Solving Fredholm Integral Equations of the First Kind via Wasserstein Gradient Flows

Francesca R. Crucinio (CREST, ENSAE)

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

Introduction and motivation

Fredholm integral of the first kind

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

$$\mu = \pi \mathrm{K} \, , \quad \pi \mathrm{K}(A) = \int_{\mathbb{R}^d} \mathrm{K}(x, A) \mathrm{d}\pi(x) \, .$$

  - $\mu =$ observed probability measure over $\mathbb{R}^p$ – *known*
  - $\mathrm{K} =$ Markov kernel – *known*
  - $\pi =$ probability measure to recover (over $\mathbb{R}^d$) – *unknown*

- Model the **inverse problem** of reconstructing a signal from distorted/noisy observations
- Applications in applied maths (e.g. electromagnetic scattering), engineering (e.g. image reconstruction), statistics (e.g. density deconvolution, epidemiology), ...

Regularization

Fredholm integral equations are **ill-posed** (non-uniqueness of the solution)

- **Tikhonov regularization**

$$\boxed{\pi^\star = \arg\min\left\{\|\mu - \pi\mathrm{K}\|^2 + \alpha\,\|\pi - \pi_0\|^2 \;:\; \pi \in \mathbb{L}^2(\mathbb{R}^d)\right\}\ .}$$

- We consider a **probabilistic formulation**

$$\boxed{\pi^\star = \arg\min\left\{\mathscr{F}_\alpha(\pi) = \mathrm{KL}\left(\mu|\pi\mathrm{K}\right) + \alpha\,\mathrm{KL}\left(\pi|\pi_0\right) \;:\; \pi \in \mathcal{P}(\mathbb{R}^d)\right\}\ .}$$

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

Previous work and goals

Standard techniques[1]

- 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)

Our goal: deriving a method which

- does not require a finite dimensional decomposition
- can be naturally implemented when we only have samples from μ
- has the potential to tackle higher dimensional problems
- allows the introduction of prior knowledge on the solution

[1] With the exception of SMC-EMS (Crucinio et al., 2021)

Entropic regularization and stability

Assumptions

- μ has finite second moment
- K
  - admits density w.r.t. Leb, k
  - $\mathrm{k} \in \mathrm{C}^3(\mathbb{R}^d \times \mathbb{R}^p, [0, +\infty))$, bounded and with bounded derivatives
  - Gaussian-like: $\mathrm{k}(x, y) \geq C_1^{-1} \exp[-C_1(1 + \|x\|^2 + \|y\|^2)]$
- 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
  - Gaussian-like: $-C_2 + \tau \|x\|^2 \leq U(x) \leq C_2 + \tau^{-1} \|x\|^2$

Workflow

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

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

Another formulation

- Recall that

$$\boxed{\pi^\star = \arg\min \left\{ \mathscr{F}_\alpha(\pi) = \mathsf{KL}\left(\mu|\pi\mathrm{K}\right) + \alpha\, \mathsf{KL}\left(\pi|\pi_0\right) \;:\; \pi \in \mathcal{P}(\mathbb{R}^d) \right\}\ .}$$

- Use that $\mathsf{KL}\left(\mu|\pi\mathrm{K}\right) = \int_{\mathbb{R}^d} \log(\mu(y))\mathrm{d}\mu(y) - \int_{\mathbb{R}^d} \log(\pi\mathrm{K}(y))\mathrm{d}\mu(y)$ to write

$$\mathscr{G}_\alpha = -\int_{\mathbb{R}^d} \log(\pi\mathrm{K}(y))\mathrm{d}\mu(y) + \alpha\, \mathsf{KL}\left(\pi|\pi_0\right)\ .$$

- To ensure **stability** and boundedness

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

for $\eta > 0$

Minimizers

Regularity/convexity properties

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

Convergence of minimizers

Then, for any $\alpha > 0$, $\eta \geq 0$, $\pi^\star_{\alpha,\eta} \in \mathcal{P}_2(\mathbb{R}^d)$ and $\pi^\star_{\alpha,\eta}$ admits a density w.r.t the Lebesgue measure. If there exists $\pi^\star \in \mathcal{P}_2(\mathbb{R}^d)$ such that $\lim_{n\to+\infty} \alpha_n = 0$, $\lim_{n\to+\infty} \eta_n = 0$ and $\lim_{n\to+\infty} \mathcal{W}_2(\pi^\star_{\alpha_n,\eta_n}, \pi^\star) = 0$,

$$\boxed{\pi^\star \in \arg\min \left\{ \mathrm{KL}\left(\pi|\pi_0\right) \ : \ \pi \in \arg\min_{\mathcal{P}_2(\mathbb{R}^d)} \mathscr{G} \right\} \ .}$$

Stability

Stability

Let $\alpha, \eta > 0$ and for any $\nu \in \mathcal{P}(\mathbb{R}^d)$ denote $\pi_{\alpha,\eta}^{\nu,\star}$ the unique minimizer of $\mathscr{G}_\alpha^\eta$ with $\mu \leftarrow \nu$. Let $(\mu_n)_{n\in\mathbb{N}} \in (\mathcal{P}_2(\mathbb{R}^d))^{\mathbb{N}}$ such that $\lim_{n\to+\infty} \mathcal{W}_1(\mu_n, \mu) = 0$ with $\mu \in \mathcal{P}_2(\mathbb{R}^d)$. Then, we have $\lim_{n\to+\infty} \mathcal{W}_1(\pi_{\alpha,\eta}^{\mu_n,\star}, \pi_{\alpha,\eta}^{\mu,\star}) = 0$.

Justifies replacing μ by $\mu^M = (1/M)\sum_{k=1}^M \delta_{y^{k,N}}$, $y^{1:N} \sim \mu^{\otimes M}$.

Wasserstein Gradient Flow PDE

Wasserstein gradient flow

Recall that we want to **minimize**

$$\boxed{\mathscr{G}_\alpha^\eta(\pi) = -\int_{\mathbb{R}^d} \log(\pi\mathrm{K}(y) + \eta)\mathrm{d}\mu(y) + \alpha\,\mathsf{KL}\,(\pi|\pi_0)}$$

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

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

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

where

$$\boxed{\bar{b}^\eta(x,\pi) = -\int_{\mathbb{R}^p} \nabla_x \mathrm{k}(x,y)/(\pi\mathrm{K}(y) + \eta)\mathrm{d}\mu(y)\ .}$$

McKean-Vlasov equations and particle system

McKean-Vlasov SDE

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

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

where

- $(B_t)_{t \geq 0}$ Brownian motion
- $(\lambda^{\star}_t)_{t \geq 0}$ is the distribution of $X^{\star}_t$
- $\bar{b}^{\eta}(x, \pi) = -\int_{\mathbb{R}^p} \nabla_x \mathrm{k}(x, y)/(\pi\mathrm{K}(y) + \eta)\mathrm{d}\mu(y)$

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 $X_0^\star$ 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 $\mathscr{G}_\alpha^\eta$.

Interacting Particle System

Approximation via particle systems

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

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

- $\{(B_t^k)_{t \geq 0} \;:\; k \in \mathbb{N}\}$ 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 $\mathscr{G}_\alpha^\eta$.

Numerical Scheme

Discretization and numerical implementation

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

$$\tilde{X}_{n+1}^{k,N} = \tilde{X}_n^{k,N} + \frac{-\gamma \bar{b}^\eta(\tilde{X}_n^{k,N}, \lambda_n^N)}{1+\gamma\left\|\bar{b}^\eta(\tilde{X}_n^{k,N}, \lambda_n^N)\right\|} + \gamma\alpha\nabla U_0(\tilde{X}_n^{k,N}) + \sqrt{2\alpha\gamma} Z_{n+1}^k \quad .$$

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

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

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

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 and influence of M

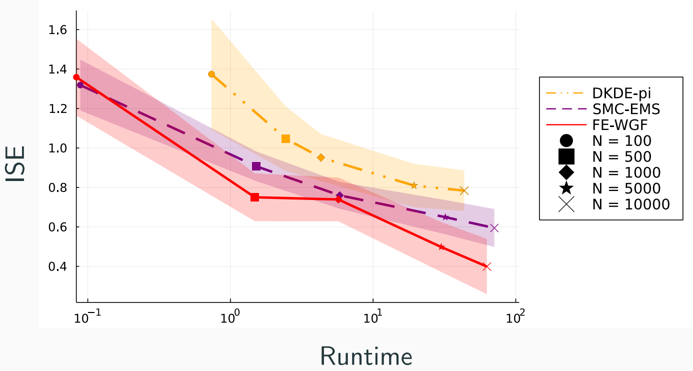
$$\mathbb{E}[\sup_{n\in\{0,\dots,n_T\}} \|X_n^\star - \tilde{X}_n^{k,N}\|] \leq C_T(N^{-1/2} + M^{-1/2} + \gamma).$$

Experiments

Density deconvolution

- convolution kernel $\mathrm{k}(x, y) = \mathrm{k}(y - x)$
- recover the density of X from observations with additive noise $Y = X + \epsilon$
- when k is Gaussian, estimators with optimal convergence rate exist: **deconvolution kernel density estimators (DKDE)**
- comparing with **SMC-EMS**, a sequential Monte Carlo implementation of Expectation Maximization Smoothing

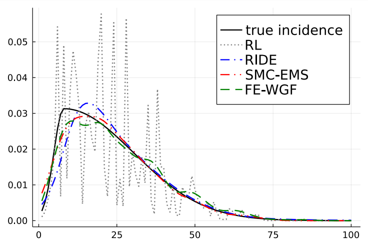
Density deconvolution

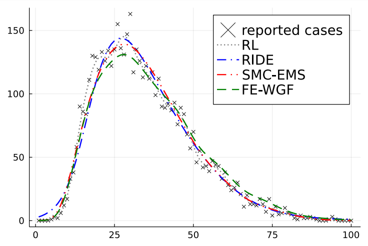


Epidemiology – effect of reference measure

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

- Comparing with **Richardson-Lucy algorithm** (i.e. EM for Poisson data) and **robust incidence deconvolution estimator (RIDE)**

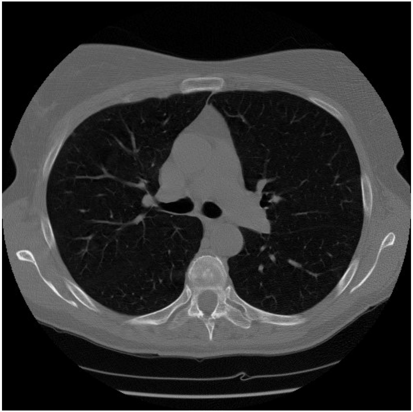




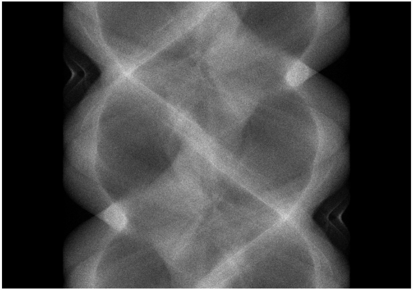
Epidemiology

	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

CT scans – data and model



CT scan



Data

$$\mu(\phi, \xi) = \int_{\mathbb{R}^2} (2\pi\sigma^2)^{-1} \exp[(x \cos\phi + y \sin\phi - \xi)^2/(2\sigma^2)]\pi(x, y)\mathrm{d}x\mathrm{d}y,$$

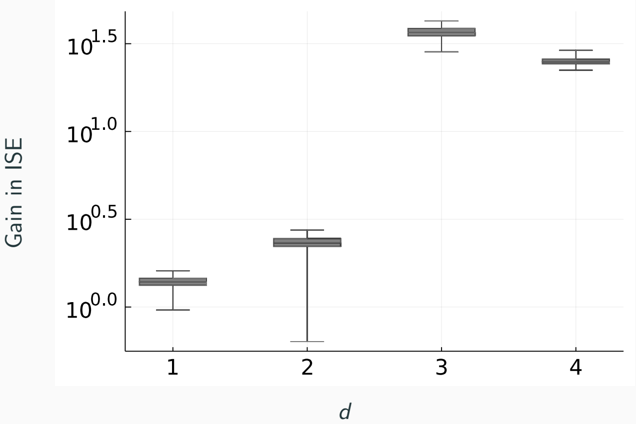
CT scans – results

CT scan FBP FE-WGF

Figure 1: Reconstruction of a lung CT scan via wGF and FBP. FBP provides reconstructions which preserve sharp edges but present speckle noise, while the reconstructions obtained with WGF are smooth but with blurry edges.

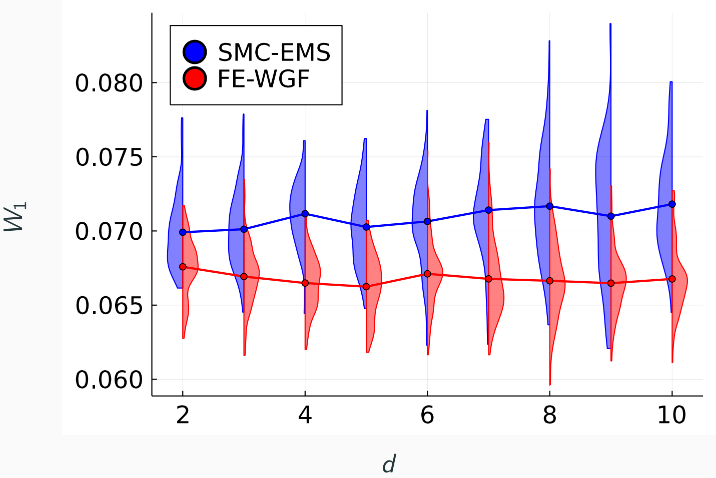
Scaling with dimension

- Multidimenesional deconvolution problem $(\mathrm{k}(x, y) = \mathrm{k}(y - x))$
- comparing with **one-step-late Expectation Maximization**



Scaling with dimension

- comparing with **SMC-EMS**



Conclusion

Conclusions

- a probabilistic formulation of the classical regularization problem for Fredholm integral equation
- a particle system to obtain approximate solutions:
  - without requiring a finite dimensional decomposition
  - can be naturally implemented when we only have samples from μ
  - performs better that its deterministic counterpart as d increases
  - minimizes a specific functional

Conclusions

- a probabilistic formulation of the classical regularization problem for Fredholm integral equation
- a particle system to obtain approximate solutions:
  - without requiring a finite dimensional decomposition
  - can be naturally implemented when we only have samples from μ
  - performs better that its deterministic counterpart as d increases
  - minimizes a specific functional

Thank you!

References

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.

Francesca R Crucinio, Arnaud Doucet, and Adam M Johansen. A Particle Method for Solving Fredholm Equations of the First Kind. *Journal of the American Statistical Association*, pages 1–11, 2021.

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.