Interacting Langevin Diffusions: Gradient Structure And Ensemble Kalman Sampler

Andrew Stuart

Computing and Mathematical Sciences

Allen Philanthropies, Earthrise Alliance, NSF, ONR, Schmidt Futures.

Le Laboratoire Jacques-Louis Lions 50, Paris, November 27th, 2019,

Overview

Wasserstein Gradient Flows

collaboration with: Garbuno-Hoffmann-Li (2019)

arXiv preprint, 1903.08866 [4]

SIAM Applied Dynamical Systems, To Appear 2020

- Jordan-Kinderlehrer-Otto 1998
- Otto 2000
- Benamou-Brenier 2000
- Ambrosio-Gigli-Savare 2008
- Villani 2008

Linear Fokker-Planck Equation

SDE & Fokker-Planck

$$
\begin{aligned}
\dot{\theta} &= -\nabla\Phi(\theta) + \sqrt{2}\dot{W} \\
\partial_t \rho &= \nabla \cdot \left(\rho \nabla \Phi\right) + \triangle \rho \,, \\
\partial_t \rho &= \nabla \cdot \left(\rho \nabla \left(\Phi + \ln \rho\right)\right) .
\end{aligned}
$$

- Define $\mathcal{P}_+ := \left\{\rho \in \mathcal{P} \,:\, \rho \geq 0 \text{ a.e.}, \ \rho \in C^\infty(\mathbb{R}^d)\right\}$.
- Tangent space at $\rho \in \mathcal{P}_+$:

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

Linear Fokker-Planck Equation 2

Wasserstein Metric $\mathcal{W}\colon \mathcal{P}_+ \times \mathcal{P}_+ \to \mathbb{R}$

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

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

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

Wasserstein Metric Tensor

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

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

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

Linear Fokker-Planck Equation 3

Theorem

The linear Fokker-Planck equation may be written as

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

This gives a gradient flow in $(\mathcal{P}_+, g_\rho)$:

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

Porous Medium Equation – Change The Lyapunov Function

Porous Medium Equation

$$\partial_t \rho = \triangle(\rho^m) \quad \star \quad \partial_t \rho = m\nabla \cdot (\rho^{m-1}\nabla\rho) \quad \star \quad \partial_t \rho = \frac{m}{m-1}\nabla \cdot (\rho\nabla\rho^{m-1}).$$

Theorem

$$\partial_t \rho = \nabla \cdot \left(\rho \nabla \frac{\delta \mathcal{E}}{\delta \rho}\right), \ \mathcal{E}(\rho) = \frac{1}{m-1}\int \rho^m \, \mathrm{d}\theta.$$

This gives a gradient flow in $(\mathcal{P}_+, g_\rho)$:

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{E}(\rho) = -\int \rho \left| \frac{m}{m-1}\nabla(\rho^{m-1})\right|^2 \mathrm{d}\theta$$
$$= -g_\rho(\partial_t\rho, \partial_t\rho).$$

Nonlinear Fokker-Planck Equation – Change The Metric

Mean Field SDE & Nonlinear Fokker-Planck Equation

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

$$\partial_t\rho = \nabla \cdot \left(\rho\,\mathcal{C}(\rho)\,\nabla\Phi\right) + \mathcal{C}(\rho) : D^2\rho\,, \qquad \mathcal{C}(\rho) = \int (\theta - \bar{\theta}) \otimes (\theta - \bar{\theta})\rho(\theta, t)d\theta$$

$$\partial_t\rho = \nabla \cdot \left(\rho\,\mathcal{C}(\rho)\,\nabla\left(\Phi + \ln\rho\right)\right).$$

- Define $\mathcal{P}_+ := \{\rho \in \mathcal{P} \,:\, \rho \geq 0 \text{ a.e.},\; \rho \in C^\infty(\mathbb{R}^d)\}$.
- Tangent space at $\rho \in \mathcal{P}_+$:

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

Nonlinear Fokker-Planck Equation 2

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

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

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

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

Kalman-Wasserstein Metric Tensor

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

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

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

Nonlinear Fokker-Planck Equation 3

Theorem

The nonlinear Fokker-Planck equation may be written as

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

This gives a gradient flow in $(\mathcal{P}_+, g_{\rho,\mathcal{C}})$:

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

Inverse Problems

- Engl-Hanke-Neubauer 1998
- Kaipio-Somersalo 2006

Inverse Problem For Parameters

Find Parameter θ From Data y

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

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

Our Setting

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

Optimization Approach

Formulation

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

$$\Phi_0(\theta; y) = \frac{1}{2\gamma^2} |y - \mathsf{G}(\theta)|^2,$$

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

Algorithms: parameter θ calibration.

Bayesian Approach

Formulation

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

$$\mathbb{P}(\theta|y) \propto \exp\Big(-\Phi_0(\theta; y)\Big) \times \exp\Big(-\frac{1}{2}\langle\theta, \Sigma^{-1}\theta\rangle\Big)$$

$$\propto \exp\Big(-\Phi(\theta; y)\Big)$$

Algorithms: parameter θ sampling.

Ensemble Kalman Methodology

collaboration with: Schillings (2017)

SIAM J Num. Analysis 2017 [2]

- Evensen 1994
- Chen-Oliver 2002
- Emerick-Reynolds 2013
- Iglesias-Law-S 2013
- Ernst-Sprungk-Starkloff 2015
- Chada-Tong-S 2019
- Garbuno-Hoffmann-Li-S 2019

EKI: Basic Discrete Time Algorithm (Iglesias et al [1] (2013))

$$\theta_{n+1}^{(j)} = \theta_n^{(j)} + C_n^{\theta y}(C_n^{yy} + \gamma^2 I)^{-1}\left(y - G(\theta_n^{(j)})\right),$$

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

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

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

Ensemble Kalman Inversion (EKI)

Continuous Time Formulation (Schillings and S [2] (2015))

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

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

Tikhonov-Regularized Ensemble Kalman Inversion (TEKI)

(Chada et al [3] (2019))

Continuous Time TEKI

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

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

The EKS & Wasserstein Gradient Flows

collaboration with: Garbuno-Hoffmann-Li (2019)

arXiv preprint, 1903.08866 [4]

SIAM Applied Dynamical Systems, To Appear 2020

- Reich 2011
- Leimkuhler-Matthews-Weare 2018
- Herty-Visconti 2018
- Wuchen Li 2018
- Nüsken-Reich 2019
- Carrillo-Vaes 2019
- Ding-Qin Li 2019

Ensemble Kalman Sampling (EKS) (Garbuno et al [4] (2019))

Continuous Time Formulation

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

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

EKS: Approximation 1

Linear Approximation (exact if G is linear)

$$(\mathrm{G}(\theta^{(k)}) - \bar{\mathrm{G}}) \approx D\mathrm{G}(\theta^{(j)})(\theta^{(k)} - \bar{\theta}), \quad C(\theta) = \frac{1}{J}\sum_{k=1}^{J}\left(\theta^{(k)} - \bar{\theta}\right) \otimes \left(\theta^{(k)} - \bar{\theta}\right).$$

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

Linear Approximation (Self-Preconditioned Langevin Equations)

$$\dot{\theta}^{(j)} = -C(\theta)\nabla\Phi_0(\theta^{(j)}) - C(\theta)\Sigma^{-1}\theta^{(j)} + \sqrt{2C(\theta)}\dot{W}^{(j)}.$$

EKS: Approximation 2

Mean Field Limit $J \to \infty$ (Preconditioned Langevin-McKean Equation)

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

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

Nonlinear Fokker-Planck equation

$$\partial_t \rho = \nabla \cdot \big(\rho\, \mathcal{C}(\rho)\, \nabla\Phi\big) + \mathcal{C}(\rho) : D^2\rho,$$

$$\Phi(\theta) = \frac{1}{2\gamma^2}\big|y - G(\theta)\big|^2 + \frac{1}{2}\big|\Sigma^{-\frac{1}{2}}\theta\big|^2.$$

Connection To Bayesian Inversion

Steady Solution

Equilibrium solution to nonlinear Fokker-Planck equation:

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

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

Theorem (Decay To Equilibrium)(Garbuno et al [4] (2019)).

Assume that there exist $\alpha, \lambda > 0$ such that, as quadratic forms,

$$\inf_{t>0} C(\rho(t)) \geq \alpha I_d\,, \qquad D^2\Phi \geq \lambda I_d\,.$$

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

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

Linear Inverse Problem

Theorem (Linear Inverse Problem: No Ensemble Collapse) (Garbuno et al [4] (2019)).

Consider the linear setting where $G(\theta) = A\theta$. 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: a) $\mathfrak{C}(t)$ is bounded away from zero; and b) $\rho(t,\theta)$ converges exponentially fast to $\rho_\infty(\theta)$ in $L^1(\mathbb{R}^d)$ as $t \to \infty$.

Applications

collaboration with: Cleary, Garbuno, Lan, Schneider & S (2019)

arXiv preprint, 1912.++++++ [5]

Example 1 – Lorenz '63

Forward Model

$$
\begin{aligned}
\dot{x} &= 10\,(y - x) \\
\dot{y} &= r\,x - y - x\,z \\
\dot{z} &= x\,y - b\,z
\end{aligned}
$$

- 2-dimensional unknown: $\theta = [r, b]$
- G comprises moments in statistical equilibrium.
- G only available to us approximately via finite time averages.

Example 2: Idealized Moist GCM (O'Gorman/Schneider '86)

Forward Model

$$\text{Moisture Conservation:} \quad \frac{\partial q}{\partial t} + \mathbf{v} \cdot \nabla q = -\frac{q - q_{ref}(T;\theta)}{\tau_q(q, T;\theta)}$$

$$\text{Energy Conservation:} \quad \frac{\partial T}{\partial t} + \mathbf{v} \cdot \nabla T = \frac{T - T_{ref}(q, T;\theta)}{\tau_T(q, T;\theta)} + \text{RAD} + \ldots$$

- Coupled with mass and momentum conservation equations.
- Model adapt to equilibrate precipitation rates: $P_q = P_T$.
- $\mathcal{O}(10^5)$ unknowns
- Employs 2 unknown parameters:
  - θ_{RH}: reference relative humidity
  - θ_τ: default relaxation timescale

Example 2: Idealized Moist GCM

Objective Function

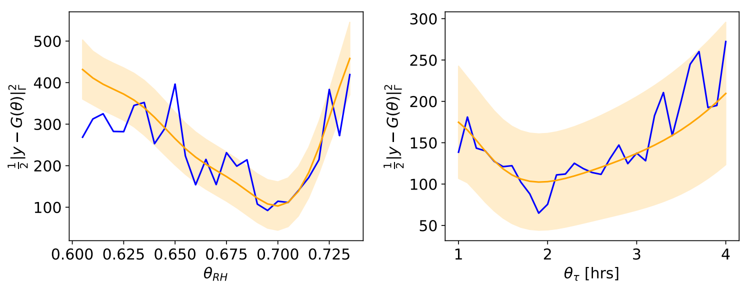

Figure: GP trained from EKI data

Dark orange: GP mean; shaded orange: GP 2 std.

Closing

Conclusions

- Wasserstein Gradient Flows – change energy or metric.
- Inverse Problems – calibration and sampling.
- EKI – calibration.
- EKS – sampling.
- Applications – climate models.

References

[1] M.A. Iglesias, K. Law, and A.M. Stuart. [EKI]

Ensemble Kalman method for inverse problems,

Inverse Problems, 2013.

[2] C. Schillings, A. M. Stuart. [EKI-Analysis]

Analysis of the ensemble Kalman filter for inverse problems,

SIAM Num. Analysis, 2017.

[3] N.K. Chada, A.M. Stuart, and X.T. Tong. [EKI-Tikhonov]

Tikhonov regularizarion within ensemble Kalman inversion,

arXiv:1901.10382).

[4] A. Garbuno-Inigo, F. Hoffmann, W. Li, and A.M. Stuart. [EKS]

Gradient structure of the ensemble kalman flow with noise.

arXiv:1903.08866.

[5] E. Cleary, A. Garbuno-Inigo, S. Lan, T. Schneider and A.M. Stuart.

Calibrate, Emulate, Sample.

arXiv:1908.+++++.

[6] Wuchen Li.

Geometry of probability simplex via optimal transport

arXiv:1803.06360.