



“Ensemblized” linear least squares (LLS)

Patrick N. Raanes



ISDA-Online
Jan. 20, 2023

LLS



EVERWHERE

LLS

Iter. ens
uses LLS





Iter. ens
uses LLS

EnKF
uses LLS



Iter. ens
uses LLS

EnKF
uses LLS

EAKF, ETKF
uses LLS



Iter. ens
uses LLS

EnKF
uses LLS

EAKF, ETKF
uses LLS

EnKF
is LLS



Iter. ens
uses LLS

EnKF
uses LLS

EAKF, ETKF
uses LLS

EnKF

is LLS

LLS ⇒

avrg. grad



Covariances and LLS regression

Let $f(\mathbf{x}) = \mathbf{y}$.

Let \mathbf{X} and \mathbf{Y} be centered ensemble matrices: n -th column is $\mathbf{x}_n - \bar{\mathbf{x}}$ and $\mathbf{y}_n - \bar{\mathbf{y}}$.

Recall covariance estimate, $\hat{\mathbf{C}}_{y,x} = \frac{1}{N-1} \sum_n (\mathbf{y}_n - \bar{\mathbf{y}})(\mathbf{x}_n - \bar{\mathbf{x}})^T$ (1)

$$= \frac{1}{N-1} \mathbf{Y} \mathbf{X}^T \quad (2)$$

Define the ensemble LLS operator $\hat{\nabla}f$ as

$$\hat{\nabla}f = \hat{\mathbf{C}}_{y,y}^{-1} \hat{\mathbf{C}}_{y,x}^T \quad (3)$$

$$= \mathbf{Y} \mathbf{X}^T (\mathbf{X} \mathbf{X}^T)^{-1} \quad (4)$$

ML if $N \leq 4$, reproduces increments

$$[\mathbf{y}_n - \bar{\mathbf{y}}] = (\hat{\nabla}f)^T [\mathbf{x}_n - \bar{\mathbf{x}}]$$

min-norm LLS, BLUE, MVUE

Relates to: finite differences (FD),
simplex derivatives



Covariances and LLS regression

Let $f(\mathbf{x}) = \mathbf{y}$.

Let \mathbf{X} and \mathbf{Y} be *centred* ensemble matrices

Recall covariance estimate, $\hat{\mathbf{C}}_{y,x} = \frac{1}{N-1} \sum_n (y_n - \bar{y})(x_n - \bar{x})^\top$ (1)

$$= \frac{1}{N-1} \mathbf{Y} \mathbf{X}^\top \quad (2)$$

Define the ensemble LLS operator $\hat{\nabla}f$ as

$$\hat{\nabla}f = \hat{\mathbf{C}}_{y,y}^{-1} \hat{\mathbf{C}}_{y,x}^\top \quad (3)$$

$$= \mathbf{Y} \mathbf{X}^\top (\mathbf{X} \mathbf{X}^\top)^{-1} \quad (4)$$

ML

If $N \leq 4$, reproduces increments

$$[y_n - \bar{y}] = (\hat{\nabla}f)^\top [x_n - \bar{x}]$$

min-norm LLS, BLUE, MVUE

Relates to: finite differences (FD),
simplex derivatives



Covariances and LLS regression

Let $f(\mathbf{x}) = \mathbf{y}$.

Let \mathbf{X} and \mathbf{Y} be *centred* ensemble matrices: n -th column is $\mathbf{x}_n - \bar{\mathbf{x}}$ and $\mathbf{y}_n - \bar{\mathbf{y}}$.

Recall covariance estimate, $\hat{\mathbf{C}}_{y,x} = \frac{1}{N-1} \sum_n (\mathbf{y}_n - \bar{\mathbf{y}})(\mathbf{x}_n - \bar{\mathbf{x}})$ (1)

$$= \frac{1}{N-1} \mathbf{Y} \mathbf{X}^T \quad (2)$$

Define the ensemble LLS operator $\hat{\nabla}f$ as

$$\hat{\nabla}f = \hat{\mathbf{C}}_{y,y}^{-1} \hat{\mathbf{C}}_{y,x}^T \quad (3)$$

$$= \mathbf{Y} \mathbf{X}^T (\mathbf{X} \mathbf{X}^T)^{-1} \quad (4)$$

ML

If $N \leq 4$, reproduces increments

$$[\mathbf{y}_n - \bar{\mathbf{y}}] = (\hat{\nabla}f)^T [\mathbf{x}_n - \bar{\mathbf{x}}]$$

min-norm LLS, BLUE, MVUE

Relates to: finite differences (FD),
simplex derivatives



Covariances and LLS regression

Let $f(\mathbf{x}) = \mathbf{y}$.

Let \mathbf{X} and \mathbf{Y} be *centred* ensemble matrices: n -th column is $\mathbf{x}_n - \bar{\mathbf{x}}$ and $\mathbf{y}_n - \bar{\mathbf{y}}$.

Recall covariance estimate, $\bar{\mathbf{C}}_{\mathbf{y}, \mathbf{x}} := \frac{1}{N-1} \sum_n (\mathbf{y}_n - \bar{\mathbf{y}})(\mathbf{x}_n - \bar{\mathbf{x}})^T$ (1)

$$\bar{\mathbf{C}}_{\mathbf{y}, \mathbf{x}} = \frac{1}{N-1} \mathbf{Y} \mathbf{X}^T \quad (2)$$

Define the ensemble LLS operator $\tilde{\nabla} f$ as

$$\tilde{\nabla} f = \bar{\mathbf{C}}_{\mathbf{y}, \mathbf{x}} \bar{\mathbf{C}}_{\mathbf{x}}^{-1} \quad (3)$$

$$= \mathbf{Y} \mathbf{X}^T (\mathbf{X} \mathbf{X}^T)^{-1} \quad (4)$$

If $N \leq 4$, reproduces increments:

$$[\mathbf{y}_n - \bar{\mathbf{y}}] = (\tilde{\nabla} f)^T [\mathbf{x}_n - \bar{\mathbf{x}}]$$

min-norm LLS, BLUE, MVUE

Relates to: finite differences (FD),
simplex derivatives



Covariances and LLS regression

Let $f(\mathbf{x}) = \mathbf{y}$.

Let \mathbf{X} and \mathbf{Y} be *centred* ensemble matrices: n -th column is $\mathbf{x}_n - \bar{\mathbf{x}}$ and $\mathbf{y}_n - \bar{\mathbf{y}}$.

Recall covariance estimate, $\bar{\mathbf{C}}_{\mathbf{y}, \mathbf{x}} := \frac{1}{N-1} \sum_n (\mathbf{y}_n - \bar{\mathbf{y}})(\mathbf{x}_n - \bar{\mathbf{x}})^\top \quad (1)$

$$= \frac{1}{N-1} \mathbf{Y} \mathbf{X}^\top \quad (2)$$

Define the ensemble LLS operator $\tilde{\nabla} f$ as

$$\tilde{\nabla} f = \bar{\mathbf{C}}_{\mathbf{y}, \mathbf{x}} \bar{\mathbf{C}}_{\mathbf{x}}^{-1} \quad (3)$$

$$= \mathbf{Y} \mathbf{X}^\top (\mathbf{X} \mathbf{X}^\top)^{-1} \quad (4)$$

If $N \leq d$, reproduces increments:

$$[\mathbf{y}_n - \bar{\mathbf{y}}] = (\tilde{\nabla} f)^\top [\mathbf{x}_n - \bar{\mathbf{x}}]$$

min-norm LLS, BLUE, MVUE

Relates to: finite differences (FD),
simplex derivatives



Covariances and LLS regression

Let $f(\mathbf{x}) = \mathbf{y}$.

Let \mathbf{X} and \mathbf{Y} be *centred* ensemble matrices: n -th column is $\mathbf{x}_n - \bar{\mathbf{x}}$ and $\mathbf{y}_n - \bar{\mathbf{y}}$.

Recall covariance estimate, $\bar{\mathbf{C}}_{\mathbf{y}, \mathbf{x}} := \frac{1}{N-1} \sum_n (\mathbf{y}_n - \bar{\mathbf{y}})(\mathbf{x}_n - \bar{\mathbf{x}})^\top \quad (1)$

$$= \frac{1}{N-1} \mathbf{Y} \mathbf{X}^\top \quad (2)$$

Define the ensemble LLS operator $\bar{\nabla}$ as

$$\bar{\nabla}f := \bar{\mathbf{C}}_{\mathbf{y}, \mathbf{x}} \bar{\mathbf{C}}_{\mathbf{x}}^+ \quad (3)$$

ML

If $N \leq d$, reproduces increments

$$[\mathbf{y}_n - \bar{\mathbf{y}}] = (\bar{\nabla}f)^\top [\mathbf{x}_n - \bar{\mathbf{x}}]$$

min-norm LLS, BLUE, MVUE

Relates to: finite differences (FD),
simplex derivatives



Covariances and LLS regression

Let $f(\mathbf{x}) = \mathbf{y}$.

Let \mathbf{X} and \mathbf{Y} be *centred* ensemble matrices: n -th column is $\mathbf{x}_n - \bar{\mathbf{x}}$ and $\mathbf{y}_n - \bar{\mathbf{y}}$.

Recall covariance estimate, $\bar{\mathbf{C}}_{\mathbf{y}, \mathbf{x}} := \frac{1}{N-1} \sum_n (\mathbf{y}_n - \bar{\mathbf{y}})(\mathbf{x}_n - \bar{\mathbf{x}})^\top \quad (1)$

$$= \frac{1}{N-1} \mathbf{Y} \mathbf{X}^\top \quad (2)$$

Define the ensemble LLS operator $\bar{\nabla}$ as

$$\bar{\nabla}f := \bar{\mathbf{C}}_{\mathbf{y}, \mathbf{x}} \bar{\mathbf{C}}_{\mathbf{x}}^+ \quad (3)$$

$$= \mathbf{Y} \mathbf{X}^\top (\mathbf{X} \mathbf{X}^\top)^+ \quad (4)$$

ML

If $N \leq d$, reproduces increments

$$[\mathbf{y}_n - \bar{\mathbf{y}}] = (\bar{\nabla}f)^\top [\mathbf{x}_n - \bar{\mathbf{x}}]$$

min-norm LLS, BLUE, MVUE

Relates to: finite differences (FD),
simplex derivatives



Covariances and LLS regression

Let $f(\mathbf{x}) = \mathbf{y}$.

Let \mathbf{X} and \mathbf{Y} be *centred* ensemble matrices: n -th column is $\mathbf{x}_n - \bar{\mathbf{x}}$ and $\mathbf{y}_n - \bar{\mathbf{y}}$.

Recall covariance estimate, $\bar{\mathbf{C}}_{\mathbf{y}, \mathbf{x}} := \frac{1}{N-1} \sum_n (\mathbf{y}_n - \bar{\mathbf{y}})(\mathbf{x}_n - \bar{\mathbf{x}})^\top \quad (1)$

$$= \frac{1}{N-1} \mathbf{Y} \mathbf{X}^\top \quad (2)$$

Define the ensemble LLS operator $\bar{\nabla}$ as

$$\bar{\nabla}f := \bar{\mathbf{C}}_{\mathbf{y}, \mathbf{x}} \bar{\mathbf{C}}_{\mathbf{x}}^+ \quad (3)$$

$$= \mathbf{Y} \mathbf{X}^\top (\mathbf{X} \mathbf{X}^\top)^+ \quad (4)$$

$$(\mathbf{A}^+ = \mathbf{A}^\top (\mathbf{A} \mathbf{A}^\top)^+)$$

ML

If $N \leq d$, reproduces increments

$$[\mathbf{y}_n - \bar{\mathbf{y}}] = (\bar{\nabla}f)^\top [\mathbf{x}_n - \bar{\mathbf{x}}]$$

min-norm LLS, BLUE, MVUE

Relates to: finite differences (FD),

simplex derivatives



Covariances and LLS regression

Let $f(\mathbf{x}) = \mathbf{y}$.

Let \mathbf{X} and \mathbf{Y} be *centred* ensemble matrices: n -th column is $\mathbf{x}_n - \bar{\mathbf{x}}$ and $\mathbf{y}_n - \bar{\mathbf{y}}$.

Recall covariance estimate, $\bar{\mathbf{C}}_{\mathbf{y}, \mathbf{x}} := \frac{1}{N-1} \sum_n (\mathbf{y}_n - \bar{\mathbf{y}})(\mathbf{x}_n - \bar{\mathbf{x}})^\top \quad (1)$

$$= \frac{1}{N-1} \mathbf{Y} \mathbf{X}^\top \quad (2)$$

Define the ensemble LLS operator $\bar{\nabla}$ as

$$\bar{\nabla}f := \bar{\mathbf{C}}_{\mathbf{y}, \mathbf{x}} \bar{\mathbf{C}}_{\mathbf{x}}^+ \quad (3)$$

$$= \mathbf{Y} \mathbf{X}^\top (\mathbf{X} \mathbf{X}^\top)^+ \quad (4)$$

$$= \mathbf{Y} \mathbf{X}^+. \quad (\mathbf{A}^+ = \mathbf{A}^\top (\mathbf{A} \mathbf{A}^\top)^+)$$



Covariances and LLS regression

Let $f(\mathbf{x}) = \mathbf{y}$.

Let \mathbf{X} and \mathbf{Y} be *centred* ensemble matrices: n -th column is $\mathbf{x}_n - \bar{\mathbf{x}}$ and $\mathbf{y}_n - \bar{\mathbf{y}}$.

Recall covariance estimate, $\bar{\mathbf{C}}_{\mathbf{y}, \mathbf{x}} := \frac{1}{N-1} \sum_n (\mathbf{y}_n - \bar{\mathbf{y}})(\mathbf{x}_n - \bar{\mathbf{x}})^\top \quad (1)$

$$= \frac{1}{N-1} \mathbf{Y} \mathbf{X}^\top \quad (2)$$

Define the ensemble LLS operator $\bar{\nabla}$ as

• ML

$$\bar{\nabla}f := \bar{\mathbf{C}}_{\mathbf{y}, \mathbf{x}} \bar{\mathbf{C}}_{\mathbf{x}}^+ \quad (3)$$

$$= \mathbf{Y} \mathbf{X}^\top (\mathbf{X} \mathbf{X}^\top)^+ \quad (4)$$

$$= \mathbf{Y} \mathbf{X}^+. \quad (\mathbf{A}^+ = \mathbf{A}^\top (\mathbf{A} \mathbf{A}^\top)^+)$$

(if $N \leq d$, reproduces increments)

$[\mathbf{y}_n - \bar{\mathbf{y}}] = (\bar{\nabla}f)^\top [\mathbf{x}_n - \bar{\mathbf{x}}]$

min-norm LLS, BLUE, MVUE

Relates to: finite differences (FD),

simplex derivatives



Covariances and LLS regression

Let $f(\mathbf{x}) = \mathbf{y}$.

Let \mathbf{X} and \mathbf{Y} be *centred* ensemble matrices: n -th column is $\mathbf{x}_n - \bar{\mathbf{x}}$ and $\mathbf{y}_n - \bar{\mathbf{y}}$.

Recall covariance estimate, $\bar{\mathbf{C}}_{\mathbf{y}, \mathbf{x}} := \frac{1}{N-1} \sum_n (\mathbf{y}_n - \bar{\mathbf{y}})(\mathbf{x}_n - \bar{\mathbf{x}})^\top \quad (1)$

$$= \frac{1}{N-1} \mathbf{Y} \mathbf{X}^\top \quad (2)$$

Define the ensemble LLS operator $\bar{\nabla}$ as

- ML

$$\bar{\nabla}f := \bar{\mathbf{C}}_{\mathbf{y}, \mathbf{x}} \bar{\mathbf{C}}_{\mathbf{x}}^+ \quad (3)$$

- If $N \leq d_x$, reproduces increments

$$= \mathbf{Y} \mathbf{X}^\top (\mathbf{X} \mathbf{X}^\top)^+ \quad (4)$$

- $\bar{\nabla}f = (\bar{\nabla}f)^\top [\mathbf{x}_n - \bar{\mathbf{x}}]$

$$= \mathbf{Y} \mathbf{X}^+. \quad (\mathbf{A}^+ = \mathbf{A}^\top (\mathbf{A} \mathbf{A}^\top)^+)$$

- min-norm LLS, BLUE, MVUE

Relates to: finite differences (FD),
simplex derivatives



Covariances and LLS regression

Let $f(\mathbf{x}) = \mathbf{y}$.

Let \mathbf{X} and \mathbf{Y} be *centred* ensemble matrices: n -th column is $\mathbf{x}_n - \bar{\mathbf{x}}$ and $\mathbf{y}_n - \bar{\mathbf{y}}$.

Recall covariance estimate, $\bar{\mathbf{C}}_{\mathbf{y}, \mathbf{x}} := \frac{1}{N-1} \sum_n (\mathbf{y}_n - \bar{\mathbf{y}})(\mathbf{x}_n - \bar{\mathbf{x}})^\top \quad (1)$

$$= \frac{1}{N-1} \mathbf{Y} \mathbf{X}^\top \quad (2)$$

Define the ensemble LLS operator $\bar{\nabla}$ as

- ML
- If $N \leq d_x$, reproduces increments:
 $[\mathbf{y}_n - \bar{\mathbf{y}}] = (\bar{\nabla}f)^\top [\mathbf{x}_n - \bar{\mathbf{x}}]$.

$$\bar{\nabla}f := \bar{\mathbf{C}}_{\mathbf{y}, \mathbf{x}} \bar{\mathbf{C}}_{\mathbf{x}}^+ \quad (3)$$

$$= \mathbf{Y} \mathbf{X}^\top (\mathbf{X} \mathbf{X}^\top)^+ \quad (4)$$

$$= \mathbf{Y} \mathbf{X}^+. \quad (\mathbf{A}^+ = \mathbf{A}^\top (\mathbf{A} \mathbf{A}^\top)^+)$$

min-norm LLS, BLUE, MVUE
Relates to: finite differences (FD),
simplex derivatives



Covariances and LLS regression

Let $f(\mathbf{x}) = \mathbf{y}$.

Let \mathbf{X} and \mathbf{Y} be *centred* ensemble matrices: n -th column is $\mathbf{x}_n - \bar{\mathbf{x}}$ and $\mathbf{y}_n - \bar{\mathbf{y}}$.

Recall covariance estimate, $\bar{\mathbf{C}}_{\mathbf{y}, \mathbf{x}} := \frac{1}{N-1} \sum_n (\mathbf{y}_n - \bar{\mathbf{y}})(\mathbf{x}_n - \bar{\mathbf{x}})^\top \quad (1)$

$$= \frac{1}{N-1} \mathbf{Y} \mathbf{X}^\top \quad (2)$$

Define the ensemble LLS operator $\bar{\nabla}$ as

$$\bar{\nabla}f := \bar{\mathbf{C}}_{\mathbf{y}, \mathbf{x}} \bar{\mathbf{C}}_{\mathbf{x}}^+ \quad (3)$$

$$= \mathbf{Y} \mathbf{X}^\top (\mathbf{X} \mathbf{X}^\top)^+ \quad (4)$$

$$= \mathbf{Y} \mathbf{X}^+. \quad (\mathbf{A}^+ = \mathbf{A}^\top (\mathbf{A} \mathbf{A}^\top)^+)$$

- ML
- If $N \leq d_x$, reproduces increments:
 $[\mathbf{y}_n - \bar{\mathbf{y}}] = (\bar{\nabla}f)^\top [\mathbf{x}_n - \bar{\mathbf{x}}]$.
- min-norm LLS

Relates to finite differences (FD),
simplex derivatives



Covariances and LLS regression

Let $f(\mathbf{x}) = \mathbf{y}$.

Let \mathbf{X} and \mathbf{Y} be *centred* ensemble matrices: n -th column is $\mathbf{x}_n - \bar{\mathbf{x}}$ and $\mathbf{y}_n - \bar{\mathbf{y}}$.

Recall covariance estimate, $\bar{\mathbf{C}}_{\mathbf{y}, \mathbf{x}} := \frac{1}{N-1} \sum_n (\mathbf{y}_n - \bar{\mathbf{y}})(\mathbf{x}_n - \bar{\mathbf{x}})^\top \quad (1)$

$$= \frac{1}{N-1} \mathbf{Y} \mathbf{X}^\top \quad (2)$$

Define the ensemble LLS operator $\bar{\nabla}$ as

$$\bar{\nabla}f := \bar{\mathbf{C}}_{\mathbf{y}, \mathbf{x}} \bar{\mathbf{C}}_{\mathbf{x}}^+ \quad (3)$$

$$= \mathbf{Y} \mathbf{X}^\top (\mathbf{X} \mathbf{X}^\top)^+ \quad (4)$$

$$= \mathbf{Y} \mathbf{X}^+. \quad (\mathbf{A}^+ = \mathbf{A}^\top (\mathbf{A} \mathbf{A}^\top)^+)$$

- ML
- If $N \leq d_x$, reproduces increments:
 $[\mathbf{y}_n - \bar{\mathbf{y}}] = (\bar{\nabla}f)^\top [\mathbf{x}_n - \bar{\mathbf{x}}]$.
- min-norm LLS, BLUE, MVUE

Relates to finite differences (FD)

simplex derivatives



Covariances and LLS regression

Let $f(\mathbf{x}) = \mathbf{y}$.

Let \mathbf{X} and \mathbf{Y} be *centred* ensemble matrices: n -th column is $\mathbf{x}_n - \bar{\mathbf{x}}$ and $\mathbf{y}_n - \bar{\mathbf{y}}$.

Recall covariance estimate, $\bar{\mathbf{C}}_{\mathbf{y}, \mathbf{x}} := \frac{1}{N-1} \sum_n (\mathbf{y}_n - \bar{\mathbf{y}})(\mathbf{x}_n - \bar{\mathbf{x}})^\top \quad (1)$

$$= \frac{1}{N-1} \mathbf{Y} \mathbf{X}^\top \quad (2)$$

Define the ensemble LLS operator $\bar{\nabla}$ as

$$\bar{\nabla}f := \bar{\mathbf{C}}_{\mathbf{y}, \mathbf{x}} \bar{\mathbf{C}}_{\mathbf{x}}^+ \quad (3)$$

$$= \mathbf{Y} \mathbf{X}^\top (\mathbf{X} \mathbf{X}^\top)^+ \quad (4)$$

$$= \mathbf{Y} \mathbf{X}^+. \quad (\mathbf{A}^+ = \mathbf{A}^\top (\mathbf{A} \mathbf{A}^\top)^+)$$

- ML
- If $N \leq d_x$, reproduces increments:
 $[\mathbf{y}_n - \bar{\mathbf{y}}] = (\bar{\nabla}f)^\top [\mathbf{x}_n - \bar{\mathbf{x}}]$.
- min-norm LLS, BLUE, MVUE
- Relates to: finite differences (FD), simplex derivatives



Stein's lemma

Let $\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{C}_x)$, i.e. $p(\mathbf{x}) \propto e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T \mathbf{C}_x^{-1} (\mathbf{x}-\boldsymbol{\mu})}$.

$$\mathbb{E} \nabla f(\mathbf{x}) = \int_{\mathbb{R}^d} \nabla f(\mathbf{x}) p(\mathbf{x}) d\mathbf{x} \quad (5)$$

$$= - \int_{\mathbb{R}^d} f(\mathbf{x}) [\nabla p(\mathbf{x})]^T d\mathbf{x} + [f(\mathbf{x}) p(\mathbf{x})]_{\mathbf{x}=\boldsymbol{\mu}} \quad (6)$$

using integration by parts, and $p(\mathbf{x}) \xrightarrow[\mathbf{x} \rightarrow \infty]{} 0$.

$$\begin{aligned} \text{But } \nabla p(\mathbf{x}) &= p(\mathbf{x}) \nabla \log p(\mathbf{x}) \\ &= p(\mathbf{x}) \mathbf{C}_x^{-1} (\mathbf{x} - \boldsymbol{\mu}), \end{aligned}$$

and so

$$\mathbb{E} \nabla f(\mathbf{x}) = \int_{\mathbb{R}^d} f(\mathbf{x}) p(\mathbf{x}) (\mathbf{x} - \boldsymbol{\mu})^T d\mathbf{x} \cdot \mathbf{C}_x^{-1} \quad (7)$$

$$= \mathbf{C}_{y,x} \mathbf{C}_x^{-1}. \quad (8)$$

Lastly, by Slutsky, $\mathbf{C}_{y,x} \mathbf{C}_x^T \xrightarrow[n \rightarrow \infty]{\text{P.s.}} \mathbf{C}_{y,x} \mathbf{C}_x^{-1}$



Stein's lemma

Let $\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{C}_x)$, i.e. $p(\mathbf{x}) \propto e^{-\|\mathbf{x}-\boldsymbol{\mu}\|_{\mathbf{C}_x}^2/2}$.

$$\mathbb{E} \nabla f(\mathbf{x}) = \int_{\mathbb{R}^d} \nabla f(\mathbf{x}) p(\mathbf{x}) d\mathbf{x} \quad (5)$$

$$= - \int_{\mathbb{R}^d} f(\mathbf{x}) [\nabla p(\mathbf{x})]^T d\mathbf{x} + [f(\mathbf{x}) p(\mathbf{x})]_{\mathbf{x}=\boldsymbol{\mu}} \quad (6)$$

using integration by parts, and $p(\mathbf{x}) \xrightarrow[\mathbf{x} \rightarrow \infty]{} 0$.

$$\begin{aligned} \nabla p(\mathbf{x}) &= p(\mathbf{x}) \nabla \log p(\mathbf{x}) \\ &= p(\mathbf{x}) \mathbf{C}_x^{-1} (\mathbf{x} - \boldsymbol{\mu}), \end{aligned}$$

and so

$$\mathbb{E} \nabla f(\mathbf{x}) = \int_{\mathbb{R}^d} f(\mathbf{x}) p(\mathbf{x}) (\mathbf{x} - \boldsymbol{\mu})^T d\mathbf{x} \cdot \mathbf{C}_x^{-1} \quad (7)$$

$$= \mathbf{C}_{y,x} \mathbf{C}_x^{-1}. \quad (8)$$

Lastly, by Slutsky, $\mathbf{C}_{y,x} \mathbf{C}_x^{-1} \xrightarrow[n \rightarrow \infty]{\text{P.s.}} \mathbf{C}_{y,x} \mathbf{C}_x^{-1}$



Stein's lemma

Let $\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{C}_x)$, i.e. $p(\mathbf{x}) \propto e^{-\|\mathbf{x}-\boldsymbol{\mu}\|_{\mathbf{C}_x}^2/2}$.

$$\mathbb{E} \nabla f(\mathbf{x}) := \int_{\mathbb{R}^d} \nabla f(\mathbf{x}) p(\mathbf{x}) d\mathbf{x} \quad (5)$$

$$= - \int_{\mathbb{R}^d} f(\mathbf{x}) [\nabla p(\mathbf{x})]^T d\mathbf{x} + [f(\mathbf{x}) p(\mathbf{x})]_{\mathbf{x}=\boldsymbol{\mu}} \quad (6)$$

using integration by parts, and $p(\mathbf{x}) \xrightarrow[\mathbf{x} \rightarrow \infty]{} 0$.

$$\text{But } \nabla p(\mathbf{x}) = p(\mathbf{x}) \nabla \log p(\mathbf{x})$$

$$= p(\mathbf{x}) \mathbf{C}_x^{-1} (\mathbf{x} - \boldsymbol{\mu}),$$

and so

$$\mathbb{E} \nabla f(\mathbf{x}) = \int_{\mathbb{R}^d} f(\mathbf{x}) p(\mathbf{x}) (\mathbf{x} - \boldsymbol{\mu})^T d\mathbf{x} \cdot \mathbf{C}_x^{-1} \quad (7)$$

$$= \mathbf{C}_{y,x} \mathbf{C}_x^{-1} \quad (8)$$

Lastly, by Slutsky, $\mathbf{C}_{y,x} \mathbf{C}_x^{-1} \xrightarrow[n \rightarrow \infty]{} \mathbf{C}_{y,x} \mathbf{C}_x^{-1}$



Stein's lemma

Let $\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{C}_x)$, i.e. $p(\mathbf{x}) \propto e^{-\|\mathbf{x}-\boldsymbol{\mu}\|_{\mathbf{C}_x}^2/2}$.

$$\mathbb{E} \nabla f(\mathbf{x}) := \int_{\mathbb{R}^d} \nabla f(\mathbf{x}) p(\mathbf{x}) d\mathbf{x} \tag{5}$$

$$= - \int_{\mathbb{R}^d} f(\mathbf{x}) (\nabla p(\mathbf{x}))^\top d\mathbf{x} + [f(\mathbf{x}) p(\mathbf{x})]_{\mathbf{x}=\boldsymbol{\mu}} \tag{6}$$

using integration by parts, and $p(\mathbf{x}) \xrightarrow[\mathbf{x} \rightarrow \boldsymbol{\mu}]{} 0$.

$$\begin{aligned} \nabla p(\mathbf{x}) &= p(\mathbf{x}) \nabla \log p(\mathbf{x}) \\ &= p(\mathbf{x}) \mathbf{C}_x^{-1} (\mathbf{x} - \boldsymbol{\mu}), \end{aligned}$$

and so

$$\mathbb{E} \nabla f(\mathbf{x}) = \int_{\mathbb{R}^d} f(\mathbf{x}) p(\mathbf{x}) (\mathbf{x} - \boldsymbol{\mu})^\top d\mathbf{x} \cdot \mathbf{C}_x^{-1} \tag{7}$$

$$= \mathbf{C}_{y,x} \mathbf{C}_x^{-1}. \tag{8}$$

Lastly, by Slutsky, $\mathbf{C}_{y,x} \mathbf{C}_x^{-1} \xrightarrow[n \rightarrow \infty]{} \mathbf{C}_{y,x} \mathbf{C}_x^{-1}$



Stein's lemma

Let $\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{C}_x)$, i.e. $p(\mathbf{x}) \propto e^{-\|\mathbf{x}-\boldsymbol{\mu}\|_{\mathbf{C}_x}^2/2}$.

$$\mathbb{E} \nabla f(\mathbf{x}) := \int_{\mathbb{R}^d} \nabla f(\mathbf{x}) p(\mathbf{x}) d\mathbf{x} \quad (5)$$

$$= - \int_{\mathbb{R}^d} f(\mathbf{x}) [\nabla p(\mathbf{x})]^\top d\mathbf{x} + [f(\mathbf{x}) p(\mathbf{x}) \mathbf{n}]_{\partial \mathbb{R}^d} \quad (6)$$

using integration by parts, and $p(\mathbf{x}) \xrightarrow[\mathbf{x} \rightarrow \infty]{} 0$

$$\text{but } \nabla p(\mathbf{x}) = p(\mathbf{x}) \nabla \log p(\mathbf{x})$$

$$= p(\mathbf{x}) \mathbf{C}_x^{-1} (\mathbf{x} - \boldsymbol{\mu}),$$

and so

$$\mathbb{E} \nabla f(\mathbf{x}) = \int_{\mathbb{R}^d} f(\mathbf{x}) p(\mathbf{x}) (\mathbf{x} - \boldsymbol{\mu})^\top d\mathbf{x} \cdot \mathbf{C}_x^{-1} \quad (7)$$

$$= \mathbf{C}_{y,x} \mathbf{C}_x^{-1}. \quad (8)$$

Lastly, by Slutsky, $\mathbf{C}_{y,x} \mathbf{C}_x^{-1} \xrightarrow[n \rightarrow \infty]{} \mathbf{C}_{y,x} \mathbf{C}_x^{-1}$



Stein's lemma

Let $\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{C}_x)$, i.e. $p(\mathbf{x}) \propto e^{-\|\mathbf{x}-\boldsymbol{\mu}\|_{\mathbf{C}_x}^2/2}$.

$$\mathbb{E} \nabla f(\mathbf{x}) := \int_{\mathbb{R}^d} \nabla f(\mathbf{x}) p(\mathbf{x}) d\mathbf{x} \quad (5)$$

$$= - \int_{\mathbb{R}^d} f(\mathbf{x}) [\nabla p(\mathbf{x})]^\top d\mathbf{x} + \cancel{[f(\mathbf{x}) p(\mathbf{x}) \mathbf{n}]_{\partial \mathbb{R}^d}} \quad (6)$$

using integration by parts, and $p(\mathbf{x}) \xrightarrow{\partial \mathbb{R}^d} 0$.

$$\text{but } \nabla p(\mathbf{x}) = p(\mathbf{x}) \nabla \log p(\mathbf{x})$$

$$= p(\mathbf{x}) \mathbf{C}_x^{-1} (\mathbf{x} - \boldsymbol{\mu}),$$

and so

$$\mathbb{E} \nabla f(\mathbf{x}) = \int_{\mathbb{R}^d} f(\mathbf{x}) p(\mathbf{x}) (\mathbf{x} - \boldsymbol{\mu})^\top d\mathbf{x} \cdot \mathbf{C}_x^{-1} \quad (7)$$

$$= \mathbf{C}_{y,x} \mathbf{C}_x^{-1}. \quad (8)$$

Lastly, by Slutsky, $\mathbf{C}_{y,x} \mathbf{C}_x^{-1} \xrightarrow[n \rightarrow \infty]{} \mathbf{C}_{y,x} \mathbf{C}_x^{-1}$



Stein's lemma

Let $\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{C}_x)$, i.e. $p(\mathbf{x}) \propto e^{-\|\mathbf{x}-\boldsymbol{\mu}\|_{\mathbf{C}_x}^2/2}$.

$$\mathbb{E} \nabla f(\mathbf{x}) := \int_{\mathbb{R}^d} \nabla f(\mathbf{x}) p(\mathbf{x}) d\mathbf{x} \quad (5)$$

$$= - \int_{\mathbb{R}^d} f(\mathbf{x}) [\nabla p(\mathbf{x})]^\top d\mathbf{x} + \cancel{[f(\mathbf{x}) p(\mathbf{x}) \mathbf{n}]_{\partial \mathbb{R}^d}} \quad (6)$$

using integration by parts, and $p(\mathbf{x}) \xrightarrow{\partial \mathbb{R}^d} 0$.

But $\nabla p(\mathbf{x}) = p(\mathbf{x}) \nabla \log p(\mathbf{x})$

$$\nabla \log p(\mathbf{x}) = \mathbf{C}_x^{-1} (\mathbf{x} - \boldsymbol{\mu})$$

and so

$$\mathbb{E} \nabla f(\mathbf{x}) = \int_{\mathbb{R}^d} f(\mathbf{x}) p(\mathbf{x}) (\mathbf{x} - \boldsymbol{\mu})^\top d\mathbf{x} \cdot \mathbf{C}_x^{-1} \quad (7)$$

$$= \mathbf{C}_{y,x} \mathbf{C}_x^{-1} \quad (8)$$

Lastly, by Slutsky, $\mathbf{C}_{y,x} \mathbf{C}_x^{-1} \xrightarrow[n \rightarrow \infty]{\text{P}} \mathbf{C}_{y,x} \mathbf{C}_x^{-1}$



Stein's lemma

Let $\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{C}_x)$, i.e. $p(\mathbf{x}) \propto e^{-\|\mathbf{x}-\boldsymbol{\mu}\|_{\mathbf{C}_x}^2/2}$.

$$\mathbb{E} \nabla f(\mathbf{x}) := \int_{\mathbb{R}^d} \nabla f(\mathbf{x}) p(\mathbf{x}) d\mathbf{x} \quad (5)$$

$$= - \int_{\mathbb{R}^d} f(\mathbf{x}) [\nabla p(\mathbf{x})]^\top d\mathbf{x} + \cancel{[f(\mathbf{x}) p(\mathbf{x}) \mathbf{n}]_{\partial \mathbb{R}^d}} \quad (6)$$

using integration by parts, and $p(\mathbf{x}) \xrightarrow{\partial \mathbb{R}^d} 0$.

$$\begin{aligned} \text{But } \nabla p(\mathbf{x}) &= p(\mathbf{x}) \nabla \log p(\mathbf{x}) \\ &= p(\mathbf{x}) \mathbf{C}_x^{-1} (\boldsymbol{\mu} - \mathbf{x}), \end{aligned}$$

and so

$$\mathbb{E} \nabla f(\mathbf{x}) = \int_{\mathbb{R}^d} f(\mathbf{x}) p(\mathbf{x}) (\mathbf{x} - \boldsymbol{\mu})^\top d\mathbf{x} \cdot \mathbf{C}_x^{-1} \quad (7)$$

$$= \mathbf{C}_{yy} \mathbf{C}_x^{-1}. \quad (8)$$

Lastly, by Slutsky, $\mathbf{C}_{yy} \mathbf{C}_x^{-1} \xrightarrow{p} \mathbf{C}_{yy} \mathbf{C}_x^{-1}$



Stein's lemma

Let $\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{C}_x)$, i.e. $p(\mathbf{x}) \propto e^{-\|\mathbf{x}-\boldsymbol{\mu}\|_{\mathbf{C}_x}^2/2}$.

$$\mathbb{E} \nabla f(\mathbf{x}) := \int_{\mathbb{R}^d} \nabla f(\mathbf{x}) p(\mathbf{x}) d\mathbf{x} \quad (5)$$

$$= - \int_{\mathbb{R}^d} f(\mathbf{x}) [\nabla p(\mathbf{x})]^\top d\mathbf{x} + \cancel{[f(\mathbf{x}) p(\mathbf{x}) \mathbf{n}]_{\partial \mathbb{R}^d}} \quad (6)$$

using integration by parts, and $p(\mathbf{x}) \xrightarrow{\partial \mathbb{R}^d} 0$.

$$\begin{aligned} \text{But } \nabla p(\mathbf{x}) &= p(\mathbf{x}) \nabla \log p(\mathbf{x}) \\ &= p(\mathbf{x}) \mathbf{C}_x^{-1} (\boldsymbol{\mu} - \mathbf{x}), \end{aligned}$$

and so

$$\mathbb{E} \nabla f(\mathbf{x}) = \int_{\mathbb{R}^d} f(\mathbf{x}) p(\mathbf{x}) (\mathbf{x} - \boldsymbol{\mu})^\top d\mathbf{x} \mathbf{C}_x^{-1} \quad (7)$$

Lastly, by Slutsky, $\mathbf{C}_{yy} \mathbf{C}_y^2 \xrightarrow{\text{a.s.}} \mathbf{C}_{yy} \mathbf{C}_y^{-2}$



Stein's lemma

Let $\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{C}_x)$, i.e. $p(\mathbf{x}) \propto e^{-\|\mathbf{x}-\boldsymbol{\mu}\|_{\mathbf{C}_x}^2/2}$.

$$\mathbb{E} \nabla f(\mathbf{x}) := \int_{\mathbb{R}^d} \nabla f(\mathbf{x}) p(\mathbf{x}) d\mathbf{x} \quad (5)$$

$$= - \int_{\mathbb{R}^d} f(\mathbf{x}) [\nabla p(\mathbf{x})]^\top d\mathbf{x} + \cancel{[f(\mathbf{x}) p(\mathbf{x}) \mathbf{n}]_{\partial \mathbb{R}^d}} \quad (6)$$

using integration by parts, and $p(\mathbf{x}) \xrightarrow{\partial \mathbb{R}^d} 0$.

$$\begin{aligned} \text{But } \nabla p(\mathbf{x}) &= p(\mathbf{x}) \nabla \log p(\mathbf{x}) \\ &= p(\mathbf{x}) \mathbf{C}_x^{-1} (\boldsymbol{\mu} - \mathbf{x}), \end{aligned}$$

and so

$$\mathbb{E} \nabla f(\mathbf{x}) = \int_{\mathbb{R}^d} f(\mathbf{x}) p(\mathbf{x}) (\mathbf{x} - \boldsymbol{\mu})^\top d\mathbf{x} \mathbf{C}_x^{-1} \quad (7)$$

$$= \mathbf{C}_{y,x} \mathbf{C}_x^{-1}. \quad (8)$$

Lastly, by Slutsky, $\mathbf{C}_{y,x} \mathbf{C}_x^{-1} \xrightarrow{\text{a.s.}} \mathbf{C}_{y,x} \mathbf{C}_x^{-1}$



Stein's lemma

Let $\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{C}_x)$, i.e. $p(\mathbf{x}) \propto e^{-\|\mathbf{x}-\boldsymbol{\mu}\|_{\mathbf{C}_x}^2/2}$.

$$\mathbb{E} \nabla f(\mathbf{x}) := \int_{\mathbb{R}^d} \nabla f(\mathbf{x}) p(\mathbf{x}) d\mathbf{x} \quad (5)$$

$$= - \int_{\mathbb{R}^d} f(\mathbf{x}) [\nabla p(\mathbf{x})]^\top d\mathbf{x} + \cancel{[f(\mathbf{x}) p(\mathbf{x}) \mathbf{n}]_{\partial \mathbb{R}^d}} \quad (6)$$

using integration by parts, and $p(\mathbf{x}) \xrightarrow{\partial \mathbb{R}^d} 0$.

$$\begin{aligned} \text{But } \nabla p(\mathbf{x}) &= p(\mathbf{x}) \nabla \log p(\mathbf{x}) \\ &= p(\mathbf{x}) \mathbf{C}_x^{-1} (\boldsymbol{\mu} - \mathbf{x}), \end{aligned}$$

and so

$$\mathbb{E} \nabla f(\mathbf{x}) = \int_{\mathbb{R}^d} f(\mathbf{x}) p(\mathbf{x}) (\mathbf{x} - \boldsymbol{\mu})^\top d\mathbf{x} \mathbf{C}_x^{-1} \quad (7)$$

$$= \bar{\mathbf{C}}_{y,x} \mathbf{C}_x^{-1}. \quad (8)$$

Lastly, by Slutsky, $\bar{\mathbf{C}}_{y,x} \bar{\mathbf{C}}_x^+ \xrightarrow[N \rightarrow \infty]{} \mathbf{C}_{y,x} \mathbf{C}_x^{-1}$.



Stein's advantages

Raanes et al. (2019):

$$\mathbf{YX}^+ =: \bar{\nabla}f \xrightarrow[N \rightarrow \infty]{} \mathbb{E} \nabla f(\mathbf{x}) \quad \text{if } \mathbf{x} \sim \mathcal{N} \quad (9)$$

Provides link to analytic gradient, ∇f .

Better than solving order- n Taylor @ mean:

$$\begin{aligned}\mathbf{Y}_n &\approx \nabla f(\bar{\mathbf{x}}) \mathbf{X}_n \\ \text{i.e. } \hat{\mathbf{C}}_{y,x} &\approx \nabla f(\bar{\mathbf{x}}) \hat{\mathbf{C}}_{x,x}\end{aligned}$$

because Stein (9)

Shows what the ODE target is.

Shows that errors vanish as $N \rightarrow \infty$.

Enables saying convergence without blushing.



Stein's advantages

Raanes et al. (2019):

$$\mathbf{YX}^+ =: \bar{\nabla}f \xrightarrow[N \rightarrow \infty]{} \mathbb{E} \nabla f(\mathbf{x}) \quad \text{if } \mathbf{x} \sim \mathcal{N} \quad (9)$$

Provides link to analytic gradient, ∇f .

Better than solving order-1 Taylor @ mean:

$$\begin{aligned} \mathbf{Y}_\lambda &\approx \nabla f(\mathbf{x}_\lambda) \mathbf{X}_\lambda \\ \text{i.e. } \hat{\mathbf{C}}_{y,x} &\approx \nabla f(\mathbf{x}_\lambda) \hat{\mathbf{C}}_{x,x} \end{aligned}$$

because Stein (9)

Shows what the OLS target is.

Shows that errors vanish as $N \rightarrow \infty$.

Enables saying *overfitting* without blushing.



Stein's advantages

Raanes et al. (2019):

$$\mathbf{Y}\mathbf{X}^+ =: \bar{\nabla}f \xrightarrow[N \rightarrow \infty]{} \mathbb{E} \nabla f(\mathbf{x}) \quad \text{if } \mathbf{x} \sim \mathcal{N} \quad (9)$$

Provides link to analytic gradient, ∇f .

Better than solving order-1 Taylor @ mean:

$$\mathbf{Y} \approx \nabla f(\bar{\mathbf{x}}) \mathbf{X},$$

i.e. $\mathbf{C}_{\text{opt}} \approx \mathbf{Y}(\mathbf{x}) \mathbf{C}_y$

because Stein (9)

Shows what the \mathbf{C}_{opt} target is.

Shows that errors vanish as $N \rightarrow \infty$.

Enables saying convergence without blushing.



Stein's advantages

Raanes et al. (2019):

$$\mathbf{Y}\mathbf{X}^+ =: \bar{\nabla}f \xrightarrow[N \rightarrow \infty]{} \mathbb{E} \nabla f(\mathbf{x}) \quad \text{if } \mathbf{x} \sim \mathcal{N} \quad (9)$$

Provides link to analytic gradient, ∇f .

Better than solving order-1 Taylor @ mean:

$$\begin{aligned} \mathbf{Y} &\approx \nabla f(\bar{\mathbf{x}}) \mathbf{X}, \\ \text{i.e. } \bar{\mathbf{C}}_{y,x} &\approx \nabla f(\bar{\mathbf{x}}) \bar{\mathbf{C}}_x, \end{aligned}$$

because Stein (9)

Shows what the estimator targets.

Shows that errors vanish as $N \rightarrow \infty$.

Enables saying something quantitative
without blushing.



Stein's advantages

Raanes et al. (2019):

$$\mathbf{Y}\mathbf{X}^+ =: \bar{\nabla}f \xrightarrow[N \rightarrow \infty]{} \mathbb{E} \nabla f(\mathbf{x}) \quad \text{if } \mathbf{x} \sim \mathcal{N} \quad (9)$$

Provides link to analytic gradient, ∇f .

Better than solving order-1 Taylor @ mean:

$$\begin{aligned} \mathbf{Y} &\approx \nabla f(\bar{\mathbf{x}}) \mathbf{X}, \\ \text{i.e. } \bar{\mathbf{C}}_{y,x} &\approx \nabla f(\bar{\mathbf{x}}) \bar{\mathbf{C}}_x, \end{aligned}$$

because Stein (9)

Shows what the estimator targets.

Shows that errors vanish as $N \rightarrow \infty$.

Enables saying something about
without blushing.



Stein's advantages

Raanes et al. (2019):

$$\mathbf{Y}\mathbf{X}^+ =: \bar{\nabla}f \xrightarrow[N \rightarrow \infty]{} \mathbb{E} \nabla f(\mathbf{x}) \quad \text{if } \mathbf{x} \sim \mathcal{N} \quad (9)$$

Provides link to analytic gradient, ∇f .

Better than solving order-1 Taylor @ mean:

$$\begin{aligned} \mathbf{Y} &\approx \nabla f(\bar{\mathbf{x}}) \mathbf{X}, \\ \text{i.e. } \bar{\mathbf{C}}_{y,x} &\approx \nabla f(\bar{\mathbf{x}}) \bar{\mathbf{C}}_x, \end{aligned}$$

because Stein (9)

- Shows what the *exact* target is.

Enables saying exact gradients
without blushing.



Stein's advantages

Raanes et al. (2019):

$$\mathbf{Y}\mathbf{X}^+ =: \bar{\nabla}f \xrightarrow[N \rightarrow \infty]{} \mathbb{E} \nabla f(\mathbf{x}) \quad \text{if } \mathbf{x} \sim \mathcal{N} \quad (9)$$

Provides link to analytic gradient, ∇f .

Better than solving order-1 Taylor @ mean:

$$\begin{aligned} \mathbf{Y} &\approx \nabla f(\bar{\mathbf{x}}) \mathbf{X}, \\ \text{i.e. } \bar{\mathbf{C}}_{y,x} &\approx \nabla f(\bar{\mathbf{x}}) \bar{\mathbf{C}}_x, \end{aligned}$$

because Stein (9)

- Shows what the *exact* target is.
- Shows that errors *cancel* as $N \rightarrow \infty$.

Enables saying *exact* without blushing.



Stein's advantages

Raanes et al. (2019):

$$\mathbf{Y}\mathbf{X}^+ =: \bar{\nabla}f \xrightarrow[N \rightarrow \infty]{} \mathbb{E} \nabla f(\mathbf{x}) \quad \text{if } \mathbf{x} \sim \mathcal{N} \quad (9)$$

Provides link to analytic gradient, ∇f .

Better than solving order-1 Taylor @ mean:

$$\begin{aligned} \mathbf{Y} &\approx \nabla f(\bar{\mathbf{x}}) \mathbf{X}, \\ \text{i.e. } \bar{\mathbf{C}}_{y,x} &\approx \nabla f(\bar{\mathbf{x}}) \bar{\mathbf{C}}_x, \end{aligned}$$

because Stein (9)

- Shows what the *exact* target is.
- Shows that errors *cancel* as $N \rightarrow \infty$.
- Enables saying *average sensitivity* without blushing.



Stein's advantages

Raanes et al. (2019):

$$\mathbf{YX}^+ =: \bar{\nabla}f \xrightarrow[N \rightarrow \infty]{} \mathbb{E} \nabla f(\mathbf{x}) \quad \text{if } \mathbf{x} \sim \mathcal{N} \quad (9)$$

Provides link to analytic gradient, ∇f .

Better than solving order-1 Taylor @ mean:

$$\begin{aligned} \mathbf{Y} &\approx \nabla f(\bar{\mathbf{x}}) \mathbf{X}, \\ \text{i.e. } \bar{\mathbf{C}}_{y,x} &\approx \nabla f(\bar{\mathbf{x}}) \bar{\mathbf{C}}_x, \end{aligned}$$

because Stein (9)

- Shows what the *exact* target is.
- Shows that errors *cancel* as $N \rightarrow \infty$.
- Enables saying *average sensitivity* without blushing.





Stein's advantages

Raanes et al. (2019):

$$\mathbf{Y}\mathbf{X}^+ =: \bar{\nabla}f \xrightarrow[N \rightarrow \infty]{} \mathbb{E} \nabla f(\mathbf{x}) \quad \text{if } x \sim \mathcal{N} \quad (9)$$

Stordal et al. (2016):

$$= \nabla \mathbb{E} f(\mathbf{x})$$

Provides link to analytic gradient, ∇f .

Better than solving order-1 Taylor @ mean:

$$\mathbf{Y} \approx \nabla f(\bar{\mathbf{x}}) \mathbf{X},$$

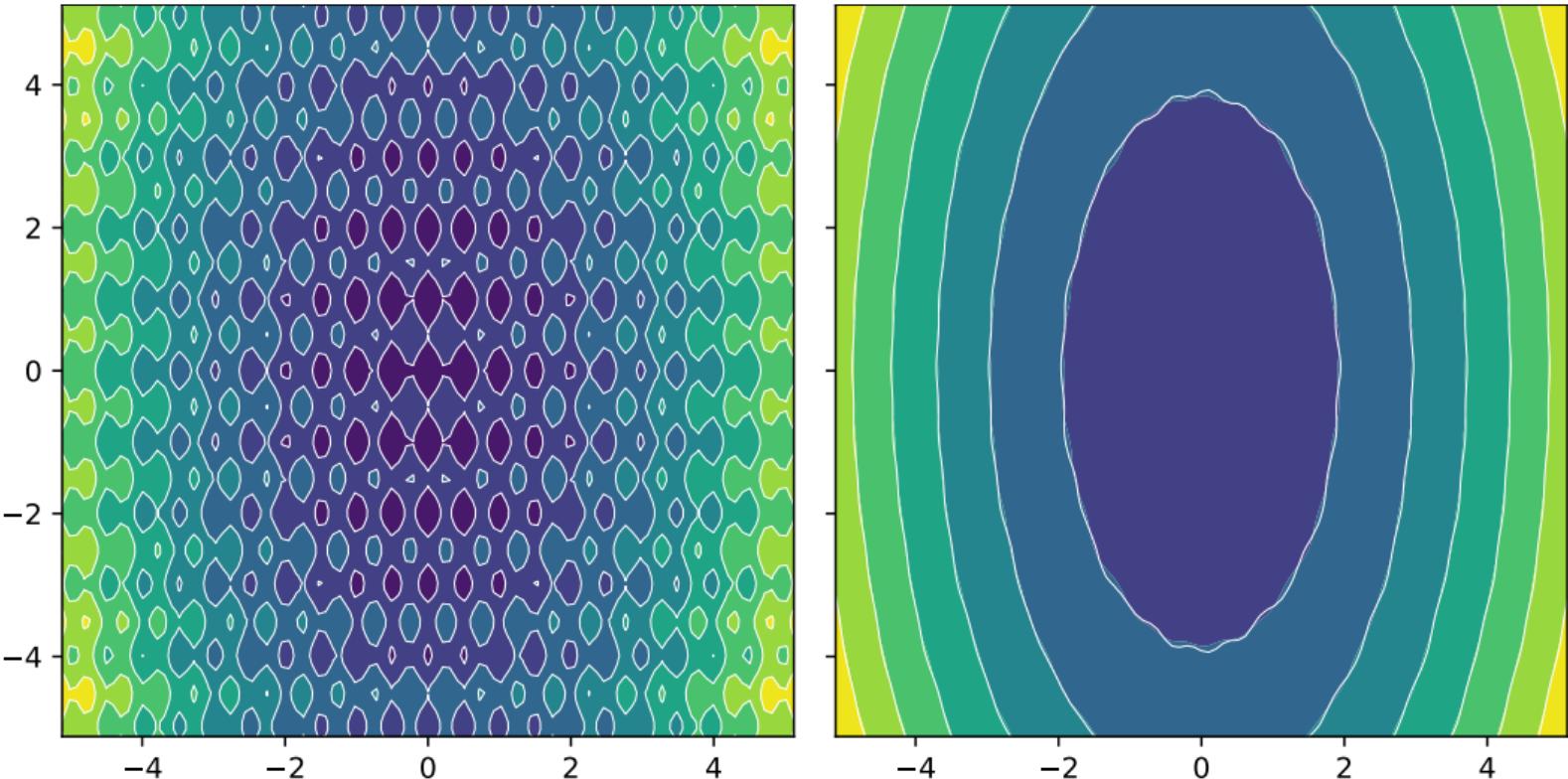
$$\text{i.e. } \bar{\mathbf{C}}_{y,x} \approx \nabla f(\bar{\mathbf{x}}) \bar{\mathbf{C}}_x,$$

because Stein (9)

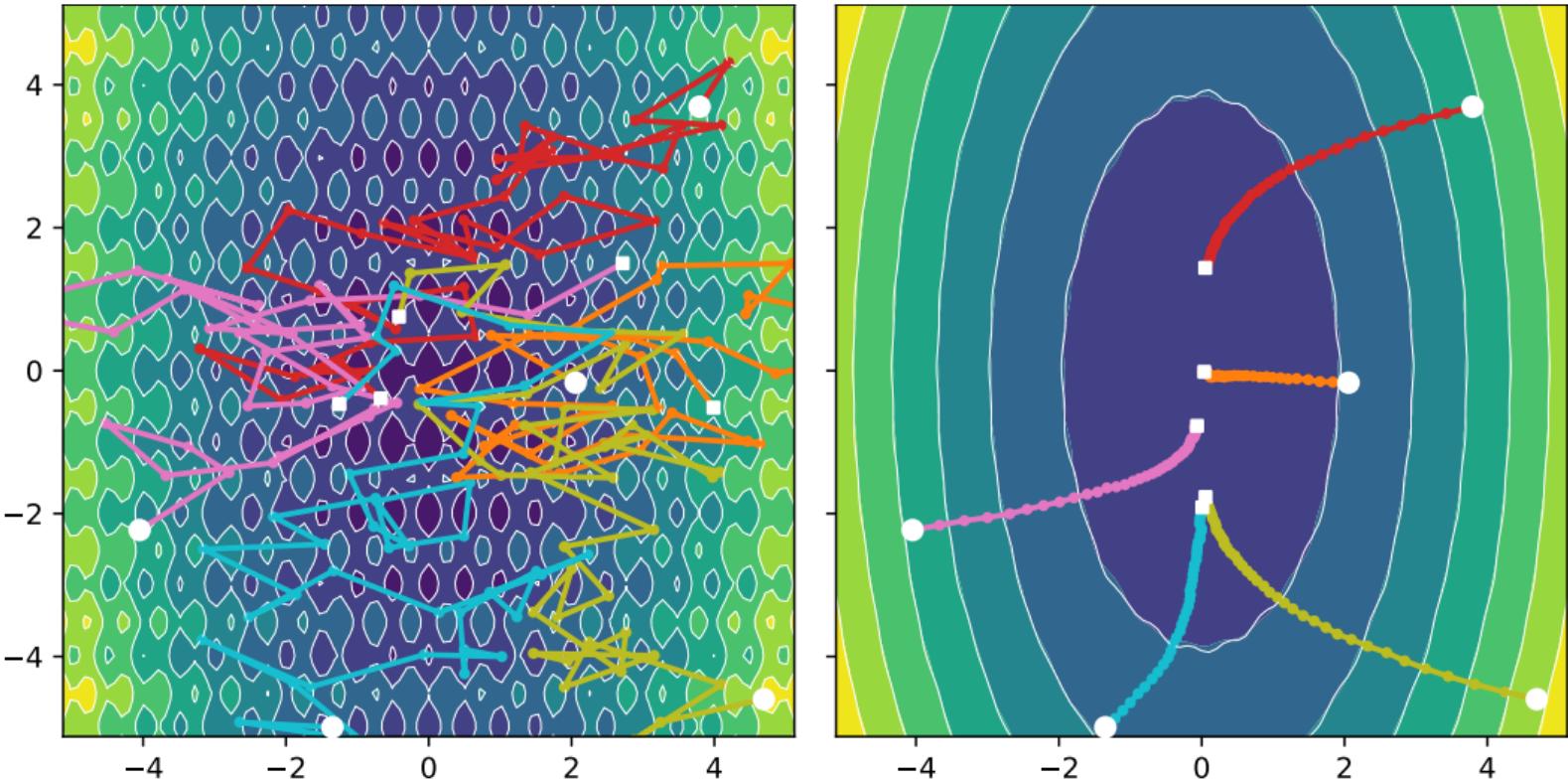
- Shows what the *exact* target is.
- Shows that errors *cancel* as $N \rightarrow \infty$.
- Enables saying *average sensitivity* without blushing.



Implicit blurring



Implicit blurring





EnKFs use LLS



Recall

EnKFs use LLS

Recall

$$\mathbf{K} := \mathbf{P} \mathbf{H}^T (\mathbf{H} \mathbf{P} \mathbf{H}^T + \mathbf{R})^{-1} \quad (10)$$



EnKFs use LLS

Recall

$$\mathbf{K} := \mathbf{P}\mathbf{H}^T(\mathbf{H}\mathbf{P}\mathbf{H}^T + \mathbf{R})^{-1} \quad (10)$$
$$\bar{\mathbf{K}} := \bar{\mathbf{C}}_{\mathbf{x}, \mathbf{y}} (\bar{\mathbf{C}}_{\mathbf{y}} + \mathbf{R})^{-1} \quad (11)$$



EnKFs use LLS

Recall

$$\mathbf{K} := \mathbf{P}\mathbf{H}^T(\mathbf{H}\mathbf{P}\mathbf{H}^T + \mathbf{R})^{-1} \quad (10)$$
$$\bar{\mathbf{K}} := \bar{\mathbf{C}}_{\mathbf{x}, \mathbf{y}} (\bar{\mathbf{C}}_{\mathbf{y}} + \mathbf{R})^{-1} \quad (11)$$
$$= \mathbf{X}\mathbf{Y}^T(\mathbf{Y}\mathbf{Y}^T + (N-1)\mathbf{R})^{-1}. \quad (12)$$



EnKFs use LLS

Recall

$$\mathbf{K} := \mathbf{P} \mathbf{H}^T (\mathbf{H} \mathbf{P} \mathbf{H}^T + \mathbf{R})^{-1} \quad (10)$$

$$\bar{\mathbf{K}} := \bar{\mathbf{C}}_{\mathbf{x}, \mathbf{y}} (\bar{\mathbf{C}}_{\mathbf{y}} + \mathbf{R})^{-1} \quad (11)$$

$$= \mathbf{X} \mathbf{Y}^T (\mathbf{Y} \mathbf{Y}^T + (N-1) \mathbf{R})^{-1}. \quad (12)$$

As in **IES**, linearise f



EnKFs use LLS

Recall

$$\mathbf{K} := \mathbf{P}\mathbf{H}^T(\mathbf{H}\mathbf{P}\mathbf{H}^T + \mathbf{R})^{-1} \quad (10)$$

$$\bar{\mathbf{K}} := \bar{\mathbf{C}}_{\mathbf{x}, \mathbf{y}} (\bar{\mathbf{C}}_{\mathbf{y}} + \mathbf{R})^{-1} \quad (11)$$

$$= \mathbf{X}\mathbf{Y}^T(\mathbf{Y}\mathbf{Y}^T + (N-1)\mathbf{R})^{-1}. \quad (12)$$

As in **IES**, linearise f , i.e. use $\begin{cases} \mathbf{P} \leftarrow \bar{\mathbf{C}}_{\mathbf{x}} = \mathbf{X}\mathbf{X}^T/(N-1) \\ \mathbf{H} \leftarrow \bar{\nabla}f = \mathbf{Y}\mathbf{X}^+ \end{cases}$ in \mathbf{K} (10).



EnKFs use LLS

Recall

$$\mathbf{K} := \mathbf{P}\mathbf{H}^T(\mathbf{H}\mathbf{P}\mathbf{H}^T + \mathbf{R})^{-1} \quad (10)$$

$$\bar{\mathbf{K}} := \bar{\mathbf{C}}_{\mathbf{x}, \mathbf{y}} (\bar{\mathbf{C}}_{\mathbf{y}} + \mathbf{R})^{-1} \quad (11)$$

$$= \mathbf{X}\mathbf{Y}^T(\mathbf{Y}\mathbf{Y}^T + (N-1)\mathbf{R})^{-1}. \quad (12)$$

As in **IES**, linearise f , i.e. use $\begin{cases} \mathbf{P} \leftarrow \bar{\mathbf{C}}_{\mathbf{x}} = \mathbf{X}\mathbf{X}^T/(N-1) \\ \mathbf{H} \leftarrow \bar{\nabla}f = \mathbf{Y}\mathbf{X}^+ \end{cases}$ in \mathbf{K} (10). Yields $\bar{\mathbf{K}}$ (12). ✓



EnKFs use LLS

Recall

$$\mathbf{K} := \mathbf{P}\mathbf{H}^T(\mathbf{H}\mathbf{P}\mathbf{H}^T + \mathbf{R})^{-1} \quad (10)$$

$$\bar{\mathbf{K}} := \bar{\mathbf{C}}_{\mathbf{x}, \mathbf{y}} (\bar{\mathbf{C}}_{\mathbf{y}} + \mathbf{R})^{-1} \quad (11)$$

$$= \mathbf{X}\mathbf{Y}^T(\mathbf{Y}\mathbf{Y}^T + (N-1)\mathbf{R})^{-1}. \quad (12)$$

As in **IES**, linearise f , i.e. use $\begin{cases} \mathbf{P} \leftarrow \bar{\mathbf{C}}_{\mathbf{x}} = \mathbf{X}\mathbf{X}^T/(N-1) \\ \mathbf{H} \leftarrow \bar{\nabla}f = \mathbf{Y}\mathbf{X}^+ \end{cases}$ in \mathbf{K} (10). Yields $\bar{\mathbf{K}}$ (12). ✓

As in **EAKF**, first update in obs-space

EnKFs use LLS

Recall

$$\mathbf{K} := \mathbf{P}\mathbf{H}^T(\mathbf{H}\mathbf{P}\mathbf{H}^T + \mathbf{R})^{-1} \quad (10)$$

$$\bar{\mathbf{K}} := \bar{\mathbf{C}}_{\mathbf{x}, \mathbf{y}} (\bar{\mathbf{C}}_{\mathbf{y}} + \mathbf{R})^{-1} \quad (11)$$

$$= \mathbf{X}\mathbf{Y}^T(\mathbf{Y}\mathbf{Y}^T + (N-1)\mathbf{R})^{-1}. \quad (12)$$

As in **IES**, linearise f , i.e. use $\begin{cases} \mathbf{P} \leftarrow \bar{\mathbf{C}}_{\mathbf{x}} = \mathbf{X}\mathbf{X}^T/(N-1) \\ \mathbf{H} \leftarrow \bar{\nabla}f = \mathbf{Y}\mathbf{X}^+ \end{cases}$ in \mathbf{K} (10). Yields $\bar{\mathbf{K}}$ (12). ✓

As in **EAKF**, first update in obs-space : $\begin{cases} \mathbf{P} \leftarrow \bar{\mathbf{C}}_{\mathbf{y}} = \mathbf{Y}\mathbf{Y}^T/(N-1) \\ \mathbf{H} \leftarrow \mathbf{I} \end{cases}$

EnKFs use LLS



Recall

$$\mathbf{K} := \mathbf{P} \mathbf{H}^T (\mathbf{H} \mathbf{P} \mathbf{H}^T + \mathbf{R})^{-1} \quad (10)$$

$$\bar{\mathbf{K}} := \bar{\mathbf{C}}_{\mathbf{x}, \mathbf{y}} (\bar{\mathbf{C}}_{\mathbf{y}} + \mathbf{R})^{-1} \quad (11)$$

$$= \mathbf{X} \mathbf{Y}^T (\mathbf{Y} \mathbf{Y}^T + (N-1) \mathbf{R})^{-1}. \quad (12)$$

As in **IES**, linearise f , i.e. use $\begin{cases} \mathbf{P} \leftarrow \bar{\mathbf{C}}_{\mathbf{x}} = \mathbf{X} \mathbf{X}^T / (N-1) \\ \mathbf{H} \leftarrow \bar{\nabla} f = \mathbf{Y} \mathbf{X}^+ \end{cases}$ in \mathbf{K} (10). Yields $\bar{\mathbf{K}}$ (12). ✓

As in **EAKF**, first update in obs-space : $\begin{cases} \mathbf{P} \leftarrow \bar{\mathbf{C}}_{\mathbf{y}} = \mathbf{Y} \mathbf{Y}^T / (N-1) \\ \mathbf{H} \leftarrow \mathbf{I} \end{cases}$, then compute
“reverse” LLS, $\bar{\nabla}^{-1} f := \mathbf{X} \mathbf{Y}^+$,

EnKFs use LLS



Recall

$$\mathbf{K} := \mathbf{P}\mathbf{H}^T(\mathbf{H}\mathbf{P}\mathbf{H}^T + \mathbf{R})^{-1} \quad (10)$$

$$\bar{\mathbf{K}} := \bar{\mathbf{C}}_{\mathbf{x}, \mathbf{y}} (\bar{\mathbf{C}}_{\mathbf{y}} + \mathbf{R})^{-1} \quad (11)$$

$$= \mathbf{X}\mathbf{Y}^T(\mathbf{Y}\mathbf{Y}^T + (N-1)\mathbf{R})^{-1}. \quad (12)$$

As in **IES**, linearise f , i.e. use $\begin{cases} \mathbf{P} \leftarrow \bar{\mathbf{C}}_{\mathbf{x}} = \mathbf{X}\mathbf{X}^T/(N-1) \\ \mathbf{H} \leftarrow \bar{\nabla}f = \mathbf{Y}\mathbf{X}^+ \end{cases}$ in \mathbf{K} (10). Yields $\bar{\mathbf{K}}$ (12). ✓

As in **EAKF**, first update in obs-space : $\begin{cases} \mathbf{P} \leftarrow \bar{\mathbf{C}}_{\mathbf{y}} = \mathbf{Y}\mathbf{Y}^T/(N-1) \\ \mathbf{H} \leftarrow \mathbf{I} \end{cases}$, then compute “reverse” LLS, $\bar{\nabla}^{-1}f := \mathbf{X}\mathbf{Y}^+$, and apply to obs-update.

EnKFs use LLS



Recall

$$\mathbf{K} := \mathbf{P}\mathbf{H}^T(\mathbf{H}\mathbf{P}\mathbf{H}^T + \mathbf{R})^{-1} \quad (10)$$

$$\bar{\mathbf{K}} := \bar{\mathbf{C}}_{\mathbf{x}, \mathbf{y}} (\bar{\mathbf{C}}_{\mathbf{y}} + \mathbf{R})^{-1} \quad (11)$$

$$= \mathbf{X}\mathbf{Y}^T(\mathbf{Y}\mathbf{Y}^T + (N-1)\mathbf{R})^{-1}. \quad (12)$$

As in **IES**, linearise f , i.e. use $\begin{cases} \mathbf{P} \leftarrow \bar{\mathbf{C}}_{\mathbf{x}} = \mathbf{X}\mathbf{X}^T/(N-1) \\ \mathbf{H} \leftarrow \bar{\nabla}f = \mathbf{Y}\mathbf{X}^+ \end{cases}$ in \mathbf{K} (10). Yields $\bar{\mathbf{K}}$ (12). ✓

As in **EAKF**, first update in obs-space : $\begin{cases} \mathbf{P} \leftarrow \bar{\mathbf{C}}_{\mathbf{y}} = \mathbf{Y}\mathbf{Y}^T/(N-1) \\ \mathbf{H} \leftarrow \mathbf{I} \end{cases}$, then compute
“reverse” LLS, $\bar{\nabla}^{-1}f := \mathbf{X}\mathbf{Y}^+$, and apply to obs-update. Yields $(\bar{\nabla}^{-1}f) \mathbf{K} = \bar{\mathbf{K}}$. ✓

EnKFs use LLS



Recall

$$\mathbf{K} := \mathbf{P}\mathbf{H}^T(\mathbf{H}\mathbf{P}\mathbf{H}^T + \mathbf{R})^{-1} \quad (10)$$

$$\bar{\mathbf{K}} := \bar{\mathbf{C}}_{\mathbf{x}, \mathbf{y}} (\bar{\mathbf{C}}_{\mathbf{y}} + \mathbf{R})^{-1} \quad (11)$$

$$= \mathbf{X}\mathbf{Y}^T(\mathbf{Y}\mathbf{Y}^T + (N-1)\mathbf{R})^{-1}. \quad (12)$$

As in **IES**, linearise f , i.e. use $\begin{cases} \mathbf{P} \leftarrow \bar{\mathbf{C}}_{\mathbf{x}} = \mathbf{X}\mathbf{X}^T/(N-1) \\ \mathbf{H} \leftarrow \bar{\nabla}f = \mathbf{Y}\mathbf{X}^+ \end{cases}$ in \mathbf{K} (10). Yields $\bar{\mathbf{K}}$ (12). ✓

As in **EAKF**, first update in obs-space : $\begin{cases} \mathbf{P} \leftarrow \bar{\mathbf{C}}_{\mathbf{y}} = \mathbf{Y}\mathbf{Y}^T/(N-1) \\ \mathbf{H} \leftarrow \mathbf{I} \end{cases}$, then compute
“reverse” LLS, $\bar{\nabla}^{-1}f := \mathbf{X}\mathbf{Y}^+$, and apply to obs-update. Yields $(\bar{\nabla}^{-1}f) \mathbf{K} = \bar{\mathbf{K}}$. ✓

As in **ETKF**, update in ens. subspace: $\overbrace{\mathbf{x} = \bar{\mathbf{x}} + \mathbf{X}\mathbf{w}}^{\phi(\mathbf{w})}$

EnKFs use LLS



Recall

$$\mathbf{K} := \mathbf{P} \mathbf{H}^T (\mathbf{H} \mathbf{P} \mathbf{H}^T + \mathbf{R})^{-1} \quad (10)$$

$$\bar{\mathbf{K}} := \bar{\mathbf{C}}_{\mathbf{x}, \mathbf{y}} (\bar{\mathbf{C}}_{\mathbf{y}} + \mathbf{R})^{-1} \quad (11)$$

$$= \mathbf{X} \mathbf{Y}^T (\mathbf{Y} \mathbf{Y}^T + (N-1) \mathbf{R})^{-1}. \quad (12)$$

As in **IES**, linearise f , i.e. use $\begin{cases} \mathbf{P} \leftarrow \bar{\mathbf{C}}_{\mathbf{x}} = \mathbf{X} \mathbf{X}^T / (N-1) \\ \mathbf{H} \leftarrow \bar{\nabla} f = \mathbf{Y} \mathbf{X}^+ \end{cases}$ in \mathbf{K} (10). Yields $\bar{\mathbf{K}}$ (12). ✓

As in **EAKF**, first update in obs-space : $\begin{cases} \mathbf{P} \leftarrow \bar{\mathbf{C}}_{\mathbf{y}} = \mathbf{Y} \mathbf{Y}^T / (N-1) \\ \mathbf{H} \leftarrow \mathbf{I} \end{cases}$, then compute “reverse” LLS, $\bar{\nabla}^{-1} f := \mathbf{X} \mathbf{Y}^+$, and apply to obs-update. Yields $(\bar{\nabla}^{-1} f) \mathbf{K} = \bar{\mathbf{K}}$. ✓

As in **ETKF**, update in ens. subspace: $\overbrace{\mathbf{x} = \bar{\mathbf{x}} + \mathbf{X} \mathbf{w}}^{\phi(\mathbf{w})}$ using $\begin{cases} \mathbf{P} \leftarrow \bar{\mathbf{C}}_{\mathbf{w}} = \mathbf{I} / (N-1) \\ \mathbf{H} \leftarrow \bar{\nabla}(f \circ \phi) = \mathbf{Y} \mathbf{I}^+ = \mathbf{Y} \end{cases}$

EnKFs use LLS



Recall

$$\mathbf{K} := \mathbf{P} \mathbf{H}^T (\mathbf{H} \mathbf{P} \mathbf{H}^T + \mathbf{R})^{-1} \quad (10)$$

$$\bar{\mathbf{K}} := \bar{\mathbf{C}}_{\mathbf{x}, \mathbf{y}} (\bar{\mathbf{C}}_{\mathbf{y}} + \mathbf{R})^{-1} \quad (11)$$

$$= \mathbf{X} \mathbf{Y}^T (\mathbf{Y} \mathbf{Y}^T + (N-1) \mathbf{R})^{-1}. \quad (12)$$

As in **IES**, linearise f , i.e. use $\begin{cases} \mathbf{P} \leftarrow \bar{\mathbf{C}}_{\mathbf{x}} = \mathbf{X} \mathbf{X}^T / (N-1) \\ \mathbf{H} \leftarrow \bar{\nabla} f = \mathbf{Y} \mathbf{X}^+ \end{cases}$ in \mathbf{K} (10). Yields $\bar{\mathbf{K}}$ (12). ✓

As in **EAKF**, first update in obs-space : $\begin{cases} \mathbf{P} \leftarrow \bar{\mathbf{C}}_{\mathbf{y}} = \mathbf{Y} \mathbf{Y}^T / (N-1) \\ \mathbf{H} \leftarrow \mathbf{I} \end{cases}$, then compute “reverse” LLS, $\bar{\nabla}^{-1} f := \mathbf{X} \mathbf{Y}^+$, and apply to obs-update. Yields $(\bar{\nabla}^{-1} f) \mathbf{K} = \bar{\mathbf{K}}$. ✓

As in **ETKF**, update in ens. subspace: $\overbrace{\mathbf{x} = \bar{\mathbf{x}} + \mathbf{X} \mathbf{w}}^{\phi(\mathbf{w})}$ using $\begin{cases} \mathbf{P} \leftarrow \bar{\mathbf{C}}_{\mathbf{w}} = \mathbf{I} / (N-1) \\ \mathbf{H} \leftarrow \bar{\nabla}(f \circ \phi) = \mathbf{Y} \mathbf{I}^+ = \mathbf{Y} \end{cases}$

Then apply ϕ to revert to state space.

EnKFs use LLS



Recall

$$\mathbf{K} := \mathbf{P} \mathbf{H}^T (\mathbf{H} \mathbf{P} \mathbf{H}^T + \mathbf{R})^{-1} \quad (10)$$

$$\bar{\mathbf{K}} := \bar{\mathbf{C}}_{\mathbf{x}, \mathbf{y}} (\bar{\mathbf{C}}_{\mathbf{y}} + \mathbf{R})^{-1} \quad (11)$$

$$= \mathbf{X} \mathbf{Y}^T (\mathbf{Y} \mathbf{Y}^T + (N-1) \mathbf{R})^{-1}. \quad (12)$$

As in **IES**, linearise f , i.e. use $\begin{cases} \mathbf{P} \leftarrow \bar{\mathbf{C}}_{\mathbf{x}} = \mathbf{X} \mathbf{X}^T / (N-1) \\ \mathbf{H} \leftarrow \bar{\nabla} f = \mathbf{Y} \mathbf{X}^+ \end{cases}$ in \mathbf{K} (10). Yields $\bar{\mathbf{K}}$ (12). ✓

As in **EAKF**, first update in obs-space : $\begin{cases} \mathbf{P} \leftarrow \bar{\mathbf{C}}_{\mathbf{y}} = \mathbf{Y} \mathbf{Y}^T / (N-1) \\ \mathbf{H} \leftarrow \mathbf{I} \end{cases}$, then compute “reverse” LLS, $\bar{\nabla}^{-1} f := \mathbf{X} \mathbf{Y}^+$, and apply to obs-update. Yields $(\bar{\nabla}^{-1} f) \mathbf{K} = \bar{\mathbf{K}}$. ✓

As in **ETKF**, update in ens. subspace: $\overbrace{\mathbf{x} = \bar{\mathbf{x}} + \mathbf{X} \mathbf{w}}^{\phi(\mathbf{w})}$ using $\begin{cases} \mathbf{P} \leftarrow \bar{\mathbf{C}}_{\mathbf{w}} = \mathbf{I} / (N-1) \\ \mathbf{H} \leftarrow \bar{\nabla}(f \circ \phi) = \mathbf{Y} \mathbf{I}^+ = \mathbf{Y} \end{cases}$
Then apply ϕ to revert to state space. Yields $(\bar{\nabla} \phi) \mathbf{K} = \bar{\mathbf{K}}$. ✓

EnKFs use LLS



Recall

$$\mathbf{K} := \mathbf{P} \mathbf{H}^T (\mathbf{H} \mathbf{P} \mathbf{H}^T + \mathbf{R})^{-1} \quad (10)$$

$$\bar{\mathbf{K}} := \bar{\mathbf{C}}_{\mathbf{x}, \mathbf{y}} (\bar{\mathbf{C}}_{\mathbf{y}} + \mathbf{R})^{-1} \quad (11)$$

$$= \mathbf{X} \mathbf{Y}^T (\mathbf{Y} \mathbf{Y}^T + (N-1) \mathbf{R})^{-1}. \quad (12)$$

As in **IES**, linearise f , i.e. use $\begin{cases} \mathbf{P} \leftarrow \bar{\mathbf{C}}_{\mathbf{x}} = \mathbf{X} \mathbf{X}^T / (N-1) \\ \mathbf{H} \leftarrow \bar{\nabla} f = \mathbf{Y} \mathbf{X}^+ \end{cases}$ in \mathbf{K} (10). Yields $\bar{\mathbf{K}}$ (12). ✓

As in **EAKF**, first update in obs-space : $\begin{cases} \mathbf{P} \leftarrow \bar{\mathbf{C}}_{\mathbf{y}} = \mathbf{Y} \mathbf{Y}^T / (N-1) \\ \mathbf{H} \leftarrow \mathbf{I} \end{cases}$, then compute
“reverse” LLS, $\bar{\nabla}^{-1} f := \mathbf{X} \mathbf{Y}^+$, and apply to obs-update. Yields $(\bar{\nabla}^{-1} f) \mathbf{K} = \bar{\mathbf{K}}$. ✓

As in **ETKF**, update in ens. subspace: $\overbrace{\mathbf{x} = \bar{\mathbf{x}} + \mathbf{X} \mathbf{w}}^{\phi(\mathbf{w})}$ using $\begin{cases} \mathbf{P} \leftarrow \bar{\mathbf{C}}_{\mathbf{w}} = \mathbf{I} / (N-1) \\ \mathbf{H} \leftarrow \bar{\nabla}(f \circ \phi) = \mathbf{Y} \mathbf{I}^+ = \mathbf{Y} \end{cases}$
Then apply ϕ to revert to state space. Yields $(\bar{\nabla} \phi) \mathbf{K} = \bar{\mathbf{K}}$. ✓

Could we have anticipated the equivalence?

EnKFs use LLS



Recall

$$\mathbf{K} := \mathbf{P} \mathbf{H}^T (\mathbf{H} \mathbf{P} \mathbf{H}^T + \mathbf{R})^{-1} \quad (10)$$

$$\bar{\mathbf{K}} := \bar{\mathbf{C}}_{\mathbf{x}, \mathbf{y}} (\bar{\mathbf{C}}_{\mathbf{y}} + \mathbf{R})^{-1} \quad (11)$$

$$= \mathbf{X} \mathbf{Y}^T (\mathbf{Y} \mathbf{Y}^T + (N-1) \mathbf{R})^{-1}. \quad (12)$$

As in **IES**, linearise f , i.e. use $\begin{cases} \mathbf{P} \leftarrow \bar{\mathbf{C}}_{\mathbf{x}} = \mathbf{X} \mathbf{X}^T / (N-1) \\ \mathbf{H} \leftarrow \bar{\nabla} f = \mathbf{Y} \mathbf{X}^+ \end{cases}$ in \mathbf{K} (10). Yields $\bar{\mathbf{K}}$ (12). ✓

As in **EAKF**, first update in obs-space : $\begin{cases} \mathbf{P} \leftarrow \bar{\mathbf{C}}_{\mathbf{y}} = \mathbf{Y} \mathbf{Y}^T / (N-1) \\ \mathbf{H} \leftarrow \mathbf{I} \end{cases}$, then compute “reverse” LLS, $\bar{\nabla}^{-1} f := \mathbf{X} \mathbf{Y}^+$, and apply to obs-update. Yields $(\bar{\nabla}^{-1} f) \mathbf{K} = \bar{\mathbf{K}}$. ✓

As in **ETKF**, update in ens. subspace: $\overbrace{\mathbf{x} = \bar{\mathbf{x}} + \mathbf{X} \mathbf{w}}^{\phi(\mathbf{w})}$ using $\begin{cases} \mathbf{P} \leftarrow \bar{\mathbf{C}}_{\mathbf{w}} = \mathbf{I} / (N-1) \\ \mathbf{H} \leftarrow \bar{\nabla}(f \circ \phi) = \mathbf{Y} \mathbf{I}^+ = \mathbf{Y} \end{cases}$
Then apply ϕ to revert to state space. Yields $(\bar{\nabla} \phi) \mathbf{K} = \bar{\mathbf{K}}$. ✓

Could we have anticipated the equivalence?

Yes, by **chain rule** (Raanes et al., 2019), i.e. $\bar{\nabla}(f \circ g) = \bar{\nabla}f \bar{\nabla}g$



EnKF is LLS

What about $\bar{\mathbf{K}}$ itself?

Then:

$$\frac{1}{\sigma^2} \mathbf{DD}^T = \mathbf{B}$$

yields

$$\bar{\mathbf{K}} = \mathbf{XY}^T (\mathbf{YY}^T + (\mathbf{B}-\mathbf{B}))^{-1} \quad (3)$$

$$\mathbf{Y}_\text{noisy} - \mathbf{Y} = \mathbf{v} \quad \mathbf{v}\mathbf{v}^T = 0$$

$$\approx \mathbf{XY}^T (\mathbf{YY}^T + \mathbf{DD}^T)^{-1} \quad (4)$$

$$\mathbf{v}\mathbf{v}^T = 0 \quad \text{yields}$$

$$\approx \mathbf{XY}_{\text{noisy}}^T (\mathbf{Y}_{\text{noisy}}\mathbf{Y}_{\text{noisy}}^T)^{-1} \quad (5)$$

$$\mathbf{A}^T (\mathbf{AA}^T)^{-1} = \mathbf{A}^T \quad \text{yields}$$

$$\approx \mathbf{XY}_{\text{noisy}}^T (\mathbf{Y}_{\text{noisy}}\mathbf{Y}_{\text{noisy}}^T)^{-1} \quad (6)$$

$$\mathbf{A}^T \mathbf{A} = \mathbf{I} \quad \text{yields}$$

$$= \mathbf{XY}_{\text{noisy}}^T \quad (7)$$

$$= \hat{\nabla}^{-1} f_{\text{noisy}} \quad (8)$$

(Snyder, 2015)

→ $\bar{\mathbf{K}} = \hat{\nabla}^{-1} f_{\text{noisy}}$ “inverts” model + noise.

⇒ Can estimate $(\mathbf{x} - \hat{\mathbf{x}})$ by applying it to $(\mathbf{y} - \hat{\mathbf{y}})$, i.e.

$$\hat{\mathbf{x}}(\mathbf{y}) = \hat{\mathbf{x}} + \bar{\mathbf{K}}(\mathbf{y} - \hat{\mathbf{y}}) \quad (9)$$



EnKF is LLS

What about $\bar{\mathbf{K}}$ itself?

Recall $\bar{\mathbf{K}} := \mathbf{X}\mathbf{Y}^T(\mathbf{Y}\mathbf{Y}^T + (N-1)\mathbf{R})^{-1}$ (13)

Then: $\frac{1}{N-1}\mathbf{DD}^T = \mathbf{R}$ yields $\approx \mathbf{XY}^T(\mathbf{YY}^T + \mathbf{DD}^T)^+$ (4)

$\mathbf{Y}\mathbf{Y}^T = \mathbf{I} - \mathbf{X}\mathbf{X}^T \Rightarrow \mathbf{Y}\mathbf{Y}^T \approx \mathbf{I}$ yields $\approx \mathbf{XY}^T(\mathbf{Y}_{\text{noisy}}\mathbf{Y}_{\text{noisy}}^T)^+$ (5)

$\mathbf{X}\mathbf{X}^T \approx \mathbf{0}$ yields $\approx \mathbf{XY}_{\text{noisy}}^T(\mathbf{Y}_{\text{noisy}}\mathbf{Y}_{\text{noisy}}^T)^+$ (6)

$\mathbf{A}^T(\mathbf{AA}^T)^+ = \mathbf{A}^T$ yields $= \mathbf{XY}_{\text{noisy}}^+$ (7)

$\mathbf{Y}_{\text{noisy}}^T \mathbf{Y}_{\text{noisy}} = \mathbf{I}$ yields $= \hat{\nabla}^{-1}f_{\text{noisy}}$ (8)

(Snyder, 2015)

→ $\mathbf{K} = \hat{\nabla}^{-1}f_{\text{noisy}}$ “inverts” model + noise.

→ Can estimate $(x - \hat{x})$ by applying it to $(y - \hat{y})$, i.e.

$$\hat{x}(y) = \hat{x} + \hat{\mathbf{K}}(y - \hat{y}) \quad (19)$$



EnKF is LLS

What about $\bar{\mathbf{K}}$ itself?

Recall $\bar{\mathbf{K}} := \mathbf{X}\mathbf{Y}^T(\mathbf{Y}\mathbf{Y}^T + (N-1)\mathbf{R})^{-1}$ (13)

Then: $\frac{1}{N-1}\mathbf{D}\mathbf{D}^T \approx \mathbf{R}$ yields $\approx \mathbf{X}\mathbf{Y}^T(\mathbf{Y}\mathbf{Y}^T + \mathbf{D}\mathbf{D}^T)^{-1}$ (4)

$\mathbf{Y}\mathbf{Y}^T = \mathbf{Y}\mathbf{Y}^T_{\text{noisy}}$ $\mathbf{Y}\mathbf{Y}^T \approx 0$ yields $\approx \mathbf{X}\mathbf{Y}^T_{\text{noisy}}(\mathbf{Y}\mathbf{Y}^T_{\text{noisy}})^{-1}$ (5)

$\mathbf{D}\mathbf{D}^T \approx 0$ yields $\approx \mathbf{X}\mathbf{Y}^T_{\text{noisy}}(\mathbf{Y}\mathbf{Y}^T_{\text{noisy}})^{-1}$ (6)

$\mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1} = \mathbf{A}^T$ yields $= \mathbf{X}\mathbf{Y}^T_{\text{noisy}}$ (7)

$\mathbf{A}\mathbf{A}^T \rightarrow \mathbf{A}^T\mathbf{A}$ yields $= \hat{\nabla}^{-1}f_{\text{noisy}}$ (8)

(Snyder, 2015)

→ $\mathbf{K} = \hat{\nabla}^{-1}f_{\text{noisy}}$ “inverts” model + noise.

→ Can estimate $(\mathbf{x} - \hat{\mathbf{x}})$ by applying it to $(y - \hat{y})$, i.e.

$$\hat{x}(y) = \hat{x} + \hat{\mathbf{K}}(y - \hat{y}) \quad (19)$$



EnKF is LLS

What about $\bar{\mathbf{K}}$ itself?

Recall

$$\bar{\mathbf{K}} := \mathbf{X}\mathbf{Y}^T(\mathbf{Y}\mathbf{Y}^T + (N-1)\mathbf{R})^{-1} \quad (13)$$

Then: $\frac{1}{N-1}\mathbf{D}\mathbf{D}^T \approx \mathbf{R}$ yields

$$\approx \mathbf{X}\mathbf{Y}^T(\mathbf{Y}\mathbf{Y}^T + \mathbf{D}\mathbf{D}^T)^+ \quad (14)$$

$$\mathbf{X}\mathbf{D}^T = \mathbf{A}\mathbf{X}^T \quad \text{yields}$$

$$\approx \mathbf{X}\mathbf{X}^T(\mathbf{X}\text{noisy} - \mathbf{Y}\text{noisy})^+ \quad (15)$$

$$\mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^+ = \mathbf{A}^T \quad \text{yields}$$

$$\approx \mathbf{X}\mathbf{Y}_{\text{noisy}}^T(\mathbf{Y}_{\text{noisy}}\mathbf{Y}_{\text{noisy}}^T)^+ \quad (16)$$

$$\mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^+ = \mathbf{A}^T \quad \text{yields}$$

$$= \mathbf{X}\mathbf{Y}_{\text{noisy}}^+ \quad (17)$$

$$\mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^+ = \mathbf{A}^T \quad \text{yields}$$

$$= \hat{\nabla}^{-1}f_{\text{noisy}} \quad (18)$$

(Snyder, 2015)

→ $\mathbf{K} = \hat{\nabla}^{-1}f_{\text{noisy}}$ “inverts” model + noise.

→ Can estimate $(\mathbf{x} - \hat{\mathbf{x}})$ by applying it to $(\mathbf{y} - \hat{\mathbf{y}})$, i.e.

$$\hat{\mathbf{x}}(\mathbf{y}) = \hat{\mathbf{x}} + \mathbf{K}(\mathbf{y} - \hat{\mathbf{y}}) \quad (19)$$



EnKF is LLS

What about $\bar{\mathbf{K}}$ itself?

Recall

$$\bar{\mathbf{K}} := \mathbf{X}\mathbf{Y}^T(\mathbf{Y}\mathbf{Y}^T + (N-1)\mathbf{R})^{-1} \quad (13)$$

Then: $\frac{1}{N-1}\mathbf{D}\mathbf{D}^T \approx \mathbf{R}$ yields

$$\approx \mathbf{X}\mathbf{Y}^T(\mathbf{Y}\mathbf{Y}^T + \mathbf{D}\mathbf{D}^T)^+ \quad (14)$$

$\mathbf{Y}_{\text{noisy}} := \mathbf{Y} + \mathbf{D}$, $\mathbf{Y}\mathbf{Y}^T \approx 0$ yields

$$\approx \mathbf{X}\mathbf{Y}_{\text{noisy}}^T(\mathbf{Y}_{\text{noisy}}\mathbf{Y}_{\text{noisy}}^T)^+ \quad (15)$$

$\mathbf{Y}\mathbf{Y}^T \approx 0$ yields

$$\approx \mathbf{X}\mathbf{Y}_{\text{noisy}}^T(\mathbf{Y}_{\text{noisy}}\mathbf{Y}_{\text{noisy}}^T)^+ \quad (16)$$

$\mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^+ = \mathbf{A}^T$ yields

$$= \mathbf{X}\mathbf{Y}_{\text{noisy}}^+ \quad (17)$$

$\mathbf{f}_{\text{true}}(\mathbf{x}) = \mathbf{f}_{\text{obs}}(\mathbf{x})$ yields

$$= \hat{\nabla}^{-1}\mathbf{f}_{\text{noisy}} \quad (18)$$

(Snyder, 2015)

→ $\mathbf{K} = \hat{\nabla}^{-1}\mathbf{f}_{\text{noisy}}$ “inverts” model + noise.

→ Can estimate $(\mathbf{x} - \hat{\mathbf{x}})$ by applying it to $(y - \tilde{y})$, i.e.

$$\hat{\mathbf{x}}(y) = \hat{\mathbf{x}} + \hat{\mathbf{K}}(y - \tilde{y}) \quad (19)$$



EnKF is LLS

What about $\bar{\mathbf{K}}$ itself?

Recall

$$\bar{\mathbf{K}} := \mathbf{X}\mathbf{Y}^T(\mathbf{Y}\mathbf{Y}^T + (N-1)\mathbf{R})^{-1} \quad (13)$$

Then: $\frac{1}{N-1}\mathbf{D}\mathbf{D}^T \approx \mathbf{R}$ yields

$$\approx \mathbf{X}\mathbf{Y}^T(\mathbf{Y}\mathbf{Y}^T + \mathbf{D}\mathbf{D}^T)^+ \quad (14)$$

$\mathbf{Y}_{\text{noisy}} := \mathbf{Y} + \mathbf{D}$, $\mathbf{Y}\mathbf{D}^T \approx \mathbf{0}$ yields

$$\approx \mathbf{X}\mathbf{Y}_{\text{noisy}}^T(\mathbf{Y}_{\text{noisy}}\mathbf{Y}_{\text{noisy}}^T)^+ \quad (15)$$

$\mathbf{Y}\mathbf{Y}^T \approx \mathbf{0}$ yields

$$\approx \mathbf{X}\mathbf{Y}_{\text{noisy}}^T(\mathbf{Y}_{\text{noisy}}\mathbf{Y}_{\text{noisy}}^T)^+ \quad (16)$$

$\mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^+ = \mathbf{A}^T$ yields

$$= \mathbf{X}\mathbf{Y}_{\text{noisy}}^+ \quad (17)$$

$\mathbf{f}_{\text{true}}(\mathbf{x}) = \mathbf{f}_{\text{obs}} + \mathbf{e}$ yields

$$= \hat{\nabla}^{-1}\mathbf{f}_{\text{noisy}} + \quad (18)$$

(Snyder, 2015)

→ $\mathbf{K} = \hat{\nabla}^{-1}\mathbf{f}_{\text{noisy}}$ “inverts” model + noise.

→ Can estimate $(\mathbf{x} - \hat{\mathbf{x}})$ by applying it to $(\mathbf{y} - \hat{\mathbf{y}})$, i.e.

$$\hat{\mathbf{x}}(\mathbf{y}) = \hat{\mathbf{x}} + \mathbf{K}(\mathbf{y} - \hat{\mathbf{y}}) \quad (19)$$



EnKF is LLS

What about $\bar{\mathbf{K}}$ itself?

Recall

$$\bar{\mathbf{K}} := \mathbf{X}\mathbf{Y}^T(\mathbf{Y}\mathbf{Y}^T + (N-1)\mathbf{R})^{-1} \quad (13)$$

Then: $\frac{1}{N-1}\mathbf{D}\mathbf{D}^T \approx \mathbf{R}$ yields

$$\approx \mathbf{X}\mathbf{Y}^T(\mathbf{Y}\mathbf{Y}^T + \mathbf{D}\mathbf{D}^T)^+ \quad (14)$$

$\mathbf{Y}_{\text{noisy}} := \mathbf{Y} + \mathbf{D}$, $\mathbf{Y}\mathbf{D}^T \approx \mathbf{0}$ yields

$$\approx \mathbf{X}\mathbf{Y}^T(\mathbf{Y}_{\text{noisy}}\mathbf{Y}_{\text{noisy}}^T)^+ \quad (15)$$

$\mathbf{Y}\mathbf{D}^T \approx \mathbf{0}$ yields

$$\approx \mathbf{X}\mathbf{Y}^T(\mathbf{Y}_{\text{noisy}}\mathbf{Y}_{\text{noisy}}^T)^+ \quad (16)$$

$\mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^+ = \mathbf{A}^T$ yields

$$= \mathbf{X}\mathbf{Y}_{\text{noisy}}^T \quad (17)$$

$\mathbf{Y}_{\text{noisy}}\mathbf{D}^T = \mathbf{Y}\mathbf{D}^T$ yields

$$= \hat{\nabla}^{-1}f_{\text{noisy}} \quad (18)$$

(Snyder, 2015)

→ $\mathbf{K} = \hat{\nabla}^{-1}f_{\text{noisy}}$ “inverts” model + noise.

→ Can estimate $(\mathbf{x} - \hat{\mathbf{x}})$ by applying it to $(y - \hat{y})$, i.e.

$$\hat{\mathbf{x}}(y) = \hat{\mathbf{x}} + \mathbf{K}(y - \hat{y}) \quad (19)$$



EnKF is LLS

What about $\bar{\mathbf{K}}$ itself?

Recall

$$\bar{\mathbf{K}} := \mathbf{X}\mathbf{Y}^T(\mathbf{Y}\mathbf{Y}^T + (N-1)\mathbf{R})^{-1} \quad (13)$$

Then: $\frac{1}{N-1}\mathbf{D}\mathbf{D}^T \approx \mathbf{R}$ yields

$$\approx \mathbf{X}\mathbf{Y}^T(\mathbf{Y}\mathbf{Y}^T + \mathbf{D}\mathbf{D}^T)^+ \quad (14)$$

$\mathbf{Y}_{\text{noisy}} := \mathbf{Y} + \mathbf{D}$, $\mathbf{Y}\mathbf{D}^T \approx \mathbf{0}$ yields

$$\approx \mathbf{X}\mathbf{Y}^T(\mathbf{Y}_{\text{noisy}}\mathbf{Y}_{\text{noisy}}^T)^+ \quad (15)$$

$\mathbf{X}\mathbf{D}^T \approx \mathbf{0}$ yields

$$\approx \mathbf{X}\mathbf{Y}^T(\mathbf{Y}\mathbf{Y}^T)^+ \quad (16)$$

$\mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^+ = \mathbf{A}^T$ yields

$$= \mathbf{X}\mathbf{Y}_{\text{noisy}}^T \quad (17)$$

$\mathbf{f}_{\text{true}}(\mathbf{x}) = \mathbf{f}_{\text{obs}} + \mathbf{e}$ yields

$$= \hat{\nabla}^{-1}\mathbf{f}_{\text{noisy}} + \quad (18)$$

(Snyder, 2015)

→ $\mathbf{K} = \hat{\nabla}^{-1}\mathbf{f}_{\text{noisy}}$ “inverts” model + noise.

→ Can estimate $(\mathbf{x} - \hat{\mathbf{x}})$ by applying it to $(\mathbf{y} - \hat{\mathbf{y}})$, i.e.

$$\hat{\mathbf{x}}(\mathbf{y}) = \hat{\mathbf{x}} + \mathbf{K}(\mathbf{y} - \hat{\mathbf{y}}) \quad (19)$$



EnKF is LLS

What about $\bar{\mathbf{K}}$ itself?

Recall

$$\bar{\mathbf{K}} := \mathbf{X}\mathbf{Y}^T(\mathbf{Y}\mathbf{Y}^T + (N-1)\mathbf{R})^{-1} \quad (13)$$

Then: $\frac{1}{N-1}\mathbf{D}\mathbf{D}^T \approx \mathbf{R}$ yields

$$\approx \mathbf{X}\mathbf{Y}^T(\mathbf{Y}\mathbf{Y}^T + \mathbf{D}\mathbf{D}^T)^+ \quad (14)$$

$\mathbf{Y}_{\text{noisy}} := \mathbf{Y} + \mathbf{D}$, $\mathbf{Y}\mathbf{D}^T \approx \mathbf{0}$ yields

$$\approx \mathbf{X}\mathbf{Y}^T(\mathbf{Y}_{\text{noisy}}\mathbf{Y}_{\text{noisy}}^T)^+ \quad (15)$$

$\mathbf{X}\mathbf{D}^T \approx \mathbf{0}$ yields

$$\approx \mathbf{X}\mathbf{Y}_{\text{noisy}}^T(\mathbf{Y}_{\text{noisy}}\mathbf{Y}_{\text{noisy}}^T)^+ \quad (16)$$

$\mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^+ = \mathbf{A}^T$ yields

$$= \mathbf{X}\mathbf{Y}_{\text{noisy}}^T \quad (17)$$

$\mathbf{J}_{\text{noisy}}(\mathbf{x}) = \mathbf{f} + \mathbf{v}$ yields

$$= \hat{\nabla}^{-1}\mathbf{f}_{\text{noisy}} + \quad (18)$$

(Snyder, 2015)

→ $\mathbf{K} = \hat{\nabla}^{-1}\mathbf{f}_{\text{noisy}}$ “inverts” model + noise.

→ Can estimate $(\mathbf{x} - \hat{\mathbf{x}})$ by applying it to $(\mathbf{y} - \hat{\mathbf{y}})$, i.e.

$$\hat{\mathbf{x}}(\mathbf{y}) = \hat{\mathbf{x}} + \bar{\mathbf{K}}(\mathbf{y} - \hat{\mathbf{y}}) \quad (19)$$



EnKF is LLS

What about $\bar{\mathbf{K}}$ itself?

Recall

$$\bar{\mathbf{K}} := \mathbf{X}\mathbf{Y}^T(\mathbf{Y}\mathbf{Y}^T + (N-1)\mathbf{R})^{-1} \quad (13)$$

Then: $\frac{1}{N-1}\mathbf{D}\mathbf{D}^T \approx \mathbf{R}$ yields

$$\approx \mathbf{X}\mathbf{Y}^T(\mathbf{Y}\mathbf{Y}^T + \mathbf{D}\mathbf{D}^T)^+ \quad (14)$$

$\mathbf{Y}_{\text{noisy}} := \mathbf{Y} + \mathbf{D}$, $\mathbf{Y}\mathbf{D}^T \approx \mathbf{0}$ yields

$$\approx \mathbf{X}\mathbf{Y}^T(\mathbf{Y}_{\text{noisy}}\mathbf{Y}_{\text{noisy}}^T)^+ \quad (15)$$

$\mathbf{X}\mathbf{D}^T \approx \mathbf{0}$ yields

$$\approx \mathbf{X}\mathbf{Y}_{\text{noisy}}^T(\mathbf{Y}_{\text{noisy}}\mathbf{Y}_{\text{noisy}}^T)^+ \quad (16)$$

$\mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^+ = \mathbf{A}^+$ yields

$$= \mathbf{X}\mathbf{Y}_{\text{noisy}}^T \quad (17)$$

$\mathbf{Y}_{\text{noisy}} = \mathbf{Y} + \mathbf{D}$ yields

$$= \hat{\nabla}^{-1}\mathbf{f}_{\text{noisy}} \quad (18)$$

(Snyder, 2015)

→ $\mathbf{K} = \hat{\nabla}^{-1}\mathbf{f}_{\text{noisy}}$ “inverts” model + noise.

→ Can estimate $(\mathbf{x} - \hat{\mathbf{x}})$ by applying it to $(\mathbf{y} - \hat{\mathbf{y}})$, i.e.

$$\hat{\mathbf{x}}(\mathbf{y}) = \hat{\mathbf{x}} + \mathbf{K}(\mathbf{y} - \hat{\mathbf{y}}) \quad (19)$$



EnKF is LLS

What about $\bar{\mathbf{K}}$ itself?

Recall

$$\bar{\mathbf{K}} := \mathbf{X}\mathbf{Y}^T(\mathbf{Y}\mathbf{Y}^T + (N-1)\mathbf{R})^{-1} \quad (13)$$

Then: $\frac{1}{N-1}\mathbf{D}\mathbf{D}^T \approx \mathbf{R}$ yields

$$\approx \mathbf{X}\mathbf{Y}^T(\mathbf{Y}\mathbf{Y}^T + \mathbf{D}\mathbf{D}^T)^+ \quad (14)$$

$\mathbf{Y}_{\text{noisy}} := \mathbf{Y} + \mathbf{D}$, $\mathbf{Y}\mathbf{D}^T \approx \mathbf{0}$ yields

$$\approx \mathbf{X}\mathbf{Y}^T(\mathbf{Y}_{\text{noisy}}\mathbf{Y}_{\text{noisy}}^T)^+ \quad (15)$$

$\mathbf{X}\mathbf{D}^T \approx \mathbf{0}$ yields

$$\approx \mathbf{X}\mathbf{Y}_{\text{noisy}}^T(\mathbf{Y}_{\text{noisy}}\mathbf{Y}_{\text{noisy}}^T)^+ \quad (16)$$

$\mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^+ = \mathbf{A}^+$ yields

$$= \mathbf{X}\mathbf{Y}_{\text{noisy}}^+ \quad (17)$$

$\mathbf{Y}_{\text{noisy}} = \mathbf{Y} + \mathbf{D}$ yields

$$\Rightarrow \bar{\mathbf{K}} = \mathbf{Y}_{\text{noisy}}^{-1} \quad (18)$$

(Snyder, 2015)

→ $\bar{\mathbf{K}} = \mathbf{Y}_{\text{noisy}}^{-1}$ "inverts" model + noise.

→ Can estimate $(\mathbf{x} - \hat{\mathbf{x}})$ by applying it to $(y - \tilde{y})$, i.e.

$$\hat{\mathbf{x}}(y) = \hat{\mathbf{x}} + \bar{\mathbf{K}}(y - \tilde{y}) \quad (19)$$



EnKF is LLS

What about $\bar{\mathbf{K}}$ itself?

Recall

$$\bar{\mathbf{K}} := \mathbf{X}\mathbf{Y}^T(\mathbf{Y}\mathbf{Y}^T + (N-1)\mathbf{R})^{-1} \quad (13)$$

Then: $\frac{1}{N-1}\mathbf{D}\mathbf{D}^T \approx \mathbf{R}$ yields

$$\approx \mathbf{X}\mathbf{Y}^T(\mathbf{Y}\mathbf{Y}^T + \mathbf{D}\mathbf{D}^T)^+ \quad (14)$$

$\mathbf{Y}_{\text{noisy}} := \mathbf{Y} + \mathbf{D}$, $\mathbf{Y}\mathbf{D}^T \approx \mathbf{0}$ yields

$$\approx \mathbf{X}\mathbf{Y}^T(\mathbf{Y}_{\text{noisy}}\mathbf{Y}_{\text{noisy}}^T)^+ \quad (15)$$

$\mathbf{X}\mathbf{D}^T \approx \mathbf{0}$ yields

$$\approx \mathbf{X}\mathbf{Y}_{\text{noisy}}^T(\mathbf{Y}_{\text{noisy}}\mathbf{Y}_{\text{noisy}}^T)^+ \quad (16)$$

$\mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^+ = \mathbf{A}^+$ yields

$$= \mathbf{X}\mathbf{Y}_{\text{noisy}}^+ \quad (17)$$

$f_{\text{noisy}}(\mathbf{x}) := f(\mathbf{x}) + \varepsilon$ yields

$$\Rightarrow \hat{\mathbf{Y}} = \nabla^{-1}f_{\text{noisy}} \quad (18)$$

(Snyder, 2015)

→ $\mathbf{K} = \nabla^{-1}f_{\text{noisy}}$ “inverts” model + noise.

→ Can estimate $(\mathbf{x} - \hat{\mathbf{x}})$ by applying it to $(\mathbf{y} - \hat{\mathbf{y}})$, i.e.

$$\hat{\mathbf{x}}(\mathbf{y}) = \hat{\mathbf{x}} + \mathbf{K}(\mathbf{y} - \hat{\mathbf{y}}) \quad (19)$$



EnKF is LLS

What about $\bar{\mathbf{K}}$ itself?

Recall

$$\bar{\mathbf{K}} := \mathbf{X}\mathbf{Y}^T(\mathbf{Y}\mathbf{Y}^T + (N-1)\mathbf{R})^{-1} \quad (13)$$

Then: $\frac{1}{N-1}\mathbf{D}\mathbf{D}^T \approx \mathbf{R}$ yields

$$\approx \mathbf{X}\mathbf{Y}^T(\mathbf{Y}\mathbf{Y}^T + \mathbf{D}\mathbf{D}^T)^+ \quad (14)$$

$\mathbf{Y}_{\text{noisy}} := \mathbf{Y} + \mathbf{D}$, $\mathbf{Y}\mathbf{D}^T \approx \mathbf{0}$ yields

$$\approx \mathbf{X}\mathbf{Y}^T(\mathbf{Y}_{\text{noisy}}\mathbf{Y}_{\text{noisy}}^T)^+ \quad (15)$$

$\mathbf{X}\mathbf{D}^T \approx \mathbf{0}$ yields

$$\approx \mathbf{X}\mathbf{Y}_{\text{noisy}}^T(\mathbf{Y}_{\text{noisy}}\mathbf{Y}_{\text{noisy}}^T)^+ \quad (16)$$

$\mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^+ = \mathbf{A}^+$ yields

$$= \mathbf{X}\mathbf{Y}_{\text{noisy}}^+ \quad (17)$$

$f_{\text{noisy}}(\mathbf{x}) := f(\mathbf{x}) + \varepsilon$ yields

$$=: \bar{\nabla}^{-1}f_{\text{noisy}}. \quad (18)$$

(Snyder, 2015)

→ $\mathbf{K} = \bar{\nabla}^{-1}f_{\text{noisy}}$ “inverts” model + noise.

→ Can estimate $(\mathbf{x} - \hat{\mathbf{x}})$ by applying it to $(\mathbf{y} - \tilde{\mathbf{y}})$, i.e.

$$\hat{\mathbf{x}}(\mathbf{y}) = \hat{\mathbf{x}} + \bar{\mathbf{K}}(\mathbf{y} - \tilde{\mathbf{y}}) \quad (19)$$



EnKF is LLS

What about $\bar{\mathbf{K}}$ itself?

Recall

$$\bar{\mathbf{K}} := \mathbf{X}\mathbf{Y}^T(\mathbf{Y}\mathbf{Y}^T + (N-1)\mathbf{R})^{-1} \quad (13)$$

Then: $\frac{1}{N-1}\mathbf{D}\mathbf{D}^T \approx \mathbf{R}$ yields

$$\approx \mathbf{X}\mathbf{Y}^T(\mathbf{Y}\mathbf{Y}^T + \mathbf{D}\mathbf{D}^T)^+ \quad (14)$$

$\mathbf{Y}_{\text{noisy}} := \mathbf{Y} + \mathbf{D}$, $\mathbf{Y}\mathbf{D}^T \approx \mathbf{0}$ yields

$$\approx \mathbf{X}\mathbf{Y}^T(\mathbf{Y}_{\text{noisy}}\mathbf{Y}_{\text{noisy}}^T)^+ \quad (15)$$

$\mathbf{X}\mathbf{D}^T \approx \mathbf{0}$ yields

$$\approx \mathbf{X}\mathbf{Y}_{\text{noisy}}^T(\mathbf{Y}_{\text{noisy}}\mathbf{Y}_{\text{noisy}}^T)^+ \quad (16)$$

$\mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^+ = \mathbf{A}^+$ yields

$$= \mathbf{X}\mathbf{Y}_{\text{noisy}}^+ \quad (17)$$

$f_{\text{noisy}}(\mathbf{x}) := f(\mathbf{x}) + \varepsilon$ yields

$$=: \bar{\nabla}^{-1}f_{\text{noisy}}. \quad (18)$$

(Snyder, 2015)

→ $\bar{\mathbf{K}} = \bar{\nabla}^{-1}f_{\text{noisy}}$ “inverts” model + noise.

→ Can estimate $(\mathbf{x} - \hat{\mathbf{x}})$ by applying it to $(y - \tilde{y})$, i.e.

$$\hat{\mathbf{x}}(y) = \hat{\mathbf{x}} + \bar{\mathbf{K}}(y - \tilde{y}) \quad (19)$$



EnKF is LLS

What about $\bar{\mathbf{K}}$ itself?

Recall

$$\bar{\mathbf{K}} := \mathbf{X}\mathbf{Y}^T(\mathbf{Y}\mathbf{Y}^T + (N-1)\mathbf{R})^{-1} \quad (13)$$

Then: $\frac{1}{N-1}\mathbf{D}\mathbf{D}^T \approx \mathbf{R}$ yields

$$\approx \mathbf{X}\mathbf{Y}^T(\mathbf{Y}\mathbf{Y}^T + \mathbf{D}\mathbf{D}^T)^+ \quad (14)$$

$\mathbf{Y}_{\text{noisy}} := \mathbf{Y} + \mathbf{D}$, $\mathbf{Y}\mathbf{D}^T \approx \mathbf{0}$ yields

$$\approx \mathbf{X}\mathbf{Y}^T(\mathbf{Y}_{\text{noisy}}\mathbf{Y}_{\text{noisy}}^T)^+ \quad (15)$$

$\mathbf{X}\mathbf{D}^T \approx \mathbf{0}$ yields

$$\approx \mathbf{X}\mathbf{Y}_{\text{noisy}}^T(\mathbf{Y}_{\text{noisy}}\mathbf{Y}_{\text{noisy}}^T)^+ \quad (16)$$

$\mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^+ = \mathbf{A}^+$ yields

$$= \mathbf{X}\mathbf{Y}_{\text{noisy}}^+ \quad (17)$$

$f_{\text{noisy}}(\mathbf{x}) := f(\mathbf{x}) + \varepsilon$ yields

$$=: \bar{\nabla}^{-1}f_{\text{noisy}}. \quad (18)$$

(Snyder, 2015)

$\implies \bar{\mathbf{K}} = \bar{\nabla}^{-1}f_{\text{noisy}}$ “inverts” model + noise.

\implies can estimate $(\hat{\mathbf{x}} - \mathbf{x})$ by applying it to $(\mathbf{y} - \tilde{\mathbf{y}})$, i.e.

$$\hat{\mathbf{x}}(\mathbf{y}) = \hat{\mathbf{x}} + \bar{\mathbf{K}}(\mathbf{y} - \tilde{\mathbf{y}}) \quad (19)$$



EnKF is LLS

What about $\bar{\mathbf{K}}$ itself?

Recall

$$\bar{\mathbf{K}} := \mathbf{X}\mathbf{Y}^T(\mathbf{Y}\mathbf{Y}^T + (N-1)\mathbf{R})^{-1} \quad (13)$$

Then: $\frac{1}{N-1}\mathbf{D}\mathbf{D}^T \approx \mathbf{R}$ yields

$$\approx \mathbf{X}\mathbf{Y}^T(\mathbf{Y}\mathbf{Y}^T + \mathbf{D}\mathbf{D}^T)^+ \quad (14)$$

$\mathbf{Y}_{\text{noisy}} := \mathbf{Y} + \mathbf{D}$, $\mathbf{Y}\mathbf{D}^T \approx \mathbf{0}$ yields

$$\approx \mathbf{X}\mathbf{Y}^T(\mathbf{Y}_{\text{noisy}}\mathbf{Y}_{\text{noisy}}^T)^+ \quad (15)$$

$\mathbf{X}\mathbf{D}^T \approx \mathbf{0}$ yields

$$\approx \mathbf{X}\mathbf{Y}_{\text{noisy}}^T(\mathbf{Y}_{\text{noisy}}\mathbf{Y}_{\text{noisy}}^T)^+ \quad (16)$$

$\mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^+ = \mathbf{A}^+$ yields

$$= \mathbf{X}\mathbf{Y}_{\text{noisy}}^+ \quad (17)$$

$f_{\text{noisy}}(\mathbf{x}) := f(\mathbf{x}) + \varepsilon$ yields

$$=: \bar{\nabla}^{-1}f_{\text{noisy}}. \quad (18)$$

(Snyder, 2015)

$\Rightarrow \bar{\mathbf{K}} = \bar{\nabla}^{-1}f_{\text{noisy}}$ “inverts” model + noise.

\Rightarrow Can estimate $(\mathbf{x} - \bar{\mathbf{x}})$ by applying it to $(\mathbf{y} - \bar{\mathbf{y}})$,

EnKF is LLS



What about $\bar{\mathbf{K}}$ itself?

$$\text{Recall } \bar{\mathbf{K}} := \mathbf{X}\mathbf{Y}^T(\mathbf{Y}\mathbf{Y}^T + (N-1)\mathbf{R})^{-1} \quad (13)$$

$$\text{Then: } \frac{1}{N-1}\mathbf{D}\mathbf{D}^T \approx \mathbf{R} \quad \text{yields} \quad \approx \mathbf{X}\mathbf{Y}^T(\mathbf{Y}\mathbf{Y}^T + \mathbf{D}\mathbf{D}^T)^+ \quad (14)$$

$$\mathbf{Y}_{\text{noisy}} := \mathbf{Y} + \mathbf{D}, \quad \mathbf{Y}\mathbf{D}^T \approx \mathbf{0} \quad \text{yields} \quad \approx \mathbf{X}\mathbf{Y}^T(\mathbf{Y}_{\text{noisy}}\mathbf{Y}_{\text{noisy}}^T)^+ \quad (15)$$

$$\mathbf{X}\mathbf{D}^T \approx \mathbf{0} \quad \text{yields} \quad \approx \mathbf{X}\mathbf{Y}_{\text{noisy}}^T(\mathbf{Y}_{\text{noisy}}\mathbf{Y}_{\text{noisy}}^T)^+ \quad (16)$$

$$\mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^+ = \mathbf{A}^+ \quad \text{yields} \quad = \mathbf{X}\mathbf{Y}_{\text{noisy}}^+ \quad (17)$$

$$f_{\text{noisy}}(\mathbf{x}) := f(\mathbf{x}) + \varepsilon \quad \text{yields} \quad =: \bar{\nabla}^{-1}f_{\text{noisy}}. \quad (18)$$

(Snyder, 2015)

$\Rightarrow \bar{\mathbf{K}} = \bar{\nabla}^{-1}f_{\text{noisy}}$ “inverts” model + noise.

\Rightarrow Can estimate $(\mathbf{x} - \bar{\mathbf{x}})$ by applying it to $(\mathbf{y} - \bar{\mathbf{y}})$, i.e.

$$\hat{\mathbf{x}}(\mathbf{y}) := \bar{\mathbf{x}} + \bar{\mathbf{K}}(\mathbf{y} - \bar{\mathbf{y}}) \checkmark \quad (19)$$



EnKF is LLS (continued)

But $\hat{x}(y) := \bar{x} + \bar{K}(y - \bar{y})$ only yields the mean update.

What about the variance?

For UD, add a random error: $\hat{x}(y) + \xi$,
so as to "explain" residuals, i.e.

$$\xi \sim \sigma - \hat{x}(y). \quad (20)$$

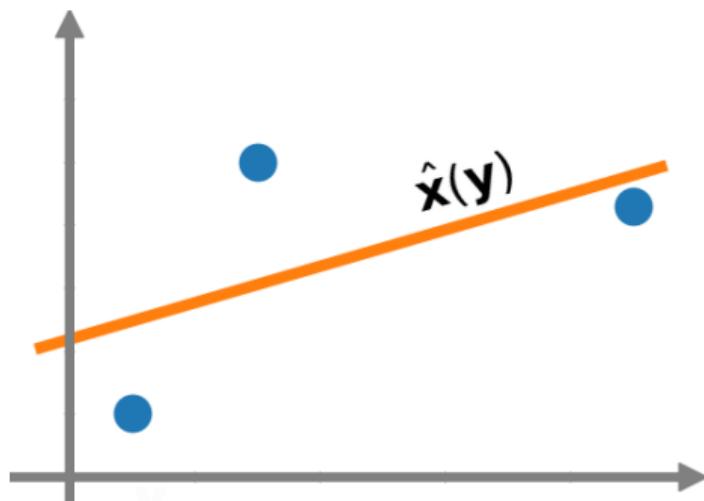
For example, for all m :

$$\xi_m = x_m - \hat{x}(y_m^{\text{noisy}}) \quad (21)$$

Then

$$\hat{x}(y) + \xi_m = x_m + \bar{K}(y - y_m^{\text{noisy}}) \quad (22)$$

Alternative: moment matching,
yielding square-root update schemes.





EnKF is LLS (continued)

But $\hat{x}(y) := \bar{x} + \bar{K}(y - \bar{y})$ only yields the mean update.

For UQ, add a random error: $\hat{x}(y) + \xi$
so as to "explain" residuals, i.e.

$$\xi \sim \sigma - \hat{\sigma}(y), \quad (20)$$

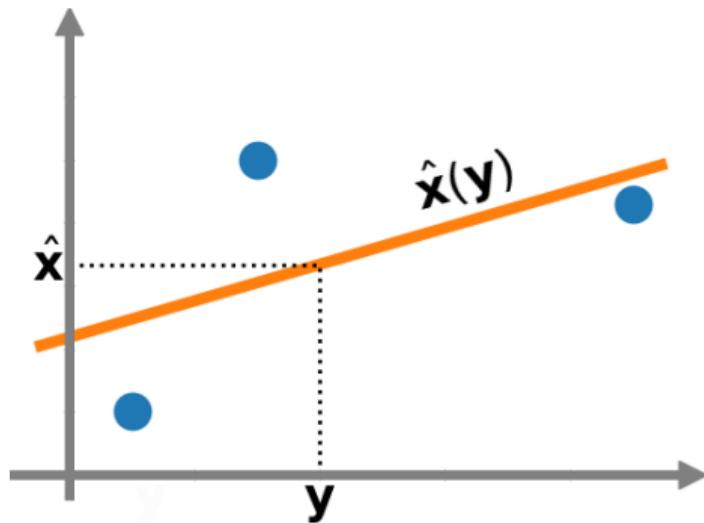
For example, for all m :

$$\xi_m = x_m - \hat{x}(y_m^{\text{noisy}}) \quad (21)$$

Then

$$\hat{x}(y) + \xi_m = x_m + \bar{K}(y - y_m^{\text{noisy}}) \quad (22)$$

Alternative: moment matching,
yielding square-root update schemes.





EnKF is LLS (continued)

But $\hat{\mathbf{x}}(\mathbf{y}) := \bar{\mathbf{x}} + \bar{\mathbf{K}}(\mathbf{y} - \bar{\mathbf{y}})$ only yields the mean update.
What about the full ensemble?

For UQ, add a random error: $\hat{\mathbf{x}}(\mathbf{y}) + \xi$
so as to "explain" residuals, i.e.

$$\xi \sim \sigma - \hat{\mathbf{x}}(\mathbf{y}), \quad (20)$$

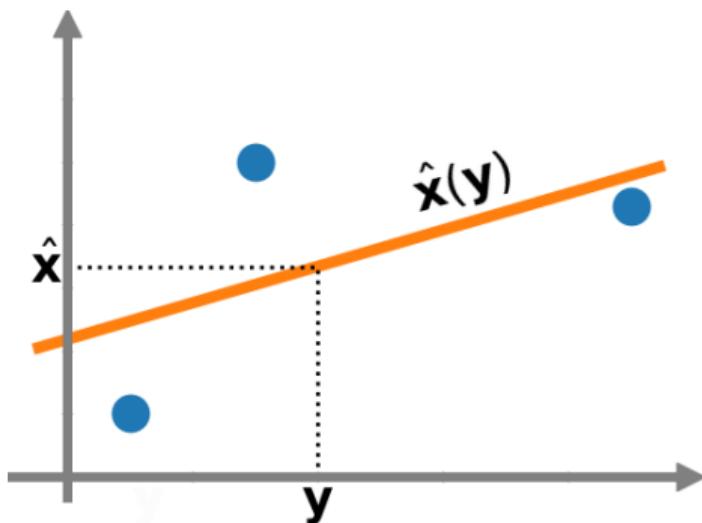
For example, for all n :

$$\xi_n = \mathbf{x}_n - \hat{\mathbf{x}}(\mathbf{y}_n^{\text{noisy}}) \quad (21)$$

Then

$$\hat{\mathbf{x}}(\mathbf{y}) + \xi_n = \mathbf{x}_n + \bar{\mathbf{K}}(\mathbf{y} - \mathbf{y}_n^{\text{noisy}}) \quad (22)$$

Alternative: moment matching,
yielding square-root update schemes.





EnKF is LLS (continued)

But $\hat{\mathbf{x}}(\mathbf{y}) := \bar{\mathbf{x}} + \bar{\mathbf{K}}(\mathbf{y} - \bar{\mathbf{y}})$ only yields the mean update.

What about the full ensemble?

For UQ, add a random error: $\hat{\mathbf{x}}(\mathbf{y}) + \boldsymbol{\xi}$,

so as to "explain" residuals, i.e.

$$\boldsymbol{\xi} \sim \mathbf{z} - \hat{\mathbf{x}}(\mathbf{y}), \quad (20)$$

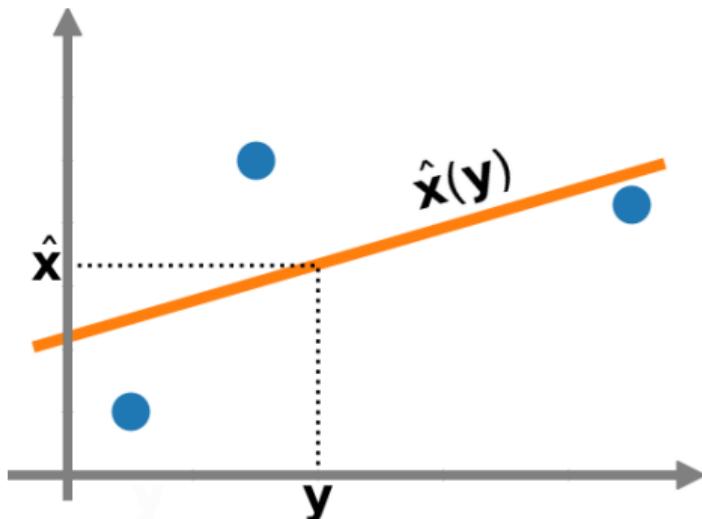
For example, for all m :

$$\xi_m = z_m - \hat{x}_m(\mathbf{y}_m^{\text{noisy}}) \quad (21)$$

Then

$$\hat{x}(\mathbf{y}) + \xi_m = z_m + \bar{\mathbf{K}}(\mathbf{y} - \mathbf{y}_m^{\text{noisy}}) \quad (22)$$

Alternative: moment matching,
yielding square-root update schemes.

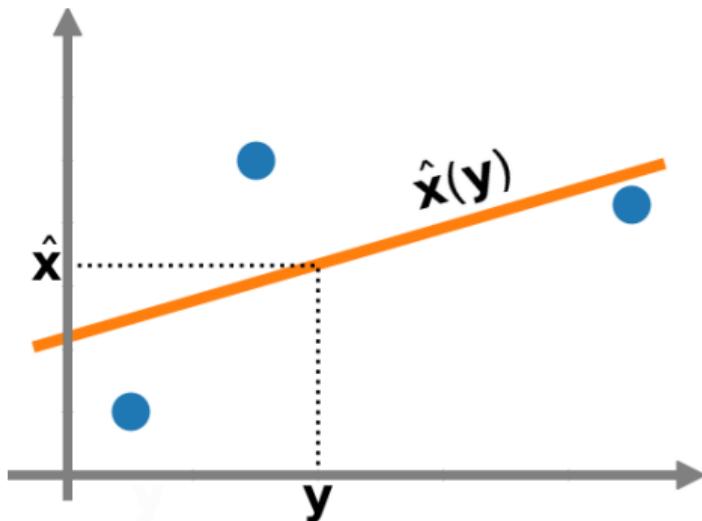




EnKF is LLS (continued)

But $\hat{\mathbf{x}}(\mathbf{y}) := \bar{\mathbf{x}} + \bar{\mathbf{K}}(\mathbf{y} - \bar{\mathbf{y}})$ only yields the mean update.
What about the full ensemble?

For UQ, add a random error: $\hat{\mathbf{x}}(\mathbf{y}) + \boldsymbol{\xi}$,
so as to “explain” residuals,



For example, for all m :

$$\xi_m = x_m - \hat{x}(y_m^{\text{noisy}}) \quad (2)$$

Then

$$\hat{x}(y) + \xi_m = x_m + \bar{K}(y - y_m^{\text{noisy}}) \quad (2)$$

Alternative: moment matching,
yielding square-root update schemes.



EnKF is LLS (continued)

But $\hat{\mathbf{x}}(\mathbf{y}) := \bar{\mathbf{x}} + \bar{\mathbf{K}}(\mathbf{y} - \bar{\mathbf{y}})$ only yields the mean update.
What about the full ensemble?

For UQ, add a random error: $\hat{\mathbf{x}}(\mathbf{y}) + \boldsymbol{\xi}$,
so as to “explain” residuals, i.e.

$$\boldsymbol{\xi} \sim \mathbf{x} - \hat{\mathbf{x}}(\mathbf{y}). \quad (20)$$

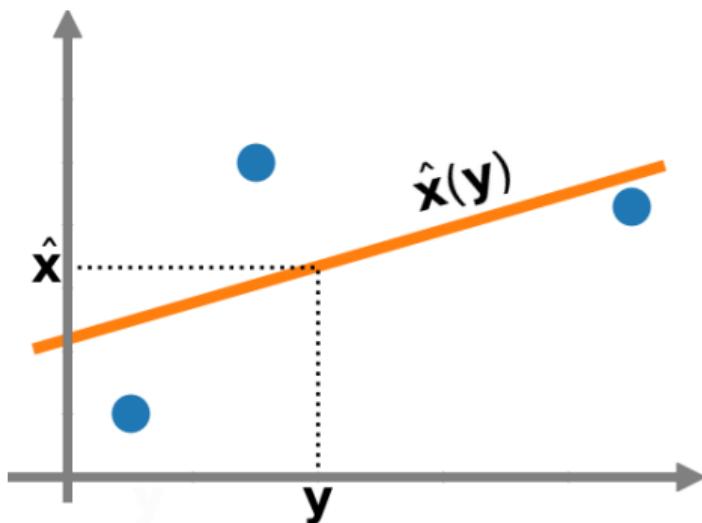
For example, for all m :

$$\xi_m = x_m - \hat{x}(y_m^{\text{noisy}}) \quad (21)$$

Then

$$\hat{x}(\mathbf{y}) + \boldsymbol{\xi}_m = x_m + \bar{\mathbf{K}}(y - y_m^{\text{noisy}}) \quad (22)$$

Alternative: moment matching,
yielding square-root update schemes.





EnKF is LLS (continued)

But $\hat{\mathbf{x}}(\mathbf{y}) := \bar{\mathbf{x}} + \bar{\mathbf{K}}(\mathbf{y} - \bar{\mathbf{y}})$ only yields the mean update.
What about the full ensemble?

For UQ, add a random error: $\hat{\mathbf{x}}(\mathbf{y}) + \boldsymbol{\xi}$,
so as to “explain” residuals, i.e.

$$\boldsymbol{\xi} \sim \mathbf{x} - \hat{\mathbf{x}}(\mathbf{y}). \quad (20)$$

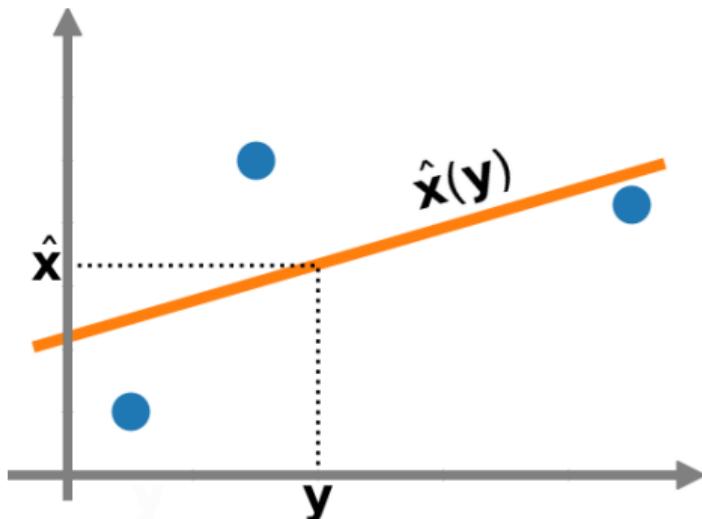
For example, for all n :

$$\boldsymbol{\xi}_n = \mathbf{x}_n - \hat{\mathbf{x}}(\mathbf{y}_n^{\text{noisy}}) \quad (21)$$

Then

$$\hat{\mathbf{x}}(\mathbf{y}) + \boldsymbol{\xi}_n = \mathbf{x}_n + \bar{\mathbf{K}}(\mathbf{y} - \mathbf{y}_n^{\text{noisy}}) \quad (22)$$

Alternative: moment matching,
yielding square-root update schemes.





EnKF is LLS (continued)

But $\hat{\mathbf{x}}(\mathbf{y}) := \bar{\mathbf{x}} + \bar{\mathbf{K}}(\mathbf{y} - \bar{\mathbf{y}})$ only yields the mean update.
What about the full ensemble?

For UQ, add a random error: $\hat{\mathbf{x}}(\mathbf{y}) + \boldsymbol{\xi}$,
so as to “explain” residuals, i.e.

$$\boldsymbol{\xi} \sim \mathbf{x} - \hat{\mathbf{x}}(\mathbf{y}). \quad (20)$$

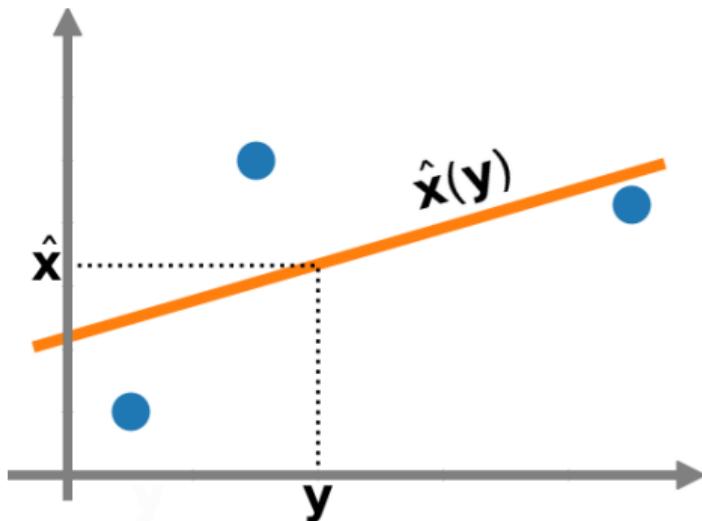
For example, for all n :

$$\boldsymbol{\xi}_n = \mathbf{x}_n - \hat{\mathbf{x}}(\mathbf{y}_n^{\text{noisy}}) \quad (21)$$

Then

$$\hat{\mathbf{x}}(\mathbf{y}) + \boldsymbol{\xi}_n = \bar{\mathbf{x}} + \bar{\mathbf{K}}(\mathbf{y} - \mathbf{y}_n^{\text{noisy}}) \quad (22)$$

achieving moment matching,
yielding square-root update schemes.





EnKF is LLS (continued)

But $\hat{\mathbf{x}}(\mathbf{y}) := \bar{\mathbf{x}} + \bar{\mathbf{K}}(\mathbf{y} - \bar{\mathbf{y}})$ only yields the mean update.
What about the full ensemble?

For UQ, add a random error: $\hat{\mathbf{x}}(\mathbf{y}) + \boldsymbol{\xi}$,
so as to “explain” residuals, i.e.

$$\boldsymbol{\xi} \sim \mathbf{x} - \hat{\mathbf{x}}(\mathbf{y}). \quad (20)$$

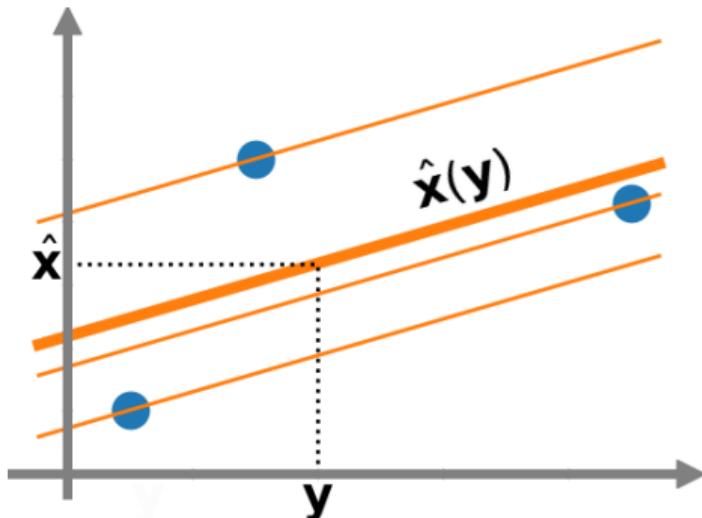
For example, for all n :

$$\boldsymbol{\xi}_n = \mathbf{x}_n - \hat{\mathbf{x}}(\mathbf{y}_n^{\text{noisy}}) \quad (21)$$

Then

$$\hat{\mathbf{x}}(\mathbf{y}) + \boldsymbol{\xi}_n = \bar{\mathbf{x}} + \bar{\mathbf{K}}(\mathbf{y} - \mathbf{y}_n^{\text{noisy}}) \quad (22)$$

Alternative moment matching,
yielding square-root update schemes.





EnKF is LLS (continued)

But $\hat{\mathbf{x}}(\mathbf{y}) := \bar{\mathbf{x}} + \bar{\mathbf{K}}(\mathbf{y} - \bar{\mathbf{y}})$ only yields the mean update.
What about the full ensemble?

For UQ, add a random error: $\hat{\mathbf{x}}(\mathbf{y}) + \boldsymbol{\xi}$,
so as to “explain” residuals, i.e.

$$\boldsymbol{\xi} \sim \mathbf{x} - \hat{\mathbf{x}}(\mathbf{y}). \quad (20)$$

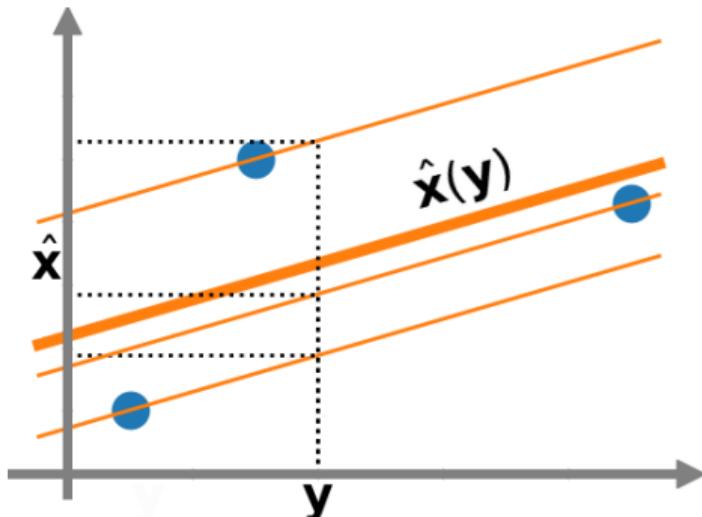
For example, for all n :

$$\boldsymbol{\xi}_n = \mathbf{x}_n - \hat{\mathbf{x}}(\mathbf{y}_n^{\text{noisy}}) \quad (21)$$

Then

$$\hat{\mathbf{x}}(\mathbf{y}) + \boldsymbol{\xi}_n = \bar{\mathbf{x}} + \bar{\mathbf{K}}(\mathbf{y} - \mathbf{y}_n^{\text{noisy}}) \quad (22)$$

achieving moment matching,
yielding square-root update schemes.





EnKF is LLS (continued)

But $\hat{\mathbf{x}}(\mathbf{y}) := \bar{\mathbf{x}} + \bar{\mathbf{K}}(\mathbf{y} - \bar{\mathbf{y}})$ only yields the mean update.
What about the full ensemble?

For UQ, add a random error: $\hat{\mathbf{x}}(\mathbf{y}) + \boldsymbol{\xi}$,
so as to “explain” residuals, i.e.

$$\boldsymbol{\xi} \sim \mathbf{x} - \hat{\mathbf{x}}(\mathbf{y}). \quad (20)$$

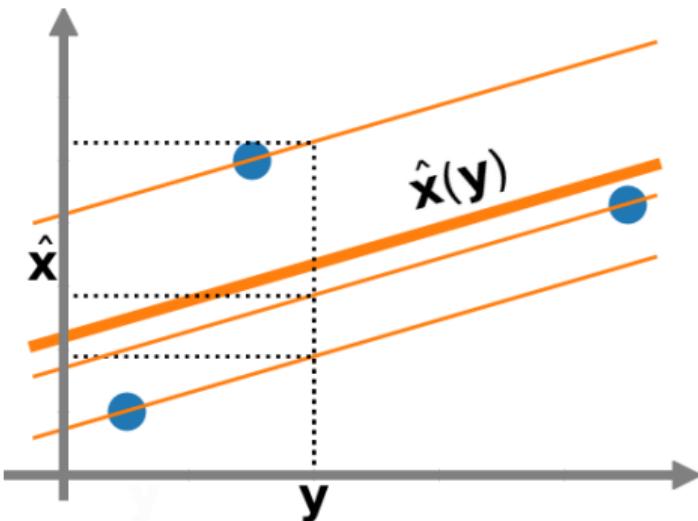
For example, for all n :

$$\boldsymbol{\xi}_n = \mathbf{x}_n - \hat{\mathbf{x}}(\mathbf{y}_n^{\text{noisy}}) \quad (21)$$

Then

$$\hat{\mathbf{x}}(\mathbf{y}) + \boldsymbol{\xi}_n = \mathbf{x}_n + \bar{\mathbf{K}}(\mathbf{y} - \mathbf{y}_n^{\text{noisy}}). \quad (22)$$

ANOTHER TRICK IN ENSEMBLE
yielding square-root update schemes.





EnKF is LLS (continued)

But $\hat{\mathbf{x}}(\mathbf{y}) := \bar{\mathbf{x}} + \bar{\mathbf{K}}(\mathbf{y} - \bar{\mathbf{y}})$ only yields the mean update.
What about the full ensemble?

For UQ, add a random error: $\hat{\mathbf{x}}(\mathbf{y}) + \boldsymbol{\xi}$,
so as to “explain” residuals, i.e.

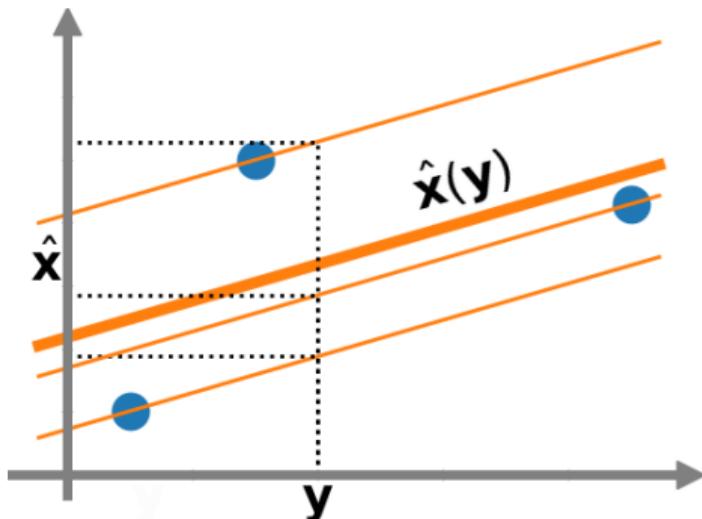
$$\boldsymbol{\xi} \sim \mathbf{x} - \hat{\mathbf{x}}(\mathbf{y}). \quad (20)$$

For example, for all n :

$$\boldsymbol{\xi}_n = \mathbf{x}_n - \hat{\mathbf{x}}(\mathbf{y}_n^{\text{noisy}}) \quad (21)$$

Then

$$\hat{\mathbf{x}}(\mathbf{y}) + \boldsymbol{\xi}_n = \mathbf{x}_n + \bar{\mathbf{K}}(\mathbf{y} - \mathbf{y}_n^{\text{noisy}}). \checkmark \quad (22)$$



yielding square-root update schemes.



EnKF is LLS (continued)

But $\hat{\mathbf{x}}(\mathbf{y}) := \bar{\mathbf{x}} + \bar{\mathbf{K}}(\mathbf{y} - \bar{\mathbf{y}})$ only yields the mean update.
What about the full ensemble?

For UQ, add a random error: $\hat{\mathbf{x}}(\mathbf{y}) + \boldsymbol{\xi}$,
so as to “explain” residuals, i.e.

$$\boldsymbol{\xi} \sim \mathbf{x} - \hat{\mathbf{x}}(\mathbf{y}). \quad (20)$$

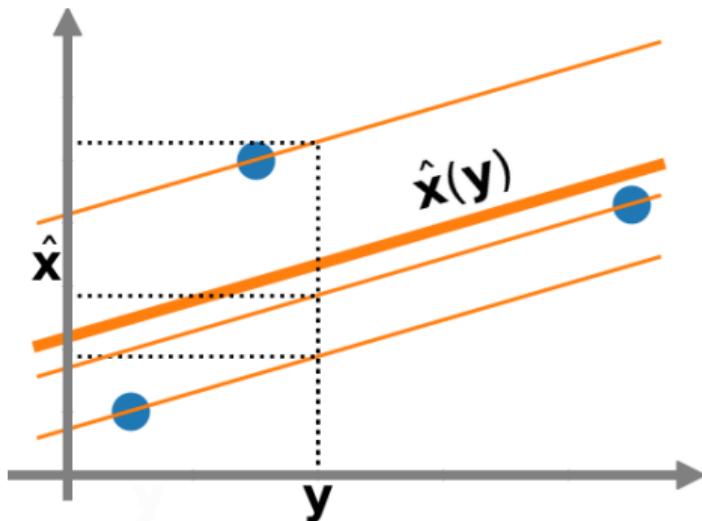
For example, for all n :

$$\boldsymbol{\xi}_n = \mathbf{x}_n - \hat{\mathbf{x}}(\mathbf{y}_n^{\text{noisy}}) \quad (21)$$

Then

$$\hat{\mathbf{x}}(\mathbf{y}) + \boldsymbol{\xi}_n = \mathbf{x}_n + \bar{\mathbf{K}}(\mathbf{y} - \mathbf{y}_n^{\text{noisy}}). \checkmark \quad (22)$$

Alternative: moment matching,
yielding square-root update schemes.





EnKF is LLS (continued)

But $\hat{\mathbf{x}}(\mathbf{y}) := \bar{\mathbf{x}} + \bar{\mathbf{K}}(\mathbf{y} - \bar{\mathbf{y}})$ only yields the mean update.
What about the full ensemble?

For UQ, add a random error: $\hat{\mathbf{x}}(\mathbf{y}) + \boldsymbol{\xi}$,
so as to “explain” residuals, i.e.

$$\boldsymbol{\xi} \sim \mathbf{x} - \hat{\mathbf{x}}(\mathbf{y}). \quad (20)$$

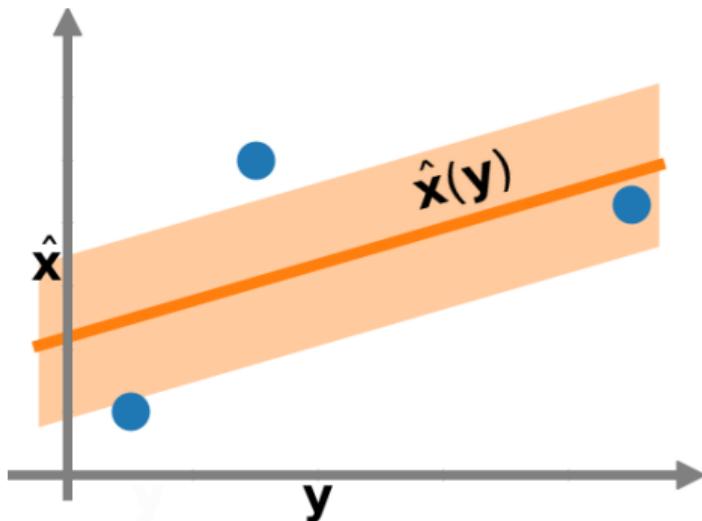
For example, for all n :

$$\boldsymbol{\xi}_n = \mathbf{x}_n - \hat{\mathbf{x}}(\mathbf{y}_n^{\text{noisy}}) \quad (21)$$

Then

$$\hat{\mathbf{x}}(\mathbf{y}) + \boldsymbol{\xi}_n = \mathbf{x}_n + \bar{\mathbf{K}}(\mathbf{y} - \mathbf{y}_n^{\text{noisy}}). \checkmark \quad (22)$$

Alternative: moment matching,
yielding square-root update schemes.





EnKF is LLS (continued)

But $\hat{\mathbf{x}}(\mathbf{y}) := \bar{\mathbf{x}} + \bar{\mathbf{K}}(\mathbf{y} - \bar{\mathbf{y}})$ only yields the mean update.
What about the full ensemble?

For UQ, add a random error: $\hat{\mathbf{x}}(\mathbf{y}) + \boldsymbol{\xi}$,
so as to “explain” residuals, i.e.

$$\boldsymbol{\xi} \sim \mathbf{x} - \hat{\mathbf{x}}(\mathbf{y}). \quad (20)$$

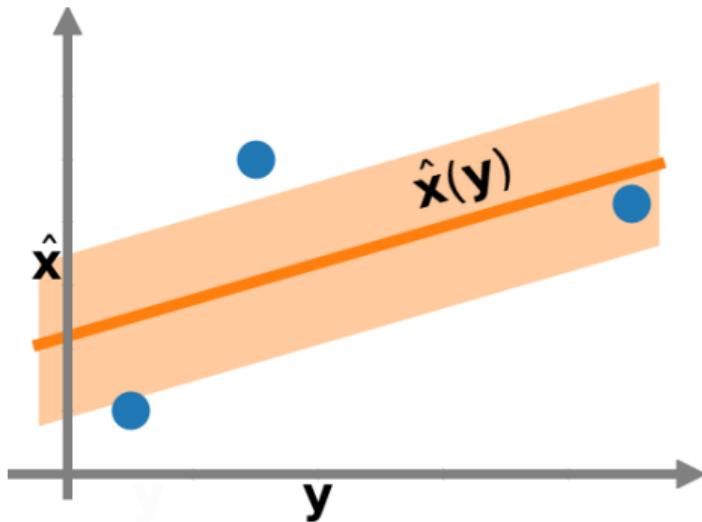
For example, for all n :

$$\boldsymbol{\xi}_n = \mathbf{x}_n - \hat{\mathbf{x}}(\mathbf{y}_n^{\text{noisy}}) \quad (21)$$

Then

$$\hat{\mathbf{x}}(\mathbf{y}) + \boldsymbol{\xi}_n = \mathbf{x}_n + \bar{\mathbf{K}}(\mathbf{y} - \mathbf{y}_n^{\text{noisy}}). \checkmark \quad (22)$$

Alternative: moment matching,
yielding square-root update schemes.





Summary



Summary

Iter. ens
uses LLS



EnKF
uses LLS



EAKF, ETKF
uses LLS



EnKF
is LLS
LLS \Rightarrow
avrg. grad





Patrick Nima Raanes, Andreas Størksen Stordal, and Geir Evensen. Revising the stochastic iterative ensemble smoother. *Nonlinear Processes in Geophysics*, 26 (3):325–338, 2019. DOI: [10.5194/npg-26-325-2019](https://doi.org/10.5194/npg-26-325-2019).

C. Snyder. Introduction to the Kalman filter. In Eric Blayo, Marc Bocquet, Emmanuel Cosme, and Leticia F. Cugliandolo, editors, *Lecture Notes of the Les Houches School of Physics: Special Issue, 28 May–15 June 2012: Advanced Data Assimilation for Geosciences*, pages 75–120. Oxford University Press, 2015.

Andreas S. Stordal, Slawomir P. Szklarz, and Olwijn Leeuwenburgh. A theoretical look at ensemble-based optimization in reservoir management. *Mathematical Geosciences*, 48(4):399–417, 2016. DOI: [10.1007/s11004-015-9598-6](https://doi.org/10.1007/s11004-015-9598-6).