Advancements of the EnKF: inverse problems and optimal transport

Neil Chada

National University of Singapore

IMS programme: Bayesian Computation for High-Dimensional Statistical Models

Marco Iglesias (Nottingham), Lassi Roininen (LUT), Claudia Schillings (Mannheim), Andrew Stuart (Caltech), Alex Thiéry (NUS), Simon Weissemann (Mannheim)

September 10, 2018

National University of Singapore

Common problem: How to blend underlying dynamics (model) with noisy observations (data)?

Common problem: How to blend underlying dynamics (model) with noisy observations (data)?

Original solution: Kalman!

Kalman Filter

State Space Model

$$
\begin{aligned}
\text{Dynamics Model:}\quad & v_{n+1} = M v_n + \xi_n, \quad n \in \mathbb{Z}^+ \\
\text{Data Model:}\quad & y_{n+1} = H v_{n+1} + \eta_{n+1}, \quad n \in \mathbb{Z}^+ \\
\text{Probabilistic Structure:}\quad & v_0 \sim N(m_0, C_0), \quad \xi_n \sim N(0, \Sigma), \quad \eta_n \sim N(0, \Gamma) \\
\text{Probabilistic Structure:}\quad & v_0 \perp \{\xi_n\} \perp \{\eta_n\} \text{ independent}
\end{aligned}
$$

- J. Basic Engineering **82**(1960); see [1].
- 27,307 Google Scholar citations.
- Kalman: National Medal of Science, 2008.
- $Y_n = \{y_\ell\}_{\ell=1}^n$.
- $v_n | Y_n \sim N(m_n, C_n)$.
- $(m_n, C_n) \mapsto (m_{n+1}, C_{n+1})$.
- Navigational and guidance systems.
- Apollo 11.

Kalman Filter

Sequential Optimization Perspective

$$
\begin{aligned}
\text{Predict:}\quad & \hat{m}_{n+1} = M m_n, \quad n \in \mathbb{Z}^+ \\
\text{Model/Data Compromise:}\quad & J_n(m) = \frac{1}{2}|m - \hat{m}_{n+1}|^2_{\hat{C}_{n+1}} + \frac{1}{2}|y_{n+1} - Hm|^2_\Gamma \\
\text{Optimize:}\quad & m_{n+1} = \operatorname{argmin}_m J_n(m).
\end{aligned}
$$

- $|\cdot|_A = |A^{-\frac{1}{2}} \cdot|$ for $A > 0$.
- Updating $\hat{C}_{n+1}$ is expensive: $\mathcal{O}(d^2)$ storage.
- d the state space dimension.

Ensemble Kalman Filter

State Space Model

$$\begin{aligned} \text{Dynamics Model:} \quad & v_{n+1} = \Psi(v_n) + \xi_n, \quad n \in \mathbb{Z}^+ \\ \text{Data Model:} \quad & y_{n+1} = Hv_{n+1} + \eta_{n+1}, \quad n \in \mathbb{Z}^+ \\ \text{Probabilistic Structure:} \quad & v_0 \sim N(m_0, C_0), \quad \xi_n \sim N(0, \Sigma), \quad \eta_j \sim N(0, \Gamma) \\ \text{Probabilistic Structure:} \quad & v_0 \perp \{\xi_n\} \perp \{\eta_n\} \text{ independent} \end{aligned}$$

- Motivated by extended Kalman filter; see [2].
- Jazwinski (1970) [3], Ghil et al (1981) [4].
- J. Geophysical Research **99**(1994).
- Original paper in ocean dynamics.
- Used in weather forecasting centres worldwide.
- $\{u_n^{(k)}\}_{k=1}^K \mapsto \{u_{n+1}^{(k)}\}_{k=1}^K$.

Ensemble Kalman Filter

Sequential Optimization Perspective

$$\text{Predict:} \quad \hat{v}_{n+1}^{(k)} = \Psi(v_n^{(k)}) + \xi_n^{(k)}, \quad n \in \mathbb{Z}^+$$

$$\text{Model/Data Compromise:} \quad J_n^{(k)}(v) = \frac{1}{2}|v - \hat{v}_{n+1}^{(k)}|^2_{\hat{C}_{n+1}} + \frac{1}{2}|y_{n+1}^{(k)} - Hv|^2_\Gamma$$

$$\text{Optimize:} \quad v_{n+1} = \mathrm{argmin}_v J_n^{(k)}(v).$$

- $\hat{C}_{n+1}$ is empirical covariance of the $\{\hat{v}_{n+1}^{(k)}\}$.
- Updating $\hat{C}_n$ requires only $\mathcal{O}(Kd)$ storage.
- The simplicity and cost of the EnKF makes its applicable!

Inverse Problems

- **Aim**: The recovery of an unknown $u \in \mathcal{X}$ from perturbed noisy measurements of data $y \in \mathcal{Y}$ where

$$y = \mathcal{G}(u) + \eta, \quad \eta \sim \mathcal{N}(0, \Gamma) \tag{1}$$

- **Bayesian approach**: Unknown is now a probabilistic distribution of the random variable $u|y$ using Bayes' formula

$$\mathbb{P}(u|y) \propto \mathbb{P}(y|u)\mathbb{P}(u).$$

- We consider a posterior measure μ^y described through Radon-Nikodym derivative

$$\frac{d\mu^y}{d\mu_0}(u) = \frac{1}{Z}\exp(-\Phi(u;y)),$$

with misfit functional

$$\Phi(u;y) = \frac{1}{2}|y - \mathcal{G}(u)|_\Gamma^2.$$

Ensemble Kalman Inversion (EKI)

Problem Statement

Find u from y where $\mathcal{G} : \mathcal{X} \mapsto \mathcal{Y}$, η is noise and

$$y = \mathcal{G}(u) + \eta.$$

- Origins in data assimilation applications. (Oliver et al. [2008]).
- Application to PDE-constrained inverse problems. (Iglesias et al. [2013]).
- Noise controlled system, i.e.
- Requires only black box forward model $\mathcal{G}$.
- Very limited theoretical understanding.

Lemma

For all $n \in \mathbb{Z}^+, j \in \{1, \dots, J\}$, $u_n^{(j)} \in \mathrm{span}\big(\{u_0^{(j)}\}_{j=1}^J\big)$.

i.e. the updated ensemble $\{u_n^{(j)}\}_{j=1}^J$ *spanned by the linear span of the initial ensemble* $\{u_0^{(j)}\}_{j=1}^J$. *Limitation*

The Basic Algorithm

- Initial Ensemble $\{u_0^{(j)}\}_{j=1}^{J} \subset X$.
- Ensemble First and Second Order Moments

Means:

$$\overline{u}_n = \frac{1}{J}\sum_{j=1}^{J} u_n^{(j)}, \quad \overline{w}_n = \frac{1}{J}\sum_{j=1}^{J} \mathcal{G}(u_n^{(j)}).$$

Covariances:

$$C_n^{ww} = \frac{1}{J-1}\sum_{j=1}^{J}\big(\mathcal{G}(u_n^{(j)}) - \overline{w}_n\big) \otimes \big(\mathcal{G}(u_n^{(j)}) - \overline{w}_n\big),$$

$$C_n^{uw} = \frac{1}{J-1}\sum_{j=1}^{J}\big(u_n^{(j)} - \overline{u}_n\big) \otimes \big(\mathcal{G}(u_n^{(j)}) - \overline{w}_n\big).$$

- Update step
  $n \mapsto n+1$:

$$u_{n+1}^{(j)} = u_n^{(j)} + C_n^{uw}\big(C_n^{ww} + \Gamma\big)^{-1}\big(y^{(j)} - \mathcal{G}(u_n^{(j)})\big)$$

Whittle-Matérn Initial Ensembles

- Create initial ensemble of functions via Gaussian random fields.
- Common choice: Whittle-Matérn family

$$c_{\sigma,\nu,\tau}(x,x') := \sigma^2 \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\tau|x-x'|\right)^{\nu} K_{\nu}\left(\tau|x-x'|\right).$$

  - Smoothness parameter: $\nu \in \mathbb{R}^+$.
  - Inverse length-scale parameter: $\tau \in \mathbb{R}^+$.
  - Amplitude parameter: $\sigma \in \mathbb{R}$.

- Corresponding covariance operator

$$\mathcal{C}_{\sigma,\nu,\tau} \propto \sigma^2 \tau^{2\nu} (\tau^2 I - \triangle)^{-\nu-\frac{d}{2}}.$$

  - $\nu = \alpha - \frac{d}{2}$.

- Hierarchical: invert for parameters such as σ, ν, τ as well as field itself.

Centred vs Non-centred

- Define $\theta = (\alpha, \tau)$.
- Generate samples v by solving the SPDE

$$(\tau^2 I - \triangle)^{\frac{\alpha}{2}} v = \sigma \tau^{\alpha - \frac{d}{2}} \xi,$$

  where $\xi \sim \mathrm{N}(0, I)$ is white noise. $u = T(\xi, \theta)$.
  See [4] Lindgren et al (2011).
- Hierarchical: invert for parameters θ as well as field v.
- Centred approach:
  - view (v, θ) as unknowns;
  - initial ensemble samples $\mathbb{P}(v|\theta)\mathbb{P}(\theta)$;
  - $y = \mathcal{G}(v) + \eta$.
- Non-centred approach:
  - view (ξ, θ) as unknowns;
  - initial ensemble samples $\mathbb{P}(\xi)\mathbb{P}(\theta)$;
  - $y = \mathcal{G}\big(T(\xi, \theta)\big) + \eta$.

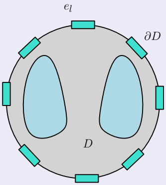
  See [11] Papaspiliopoulos et al (2007).

Electrical Impedance Tomography (EIT)

Forward Problem

Given $(\kappa, I) \in L^\infty(D; \mathbb{R}^+) \times \mathbb{R}^{m_e}$ find $(\nu, V) \in H^1(D) \times \mathbb{R}^{m_e}$:

$$\begin{aligned} -\nabla \cdot (\kappa \nabla \nu) = 0 \quad &\in \quad D, \\ \nu + z_l \kappa \nabla \nu \cdot n = V_l \quad &\in \quad e_l, \quad l = 1, \dots, m_e, \\ \nabla \nu \cdot n = 0 \quad &\in \quad \partial D \backslash \cup_{l=1}^{m_e} e_l, \\ \int \kappa \nabla \nu \cdot n \, ds = I_l \quad &\in \quad e_l, \quad l = 1, \dots, m_e. \end{aligned}$$



Ohms Law:
$V = R(\kappa) \times I$

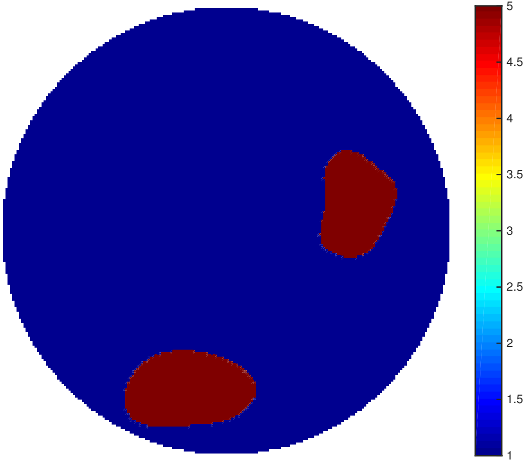
Inverse Problem

Given $\kappa = \exp(u)$ and a set of m measurements of voltage V_l, $\mathcal{G}(u) \in \mathbb{R}^{m_e \times m}$ find u from y where:

$$y = \mathcal{G}(u) + \eta, \quad \eta \sim \mathrm{N}(0, \Gamma).$$

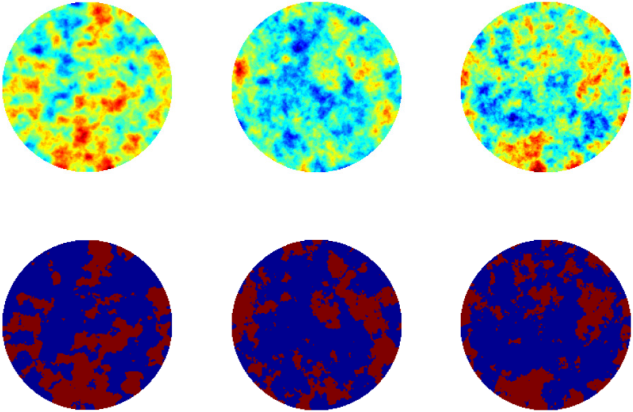
Numerical Example: EIT Revisited

- True Conductivity.
- level set formulation.
- Hyperparameters uniformly chosen: $\tau, \alpha \sim \mathcal{U}[\cdot, \cdot]$.



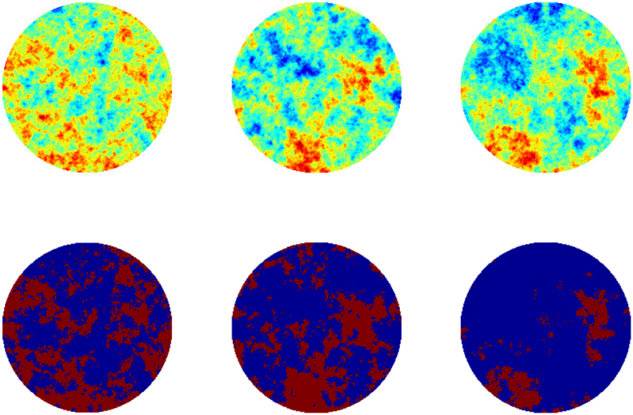
Numerical Example: EIT Revisited

- Non-Hierarchical.
- Reconstructed level set function at three different iteration steps.
- Reconstructed conductivity at three different iteration steps.



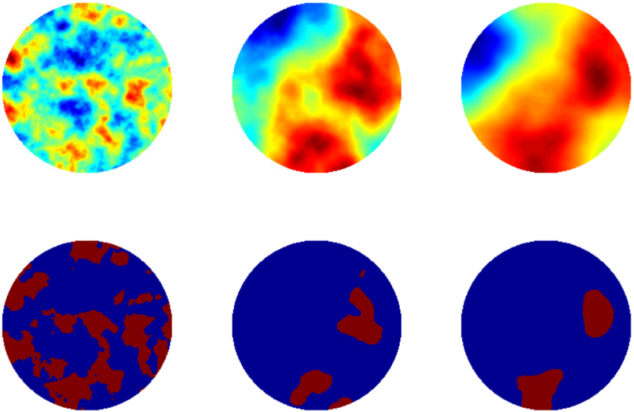
Numerical Example: EIT Revisited

- Centred Hierarchical.
- Reconstructed level set function at three different iteration steps.
- Reconstructed conductivity at three different iteration steps.



Numerical Example: EIT Revisited

- Non-centred Hierarchical.
- Reconstructed level set function at three different iteration steps.
- Reconstructed conductivity at three different iteration steps.



Non-Stationary Hyperparameters

- Treating the length scale $\tau^{-1} = \ell$ as a field.
- To ensure positivity write

$$\ell(x) := \exp(v(x)).$$

- Now we generate samples ξ from the SPDE

$$(1 - \ell(x; v)^2 \Delta)^{\alpha/2} u = \sigma \sqrt{\ell(x; v)^d} \xi.$$

- Centred approach: view (u, v) as unknown.
- Non-centred approach: view (ξ, v) as unknown.
- Gaussian and Cauchy (next weeks talk) length scale.
- Details found in Chada et al. [2018].

Continuum limits

- *Non-hierarchical*:

$$u_{n+1}^{(j)} = u_n^{(j)} + C_n^{uw}\left(C_n^{ww} + h^{-1}\Gamma\right)^{-1}\left(y^{(j)} - \mathcal{G}(u_n^{(j)})\right)$$

- *Hierarchical*: through Euler-Maruyama discretization we have $(h \to 0)$

$$\frac{d\xi^{(j)}}{dt} = \frac{1}{J}\sum_{k=1}^{J}\left\langle \mathcal{G}^T(\xi^{(k)}, \theta^{(k)}) - \overline{\mathcal{G}^T}, y - \mathcal{G}^T(\xi^{(j)}, \theta^{(j)}) + \sqrt{\Gamma}\frac{dW^{(j)}}{dt}\right\rangle_{\Gamma}(\xi^{(k)} - \bar{\xi})$$

- In the linear $\mathcal{G}^T(\xi, \theta) = Au$, noise-free case we have

$$\frac{du^{(k)}}{dt} = -C\nabla\Phi_u(u^{(k)}),$$

$$C = \frac{1}{K}\sum_{k=1}^{K}(u^{(k)} - \overline{u}) \otimes (u^{(k)} - \overline{u}). \tag{2}$$

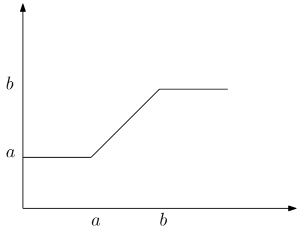
Theorem

For every $(n, j) \in \mathbb{N} \times \{1, \ldots, J\}$ we have $\xi_{n+1}^{(j)}$, $\theta_{n+1}^{(j)} \in \mathcal{A}$ and ξ_{n+1}, $\theta_{n+1} \in \mathcal{A}$ hence $u_{n+1} \notin T\mathcal{A}$, where $T\mathcal{A}$ is the space containing the transformed ensemble of particles.

Box-constrained optimization

- *Issue*: No guarantee the ensemble will stay within a region of interest.
- Motivates box-constrained optimization of the ensemble.
- Let $\mathcal{B} = [a_1, b_1] \times \cdots \times [a_n, b_n]$ denote a box, where we assume that $\mathcal{X} = \mathbb{R}^n$, with a general projection as $\mathcal{P}_{\mathcal{B}} : \mathbb{R}^n \to \mathcal{B}$.

$$(\mathcal{P}(x))_i = \begin{cases} a_i, & \text{if } \; x_i < a_i \\ x_i, & \text{if } \; x_i \in [a_i, b_i], \\ b_i, & \text{if } \; x_i > b_i, \end{cases}, \qquad i = 1, \dots, n,$$



- We consider the ensemble square roof filter (ESRF)

$$\begin{cases} \tilde{u}^{(j)}_{n+1,\mathcal{P}} = u^{(j)}_{n,\mathcal{P}} + C^{up}_{n,\mathcal{P}}(C^{pp}_{n,\mathcal{P}} + h^{-1}\Gamma)^{-1}(y - \frac{1}{2}\mathcal{G}(u^{(j)}_{n.\mathcal{P}}) - \frac{1}{2}\bar{\mathcal{G}}_{\mathcal{P}}), \\ u^{(j)}_{n+1,\mathcal{P}} = \mathcal{P}(\tilde{u}^{(j)}_{n+1,\mathcal{P}}). \end{cases}$$

- Corresponding limit as $h \to 0$

$$(u^{(j)}_{n+1})_i - (u^{(j)}_n)_i = \mathcal{P}\Big([\mathcal{P}(\tilde{u}^{(j)}_n)]_i + h\Big[C^{up}_{n,\mathcal{P}}(hC^{pp}_{n,\mathcal{P}} + \Gamma)^{-1}(y - \frac{1}{2}\mathcal{G}(\mathcal{P}(u^{(j)}_n)) - \frac{1}{2}\bar{\mathcal{G}}_{\mathcal{P}})\Big]_i\Big) - \mathcal{P}\Big([\mathcal{P}(\tilde{u}^{(j)}_n)]_i\Big)$$

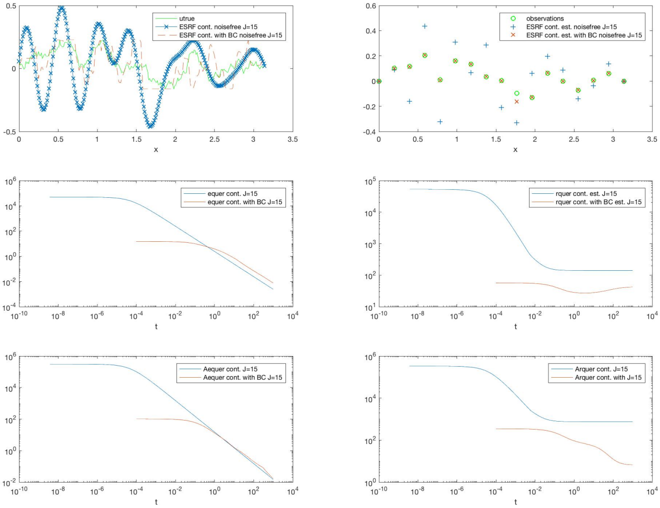
- State in terms of directional derivatives

$$\lim_{h\to 0} \frac{(u^{(j)}_{n+1})_i - (u^{(j)}_n)_i}{h} = \begin{cases} v_i & , (\tilde{u}^{(j)}_n)_i \in (a_i, b_i) \\ 1_{[0,\infty)}(v_i)v_i & , (\tilde{u}^{(j)}_n)_i \leq a_i \\ 1_{(-\infty,0]}(v_i)v_i & , (\tilde{u}^{(j)}_n)_i \geq b_i. \end{cases}$$

We observe that

- $(\tilde{u}^{(j)}_n)_i \in (a_i, b_i)$ if and only if $(u^{(j)}_n)_i \in (a_i, b_i)$,
- $(\tilde{u}^{(j)}_n)_i \leq a_i$ if and only if $(u^{(j)}_n)_i = a_i$,
- $(\tilde{u}^{(j)}_n)_i \geq b_i$ if and only if $(u^{(j)}_n)_i = b_i$.

Numerics (1D cont flow - elliptic PDE)



Optimal transport

- Understand the EnKF through optimal transport (OT).
- Cheung, Cotter, Reich [2012, 2013].
- Arises through coupling of μ and ν.

Kantorovich problem

Given two distributions μ, ν and a set of all couplings $\mathcal{P}(\mu, \nu)$ with cost function $c : \mathcal{X} \times \mathcal{X} \to [0, \infty)$, the OT problem seeks to minimize

$$\mathbb{E}_\gamma[c(u, v)]$$

over the set of couplings γ

$$\gamma^{\mathrm{OT}} := \operatorname{argmin}\{\gamma \mapsto \mathbb{E}_\gamma[c(u, v)], \quad \gamma \in \mathcal{P}(\mu, \nu), \quad (u, v) \sim \gamma\}$$

- Treat μ and ν as prior and posterior.
- Consider gradient flow structure.

Entropic regularization

- Require regularization

$$u^{(\ell+1} := \mathrm{argmin}_{u\in\mathcal{X}} \mathcal{W}_p(u, u^{(\ell)})^p + \tau \; F(u).$$

- Know result: $\tau \to 0$, $\implies$ Fokker-Planck equation.
- Consider a Lagrangian discretization (set of measures and their locations) over time

$$\mu_t := \frac{1}{N}\sum_{i=1}^{N} \delta_{x_i}(t). \tag{3}$$

- equal-weight particle filter.
- Idea: $\mu_t \equiv \mu \to \mu_{t+1} \equiv \nu$, i.e. represent (3) as an N coupled ODEs of particles $X(t) = (x_i(t))_i$.
- Hope: $X(t) = -\nabla\mathcal{F}(X(t))$ where $\mathcal{F}(X) = F(\frac{1}{n}\sum_{i=1}^{n} \delta_{x_i})$.

Gaussian mixture model

- μ, ν are assume Gaussian.
- Gradient flow evolution of Lagrangian discretization.
- Interested in stationary distribution.

Discussion

- The EnKF is a simple but highly applicable algorithm.
- Applications: geosciences, NWP, oceanography, machine learning.
- Considered both inverse problems and optimal transport.
- Blends MC approaches with optimization techniques.

- Project-based optimization, for EKI and HEKI.
- Understanding the EnKF as an optimizer.
- Additional forms of regularization to be included!
- Applications to large-scale problems.

References

G. Evensen
Data Assimilation: the Ensemble Kalman Filter,
Springer, 2009.

N. K. Chada, M. Iglesias, L. Roininen, A. M. Stuart
Paramterizations of ensemble Kalman inversion,
Inverse Problems, 34 055009, 2018.

M. A. Iglesias, K. Law, A. M. Stuart
Ensemble Kalman method for inverse problems,
Inverse Problems, 29 (4), 045001, 2013.

O. Papaspiliopoulos, G. O. Roberts, M. Sköld
A general framework for the parametrization of hierarchical models,
Statistical Science, pages 59-73, 2007.

L. Roininen, M. Girolami, S. Lasanen, M. Markkanen
Hyperpriors for Matern fields with applications in Bayesian inversion,
To appear in Inverse Problems and Imaging, 2018.

G. Peyre, M. Cuturi
Computational Optimal Transport,
arXiv:1803.00567, 2018.

Yuan Cheng, Sebastian Reich
A McKean optimal transportation perspective on Feynman-Kac formulae with application to data assimilation,
arXiv preprint arXiv:1311.6300, 2013.

C. Schillings, A. M. Stuart
Analysis of the ensemble Kalman filter for inverse problems,
SIAM. J. Num. Anal., 2018