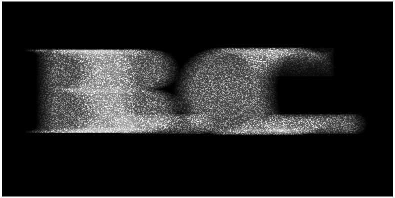
Solving Fredholm Integral Equations through Wasserstein Gradient Flows

Francesca R. Crucinio
with Adam M. Johansen & Arnaud Doucet

25 June 2020

Fredholm Integral Equations of the First Kind (FIE)

$$\mu(d\mathbf{u}) = \int K(\mathbf{x}, d\mathbf{u})\rho(d\mathbf{x})$$



constant speed horizontal motion

Fredholm Integral Equations of the First Kind (FIE)

- Applications: PDEs, indirect density estimation, signal reconstruction, causal inference, ...
- Ill posed inverse problems
- Solution methods often require discretisation/strong assumptions on ρ

FIE - Solution method

$$\mu(dy) = \int \rho(dx) K(x, dy),$$

Take ρ, μ probability measures and K a Markov transition kernel

$$\rho^\star := \operatorname{argmin}\ \mathrm{KL}(\mu, \rho K) - \alpha\, \mathrm{ent}(\rho)$$

for some $\alpha > 0$.

FIE - Solution method

$$\mu(dy) = \int \rho(dx) K(x, dy),$$

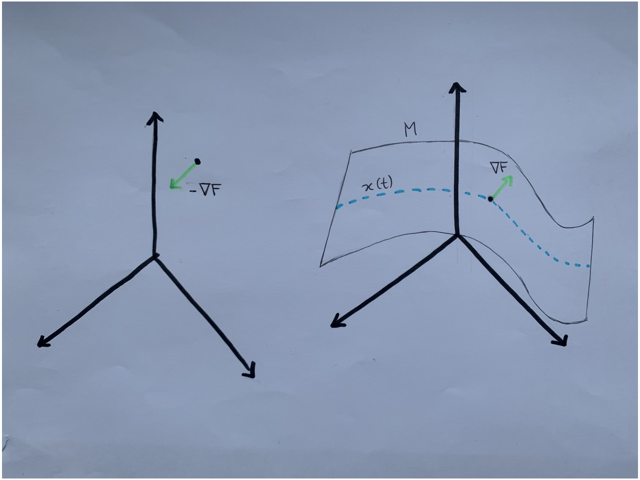
Take ρ, μ probability measures and K a Markov transition kernel

$$\rho^\star := \text{argmin } \; \mathrm{KL}(\mu, \rho K) - \alpha\, \mathrm{ent}(\rho)$$

for some $\alpha > 0$.
How?

$$\text{minimisation} \longrightarrow \text{PDE} \longrightarrow \text{SDE}$$

Gradient Flows



$$x'(t) = -\nabla F(x(t))$$

Wasserstein Gradient Flows - Space

- Extension of gradient flows to probability measures
- In particular, probability measures μ with finite second moment

$$\int \|x\|_2^2 \, d\mu(x) < \infty$$

and absolutely continuous w.r.t. Lebesgue $\mu \ll \mathcal{L}$: $\mathcal{P}_2^{ac}(\mathbb{R}^d)$

Wasserstein Gradient Flows - Distance

2-Wasserstein distance

$$W_2(\mu, \nu) := \left(\inf_{\pi \in \Pi(\mu,\nu)} \int \|x - y\|_2^2 \, d\pi(x, y) \right)^{1/2}$$

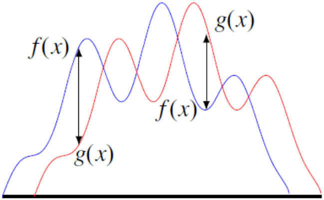
[1] Santambrogio (2017)

Wasserstein Gradient Flows - Distance

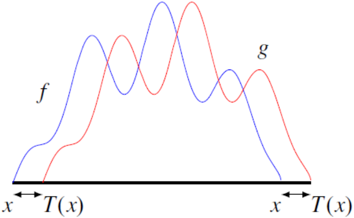
2-Wasserstein distance

$$W_2(\mu,\nu) := \left(\inf_{\pi\in\Pi(\mu,\nu)} \int \|x-y\|_2^2 \, d\pi(x,y)\right)^{1/2}$$

e.g. 1D densities[1]



$\mathbb{L}_2$



W_2

[1] Santambrogio (2017)

Wasserstein Gradient Flows - Optimal transport

In $\mathcal{P}_2^{ac}(\mathbb{R}^d)$, we have

$$W_2(\mu,\nu) = \left(\int \|x - t_\mu^\nu(x)\|_2^2 \; d\mu(x) \right)^{1/2} \qquad \text{with } (t_\mu^\nu)\#\mu = \nu,$$

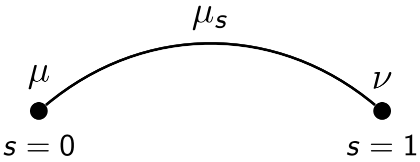
where $T_\#\mu$ denotes the push-forward measure $T_\#\mu(A) = \mu(T^{-1}(A))$.

Wasserstein Gradient Flows - Curves

Curves in $\left(\mathcal{P}_2^{ac}(\mathbb{R}^d), W_2\right)$ are geodesics

$$\mu_s = ((1-s)Id + s t_\mu^\nu)_\# \mu$$

with t_μ^ν the unique optimal transport map between μ and ν.



Wasserstein Gradient Flows - PDE

For $F : M \subset \mathbb{R}^d \to \mathbb{R}$ the gradient flow equation is

$$x'(t) = -\nabla F(x(t)).$$

For $F : \mathcal{P}_2^{ac}(\mathbb{R}^d) \to \mathbb{R}$ we have[2]

$$\partial_t \rho_t = \nabla \cdot \left(\rho_t \frac{\delta F}{\delta \rho_t} \right) \tag{1}$$

with

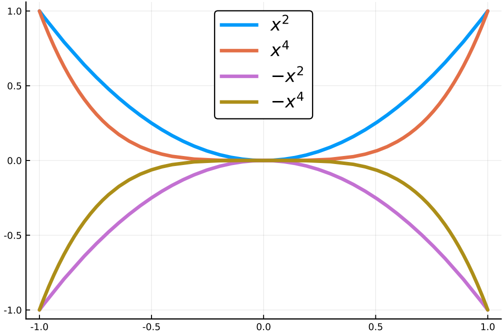
$$\frac{\delta F}{\delta \rho_t} := \lim_{\epsilon \to 0} \frac{F(\rho + \epsilon \chi) - F(\rho)}{\epsilon}.$$

[2] Jordan, Kinderlehrer, Otto (1998)

When is (1) well-behaved?

$F : \mathbb{R}^d \to \mathbb{R}$ is λ-convex if for $s \in [0, 1]$

$$F(sx + (1-s)y) \leq sF(x) + (1-s)F(y) - \frac{\lambda}{2}s(1-s)\|x-y\|_2^2$$



When is (1) well-behaved?

We need:

- F continuous, $F(\rho) < +\infty$ for some ρ
- F $\lambda-$geodesically convex w.r.t. W_2 with $\lambda \geq 0$
- F is coercive: exists $r > 0$ s.t.

$$\inf \left\{ F(\rho) : \rho \in \mathcal{P}_2^{ac}(\mathbb{R}^d),\ \int \|x\|_2^2 \, d\rho(x) \leq r \right\} > -\infty$$

Then...

The gradient flow PDE (1) has a unique solution for a given initial condition ρ_0.

For two initial conditions ρ_0^1, ρ_0^2 we have the estimate

$$W_2(\rho_t^1, \rho_t^2) \leq e^{-\lambda t} W_2(\rho_0^1, \rho_0^2).$$

Gradient Flow for FIE - Assumptions

$$\mu(y) = \int_{\mathbb{X}} \rho(y) K(x, y) \; dx \qquad \forall y \in \mathbb{Y},$$

1. $\mathbb{X}, \mathbb{Y}$ subsets of some Euclidean spaces with Borel σ-algebras;
2. $\mu, \rho \in \mathcal{P}_2^{ac}(\mathbb{X})$;
3. K bounded and bounded away from 0, Lipschitz continuous with Lipschitz continuous gradient and for all fixed $y \in \mathbb{Y}$, λ-concave in x for some $\lambda > 0$ (potentially dependent on y).

Gradient Flow for FIE

$$F(\rho) := \mathrm{KL}(\mu, \rho K) - \alpha\,\mathrm{ent}(\rho)$$

is continuous, coercive and λ-geodesically convex with $\lambda = 0$ and has first variation

$$\frac{\delta F}{\delta \rho}(x) = -\int \mu(dy)\,\frac{K(x,y)}{\rho K(y)} + \alpha\,(1 + \log \rho(x))\,.$$

the gradient flow exist and is unique for each initial condition ρ_0

Gradient Flow for FIE - PDE

$$
\begin{aligned}
\partial_t \rho_t &= \nabla \cdot \left(\rho_t \nabla \frac{\delta F}{\delta \rho_t} \right) \\
&= \nabla \cdot \left(\rho_t \left[- \int \mu(dy) \frac{\nabla K(x,y)}{\rho K(y)} + \alpha \nabla \log \rho(x) \right] \right) \\
&= -\nabla \cdot \left(\rho_t \int \mu(dy) \frac{\nabla K(x,y)}{\rho_t K(y)} \right) + \alpha \triangle \rho_t
\end{aligned}
$$

is a Fokker-Plank equation with corresponding SDE...

Gradient Flow for FIE - SDE

$$dX_t = \int \mu\left(dy\right) \frac{\nabla K(X_t, y)}{\rho_t K(y)} dt + \sqrt{2\alpha} dW_t, \quad X_0 \sim \rho_0$$

- McKean-Vlasov SDE
- a strong solution exists and is unique
- requires discretisation in time and space

SDE - Implementation

- discretisation in space: take N copies of the SDE and use the resulting empirical measure instead of ρ_t

$$dX_t^i = \int \mu(dy)\,\frac{\nabla K(X_t^i, y)}{\rho_t^N K(y)} dt + \sqrt{2\alpha}\, dW_t^i, \quad \rho_t^N = \frac{1}{N}\sum_{i=1}^N \delta_{X_t^i}$$

- discretisation in time: Euler scheme

$$X_{n+1}^i = X_n^i + \int \mu(dy)\,\frac{\nabla K(X_n^i, y)}{\rho_n^N K(y)} \Delta t + \sqrt{2\alpha}\,\Delta W_n^i,$$
$$\rho_n^N = \frac{1}{N}\sum_{i=1}^N \delta_{X_n^i}$$

Toy Example

$$\mu(dy) = \int_{\mathbb{X}} \rho(dx) K(x, dy),$$

with

$$\begin{aligned} \mu(y) &= \mathcal{N}(y; m, \sigma_\mu^2 := \sigma_K^2 + \sigma_\rho^2) \\ K(x, y) &= \mathcal{N}(y; x, \sigma_K^2) \\ \rho(x) &= \mathcal{N}(x; m, \sigma_\rho^2). \end{aligned}$$

The unique minimiser of E is

$$\rho_\alpha(x) = \mathcal{N}(x; m, \sigma_\alpha^2)$$

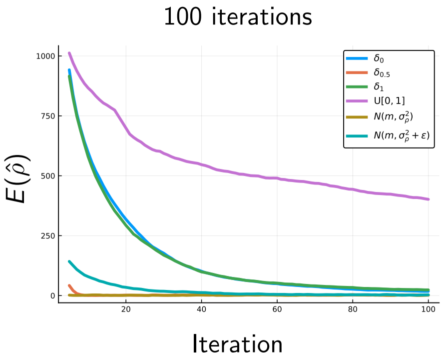
for $\alpha \in [0, 1)$.

A number of quantities need to be specified

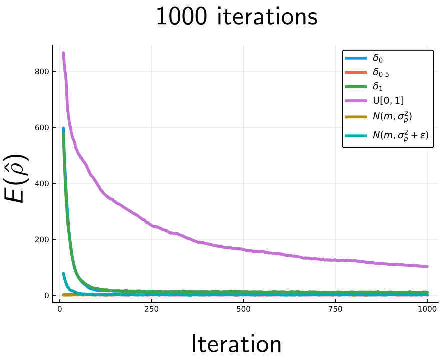
- regularisation: α
- gradient flow set up: ρ_0
- numerical approximation of SDE: N (500/1000), Δt (10^{-3} is enough)

And, how does it compare to other methods?

ρ_0

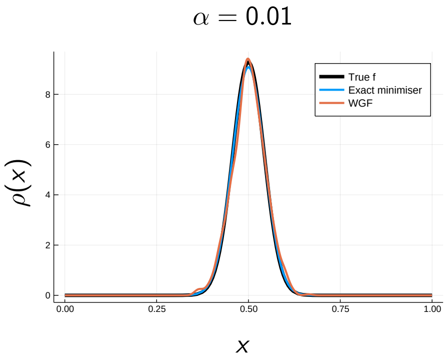


100 iterations
$E(\hat{\rho})$
Iteration
δ_0
$\delta_{0.5}$
δ_1
U[0, 1]
$N(m, \sigma_\rho^2)$
$N(m, \sigma_\rho^2 + \varepsilon)$

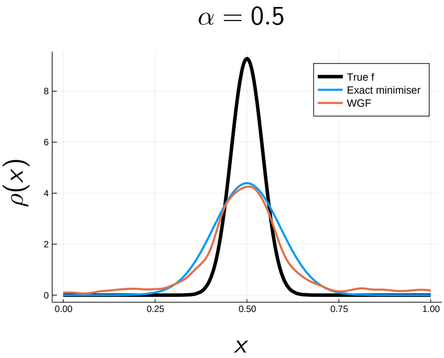


1000 iterations
$E(\hat{\rho})$
Iteration
δ_0
$\delta_{0.5}$
δ_1
U[0, 1]
$N(m, \sigma_\rho^2)$
$N(m, \sigma_\rho^2 + \varepsilon)$

α

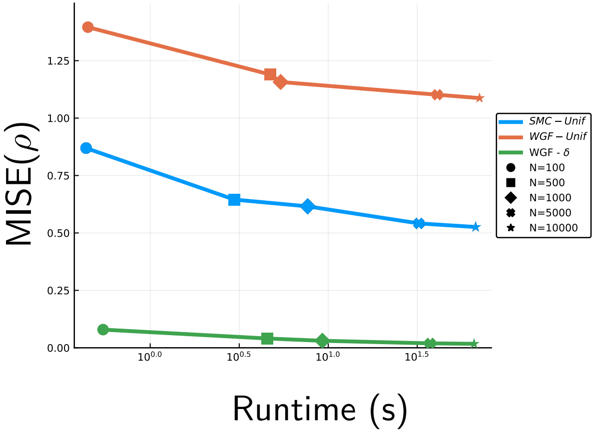


$\alpha = 0.01$
True f
Exact minimiser
WGF
$\rho(x)$
x
0.00
0.25
0.50
0.75
1.00
0
2
4
6
8



$\alpha = 0.5$
True f
Exact minimiser
WGF
$\rho(x)$
x
0.00
0.25
0.50
0.75
1.00
0
2
4
6
8

Comparison

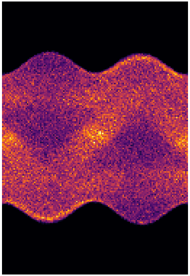


MISE(ρ)
1.25
1.00
0.75
0.50
0.25
0.00
$10^{0.0}$
$10^{0.5}$
$10^{1.0}$
$10^{1.5}$
Runtime (s)
$SMC - Unif$
$WGF - Unif$
WGF - δ
N=100
N=500
N=1000
N=5000
N=10000

PET

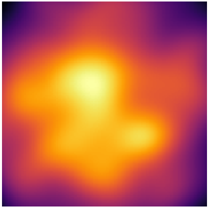
The reconstruction of a cross-section of the brain (a) from the data image provided by PET scanners (b) is described by a 2D Fredholm integral equation of the first kind.

(a) 128-pixels Shepp-Logan phantom

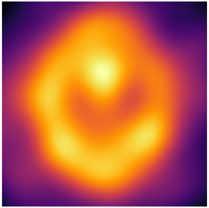


(b) Sinogram + noise

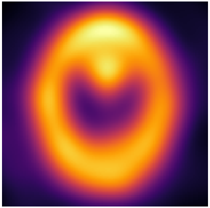
PET



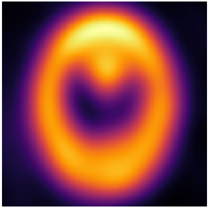
Iteration 1



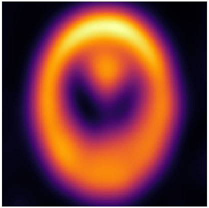
Iteration 10



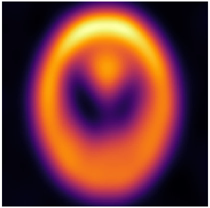
Iteration 50



Iteration 100



Iteration 300



Iteration 500