Solving Fredholm Integral Equations via Wasserstein Gradient Flows

Francesca R. Crucinio (CREST, ENSAE, IP Paris)

Joint work with Valentin De Bortoli (CNRS and ENS), Arnaud Doucet (Oxford) and Adam Johansen (Warwick)

More info: arXiv preprint arXiv:2209.09936

Introduction and motivation

Fredholm equations

- Fredholm integral equation of the **first kind**

$$\mu(y) = \int_{\mathbb{R}^d} \mathrm{k}(x, y) \mathrm{d}\pi(x) := \pi[\mathrm{k}(\cdot, y)], \qquad y \in \mathbb{R}^p$$

  model: electromagnetic scattering, image reconstruction, density deconvolution...

- Fredholm integral equation of the **second kind**

$$\pi(y) = \lambda \int_{\mathbb{R}^d} \mathrm{k}(x, y) \mathrm{d}\pi(x) + \varphi(x)$$

  model: reinforcement learning, optimal control, light transport...

Fredholm equations

- Fredholm integral equation of the **first kind**

$$\mu(y) = \int_{\mathbb{R}^d} \mathrm{k}(x,y)\mathrm{d}\pi(x) := \pi[\mathrm{k}(\cdot,y)], \qquad y \in \mathbb{R}^p$$

  model: electromagnetic scattering, image reconstruction, density deconvolution...

- Fredholm integral equation of the **second kind**

$$\pi(y) = \lambda \int_{\mathbb{R}^d} \mathrm{k}(x,y)\mathrm{d}\pi(x) + \varphi(x)$$

  model: reinforcement learning, optimal control, light transport...

- μ = observed probability measure over $\mathbb{R}^p$ – *known*
- φ = forcing function $\mathbb{R}^d$ – *known*
- $\lambda \in \mathbb{R}$ – *known*
- k = integral kernel – *known*
- π = probability measure to recover (over $\mathbb{R}^d$) – *unknown*

The functional

- Fredholm integral equation of the **first kind**

$$\mathsf{KL}\left(\mu \middle| \int_{\mathbb{R}^d} \mathrm{k}(\cdot, y)\mathrm{d}\pi(y)\right) + \alpha\, \mathsf{KL}\left(\pi|\pi_0\right)$$

- Fredholm integral equation of the **second kind**

$$\mathsf{KL}\left(\pi \middle| \varphi + \lambda \int_{\mathbb{R}^d} \mathrm{k}(\cdot, y)\mathrm{d}\pi(y)\right) + \alpha\, \mathsf{KL}\left(\pi|\pi_0\right)$$

where $\mathsf{KL}\left(\nu_1|\nu_2\right) = \int_{\mathbb{R}^d} \log(\nu_1(z))\mathrm{d}\nu_1(z) - \int_{\mathbb{R}^d} \log(\nu_2(z))\mathrm{d}\nu_1(z)$.

Assumptions

- k smooth with bounded derivatives
- π_0
  - admits density w.r.t. Leb, $(\mathrm{d}\pi_0/\mathrm{dLeb}_d)(x) \propto \exp[-U(x)]$
  - with Lipschitz continuous first and second derivatives
- μ has finite second moment
- φ is positive, smooth with bounded derivatives
- λ is positive

Workflow

$\mathscr{F}_\alpha(\pi)$ $\xrightarrow{\text{Wasserstein gradient flow}}$ PDE $\xrightarrow{\text{Itô's lemma}}$ SDE

SDE $\xrightarrow{\text{Particles}}$ Interacting Particle System $\xrightarrow{\text{Time Discretization}}$ Numerical Scheme

Another formulation

Recall that we want to minimize

$$\mathsf{KL}\left(\mu \middle| \int_{\mathbb{R}^d} \mathrm{k}(\cdot, y)\mathrm{d}\pi(y)\right) + \alpha\,\mathsf{KL}\left(\pi|\pi_0\right).$$

We consider instead

$$\boxed{\mathscr{F}_\alpha^\eta(\pi) = -\int_{\mathbb{R}^d} \log(\pi[\mathrm{k}(\cdot, y)] + \eta)\mathrm{d}\mu(y) + \alpha\,\mathsf{KL}\left(\pi|\pi_0\right)}$$

with $\eta > 0$ to ensure **stability** and boundedness.

Minimizers

Regularity/convexity properties

For any $\alpha, \eta > 0$, $\mathscr{F}_\alpha^\eta$ is proper, strictly convex, coercive and lower semi-continuous. In particular, $\mathscr{F}_\alpha^\eta$ admits a **unique minimizer** $\pi^\star_{\alpha,\eta} \in \mathcal{P}(\mathbb{R}^d)$.

Wasserstein Gradient Flow and McKean–Vlasov SDE

Subdifferential

Recall that we want to **minimize**

$$\boxed{\mathscr{F}_\alpha^\eta(\pi) = -\int_{\mathbb{R}^d} \log(\pi[\mathrm{k}(\cdot, y)] + \eta)\mathrm{d}\mu(y) + \alpha\,\mathrm{KL}\,(\pi|\pi_0)}$$

(we focus on the case where $\alpha, \eta > 0$).

Subdifferential of $\mathscr{F}_\alpha^\eta$ is given by

$$\boxed{\partial_{\mathrm{s}}\mathscr{F}_\alpha^\eta(x) = -\int_{\mathbb{R}^p} \nabla_1\mathrm{k}(x, y)/(\pi[\mathrm{k}(\cdot, y)] + \eta)\mathrm{d}\mu(y) + \alpha\nabla\log\pi(x)/\pi_0(x)\,.}$$

Wasserstein gradient flow

A Wasserstein gradient flow for $\mathscr{F}_\alpha^\eta$ is given by $(\pi_t)_{t\geq 0}$

$$\boxed{\partial\pi_t = -\mathrm{div}((b^\eta - \alpha\nabla U)\pi_t) + \alpha\Delta\pi_t \;.}$$

where

$$\boxed{b^\eta(x,\pi) = \int_{\mathbb{R}^p} \nabla_1 \mathrm{k}(x,y)/(\pi[\mathrm{k}(\cdot,y)] + \eta)\mathrm{d}\mu(y) \;.}$$

No guarantee of convergence via standard methods (Ambrosio et al., 2008) since $\mathscr{F}_\alpha^\eta$ is not strongly geodesically convex

McKean-Vlasov SDE whose law **converges to the unique minimizer** of $\mathscr{F}_{\alpha}^{\eta}$:

$$\boxed{\mathrm{d}X_t^\star = \left\{b^\eta(X_t^\star, \lambda_t^\star) - \alpha \nabla U(X_t^\star)\right\} \mathrm{d}t + \sqrt{2\alpha}\mathrm{d}B_t \ ,}$$

where

- $(B_t)_{t\geq 0}$ Brownian motion
- $(\lambda_t^\star)_{t\geq 0}$ is the distribution of $X_t^\star$
- $b^\eta(x, \pi) = \int_{\mathbb{R}^p} \nabla_1 \mathrm{k}(x, y)/(\pi[\mathrm{k}(\cdot, y)] + \eta)\mathrm{d}\mu(y)$

Convergence of the McKean-Vlasov process

Existence and uniqueness

Under the previous assumptions, there exists a unique strong solution to the McKean-Vlasov equation for any initial condition X_0 with $\mathbb{E}[\|X_0\|^2] < +\infty$.

Convergence of the McKean-Vlasov process

Under the previous assumptions we have

$$\lim_{t \to +\infty} \mathcal{W}_2(\lambda_t^\star, \pi_{\alpha,\eta}^\star) = 0 \; .$$

Results due to Hu et al. (2019). Contrary to previous works use the fact that $\lambda_t^\star$ is a gradient flow for $\mathscr{F}_\alpha^\eta$ and that $\alpha > 0$.

Interacting Particle System and Numerical Scheme

Approximation via particle systems

For any $N \in \mathbb{N}$ and $k \leq N$

$$\boxed{\mathrm{d}X_t^{k,N} = \left\{ b^\eta(X_t^{k,N}, \lambda_t^N) - \alpha \nabla U(X_t^{k,N}) \right\} \mathrm{d}t + \sqrt{2\alpha}\mathrm{d}B_t^k \ ,}$$

- $\{(B_t^k)_{t\geq 0} \ : \ k \in \mathbb{N}\}$ independent Brownian motion
- $\lambda_t^N = (1/N) \sum_{k=1}^N \delta_{X_t^{k,N}}$ is the **empirical measure.**

Classical **propagation of chaos** results (Sznitman, 1991). Particle systems approximate McKean-Vlasov for large $N \in \mathbb{N}$ for any finite time horizon, $\lim_{N\to+\infty} \mathcal{L}(X_t^{1,N}) = \mathcal{L}(X_t^\star)$ at rate $N^{-1/2}$.

Geometric ergodicity and approximation

For any $N \in \mathbb{N}$ **geometric ergodicity** holds

Geometric ergodicity

Under the previous assumptions, for any $N \in \mathbb{N}$, there exist $C_N \geq 0$, $\rho_N \in [0, 1)$ such that for any $t \geq 0$

$$\mathcal{W}_1(\lambda_t^N(x_1^{1:N}), \lambda_t^N(x_2^{1:N})) \leq C_N \rho_N^t \|x_1^{1:N} - x_2^{1:N}\|.$$

In particular, the particle system admits a unique invariant probability measure π^N.

- $\lim_{N \to +\infty} C_N = +\infty$ and $\lim_{N \to +\infty} \rho_N = 1$

Approximation of the target measure

Under the previous assumptions, $\lim_{N \to +\infty} \mathcal{W}_1(\pi^N, \pi^\star_{\alpha,\eta}) = 0$, the unique minimizer of $\mathscr{F}_\alpha^\eta$.

Discretization and numerical implementation

Euler-Maruyama discretization. For any $N \in \mathbb{N}$ and $k \leq N$

$$\tilde{X}_{n+1}^{k,N} = \tilde{X}_n^{k,N} + \frac{\gamma b^\eta(\tilde{X}_n^{k,N}, \lambda_n^N)}{1+\gamma \left\| b^\eta(\tilde{X}_n^{k,N}, \lambda_n^N) \right\|} - \gamma\alpha \nabla U(\tilde{X}_n^{k,N}) + \sqrt{2\alpha\gamma} Z_{n+1}^k \quad .$$

For stability issues, we consider a **tamed version**

Strong convergence (Bao et al., 2020)

Under the previous assumptions, for any $N \in \mathbb{N}$, any $\eta, \alpha > 0$ and any $T \geq 0$ there exists $C_T \geq 0$ such that

$$\mathbb{E}[\sup_{n \in \{0,\dots,n_T\}} \|\tilde{X}_n^{k,N} - X_n^{k,N}\|] \leq C_T \gamma.$$

for all $k \in \{1, \dots, N\}$

Practicalities

- smooth reconstructions obtained by kernel density estimation
- π_0 used as "prior" to guarantee smoothness/sparsity (influences shape of reconstruction not speed of convergence)
- α selected by cross validation
- choice of N, γ

$$\mathbb{E}[\sup_{n \in \{0,\dots,n_T\}} \|X_n^\star - \tilde{X}_n^{k,N}\|] \leq C_T(N^{-1/2} + \gamma).$$

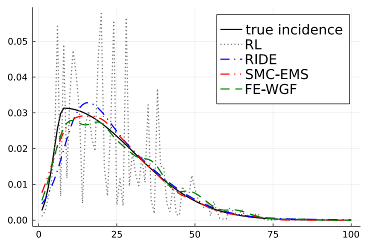
Experiments

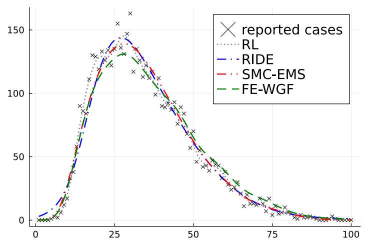
First kind: Epidemiology

$$\mu(y) = \int_{\mathbb{R}^d} \mathrm{k}(x, y) \mathrm{d}\pi(x) := \pi[\mathrm{k}(\cdot, y)], \qquad y \in \mathbb{R}^p$$

- μ = distribution of hospitalisations over time
- k = delay between infection and hospitalisation
- π = distribution of infections over time

- Comparing with **Richardson-Lucy algorithm/ EM**, **robust incidence deconvolution estimator (RIDE)** and **SMC-EMS**, a sequential Monte Carlo implementation of EM + Smoothing





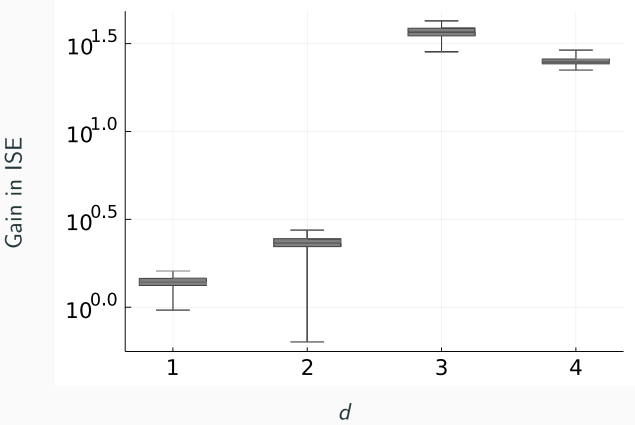
First kind: Epidemiology

	Well-specified		
Method	ISE(π)	ISE(μ)	runtime (s)
RIDE	$9.0 \cdot 10^{-4}$	$3.4 \cdot 10^{-4}$	58
SMC-EMS	$3.3 \cdot 10^{-4}$	$2.5 \cdot 10^{-4}$	3
FE-WGF	$2.7 \cdot 10^{-4}$	$2.5 \cdot 10^{-4}$	96

	Misspecified		
Method	ISE(π)	ISE(μ)	runtime (s)
RIDE	$1.0 \cdot 10^{-3}$	$3.4 \cdot 10^{-4}$	58
SMC-EMS	$3.7 \cdot 10^{-4}$	$2.5 \cdot 10^{-4}$	3
FE-WGF	$3.1 \cdot 10^{-4}$	$2.5 \cdot 10^{-4}$	95

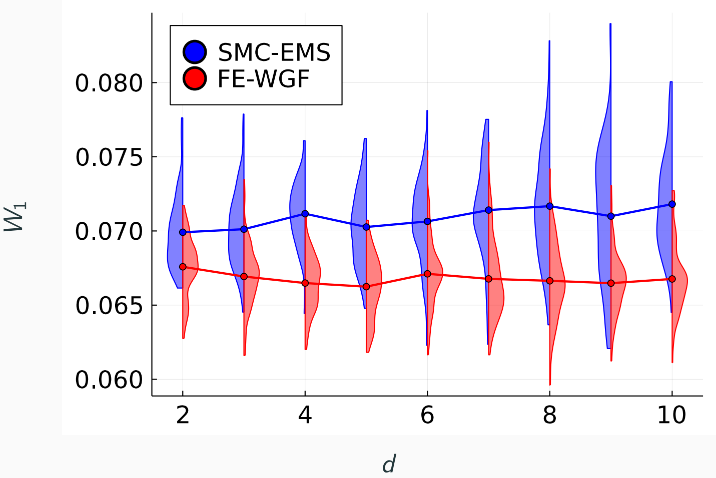
First kind: Scaling with dimension

- Multidimenesional deconvolution problem $(\mathrm{k}(x, y) = \mathrm{k}(y - x))$
- recover the density of X from observations with additive noise $Y = X + \epsilon$
- comparing with **one-step-late Expectation Maximization**



First kind: Scaling with dimension

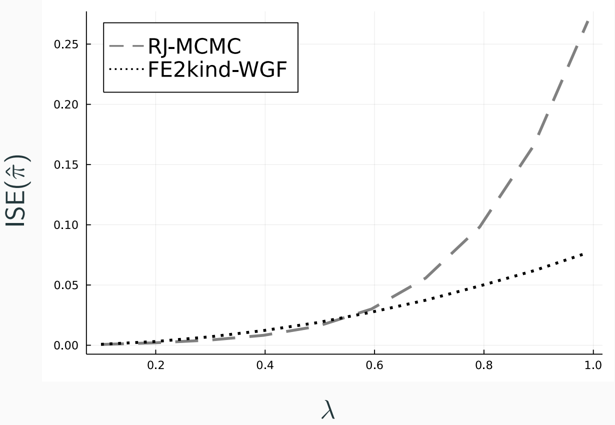
- comparing with **SMC-EMS**



Second kind: Toy model

$$\pi(x) = \varphi(x) + \lambda \int_{\mathbb{R}^d} \mathrm{k}(x, y)\mathrm{d}\pi(y)$$

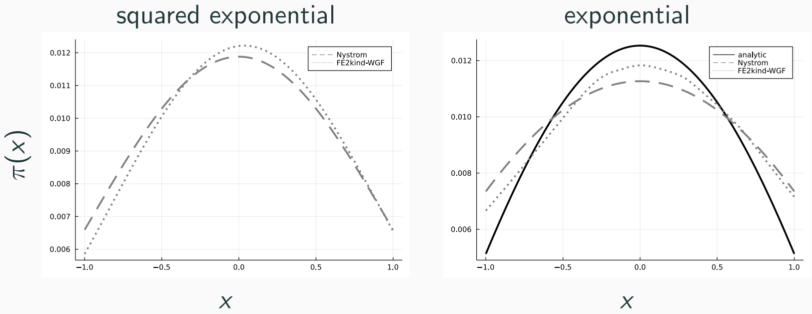
- comparing with **Von-Neumann expansion**



Second kind: Karhunen–Loeve Expansions

$$\pi(x) = \lambda \int_{\mathbb{R}^d} \mathrm{k}(x, y) \mathrm{d}\pi(y)$$

- comparing with **Nyström method**



Connections

Tikhonov's Regularization

Minimizing

$$\mathsf{KL}\left(\mu \middle| \int_{\mathbb{R}^d} \mathrm{k}(\cdot, y)\mathrm{d}\pi(y)\right) + \alpha\,\mathsf{KL}\left(\pi|\pi_0\right)$$

is a probabilistic analogous to **Tikhonov regularization**

$$\min\left\{\left\|\mu - \int_{\mathbb{R}^d} \mathrm{k}(\cdot, y)\mathrm{d}\pi(y)\right\|^2 + \alpha\left\|\pi - \pi_0\right\|^2 \,:\, \pi \in \mathbb{L}^2(\mathbb{R}^d)\right\}\,.$$

In the limit $\lim_{n\to+\infty}\alpha_n = 0$, $\lim_{n\to+\infty}\eta_n = 0$ we have

$$\pi^\star \in \arg\min\{\mathsf{KL}\left(\pi|\pi_0\right) \,:\, \pi \in \arg\min_{\mathcal{P}_2(\mathbb{R}^d)} \mathscr{F}_0^0\}.$$

Maxent methods

The functional $\mathscr{F}_\alpha$ can be seen as the Lagrangian associated with the following primal problem

$$\arg\min\{\mathsf{KL}\left(\pi|\pi_0\right) \,:\, \pi \in \mathcal{P}(\mathbb{R}^d),\; \mathsf{KL}\left(\mu\middle|\int_{\mathbb{R}^d} \mathrm{k}(\cdot, y)\mathrm{d}\pi(y)\right) = 0\}.$$

Closely related to

$$\arg\max\{\mathrm{H}(\pi) \,:\, \pi \in \mathcal{P}_{\mathrm{H}}(\mathbb{R}^d),\; \mathsf{KL}\left(\mu\middle|\int_{\mathbb{R}^d} \mathrm{k}(\cdot, y)\mathrm{d}\pi(y)\right) = 0\},$$

Conclusion

Conclusions

Standard techniques

- require discretization of the domain and/or approximate π with a linear combination of basis functions
- require discretization of μ
- impractical as dimension increases
- Require a specific form of k (e.g. convolution kernel)

Gradient flows allow

- adaptive stochastic discretizations
- natural implementation when we only have samples from μ
- tackling higher dimensional problems

Conclusions

Standard techniques

- require discretization of the domain and/or approximate π with a linear combination of basis functions
- require discretization of μ
- impractical as dimension increases
- Require a specific form of k (e.g. convolution kernel)

Gradient flows allow

- adaptive stochastic discretizations
- natural implementation when we only have samples from μ
- tackling higher dimensional problems

Thank you!

References

Luigi Ambrosio, Nicola Gigli, and Giuseppe Savaré. *Gradient flows: in metric spaces and in the space of probability measures*. Springer Science & Business Media, 2008.

J Bao, C Reisinger, P Ren, and W Stockinger. First order convergence of Milstein schemes for McKean-Vlasov equations and interacting particle systems. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 2020.

Kaitong Hu, Zhenjie Ren, David Siska, and Lukasz Szpruch. Mean-field langevin dynamics and energy landscape of neural networks. *arXiv preprint arXiv:1905.07769*, 2019.

Alain-Sol Sznitman. Topics in propagation of chaos. In *Ecole d'été de probabilités de Saint-Flour XIX—1989*, volume 1464 of *Lecture Notes in Mathematics*, pages 165–251. Springer, Berlin, 1991.