Exponential convergence of Sinkhorn algorithm for quadratic entropic optimal transport

Giovanni Conforti

Workshop Mesa, March 21st 2023

École Polytechnique

Introduction

Schrödinger's thought experiment

"Imaginez que vous observez un **système de particules en diffusion**, *qui soient en* **équilibre thermodynamique.** *Admettons qu'à un instant donné* 0 *vous les ayez trouvées en* **répartition à peu près uniforme** *et qu'à* T *vous ayez trouvé un* **écart spontané et considérable par rapport à cette uniformité**. *On vous demande de quelle manière cet écart s'est produit. Quelle en est la* **manière la plus probable ?"**

- Erwin Schrödinger. "La théorie relativiste de l'électron et l' interprétation de la mécanique quantique". In: *Ann. Inst Henri Poincaré* 2 (1932), pp. 269–310

Schrödinger problem

Schrödinger problem(Entropic Optimal Transport)

$$\inf_{\pi \in \Pi(\mu,\nu)} \int \log \frac{\mathrm{d}\pi}{\mathrm{dLeb}}(x,y)\pi(\mathrm{d}x\mathrm{d}y) + \frac{1}{2T}\int |x-y|^2 \pi(\mathrm{d}x\mathrm{d}y)$$

with

$$\Pi(\mu,\nu) = \{\pi : \pi(A \times \mathbb{R}^d) = \mu(A), \pi(\mathbb{R}^d \times A) = \nu(A)\ \forall A\}.$$

The optimal coupling $\pi^\star$ is the **Schrödinger bridge**.

Monge-Kantorovich problem (Optimal transport)

$$\inf_{\pi \in \Pi(\mu,\nu)} \int_{\mathbb{R}^d \times \mathbb{R}^d} |x-y|^2 \pi(\mathrm{d}x\mathrm{d}y)$$

Schrödinger problem

Schrödinger problem(Entropic Optimal Transport)

$$\inf_{\pi\in\Pi(\mu,\nu)}\int \log\frac{\mathrm{d}\pi}{\mathrm{dLeb}}(x,y)\pi(\mathrm{d}x\mathrm{d}y)+\frac{1}{2T}\int |x-y|^2\pi(\mathrm{d}x\mathrm{d}y)$$

with

$$\Pi(\mu,\nu)=\{\pi:\pi(A\times\mathbb{R}^d)=\mu(A),\pi(\mathbb{R}^d\times A)=\nu(A)\ \forall A\}.$$

The optimal coupling $\pi^\star$ is the **Schrödinger bridge**.

Monge-Kantorovich problem (Optimal transport)

$$\inf_{\pi\in\Pi(\mu,\nu)}\int_{\mathbb{R}^d\times\mathbb{R}^d}|x-y|^2\pi(\mathrm{d}x\mathrm{d}y)$$

Schrödinger potentials

Theorem

Let μ, ν have finite entropy. Then there exist $\varphi^\star, \psi^\star : \mathbb{R}^d \longrightarrow \mathbb{R}$ such that

$$\frac{\mathrm{d}\pi^\star}{\mathrm{d}R_{0T}}(x, y) = \exp(-\varphi^\star(x) - \psi^\star(y))$$

$\varphi^\star, \psi^\star$ are the **Schrödinger potentials**.

Theorem (Nutz, Wiesel '21)

Let φ^{Br} be a **Kantorovich potential** and μ, ν have finite entropy. Then

$$T\varphi^\star \xrightarrow{L^1(\mu)} \varphi^{\mathrm{Br}}$$

Theorem (Chiarini,C., Greco,Tamanini '22)

Assume that μ, ν have finite Fisher information.Then

$$T\nabla\varphi^\star \xrightarrow{L^2(\mu)} \nabla\varphi^{\mathrm{Br}},$$

where $\nabla\varphi^{\mathrm{Br}}$ is the Brenier map.

Schrödinger potentials

Theorem

Let μ, ν have finite entropy. Then there exist $\varphi^\star, \psi^\star : \mathbb{R}^d \longrightarrow \mathbb{R}$ such that

$$\frac{\mathrm{d}\pi^\star}{\mathrm{d}R_{0T}}(x,y) = \exp(-\varphi^\star(x) - \psi^\star(y))$$

$\varphi^\star, \psi^\star$ are the **Schrödinger potentials**.

Theorem (Nutz, Wiesel '21)

Let φ^{Br} be a **Kantorovich potential** and μ, ν have finite entropy. Then

$$T\varphi^\star \xrightarrow{L^1(\mu)} \varphi^{\mathrm{Br}}$$

Theorem (Chiarini,C., Greco,Tamanini '22)

Assume that μ, ν have finite Fisher information.Then

$$T\nabla\varphi^\star \xrightarrow{L^2(\mu)} \nabla\varphi^{\mathrm{Br}},$$

where $\nabla\varphi^{\mathrm{Br}}$ is the Brenier map.

Schrödinger potentials

Theorem

Let μ, ν have finite entropy. Then there exist $\varphi^\star, \psi^\star : \mathbb{R}^d \longrightarrow \mathbb{R}$ such that

$$\frac{\mathrm{d}\pi^\star}{\mathrm{d}R_{0T}}(x,y) = \exp(-\varphi^\star(x) - \psi^\star(y))$$

$\varphi^\star, \psi^\star$ are the **Schrödinger potentials**.

Theorem (Nutz, Wiesel '21)

Let φ^{Br} be a **Kantorovich potential** and μ, ν have finite entropy. Then

$$T\varphi^\star \xrightarrow{L^1(\mu)} \varphi^{\mathrm{Br}}$$

Theorem (Chiarini,C., Greco,Tamanini '22)

Assume that μ, ν have finite Fisher information.Then

$$T\nabla\varphi^\star \xrightarrow{L^2(\mu)} \nabla\varphi^{\mathrm{Br}},$$

where $\nabla\varphi^{\mathrm{Br}}$ is the Brenier map.

The Schrödinger system

Theorem (Fortet'40...)

Let

$$\mu(\mathrm{d}x) = \exp(-U^{\mu}(x))\mathrm{d}x \quad \text{and} \quad \nu(\mathrm{d}y) = \exp(-U^{\nu}(y))\mathrm{d}y.$$

Then φ, ψ solve the **Schrödinger system**

$$\begin{cases} \varphi(x) = U^{\mu}(x) + \log \mathrm{P}_T \exp(-\psi)(x) \\ \psi(y) = U^{\nu}(y) + \log \mathrm{P}_T \exp(-\varphi)(y) \end{cases}$$

where P_T is the heat semigroup.

$$\mathrm{P}_T f(x) = \frac{1}{(2\pi T)^{d/2}} \int f(y) e^{-\frac{|y-x|^2}{2T}} \mathrm{d}y$$

Sinkhorn's algorithm

Sinkhorn's iteration

- Choose (φ^0, ψ^0)
- Given (φ^n, ψ^n) compute

$$\varphi^{n+1} = U^\mu + \log \mathrm{P}_T \exp(-\psi^n)$$
$$\psi^{n+1} = U^\nu + \log \mathrm{P}_T \exp(-\varphi^{n+1})$$

$$\varphi^n \longrightarrow \varphi^\star \quad \text{and} \quad \psi^n \longrightarrow \psi^\star$$

- A.k.a. Iterative Proportional Fitting Procedure (IPFP)
- Successful applications in statistical ML
- Marco Cuturi. "Sinkhorn distances: Lightspeed computation of optimal transport". In: *Advances in Neural Information Processing Systems*. 2013, pp. 2292–2300

Schrödinger system and HJB equations

Feynman-Kac formula

We have

$$-\log \mathrm{P}_T \exp(-g) = U_0^{T,g}$$

where $(U_t^{T,g})_{t\leq T}$ is the only (classical) solution of

$$\begin{cases} \partial_t \varphi_t + \frac{1}{2}\Delta\varphi_t - \frac{1}{2}|\nabla\varphi_t|^2 = 0 \\ \varphi_T = g \end{cases} \tag{HJB}$$

Schrödinger system and Sinkhorn's algorithm

$$\begin{cases} \psi = U^\nu - U_0^{T,\varphi} \\ \varphi = U^\mu - U_0^{T,\psi} \end{cases}$$

$$\varphi^{n+1} = U^\mu - U_0^{T,\psi^{n+1}}, \quad \psi^{n+1} = U^\nu - U^{T,\varphi^{n+1}}$$

Schrödinger system and HJB equations

Feynman-Kac formula

We have

$$-\log \mathrm{P}_T \exp(-g) = U_0^{T,g}$$

where $(U_t^{T,g})_{t \leq T}$ is the only (classical) solution of

$$\begin{cases} \partial_t \varphi_t + \frac{1}{2}\Delta\varphi_t - \frac{1}{2}|\nabla\varphi_t|^2 = 0 \\ \varphi_T = g \end{cases} \tag{HJB}$$

Schrödinger system and Sinkhorn's algorithm

$$\begin{cases} \psi = U^\nu - U_0^{T,\varphi} \\ \varphi = U^\mu - U_0^{T,\psi} \end{cases}$$

$$\varphi^{n+1} = U^\mu - U_0^{T,\psi^{n+1}}, \quad \psi^{n+1} = U^\nu - U^{T,\varphi^{n+1}}$$

Exponential convergence of Sinkhorn's algorithm

Convergence of Sinkhorn algorithm

Exponential convergence in n

- ✓ Bounded costs or compact manifolds
- ✓ Compactly supported marginals

For unbounded costs and marginals only *polynomial* convergence rates are known!

Current setup

- Quadratic cost
- Weakly log-concave marginals

Integrated modulus of convexity

Definition

Let $U : \mathbb{R}^d \longrightarrow \mathbb{R}$

$$\kappa_U(r) = \inf\{r^{-2}\langle \nabla U(x) - \nabla U(\hat{x}), x - \hat{x}\rangle : |x - \hat{x}| = r\}.$$

Relation with semiconvexity

$$\kappa_U(r) \geq \alpha \quad \forall r > 0 \quad \Leftrightarrow \nabla^2 U(x) \succeq \alpha \mathrm{I} \quad \forall x \in \mathbb{R}^d.$$

Characterization of $\kappa_U(r) \geq \alpha$

$$\int_0^r \langle v, \nabla^2 U(x + \theta v), v\rangle d\theta \geq \alpha\, r \quad \forall x, v \in \mathbb{R}^d,\ |v| = 1.$$

Integrated modulus of convexity

Definition

Let $U : \mathbb{R}^d \longrightarrow \mathbb{R}$

$$\kappa_U(r) = \inf\{r^{-2}\langle \nabla U(x) - \nabla U(\hat{x}), x - \hat{x}\rangle : |x - \hat{x}| = r\}.$$

Relation with semiconvexity

$$\kappa_U(r) \geq \alpha \quad \forall r > 0 \quad \Leftrightarrow \nabla^2 U(x) \succeq \alpha \mathrm{I} \quad \forall x \in \mathbb{R}^d.$$

Characterization of $\kappa_U(r) \geq \alpha$

$$\int_0^r \langle v, \nabla^2 U(x + \theta v), v\rangle d\theta \geq \alpha\, r \quad \forall x, v \in \mathbb{R}^d,\ |v| = 1.$$

Integrated modulus of convexity

Definition

Let $U : \mathbb{R}^d \longrightarrow \mathbb{R}$

$$\kappa_U(r) = \inf\{r^{-2}\langle \nabla U(x) - \nabla U(\hat{x}), x - \hat{x}\rangle : |x - \hat{x}| = r\}.$$

Relation with semiconvexity

$$\kappa_U(r) \geq \alpha \quad \forall r > 0 \quad \Leftrightarrow \nabla^2 U(x) \succeq \alpha \mathrm{I} \quad \forall x \in \mathbb{R}^d.$$

Characterization of $\kappa_U(r) \geq \alpha$

$$\int_0^r \langle v, \nabla^2 U(x + \theta v), v\rangle d\theta \geq \alpha\, r \quad \forall x, v \in \mathbb{R}^d,\ |v| = 1.$$

Weakly log-concave probability measures

Definition

$\mathrm{d}\nu = e^{-U_\nu}\mathrm{dLeb}$ is weakly log-concave if there exist $\alpha_\nu > 0$, $L_\nu, R_\nu \geq 0$ s.t.

$$\kappa_{U^\nu}(r) \geq \begin{cases} \alpha_\nu - L_\nu & \text{if } r \leq R_\nu \\ \alpha_\nu, & \text{if } r \geq R_\nu \end{cases}$$

- U^ν is $\alpha_\nu - L^\nu$-semiconvex
- U^ν behaves like a α_ν-strongly convex function for distant points

Example

$$U^\nu = U_c + U_l, \quad \text{with } U_c \text{ strongly convex, } U_l \text{ Lipschitz+semiconvex}$$

- More general than

$$\begin{cases} \text{"convex outside a ball"} \\ \text{"bounded pert. of log-concave"} \end{cases}$$

Weakly log-concave probability measures

Definition

$\mathrm{d}\nu = e^{-U_\nu}\mathrm{dLeb}$ is weakly log-concave if there exist $\alpha_\nu > 0$, $L_\nu, R_\nu \geq 0$ s.t.

$$\kappa_{U^\nu}(r) \geq \begin{cases} \alpha_\nu - L_\nu & \text{if } r \leq R_\nu \\ \alpha_\nu, & \text{if } r \geq R_\nu \end{cases}$$

- U^ν is $\alpha_\nu - L^\nu$-semiconvex
- U^ν behaves like a α_ν-strongly convex function for distant points

Example

$$U^\nu = U_c + U_l, \quad \text{with } U_c \text{ strongly convex, } U_l \text{ Lipschitz+semiconvex}$$

- More general than

$$\begin{cases} \text{"convex outside a ball"} \\ \text{"bounded pert. of log-concave"} \end{cases}$$

Weakly log-concave probability measures

Definition

$\mathrm{d}\nu = e^{-U_\nu}\mathrm{dLeb}$ is weakly log-concave if there exist $\alpha_\nu > 0$, $L_\nu, R_\nu \geq 0$ s.t.

$$\kappa_{U^\nu}(r) \geq \begin{cases} \alpha_\nu - L_\nu & \text{if } r \leq R_\nu \\ \alpha_\nu, & \text{if } r \geq R_\nu \end{cases}$$

- U^ν is $\alpha_\nu - L^\nu$-semiconvex
- U^ν behaves like a α_ν-strongly convex function for distant points

Example

$$U^\nu = U_c + U_l, \quad \text{with } U_c \text{ strongly convex}, \ U_l \text{ Lipschitz+semiconvex}$$

- More general than

$$\begin{cases} \text{"convex outside a ball"} \\ \text{"bounded pert. of log-concave"} \end{cases}$$

Exponential convergence (simplified)

Theorem(C.,Durmus,Greco'23)

Assume that μ, ν are weakly log-concave. Then there exist T_0 such that for all $T > T_0$ there exists a rate $\lambda_T > 0$ s.t.

$$\|\nabla\varphi^n - \nabla\varphi^\star\|_{L^1(\mu)} + \|\nabla\psi^n - \nabla\psi^\star\|_{L^1(\nu)} \leq C\exp(-\lambda_T\, n) \quad \forall n \geq 0.$$

- Explicit constants
- Weaker assumptions
- *Pointwise* exponential convergence for large T

Exponential convergence results for

- Optimal plans
- Hessians of potentials

Exponential convergence (simplified)

Theorem(C.,Durmus,Greco'23)

Assume that μ, ν are weakly log-concave. Then there exist T_0 such that for all $T > T_0$ there exists a rate $\lambda_T > 0$ s.t.

$$\|\nabla\varphi^n - \nabla\varphi^\star\|_{L^1(\mu)} + \|\nabla\psi^n - \nabla\psi^\star\|_{L^1(\nu)} \leq C\exp(-\lambda_T\, n) \quad \forall n \geq 0.$$

- Explicit constants
- Weaker assumptions
- *Pointwise* exponential convergence for large T

Exponential convergence results for

- Optimal plans
- Hessians of potentials

Exponential convergence (simplified)

Theorem(C.,Durmus,Greco'23)

Assume that μ, ν are weakly log-concave. Then there exist T_0 such that for all $T > T_0$ there exists a rate $\lambda_T > 0$ s.t.

$$\|\nabla\varphi^n - \nabla\varphi^\star\|_{L^1(\mu)} + \|\nabla\psi^n - \nabla\psi^\star\|_{L^1(\nu)} \leq C\exp(-\lambda_T\, n) \quad \forall n \geq 0.$$

- Explicit constants
- Weaker assumptions
- *Pointwise* exponential convergence for large T

Exponential convergence results for

- Optimal plans
- Hessians of potentials

Sketch of proof

Representation of HJB gradient

$$\nabla U_0^{T,g}(x) = \frac{1}{T}\int (y-x)\,\pi_T^{x,g}(\mathrm{d}y)$$

where

$$\pi_T^{x,g}(\mathrm{d}y) \propto \exp\Big(-\frac{|y-x|^2}{2T} - g(y)\Big)\mathrm{d}y$$

Apply to Sinkhorn iterates

$$|\nabla\varphi^{n+1} - \nabla\varphi^{\star}|(x) \leq \frac{1}{T} W_1(\pi_T^{x,\psi^n}, \pi_T^{x,\psi^\star})$$

with W_1 the Wasserstein distance of order one.

We need toolbox for estimating transport distances

- Coupling methods
- Stein's method

Sketch of proof

Representation of HJB gradient

$$\nabla U_0^{T,g}(x) = \frac{1}{T}\int (y-x)\,\pi_T^{x,g}(\mathrm{d}y)$$

where

$$\pi_T^{x,g}(\mathrm{d}y) \propto \exp\Big(-\frac{|y-x|^2}{2T} - g(y)\Big)\mathrm{d}y$$

Apply to Sinkhorn iterates

$$|\nabla\varphi^{n+1} - \nabla\varphi^\star|(x) \le \frac{1}{T} W_1(\pi_T^{x,\psi^n}, \pi_T^{x,\psi^\star})$$

with W_1 the Wasserstein distance of order one.

We need toolbox for estimating transport distances

- Coupling methods
- Stein's method

Sketch of proof

Key estimate

$$W_1(\pi_T^{x,\psi^n}, \pi_T^{x,\psi^\star}) \leq \gamma_n^\nu \|\nabla\psi^n - \nabla\psi^\star\|_{L^1(\pi_T^{x,\psi^\star})}$$

Hold if π_T^{x,ψ^n} is weakly log-concave.

Averaging step

$$\|\nabla\varphi^{n+1} - \nabla\varphi^\star\|_{L^1(\mu)} \leq T^{-1}\gamma_n^\nu \|\nabla\psi^n - \nabla\psi^\star\|_{L^1(\nu)}$$

Iterate the argument

$$\|\nabla\varphi^{n+1} - \nabla\varphi^\star\|_{L^1(\mu)} \leq T^{-2}\gamma_n^\mu\gamma_n^\nu \|\nabla\varphi^n - \nabla\varphi^\star\|_{L^1(\mu)}$$

Key question is then...

How (semi)convex are the potentials ψ^n?

Sketch of proof

Key estimate

$$W_1(\pi_T^{x,\psi^n}, \pi_T^{x,\psi^\star}) \leq \gamma_n^\nu \|\nabla\psi^n - \nabla\psi^\star\|_{L^1(\pi_T^{x,\psi^\star})}$$

Hold if π_T^{x,ψ^n} is weakly log-concave.

Averaging step

$$\|\nabla\varphi^{n+1} - \nabla\varphi^\star\|_{L^1(\mu)} \leq T^{-1}\gamma_n^\nu \|\nabla\psi^n - \nabla\psi^\star\|_{L^1(\nu)}$$

Iterate the argument

$$\|\nabla\varphi^{n+1} - \nabla\varphi^\star\|_{L^1(\mu)} \leq T^{-2}\gamma_n^\mu\gamma_n^\nu \|\nabla\varphi^n - \nabla\varphi^\star\|_{L^1(\mu)}$$

Key question is then...

How (semi)convex are the potentials ψ^n?

Sketch of proof

Key estimate

$$W_1(\pi_T^{x,\psi^n}, \pi_T^{x,\psi^\star}) \leq \gamma_n^\nu \|\nabla\psi^n - \nabla\psi^\star\|_{L^1(\pi_T^{x,\psi^\star})}$$

Hold if π_T^{x,ψ^n} is weakly log-concave.

Averaging step

$$\|\nabla\varphi^{n+1} - \nabla\varphi^\star\|_{L^1(\mu)} \leq T^{-1}\gamma_n^\nu \|\nabla\psi^n - \nabla\psi^\star\|_{L^1(\nu)}$$

Iterate the argument

$$\|\nabla\varphi^{n+1} - \nabla\varphi^\star\|_{L^1(\mu)} \leq T^{-2}\gamma_n^\mu\gamma_n^\nu \|\nabla\varphi^n - \nabla\varphi^\star\|_{L^1(\mu)}$$

Key question is then...

How (semi)convex are the potentials ψ^n?

Sketch of proof

Key estimate

$$W_1(\pi_T^{x,\psi^n}, \pi_T^{x,\psi^\star}) \leq \gamma_n^\nu \|\nabla\psi^n - \nabla\psi^\star\|_{L^1(\pi_T^{x,\psi^\star})}$$

Hold if π_T^{x,ψ^n} is weakly log-concave.

Averaging step

$$\|\nabla\varphi^{n+1} - \nabla\varphi^\star\|_{L^1(\mu)} \leq T^{-1}\gamma_n^\nu \|\nabla\psi^n - \nabla\psi^\star\|_{L^1(\nu)}$$

Iterate the argument

$$\|\nabla\varphi^{n+1} - \nabla\varphi^\star\|_{L^1(\mu)} \leq T^{-2}\gamma_n^\mu\gamma_n^\nu \|\nabla\varphi^n - \nabla\varphi^\star\|_{L^1(\mu)}$$

Key question is then...

How (semi)convex are the potentials ψ^n?

Propagation of weak log-concavity

Two basic observations

HJB preserves convexity

$$\varphi \text{ convex} \Rightarrow U_0^{T,\varphi} \text{ convex}$$

HJB preserves concavity

$$\psi \text{ concave} \Rightarrow U_0^{T,\psi} \text{ concave}$$

- **Idea:** Implement these observations along Sinkhorn's algorithm
- Max Fathi, Nathael Gozlan, and Maxime Prodhomme. "A proof of the Caffarelli contraction theorem via entropic regularization". In: *Calculus of Variations and Partial Differential Equations* 59.96 (2020)

- concavity $\hookrightarrow$ Jensen/Cramer-Rao inequality
- convexity $\hookrightarrow$ Prékopa-Leindler/Brascamp-Lieb inequality

Propagation of convexity along Sinkhorn's iterations

Assume U^μ concave and U^ν convex.

ψ^n **convex** $\Rightarrow \varphi^{n+1}$ **concave**

$$\varphi^{n+1} = U^\mu - \underbrace{U_0^{T,\psi^n}}_{\text{HJB propagates convexity}}$$

φ^{n+1} **concave** $\Rightarrow \psi^{n+1}$ **convex**

$$\psi^{n+1} = U^\nu - \underbrace{U_0^{T,\varphi^{n+1}}}_{\text{HJB propagates concavity}}$$

Therefore we have

$$\varphi_n \text{ concave} \quad \Rightarrow \varphi_{n+1} \text{ concave} \quad \Rightarrow \varphi \text{ concave}$$

What can we do in absence of *pointwise* convexity?

Propagation of convexity along Sinkhorn's iterations

Assume U^μ concave and U^ν convex.

ψ^n **convex** $\Rightarrow \varphi^{n+1}$ **concave**

$$\varphi^{n+1} = U^\mu - \underbrace{U_0^{T,\psi^n}}_{\text{HJB propagates convexity}}$$

φ^{n+1} **concave** $\Rightarrow \psi^{n+1}$ **convex**

$$\psi^{n+1} = U^\nu - \underbrace{U_0^{T,\varphi^{n+1}}}_{\text{HJB propagates concavity}}$$

Therefore we have

$$\varphi_n \text{ concave} \quad \Rightarrow \varphi_{n+1} \text{ concave} \quad \Rightarrow \varphi \text{ concave}$$

What can we do in absence of *pointwise* convexity?

Propagation of convexity along Sinkhorn's iterations

Assume U^μ concave and U^ν convex.

ψ^n **convex** $\Rightarrow \varphi^{n+1}$ **concave**

$$\varphi^{n+1} = U^\mu - \underbrace{U_0^{T,\psi^n}}_{\text{HJB propagates convexity}}$$

φ^{n+1} **concave** $\Rightarrow$ ψ^{n+1} **convex**

$$\psi^{n+1} = U^\nu - \underbrace{U_0^{T,\varphi^{n+1}}}_{\text{HJB propagates concavity}}$$

Therefore we have

$$\varphi_n \,\text{concave} \quad \Rightarrow \varphi_{n+1} \,\text{concave} \quad \Rightarrow \varphi \,\text{concave}$$

What can we do in absence of *pointwise* convexity?

A class of invariant weakly convex functions

Problem

Find **invariant** sets of "approximately" convex functions for

$$g \mapsto U_0^{T,g}$$

Semiconvexity does not propagate

$$\nabla^2 g \succeq -\varepsilon \mathrm{I} \ \Rightarrow \ \nabla^2 U_0^{T,g} \succeq -\varepsilon_T,$$
$$\varepsilon_T = \frac{1}{T - \varepsilon^{-1}}.$$

- The estimate is sharp for quadratic functions
- ε_T blows up in finite time!

- **Idea:** Look for sets of the form $\{\kappa_g \geq h\}$, and find good h!

A class of invariant weakly convex functions

Problem

Find **invariant** sets of "approximately" convex functions for

$$g \mapsto U_0^{T,g}$$

Semiconvexity does not propagate

$$\nabla^2 g \succeq -\varepsilon \mathrm{I} \;\Rightarrow\; \nabla^2 U_0^{T,g} \succeq -\varepsilon_T,$$
$$\varepsilon_T = \frac{1}{T - \varepsilon^{-1}}.$$

- The estimate is sharp for quadratic functions
- ε_T blows up in finite time!

- **Idea:** Look for sets of the form $\{\kappa_g \geq h\}$, and find good h!

A class of invariant weakly convex functions

Problem

Find **invariant** sets of "approximately" convex functions for

$$g \mapsto U_0^{T,g}$$

Semiconvexity does not propagate

$$\nabla^2 g \succeq -\varepsilon \mathrm{I} \ \Rightarrow \ \nabla^2 U_0^{T,g} \succeq -\varepsilon_T,$$
$$\varepsilon_T = \frac{1}{T - \varepsilon^{-1}}.$$

- The estimate is sharp for quadratic functions
- ε_T blows up in finite time!

- **Idea:** Look for sets of the form $\{\kappa_g \geq h\}$, and find good h!

Classes of invariant weakly convex functions

Theorem (C.'22)

For any $L > 0$ consider

$$f_L : [0, +\infty] \longrightarrow [0, +\infty], \quad f_L(r) = (2L)^{1/2} \tanh\left(\frac{1}{2}(2L)^{1/2} r\right).$$

and define

$$\mathcal{F}_L = \{g \in C^1(\mathbb{R}^d) : \kappa_g(r) \geq -r^{-1} f_L(r) \quad \forall r > 0\}.$$

Then we have

$$g \in \mathcal{F}_L \Rightarrow U_0^{T,g} \in \mathcal{F}_L.$$

Are the $\mathcal{F}_L$ rich enough?

Yes! μ weakly log-concave $\Rightarrow U^\mu \in \mathcal{F}_L$ for some L.

Classes of invariant weakly convex functions

Theorem (C.'22)

For any $L > 0$ consider

$$f_L : [0, +\infty] \longrightarrow [0, +\infty], \quad f_L(r) = (2L)^{1/2} \tanh\left(\frac{1}{2}(2L)^{1/2} r\right).$$

and define

$$\mathcal{F}_L = \{g \in C^1(\mathbb{R}^d) : \kappa_g(r) \geq -r^{-1} f_L(r) \quad \forall r > 0\}.$$

Then we have

$$g \in \mathcal{F}_L \Rightarrow U_0^{T,g} \in \mathcal{F}_L.$$

Are the $\mathcal{F}_L$ rich enough?

Yes! μ weakly log-concave $\Rightarrow U^\mu \in \mathcal{F}_L$ for some L.

Classes of invariant weakly convex functions

Theorem (C.'22)

For any $L > 0$ consider

$$f_L : [0, +\infty] \longrightarrow [0, +\infty], \quad f_L(r) = (2L)^{1/2} \tanh\Big(\frac{1}{2}(2L)^{1/2} r\Big).$$

and define

$$\mathcal{F}_L = \{g \in C^1(\mathbb{R}^d) : \kappa_g(r) \geq -r^{-1} f_L(r) \quad \forall r > 0\}.$$

Then we have

$$g \in \mathcal{F}_L \Rightarrow U_0^{T,g} \in \mathcal{F}_L.$$

Are the $\mathcal{F}_L$ rich enough?

Yes! μ weakly log-concave $\Rightarrow U^\mu \in \mathcal{F}_L$ for some L.

Classes of invariant weakly convex functions

Theorem (C.'22)

For any $L > 0$ consider

$$f_L : [0, +\infty] \longrightarrow [0, +\infty], \quad f_L(r) = (2L)^{1/2} \tanh\Big(\frac{1}{2}(2L)^{1/2} r\Big).$$

and define

$$\mathcal{F}_L = \{g \in C^1(\mathbb{R}^d) : \kappa_g(r) \geq -r^{-1} f_L(r) \quad \forall r > 0\}.$$

Then we have

$$g \in \mathcal{F}_L \Rightarrow U_0^{T,g} \in \mathcal{F}_L.$$

Are the $\mathcal{F}_L$ rich enough?

Yes! μ weakly log-concave $\Rightarrow U^{\mu} \in \mathcal{F}_L$ for some L.

Propagation of weak convexity along Sinkhorn

Theorem(C.,Durmus, Greco'23)

Assume

- ν weakly log-concave
- There exists $\beta_\mu \in (0, +\infty]$ s.t.

$$\nabla^2 U^\mu(x) \preceq \beta_\mu \mathrm{I} \quad \forall x \in \mathbb{R}^d.$$

Then π_T^{x,ψ^n} is weakly log-concave uniformly in x, n where ψ^n is the n-th Sinkhorn iterate and

$$\pi_T^{x,\psi^n}(\mathrm{d}y) \propto \exp\Big(- \frac{|y-x|^2}{2T} - \psi^n(y) \Big)$$

- Key estimate for exponential convergence
- $\beta_\mu = +\infty$ is allowed!

Propagation of weak convexity along Sinkhorn

Theorem(C.,Durmus, Greco'23)

Assume

- ν weakly log-concave
- There exists $\beta_\mu \in (0, +\infty]$ s.t.

$$\nabla^2 U^\mu(x) \preceq \beta_\mu \mathrm{I} \quad \forall x \in \mathbb{R}^d.$$

Then π_T^{x,ψ^n} is weakly log-concave uniformly in x, n where ψ^n is the n-th Sinkhorn iterate and

$$\pi_T^{x,\psi^n}(\mathrm{d}y) \propto \exp\Big(- \frac{|y-x|^2}{2T} - \psi^n(y) \Big)$$

- Key estimate for exponential convergence
- $\beta_\mu = +\infty$ is allowed!

A second order analysis of coupling by reflection along characterstics

HJB equation and characteristics

Characteristics of the HJB equation

$$\mathrm{d}X_t = -\nabla U_t^{T,g}(X_t)\mathrm{d}t + \mathrm{d}B_t$$

Stochastic Maximum Principle

Let $(X_t)_{t\leq T}$ be a characteristic. Then

$$\mathrm{d}\nabla U_t^{T,g}(X_t) = \nabla^2 U_t^{T,g}(X_t) \cdot \mathrm{d}B_t$$

i.e. the drift is a **martingale**

Propagation of convexity via characteristics

Let $(X_t, \hat{X}_t)_{t\leq T}$ be **synchronous coupling** of two characteristics. Then

$$\mathcal{U}_t = |X_t - \hat{X}_t|^{-1}\langle \nabla U_t^{T,g}(X_t) - \nabla U_t^{T,g}(\hat{X}_t), X_t - \hat{X}_t\rangle$$

is a supermartingale.

HJB equation and characteristics

Characteristics of the HJB equation

$$\mathrm{d}X_t = -\nabla U_t^{T,g}(X_t)\mathrm{d}t + \mathrm{d}B_t$$

Stochastic Maximum Principle

Let $(X_t)_{t\leq T}$ be a characteristic. Then

$$\mathrm{d}\nabla U_t^{T,g}(X_t) = \nabla^2 U_t^{T,g}(X_t)\cdot \mathrm{d}B_t$$

i.e. the drift is a **martingale**

Propagation of convexity via characteristics

Let $(X_t, \hat{X}_t)_{t\leq T}$ be **synchronous coupling** of two characteristics. Then

$$\mathcal{U}_t = |X_t - \hat{X}_t|^{-1}\langle \nabla U_t^{T,g}(X_t) - \nabla U_t^{T,g}(\hat{X}_t), X_t - \hat{X}_t\rangle$$

is a supermartingale.

HJB equation and characteristics

Characteristics of the HJB equation

$$\mathrm{d}X_t = -\nabla U_t^{T,g}(X_t)\mathrm{d}t + \mathrm{d}B_t$$

Stochastic Maximum Principle

Let $(X_t)_{t \leq T}$ be a characteristic. Then

$$\mathrm{d}\nabla U_t^{T,g}(X_t) = \nabla^2 U_t^{T,g}(X_t) \cdot \mathrm{d}B_t$$

i.e. the drift is a **martingale**

Propagation of convexity via characteristics

Let $(X_t, \hat{X}_t)_{t \leq T}$ be **synchronous coupling** of two characteristics. Then

$$\mathcal{U}_t = |X_t - \hat{X}_t|^{-1} \langle \nabla U_t^{T,g}(X_t) - \nabla U_t^{T,g}(\hat{X}_t), X_t - \hat{X}_t \rangle$$

is a supermartingale.

Proof idea

1) Build **coupling by reflection** $(X_t, \hat{X}_t)$ of two characteristics with initial states $x, \hat{x}$
2) Introduce processes $(\Gamma_t, \mathcal{U}_t)_{t \leq T}$

$$\mathcal{U}_t = |X_t - \hat{X}_t|^{-1} \langle \nabla U_t^{T,g}(X_t) - \nabla U_t^{T,g}(\hat{X}_t), X_t - \hat{X}_t \rangle$$
$$\Gamma_t = \exp\left(\int_0^t f_L'(|X_s - \hat{X}_s|)\mathrm{d}s\right)(\mathcal{U}_t + f_L(|X_t - \hat{X}_t|))$$

3) Show that $(\Gamma_t)_{t \leq T}$ is a supermartingale
4) Impose $\mathbb{E}[\Gamma_0] \geq \mathbb{E}[\Gamma_T]$

$$\mathbb{E}[\Gamma_0] = |x - \hat{x}|^{-1} \langle \nabla U_0^{T,g}(x) - \nabla U_0^{T,g}(\hat{x}), x - \hat{x} \rangle + f_L(|x - \hat{x}|)$$
$$\mathbb{E}[\Gamma_T] \geq 0 \quad \text{since } g \in \mathcal{F}_L$$

Proof idea

1) Build **coupling by reflection** $(X_t, \hat{X}_t)$ of two characteristics with initial states $x, \hat{x}$
2) Introduce processes $(\Gamma_t, \mathcal{U}_t)_{t \leq T}$

$$\mathcal{U}_t = |X_t - \hat{X}_t|^{-1} \langle \nabla U_t^{T,g}(X_t) - \nabla U_t^{T,g}(\hat{X}_t), X_t - \hat{X}_t \rangle$$

$$\Gamma_t = \exp\left(\int_0^t f_L'(|X_s - \hat{X}_s|) \mathrm{d}s \right) (\mathcal{U}_t + f_L(|X_t - \hat{X}_t|))$$

3) Show that $(\Gamma_t)_{t \leq T}$ is a supermartingale
4) Impose $\mathbb{E}[\Gamma_0] \geq \mathbb{E}[\Gamma_T]$

$$\mathbb{E}[\Gamma_0] = |x - \hat{x}|^{-1} \langle \nabla U_0^{T,g}(x) - \nabla U_0^{T,g}(\hat{x}), x - \hat{x} \rangle + f_L(|x - \hat{x}|)$$

$$\mathbb{E}[\Gamma_T] \geq 0 \quad \text{since } g \in \mathcal{F}_L$$

Proof idea

1) Build **coupling by reflection** $(X_t, \hat{X}_t)$ of two characteristics with initial states $x, \hat{x}$
2) Introduce processes $(\Gamma_t, \mathcal{U}_t)_{t\leq T}$

$$\mathcal{U}_t = |X_t - \hat{X}_t|^{-1}\langle \nabla U_t^{T,g}(X_t) - \nabla U_t^{T,g}(\hat{X}_t), X_t - \hat{X}_t\rangle$$
$$\Gamma_t = \exp\left(\int_0^t f_L'(|X_s - \hat{X}_s|)\mathrm{d}s\right)(\mathcal{U}_t + f_L(|X_t - \hat{X}_t|))$$

3) Show that $(\Gamma_t)_{t\leq T}$ is a supermartingale
4) Impose $\mathbb{E}[\Gamma_0] \geq \mathbb{E}[\Gamma_T]$

$$\mathbb{E}[\Gamma_0] = |x - \hat{x}|^{-1}\langle \nabla U_0^{T,g}(x) - \nabla U_0^{T,g}(\hat{x}), x - \hat{x}\rangle + f_L(|x - \hat{x}|)$$
$$\mathbb{E}[\Gamma_T] \geq 0 \quad \text{since } g \in \mathcal{F}_L$$

Proof idea

1) Build **coupling by reflection** $(X_t, \hat{X}_t)$ of two characteristics with initial states $x, \hat{x}$
2) Introduce processes $(\Gamma_t, \mathcal{U}_t)_{t \le T}$

$$\mathcal{U}_t = |X_t - \hat{X}_t|^{-1} \langle \nabla U_t^{T,g}(X_t) - \nabla U_t^{T,g}(\hat{X}_t), X_t - \hat{X}_t \rangle$$
$$\Gamma_t = \exp\left(\int_0^t f_L'(|X_s - \hat{X}_s|)\mathrm{d}s \right) (\mathcal{U}_t + f_L(|X_t - \hat{X}_t|))$$

3) Show that $(\Gamma_t)_{t \le T}$ is a supermartingale
4) Impose $\mathbb{E}[\Gamma_0] \ge \mathbb{E}[\Gamma_T]$

$$\mathbb{E}[\Gamma_0] = |x - \hat{x}|^{-1} \langle \nabla U_0^{T,g}(x) - \nabla U_0^{T,g}(\hat{x}), x - \hat{x} \rangle + f_L(|x - \hat{x}|)$$
$$\mathbb{E}[\Gamma_T] \ge 0 \quad \text{since } g \in \mathcal{F}_L$$

Coupling by reflection for HJB characteristics

Coupling by reflection

Let $B_\cdot$ be a BM.

$$\begin{cases} \mathrm{d}X_t = -\nabla U_t^{T,g}(X_t)\mathrm{d}s + \mathrm{d}B_t \\ \mathrm{d}\hat{X}_t = -\nabla U_t^{T,g}(\hat{X}_t)\mathrm{d}s + \mathrm{d}\hat{B}_t \end{cases} \qquad \begin{cases} \mathrm{d}\hat{B}_t = (\mathrm{I} - 2\mathrm{e}_t\mathrm{e}_t^\top)\cdot \mathrm{d}B_t \\ \mathrm{e}_t = \frac{X_t - \hat{X}_t}{|X_t - \hat{X}_t|} \end{cases}$$

B vs $\hat{\mathbf{B}}$

- Increments $\perp$ to $X_t - \hat{X}_t$ are the same
- Increments $\parallel$ to $X_t - \hat{X}_t$ are reflected

The angle process

- e_t = angle under which $\hat{X}_t$ sees X_t

Coupling by reflection for HJB characteristics

Coupling by reflection

Let $B_\cdot$ be a BM.

$$\begin{cases} \mathrm{d}X_t = -\nabla U_t^{T,g}(X_t)\mathrm{d}s + \mathrm{d}B_t \\ \mathrm{d}\hat{X}_t = -\nabla U_t^{T,g}(\hat{X}_t)\mathrm{d}s + \mathrm{d}\hat{B}_t \end{cases} \quad \begin{cases} \mathrm{d}\hat{B}_t = (\mathrm{I} - 2\mathrm{e}_t\mathrm{e}_t^\top)\cdot \mathrm{d}B_t \\ \mathrm{e}_t = \frac{X_t-\hat{X}_t}{|X_t-\hat{X}_t|} \end{cases}$$

B vs $\hat{\mathbf{B}}$

- Increments $\perp$ to $X_t - \hat{X}_t$ are the same
- Increments $\parallel$ to $X_t - \hat{X}_t$ are reflected

The angle process

- e_t = angle under which $\hat{X}_t$ sees X_t

Coupling by reflection for HJB characteristics

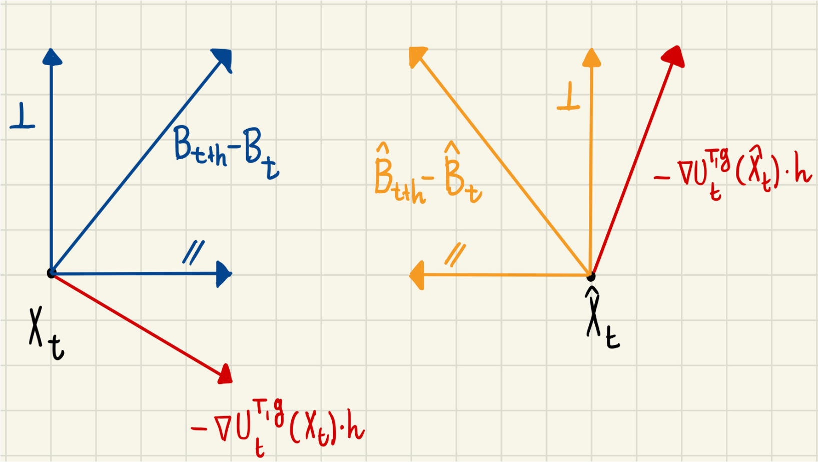


Figure 1: Discretisation of coup. by ref. between times t and $t + h$

Angle process

Dynamics of the angle process

$$\mathrm{d}\mathrm{e}_t = -|X_t - \hat{X}_t|^{-1}\mathrm{proj}_{\mathrm{e}_t^{\perp}}(\nabla U_t^{T,g}(X_t) - \nabla U_t^{T,g}(\hat{X}_t))\mathrm{d}t$$

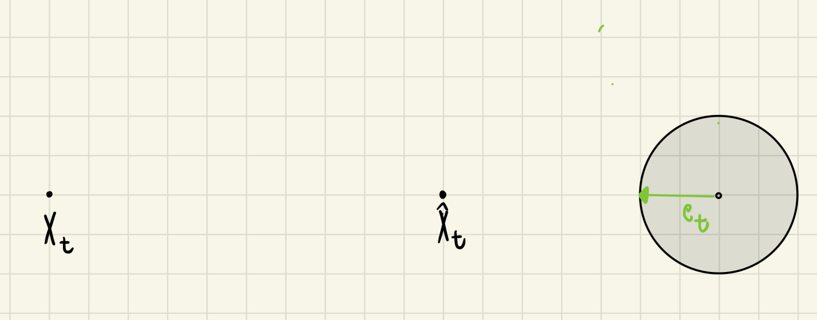
- No $\mathrm{d}B_t$ part!

Angle process

Dynamics of the angle process

$$\mathrm{d}\mathrm{e}_t = -|X_t - \hat{X}_t|^{-1}\mathrm{proj}_{\mathrm{e}_t^\perp}(\nabla U_t^{T,g}(X_t) - \nabla U_t^{T,g}(\hat{X}_t))\mathrm{d}t$$

- No $\mathrm{d}B_t$ part!

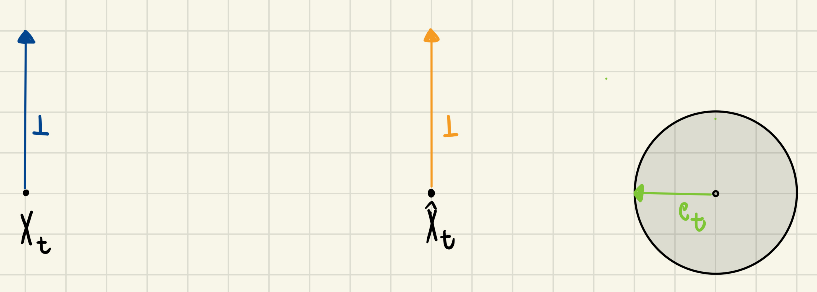


Angle process

Dynamics of the angle process

$$\mathrm{d}\mathrm{e}_t = -|X_t - \hat{X}_t|^{-1} \mathrm{proj}_{\mathrm{e}_t^\perp}(\nabla U_t^{T,g}(X_t) - \nabla U_t^{T,g}(\hat{X}_t))\mathrm{d}t$$

- No $\mathrm{d}B_t$ part!

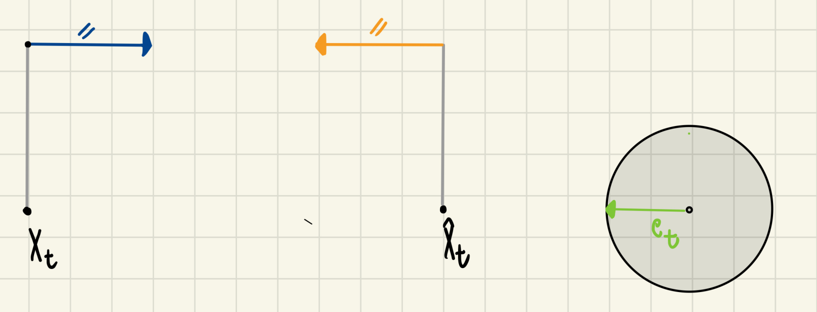


Angle process

Dynamics of the angle process

$$\mathrm{d}\mathrm{e}_t = -|X_t - \hat{X}_t|^{-1}\mathrm{proj}_{\mathrm{e}_t^{\perp}}(\nabla U_t^{T,g}(X_t) - \nabla U_t^{T,g}(\hat{X}_t))\mathrm{d}t$$

- No $\mathrm{d}B_t$ part!

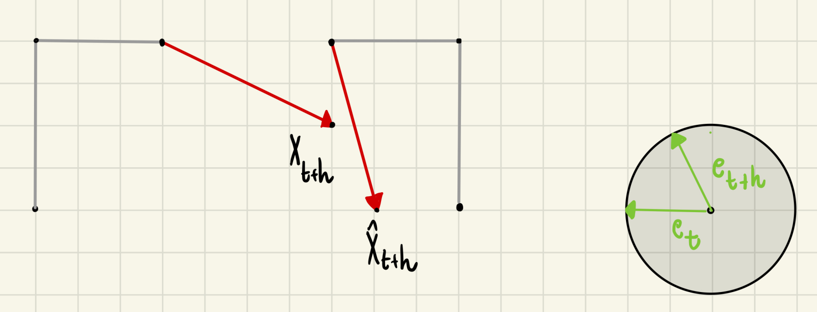


Angle process

Dynamics of the angle process

$$\mathrm{d}\mathrm{e}_t = -|X_t - \hat{X}_t|^{-1} \mathrm{proj}_{\mathrm{e}_t^\perp} (\nabla U_t^{T,g}(X_t) - \nabla U_t^{T,g}(\hat{X}_t))\mathrm{d}t$$

- No $\mathrm{d}B_t$ part!



$\mathcal{U}_t$ is a supermartingale

Recall that

$$\mathcal{U}_t = \langle \nabla U_t^{T,g}(X_t) - \nabla U_t^{T,g}(\hat{X}_t), \mathrm{e}_t \rangle$$

Lemma

Proof.

Itô+ angle process differential + Stochastic Maximum Principle □

$\mathcal{U}_t$ is a supermartingale

Recall that

$$\mathcal{U}_t = \langle \nabla U_t^{T,g}(X_t) - \nabla U_t^{T,g}(\hat{X}_t), \mathrm{e}_t \rangle$$

Lemma

$$\begin{aligned} \mathrm{d}\mathcal{U}_t = {} & \langle \nabla U_t^{T,g}(X_t) - \nabla U_t^{T,g}(\hat{X}_t), \mathrm{d}\mathrm{e}_t \rangle \\ & + \langle \mathrm{e}_t, \mathrm{d}\big(\nabla U_t^{T,g}(X_t) - \nabla U_t^{T,g}(\hat{X}_t)\big) \rangle \\ & + \underbrace{0}_{\text{No covariation Itô}} \end{aligned}$$

Proof.

Itô+ angle process differential + Stochastic Maximum Principle □

$\mathcal{U}_t$ is a supermartingale

Recall that

$$\mathcal{U}_t = \langle \nabla U_t^{T,g}(X_t) - \nabla U_t^{T,g}(\hat{X}_t), \mathrm{e}_t \rangle$$

Lemma

$$\mathrm{d}\mathcal{U}_t = \underbrace{-|X_t - \hat{X}_t|^{-1} |\mathrm{proj}_{\mathrm{e}_t^{\perp}}(\nabla U_t^{T,g}(X_t) - \nabla U_t^{T,g}(\hat{X}_t))|^2}_{\text{angle process}} \,\mathrm{d}t$$

$$+ \underbrace{\langle e_t, \nabla^2 U_t^{T,g}(X_t) \cdot \mathrm{d}B_t + \nabla^2 U_t^{T,g}(\hat{X}_t) \cdot \mathrm{d}\hat{B}_t \rangle}_{\text{SMP}}$$

Proof.

Itô+ angle process differential + Stochastic Maximum Principle □

$\mathcal{U}_t$ is a supermartingale

Recall that

$$\mathcal{U}_t = \langle \nabla U_t^{T,g}(X_t) - \nabla U_t^{T,g}(\hat{X}_t), \mathrm{e}_t \rangle$$

Lemma

$$\mathrm{d}\mathcal{U}_t = \underbrace{-|X_t - \hat{X}_t|^{-1} |\mathrm{proj}_{\mathrm{e}_t^\perp}(\nabla U_t^{T,g}(X_t) - \nabla U_t^{T,g}(\hat{X}_t))|^2}_{\leq 0 \quad \text{angle process}} \mathrm{d}t$$

$$+ \underbrace{\langle e_t, \nabla^2 U_t^{T,g}(X_t) \cdot \mathrm{d}B_t + \nabla^2 U_t^{T,g}(\hat{X}_t) \cdot \mathrm{d}\hat{B}_t \rangle}_{\mathrm{d}M_t^1 \quad \text{SMP}}$$

Proof.

Itô+ angle process differential + Stochastic Maximum Principle □

$\mathcal{U}_t$ is a supermartingale

Recall that

$$\mathcal{U}_t = \langle \nabla U_t^{T,g}(X_t) - \nabla U_t^{T,g}(\hat{X}_t), \mathrm{e}_t \rangle$$

Lemma

$$\mathrm{d}\mathcal{U}_t = \underbrace{-|X_t - \hat{X}_t|^{-1} |\mathrm{proj}_{\mathrm{e}_t^{\perp}}(\nabla U_t^{T,g}(X_t) - \nabla U_t^{T,g}(\hat{X}_t))|^2}_{\leq 0 \quad \text{angle process}} \mathrm{d}t$$

$$+ \underbrace{\langle e_t, \nabla^2 U_t^{T,g}(X_t) \cdot \mathrm{d}B_t + \nabla^2 U_t^{T,g}(\hat{X}_t) \cdot \mathrm{d}\hat{B}_t \rangle}_{\mathrm{d}M_t^1 \quad \text{SMP}}$$

Proof.

Itô+ angle process differential + Stochastic Maximum Principle □

The dynamics of $f_L(|X_t - \hat{X}_t|)$

Following

- Andreas Eberle. "Reflection couplings and contraction rates for diffusions". In: *Probability Theory and Related Fields* 166.3-4 (2016), pp. 851–886

$$\mathrm{d}f_L(|X_t - \hat{X}_t|) = \Big(- f_L'(|X_t - \hat{X}_t|)\,\mathcal{U}_t + 2f_L''(|X_t - \hat{X}_t|)\Big)\mathrm{d}t + \mathrm{d}M_t^2$$

By construction

$$2f_L'' = -f_L f_L'$$

Lemma

$$\mathrm{d}f_L(|X_t - \hat{X}_t|) = -f_L'(|X_t - \hat{X}_t|)\Big(\mathcal{U}_t + f_L(|X_t - \hat{X}_t|)\Big)\mathrm{d}t + \mathrm{d}M_t^2$$

The dynamics of $f_L(|X_t - \hat{X}_t|)$

Following

- Andreas Eberle. "Reflection couplings and contraction rates for diffusions". In: *Probability Theory and Related Fields* 166.3-4 (2016), pp. 851–886

$$\mathrm{d}f_L(|X_t - \hat{X}_t|) = \Big(- f_L'(|X_t - \hat{X}_t|)\,\mathcal{U}_t + 2f_L''(|X_t - \hat{X}_t|)\Big)\mathrm{d}t + \mathrm{d}M_t^2$$

By construction

$$2f_L'' = -f_L f_L'$$

Lemma

$$\mathrm{d}f_L(|X_t - \hat{X}_t|) = -f_L'(|X_t - \hat{X}_t|)\Big(\mathcal{U}_t + f_L(|X_t - \hat{X}_t|)\Big)\mathrm{d}t + \mathrm{d}M_t^2$$

Γ_t is a supermartingale.

$$\mathrm{d}(\mathcal{U}_t + f_L(|X_t - \hat{X}_t|)) \leq -f_L'(|X_t - \hat{X}_t|)\Big(\mathcal{U}_t + f_L(|X_t - \hat{X}_t|)\Big)\mathrm{d}t$$
$$+\mathrm{d}M_t^2 + \mathrm{d}M_t^1$$

And therefore

$$\mathrm{d}\Gamma_t \leq \Gamma_t(\mathrm{d}M_t^1 + \mathrm{d}M_t^2)$$

Propagation of weak convexity is a consequence of

$$\mathbb{E}[\Gamma_0] \geq \mathbb{E}[\Gamma_T]$$

- Giovanni Conforti. "Weak semiconvexity estimates for Schrödinger potentials and logarithmic Sobolev inequality for Schrödinger bridges". In: *preprint arXiv:2301.00083* (2022)

Invariant sets of weakly convex functions

Γ_t is a supermartingale.

$$\mathrm{d}(\mathcal{U}_t + f_L(|X_t - \hat{X}_t|)) \leq -f_L'(|X_t - \hat{X}_t|)\Big(\mathcal{U}_t + f_L(|X_t - \hat{X}_t|)\Big)\mathrm{d}t$$
$$+\mathrm{d}M_t^2 + \mathrm{d}M_t^1$$

And therefore

$$\mathrm{d}\Gamma_t \leq \Gamma_t(\mathrm{d}M_t^1 + \mathrm{d}M_t^2)$$

Propagation of weak convexity is a consequence of

$$\mathbb{E}[\Gamma_0] \geq \mathbb{E}[\Gamma_T]$$

- Giovanni Conforti. "Weak semiconvexity estimates for Schrödinger potentials and logarithmic Sobolev inequality for Schrödinger bridges". In: *preprint arXiv:2301.00083* (2022)

Invariant sets of weakly convex functions

Γ_t is a supermartingale.

$$\mathrm{d}(\mathcal{U}_t + f_L(|X_t - \hat{X}_t|)) \leq -f_L'(|X_t - \hat{X}_t|)\Big(\mathcal{U}_t + f_L(|X_t - \hat{X}_t|)\Big)\mathrm{d}t$$
$$+\mathrm{d}M_t^2 + \mathrm{d}M_t^1$$

And therefore

$$\mathrm{d}\Gamma_t \leq \Gamma_t(\mathrm{d}M_t^1 + \mathrm{d}M_t^2)$$

Propagation of weak convexity is a consequence of

$$\mathbb{E}[\Gamma_0] \geq \mathbb{E}[\Gamma_T]$$

- Giovanni Conforti. "Weak semiconvexity estimates for Schrödinger potentials and logarithmic Sobolev inequality for Schrödinger bridges". In: *preprint arXiv:2301.00083* (2022)

Conclusion and perspectives

What is weak convexity propagation useful for?

Theoretical developments

- 🔧 Construction of Lipschitz transports between non log-concave probability measures
- ✓ Functional inequalities for Schrödinger bridges
- 🔧 Quantitative stability w.r.t. marginals
- 🔧 McKean-Vlasov dynamics for Schrödinger bridges

Quantitative convergence and error estimates for learning algorithms

- ✓ Sinkhorn's algorithm
- ? Mean field Sinkhorn algorithm
- 🔧 Score-based generative models

What is weak convexity propagation useful for?

Theoretical developments

- 🔧 Construction of Lipschitz transports between non log-concave probability measures
- ✓ Functional inequalities for Schrödinger bridges
- 🔧 Quantitative stability w.r.t. marginals
- 🔧 McKean-Vlasov dynamics for Schrödinger bridges

Quantitative convergence and error estimates for learning algorithms

- ✓ Sinkhorn's algorithm
- ? Mean field Sinkhorn algorithm
- 🔧 Score-based generative models

What is weak convexity propagation useful for?

Theoretical developments

- 🔧 Construction of Lipschitz transports between non log-concave probability measures
- ✓ Functional inequalities for Schrödinger bridges
- 🔧 Quantitative stability w.r.t. marginals
- 🔧 McKean-Vlasov dynamics for Schrödinger bridges

Quantitative convergence and error estimates for learning algorithms

- ✓ Sinkhorn's algorithm
- ? Mean field Sinkhorn algorithm
- 🔧 Score-based generative models

Thank you for the attention!

References

Giovanni Conforti. “Weak semiconvexity estimates for Schrödinger potentials and logarithmic Sobolev inequality for Schrödinger bridges”. In: *preprint arXiv:2301.00083* (2022).

Marco Cuturi. “Sinkhorn distances: Lightspeed computation of optimal transport”. In: *Advances in Neural Information Processing Systems*. 2013, pp. 2292–2300.

Andreas Eberle. “Reflection couplings and contraction rates for diffusions”. In: *Probability Theory and Related Fields* 166.3-4 (2016), pp. 851–886.

Max Fathi, Nathael Gozlan, and Maxime Prodhomme. “A proof of the Caffarelli contraction theorem via entropic regularization”. In: *Calculus of Variations and Partial Differential Equations* 59.96 (2020).

Erwin Schrödinger. “La théorie relativiste de l’électron et l’ interprétation de la mécanique quantique”. In: *Ann. Inst Henri Poincaré* 2 (1932), pp. 269–310.