Optimization And Sampling Without Derivatives

Andrew Stuart

Computing and Mathematical Sciences
California Institute of Technology

NSF, ONR, Allen Philanthropies,
Mission Control for Earth, Schmidt Futures

⋆ ⋆ ⋆

CSE Seminar MIT

December 9th 2021

Overview

Main Ideas

- Ensemble Kalman: Derivative-Free Optimization and Sampling
- Ensemble Kalman: Filtering and Inverse Problems
- Insights From: Mean Field Derivation
- Insights From: Continuous Time Limits
- Applications: Weather Forecasting, Medical Imaging

Main Ideas

- Ensemble Kalman: Derivative-Free Optimization and Sampling
- Ensemble Kalman: Filtering and Inverse Problems
- Insights From: Mean Field Derivation
- Insights From: Continuous Time Limits
- Applications: Weather Forecasting, Medical Imaging

Alternative Mean-Field Approaches (Consensus) Carrillo et al [7], [5]

Collaborators

- Alex Beskos (UCL)
- Edoardo Calvello (Caltech)
- Alfredo Garbuno-Inigo (Instituto Tecnológico Autónomo de México)
- Franca Hoffmann (Bonn)
- Marco Iglesias (Nottingham)
- Ajay Jasra (KAUST)
- Kody Law (Manchester)
- Wuchen Li (University of South Carolina)
- Sebastian Reich (Potsdam)
- Claudia Schillings (Mannheim)

Filtering

- Practical Considerations Doucet et al [14]
- Continuous Time Bain and Crisan [2]

Hidden Markov Model

Dynamics and Data

$$\begin{aligned}
\text{Dynamics Model:} \quad & v^{\dagger}_{n+1} = \Psi(v^{\dagger}_{n}) + \xi_n, \quad n \in \mathbb{Z}^+ \\
\text{Data Model:} \quad & y^{\dagger}_{n+1} = h(v^{\dagger}_{n+1}) + \eta_{n+1}, \quad n \in \mathbb{Z}^+ \\
\text{Probabilistic Structure:} \quad & v^{\dagger}_0 \sim N(m_0, C_0), \quad \xi_n \sim N(0, \Sigma), \quad \eta_n \sim N(0, \Gamma) \\
\text{Probabilistic Structure:} \quad & v^{\dagger}_0 \perp \{\xi_n\} \perp \{\eta_n\} \text{ independent}
\end{aligned}$$

Hidden Markov Model

Probabilistic Picture

$$\begin{aligned} \text{Dynamics Model (Prediction):} \quad & \widehat{\mu}_{n+1} = P\mu_n, \\ \text{Data Model (Bayes):} \quad & \mu_{n+1} = L_n\widehat{\mu}_{n+1}, \end{aligned}$$

$$Y_n = \{y_1^\dagger, \ldots, y_n^\dagger\}; \quad v_n^\dagger | Y_n \sim \mu_n.$$

Particle Filter

Maps On Probability Measures

$$\begin{aligned} \text{True Model:} \quad & \mu_{n+1} = L_n P \mu_n, \\ \text{Particle Approximation:} \quad & \mu^J_{n+1} = L_n S^J P \mu^J_n, \end{aligned}$$

$$S^J \pi = \frac{1}{J} \sum_{j=1}^{J} \delta_{u^{(j)}}, \quad u^{(j)} \sim \pi, \quad \text{i.i.d.}.$$

Particle Filter

Theorem Rebeschini and Van Handel [33]

Assume h is bounded. Then there is $C(N) > 0$ such that, for all $1 \leq n \leq N$,

$$d(\mu_n, \mu_n^J) \leq C(N)\frac{1}{\sqrt{J}}.$$

$$d(\pi, \pi') = \sup_{|f|_\infty \leq 1}\Big(\mathbb{E}\Big[\big(\pi(f) - \pi'(f)\big)^2\Big]\Big)^{1/2},$$

Particle Filter

Theorem Rebeschini and Van Handel [33]

Assume h is bounded. Then there is $C(N) > 0$ such that, for all $1 \leq n \leq N$,

$$d(\mu_n, \mu_n^J) \leq C(N)\frac{1}{\sqrt{J}}.$$

$$d(\pi, \pi') = \sup_{|f|_\infty \leq 1} \left(\mathbb{E}\left[\left(\pi(f) - \pi'(f)\right)^2\right]\right)^{1/2},$$

$C(N)$ depends badly on dimension:
Weight Collapse In High Dimension

Filtering and Mean Field Dynamics

- Discrete Time Daum et al [12]
- Continuous Time Crisan and Xiong [11]
- Continuous Time Yang et al [41]
- Optimal Transport Reich [34]
- Transport Spantini et al [38]

Mean Field Dynamics

Prediction and Transport – Nonlinear Markov Process

$$\begin{aligned} \text{Dynamics Prediction:} \quad & \widehat{v}_{n+1} = \Psi(v_n) + \xi_n, \\ \text{Data Prediction:} \quad & \widehat{y}_{n+1} = h(\widehat{v}_{n+1}) + \eta_{n+1}, \\ \text{Perfect Transport:} \quad & v_{n+1} = T^S(\widehat{v}_{n+1}, \widehat{y}_{n+1}; \nu_{n+1}, y^\dagger_{n+1}). \end{aligned}$$

Mean Field Dynamics

Prediction and Transport – Nonlinear Markov Process

$$
\begin{aligned}
\text{Dynamics Prediction:} \quad & \widehat{v}_{n+1} = \Psi(v_n) + \xi_n, \\
\text{Data Prediction:} \quad & \widehat{y}_{n+1} = h(\widehat{v}_{n+1}) + \eta_{n+1}, \\
\text{Perfect Transport:} \quad & v_{n+1} = T^S(\widehat{v}_{n+1}, \widehat{y}_{n+1}; \nu_{n+1}, y^\dagger_{n+1}).
\end{aligned}
$$

Transport Chosen To Effect Conditioning

$$
\begin{aligned}
\text{Assumption:} \quad & v_n \sim \mu_n \qquad (v_n^\dagger | Y_n) \\
\text{Dynamics and Data:} \quad & (\widehat{v}_{n+1}, \widehat{y}_{n+1}) \sim \nu_{n+1} \\
\text{Conditioning:} \quad & v_{n+1} \sim \mu_{n+1} \qquad (v_{n+1}^\dagger | Y_{n+1})
\end{aligned}
$$

Transport-based Inversion and Conditioning Marzouk et al [31], [27]

Mean Field Dynamics

Particle Approximation – Linear Markov Process

$$\begin{aligned}
\widehat{v}_{n+1}^{(j)} &= \Psi(v_n^{(j)})) + \xi_n^{(j)}, \\
\widehat{y}_{n+1}^{(j)} &= h(\widehat{v}_{n+1}^{(j)})) + \eta_{n+1}^{(j)}, \\
v_{n+1}^{(j)} &= T^S(\widehat{v}_{n+1}^{(j)}, \widehat{y}_{n+1}^{(j)}; \nu_{n+1}^J, y_{n+1}^\dagger), \\
\nu_{n+1}^J &= \frac{1}{J}\sum_{j=1}^{J} \delta_{(\widehat{v}_{n+1}^{(j)}, \widehat{y}_{n+1}^{(j)})}
\end{aligned}$$

$$(v_n^\dagger | Y_n) \qquad \mu_n \approx \mu_n^J = \frac{1}{J}\sum_{j=1}^{J} \delta_{v_n^{(j)}}$$

Mean Field Dynamics

Particle Approximation – Linear Markov Process

$$
\begin{aligned}
\widehat{v}_{n+1}^{(j)} &= \Psi(v_n^{(j)})) + \xi_n^{(j)},\\
\widehat{y}_{n+1}^{(j)} &= h(\widehat{v}_{n+1}^{(j)})) + \eta_{n+1}^{(j)},\\
v_{n+1}^{(j)} &= T^S(\widehat{v}_{n+1}^{(j)}, \widehat{y}_{n+1}^{(j)}; \nu_{n+1}^J, y_{n+1}^\dagger),\\
\nu_{n+1}^J &= \frac{1}{J}\sum_{j=1}^J \delta_{(\widehat{v}_{n+1}^{(j)}, \widehat{y}_{n+1}^{(j)})}
\end{aligned}
$$

$$
(v_n^\dagger | Y_n) \qquad \mu_n \approx \mu_n^J = \frac{1}{J}\sum_{j=1}^J \delta_{v_n^{(j)}}
$$

Equal Weights: No Collapse

Mean Field Dynamics

Particle Approximation – Linear Markov Process

$$
\begin{aligned}
\widehat{v}_{n+1}^{(j)} &= \Psi(v_n^{(j)})) + \xi_n^{(j)}, \\
\widehat{y}_{n+1}^{(j)} &= h(\widehat{v}_{n+1}^{(j)})) + \eta_{n+1}^{(j)}, \\
v_{n+1}^{(j)} &= T^S(\widehat{v}_{n+1}^{(j)}, \widehat{y}_{n+1}^{(j)}; \nu_{n+1}^J, y_{n+1}^{\dagger}), \\
\nu_{n+1}^J &= \frac{1}{J}\sum_{j=1}^{J} \delta_{(\widehat{v}_{n+1}^{(j)}, \widehat{y}_{n+1}^{(j)})}
\end{aligned}
$$

$$
(v_n^{\dagger} | Y_n) \qquad \mu_n \approx \mu_n^J = \frac{1}{J}\sum_{j=1}^{J} \delta_{v_n^{(j)}}
$$

Equal Weights: No Collapse

But: Computation of T^S Prohibitive

Filtering and Ensemble Kalman

- Original Kalman Paper Kalman [26]
- Original Ensemble Kalman Paper Evensen [19]
- Link To Transport Reich [34]

Mean Field Kalman Dynamics

Prediction and Kalman Transport – Nonlinear Markov Process

$$\begin{aligned} \text{Dynamics Prediction:} \quad & \widehat{v}_{n+1} = \Psi(v_n) + \xi_n, \\ \text{Data Prediction:} \quad & \widehat{y}_{n+1} = h(\widehat{v}_{n+1}) + \eta_{n+1}, \end{aligned}$$

Mean Field Kalman Dynamics

Prediction and Kalman Transport – Nonlinear Markov Process

$$\text{Dynamics Prediction:}\quad \widehat{v}_{n+1} = \Psi(v_n) + \xi_n,$$
$$\text{Data Prediction:}\quad \widehat{y}_{n+1} = h(\widehat{v}_{n+1}) + \eta_{n+1},$$

Transport $\mathbb{E} := \mathbb{E}^{\nu_{n+1}}$

$$\text{Transport:}\quad v_{n+1} = \widehat{v}_{n+1} + \widehat{C}^{vy}_{n+1}\big(\widehat{C}^{yy}_{n+1}\big)^{-1}\big(y^{\dagger}_{n+1} - \widehat{y}_{n+1}\big).,$$
$$\text{Data Covariance:}\quad \widehat{C}^{yy}_{n+1} = \mathbb{E}\Big(\big(\widehat{y}_{n+1} - \mathbb{E}\widehat{y}_{n+1}\big) \otimes \big(\widehat{y}_{n+1} - \mathbb{E}\widehat{y}_{n+1}\big)\Big),$$
$$\text{Cross Covariance:}\quad \widehat{C}^{vy}_{n+1} = \mathbb{E}\Big(\big(\widehat{v}_{n+1} - \mathbb{E}\widehat{v}_{n+1}\big) \otimes \big(\widehat{y}_{n+1} - \mathbb{E}\widehat{y}_{n+1}\big)\Big).$$

Perfect Conditioning Via Transport For Gaussian ν_{n+1}.

Mean Field Kalman Dynamics

Particle Approximation – Linear Markov Process

$$\begin{aligned}
\widehat{v}^{(j)}_{n+1} &= \Psi(v^{(j)}_n) + \xi^{(j)}_n, \\
\widehat{y}^{(j)}_{n+1} &= h(\widehat{v}^{(j)}_{n+1}) + \eta^{(j)}_{n+1}, \\
\nu^J_{n+1} &= \frac{1}{J}\sum_{j=1}^{J} \delta_{(\widehat{v}^{(j)}_{n+1}, \widehat{y}^{(j)}_{n+1})}.
\end{aligned}$$

Empirical Covariances; $\mathbb{E} := \mathbb{E}^{\nu^J_{n+1}}$

Kalman Transport: $v^{(j)}_{n+1} = \widehat{v}^{(j)}_{n+1} + \widehat{C}^{vy}_{n+1}(\widehat{C}^{yy}_{n+1})^{-1}(y^{\dagger}_{n+1} - \widehat{y}^{(j)}_{n+1})$,

Data Covariance: $\widehat{C}^{yy}_{n+1} = \mathbb{E}\Big(\big(\widehat{y}_{n+1} - \mathbb{E}\widehat{y}_{n+1}\big) \otimes \big(\widehat{y}_{n+1} - \mathbb{E}\widehat{y}_{n+1}\big)\Big)$,

Cross Covariance: $\widehat{C}^{vy}_{n+1} = \mathbb{E}\Big(\big(\widehat{v}_{n+1} - \mathbb{E}\widehat{v}_{n+1}\big) \otimes \big(\widehat{y}_{n+1} - \mathbb{E}\widehat{y}_{n+1}\big)\Big)$.

Particle Filter

Theorem Le Gland et al [30]

Assume Ψ, h are linear. Then there is $C(N) > 0$ such that, for all $1 \leq n \leq N$,

$$d_\phi(\mu_n, \mu_n^J) \leq C(N) \frac{1}{\sqrt{J}}.$$

$$\mu_n \approx \mu_n^J = \frac{1}{J} \sum_{j=1}^{J} \delta_{\widehat{v}_n^{(j)}}$$

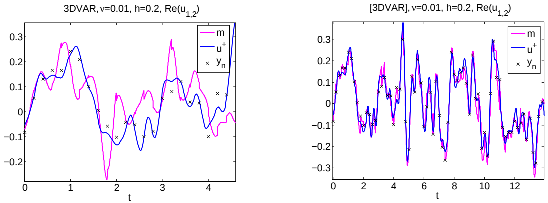
For locally Lipschitz ϕ, with polynomial growth:

$$d_\phi(\pi, \pi') = \left(\mathbb{E}\Big[\big(\pi(f) - \pi'(f)\big)^p\Big]\right)^{1/p},$$

Weather Forecasting

- Evaluation of Filters Law and AMS [29]
- Filters in Geophysical Applications van Leeuwen et al [40]

3DVAR ($\equiv$ Averaged ExKF) Overcomes Butterfly Effect

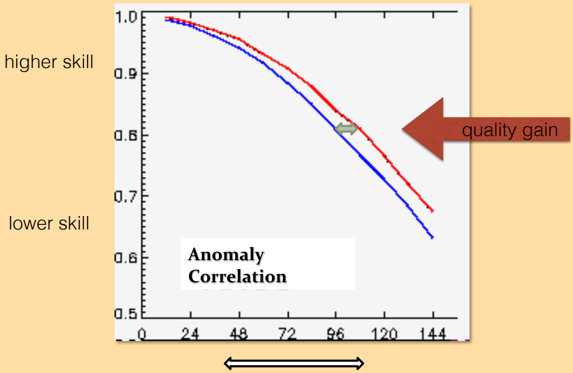


ExKF Jazwinski [23] 3DVAR Law et al [28]

Impact of EnKF over 3DVAR

courtesy Roland Potthast(DWD)

Ensemble Kalman Filter (red) versus 3DVAR (blue)



forecast lead time in hours

Filtering and Inverse Problems

- Optimization Approach Engl et al [17]
- Bayesian Approach Kaipio and Somersalo [25]
- Bayesian Approach (Banach Space) AMS [39]
- Ensemble Sampling and Optimization Reich [34]
- Ensemble Sampling Chen and Oliver [9]
- Ensemble Sampling Emerick and Reynolds [16]
- Ensemble Optimization Iglesias et al [22]
- Ensemble Optimization With Constraints Albers et al [1]
- Analysis of Ensemble Sampling Ernst et al [18]

Inverse Problem

Problem Statement

Find u from y where $\mathsf{G} : \mathcal{U} \mapsto \mathcal{Y}$, $\eta \sim N(0, \Gamma)$ is noise and

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

- $|\cdot|$ Euclidean norm.
- $|\cdot|_A = |A^{-\frac{1}{2}} \cdot|$ for $A > 0$.

Inverse Problem

Bayesian Approach

$$\text{Objective} \quad \Phi_0(u) = \frac{1}{2}|y - G(u)|_\Gamma^2,$$

$$\text{Prior} \quad \mu_0(du),$$

$$\text{Posterior} \quad \mu(du) = \frac{1}{Z}\exp\big(-\Phi_0(u)\big)\mu_0(du).$$

- $|\cdot|$ Euclidean norm.
- $|\cdot|_A = |A^{-\frac{1}{2}}\cdot|$ for $A > 0$.

Inverse Problem

Sequential Monte Carlo – SMC

$$\text{Sequential Updates 1} \quad \mu_n(du) = \frac{1}{Z_n} \exp\big(-nh\Phi_0(u)\big)\mu_0(du),$$

$$\text{Sequential Updates 2} \quad \mu_{n+1}(du) = \frac{Z_n}{Z_{n+1}} \exp\big(-h\Phi_0(u)\big)\mu_n(du),$$

$$\text{Posterior} \quad \mu(du) = \mu_N(du), \quad Nh = 1.$$

Del Moral et al [13] Beskos et al [4] Chopin and Papaspiliopoulos [10]

Hidden Markov Model

Dynamics and Data

$$\text{Dynamics Model:} \quad v^\dagger_{n+1} = v^\dagger_n, \quad n \in \mathbb{Z}^+$$

$$\text{Data Model:} \quad y^\dagger_{n+1} = G(v^\dagger_{n+1}) + \eta_{n+1}, \quad n \in \mathbb{Z}^+$$

$$\text{Probabilistic Structure:} \quad v^\dagger_0 \sim \mu_0, \quad \eta_n \sim N(0, \frac{1}{h}\Gamma)$$

$$\text{Probabilistic Structure:} \quad v^\dagger_0 \perp \{\eta_n\} \text{ independent}$$

$$v^\dagger_n | Y_n \sim \mu_n$$

$$v^\dagger_N | Y_N \sim \mu$$

Continuous Time Limit

$$Nh = 1, \quad h \to 0; \quad nh = t.$$

$$\mu_n \approx \mu(t), \quad v_n \approx u(t).$$

- Optimization Schillings and AMS [37]
- Sampling Garbuno-Inigo et al [20]

Ensemble Kalman Inversion (EKI)

Continuous Time Formulation $\mathbb{E} = \mathbb{E}^{u' \sim \mu}$ Schillings and AMS [37]

$$\dot{u} = -\mathbb{E}\Big(\Big\langle \mathsf{G}(u') - \bar{\mathsf{G}}, \mathsf{G}(u) - y + \sqrt{\Gamma'}\dot{B}\Big\rangle_{\Gamma} (u' - \bar{u})\Big), \quad u(0) \sim \mu_0$$

$$\bar{u} = \mathbb{E}u' \quad \bar{\mathsf{G}} = \mathbb{E}(u').$$

Theorem Reich [34] Garbuno-Inigo et al [20]

Let G be linear and $\Gamma' = \Gamma$. Then $\mu|_{t=1} = \mu$, solution of the Bayesian inverse problem.

- $\Gamma' = \Gamma$ is continuous limit of ensemble Kalman SMC.
- $\Gamma' = 0$ gives optimization – EKI.

Connection to Optimization – Linear Approximation

Linear Approximation

$$(\mathsf{G}(u') - \bar{\mathsf{G}}) \approx D\mathsf{G}(u)(u' - \bar{u}).$$

EKI As Self-Preconditioned Gradient Descent See [34], [37]

$$\dot{u} = -C(\mu)\nabla\Phi_0(u),$$
$$C(\mu) = \mathbb{E}\Big((u' - \bar{u}) \otimes (u' - \bar{u})\Big),$$
$$u \sim \mu, \quad \Phi_0(u) = \frac{1}{2}\big|y - G(u)\big|_\Gamma^2.$$

Ensemble Kalman Sampling (EKS)

Continuous Time Formulation: Put EKI in a heat bath

$$\dot{u} = -\mathbb{E}\Big(\Big\langle \mathsf{G}(u') - \bar{\mathsf{G}}, \mathsf{G}(u) - y + \sqrt{\Gamma'}\dot{B}\Big\rangle_{\Gamma} (u' - \bar{u})\Big) - C(\mu)\Sigma^{-1}u + \sqrt{2C(\mu)}\dot{W},$$

$$C(\mu) = \mathbb{E}\Big((u' - \bar{u}) \otimes (u' - \bar{u})\Big).$$

Ensemble Kalman Sampling – Linear Approximation

Linear Approximation

$$(\mathsf{G}(u') - \bar{\mathsf{G}}) \approx D\mathsf{G}(u)(u' - \bar{u}),$$
$$\mu_0 = N(0, \Sigma).$$

EKS As Self-Preconditioned Langevin Equation See [20], [21]

$$\dot{u} = -C(\mu)\nabla\Phi(u) + \sqrt{2C(\mu)}\dot{W},$$
$$C(\mu) = \mathbb{E}\Big((u' - \bar{u}) \otimes (u' - \bar{u})\Big),$$
$$\Phi(u) = \frac{1}{2}|y - G(u)|_\Gamma^2 + \frac{1}{2}|u|_\Sigma^2,$$
$$= \Phi_0(u) + \frac{1}{2}|u|_\Sigma^2.$$

Nonlinear Nonlocal Fokker-Planck Equation

Theorem Garbuno-Inigo et al [20]

Measure μ has density ρ solving a nonlinear, nonlocal Fokker-Planck equation:

$$\partial_t \rho = \nabla \cdot \left(\rho \mathcal{C}(\rho) \nabla \frac{\delta \mathcal{E}}{\delta \rho} \right) \, , \mathcal{E}(\rho) = \int (\Phi + \ln \rho)\, \rho \, \mathrm{d}u.$$

Gradient flow in $\mathcal{P}_+$ (probability measures) w.r.t. metric $g_{\rho,\mathcal{C}}$ (on the tangent space):

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t} \mathcal{E}(\rho) &= -\int \rho \left| \mathcal{C}(\rho)^{\frac{1}{2}} \nabla (\Phi + \ln \rho) \right|^2 \mathrm{d}u \\ &= -g_{\rho,\mathcal{C}}(\partial_t \rho, \partial_t \rho). \end{aligned}$$

Builds on work of: Otto: [24, 32]; Cotter and Reich: [35]

Nonlinear Nonlocal Fokker-Planck Equation

Theorem Garbuno-Inigo et al [20]

Let G be linear and $\mu(0)$ be Gaussian. Then $\mu(t) \to \mu$ in L^1 (μ solution of the Bayesian inverse problem) at universal rate $\exp(-t)$.

Extension to non-Gaussian initialization: Carrillo and Vaes [6]

Electrical Impedence Tomography

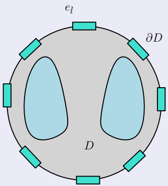
- Bayesian Formulation Dunlop and AMS [15]
- Ensemble Kalman Approach Chada et al [8]

Electrical Impedance Tomography (EIT) 1

Forward Problem

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

$$\begin{aligned} -\nabla \cdot (\kappa \nabla \nu) = 0 \quad &\in \quad D, \\ \nu + z_\ell \kappa \nabla \nu \cdot n = V_\ell \quad &\in \quad e_\ell, \quad \ell = 1, \dots, m, \\ \nabla \nu \cdot n = 0 \quad &\in \quad \partial D \backslash \cup_{\ell=1}^m e_\ell, \\ \int \kappa \nabla \nu \cdot n \, ds = I_\ell \quad &\in \quad e_\ell, \quad \ell = 1, \dots, m. \end{aligned}$$



Ohm's Law: $V = R(\kappa) \times I$.

Inverse Problem

Set $\kappa = \exp(u)$. Given a set of K noisy measurements of voltage $V(k)$ from currents $I(k)$, and $\mathsf{G}_k(u) = R(\exp(u)) \times I(k)$, find u from y where:

$$y(k) = \mathsf{G}_k(u) + \eta, \quad \eta \sim \mathsf{N}(0, \gamma^2), \quad k = 1, \dots, K.$$

EIT 2

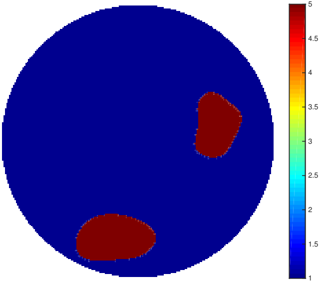


Figure: True Conductivity.

Parameterization

- Continuous level set function.
- Lengthscale of level set function.
- Smoothness of level set function.

EIT 3

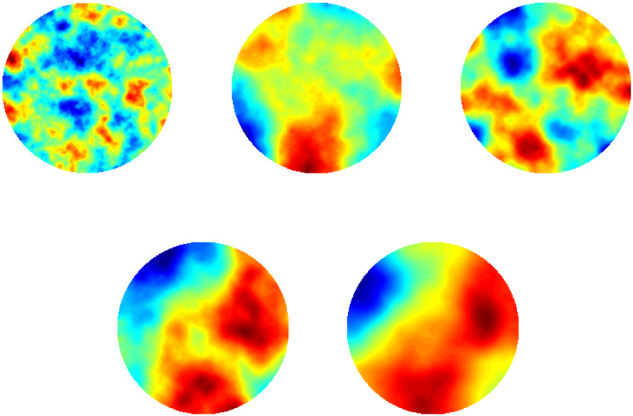


Figure: Five succesive iterations: level set function.

EIT 4

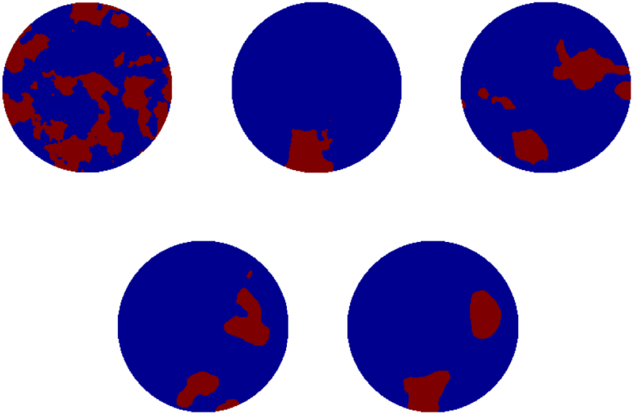


Figure: Five succesive iterations: thresholded level set function.

Closing

Conclusions: Ensemble Kalman Methodologies

- Kalman Filtering: 1960, Rudolph Kalman.
- Ensemble Kalman Filtering: 1994, Geir Evensen.
- Applications in numerous fields:
  - Weather forecasting;
  - Oceanography;
  - Hydrology, Subsurface flow;
  - Medical imaging, Machine learning $\cdots$.
- Developing as a general methodology for state estimation.
- Developing as a general methodology for inverse problems:
  - Gradient flow structure: parameter space;
  - Gradient flow structure: probability space.
- Connections to Wasserstein gradient flows, optimal transport.
- Many open mathematical questions.

References I

[1] D. J. Albers, P.-A. Blancquart, M. E. Levine, E. E. Seylabi, and A. Stuart.
Ensemble Kalman methods with constraints.
Inverse Problems, 35(9):095007, 2019.

[2] A. Bain and D. Crisan.
Fundamentals of stochastic filtering, volume 60.
Springer Science & Business Media, 2008.

[3] J.-D. Benamou and Y. Brenier.
A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem.
Numerische Mathematik, 84(3):375–393, 2000.

[4] A. Beskos, A. Jasra, E. A. Muzaffer, and A. M. Stuart.
Sequential Monte Carlo methods for bayesian elliptic inverse problems.
Statistics and Computing, 25(4):727–737, 2015.

[5] J. Carrillo, F. Hoffmann, A. Stuart, and U. Vaes.
Consensus based sampling.
arXiv preprint arXiv:2106.02519, 2021.

[6] J. Carrillo and U. Vaes.
Wasserstein stability estimates for covariance-preconditioned fokker–planck equations.
Nonlinearity, 34(4):2275, 2021.

[7] J. A. Carrillo, Y.-P. Choi, C. Totzeck, and O. Tse.
An analytical framework for consensus-based global optimization method.
Mathematical Models and Methods in Applied Sciences, 28(06):1037–1066, 2018.

[8] N. K. Chada, M. A. Iglesias, L. Roininen, and A. M. Stuart.
Parameterizations for ensemble Kalman inversion.
Inverse Problems, 34(5):055009, 2018.

References II

[9] Y. Chen and D. S. Oliver.
Ensemble randomized maximum likelihood method as an iterative ensemble smoother.
Mathematical Geosciences, 44(1):1–26, 2012.

[10] N. Chopin, O. Papaspiliopoulos, et al.
An introduction to sequential Monte Carlo, volume 4.
Springer, 2020.

[11] D. Crisan and J. Xiong.
Approximate mckean–vlasov representations for a class of SPDEs.
Stochastics An International Journal of Probability and Stochastics Processes, 82(1):53–68, 2010.

[12] F. Daum, J. Huang, and A. Noushin.
Exact particle flow for nonlinear filters.
In *Signal processing, sensor fusion, and target recognition XIX*, volume 7697, page 769704. International society for optics and photonics, 2010.

[13] P. Del Moral, A. Doucet, and A. Jasra.
Sequential monte carlo samplers.
Journal of the Royal Statistical Society: Series B (Statistical Methodology), 68(3):411–436, 2006.

[14] A. Doucet, N. De Freitas, N. J. Gordon, et al.
Sequential Monte Carlo methods in practice, volume 1.
Springer, 2001.

[15] M. M. Dunlop and A. M. Stuart.
The bayesian formulation of EIT: analysis and algorithms.
Inverse Problems and Imaging, 10(4):1007–1036, 2016.

[16] A. A. Emerick and A. C. Reynolds.
Ensemble smoother with multiple data assimilation.
Computers & Geosciences, 55:3–15, 2013.

References III

[17] H. W. Engl, M. Hanke, and A. Neubauer.
Regularization of inverse problems, volume 375.
Springer Science & Business Media, 1996.

[18] O. G. Ernst, B. Sprungk, and H.-J. Starkloff.
Analysis of the ensemble and polynomial chaos Kalman filters in bayesian inverse problems.
SIAM/ASA Journal on Uncertainty Quantification, 3(1):823–851, 2015.

[19] G. Evensen.
Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics.
Journal of Geophysical Research: Oceans, 99(C5):10143–10162, 1994.

[20] A. Garbuno-Inigo, F. Hoffmann, W. Li, and A. M. Stuart.
Interacting Langevin diffusions: Gradient structure and ensemble Kalman sampler.
SIAM Journal on Applied Dynamical Systems, 19(1):412–441, 2020.

[21] A. Garbuno-Inigo, N. Nüsken, and S. Reich.
Affine invariant interacting Langevin dynamics for bayesian inference.
SIAM Journal on Applied Dynamical Systems, 19(3):1633–1658, 2020.

[22] M. A. Iglesias, K. J. Law, and A. M. Stuart.
Ensemble Kalman methods for inverse problems.
Inverse Problems, 29(4):045001, 2013.

[23] A. Jazwinski.
Stochastic Processes and Filtering Theory.
Academic Press, 1970.

[24] R. Jordan, D. Kinderlehrer, and F. Otto.
The variational formulation of the fokker–planck equation.
SIAM journal on mathematical analysis, 29(1):1–17, 1998.

References IV

[25] J. Kaipio and E. Somersalo.
Statistical and computational inverse problems, volume 160.
Springer Science & Business Media, 2006.

[26] R. Kalman.
A new approach to linear filtering and prediction problems.
Journal of Basic Engineering, 82:35–45, 1960.

[27] N. Kovachki, R. Baptista, B. Hosseini, and Y. Marzouk.
Conditional sampling with monotone gans.
arXiv preprint arXiv:2006.06755, 2020.

[28] K. Law, A. Stuart, and K. Zygalakis.
Data assimilation.
Cham, Switzerland: Springer, 2015.

[29] K. J. Law and A. M. Stuart.
Evaluating data assimilation algorithms.
Monthly Weather Review, 140(11):3757–3782, 2012.

[30] F. Le Gland, V. Monbet, and V.-D. Tran.
Large sample asymptotics for the ensemble Kalman filter.
2009.

[31] Y. Marzouk, T. Moselhy, M. Parno, and A. Spantini.
Sampling via measure transport: An introduction.
Handbook of uncertainty quantification, pages 1–41, 2016.

[32] F. Otto.
The geometry of dissipative evolution equations: the porous medium equation.
2001.

References V

[33] P. Rebeschini and R. Van Handel.
Can local particle filters beat the curse of dimensionality?
The Annals of Applied Probability, 25(5):2809–2866, 2015.

[34] S. Reich.
A dynamical systems framework for intermittent data assimilation.
BIT Numerical Mathematics, 51(1):235–249, 2011.

[35] S. Reich and C. Cotter.
Probabilistic forecasting and Bayesian data assimilation.
Cambridge University Press, 2015.

[36] S. Reich and C. J. Cotter.
Ensemble filter techniques for intermittent data assimilation.
Large Scale Inverse Problems. Computational Methods and Applications in the Earth Sciences, 13:91–134, 2013.

[37] C. Schillings and A. M. Stuart.
Analysis of the ensemble Kalman filter for inverse problems.
SIAM Journal on Numerical Analysis, 55(3):1264–1290, 2017.

[38] A. Spantini, R. Baptista, and Y. Marzouk.
Coupling techniques for nonlinear ensemble filtering.
arXiv preprint arXiv:1907.00389, 2019.

[39] A. M. Stuart.
Inverse problems: a bayesian perspective.
Acta numerica, 19:451–559, 2010.

[40] P. J. Van Leeuwen, H. R. Künsch, L. Nerger, R. Potthast, and S. Reich.
Particle filters for high-dimensional geoscience applications: A review.
Quarterly Journal of the Royal Meteorological Society, 145(723):2335–2365, 2019.

References VI

[41] T. Yang, P. G. Mehta, and S. P. Meyn.
Feedback particle filter.
IEEE transactions on Automatic control, 58(10):2465–2480, 2013.

Metric For Gradient Structure

Otto: [24, 32], Cotter and Reich: [35]

Kalman-Wasserstein Metric Tensor (Otto [36],[20])

Define $g_{\rho,\mathcal{C}} : T_\rho\mathcal{P}_+ \times T_\rho\mathcal{P}_+ \to \mathbb{R}$ by

$$g_{\rho,\mathcal{C}}(\sigma_1,\sigma_2) := \int_\Omega \langle \nabla\psi_1 \, , \, \mathcal{C}(\rho)\nabla\psi_2 \rangle \; \rho \, \mathrm{d}x,$$

where $\sigma_i = -\nabla \cdot (\rho\mathcal{C}(\rho)\nabla\psi_i) \in T_\rho\mathcal{P}_+$ for $i = 1, 2$.

Kalman-Wasserstein Metric (Benamou-Brenier [3])

For ρ^0, $\rho^1 \in \mathcal{P}_+$, $\mathcal{W}_\mathcal{C} : \mathcal{P}_+ \times \mathcal{P}_+ \to \mathbb{R}$

$$\mathcal{W}_\mathcal{C}(\rho^0,\rho^1)^2 := \inf_{(\rho_t,\psi_t)} \int_0^1 \int_\Omega \langle \nabla\psi_t \, , \, \mathcal{C}(\rho_t)\nabla\psi_t \rangle \; \rho_t \, \mathrm{d}x$$

$$\text{subject to} \quad \partial_t\rho_t + \nabla \cdot (\rho_t\mathcal{C}(\rho_t)\nabla\psi_t) = 0, \;\; \rho_0 = \rho^0, \;\; \rho_1 = \rho^1,$$