The Gradient Flow Interpretation of the Fokker-Planck Equation

Gianluca Covini

Presentation for the exam of *Introduction to Real Analysis II*

Table of contents

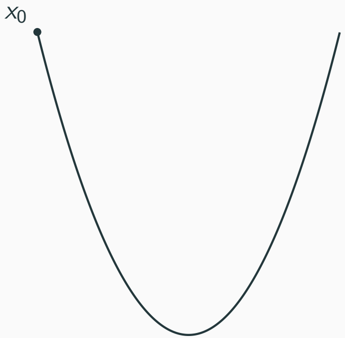
1. Gradient Flows in Euclidean Space
2. JKO Scheme for the Fokker-Planck Equation
3. Geometric Interlude
4. Convergence to the Fokker-Planck Solutions
5. Hints of Extension to the Metric Case
6. Conclusion

Gradient Flows in Euclidean Space

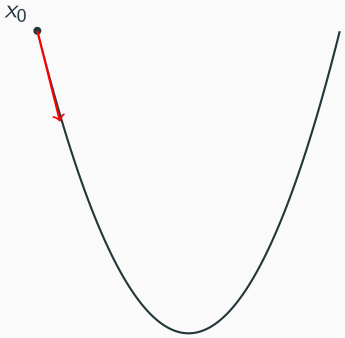
Gradient Flows in $\mathbb{R}^d$

Starting from a point x_0 you want to move towards the minimum of a function F.

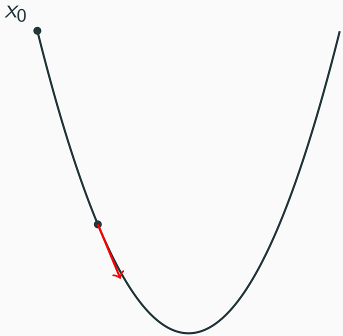
Idea: move towards the steepest descent of F, i.e. the negative gradient direction.



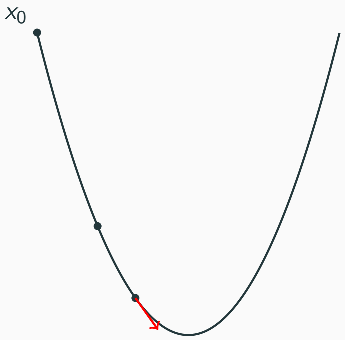
Gradient Flows in $\mathbb{R}^d$



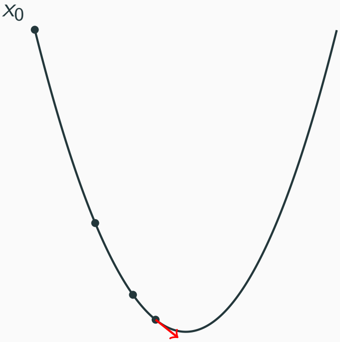
Gradient Flows in $\mathbb{R}^d$



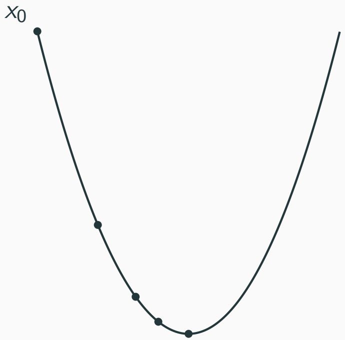
Gradient Flows in $\mathbb{R}^d$



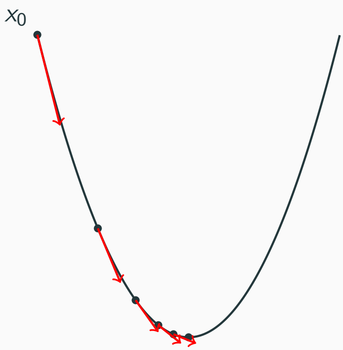
Gradient Flows in $\mathbb{R}^d$



Gradient Flows in $\mathbb{R}^d$

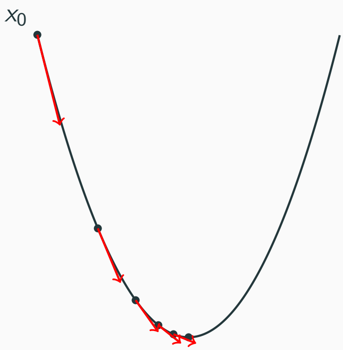


Gradient Flows in $\mathbb{R}^d$



You can imagine a continuous curve $x(t)$ which at each instant moves towards the steepest decrease of F.

Gradient Flows in $\mathbb{R}^d$



You can imagine a continuous curve $x(t)$ which at each instant moves towards the steepest decrease of F.

Gradient Flow in $\mathbb{R}^d$

Given $F : \Omega \subseteq \mathbb{R}^d \to \mathbb{R}$ *nice* enough, $x_0 \in \Omega$, a *gradient flow of F* is a solution of the Cauchy problem

$$\begin{cases} x'(t) & = -\nabla F(x(t)) \\ x(0) & = x_0 \end{cases}$$

In general, $x' \in -\partial F(x(t))$.

Discrete Scheme

For numerical reasons or to relax the regularity of F, we can define a discrete scheme.

Minimizing Movement Scheme

Fix the time step $\tau > 0$, for $k = 0, 1, \dots$

$$x_{k+1}^{\tau} = \arg\min_{x \in \Omega} \left(F(x) + \frac{\overbrace{|x - x_k^{\tau}|}^{\text{distance}}{}^2}{2\tau} \right)$$

Discrete Scheme

For numerical reasons or to relax the regularity of F, we can define a discrete scheme.

Minimizing Movement Scheme

Fix the time step $\tau > 0$, for $k = 0, 1, \ldots$

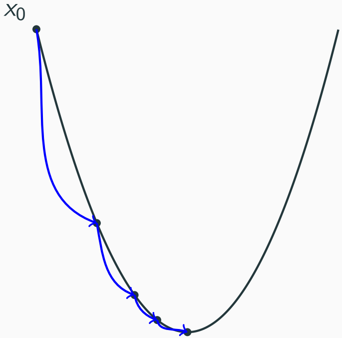
$$x_{k+1}^{\tau} = \underset{x \in \Omega}{\arg\min} \left(F(x) + \frac{\overbrace{|x - x_k^{\tau}|}^{\text{distance}}{}^2}{2\tau} \right)$$

Optimality Conditions

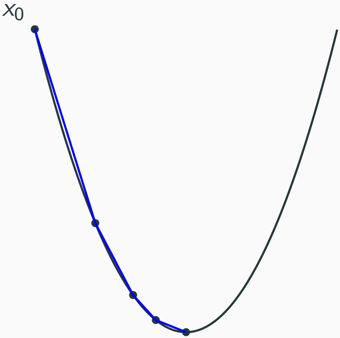
By the optimality condition $\nabla \left(F(x_{k+1}^{\tau}) + \frac{|x_{k+1}^{\tau} - x_k^{\tau}|^2}{2\tau} \right) = 0$, you recover *Implicit Euler Scheme*.

Convergence

We can interpolate the sequence from the discrete scheme $\{x_k^\tau\}_k$ in two ways.



Piecewise constant interpolation



Interpolation by segments

Under some hypotheses, for $\tau \downarrow 0$, the two interpolations converge to a solution of a gradient flow $x'(t) \in -\partial F(x(t))$.

Generalization

$$x_{k+1}^{\tau} = \underset{x \in \mathbb{R}^d}{\arg\min} \left(F(x) + \frac{\overbrace{|x - x_k^{\tau}|}^{\text{distance}}{}^2}{2\tau} \right)$$

Two ideas

- Replace the euclidean distance $|x-y|^2$ in the discrete scheme with the Wasserstein distance;

Generalization

$$x_{k+1}^{\tau} = \underset{x \in \mathbb{R}^d}{\arg\min} \left(F(x) + \frac{\overbrace{\left| x - x_k^{\tau} \right|}^{\text{distance}}{}^2}{2\tau} \right)$$

Two ideas

- Replace the euclidean distance $|x - y|^2$ in the discrete scheme with the Wasserstein distance;
- choose wisely the functional F to retrieve, at convergence, PDEs of particular interest in place of $x' = -\nabla F(x)$.

JKO Scheme for the Fokker-Planck Equation

Structure of the Presentation

Historical Insight

- Jordan, Kinderlehrer and Otto in 1998 in the paper *The Variational Formulation of the Fokker-Planck Equation* used for the first time the concept of gradient flows to characterize the Fokker-Planck PDE;
- comparison was mainly with the known discretization of the heat equation as *gradient flow* of the Dirichlet energy with respect to the L^2 distance in $\mathbb{R}^n$.

Structure of the Presentation

Historical Insight

- Jordan, Kinderlehrer and Otto in 1998 in the paper *The Variational Formulation of the Fokker-Planck Equation* used for the first time the concept of gradient flows to characterize the Fokker-Planck PDE;
- comparison was mainly with the known discretization of the heat equation as *gradient flow* of the Dirichlet energy with respect to the L^2 distance in $\mathbb{R}^n$.

We will present the result from Jordan, Kinderlehrer and Otto following the modern approach established after the book *Gradient Flows* of Ambrosio, Gigli and Savaré in 2005 and summarized by Santambrogio in his book *Optimal Transport for Applied Mathematicians* in 2015 and in 2017 in a survey titled *Euclidean, metric, and Wasserstein gradient flows: an overview*.

Why Fokker-Planck Equation?

Fokker-Planck Equation

$\rho = \rho_t \in \mathcal{P}_2(\Omega \subseteq \mathbb{R}^d)$,

$$\partial_t \rho_t - \Delta \rho_t - \nabla \cdot (\rho_t \nabla \Psi) = 0$$

Why Fokker-Planck Equation?

Fokker-Planck Equation

$\rho = \rho_t \in \mathcal{P}_2(\Omega \subseteq \mathbb{R}^d)$,

$$\partial_t \rho_t - \Delta \rho_t - \nabla \cdot (\rho_t \nabla \Psi) = 0$$

- It is a generalization of the heat equation $\partial_t \rho_t - \Delta \rho_t = 0$;

Why Fokker-Planck Equation?

Fokker-Planck Equation

$\rho = \rho_t \in \mathcal{P}_2(\Omega \subseteq \mathbb{R}^d)$,

$$\partial_t \rho_t - \Delta \rho_t - \nabla \cdot (\rho_t \nabla \Psi) = 0$$

- It is a generalization of the heat equation $\partial_t \rho_t - \Delta \rho_t = 0$;
- it represents the evolution of the probability density of a stochastic process $(X_t)_t$ under Langevin diffusion

$$dX_t = -\nabla \Psi(X_t) dt + \sqrt{2} dB_t$$

Why Fokker-Planck Equation?

Fokker-Planck Equation

$\rho = \rho_t \in \mathcal{P}_2(\Omega \subseteq \mathbb{R}^d)$,

$$\partial_t \rho_t - \Delta \rho_t - \nabla \cdot (\rho_t \nabla \Psi) = 0$$

- It is a generalization of the heat equation $\partial_t \rho_t - \Delta \rho_t = 0$;
- it represents the evolution of the probability density of a stochastic process $(X_t)_t$ under Langevin diffusion

$$dX_t = -\nabla \Psi(X_t) dt + \sqrt{2} dB_t$$

- its stationary distribution is the Gibbs distribution $\rho_{\mathrm{Gibbs}}(x) = e^{-\Psi(x)}/Z$;

Fokker Planck as a Gradient Flow

Last but not Least

- it is the gradient flow in the Wasserstein space of the Kullback-Leibler functional

$$\mathrm{KL}(\rho|\rho_{\mathsf{Gibbs}}) = \int_\Omega \rho(x) \log \frac{\rho(x)}{Z^{-1} e^{-\Psi(x)}} dx$$

Fokker Planck as a Gradient Flow

Last but not Least

- it is the gradient flow in the Wasserstein space of the Kullback-Leibler functional

$$\mathrm{KL}(\rho|\rho_{\mathsf{Gibbs}}) = \int_\Omega \rho(x) \log \frac{\rho(x)}{Z^{-1} e^{-\Psi(x)}} dx$$

JKO Scheme

Given a time step $\tau > 0$, we define the *JKO scheme*

$$\rho^\tau_{k+1} \in \underset{\rho \in \mathcal{P}_2(\Omega)}{\arg\min} \left(\mathrm{KL}(\rho|\rho_{\mathsf{Gibbs}}) + \frac{\mathcal{W}_2^2(\rho, \rho^\tau_k)}{2\tau} \right)$$

where, by slight abuse of notation, we write

$$\mathrm{KL}(\rho) := \begin{cases} \int_\Omega \rho \log \rho + \int_\Omega \Psi \, d\rho, & \rho << \mathcal{L}^d; \\ +\infty, & \text{otherwise.} \end{cases}$$

Results

$$\rho_{k+1}^{\tau} \in \underset{\rho \in \mathcal{P}_2(\Omega)}{\arg\min} \overbrace{\left(\mathrm{KL}(\rho) + \frac{\mathcal{W}_2^2(\rho, \rho_k^{\tau})}{2\tau} \right)}^{\widetilde{\mathrm{KL}}(\rho) :=}$$

Results

1. The JKO scheme is well posed (existence and unicity of a minimizer);
2. piecewise constant interpolation of the JKO scheme $\{\rho_k^{\tau}\}_k$ converges to the solution of the Fokker-Planck equation.

Assumptions

Ω is convex and compact; $\mathrm{KL}(\rho_0^{\tau} := \rho_0) < +\infty$.

Well-posedness of the JKO scheme

Existence of a minimizer

- Since Ω is compact, $\mathcal{P}_2(\Omega)$ is compact with respect to the $\mathcal{W}_2$ distance (and, equivalently, for the weak topology);
- $\mathrm{KL}(\rho) + \frac{\mathcal{W}_2^2(\rho,\rho_k^\tau)}{2\tau} =: \widetilde{\mathrm{KL}}(\rho)$ is lower semicontinuous, i.e.

$$\widetilde{\mathrm{KL}}(\rho) \leq \liminf_{\substack{n\to\infty \\ \rho_n\to\rho}} \widetilde{\mathrm{KL}}(\rho_n)$$

By direct method of calculus of variations, $\exists \bar{\rho} = \arg\min_{\rho\in\mathcal{P}_2(\Omega)} \widetilde{\mathrm{KL}}(\rho)$.

Unicity of the minimizer

From strict convexity of $\widetilde{\mathrm{KL}}$, it follows uniqueness of the minimizer.

Interpolations

Interpolations

Given $\{\rho_k^\tau\}_k$ solution of the JKO scheme, we can build the interpolation

- Piecewise constant interpolation:

$$\rho_t^\tau := \rho_{k+1}^\tau \quad \forall t \in (k\tau, (k+1)\tau]$$

We can associate a velocity vector

$$-v_t^\tau := \frac{T_{\rho_{k+1}^\tau \to \rho_k^\tau} - \mathrm{id}}{\tau} \quad \forall t \in (k\tau, (k+1)\tau]$$

and a momentum $E_t^\tau := \rho_t^\tau v_t^\tau$.

Interpolations

Given $\{\rho_k^\tau\}_k$ solution of the JKO scheme, we can build the interpolation

- Piecewise constant interpolation:

$$\rho_t^\tau := \rho_{k+1}^\tau \quad \forall t \in (k\tau, (k+1)\tau]$$

  We can associate a velocity vector

$$-v_t^\tau := \frac{T_{\rho_{k+1}^\tau \to \rho_k^\tau} - \mathrm{id}}{\tau} \quad \forall t \in (k\tau, (k+1)\tau]$$

  and a momentum $E_t^\tau := \rho_t^\tau v_t^\tau$.

- We would like to build another continuous interpolation, but what is the best way to do it?

Interpolations

Given $\{\rho_k^\tau\}_k$ solution of the JKO scheme, we can build the interpolation

- Piecewise constant interpolation:

$$\rho_t^\tau := \rho_{k+1}^\tau \quad \forall t \in (k\tau, (k+1)\tau]$$

  We can associate a velocity vector

$$-v_t^\tau := \frac{T_{\rho_{k+1}^\tau \to \rho_k^\tau} - \mathrm{id}}{\tau} \quad \forall t \in (k\tau, (k+1)\tau]$$

  and a momentum $E_t^\tau := \rho_t^\tau v_t^\tau$.

- We would like to build another continuous interpolation, but what is the best way to do it?

In $\mathbb{R}^d$ we could take the segment. But in the Wasserstein space?

Geometric Interlude

Curves in the Wasserstein Space

Curve

An *absolutely continuous curve* in in the Wasserstein space is a function $t \in [0,1] \to \rho_t \in \mathcal{P}_2(\Omega)$ s.t. $\mathcal{W}_2(\rho_s, \rho_t) \leq \int_s^t g(r)dr, \quad g \in L^1(0,1)$

Curves in the Wasserstein Space

Curve

An *absolutely continuous curve* in in the Wasserstein space is a function $t \in [0,1] \to \rho_t \in \mathcal{P}_2(\Omega)$ s.t. $\mathcal{W}_2(\rho_s, \rho_t) \leq \int_s^t g(r)dr, \quad g \in L^1(0,1)$

Theorem

For any $(\rho_t)_t$ absolutely continuous curve in the Wasserstein space, $\exists (v_t)_t \subseteq L^2_{\rho_t}(\mathbb{R}^d)$ velocity field such that, for a.e. t,

(i) $\partial_t \rho_t + \nabla \cdot (v_t \rho_t) = 0$ (continuity equation);

(ii) $\|v_t\|_{L^2_{\rho_t}(\mathbb{R}^d)} = |\rho'|(t) := \lim_{h \downarrow 0} \frac{\mathcal{W}_2(\rho_{t+h}, \rho_t)}{h}$.

Curves in the Wasserstein Space

Curve

An *absolutely continuous curve* in in the Wasserstein space is a function $t \in [0,1] \to \rho_t \in \mathcal{P}_2(\Omega)$ s.t. $\mathcal{W}_2(\rho_s, \rho_t) \leq \int_s^t g(r)dr, \quad g \in L^1(0,1)$

Theorem

For any $(\rho_t)_t$ absolutely continuous curve in the Wasserstein space, $\exists (v_t)_t \subseteq L^2_{\rho_t}(\mathbb{R}^d)$ velocity field such that, for a.e. t,

(i) $\partial_t \rho_t + \nabla \cdot (v_t \rho_t) = 0$ (continuity equation);

(ii) $\|v_t\|_{L^2_{\rho_t}(\mathbb{R}^d)} = |\rho'|(t) := \lim_{h \downarrow 0} \frac{\mathcal{W}_2(\rho_{t+h}, \rho_t)}{h}$.

Definitions

Length of $(\rho_t)_t := \int_0^1 |\rho'|(t)dt$;

Curves in the Wasserstein Space

Curve

An *absolutely continuous curve* in in the Wasserstein space is a function $t \in [0,1] \to \rho_t \in \mathcal{P}_2(\Omega)$ s.t. $\mathcal{W}_2(\rho_s, \rho_t) \leq \int_s^t g(r)dr, \quad g \in L^1(0,1)$

Theorem

For any $(\rho_t)_t$ absolutely continuous curve in the Wasserstein space, $\exists (v_t)_t \subseteq L^2_{\rho_t}(\mathbb{R}^d)$ velocity field such that, for a.e. t,

(i) $\partial_t \rho_t + \nabla \cdot (v_t \rho_t) = 0$ (continuity equation);

(ii) $\|v_t\|_{L^2_{\rho_t}(\mathbb{R}^d)} = |\rho'|(t) := \lim_{h \downarrow 0} \frac{\mathcal{W}_2(\rho_{t+h}, \rho_t)}{h}$.

Definitions

Length of $(\rho_t)_t := \int_0^1 |\rho'|(t)dt$;

We call *constant-speed geodesic* between two probabilities measures μ and ν the curve $(\rho_t)_t$ of minimum length between μ and ν with constant speed $|\rho'|$.

Two Nice Geometric Results

Benamou-Brenier Formulation of OT

$$\mathcal{W}_2^2(\mu,\nu) = \min\left\{\int_0^1 \|v_t\|_{L^2}^2 dt : \partial_t \rho_t + \nabla\cdot(\rho_t v_t) = 0,\ \rho_0 = \mu, \rho_1 = \nu\right\}$$

Two Nice Geometric Results

Benamou-Brenier Formulation of OT

$$\mathcal{W}_2^2(\mu,\nu) = \min \left\{ \int_0^1 \|v_t\|_{L^2}^2 dt : \partial_t \rho_t + \nabla \cdot (\rho_t v_t) = 0,\ \rho_0 = \mu, \rho_1 = \nu \right\}$$

Geodesic

The constant-speed geodesic between ρ_k^τ and ρ_{k+1}^τ is

$$((1-t)\mathrm{id} + tT_{\rho_k^\tau \to \rho_{k+1}^\tau})_\sharp \rho_k^\tau$$

Note: the constant-speed geodesic exists unique between $\rho_k^\tau << \mathcal{L}^d$.

Convergence to the Fokker-Planck Solutions

Convergence

Interpolation along geodesics

We define the interpolation along geodesics

$$\tilde{\rho}_t^\tau := ((1-t)\mathrm{id} + tT_{\rho_k^\tau \to \rho_{k+1}^\tau})_\sharp \rho_k^\tau \quad t \in (k\tau, (k+1)\tau]$$

We call $(\tilde{v}_t^\tau)_t$ the corresponding velocity field and $\tilde{E}^\tau = \tilde{\rho}^\tau \tilde{v}^\tau$ the momentum.

Convergence

Interpolation along geodesics

We define the interpolation along geodesics

$$\tilde{\rho}_t^\tau := ((1-t)\mathrm{id} + tT_{\rho_k^\tau \to \rho_{k+1}^\tau})_\sharp \rho_k^\tau \quad t \in (k\tau, (k+1)\tau]$$

We call $(\tilde{v}_t^\tau)_t$ the corresponding velocity field and $\tilde{E}^\tau = \tilde{\rho}^\tau \tilde{v}^\tau$ the momentum.

Bound

We can show that

$$\mathcal{W}_2(\tilde{\rho}_t^\tau, \tilde{\rho}_s^\tau) \leq C(t-s)^{1/2}$$

Convergence

Interpolation along geodesics

We define the interpolation along geodesics

$$\tilde{\rho}_t^\tau := ((1-t)\mathrm{id} + tT_{\rho_k^\tau \to \rho_{k+1}^\tau})_\sharp \rho_k^\tau \quad t \in (k\tau, (k+1)\tau]$$

We call $(\tilde{v}_t^\tau)_t$ the corresponding velocity field and $\tilde{E}^\tau = \tilde{\rho}^\tau \tilde{v}^\tau$ the momentum.

Bound

We can show that

$$\mathcal{W}_2(\tilde{\rho}_t^\tau, \tilde{\rho}_s^\tau) \leq C(t-s)^{1/2}$$

Corollaries

- By Ascoli-Arzelà theorem, $\exists \rho$ s.t. $\tilde{\rho}^\tau \rightharpoonup \rho$;
- $\mathcal{W}_2(\tilde{\rho}_t^\tau, \rho_t^\tau) \to 0$ for $\tau \downarrow 0$.

One can also show that $w - \lim_{\tau \downarrow 0} E^\tau = E = w - \lim_{\tau \downarrow 0} \tilde{E}^\tau$ weakly.

Solution of the Fokker-Planck Equation

We can define (ρ, E) as same weak limit of both the interpolations (ρ^τ, E^τ) and $(\tilde{\rho}^\tau, \tilde{E}^\tau)$.

Theorem

(ρ, E) solve in the distributional sense

$$\begin{cases} \partial_t \rho_t + \nabla \cdot E_t = 0 \\ E_t = -\nabla \rho_t - \rho_t \nabla \Psi \\ \rho(0) = \rho_0 \end{cases} \quad \rightsquigarrow \quad \begin{cases} \partial_t \rho_t - \Delta \rho_t - \nabla \cdot (\rho_t \nabla \Psi) = 0 \\ \rho(0) = \rho_0 \end{cases}$$

Proof

$$\begin{cases} \partial_t \rho_t + \nabla \cdot E_t = 0 \\ E_t = -\nabla \rho_t - \rho_t \nabla \Psi \\ \rho(0) = \rho_0 \end{cases} \rightsquigarrow \begin{cases} \partial_t \rho_t - \Delta \rho_t - \nabla \cdot (\rho_t \nabla \Psi) = 0 \\ \rho(0) = \rho_0 \end{cases}$$

First and third equation

- We know $(\tilde{\rho}^\tau, \tilde{E}^\tau)$ solves the continuity equation $\partial_t \tilde{\rho}_t^\tau + \nabla \cdot \tilde{E}_t^\tau = 0$ with $\tilde{\rho}^\tau(0) = \rho_0$;
- by weak convergence and continuity, $\partial_t \rho_t + \nabla \cdot E_t = 0$ and $\rho(0) = \rho_0$.

Proof

$$\begin{cases} \partial_t \rho_t + \nabla \cdot E_t = 0 \\ E_t = -\nabla \rho_t - \rho_t \nabla \Psi \\ \rho(0) = \rho_0 \end{cases} \rightsquigarrow \begin{cases} \partial_t \rho_t - \Delta \rho_t - \nabla \cdot (\rho_t \nabla \Psi) = 0 \\ \rho(0) = \rho_0 \end{cases}$$

Second equation

- If we show that

$$E_t^\tau = \rho_t^\tau v_t^\tau \overset{(*)}{=} -\rho_t^\tau (\nabla(\log \rho_t^\tau + \Psi)) = -\nabla \rho_t^\tau - \rho_t^\tau \nabla \Psi$$

- By weak convergence, it follows $E_t = -\nabla \rho_t - \rho_t \nabla \Psi$.

Proof

$$\begin{cases} \partial_t \rho_t + \nabla \cdot E_t = 0 \\ E_t = -\nabla \rho_t - \rho_t \nabla \Psi \\ \rho(0) = \rho_0 \end{cases} \quad \rightsquigarrow \quad \begin{cases} \partial_t \rho_t - \Delta \rho_t - \nabla \cdot (\rho_t \nabla \Psi) = 0 \\ \rho(0) = \rho_0 \end{cases}$$

Second equation

- If we show that

$$E_t^\tau = \rho_t^\tau v_t^\tau \stackrel{(*)}{=} -\rho_t^\tau (\nabla(\log \rho_t^\tau + \Psi)) = -\nabla \rho_t^\tau - \rho_t^\tau \nabla \Psi$$

- By weak convergence, it follows $E_t = -\nabla \rho_t - \rho_t \nabla \Psi$.

We just need to show $-v^\tau := \frac{T - \mathrm{id}}{\tau} = \nabla(\log \rho^\tau + \Psi)$.

Last Step

But proving that $\frac{T-\mathrm{id}}{\tau} = \nabla(\log \rho^\tau + \Psi)$ is not trivial, it is actually the cornerstone of the argument by Jordan, Kinderlehrer and Otto, because it is the optimality condition for the JKO scheme.

Last Step

But proving that $\frac{T-\mathrm{id}}{\tau} = \nabla(\log \rho^\tau + \Psi)$ is not trivial, it is actually the cornerstone of the argument by Jordan, Kinderlehrer and Otto, because it is the optimality condition for the JKO scheme.

First Variation

Let $F : \mathcal{P}_2(\Omega) \to \mathbb{R} \cup \{+\infty\}$, we define as *first variation* $\frac{\delta F}{\delta \rho}(\rho)$ if it exists any measurable function such that $\forall \rho' \in L_c^\infty(\Omega) \cap \mathcal{P}_2(\Omega), \varepsilon \in [0,1]$,

$$\frac{d}{d\varepsilon} F(\rho + \varepsilon(\rho' - \rho))_{|\varepsilon=0} = \int \frac{\delta F}{\delta \rho}(\rho) d(\rho' - \rho)$$

The first variation is defined for any ρ s.t. $F(\rho + \varepsilon(\rho' - \rho)) < +\infty$.

Last Step

Claim: $\frac{T - \mathrm{id}}{\tau} = \nabla(\log \rho^\tau + \Psi)$

First Variation

$$\frac{d}{d\varepsilon} F(\rho + \varepsilon(\rho' - \rho))_{|\varepsilon=0} = \int \frac{\delta F}{\delta \rho}(\rho) d(\rho' - \rho)$$

Optimality conditions

The optimality condition of the JKO scheme can be seen as a condition on the first variation of the functional $\widetilde{\mathrm{KL}}$ we are minimizing

$$\nabla \left(\frac{\delta \widetilde{\mathrm{KL}}}{\delta \rho} (\rho^\tau_{k+1}) \right) = 0$$

Last Step

Claim: $\frac{T - \mathrm{id}}{\tau} = \nabla(\log \rho^\tau + \Psi)$

Tool: $\nabla\left(\frac{\delta \widetilde{\mathrm{KL}}}{\delta \rho}(\rho^\tau_{k+1})\right) = 0$

First Variations

- $\frac{\delta \mathrm{KL}}{\delta \rho}(\rho^\tau_{k+1}) = 1 + \log(\rho^\tau_{k+1}) + \Psi$;
- $\frac{\delta(\mathcal{W}_2^2(\cdot, \rho^\tau_k)/2\tau)}{\delta \rho}(\rho^\tau_{k+1}) = \frac{\varphi}{\tau}$ where φ is the Kantorovich potential for the associated cost functional for the transport from ρ^τ_{k+1} to ρ^τ_k.

Last Step

Claim: $\frac{T-\text{id}}{\tau} = \nabla(\log \rho^\tau + \Psi)$

Tool: $\nabla\left(\frac{\delta \widetilde{\mathrm{KL}}}{\delta \rho}(\rho^\tau_{k+1})\right) = 0$

First Variations

- $\frac{\delta \mathrm{KL}}{\delta \rho}(\rho^\tau_{k+1}) = 1 + \log(\rho^\tau_{k+1}) + \Psi$;
- $\frac{\delta(\mathcal{W}_2^2(\cdot,\rho^\tau_k)/2\tau)}{\delta \rho}(\rho^\tau_{k+1}) = \frac{\varphi}{\tau}$ where φ is the Kantorovich potential for the associated cost functional for the transport from ρ^τ_{k+1} to ρ^τ_k.

Optimality Condition

By writing the optimality condition and using Brenier's theorem, we obtain the equality we needed:

$$\frac{\text{id} - T_{\rho^\tau_{k+1} \to \rho^\tau_k}}{\tau} = \frac{\nabla \varphi}{\tau} = -\nabla(\log(\rho^\tau_{k+1}) + \Psi)$$

□

Hints of Extension to the Metric Case

Generalization

General JKO scheme

The JKO scheme gives a intuitive way to its generalization to metric spaces

$$\rho_{k+1}^{\tau} \in \operatorname*{arg\,min}_{\rho \in \mathcal{P}_2(\Omega)} \left(\mathrm{KL}(\rho) + \frac{\mathcal{W}_2^2(\rho, \rho_k^{\tau})}{2\tau} \right) \quad \rightsquigarrow$$

$$\rightsquigarrow \quad x_{k+1}^{\tau} \in \operatorname*{arg\,min}_{x \in X} \left(F(x) + \frac{d^2(x, x_k^{\tau})}{2\tau} \right)$$

under F l.s.c. and suitable compactness.

Generalization

General JKO scheme

The JKO scheme gives a intuitive way to its generalization to metric spaces

$$\rho_{k+1}^{\tau} \in \operatorname*{arg\,min}_{\rho \in \mathcal{P}_2(\Omega)} \left(\mathrm{KL}(\rho) + \frac{\mathcal{W}_2^2(\rho, \rho_k^{\tau})}{2\tau} \right) \quad \rightsquigarrow$$

$$\rightsquigarrow \quad x_{k+1}^{\tau} \in \operatorname*{arg\,min}_{x \in X} \left(F(x) + \frac{d^2(x, x_k^{\tau})}{2\tau} \right)$$

under F l.s.c. and suitable compactness.

Also the piecewise constant interpolation makes sense

$$x_t^{\tau} := x_{k+1}^{\tau} \quad \forall t \in (k\tau, (k+1)\tau]$$

Generalization

General JKO scheme

The JKO scheme gives a intuitive way to its generalization to metric spaces

$$\rho_{k+1}^\tau \in \operatorname*{arg\,min}_{\rho \in \mathcal{P}_2(\Omega)} \left(\mathrm{KL}(\rho) + \frac{\mathcal{W}_2^2(\rho, \rho_k^\tau)}{2\tau} \right) \quad \rightsquigarrow$$

$$\rightsquigarrow \quad x_{k+1}^\tau \in \operatorname*{arg\,min}_{x \in X} \left(F(x) + \frac{d^2(x, x_k^\tau)}{2\tau} \right)$$

under F l.s.c. and suitable compactness.

Also the piecewise constant interpolation makes sense

$$x_t^\tau := x_{k+1}^\tau \quad \forall t \in (k\tau, (k+1)\tau]$$

But what does it mean convergence to a gradient flow in a metric space?

Geodesic Spaces

We can consider a slightly more regular space.

Geodesic Spaces

(X, d) is a geodesic space if there always exists a geodesic between two arbitrary points and the length of the geodesic is equal to the distance between the two points.

Geodesic Spaces

We can consider a slightly more regular space.

Geodesic Spaces

(X, d) is a geodesic space if there always exists a geodesic between two arbitrary points and the length of the geodesic is equal to the distance between the two points.

Geodesic Interpolation

We can define the geodesic interpolation $\tilde{x}^{\tau}(t)$ of the solutions of the generalized JKO scheme $\{x_k^{\tau}\}_k$.

Geodesic Spaces

We can consider a slightly more regular space.

Geodesic Spaces

(X, d) is a geodesic space if there always exists a geodesic between two arbitrary points and the length of the geodesic is equal to the distance between the two points.

Geodesic Interpolation

We can define the geodesic interpolation $\tilde{x}^\tau(t)$ of the solutions of the generalized JKO scheme $\{x_k^\tau\}_k$.

Geodesic λ-convexity

F is λ-geodesically convex if, given $x(0), x(1) \in X$ and $x(t)$ the constant-speed geodesic between them,

$$F(x(t)) \leq (1-t)F(x(0)) + tF(x(1)) - \lambda \frac{t(1-t)}{2} d^2(x(0), x(1))$$

Gradient Flows in Geodesic Spaces: EDE

There are several notions of generalization of gradient flows in metric spaces, we will define only one.

Gradient Flows in Geodesic Spaces: EDE

There are several notions of generalization of gradient flows in metric spaces, we will define only one.

In the Euclidean setting, assuming enough smoothness,

$$F(x(s)) - F(x(t)) = \int_s^t -\nabla F(x(r)) \cdot x'(r) dr = \int_s^t |x'(r)||\nabla F(x(r))| dr =$$
$$= \int_s^t \left(\frac{1}{2}|x'(r)|^2 + \frac{1}{2}|\nabla F(x(r))|^2 \right) dr$$

The **equality** holds if and only if $x'(r) = -\nabla F(x(r))$.

Gradient Flows in Geodesic Spaces: EDE

In the Euclidean setting, assuming enough smoothness,

$$F(x(s)) - F(x(t)) = \int_s^t -\nabla F(x(r)) \cdot x'(r) dr = \int_s^t |x'(r)||\nabla F(x(r))| dr =$$
$$= \int_s^t \left(\frac{1}{2}|x'(r)|^2 + \frac{1}{2}|\nabla F(x(r))|^2 \right) dr$$

The **equality** holds if and only if $x'(r) = -\nabla F(x(r))$.

EDE

F λ-geodesically convex, satisfies the **EDE** if

$$F(x(s)) - F(x(t)) = \int_s^t \left(\frac{1}{2}|x'|^2(r) + \frac{1}{2}|\nabla^- F|^2(x(r)) \right) dr,$$

where

- $|x'|(r) := \lim_{h\downarrow 0} \frac{d(x(r+h),x(r))}{h}$ (metric derivative);
- $|\nabla^- F|(x) := \limsup_{\substack{y\to x \\ y\neq x}} \frac{[F(x)-F(y)]_+}{d(x,y)}$ (descending slope).

Convergence of the general JKO scheme

$$x_{k+1}^{\tau} \in \operatorname*{arg\,min}_{x \in X} \left(F(x) + \frac{d^2(x, x_k^{\tau})}{2\tau} \right)$$

A Condition for Convergence

If F is λ-geodesically convex, lower semicontinuous and with compact sublevels, we have that the generalized JKO scheme under geodesic interpolation converges for $\tau \downarrow 0$ to a curve $(x(t))_t$ satisfying the EDE definition of gradient flow

$$F(x(s)) - F(x(t)) = \int_s^t \left(\frac{1}{2}|x'|^2(r) + \frac{1}{2}|\nabla^- F|^2(x(r)) \right) dr$$

Conclusion

Conclusion

Takeaways

- The JKO scheme gives a general scheme to deal with gradient flows in discrete setting,

$$x_{k+1}^{\tau} \in \arg\min_{x \in X} \left(F(x) + \frac{d^2(x, x_k^{\tau})}{2\tau} \right)$$

Conclusion

Takeaways

- The JKO scheme gives a general scheme to deal with gradient flows in discrete setting,

$$x_{k+1}^{\tau} \in \underset{x \in X}{\arg\min} \left(F(x) + \frac{d^2(x, x_k^{\tau})}{2\tau} \right)$$

- By choosing wisely the functional F, an interpolation of the JKO scheme solutions converge to a curve which models a gradient flow.

Conclusion

Takeaways

- The JKO scheme gives a general scheme to deal with gradient flows in discrete setting,

$$x_{k+1}^{\tau} \in \underset{x \in X}{\arg\min} \left(F(x) + \frac{d^2(x, x_k^{\tau})}{2\tau} \right)$$

- By choosing wisely the functional F, an interpolation of the JKO scheme solutions converge to a curve which models a gradient flow.
- In particular, if $F(\cdot) = KL(\cdot|\rho_{Gibbs})$ and the distance is the 2-Wasserstein distance, the gradient flow we obtain at convergence is the Fokker-Planck equation.

Conclusion

Takeaways

- The JKO scheme gives a general scheme to deal with gradient flows in discrete setting,

$$x_{k+1}^{\tau} \in \arg\min_{x \in X} \left(F(x) + \frac{d^2(x, x_k^{\tau})}{2\tau} \right)$$

- By choosing wisely the functional F, an interpolation of the JKO scheme solutions converge to a curve which models a gradient flow.
- In particular, if $F(\cdot) = KL(\cdot|\rho_{Gibbs})$ and the distance is the 2-Wasserstein distance, the gradient flow we obtain at convergence is the Fokker-Planck equation.

Thank you for your attention!