Inverse Problems Without Adjoint: Ensemble Approaches

Andrew Stuart

Computing and Mathematical Sciences
California Institute of Technology

AFOSR, DARPA, EPSRC, NSF, ONR
Allen Philanthropies, Mission Control for Earth, Schmidt Futures

***

Sayas Numerics Seminar

April 13th 2021
Overview

Ensemble Kalman Methods For Inverse Problems

Mathematical Structure

Guiding Examples

Rough Forward Models

Refineable Ensemble Methods

Closing
Ensemble Kalman Methods
For Inverse Problems

- Reich [38] (Assimilation step in data assimilation)
- Chen & Oliver [11] (Randomized maximum likelihood)
- Emerick and Reynolds [14] (Iterative ensemble smoother)
- Ernst, Sprungk and Starkloff [15] (Limitation in non-Gaussian setting)
- Iglesias [24] (Stopping rules for EKI)
- Evensen [16] (Iterative ensemble smoothers)
- Blömker, Schillings and Wacker [8], [9] (Numerical analysis perspective)
- Schneider, S and Wu [44] (Learning SDEs w/EKI)
- Schneider, S and Wu [43] (Sparsity w/EKI)
Inverse Problem

Problem Statement

Find \( u \) from \( y \) where \( G : \mathcal{U} \rightarrow \mathcal{Y} \), \( \eta \sim \mathcal{N}(0, \Gamma) \) is noise and \( y = G(u) + \eta \).

Main Approaches

Optimization
\[
\Phi(u) = \frac{1}{2} |y - G(u)|_F^2 + \frac{1}{2} |u|_\Sigma^2;
\]

Probability
\[
P(u|y) \propto \exp(-\Phi(u)).
\]
Inverse Problem

Problem Statement

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

$$y = G(u) + \eta.$$

Main Approaches

Optimization

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

Probability

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

Here $\langle \cdot, \cdot \rangle_A = \langle \cdot, A^{-1} \cdot \rangle$ and $|\cdot|_A = |A^{-\frac{1}{2}} \cdot|$. 
Inverse Problem

Filtering Dynamical Systems

Dynamics Model: \( v_{n+1} = \Psi(v_n) + N(0, \Sigma), \quad n \in \mathbb{Z}^+ \)

Data Model: \( y_{n+1} = Hv_{n+1} + N(0, \Gamma), \quad n \in \mathbb{Z}^+ \)

State Estimation: \( v_n \mid \{ y_\ell \}_{\ell=1}^n \).

Inverse Problem: Dynamical Formulation

Dynamics Model: \( u_{n+1} = u_n, \quad n \in \{0, \cdots, M - 1\} \)

Dynamics Model: \( w_{n+1} = G(u_n), \quad n \in \{0, \cdots, M - 1\} \)

Data Model: \( y_{n+1} = w_{n+1} + N(0, M\Gamma), \quad n \in \{0, \cdots, M - 1\} \)

Parameter Estimation: \( u_n \mid \{ y_\ell = y \}_{\ell=1}^n \)
Discrete Time: Ensemble Kalman Inversion

Covariances

\[ C_{ww}^{n} = \frac{1}{J} \sum_{j=1}^{J} (G(u_{n}^{(j)}) - \bar{w}_{n}) \otimes (G(u_{n}^{(j)}) - \bar{w}_{n}), \quad \bar{w}_{n} = \frac{1}{J} \sum_{j=1}^{J} G(u_{n}^{(j)}), \]

\[ C_{uw}^{n} = \frac{1}{J} \sum_{j=1}^{J} (u_{n}^{(j)} - \bar{u}_{n}) \otimes (G(u_{n}^{(j)}) - \bar{w}_{n}), \quad \bar{u}_{n} = \frac{1}{J} \sum_{j=1}^{J} u_{n}^{(j)}. \]

Iteration \( n \mapsto n + 1 \)

\[ u_{n+1}^{(j)} = u_{n}^{(j)} + C_{n}^{uw} (C_{n}^{ww} + M\Gamma)^{-1} (y - G(u_{n}^{(j)})) \]
Continuous Time: Ensemble Kalman Inversion

Means

\[ \bar{u}(t) = \frac{1}{J} \sum_{j=1}^{J} u^{(j)}(t), \]

\[ \bar{G}(t) = \frac{1}{J} \sum_{j=1}^{J} G(u^{(j)}(t)). \]

Continuous Time Limit

\[ u^{(j)}_n \approx u^{(j)}(t) \big|_{t=n/M} \]

\[ \dot{u}^{(j)} = -\frac{1}{J} \sum_{k=1}^{J} \left< G(u^{(k)}) - \bar{G}, G(u^{(j)}) - y \right> \Gamma \left( u^{(k)} - \bar{u} \right) \]
Mathematical Structure
Gradient Flow In Parameter Space

- Ensemble Filtering Continuous Time: Reich (2011) [38]
- Connection to Foais/Prodi: Titi and coworkers [21, 2]
- Ensemble Inversion Continuous Time: Schillings & S (2017) [42]
- Ensemble Filtering Continuous Time: Lange & Stannat [30]
- Ensemble Square Root Filtering Continuous Time: Lange & Stannat [29]
- Tikhonov Regularization: Chada, S & Tong [10]
Gradient Flow In Space
Of Probability Measures

- Jordan, Kinderlehrer & Otto 1998 [26]
- Otto 2001 [34]
- Benamou & Brenier 2000 [3]
- Ambrosio, Gigli & Savare 2008 [1]
- Villani 2008 [45]
- Reich & Cotter 2013 [40]
- Garbuno-Inigo, Hoffmann, Li & Stuart 2020 [17]
- Garbuno-Inigo, Nüsken & Reich [18]
Ensemble Kalman Sampling (EKS)

Continuous Time Formulation: Put EKI in a heat bath

\[ \dot{u}^{(j)} = -\frac{1}{J} \sum_{k=1}^{J} \langle G(u^{(k)}) - \bar{G}, G(u^{(j)}) - y \rangle_{\Gamma} \left( u^{(k)} - \bar{u} \right) \]
\[ - C(u)\Sigma^{-1} u^{(j)} + \sqrt{2C(u)} \dot{W}^{(j)}. \]

EKS Is Self-Preconditioned Langevin Equation (Linear G)

\[ \dot{u}^{(j)} = -C(u)\nabla \Phi(u^{(j)}) + \sqrt{2C(u)} \dot{W}^{(j)}, \quad \Phi(u) = \frac{1}{2} |y - G(u)|_{\Gamma}^2 + \frac{1}{2} |u|_{\Sigma}^2, \]
\[ \bar{u} = \frac{1}{J} \sum_{k=1}^{J} u^{(k)}, \quad C(u) = \frac{1}{J} \sum_{k=1}^{J} \left( u^{(k)} - \bar{u} \right) \otimes \left( u^{(k)} - \bar{u} \right). \]
Self-Preconditioned Langevin Equation \[17\]

**Mean Field Limit: Nonlinear Nonlocal Fokker-Planck Eq.**

\[
\dot{u} = -C(\rho) \nabla \Phi(u) + \sqrt{2C(\rho)} \dot{W}, \\
C(\rho) = \int (u - \bar{u}) \otimes (u - \bar{u}) \rho(u, t) du, \hspace{1cm} \bar{u} = \int u \rho(u, t) du, \\
\partial_t \rho = \nabla \cdot (\rho C(\rho) \nabla \Phi) + C(\rho) : D^2 \rho, \hspace{1cm} \rho(0) = \rho_0.
\]

**Theorem** \[26],[17\]

The nonlinear nonlocal Fokker-Planck equation may be written as

\[
\partial_t \rho = \nabla \cdot \left( \rho C(\rho) \nabla \frac{\delta \mathcal{E}}{\delta \rho} \right), \hspace{1cm} \mathcal{E}(\rho) = \int (\Phi + \ln \rho) \rho du.
\]

\[\rho_\infty(u) := \exp(-\Phi(u))\] is a steady-state of the Fokker-Planck equation.

**Theorem** \[17\]

\[\text{G linear then } \|\rho(\cdot, t) - \rho_\infty\|_{L^1} \leq C \exp(-t) \quad \text{(independent of G)}\]
Guiding Examples

- Learn dynamical systems from time-averaged data [12]
- Simplified GCM [35, 22]
- Unscented Kalman Filtering [27]
- Unscented Kalman Inversion [23]
Data From Dynamics

Time-Averaged Data

\[
\frac{dv}{dt} = F(v; u), \quad v(0) = v_0,
\]

\[
y = G_T(u; v_0) = \frac{1}{T} \int_0^T \varphi(v(t)) \, dt.
\]

Central Limit Theorem

\[
G_T(u; v_0) = G(u) + \frac{1}{\sqrt{T}} N(0, \Sigma),
\]

\[
y = G(u) + \frac{1}{\sqrt{T}} N(0, \Sigma).
\]
### Example 1 – 3D NS With Hydrostatic Assumption

#### Governing Dynamics

\[
\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{v}) + \frac{\partial \rho \omega}{\partial z} = 0
\]

\[
D \frac{\mathbf{v}}{dt} + \Omega \mathbf{k} \times \mathbf{v} + \nabla p \rho + \nabla \Phi = F
\]

\[
\frac{D T}{dt} - \frac{RT \omega}{C_p \rho} = Q
\]

\[
\frac{\partial p}{\partial z} = -\rho g
\]

\[
p = \rho RT
\]

- \(\rho\) : fluid density; \(\mathbf{v}, \omega\) : horizontal and vertical velocities;
- \(T\) : temperature; \(p\) : pressure, \(\Phi\) : geopotential; \(k\) unit vertical.
- \(D/Dt\) represents the derivative following a fluid parcel.
- \(Q\) : radiation, to be learned.
Closure Model (Radiation Model)

\[ Q = -k_T(\phi, \sigma)(T - T_{eq}(\phi, p)) \]

\[ k_T = k_a + (k_s - k_a) \max\left(0, \frac{\sigma - \sigma_b}{1 - \sigma_b}\right) \cos^4 \phi \]

\[ T_{eq} = \max\left\{ 200K, \left[315K - \Delta T_y \sin^2 \phi - \Delta \theta_z \log\left(\frac{p}{p_0}\right) \cos^2 \phi\right]\left(\frac{p}{p_0}\right)^\kappa \right\} \]

\[ k_a = 1/40 \text{ day}^{-1} \quad k_s = 1/4 \text{ day}^{-1} \quad \Delta T_y = 60 \text{ K} \quad \Delta \theta_z = 10 \text{ K} \]

Here \( \sigma \) is the vertical coordinate, \( \phi \) is the latitude, \( p_0 = 10^5 Pa \) is the reference sea-level pressure, and \( \kappa = \frac{R}{C_p} \).
Zonally/Temporally Averaged Temperature

Data used for training
Convergence History

Behavior of UKI
Rough Forward Models

- Affine invariance and ensemble samplers: Goodman and Weare [19]
- Other ensemble samplers: Leimkuhler, Matthews and Weare [32]
- Ensemble GP samplers: Reich and co-workers [33, 41]
- Multiscale analysis: Duncan, S & Wolfram [13]
- Related analysis for MCMC: Plechac and Simpson [37]
Problem Statement

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

$$y = G(u) + \eta.$$  

Main Approaches

**Optimization**

$$\Phi(u) = \frac{1}{2} |y - G(u)|^2_f + \frac{1}{2} |u|^2_\Sigma;$$

**Probability**

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

Here $\langle \cdot, \cdot \rangle_A = \langle \cdot, A^{-1} \cdot \rangle$ and $| \cdot |_A = |A^{-\frac{1}{2}} \cdot |.$
Sample Path

\[ \dot{u} = F(u, \rho; G) + \sqrt{2C(\rho)} \dot{W}, \]

Fokker-Planck

\[ \partial_t \rho = \nabla \cdot \left( \nabla \cdot (C(\rho) \rho) - F(u, \rho; G) \rho \right). \]
Assumption

The forward model \( G = G_\epsilon \) where

\[
G_\epsilon(u) = G_0(u) + G_1(u/\epsilon),
\]

\( G_0 \in C^1(\mathbb{R}^d, \mathbb{R}^K), \ G_1 \in C^1(\mathbb{T}^d, \mathbb{R}^K) \) and \( \int_{\mathbb{T}^d} G_1(y) \, dy = 0. \)

Multiscale Expansion Result

In limit \( \epsilon \to 0 \)

\[
\rho(u, t; G_\epsilon) \to \rho(u, t; G_0).
\]
Self-Preconditioned Langevin Equation

Sample Path

\[ \dot{u} = -C(\rho) \nabla \Phi(u) + \sqrt{2C(\rho)} \dot{W} \]

Fokker-Planck

\[ \partial_t \rho = \nabla \cdot (C(\rho) (\nabla V \rho + \nabla \rho)) \]
Self-Preconditioned Langevin Equation

Assumption

The forward model \( G = G_\epsilon \) where

\[
G_\epsilon(u) = G_0(u) + G_1(u/\epsilon),
\]

\( G_0 \in C^1(\mathbb{R}^d, \mathbb{R}^K) \), \( G_1 \in C^1(\mathbb{T}^d, \mathbb{R}^K) \) and \( \int_{\mathbb{T}^d} G_1(y) \, dy = 0 \).

Multiscale Expansion Result

In limit \( \epsilon \to 0 \)

\[
\rho(u, t; \Phi_\epsilon, C) \to \rho(u, t; \Phi_*, D).
\]

Here \( \Phi_* \neq \Phi_0 \) and \( C \supseteq D \).
Example 2 – Linear + Periodic

Available Forward Model $G_\epsilon(\cdot)$

\[
G_\epsilon(u) = Au + \left[ \sin \left( \frac{2\pi u_1}{\epsilon} \right), \sin \left( \frac{2\pi u_2}{\epsilon} \right) \right]^\top \quad \text{with} \quad A = \begin{pmatrix} -1 & 0 \\ 0 & 2 \end{pmatrix}.
\]

Desired Forward Model $G_0(\cdot)$

\[
G_0(u) = Au \quad \text{with} \quad A = \begin{pmatrix} -1 & 0 \\ 0 & 2 \end{pmatrix}.
\]
Noisy Misfit

Linear + Periodic – Ensemble Langevin sampler
Noisy Misfit

Linear + Periodic – Ensemble Kalman sampler
Noisy Misfit

Linear + Periodic – Misfit along iteration
Example 3 – Lorenz ’63

Governing Dynamics

\[
\begin{align*}
\dot{x} &= 10 (y - x) \\
\dot{y} &= rx - y - xz \\
\dot{z} &= xy - bz
\end{align*}
\]

- 2-dimensional unknown: \( u = [r, b]^T \)
- Forward map \( G \) only available to us approximately via \( G_T \).
- Noise \( \eta \) only available to us approximately via \( G_T \).
Noisy Misfit

Lorenz '63 – Misfit versus parameter
Noisy Misfit

Lorenz ’63 – Misfit along iteration
Refineable Ensemble Methods

- Multiscale particles: Pavliotis, S & Vaes (2021) [36]
Multiscale-EKS

Sample Path

\[ \dot{u} = -\frac{1}{J\sigma^2} \sum_{j=1}^{J} \langle G(u^{(j)}) - G(u), G(u) - y \rangle \Gamma(u^{(j)} - u) \]

\[ \quad - C(\Xi)\Sigma^{-1} u + \nu \sqrt{2C(\Xi)} \dot{w}, \]

\[ u^{(j)} = u + \sigma \xi^{(j)}, \quad j = 1, \ldots, J, \]

\[ \dot{\xi}^{(j)} = -\frac{1}{\delta^2} \xi^{(j)} + \sqrt{\frac{2}{\delta^2}} \dot{w}^{(j)}, \quad \xi^{(j)}(0) \sim \mathcal{N}(0, I_d), \quad j = 1, \ldots, J, \]

Covariance

\[ C(\Xi) = \frac{1}{J} \sum_{j=1}^{J} (\xi^{(j)} \otimes \xi^{(j)}). \]
Sample Path

\[ u_t = -\nabla \Phi_R(u_t) + \nu \sqrt{2} \dot{w}_t. \]

Define the exponent \( \beta \) is defined as follows:

\[ \beta = \begin{cases} 
1 & \text{if } G \in C^2(T^d, \mathbb{R}^K), \\
2 & \text{if } G \in C^3(T^d, \mathbb{R}^K). 
\end{cases} \]

Theorem [36]

Let \( p > 1 \). Then

\[ \mathbb{E} \left( \sup_{0 \leq t \leq T} \| u(t) - u(t) \|^p \right) \leq C(\delta^p + \sigma^{\beta p}). \]
Example 3: Darcy Flow

Problem Setting

- **Forward**: Find pressure $p(\cdot)$ from permeability $a(\cdot)$:

  \[-\nabla \cdot (a(x) \nabla p(x)) = f(x), \quad x \in D\]
  \[p(x) = 0, \quad x \in \partial D.\]

- **Inverse**: Find $a$ from linear functionals $\{\ell_j\}$ of $p$.

- **Prior on $a$**: $C = (-\Delta + \tau^2 I)^{-\alpha}$, $C \varphi_j = \lambda_j \varphi_j$, $\log a \sim N(0, C)$:

  \[
  \log a(x) = \sum_{j \in \mathbb{Z}_+^2} u_j \sqrt{\lambda_j} \varphi_j(x), \quad u_j \sim N(0, 1) \text{ i.i.d.}
  \]

- **Likelihood** $y | u \sim N(G(u), \gamma^2 I)$,

  \[G_j(u) = \ell_j(p(\cdot; u)), \quad j = 1, \cdots, 50.\]

- **Posterior** $u | y$. 
Posterior Mean

True permeability (left); Posterior mode permeability (right)
Posterior Variability

Posterior in Karhunen-Loève coefficients $u$
Closing
Conclusions

1. Inverse problems of increasing importance.
   ▶ Often forward model is expensive.
   ▶ Often forward model adjoints impossible/expensive.
   ▶ Sometimes only rough forward model available.

2. Ensemble methods attractive in this setting.
   ▶ Gradient flow structure: parameter space;
   ▶ Gradient flow structure: probability space.
   ▶ Multiscale analysis of rough forward models.
   ▶ Multiscale approach to refineable approximations.

3. Many open mathematical questions.
References I

*Gradient flows: in metric spaces and in the space of probability measures.*  

[2] A. Azouani, E. Olson, and E. S. Titi.  
Continuous data assimilation using general interpolant observables.  

A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem.  

A localization technique for ensemble Kalman filters.  
References II

A mollified ensemble Kalman filter.

An ensemble Kalman-Bucy filter for continuous data assimilation.

Accuracy and stability of the continuous-time 3DVAR filter for the navier–stokes equation.

A strongly convergent numerical scheme from ensemble kalman inversion.
References III

Well posedness and convergence analysis of the ensemble kalman inversion.

Tikhonov regularization within ensemble Kalman inversion.

Ensemble randomized maximum likelihood method as an iterative ensemble smoother.

Calibrate, emulate, sample.
Ensemble inference methods for models with noisy and expensive likelihoods.

Ensemble smoother with multiple data assimilation.

Analysis of the ensemble and polynomial chaos kalman filters in bayesian inverse problems.

Analysis of iterative ensemble smoothers for solving inverse problems.
Interacting Langevin diffusions: Gradient structure and ensemble Kalman sampler.

Affine invariant interacting Langevin dynamics for bayesian inference.

Ensemble samplers with affine invariance.

Never look back-a modified enkf method and its application to the training of neural networks without back propagation.
[21] K. Hayden, E. Olson, and E. S. Titi.
Discrete data assimilation in the lorenz and 2d navier–stokes equations.

A proposal for the intercomparison of the dynamical cores of atmospheric general circulation models.

Unscented Kalman inversion.

A regularizing iterative ensemble kalman method for pde-constrained inverse problems.
References VII

Ensemble Kalman methods for inverse problems.  

The variational formulation of the fokker–planck equation.  

A new approach for filtering nonlinear systems.  

Well-posedness and accuracy of the ensemble Kalman filter in discrete and continuous time.  
References VIII

[29] T. Lange and W. Stannat.
On the continuous time limit of ensemble square root filters.

On the continuous time limit of the ensemble Kalman filter.

Data assimilation.
Cham, Switzerland: Springer, 2015.

Ensemble preconditioning for markov chain monte carlo simulation.

Interacting particle solutions of Fokker-Planck equations through
gradient-log-density estimation.
Entropy, 22(8), 2020.
References IX

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

The hydrological cycle over a wide range of climates simulated with an idealized gcm.

Derivative-free bayesian inversion using multiscale dynamics.

Sampling from rough energy landscapes.
References

[38] S. Reich.
A dynamical systems framework for intermittent data assimilation.

*Probabilistic forecasting and Bayesian data assimilation.*

Ensemble filter techniques for intermittent data assimilation.

Fokker-planck particle systems for bayesian inference: Computational approaches.
References XI

Analysis of the ensemble Kalman filter for inverse problems.

Imposing sparsity within ensemble kalman inversion.

Learning stochastic closures using ensemble kalman inversion.

*Optimal transport: old and new*, volume 338.