"Ensemblized" linear least squares (LLS)

Patrick N. Raanes

EnKF workshop 2023
May. 5, 2023

Iter. ens uses LLS

EnKF uses LLS

EAKF, ETKF uses LLS

EnKF is LLS

LLS ⇒ avrg. grad

Covariances and LLS regression

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

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

Recall covariance estimate,

$$\bar{\mathbf{C}}_{\boldsymbol{y},\boldsymbol{x}} := \frac{1}{N-1} \sum_n (\boldsymbol{y}_n - \bar{\boldsymbol{y}})(\boldsymbol{x}_n - \bar{\boldsymbol{x}})^\mathsf{T} \tag{1}$$

$$= \frac{1}{N-1} \mathbf{Y}\,\mathbf{X}^\mathsf{T} \tag{2}$$

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

$$\bar{\nabla} f := \bar{\mathbf{C}}_{\boldsymbol{y},\boldsymbol{x}}\, \bar{\mathbf{C}}_{\boldsymbol{x}}^{+} \tag{3}$$

$$= \mathbf{Y}\,\mathbf{X}^\mathsf{T}(\mathbf{X}\,\mathbf{X}^\mathsf{T})^{+} \tag{4}$$

$$= \mathbf{Y}\mathbf{X}^{+}\,. \qquad (\mathbf{A}^{+} = \mathbf{A}^\mathsf{T}(\mathbf{A}\mathbf{A}^\mathsf{T})^{+})$$

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

Stein's lemma

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

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

$$= -\int_{\mathbb{R}^d} f(\boldsymbol{x})\, [\nabla p(\boldsymbol{x})]^{\mathsf{T}}\, d\boldsymbol{x} + \cancel{\left[f(\boldsymbol{x})\, p(\boldsymbol{x})\, \boldsymbol{n}\right]_{\partial\mathbb{R}^d}} \tag{6}$$

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

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

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

and so

$$\mathbb{E}\,\nabla f(\boldsymbol{x}) = \int_{\mathbb{R}^d} f(\boldsymbol{x})\, p(\boldsymbol{x})\, (\boldsymbol{x} - \boldsymbol{\mu})^{\mathsf{T}}\, d\boldsymbol{x} \;\; \mathbf{C}_{\boldsymbol{x}}^{-1} \tag{7}$$

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

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

Stein's advantages

Raanes et al. (2019): $\mathbf{Y}\mathbf{X}^{+} =: \bar{\nabla} f \xrightarrow[N\to\infty]{} \mathbb{E}\, \nabla f(\boldsymbol{x})$ if $x \sim \mathcal{N}$ (9)

Stordal et al. (2016): $= \nabla\, \mathbb{E}\, f(\boldsymbol{x})$

Provides link to analytic gradient, ∇f.

Better than solving order-1 Taylor @ mean:

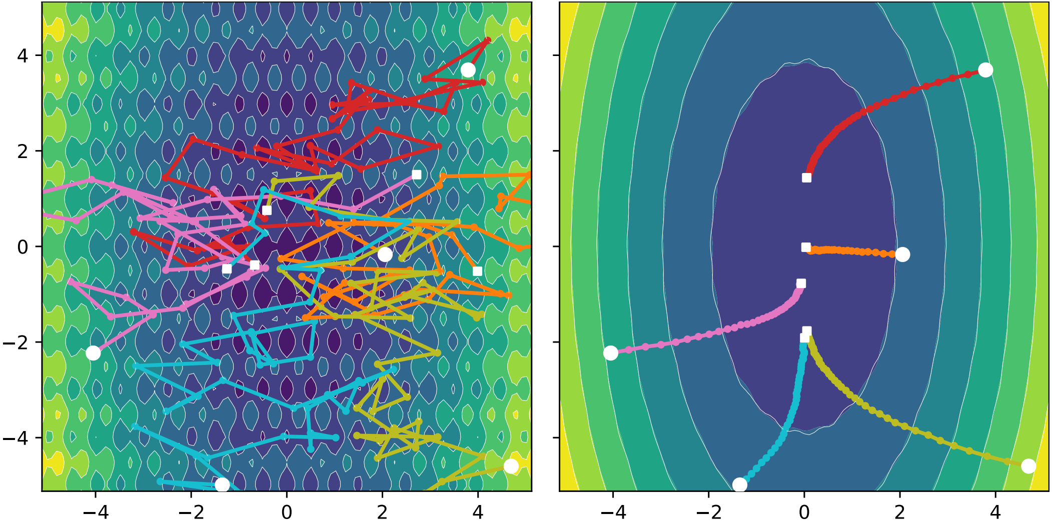
$$\mathbf{Y} \approx \nabla f(\bar{\boldsymbol{x}})\, \mathbf{X}\,,$$

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

because Stein (9)

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

Implicit blurring




EnKF
Monte-Carlo
+ LLS
imgflip.com

EnKFs use LLS

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

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

As in ***EAKF***, first update in obs-space : $\begin{cases} \mathbf{P} \\ \mathbf{H} \end{cases} \leftarrow \begin{matrix} \bar{\mathbf{C}}_{\boldsymbol{y}} = \mathbf{Y}\mathbf{Y}^\mathsf{T}/(N-1) \\ \mathbf{I} \end{matrix}$ in (10), 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: $\overbrace{\boldsymbol{x} = \bar{\boldsymbol{x}} + \mathbf{X}\boldsymbol{w}}^{\boldsymbol{\phi}(\boldsymbol{w})}$ using $\begin{cases} \mathbf{P} \\ \mathbf{H} \end{cases} \leftarrow \begin{matrix} \bar{\mathbf{C}}_{\boldsymbol{w}} = \mathbf{I}/(N-1) \\ \bar{\nabla}(f \circ \boldsymbol{\phi}) = \mathbf{Y}\,\mathbf{I}^+ = \mathbf{Y} \end{matrix}$

Then $\boldsymbol{\phi}$ "applied to" $\mathbf{K}$ to revert to state space. Yields $\bar{\mathbf{K}}$. ✓

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

Then:

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

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

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

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

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

(Snyder, 2015)

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

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

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

EnKF *is* LLS (continued)

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

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

$$\boldsymbol{\xi} \sim \boldsymbol{x} - \hat{\boldsymbol{x}}(\boldsymbol{y}) \,. \tag{20}$$

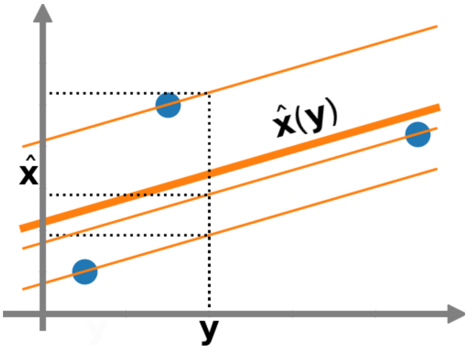
For example, for all n:

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

Then

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

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



Summary

Iter. ens
uses LLS

EnKF
uses LLS

EAKF, ETKF
uses LLS

EnKF
is LLS

LLS ⇒
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.

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.