Kalman-Wasserstein Gradient Flows

Franca Hoffmann

Computing and Mathematical Sciences
California Institute of Technology

Institut für Mathematik der Universität Potsdam.

SFB-Seminar
Potsdam, August 13th, 2019

Overview

collaboration with:

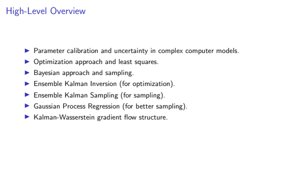
Alfredo Garbuno-Inigo, Wuchen Li, Andrew Stuart (2019)

arXiv preprint, 1903.08866

The Big Picture

High-Level Overview

- Parameter calibration and uncertainty in complex computer models.
- Optimization approach and least squares.
- Bayesian approach and sampling.
- Ensemble Kalman Inversion (for optimization).
- Ensemble Kalman Sampling (for sampling).
- Gaussian Process Regression (for better sampling).
- Kalman-Wasserstein gradient flow structure.



- want to solve an inverse problem
- with randomness on observed data (e.g. climate models), we are not just interested to invert, but also in UQ, i.e. to understand what the posterior distribution looks like
- In many applications in science & engineering: need to solve inverse problems without using derivatives/adjoints of the forward model
- Goal 1: construct derivative-free methods which generate approximate samples from the posterior that solves inverse problem
- Goal 2: establish mathematical framework to do meaningful analysis

- Parameter calibration and uncertainty in complex computer models.
- Optimization approach and least squares.
- Bayesian approach and sampling.
- Ensemble Kalman Inversion (for optimization).
- Ensemble Kalman Sampling (for sampling).
- Gaussian Process Regression (for better sampling).
- Kalman-Wasserstein gradient flow structure.

Roadmap of arguments

- continuous time limit of EKI
- introduce specific term of noise and derive SDE: new algorithm EKS
- approximate this SDE replacing differences by gradients
- mean-field limit of approximate SDE gives a non-lin FP eqn
- identify novel GF structure of non-linear FP eqn (augmenting W2 GF with the covariance of the noisy flow)
- using GF structure, investigate long-time behavior. Limit=posterior of underlying inverse problem!
- investigate numerically EKS: effect of adding this type of noise to original derivative-free EKI. Get good approximate samples from posterior distribution!

Kalman-Wasserstein space

Gradient Flow Structure

$$\partial_t \rho = \nabla \cdot \left(\rho \, \mathcal{C}(\rho) \, \nabla \frac{\delta \mathcal{E}(\rho)}{\delta \rho} \right)$$

$$\mathcal{C}(\rho) = \int (\theta - \bar{\theta}) \otimes (\theta - \bar{\theta}) \rho(\theta, t) d\theta, \quad \bar{\theta} = \int \theta \rho(\theta, t) d\theta.$$

- Gradient flow in $\mathcal{C}(\rho)$-weighted metric.
- Suitable energy $\mathcal{E}$: unique attractor is the posterior for an underlying inverse problem.
- Inspired new efficient derivative-free algorithm to solve inverse problems.

Inverse Problem For Parameters

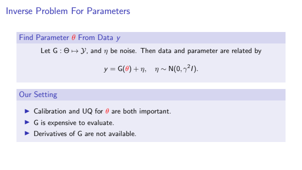
Find Parameter θ From Data y

Let $\mathrm{G} : \Theta \mapsto \mathcal{Y}$, and η be noise. Then data and parameter are related by

$$y = \mathrm{G}(\theta) + \eta, \quad \eta \sim \mathrm{N}(0, \gamma^2 I).$$

Our Setting

- Calibration and UQ for θ are both important.
- G is expensive to evaluate.
- Derivatives of G are not available.



- Goal: find out info about true θ that generated data, given observations y, noise γ and forward map G. I.e. need to invert G.
- the spaces $\mathcal{U}$ and $\mathcal{Y}$ can be very high dimensional, and usually are in applications
- forward map G could be horrible (non-linear, not practical to differentiate). For ex think of θ as the coefficients and initial conditions of a complicated climate model, and G the solution map, where y are measurements taken

Optimization Approach

Find Parameter θ From Data y

Let $\mathsf{G} : \Theta \mapsto \mathcal{Y}$, and η be noise. Then data and parameter are related by

$$y = \mathsf{G}(\theta) + \eta, \quad \eta \sim \mathsf{N}(0, \gamma^2 I).$$

Mathematical Formulation

$$\theta^\star = \mathrm{argmin}_{\theta \in \Theta} \Phi(\theta; y),$$

$$\Phi(\theta; y) = \frac{1}{2\gamma^2}|y - \mathsf{G}(\theta)|^2, \qquad \Phi_R(\theta; y) = \frac{1}{2\gamma^2}|y - \mathsf{G}(\theta)|^2 + \frac{1}{2}\langle \theta, \Sigma_0^{-1}\theta \rangle.$$

Algorithms: parameter θ calibration.

Optimization Approach

Find Parameter θ From Data y

Let $\mathsf{G} : \Theta \mapsto \mathcal{Y}$, and η be noise. Then data and parameter are related by

$$y = \mathsf{G}(\theta) + \eta, \quad \eta \sim \mathsf{N}(0, \gamma^2 I).$$

Mathematical Formulation

$$\theta^\star = \mathrm{argmin}_{\theta \in \Theta} \Phi(\theta; y),$$

$$\Phi(\theta; y) = \frac{1}{2\gamma^2}|y - \mathsf{G}(\theta)|^2, \qquad \Phi_R(\theta; y) = \frac{1}{2\gamma^2}|y - \mathsf{G}(\theta)|^2 + \frac{1}{2}\langle \theta, \Sigma_0^{-1}\theta \rangle.$$

Algorithms: parameter θ calibration.

- Optimization approach: find most likely θ.
- As in any optimization problem, need a loss function. Here, it is natural to choose Φ.
- how to find $\theta^\star$? Recall that we want to avoid taking gradients and evaluations of G are costly
- need efficient algorithms that get around this challenge

Bayesian Approach

- Prior: $\mathbb{P}(\theta)$
- Likelihood: $\mathbb{P}(y|\theta)$
- Posterior: $\mathbb{P}(\theta|y)$

- Prior: $\theta \sim \mathrm{N}(0, \Sigma_0)$
- Likelihood: $y - \mathrm{G}(\theta) \sim \mathrm{N}(0, \gamma^2 I)$
- Posterior: $\theta|y \sim ?$

Mathematical Formulation

$$\mathbb{P}(\theta|y) \propto \mathbb{P}(y|\theta) \times \mathbb{P}(\theta),$$

$$\begin{aligned}\mathbb{P}(\theta|y) &\propto \exp\Big(-\Phi(\theta; y)\Big) \times \exp\Big(-\frac{1}{2}\langle\theta, \Sigma_0^{-1}\theta\rangle\Big) \\ &\propto \exp\Big(-\Phi_R(\theta; y)\Big)\end{aligned}$$

Algorithms: parameter θ sampling.

Bayesian Approach

- Prior: $\mathbb{P}(\theta)$
- Likelihood: $\mathbb{P}(y|\theta)$
- Posterior: $\mathbb{P}(\theta|y)$

- Prior: $\theta \sim \mathrm{N}(0, \Sigma_0)$
- Likelihood: $y - \mathrm{G}(\theta) \sim \mathrm{N}(0, \gamma^2 I)$
- Posterior: $\theta | y \sim ?$

Mathematical Formulation

$$\begin{aligned}
\mathbb{P}(\theta|y) &\propto \mathbb{P}(y|\theta) \times \mathbb{P}(\theta), \\
\mathbb{P}(\theta|y) &\propto \exp\Big(-\Phi(\theta; y)\Big) \times \exp\Big(-\frac{1}{2}\langle \theta, \Sigma_0^{-1}\theta \rangle\Big) \\
&\propto \exp\Big(-\Phi_R(\theta; y)\Big)
\end{aligned}$$

Algorithms: parameter θ sampling.

- with randomness on observed data (e.g. climate models), we are not just interested to invert, but also in UQ, i.e. to understand what the posterior distribution looks like
- Bayesian approach to inverse problems: get your hands on posterior

Bayesian Approach

Mathematical Formulation

$$\mathbb{P}(\theta|y) \propto \exp\Big(-\Phi_R(\theta; y)\Big)$$

$$\Phi_R(\theta; y) = \frac{1}{2\gamma^2}|y - \mathrm{G}(\theta)|^2 + \frac{1}{2}\langle\theta, \Sigma_0^{-1}\theta\rangle$$

- maximizing $\mathbb{P}(\theta|y)$ = minimizing $\Phi_R(\theta; y)$.
- prior term introduces regularization of Tikhonov-Phillips form.
- UQ: want more information on posterior than just optimum.

Algorithms: parameter θ sampling.

Bayesian Approach

Mathematical Formulation

$$\mathbb{P}(\theta|y) \propto \exp\Big(-\Phi_R(\theta; y)\Big)$$

$$\Phi_R(\theta; y) = \frac{1}{2\gamma^2}|y - \mathrm{G}(\theta)|^2 + \frac{1}{2}\langle\theta, \Sigma_0^{-1}\theta\rangle$$

- maximizing $\mathbb{P}(\theta|y)$ = minimizing $\Phi_R(\theta; y)$.
- prior term introduces regularization of Tikhonov-Phillips form.
- UQ: want more information on posterior than just optimum.

Algorithms: parameter θ sampling.

- to find most likely θ given observed data: maximize posterior. This is the same as minimizing Φ_R.
- Note that this corresponds to the previous optimization problem, but with an additional regularization term of Tikhonov-Phillips form.
- UQ: want more information on posterior than just its max.

Bayesian Approach

Mathematical Formulation

$$\mathbb{P}(\theta|y) \propto \exp\Big(-\Phi_R(\theta; y)\Big) \qquad \Phi_R(\theta; y) = \frac{1}{2\gamma^2}|y - \mathsf{G}(\theta)|^2 + \frac{1}{2}\langle\theta, \Sigma_0^{-1}\theta\rangle$$

Our Setting

- G is expensive to evaluate.
- Derivatives of G are not available.

Goals

- Methods to generate approximate samples from posterior.
- Mathematical framework for analysis using gradient flow structure.

Bayesian Approach

Mathematical Formulation

$$\mathbb{P}(\theta|y) \propto \exp\Big(-\Phi_R(\theta; y)\Big) \qquad \Phi_R(\theta; y) = \frac{1}{2\gamma^2}|y - \mathsf{G}(\theta)|^2 + \frac{1}{2}\langle\theta, \Sigma_0^{-1}\theta\rangle$$

Our Setting

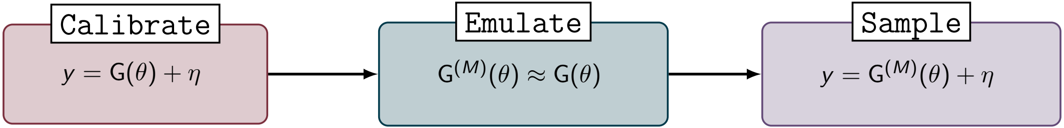
- G is expensive to evaluate.
- Derivatives of G are not available.

Goals

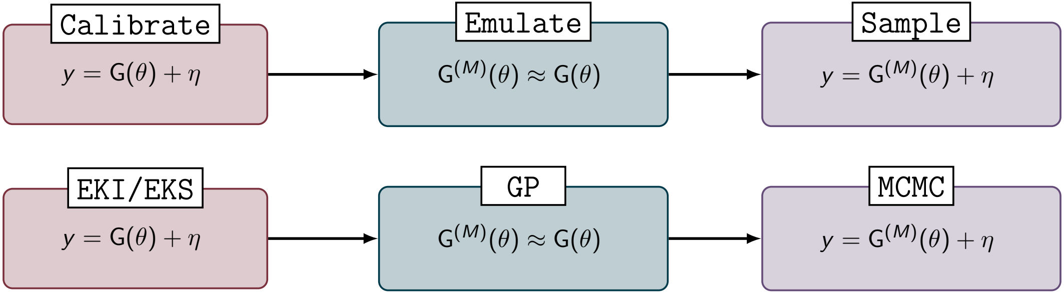
- Methods to generate approximate samples from posterior.
- Mathematical framework for analysis using gradient flow structure.

- Since posterior is expensive to evaluate, need to find efficient ways to sample from the posterior.
- we will propose here two new avenues: (1) new algorithm to sample from posterior, (2) mathematical framework using a novel gradient flow structure.

Calibrate, Emulate, Sample



Calibrate, Emulate, Sample



EKI And EKS: Calibration

collaboration with: Garbuno, Hoffmann, Li, Stuart (2019)

arXiv preprint, 1903.08866 [4]

Calibrate

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

$\longrightarrow$

Emulate

$$\mathsf{G}^{(M)}(\theta) \approx \mathsf{G}(\theta)$$

$\longrightarrow$

Sample

$$y = \mathsf{G}^{(M)}(\theta) + \eta$$

Ensemble Kalman Inversion (EKI)

Continuous Time Formulation

$$\dot{\theta}^{(j)} = -\frac{1}{J}\sum_{k=1}^{J}\frac{1}{\gamma^2}\left\langle \mathsf{G}(\theta^{(k)}) - \bar{\mathsf{G}}, \mathsf{G}(\theta^{(j)}) - y\right\rangle \left(\theta^{(k)} - \bar{\theta}\right),$$

$$\bar{\theta} = \frac{1}{J}\sum_{k=1}^{J}\theta^{(k)}, \qquad \bar{\mathsf{G}} = \frac{1}{J}\sum_{k=1}^{J}\mathsf{G}\big(\theta^{(k)}\big).$$

- Sample from prior instead of posterior.
- Evolve particle ensemble: derivative-free algorithm.
- Tool for optimization: collapses to max $\mathbb{P}(\theta|y)$.

Ensemble Kalman Inversion (EKI)

Continuous Time Formulation

$$\dot{\theta}^{(j)} = -\frac{1}{J}\sum_{k=1}^{J}\frac{1}{\gamma^2}\left\langle \mathsf{G}(\theta^{(k)}) - \bar{\mathsf{G}}, \mathsf{G}(\theta^{(j)}) - y \right\rangle \left(\theta^{(k)} - \bar{\theta}\right),$$

$$\bar{\theta} = \frac{1}{J}\sum_{k=1}^{J}\theta^{(k)}, \quad \bar{\mathsf{G}} = \frac{1}{J}\sum_{k=1}^{J}\mathsf{G}(\theta^{(k)}).$$

- Sample from prior instead of posterior.
- Evolve particle ensemble: derivative-free algorithm.
- Tool for optimization: collapses to max $\mathbb{P}(\theta|y)$.

- Let me begin by telling you about an existing algorithm to sample the posterior: EKI
- main idea: instead of sampling from posterior directly, which is costly, instead sample from prior, which can be done efficiently, and then evolve samples under a suitable particle model to include information on observations.
- **Ensemble** because we evolve an ensemble of particles in parameter space. This allows to get around the issue of computing derivatives. EKI is derivative-free!
- first term in brackets: drives particles to consensus. second term in brackets: drives particles to match observations. mean cancels.

Ensemble Kalman Inversion (EKI)

Continuous Time Formulation

$$\dot{\theta}^{(j)} = -\frac{1}{J}\sum_{k=1}^{J}\frac{1}{\gamma^2}\left\langle \mathsf{G}(\theta^{(k)}) - \bar{\mathsf{G}}, \mathsf{G}(\theta^{(j)}) - y \right\rangle \left(\theta^{(k)} - \bar{\theta}\right),$$

$$\bar{\theta} = \frac{1}{J}\sum_{k=1}^{J}\theta^{(k)}, \qquad \bar{\mathsf{G}} = \frac{1}{J}\sum_{k=1}^{J}\mathsf{G}\big(\theta^{(k)}\big).$$

Optimization Approach

$$\theta^{\star} = \operatorname{argmin}_{\theta\in\Theta}\Phi(\theta; y),$$

$$\Phi(\theta; y) = \frac{1}{2\gamma^2}|y - \mathsf{G}(\theta)|^2.$$

Ensemble Kalman Inversion (EKI)

Continuous Time Formulation

$$\dot{\theta}^{(j)} = -\frac{1}{J}\sum_{k=1}^{J}\frac{1}{\gamma^2}\left\langle \mathsf{G}(\theta^{(k)}) - \bar{\mathsf{G}}, \mathsf{G}(\theta^{(j)}) - y \right\rangle \left(\theta^{(k)} - \bar{\theta}\right),$$

$$\bar{\theta} = \frac{1}{J}\sum_{k=1}^{J}\theta^{(k)}, \quad \bar{\mathsf{G}} = \frac{1}{J}\sum_{k=1}^{J}\mathsf{G}(\theta^{(k)}).$$

Optimization Approach

$$\theta^{\star} = \mathrm{argmin}_{\theta\in\Theta}\,\Phi(\theta; y),$$

$$\Phi(\theta; y) = \frac{1}{2\gamma^2}|y - \mathsf{G}(\theta)|^2.$$

- issue with EKI: collapses very fast to a point even when adding noise to the observations. so not so useful to reproduce posterior, see [Ernst, Sprungk and Starkloff, 2015]. However, tool for optimization as ensemble collapses to solution of our initial optimization problem. Used widely in applications.
- To do UQ: use framework of Bayesian inverse problems and try to understand better shape of posterior. We suggest a new noisy EKI that allows to sample from the posterior. This is why we call it **Ensemble Kalman Sampling (EKS)**.
- Recall: goal is to derive efficient methods to approximate samples from posterior without using gradients/adjoints.

Ensemble Kalman Inversion (EKI)

Continuous Time Formulation

$$\dot{\theta}^{(j)} = -\frac{1}{J}\sum_{k=1}^{J}\frac{1}{\gamma^2}\left\langle \mathsf{G}(\theta^{(k)}) - \bar{\mathsf{G}}, \mathsf{G}(\theta^{(j)}) - y \right\rangle \left(\theta^{(k)} - \bar{\theta}\right),$$
$$\bar{\theta} = \frac{1}{J}\sum_{k=1}^{J}\theta^{(k)}, \quad \bar{\mathsf{G}} = \frac{1}{J}\sum_{k=1}^{J}\mathsf{G}(\theta^{(k)}).$$

Optimization Approach

$$\theta^{\star} = \mathrm{argmin}_{\theta\in\Theta}\,\Phi(\theta; y),$$
$$\Phi(\theta; y) = \frac{1}{2\gamma^2}|y - \mathsf{G}(\theta)|^2.$$

Just to remind myself:

- Linear G: can think of EKI as a Monte Carlo approximation of the Kalman filter (KF).
- Non-linear G: EKI does not converge to posterior for large number of particles ! [Ernst, Sprungk and Starkloff, 2015]
- The KF was proposed in the 60s for linear forward maps and Gaussian noise as a method for state estimation (took us to the moon: track rocket and correct trajectory given noisy observations).
- **Filter**: filter out noise to make predictions. **Inversion**: easier than filtering, we only invert once.

Ensemble Kalman Sampling (EKS)

Continuous Time Formulation

$$\dot{\theta}^{(j)} = -\frac{1}{J}\sum_{k=1}^{J}\frac{1}{\gamma^2}\left\langle \mathsf{G}(\theta^{(k)}) - \bar{\mathsf{G}}, \mathsf{G}(\theta^{(j)}) - y \right\rangle \left(\theta^{(k)} - \bar{\theta}\right) - C(\theta)\Sigma_0^{-1}\theta^{(j)} + \sqrt{2C(\theta)}\dot{W}^{(j)}$$

$$C(\theta) = \frac{1}{J}\sum_{k=1}^{J}\left(\theta^{(k)} - \bar{\theta}\right) \otimes \left(\theta^{(k)} - \bar{\theta}\right).$$

- add damping term related to prior [Chada,Stuart,Tong 2019].
- new noise covariance structure.

Ensemble Kalman Sampling (EKS)

Continuous Time Formulation

$$\dot{\theta}^{(j)} = -\frac{1}{J}\sum_{k=1}^{J} \frac{1}{\gamma^2} \left\langle \mathrm{G}(\theta^{(k)}) - \bar{\mathrm{G}}, \mathrm{G}(\theta^{(j)}) - y \right\rangle \left(\theta^{(k)} - \bar{\theta}\right) - C(\theta)\Sigma_0^{-1}\theta^{(j)} + \sqrt{2C(\theta)}\dot{W}^{(j)}$$

$$C(\theta) = \frac{1}{J}\sum_{k=1}^{J}\left(\theta^{(k)} - \bar{\theta}\right) \otimes \left(\theta^{(k)} - \bar{\theta}\right).$$

- add damping term related to prior [Chada,Stuart,Tong 2019].
- new noise covariance structure.

- We propose a new algorithm inspired by a novel gradient flow structure to capture posterior.
- add damping term (related to prior). This term has already been introduced in [Chada,Stuart,Tong 2019].
- changed noise covariance structure: new to use this covariance for the noise. Choice of this noise is informed by optimal transport (GF structure).
- idea: perturb particles instead of data (classical noisy EKI has noise on observations). This idea has already been introduced in [Kovachki, Stuart 2018], but in the context of optimization instead of sampling.
- output of EKS: approximate samples from posterior.

The EKS And Mean Field Limits

EKS: Approximation 1

$$\dot{\theta}^{(j)} = -\frac{1}{J}\sum_{k=1}^{J}\frac{1}{\gamma^2}\left\langle \mathsf{G}(\theta^{(k)}) - \bar{\mathsf{G}}, \mathsf{G}(\theta^{(j)}) - y\right\rangle \left(\theta^{(k)} - \bar{\theta}\right) - C(\theta)\Sigma_0^{-1}\theta^{(j)} + \sqrt{2C(\theta)}\dot{W}^{(j)},$$

$$C(\theta) = \frac{1}{J}\sum_{k=1}^{J}\left(\theta^{(k)} - \bar{\theta}\right) \otimes \left(\theta^{(k)} - \bar{\theta}\right).$$

Linear Approximation

$$\left(\mathsf{G}(\theta^{(k)}) - \bar{\mathsf{G}}\right) \approx A(\theta^{(k)} - \bar{\theta})\,, \qquad A := D\mathsf{G}(\theta^{(j)})\,.$$

EKS: Approximation 1

Linear Approximation of EKS

$$\dot{\theta}^{(j)} = -\frac{1}{J\gamma^2}\sum_{k=1}^{J}\left\langle A\left(\theta^{(k)}-\bar{\theta}\right), \left(\mathsf{G}(\theta^{(j)})-y\right)\right\rangle \left(\theta^{(k)}-\bar{\theta}\right) \\ - C(\theta)\Sigma_0^{-1}\theta^{(j)} + \sqrt{2C(\theta)}\,\dot{W}^{(j)}\,.$$

EKS: Approximation 1

- $\nabla\Phi_R(\theta) = \frac{1}{\gamma^2}A^T(A\theta - y) + \Sigma_0^{-1}\theta.$
- $\left\langle DG(\theta^{(j)})(\theta^{(k)} - \bar{\theta}), (G(\theta^{(j)}) - y)\right\rangle = \left\langle (\theta^{(k)} - \bar{\theta}), A^T(A\theta^{(j)} - y)\right\rangle.$

Preconditioned Langevin Equation

$$\dot{\theta}^{(j)} = -C(\theta)\nabla\Phi_R(\theta^{(j)}) + \sqrt{2C(\theta)}\dot{W}^{(j)}.$$

- If G linear, this is EKS.
- If G non-linear, expect solution to be close to EKS solution if particles close together (conjecture!).

EKS: Approximation 1

- $\nabla\Phi_R(\theta) = \frac{1}{\gamma^2}A^T(A\theta - y) + \Sigma_0^{-1}\theta$.
- $\left\langle DG(\theta^{(j)})(\theta^{(k)} - \bar{\theta}), (G(\theta^{(j)}) - y)\right\rangle = \left\langle (\theta^{(k)} - \bar{\theta}), A^T(A\theta^{(j)} - y)\right\rangle$.

Preconditioned Langevin Equation

$$\dot{\theta}^{(j)} = -C(\theta)\nabla\Phi_R(\theta^{(j)}) + \sqrt{2C(\theta)}\dot{W}^{(j)}.$$

- If G linear, this is EKS.
- If G non-linear, expect solution to be close to EKS solution if particles close together (conjecture!).

- we propose EKS as new algorithm to sample posterior (derivative-free!)
- last SDE: natural object of analysis.

Proof to obtain last SDE:

$$\begin{aligned}
&\frac{1}{J}\sum_{k=1}^{J}\left\langle DG(\theta^{(j)})(\theta^{(k)} - \bar{\theta}), (G(\theta^{(j)}) - y)\right\rangle \left(\theta^{(k)} - \bar{\theta}\right) \\
&\quad = \frac{1}{J}\sum_{k=1}^{J}\left\langle (\theta^{(k)} - \bar{\theta}), A^T\left(A\theta^{(j)} - y\right)\right\rangle \left(\theta^{(k)} - \bar{\theta}\right) \\
&\quad = C(\theta)A^T\left(A\theta^{(j)} - y\right).
\end{aligned}$$

EKS: Approximation 2

$$\mathcal{C}(\rho) := \int (\theta - \bar{\theta}) \otimes (\theta - \bar{\theta}) \rho(\theta, t) d\theta, \qquad \bar{\theta} := \int \theta \rho(\theta, t) d\theta.$$

Mean Field Limit

$$\dot{\theta} = -\mathcal{C}(\rho) \nabla \Phi_R(\theta) + \sqrt{2C(\rho)} \dot{W} \, .$$

Nonlinear Fokker-Planck equation

$$\partial_t \rho = \nabla \cdot \left(\rho \, \mathcal{C}(\rho) \, \nabla \Phi_R \right) + \mathcal{C}(\rho) : D^2 \rho,$$

$$\Phi_R(\theta) = \frac{1}{2\gamma^2} \left| y - G(\theta) \right|^2 + \frac{1}{2} \left| \Sigma_0^{-\frac{1}{2}} \theta \right|^2.$$

EKS: Approximation 2

$$\mathcal{C}(\rho) := \int (\theta - \bar{\theta}) \otimes (\theta - \bar{\theta}) \rho(\theta, t) d\theta, \qquad \bar{\theta} := \int \theta \rho(\theta, t) d\theta.$$

Mean Field Limit

$$\dot{\theta} = -\mathcal{C}(\rho) \nabla \Phi_R(\theta) + \sqrt{2\mathcal{C}(\rho)} \dot{W}\,.$$

Nonlinear Fokker-Planck equation

$$\partial_t \rho = \nabla \cdot \left(\rho\, \mathcal{C}(\rho)\, \nabla \Phi_R\right) + \mathcal{C}(\rho) : D^2 \rho,$$

$$\Phi_R(\theta) = \frac{1}{2\gamma^2} \left|y - G(\theta)\right|^2 + \frac{1}{2} \left|\Sigma_0^{-\frac{1}{2}} \theta\right|^2.$$

- rigorous proof of mean-field limit still open.
- Can consider both of these equations for non-linear G also!
- If G linear, this corresponds to EKS.

Connection To Bayesian Inversion

$$\partial_t \rho = \nabla \cdot \left(\rho\, \mathcal{C}(\rho)\, \nabla \left(\Phi_R + \ln \rho\right)\right) , \qquad \mathcal{C}(\rho) = \int \left(\theta - \bar{\theta}\right) \otimes \left(\theta - \bar{\theta}\right) \rho(\theta, t) d\theta .$$

Manifold of Stationary states

$$\rho(\theta) = \delta_v(\theta) \text{ for some } v \in \mathbb{R}^d \quad \Leftrightarrow \quad \mathcal{C}(\rho) = 0 .$$

Steady state

Equilibrium solution to non-linear Fokker-Planck equation:

$$\rho_\infty(\theta) := \frac{e^{-\Phi_R(\theta)}}{\int e^{-\Phi_R(\theta)}\, \mathrm{d}\theta} .$$

This is the density of $\mathbb{P}(\theta | y)$.

Connection To Bayesian Inversion

$$\partial_t \rho = \nabla \cdot \left(\rho \mathcal{C}(\rho) \nabla \left(\Phi_R + \ln \rho\right)\right) , \qquad \mathcal{C}(\rho) = \int (\theta - \bar{\theta}) \otimes (\theta - \bar{\theta}) \rho(\theta, t) d\theta \, .$$

Manifold of Stationary states

$$\rho(\theta) = \delta_v(\theta) \text{ for some } v \in \mathbb{R}^d \quad \Leftrightarrow \quad \mathcal{C}(\rho) = 0 \, .$$

Steady state

Equilibrium solution to non-linear Fokker-Planck equation:

$$\rho_\infty(\theta) := \frac{e^{-\Phi_R(\theta)}}{\int e^{-\Phi_R(\theta)} \, \mathrm{d}\theta} \, .$$

This is the density of $\mathbb{P}(\theta|y)$.

Proof of Lemma:

- $(\Rightarrow)$ by direct substitution.
- $(\Leftarrow)$ $\mathcal{C}(\rho) = 0$ implies $\int |u|^2 \rho \, \mathrm{d}u = \left(\int u \rho \, \mathrm{d}u\right)^2$, which is the equality case of Jensen's inequality, and therefore only holds if ρ is the law of a constant random variable.

We expect that ρ_∞ is the unique attractor:
always converge to ρ_∞ avoiding Diracs as long as we start away from Diracs. More on this later.

Kalman-Wasserstein Space

Gradient Flow Structure

- For $\Omega \subseteq \mathbb{R}^d$ convex set, define $\mathcal{P}_+ := \{\rho \in \mathcal{P} \, : \, \rho > 0 \text{ a.e. } , \rho \in C^\infty(\Omega)\}$.
- Tangent space at $\rho \in \mathcal{P}_+$:

$$T_\rho \mathcal{P}_+ = \left\{ \sigma \in C^\infty(\Omega) \, : \, \int \sigma dx = 0 \right\} .$$

- Onsager operator $\Delta_{\rho,\mathcal{C}} : T_\rho^* \mathcal{P}_+ \to T_\rho \mathcal{P}_+$:

$$(-\Delta_{\rho,\mathcal{C}})\, \phi := -\nabla \cdot (\rho \mathcal{C}(\rho) \nabla \phi)$$

- $\Delta_{\rho,\mathcal{C}}$ degenerate if ρ is a Dirac.
- For $\rho \in \mathcal{P}_+$ bounded away from zero, $\Delta_{\rho,\mathcal{C}}$ is well-defined, non-singular and invertible: $\sigma \mapsto (-\Delta_{\rho,\mathcal{C}})^{-1} \sigma$ for $\sigma \in T_\rho \mathcal{P}_+$ (else, use pseudo-inverse).

Gradient Flow Structure

Kalman-Wasserstein metric tensor

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

$$g_{\rho,\mathcal{C}}(\sigma_1, \sigma_2) := \int_\Omega \langle \nabla \phi_1 \,,\, \mathcal{C}(\rho) \nabla \phi_2 \rangle \; \rho \, \mathrm{d}x,$$

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

Gradient Flow Structure

Kalman-Wasserstein metric $\mathcal{W}_{\mathcal{C}}: \mathcal{P}_+ \times \mathcal{P}_+ \to \mathbb{R}$

For $\rho^0,\ \rho^1 \in \mathcal{P}_+$,

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

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

Gradient Flow Structure

Theorem

Given a finite functional $\mathcal{F} : \mathcal{P}_+ \to \mathbb{R}$, the gradient flow of $\mathcal{F}(\rho)$ in $(\mathcal{P}_+, g_{\rho,\mathcal{C}})$ satisfies

$$\partial_t \rho = \nabla \cdot \left(\rho\, \mathcal{C}(\rho) \nabla \frac{\delta \mathcal{F}}{\delta \rho} \right) .$$

2019-08-13

Gradient Flow Structure

Theorem

Given a finite functional $\mathcal{F} : \mathcal{P}_+ \to \mathbb{R}$, the gradient flow of $\mathcal{F}(\rho)$ in $(\mathcal{P}_+, g_{\rho,\mathcal{C}})$ satisfies

$$\partial_t \rho = \nabla \cdot \left(\rho \mathcal{C}(\rho) \nabla \frac{\delta \mathcal{F}}{\delta \rho} \right) .$$

- This is just formal. To make it rigrous, need theory on all of $\mathcal{P}$.

KW GF for EKS

- **Energy:** Kullback-Leibler divergence

$$\begin{aligned}\mathcal{E}(\rho) &= \int (\Phi_R + \ln \rho(t))\, \rho(t)\, \mathrm{d}\theta \\ &= \int \frac{\rho(t)}{\rho_\infty} \ln\left(\frac{\rho(t)}{\rho_\infty}\right) \rho_\infty\, \mathrm{d}\theta + \ln\left(\int e^{-\Phi_R(\theta)}\, \mathrm{d}\theta\right) \\ &= \mathrm{KL}(\rho(t) \| \rho_\infty) + c\end{aligned}$$

- **Euler-Lagrange condition:**

$$\frac{\delta \mathcal{E}}{\delta \rho} = \Phi_R(\theta) + \ln \rho(\theta) = c \qquad \text{on supp } (\rho)$$

- **Unique solution:** posterior

$$\rho_\infty(\theta) := \frac{e^{-\Phi_R(\theta)}}{\int e^{-\Phi_R(\theta)}\, \mathrm{d}\theta} .$$

- **Energy dissipation:**

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{E}(\rho) = -\int \rho \left|\mathcal{C}(\rho)^{\frac{1}{2}}\nabla(\Phi_R + \ln\rho)\right|^2 \mathrm{d}\theta\,.$$

  Hence $\mathcal{E} \searrow$ along paths until $\mathcal{C}(\rho) = 0$ or $\rho = \rho_\infty$.

- **Fisher-Information:** for any covariance matrix Λ,

$$\mathcal{I}_\Lambda(\rho\|\rho_\infty) := \int \rho \left\langle \nabla \ln\left(\frac{\rho}{\rho_\infty}\right), \Lambda\nabla \ln\left(\frac{\rho}{\rho_\infty}\right)\right\rangle \mathrm{d}\theta\,.$$

- **Kalman-Fisher information:** For $\Lambda = \mathcal{C}(\rho)$,

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathrm{KL}(\rho(t)\|\rho_\infty) = -\mathcal{I}_{\mathcal{C}}(\rho(t)\|\rho_\infty)\,.$$

Convergence to Equilibrium

Non-linear Inverse Problem

Theorem (Decay To Equilibrium)

Assume that $\nabla\Phi_R$, $\alpha > 0$ and $\lambda > 0$ exists such that

$$\mathcal{C}(\rho(t)) \geq \alpha I_d\,, \qquad D^2\Phi_R \geq \lambda I_d\,.$$

If $\mathrm{KL}(\rho_0\|\rho_\infty) < \infty$ then there is $c > 0$ such that

$$\|\rho(t) - \rho_\infty\|_{L^1(\mathbb{R}^d)} \leq c e^{-\alpha\lambda t}\,.$$

- Effect of $\mathcal{C}(\rho)$ in FP: faster convergence.

Proof

- $D^2\Phi_R \geq \lambda I_d$ guarantees log Sobolev inequality [Bakry,Émery 1985]:

$$\mathrm{KL}(\rho(t)\|\rho_\infty) \leq \frac{1}{2\lambda}\mathcal{I}_{I_d}(\rho(t)\|\rho_\infty) \qquad \forall \rho\,.$$

- $C(\rho(t)) \geq \alpha I_d$ gives

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\mathrm{KL}(\rho(t)\|\rho_\infty) &= -\mathcal{I}_{\mathcal{C}}(\rho(t)\|\rho_\infty) \\ &\leq -\alpha\mathcal{I}_{I_d}(\rho(t)|\rho_\infty) \leq -2\alpha\lambda\mathrm{KL}(\rho(t)\|\rho_\infty)\,.\end{aligned}$$

- By Csiszár-Kullback inequality:

$$\frac{1}{2}\|\rho(t)-\rho_\infty\|^2_{L^1(\mathbb{R}^d)} \leq \mathrm{KL}(\rho(t)\|\rho_\infty) \leq \mathrm{KL}(\rho_0\|\rho_\infty)e^{-2\alpha\lambda t}\,.$$

Linear Inverse Problem

We consider the setting where $G(\theta) = A\theta$.

Theorem (Linear Inverse Problem).

The mean field limt of EKS has the property that, if

$$\rho_0(\theta) := \frac{1}{(2\pi)^{d/2}} (\det \mathcal{C}_0)^{-1/2} \exp\left(-\frac{1}{2}\|\theta\|^2_{\mathcal{C}_0}\right)$$

then the solution of the nonlinear Fokker-Planck equation is

$$\rho(t,\theta) := \frac{1}{(2\pi)^{d/2}} (\det \mathfrak{C}(t))^{-1/2} \exp\left(-\frac{1}{2}\|\theta - \mathfrak{m}(t)\|^2_{\mathfrak{C}(t)}\right)$$

where $\mathfrak{m}(t)$ and $\mathfrak{C}(t)$ satisfy explicit ODEs. In particular $\rho(t,\theta)$ converges exponentially fast to $\rho_\infty(\theta)$ in $L^1(\mathbb{R}^d)$ as $t \to \infty$.

- Gaussians remain Gaussians.

Linear Inverse Problem

We consider the setting where $G(\theta) = A\theta$.

Theorem (Linear Inverse Problem).

The mean field limt of EKS has the property that, if

$$\rho_0(\theta) := \frac{1}{(2\pi)^{d/2}}(\det \mathcal{C}_0)^{-1/2} \exp\left(-\frac{1}{2}\|\theta\|^2_{\mathcal{C}_0}\right)$$

then the solution of the nonlinear Fokker-Planck equation is

$$\rho(t,\theta) := \frac{1}{(2\pi)^{d/2}}(\det \mathfrak{C}(t))^{-1/2} \exp\left(-\frac{1}{2}\|\theta - \mathfrak{m}(t)\|^2_{\mathfrak{C}(t)}\right)$$

where $\mathfrak{m}(t)$ and $\mathfrak{C}(t)$ satisfy explicit ODEs. In particular $\rho(t,\theta)$ converges exponentially fast to $\rho_\infty(\theta)$ in $L^1(\mathbb{R}^d)$ as $t \to \infty$.

- Gaussians remain Gaussians.

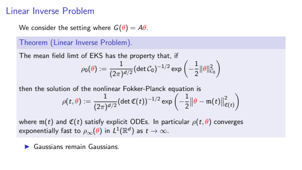
- Main result: FP eqn has exact Gaussian solutions.
- useful since for applications we are often interested in propagating Gaussians.

Proof:

- derive exact closed equations for mean and covariance.
- FP eqn becomes linear.
- explicitly solvable for Gaussian IC.

Consequences:

- For Gaussian IC: $\rho(t) \to \rho_\infty$ since $\mathfrak{m}(t)$ and $\mathfrak{C}(t)$ converge
- Can get same result for non-Gaussian initial conditions!



More details: in linear case,

$$\nabla \Phi_R(\theta) = B^{-1}\theta - r\,,$$

$$r := \frac{1}{\gamma^2} A^\top y \in \mathbb{R}^d\,, \quad B := \left(\frac{1}{\gamma^2} A^\top A + \Gamma_0^{-1}\right)^{-1} \in \mathbb{R}^{d\times d}.$$

- $\frac{1}{\gamma^2} A^\top A + \Gamma_0^{-1}$ invertible since Σ_0^{-1} strictly positive definite (Tikhonov regularizarion).
- $\theta_0 := Br$ solves $\nabla \Phi_R(\theta) = 0$ (solution to regularized normal equations)
- i.e. $\theta_0 = \min \Phi_R$, i.e. $\theta_0 = \max \mathbb{P}(\theta|y)$.
- Note that $\rho_\infty(\theta) \propto \exp\left(-\frac{1}{2}\|\theta - \theta_0\|_B^2\right)$.

Linear Inverse Problem

We consider the setting where $G(\theta) = A\theta$.

Theorem (Linear Inverse Problem).

The mean field limt of EKS has the property that, if

$$\rho_0(\theta) := \frac{1}{(2\pi)^{d/2}} (\det \mathcal{C}_0)^{-1/2} \exp\left(-\frac{1}{2}\|\theta\|^2_{\mathcal{C}_0}\right)$$

then the solution of the nonlinear Fokker-Planck equation is

$$\rho(t,\theta) := \frac{1}{(2\pi)^{d/2}} (\det \mathfrak{C}(t))^{-1/2} \exp\left(-\frac{1}{2}\|\theta - \mathfrak{m}(t)\|^2_{\mathfrak{C}(t)}\right)$$

where $\mathfrak{m}(t)$ and $\mathfrak{C}(t)$ satisfy explicit ODEs. In particular $\rho(t,\theta)$ converges exponentially fast to $\rho_\infty(\theta)$ in $L^1(\mathbb{R}^d)$ as $t \to \infty$.

- Gaussians remain Gaussians.

Closed eqns for moments in linear case:

- For any solution $\rho(t)$, its mean and covariance satisfy

$$\dot{\mathfrak{m}}(t) = -\mathfrak{C}(t)\big(B^{-1}\mathfrak{m}(t) - r\big)\,,$$
$$\dot{\mathfrak{C}}(t) = -2\mathfrak{C}(t)B^{-1}\mathfrak{C}(t) + 2\mathfrak{C}(t)\,.$$

- We derive

$$\frac{\mathrm{d}}{\mathrm{d}t}\det\mathfrak{C}(t) = -2\,(\det\mathfrak{C}(t))\,\mathrm{Tr}\left[B^{-1}\mathfrak{C}(t) - I_d\right]\,,$$

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\mathfrak{C}(t)^{-1}\right) = 2B^{-1} - 2\mathfrak{C}(t)^{-1}.$$

- As a consequence: $\mathfrak{C}(t) \to B$ and $\mathfrak{m}(t) \to \theta_0$ exponentially as $t \to \infty$.
- Solve last eqn explicitly: $\mathfrak{C}(t)^{-1} = \left(\mathfrak{C}(0)^{-1} - B^{-1}\right)e^{-2t} + B^{-1}$.
- From first eqn: $\|\mathfrak{m}(t) - \theta_0\|_{\mathfrak{C}(t)} = \|\mathfrak{m}(0) - \theta_0\|_{\mathfrak{C}(0)}e^{-t}$.

Linear Inverse Problem

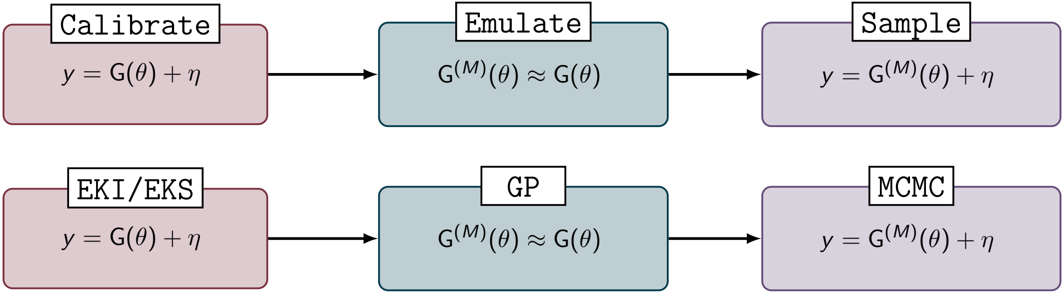
We consider the setting where $G(\theta) = A\theta$.

Also obtain convergence for non-Gaussian IC:

- $D^2\Phi_R(\theta)$ bounded below and independent of θ.
- $\mathfrak{C}(t)$ bounded below as an operator.

$\Longrightarrow$ If $\mathrm{KL}(\rho_0 \| \rho_\infty) < \infty$, then $\rho(t) \to \rho_\infty$ exponentially in $L^1(\mathbb{R}^d)$ and in entropy.

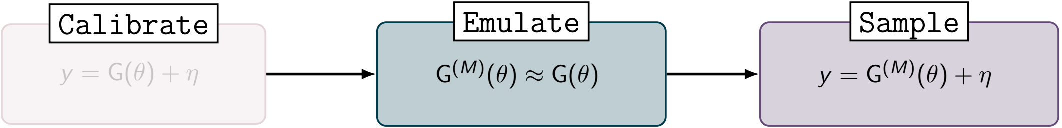
Calibrate, Emulate, Sample



Sample: GP-MCMC

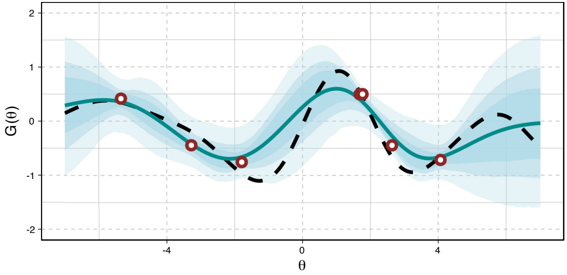
Cleary, Garbuno, Lan, Schneider, Stuart (2019)

arXiv preprint, 1908.+++++ [6]



Gaussian Process Accelerated Sampling

- From EKS we generate approximate posterior samples $\left\{\theta^{(i)}, \mathsf{G}(\theta^{(i)})\right\}_{i=1}^{J}$.
- Use parameter-output pairs to train a Gaussian Process (GP) emulator $\mathsf{G}_J(\cdot)$.

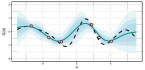


- Define $\Phi_J(\theta; y) = \frac{1}{2\gamma^2}|y - \mathsf{G}_J(\theta)|^2$. Evaluation of Φ_J is fast.
- Sample approximate posterior

$$\mathbb{P}_J(\theta|y) \propto \exp\Big(-\Phi_J(\theta; y)\Big) \times \exp\Big(-\frac{1}{2}\langle\theta, \Sigma_0^{-1}\theta\rangle\Big).$$

Gaussian Process Accelerated Sampling

- From EKS we generate approximate posterior samples $\left\{\theta^{(i)}, \mathrm{G}(\theta^{(i)})\right\}_{i=1}^{J}$.
- Use parameter-output pairs to train a Gaussian Process (GP) emulator $\mathrm{G}_J(\cdot)$.



- Define $\Phi_J(\theta; y) = \frac{1}{2\gamma^2}|y - \mathrm{G}_J(\theta)|^2$. Evaluation of Φ_J is fast.
- Sample approximate posterior

$$\mathbb{P}_J(\theta|y) \propto \exp\Big(-\Phi_J(\theta; y)\Big) \times \exp\Big(-\frac{1}{2}\langle\theta, \Sigma_0^{-1}\theta\rangle\Big).$$

- only need $O(100)$ samples from posterior.
- then interpolate using Gaussian processes.
- to compare, MCMC needs around $O(10^5)$.

Numerics:
EKI vs EKS vs MCMC

Example: Elliptic BVP

- 1D problem for $x \in [0, 1]$,

$$-\frac{\mathrm{d}}{\mathrm{d}x}\left(\exp(\theta_1)\frac{\mathrm{d}}{\mathrm{d}x}p(x)\right) = 1,$$

  with boundary conditions $p(0) = 0$ and $p(1) = \theta_2$.

- Explicit solution is available and we define

$$G(\theta) = \begin{pmatrix} p(x_1) \\ p(x_2) \end{pmatrix}.$$

- Run EKI with $J = 1000$.

Example: Elliptic BVP

- 1D problem for $x \in [0,1]$,

$$-\frac{\mathrm{d}}{\mathrm{d}x}\left(\exp(\theta_1)\frac{\mathrm{d}}{\mathrm{d}x}p(x)\right) = 1,$$

with boundary conditions $p(0) = 0$ and $p(1) = \theta_2$.

- Explicit solution is available and we define

$$G(\theta) = \begin{pmatrix} p(x_1) \\ p(x_2) \end{pmatrix}.$$

- Run EKI with $J = 1000$.

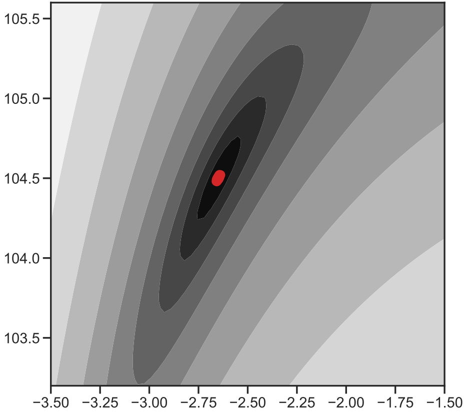


Figure: Contour plots: $\Phi(\theta) = \frac{1}{2\gamma^2}|y - G(\theta)|^2$.

Example: Elliptic BVP

- 1D problem for $x \in [0, 1]$,
$$-\frac{\mathrm{d}}{\mathrm{d}x}\left(\exp(\theta_1)\frac{\mathrm{d}}{\mathrm{d}x}p(x)\right) = 1,$$
with boundary conditions $p(0) = 0$ and $p(1) = \theta_2$.
- Explicit solution is available and we define
$$G(\theta) = \begin{pmatrix} p(x_1) \\ p(x_2) \end{pmatrix}.$$
- Run EKI with $J = 1000$.

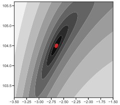


Figure: Contour plots: $\Phi(\theta) = \frac{1}{2\gamma^2}|y - G(\theta)|^2$.

- explicit solution: $p(x) = \theta_2 x + \exp(-\theta_1)\left(-\frac{x^2}{2} + \frac{x}{2}\right)$
- want $\theta = (\theta_1, \theta_2)^\top$ given noisy data $y = (27.5, 79.7)^\top$ at $x_1 = 0.25$ and $x_2 = 0.75$.
- find $\mathbb{P}(\theta|y)$ assuming additive Gaussian noise $\eta \sim \mathbb{N}(0, \Gamma)$, where $\gamma = 0.1$
- prior: $\mathbb{N}(0, \Gamma_0)$, $\Gamma_0 = \sigma^2 I_2$ with $\sigma = 10$.

Example: Elliptic BVP

- 1D problem for $x \in [0,1]$,

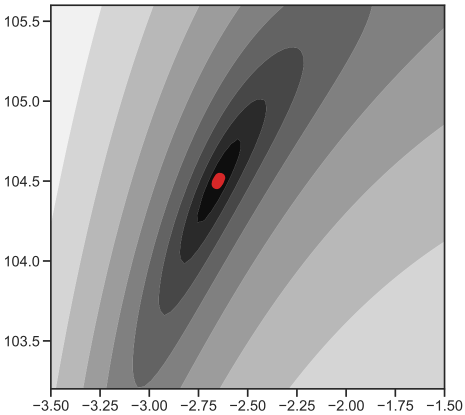
$$-\frac{\mathrm{d}}{\mathrm{d}x}\left(\exp(\theta_1)\frac{\mathrm{d}}{\mathrm{d}x}p(x)\right) = 1,$$

  with boundary conditions $p(0) = 0$ and $p(1) = \theta_2$.

- Explicit solution is available and we define

$$G(\theta) = \begin{pmatrix} p(x_1) \\ p(x_2) \end{pmatrix}.$$

- Run EKS with $J = 1000$.
- Compare with exact MCMC.



Example: Elliptic BVP

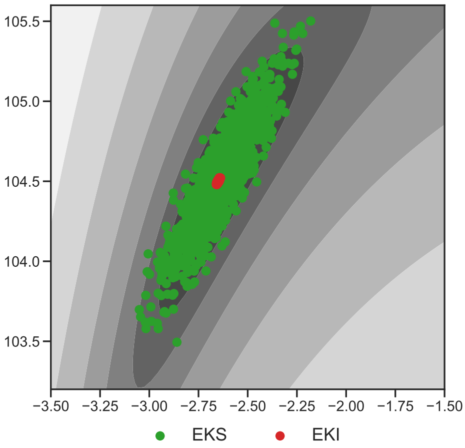
- 1D problem for $x \in [0,1]$,

$$-\frac{\mathrm{d}}{\mathrm{d}x}\left(\exp(\theta_1)\,\frac{\mathrm{d}}{\mathrm{d}x}p(x)\right) = 1,$$

  with boundary conditions $p(0) = 0$ and $p(1) = \theta_2$.
- Explicit solution is available and we define

$$G(\theta) = \begin{pmatrix} p(x_1) \\ p(x_2) \end{pmatrix}.$$

- Run EKS with $J = 1000$.
- Compare with exact MCMC.



Example: Elliptic BVP

- 1D problem for $x \in [0, 1]$,

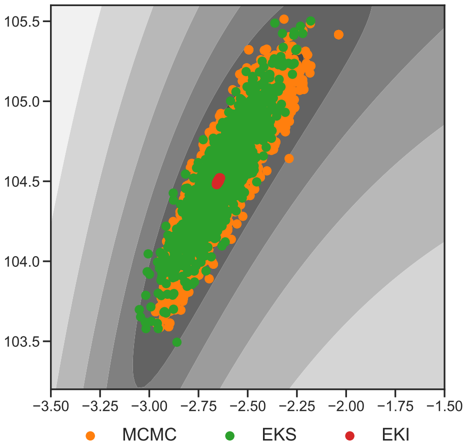
$$-\frac{\mathrm{d}}{\mathrm{d}x}\left(\exp(\theta_1)\frac{\mathrm{d}}{\mathrm{d}x}p(x)\right) = 1,$$

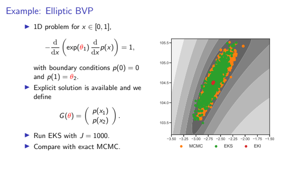
  with boundary conditions $p(0) = 0$ and $p(1) = \theta_2$.

- Explicit solution is available and we define

$$G(\theta) = \begin{pmatrix} p(x_1) \\ p(x_2) \end{pmatrix}.$$

- Run EKS with $J = 1000$.
- Compare with exact MCMC.





EKI:

- collapses to a point very quickly.
- used EKI as propose in [Kovachki, Stuart 2018]

EKS:

- Following [Herty,Visconti 2018], we consider an initial ensemble drawn from $\mathbb{N}(0, 1) \times \mathrm{U}(90, 110)$.
- run for 30 iterations.

MCMC:

- use Random Walk Metropolis Hastings (RWMH) algorithm with $N = 10^5$ samples.
- start of Markov chain = mean of the last step of EKS.
- proposal: Gaussian centered at the current state of the Markov chain with covariance given by $\Sigma = \tau \times C(\theta^*)$, where $C(\theta^*)$ is the

Conclusions

- Introduced an algorithm to generate approximate posterior samples for Inverse Problems inspired by a suitable gradient structure in the mean-field limit
- Related Kalman-Wasserstein metric is a generalization of the Wasserstein distance that depends on the covariance matrix of the particle distribution.
- Described a framework to enable Bayesian inference on expensive and noisy forward models (**Calibrate-Emulate-Sample**).

Conclusions

- Introduced an algorithm to generate approximate posterior samples for Inverse Problems inspired by a suitable gradient structure in the mean-field limit
- Related Kalman-Wasserstein metric is a generalization of the Wasserstein distance that depends on the covariance matrix of the particle distribution.
- Described a framework to enable Bayesian inference on expensive and noisy forward models (**Calibrate-Emulate-Sample**).

- Optimize: Use the algorithm to minimize the flow and reach the equilibrium distribution.
- Learn: Use the samples from the algorithm to learn a surrogate model (a cheaper but equally accurate model as the original) for the expensive forward model G.
- Sample: Use the surrogate model to correct approximation done by the optimization step due to small ensemble size and inherent linear approximation of the method.

Future Directions

- How close are dynamics of EKS to KW flow for non-linear G?
- Application of **Calibrate-Emulate-Sample** to climate models and large-scale inverse problems with expensive noisy forward models.
- Rigorous justification of mean-field limit.
- Properties of Kalman-Wasserstein space and related functional inequalities.
- Convergence to equilibrium in $W_{\mathcal{C}}$ with an explicit rate.
- Instead of $\mathcal{C}(\rho)$ choose general matrix $K(\rho)$. Which choice of $K(\rho)$ provides optimal rates of convergence?

References

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

[2] G. Evensen.
Data Assimilation: The Ensemble Kalman Filter. Springer, 2006.

[3] D.S Oliver, A.C. Reynolds and N. Liu.
Inverse Theory For Petroleum Reservoir Characterization And History Matching. CUP 2008.

[4] M.A. Iglesias, K. Law, and A.M. Stuart.
Ensemble Kalman method for inverse problems, *Inverse Problems*, 29 (4), 2018.

[5] S. Reich.
A non-parametric ensemble transform method for Bayesian inference, *SIAM J. Sci. Comput.*, 35(4), A2013A2024.

[6] 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) 823-851, 2015.

[7] M. Herty and G. Visconti.
Kinetic methods for inverse problems. *arXiv preprint arXiv:1811.09387*, 2018.

[8] A. Garbuno-Inigo, F. Hoffmann, W. Li, and A.M. Stuart.
Interacting Langevin Diffusions: Gradient Structure And Ensemble Kalman Sampler, *arXiv preprint arXiv:1903.08866*, 2019.

[9] W. Li.
Geometry of Probability Simplex via Optimal Transport. *arXiv preprint arXiv:1803.06360*, 2018.