i\left(x_i\right)\delta x_i\right\|^2_{\boldsymbol{R}_i^{-1}}$$

$$+\sum_{i=1}^{k} \left\|x_i + \delta x_i - \mathcal{M}_i\left(x_{i-1}\right) - \mathcal{M}'_i\left(x_{i-1}\right)\delta x_{i-1}\right\|^2_{\boldsymbol{Q}_i^{-1}}$$

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

Incremental 4DVar

$$\left\|x_0 + \delta x_0 - x_{\mathrm{b}}\right\|^2_{\boldsymbol{B}^{-1}} + \sum_{i=1}^{k} \left\|d_i - \mathcal{H}_i\left(x_i\right) - \mathcal{H}_i'\left(x_i\right)\delta x_i\right\|^2_{\boldsymbol{R}_i^{-1}}$$

$$+\sum_{i=1}^{k} \left\|x_i + \delta x_i - \mathcal{M}_i\left(x_{i-1}\right) - \mathcal{M}_i'\left(x_{i-1}\right)\delta x_{i-1}\right\|^2_{\boldsymbol{Q}_i^{-1}}$$

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

Incremental 4DVar

$$\left\|x_0 + \delta x_0 - x_{\mathrm{b}}\right\|^2_{\boldsymbol{B}^{-1}} + \sum_{i=1}^{k} \left\|d_i - \mathcal{H}_i\left(x_i\right) - \mathcal{H}_i'\left(x_i\right)\delta x_i\right\|^2_{\boldsymbol{R}_i^{-1}}$$

$$+\sum_{i=1}^{k} \left\|x_i + \delta x_i - \mathcal{M}_i\left(x_{i-1}\right) - \mathcal{M}_i'\left(x_{i-1}\right)\delta x_{i-1}\right\|^2_{\boldsymbol{Q}_i^{-1}}$$

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

Outline

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 :

$$\arg\min_{x \in \mathcal{R}^n} F(x) = \|f(x)\|^2,$$

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

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

$$FL(x_j + \delta x) = \|f(x_j) + J_f(x_j)\delta x\|^2 + \gamma\|\delta x\|^2 \to \min_{\delta x},$$

where x_j is the j-th iterate, $J_f(x_j)$ is the Jacobian of f at x_j and γ 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 :

$$\arg\min_{x \in \mathcal{R}^n} F(x) = \|f(x)\|^2,$$

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

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

$$FL(x_j + \delta x) = \|f(x_j) + J_f(x_j)\delta x\|^2 + \gamma\|\delta x\|^2 \to \min_{\delta x},$$

where x_j is the j-th iterate, $J_f(x_j)$ is the Jacobian of f at x_j and γ 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 :

$$\arg\min_{x \in \mathcal{R}^n} F(x) = \|f(x)\|^2,$$

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

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

$$FL(x_j + \delta x) = \|f(x_j) + J_f(x_j)\delta x\|^2 + \gamma\|\delta x\|^2 \to \min_{\delta x},$$

where x_j is the j-th iterate, $J_f(x_j)$ is the Jacobian of f at x_j and γ is the regularization parameter,

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, $\Rightarrow$ 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, $\Rightarrow$ 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, $\Rightarrow$ 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, $\Rightarrow$ The term $\gamma \|\delta x\|^2$ controls the step size as well as rotates the step direction towards the steepest descent.

- In LM method γ may remain constant over the iterations, or adaptive :
  1. γ remains constant, it must be chosen large enough to ensure the convergence.
  2. γ 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 γ,
  - Else, we increase γ,

- In LM method γ may remain constant over the iterations, or adaptive :
  1. γ remains constant, it must be chosen large enough to ensure the convergence.
  2. γ 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 γ,
- Else, we increase γ,

- In LM method γ may remain constant over the iterations, or adaptive :
  1. γ remains constant, it must be chosen large enough to ensure the convergence.
  2. γ 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 γ,
- Else, we increase γ,

- In LM method γ may remain constant over the iterations, or adaptive :
  1. γ remains constant, it must be chosen large enough to ensure the convergence.
  2. γ 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 γ,
  - Else, we increase γ,

- In LM method γ may remain constant over the iterations, or adaptive :
  1. γ remains constant, it must be chosen large enough to ensure the convergence.
  2. γ 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 γ,
  - Else, we increase γ,

Levenberg-Marquart Method

- In 4DVar linearized problem we add regularization as follows,

$$\tilde{J}(x_{0:k}) = \|x_0 + \delta x_0 - x_\mathrm{b}\|^2_{\boldsymbol{B}^{-1}} + \sum_{i=1}^{k} \left\|x_i + \delta x_i - \mathcal{M}_i\left(x_{i-1}\right) - \mathcal{M}_i'\left(x_{i-1}\right)\delta x_{i-1}\right\|^2_{\boldsymbol{Q}_i^{-1}} + \sum_{i=1}^{k} \left\|d_i - \mathcal{H}_i\left(x_i\right) - \mathcal{H}_i'\left(x_i\right)\delta x_i\right\|^2_{\boldsymbol{R}_i^{-1}} + \gamma \sum_{i=0}^{k} \|\delta x_i\|^2_{\boldsymbol{S}_i^{-1}}$$

- Assume the model and observation operator are regular and that γ 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,

$$\tilde{J}(x_{0:k}) = \|x_0 + \delta x_0 - x_\mathrm{b}\|^2_{\boldsymbol{B}^{-1}} + \sum_{i=1}^{k} \left\|x_i + \delta x_i - \mathcal{M}_i\left(x_{i-1}\right) - \mathcal{M}_i'\left(x_{i-1}\right)\delta x_{i-1}\right\|^2_{\boldsymbol{Q}_i^{-1}} + \sum_{i=1}^{k} \left\|d_i - \mathcal{H}_i\left(x_i\right) - \mathcal{H}_i'\left(x_i\right)\delta x_i\right\|^2_{\boldsymbol{R}_i^{-1}} + \gamma \sum_{i=0}^{k} \|\delta x_i\|^2_{\boldsymbol{S}_i^{-1}}$$

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

Linearized 4DVar as Kalman smoother

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

$$\|z_0 - z_\mathrm{b}\|^2_{\boldsymbol{B}^{-1}} + \sum_{i=1}^{k} \|z_i - \boldsymbol{M}_i z_{i-1} - m_i\|^2_{\boldsymbol{Q}_i^{-1}} + \sum_{i=1}^{k} \|d_i - \boldsymbol{H}_i z_i\|^2_{\boldsymbol{R}_i^{-1}}$$

$$z_\mathrm{b} = x_\mathrm{b} - x_0, \quad m_i = \mathcal{M}_i\left(x_{i-1}\right) - x_i, \quad d_i = d_i - \mathcal{H}_i\left(x_i\right),$$

$$\boldsymbol{M}_i = \mathcal{M}_i'\left(x_{i-1}\right), \quad \boldsymbol{H}_i = \mathcal{H}_i'\left(x_i\right)$$

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

$$\begin{aligned} Z_0 &= z_\mathrm{b} & &+V_0, & V_0 &\sim N(0, \boldsymbol{B}) \\ Z_i &= \boldsymbol{M}_i Z_{i-1} + m_i & &+V_i, & V_i &\sim N(0, \boldsymbol{Q}_i) \\ d_i &= \boldsymbol{H}_i Z_i & &+W_i, & W_i &\sim N(0, \boldsymbol{R}_i) \end{aligned}$$

Linearized 4DVar as Kalman smoother

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

$$\|z_0 - z_\mathrm{b}\|^2_{\boldsymbol{B}^{-1}} + \sum_{i=1}^{k} \|z_i - \boldsymbol{M}_i z_{i-1} - m_i\|^2_{\boldsymbol{Q}_i^{-1}} + \sum_{i=1}^{k} \|d_i - \boldsymbol{H}_i z_i\|^2_{\boldsymbol{R}_i^{-1}}$$

$$z_\mathrm{b} = x_\mathrm{b} - x_0, \quad m_i = \mathcal{M}_i\left(x_{i-1}\right) - x_i, \quad d_i = d_i - \mathcal{H}_i\left(x_i\right),$$
$$\boldsymbol{M}_i = \mathcal{M}_i'\left(x_{i-1}\right), \quad \boldsymbol{H}_i = \mathcal{H}_i'\left(x_i\right)$$

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

$$\begin{array}{lll} Z_0 = z_\mathrm{b} & +V_0, & V_0 \sim N\left(0, \boldsymbol{B}\right) \\ Z_i = \boldsymbol{M}_i Z_{i-1} + m_i & +V_i, & V_i \sim N\left(0, \boldsymbol{Q}_i\right) \\ d_i = \boldsymbol{H}_i Z_i & +W_i, & W_i \sim N\left(0, \boldsymbol{R}_i\right) \end{array}$$

Ensemble Kalman filter (EnKF) and smoother (EnKS)

$Z_{i|k}^N = [z_{i|k}^1, \ldots, z_{i|k}^N]$ is ensemble of states at time i, conditioned on all data up to time k.

Algorithm (EnKF)

1. *Initialize*

$$z_{0|0}^\ell \sim N\left(z_\mathrm{b}, \boldsymbol{B}\right), \quad \ell = 1, \ldots, N. \tag{1}$$

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

$$z_{i|i-1}^\ell = \boldsymbol{M}_i z_{i-1|i-1}^\ell + m_i + v_i^\ell, \quad v_i^\ell \sim N\left(0, \boldsymbol{Q}_i\right), \tag{2}$$

$$\begin{aligned} z_{i|i}^\ell =& z_{i|i-1}^\ell - \boldsymbol{P}_i^N \boldsymbol{H}_i^\mathrm{T} (\boldsymbol{H}_i \boldsymbol{P}_i^N \boldsymbol{H}_i^\mathrm{T} + \boldsymbol{R}_i)^{-1} \\ & \cdot (\boldsymbol{H}_i z_{i|i-1}^\ell - d_i - w_i^\ell), \quad w_i^\ell \sim N\left(0, \boldsymbol{R}_i\right), \end{aligned} \tag{3}$$

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 $i-1$,

$$Z^N_{0:i|i-1} = \begin{bmatrix} Z^N_{0|i-1} \\ \vdots \\ Z^N_{i|i-1} \end{bmatrix}.$$

in the place of $Z_{i|i-1}$.

- The observation term $\boldsymbol{H}_i Z^N_{i|i-1} - d_i$ becomes

$$[0, \ldots, \boldsymbol{H}_i]\, Z^N_{0:i|i-1} - d_i = \boldsymbol{H}_i Z^N_{i|i-1} - d_i. \tag{5}$$

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 $i-1$,

$$Z^N_{0:i|i-1} = \begin{bmatrix} Z^N_{0|i-1} \\ \vdots \\ Z^N_{i|i-1} \end{bmatrix}.$$

  in the place of $Z_{i|i-1}$.
- The observation term $\boldsymbol{H}_i Z^N_{i|i-1} - d_i$ becomes

$$[0, \ldots, \boldsymbol{H}_i]\, Z^N_{0:i|i-1} - d_i = \boldsymbol{H}_i Z^N_{i|i-1} - d_i. \quad (5)$$

Ensemble Kalman filter (EnKF) and smoother (EnKS)

Algorithm (EnKS)

Given z_{b},

1. *Initialize* $z^{\ell}_{0|0} \sim N\left(z_{\mathrm{b}}, \boldsymbol{B}\right), \quad \ell = 1, \ldots, N.$
2. *For* $i = 1, \ldots, k$, *advance in time,*

$$z^{\ell}_{i|i-1} = \boldsymbol{M}_i z^{\ell}_{i-1|i-1} + m_i + v^{\ell}_i, \quad v^{\ell}_i \sim N\left(0, \boldsymbol{Q}_i\right), \quad (6)$$

$$Z^N_{0:i|i} = Z^N_{0:i|i-1} - \boldsymbol{P}^N_{0:i,0:i-1} \widetilde{\boldsymbol{H}}^{\mathrm{T}}_{0:i} (\widetilde{\boldsymbol{H}}_{0:i} \boldsymbol{P}_{0:i,0:i-1} \widetilde{\boldsymbol{H}}^{\mathrm{T}}_{0:i} + \boldsymbol{R}_i)^{-1} \quad (7)$$

$$\cdot\, (\widetilde{\boldsymbol{H}}_{0:i} Z^N_{i|i-1} - d_i - w_i), \quad w_i \sim N\left(0, \boldsymbol{R}_i\right), \quad (8)$$

where $\widetilde{\boldsymbol{H}}_{0:i} = [0, \ldots, \boldsymbol{H}_i]$, *and* $\boldsymbol{P}^N_{0:i,0:i-1}$ *is the sample covariance matrix of* $Z^N_{0:i|i-1}$.

Derivative-free implementation of the EnKS - model

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

$$\boldsymbol{M}_i \delta x_{i-1}^n + m_i = \mathcal{M}_i'(x_{i-1})\,\delta x_{i-1}^n + \mathcal{M}_i(x_{i-1}) - x_i,$$

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

$$\boldsymbol{M}_i \delta x_{i-1}^n + m_i \approx \frac{\mathcal{M}_i(x_{i-1} + \tau \delta x_{i-1}^n) - \mathcal{M}_i(x_{i-1})}{\tau} + \mathcal{M}_i(x_{i-1}) - x_i,$$

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

Derivative-free implementation of the EnKS - model

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

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

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

$$\boldsymbol{M}_i \delta x^n_{i-1} + m_i \approx \frac{\mathcal{M}_i\left(x_{i-1} + \tau \delta x^n_{i-1}\right) - \mathcal{M}_i(x_{i-1})}{\tau} + \mathcal{M}_i(x_{i-1}) - x_i,$$

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

Derivative-free implementation of the EnKS - model

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

$$\boldsymbol{M}_i \delta x_{i-1}^n + m_i = \mathcal{M}_i'(x_{i-1})\, \delta x_{i-1}^n + \mathcal{M}_i(x_{i-1}) - x_i,$$

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

$$\boldsymbol{M}_i \delta x_{i-1}^n + m_i \approx \frac{\mathcal{M}_i(x_{i-1} + \tau \delta x_{i-1}^n) - \mathcal{M}_i(x_{i-1})}{\tau} + \mathcal{M}_i(x_{i-1}) - x_i,$$

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

Derivative-free implementation of the EnKS - model

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

$$\boldsymbol{M}_i \delta x_{i-1}^n + m_i = \mathcal{M}_i'(x_{i-1})\, \delta x_{i-1}^n + \mathcal{M}_i(x_{i-1}) - x_i,$$

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

$$\boldsymbol{M}_i \delta x_{i-1}^n + m_i \approx \frac{\mathcal{M}_i\left(x_{i-1} + \tau \delta x_{i-1}^n\right) - \mathcal{M}_i(x_{i-1})}{\tau} + \mathcal{M}_i(x_{i-1}) - x_i,$$

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

Derivative-free implementation of the EnKS - model

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

$$\boldsymbol{M}_i \delta x_{i-1}^n + m_i = \mathcal{M}_i'(x_{i-1})\, \delta x_{i-1}^n + \mathcal{M}_i(x_{i-1}) - x_i,$$

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

$$\boldsymbol{M}_i \delta x_{i-1}^n + m_i \approx \frac{\mathcal{M}_i\left(x_{i-1} + \tau \delta x_{i-1}^n\right) - \mathcal{M}_i(x_{i-1})}{\tau} + \mathcal{M}_i(x_{i-1}) - x_i,$$

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

Derivative-free implementation of the EnKS - model

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

$$\boldsymbol{M}_i \delta x_{i-1}^n + m_i = \mathcal{M}_i'(x_{i-1})\, \delta x_{i-1}^n + \mathcal{M}_i(x_{i-1}) - x_i,$$

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

$$\boldsymbol{M}_i \delta x_{i-1}^n + m_i \approx \frac{\mathcal{M}_i\left(x_{i-1} + \tau \delta x_{i-1}^n\right) - \mathcal{M}_i(x_{i-1})}{\tau} + \mathcal{M}_i(x_{i-1}) - x_i,$$

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

Derivative-free implementation of the EnKS - model

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

$$\boldsymbol{M}_i \delta x_{i-1}^n + m_i = \mathcal{M}_i'(x_{i-1})\,\delta x_{i-1}^n + \mathcal{M}_i(x_{i-1}) - x_i,$$

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

$$\boldsymbol{M}_i \delta x_{i-1}^n + m_i \approx \frac{\mathcal{M}_i\left(x_{i-1} + \tau \delta x_{i-1}^n\right) - \mathcal{M}_i(x_{i-1})}{\tau} + \mathcal{M}_i(x_{i-1}) - x_i,$$

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

Derivative-free implementation of the EnKS - observation

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

$$\boldsymbol{H}_i Z^N = \left[\boldsymbol{H}_i \delta x^1, \ldots, \boldsymbol{H}_i \delta x^N\right].$$

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

$$\boldsymbol{H}_i \delta x_i^n \approx \frac{\mathcal{H}_i\left(x_{i-1} + \tau \delta x_{i-1}^n\right) - \mathcal{H}_i\left(x_{i-1}\right)}{\tau},$$

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

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

Derivative-free implementation of the EnKS - observation

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

$$\boldsymbol{H}_i Z^N = \left[\boldsymbol{H}_i \delta x^1, \ldots, \boldsymbol{H}_i \delta x^N\right].$$

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

$$\boldsymbol{H}_i \delta x_i^n \approx \frac{\mathcal{H}_i\left(x_{i-1} + \tau \delta x_{i-1}^n\right) - \mathcal{H}_i\left(x_{i-1}\right)}{\tau},$$

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

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

Derivative-free implementation of the EnKS - observation

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

$$\boldsymbol{H}_i Z^N = \left[\boldsymbol{H}_i \delta x^1, \dots, \boldsymbol{H}_i \delta x^N\right].$$

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

$$\boldsymbol{H}_i \delta x_i^n \approx \frac{\mathcal{H}_i\left(x_{i-1} + \tau \delta x_{i-1}^n\right) - \mathcal{H}_i\left(x_{i-1}\right)}{\tau},$$

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

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

Derivative-free implementation of the EnKS - observation

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

$$\boldsymbol{H}_i Z^N = \left[\boldsymbol{H}_i \delta x^1, \ldots, \boldsymbol{H}_i \delta x^N\right].$$

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

$$\boldsymbol{H}_i \delta x_i^n \approx \frac{\mathcal{H}_i\left(x_{i-1} + \tau \delta x_{i-1}^n\right) - \mathcal{H}_i\left(x_{i-1}\right)}{\tau},$$

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

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

Derivative-free implementation of the EnKS - observation

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

$$\boldsymbol{H}_i Z^N = \left[\boldsymbol{H}_i \delta x^1, \ldots, \boldsymbol{H}_i \delta x^N\right].$$

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

$$\boldsymbol{H}_i \delta x_i^n \approx \frac{\mathcal{H}_i\left(x_{i-1} + \tau \delta x_{i-1}^n\right) - \mathcal{H}_i\left(x_{i-1}\right)}{\tau},$$

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

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

Outline

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

A LM-EnKS method

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

For outer loop = 1,2,...

- Initialize $z^{\ell}_{0|0} \sim N(0, \boldsymbol{B})$, for $\ell = 1, \ldots, N$
- for $i = 1, \ldots, k$ advance z^{ℓ} in time following (2), with the linearized operator approximated by finite differences :

$$z^{\ell}_{i|i-1} = \frac{\mathcal{M}_i\left(x_{i-1} + \tau z^{\ell}_{i-1|i-1}\right) - \mathcal{M}_i\left(x_{i-1}\right)}{\tau} + \mathcal{M}_i\left(x_{i-1}\right) - x_i + v^{\ell}_i, \quad v^{\ell}_i \sim N\left(0, \boldsymbol{Q}_i\right)$$

followed by the smoother analysis step with matrix-vector products $\boldsymbol{H}_i z_i$ approximated by finite differences.

A LM-EnKS method

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

For outer loop = 1,2,...

- Initialize $z_{0|0}^{\ell} \sim N(0, \boldsymbol{B})$, for $\ell = 1, \ldots, N$
- for $i = 1, \ldots, k$ advance z^{ℓ} in time following (2), with the linearized operator approximated by finite differences :

$$z_{i|i-1}^{\ell} = \frac{\mathcal{M}_i\left(x_{i-1} + \tau z_{i-1|i-1}^{\ell}\right) - \mathcal{M}_i\left(x_{i-1}\right)}{\tau} + \mathcal{M}_i\left(x_{i-1}\right) - x_i + v_i^{\ell}, \quad v_i^{\ell} \sim N\left(0, \boldsymbol{Q}_i\right)$$

followed by the smoother analysis step with matrix-vector products $\boldsymbol{H}_i z_i$ approximated by finite differences.

- Tikhonov regularization is considered as a further observations

$$\tilde{d}_i = 0 = z_i + \tilde{W}_i \quad \tilde{W}_i \sim N\left(0, \frac{1}{\gamma}\boldsymbol{S}_i\right),$$

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

- $x_i \leftarrow x_i + \frac{1}{N}\sum_{\ell=1}^{N} z_{i|k}^{\ell}$, $i = 1, \ldots, k$

- Tikhonov regularization is considered as a further observations

$$\tilde{d}_i = 0 = z_i + \tilde{W}_i \quad \tilde{W}_i \sim N\left(0, \frac{1}{\gamma}\boldsymbol{S}_i\right),$$

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

- $x_i \leftarrow x_i + \frac{1}{N}\sum_{\ell=1}^{N} z_{i|k}^{\ell}$, $i = 1, \ldots, k$

Towards a convergence theory

- 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.
- When $\tau \to 0$ and $N \to \infty$, we prove that the gradient of the iterates goes to 0 for any initial iterate.

Towards a convergence theory

- 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.
- When $\tau \to 0$ and $N \to \infty$, we prove that the gradient of the iterates goes to 0 for any initial iterate.

Towards a convergence theory

- 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.
- When $\tau \to 0$ and $N \to \infty$, we prove that the gradient of the iterates goes to 0 for any initial iterate.

Towards a convergence theory

- 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.
- When $\tau \to 0$ and $N \to \infty$, we prove that the gradient of the iterates goes to 0 for any initial iterate.

Towards a convergence theory

- 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.
- When $\tau \to 0$ and $N \to \infty$, we prove that the gradient of the iterates goes to 0 for any initial iterate.

Outline

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

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 $\mathcal{H}_i$ 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 $\mathcal{H}_i$ to the truth and then adding a gaussian perturbation,
- Try to recover the truth using LM-EnKS method.

Computational results

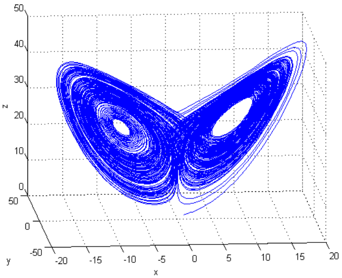
- To evaluate the performance of the method, we use the twin experiment technique.
- Truth = an integration of the model over time,
- We obtain the background by adding a gaussian perturbation to the initial state,
- We obtain the data d_i by applying the observation operator $\mathcal{H}_i$ to the truth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

Computational results

Lorenz 63 model



$$\frac{dx}{dt} = -\sigma(x - y)$$
$$\frac{dy}{dt} = \rho x - y - xz$$
$$\frac{dz}{dt} = xy - \beta z$$

σ, ρ and β 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.
-
$$\boldsymbol{B} = \sigma_b^2 \operatorname{diag}\left(1, \frac{1}{4}, \frac{1}{9}\right), \quad R_i = \sigma_r^2 \boldsymbol{I},$$
$$\mathcal{H}_i\left(x, y, z\right) = \left(x^2, y^2, z^2\right).$$
- $\boldsymbol{Q}_i = \varepsilon \boldsymbol{I}$, $\sigma_b = 1, \sigma_r = 1$, and $\epsilon = 0.0001$.

Computational results

Lorenz 63 model

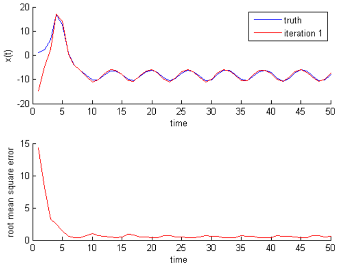


Figure : The first component $x(t)$ of the truth and five iterations of LM-EnKS. The initial conditions for the truth are $x(0) = 1$, $y(0) = 1$, and $z(0) = 1$, time step $dt = 0.1$, observations are the full state at each time, ensemble size is 100. And Root mean square error of LM-EnKS iterations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

Computational results

Lorenz 63 model

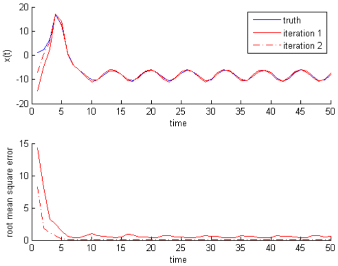


Figure : The first component $x(t)$ of the truth and five iterations of LM-EnKS. The initial conditions for the truth are $x(0) = 1$, $y(0) = 1$, and $z(0) = 1$, time step $dt = 0.1$, observations are the full state at each time, ensemble size is 100. And Root mean square error of LM-EnKS iterations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

Computational results
Lorenz 63 model

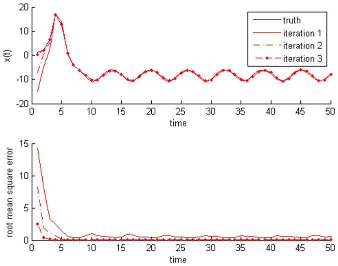


Figure : The first component $x(t)$ of the truth and five iterations of LM-EnKS. The initial conditions for the truth are $x(0) = 1$, $y(0) = 1$, and $z(0) = 1$, time step $dt = 0.1$, observations are the full state at each time, ensemble size is 100. And Root mean square error of LM-EnKS iterations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

Computational results

Lorenz 63 model

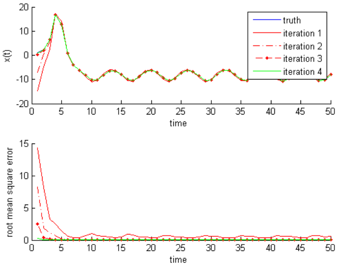


Figure : The first component $x(t)$ of the truth and five iterations of LM-EnKS. The initial conditions for the truth are $x(0) = 1$, $y(0) = 1$, and $z(0) = 1$, time step $dt = 0.1$, observations are the full state at each time, ensemble size is 100. And Root mean square error of LM-EnKS iterations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

Computational results

Lorenz 63 model

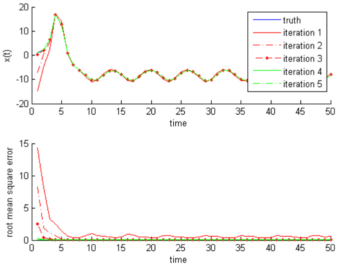
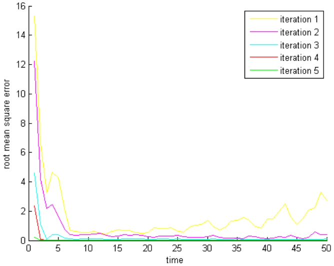


Figure : The first component $x(t)$ of the truth and five iterations of LM-EnKS. The initial conditions for the truth are $x(0) = 1$, $y(0) = 1$, and $z(0) = 1$, time step $dt = 0.1$, observations are the full state at each time, ensemble size is 100. And Root mean square error of LM-EnKS iterations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

LM-EnKS for Lorenz 63 model



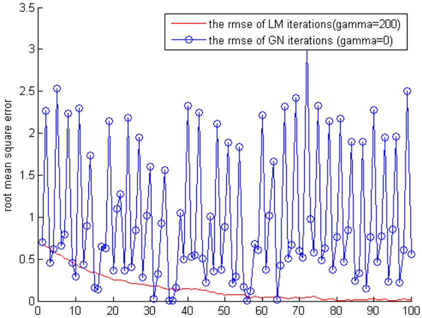
Root mean square error of LM-EnKS iterations over 50 time steps

Iteration	1	2	3	4	5	6
RMSE	20.16	15.37	3.73	2.53	0.09	0.09

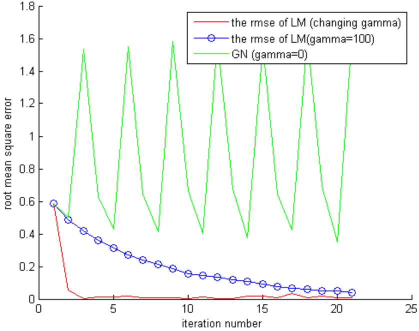
An example where Gauss-Newton does not converge

$$(x_0 - 2)^2 + (3 + x_1^3)^2 + 10^6 (x_0 - x_1)^2 \to \min$$

Could be seen as 4DVar problem with $x_\mathrm{b} = 2$, $\boldsymbol{B} = \boldsymbol{I}$, $M_1 = I$, $\mathcal{H}_1(x) = -x^3$, $d_1 = 3$, $\boldsymbol{Q}_1 = 10^{-6}$



- Adaptive gamma is better than fix gamma.



Outline

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.

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.

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 backgroud 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

Bell, B., 1994: The iterated Kalman smoother as a Gauss-Newton method. *SIAM Journal on Optimization*, **4 (3)**, 626–636, doi:10.1137/0804035.

Bocquet, M. and P. Sakov, 2012: Combining inflation-free and iterative ensemble kalman filters for strongly nonlinear systems. *Nonlinear Processes in Geophysics*, **19 (3)**, 383–399, doi:10.5194/npg-19-383-2012.

Courtier, P., J.-N. Thépaut, and A. Hollingsworth, 1994: A strategy for operational implementation of 4d-var, using an incremental approach. *Quarterly Journal of the Royal Meteorological Society*, **120 (519)**, 1367–1387, doi:10.1002/qj.49712051912.

Evensen, G., 2009: *Data Assimilation: The Ensemble Kalman Filter*. 2d ed., Springer, xxiv+307 pp., doi:10.1007/978-3-642-03711-5.

Fisher, M., M. Leutbecher, and G. A. Kelly, 2005: On the equivalence between Kalman smoothing and weak-constraint four-dimensional variational data assimilation. *Quarterly Journal of the Royal Meteorological Society*, **131 (613, Part c)**, 3235–3246, doi:{10.1256/qj.04.142}.

Le Gland, F., V. Monbet, and V.-D. Tran, 2011: Large sample asymptotics for the ensemble Kalman filter. *The Oxford Handbook of Nonlinear Filtering*, D. Crisan and B. Rozovskii, Eds., Oxford University Press, INRIA Report 7014, August 2009.

Mandel, J., L. Cobb, and J. D. Beezley, 2009: On the convergence of the ensemble Kalman filter. arXiv:0901.2951.

Rauch, H. E., F. Tung, and C. T. Striebel, 1965: Maximum likelihood estimates of linear dynamic systems. *AIAA Journal*, **3 (8)**, 1445–1450.

Sakov, P., D. S. Oliver, and L. Bertino, 2012: An iterative EnKF for strongly nonlinear systems. *Monthly Weather Review*, **140 (6)**, 1988–2004, doi:10.1175/MWR-D-11-00176.1.

Trémolet, Y., 2007: Model-error estimation in 4D-Var. *Quarterly Journal of the Royal Meteorological Society*, **133 (626)**, 1267–1280, doi:10.1002/qj.94.

Tshimanga, J., S. Gratton, A. T. Weaver, and A. Sartenaer, 2008: Limited-memory preconditioners, with application to incremental four-dimensional variational data assimilation. *Quarterly Journal of the Royal Meteorological Society*, **134 (632)**, 751–769, doi:10.1002/qj.228.