Optimal transport and Fokker-Planck equations with monotone drifts

Giuseppe Savaré
http://www.imati.cnr.it/∼savare

Department of Mathematics, University of Pavia, Italy

Hammamet, *May 4, 2010*

Jointly with *Luca Natile* (Pavia) and *Mark A. Peletier* (Eindhoven)

Outline

Outline

Outline

Outline

Outline

Measure-valued solutions to Fokker-Planck equations

We consider nonnegative measure-valued solutions to the Fokker-Planck equation

$$\boxed{\partial_t \rho - \Delta \rho - \nabla \cdot (\rho B) = 0, \quad \rho\big|_{t=0} = \rho_0,} \tag{FP}$$

where $B : \mathbb{R}^d \to \mathbb{R}^d$ is a Borel $\boldsymbol{\lambda}$-monotone operator, $\boldsymbol{\lambda} \in \mathbb{R}$, i.e.

$$\langle B(x) - B(y), x - y \rangle \geq \boldsymbol{\lambda} \left| x - y \right|^2 \quad \text{for every } x, y \in \mathbb{R}^d.$$

Measure-valued solutions to Fokker-Planck equations

We consider nonnegative measure-valued solutions to the Fokker-Planck equation

$$\boxed{\partial_t \rho - \Delta \rho - \nabla \cdot (\rho B) = 0, \quad \rho\big|_{t=0} = \rho_0,} \tag{FP}$$

where $B : \mathbb{R}^d \to \mathbb{R}^d$ is a Borel $\boldsymbol{\lambda}$-monotone operator, $\boldsymbol{\lambda} \in \mathbb{R}$, i.e.

$$\langle B(x) - B(y), x - y \rangle \geq \boldsymbol{\lambda} \left| x - y \right|^2 \quad \text{for every } x, y \in \mathbb{R}^d.$$

$(\rho_t)_{t \geq 0} \subset \mathscr{P}(\mathbb{R}^d)$ is a *weakly continuous* family of probability measures satisfying the equation (FP) **in the sense of distributions**

$$\int_0^{+\infty} \int_{\mathbb{R}^d} \Big(\partial_t \zeta + \Delta \zeta - B \cdot \nabla \zeta \Big) \, \mathrm{d}\rho_t \mathrm{d}t = 0 \quad \forall \, \zeta \in C_{\mathrm{c}}^{\infty}(\mathbb{R}^d \times (0, +\infty)),$$

Measure-valued solutions to Fokker-Planck equations

We consider nonnegative measure-valued solutions to the Fokker-Planck equation

$$\boxed{\partial_t \rho - \Delta\rho - \nabla\cdot(\rho B) = 0, \quad \rho\big|_{t=0} = \rho_0,} \tag{FP}$$

where $B : \mathbb{R}^d \to \mathbb{R}^d$ is a Borel $\boldsymbol{\lambda}$-monotone operator, $\boldsymbol{\lambda} \in \mathbb{R}$, i.e.

$$\langle B(x) - B(y), x - y\rangle \geq \boldsymbol{\lambda} \left|x - y\right|^2 \quad \text{for every } x, y \in \mathbb{R}^d.$$

$(\rho_t)_{t\geq 0} \subset \mathscr{P}(\mathbb{R}^d)$ is a *weakly continuous* family of probability measures satisfying the equation (FP) **in the sense of distributions**

$$\int_0^{+\infty}\int_{\mathbb{R}^d}\Big(\partial_t\zeta + \Delta\zeta - B\cdot\nabla\zeta\Big)\,\mathrm{d}\rho_t\mathrm{d}t = 0 \quad \forall\,\zeta \in C_c^\infty(\mathbb{R}^d\times(0,+\infty)),$$

and the initial condition $\lim_{t\downarrow 0}\rho_t = \rho_0$ in $\mathscr{P}(\mathbb{R}^d)$.

Measure-valued solutions to Fokker-Planck equations

We consider nonnegative measure-valued solutions to the Fokker-Planck equation

$$\boxed{\partial_t \rho - \Delta\rho - \nabla\cdot(\rho B) = 0, \quad \rho\big|_{t=0} = \rho_0,} \tag{FP}$$

where $B : \mathbb{R}^d \to \mathbb{R}^d$ is a Borel $\boldsymbol{\lambda}$-monotone operator, $\boldsymbol{\lambda} \in \mathbb{R}$, i.e.

$$\langle B(x) - B(y), x - y\rangle \geq \boldsymbol{\lambda}\left|x-y\right|^2 \quad \text{for every } x, y \in \mathbb{R}^d.$$

$(\rho_t)_{t\geq 0} \subset \mathscr{P}(\mathbb{R}^d)$ is a *weakly continuous* family of probability measures satisfying the equation (FP) **in the sense of distributions**

$$\int_0^{+\infty}\int_{\mathbb{R}^d}\Big(\partial_t\zeta + \Delta\zeta - B\cdot\nabla\zeta\Big)\,\mathrm{d}\rho_t\mathrm{d}t = 0 \quad \forall\,\zeta \in C_c^\infty(\mathbb{R}^d \times (0,+\infty)),$$

and the initial condition $\lim_{t\downarrow 0}\rho_t = \rho_0$ in $\mathscr{P}(\mathbb{R}^d)$.

- $\boldsymbol{\lambda} = 0$: B is monotone.
  $\boldsymbol{\lambda} > 0$: B is *strongly* monotone.

Measure-valued solutions to Fokker-Planck equations

We consider nonnegative measure-valued solutions to the Fokker-Planck equation

$$\boxed{\partial_t \rho - \Delta\rho - \nabla\cdot(\rho B) = 0, \quad \rho\big|_{t=0} = \rho_0,} \tag{FP}$$

where $B : \mathbb{R}^d \to \mathbb{R}^d$ is a Borel $\boldsymbol{\lambda}$-monotone operator, $\boldsymbol{\lambda} \in \mathbb{R}$, i.e.

$$\langle B(x) - B(y), x - y\rangle \geq \boldsymbol{\lambda}\left|x-y\right|^2 \quad \text{for every } x, y \in \mathbb{R}^d.$$

$(\rho_t)_{t\geq 0} \subset \mathscr{P}(\mathbb{R}^d)$ is a *weakly continuous* family of probability measures satisfying the equation (FP) **in the sense of distributions**

$$\int_0^{+\infty}\int_{\mathbb{R}^d}\Big(\partial_t\zeta + \Delta\zeta - B\cdot\nabla\zeta\Big)\,\mathrm{d}\rho_t\mathrm{d}t = 0 \quad \forall\,\zeta \in C_c^\infty(\mathbb{R}^d \times (0,+\infty)),$$

and the initial condition $\lim_{t\downarrow 0}\rho_t = \rho_0$ in $\mathscr{P}(\mathbb{R}^d)$.

- $\boldsymbol{\lambda} = 0$: B is monotone.
  $\boldsymbol{\lambda} > 0$: B is *strongly* monotone.
  $\boldsymbol{\lambda} < 0$: B is a *Lipschitz perturbation* of a monotone operator.

Measure-valued solutions to Fokker-Planck equations

We consider nonnegative measure-valued solutions to the Fokker-Planck equation

$$\boxed{\partial_t \rho - \Delta\rho - \nabla\cdot(\rho B) = 0, \quad \rho\big|_{t=0} = \rho_0,} \tag{FP}$$

where $B : \mathbb{R}^d \to \mathbb{R}^d$ is a Borel $\boldsymbol{\lambda}$-monotone operator, $\boldsymbol{\lambda} \in \mathbb{R}$, i.e.

$$\langle B(x) - B(y), x - y\rangle \geq \boldsymbol{\lambda}\,\big|x - y\big|^2 \quad \text{for every } x, y \in \mathbb{R}^d.$$

$(\rho_t)_{t\geq 0} \subset \mathscr{P}(\mathbb{R}^d)$ is a *weakly continuous* family of probability measures satisfying the equation (FP) **in the sense of distributions**

$$\int_0^{+\infty}\int_{\mathbb{R}^d}\Big(\partial_t\zeta + \Delta\zeta - B\cdot\nabla\zeta\Big)\,\mathrm{d}\rho_t\mathrm{d}t = 0 \quad \forall\,\zeta \in C_c^\infty(\mathbb{R}^d \times (0,+\infty)),$$

and the initial condition $\lim_{t\downarrow 0}\rho_t = \rho_0$ in $\mathscr{P}(\mathbb{R}^d)$.

- $\boldsymbol{\lambda} = 0$: B is monotone.
  $\boldsymbol{\lambda} > 0$: B is *strongly* monotone.
  $\boldsymbol{\lambda} < 0$: B is a *Lipschitz perturbation* of a monotone operator.
  In the *gradient case* $B = \nabla V$:
  B is **$\boldsymbol{\lambda}$-monotone** iff V is **$\boldsymbol{\lambda}$-convex**, i.e. $\mathrm{D}^2 V \geq \boldsymbol{\lambda}\mathrm{I}$.

Measure-valued solutions to Fokker-Planck equations

We consider nonnegative measure-valued solutions to the Fokker-Planck equation

$$\boxed{\partial_t\rho - \Delta\rho - \nabla\cdot(\rho B) = 0, \quad \rho\big|_{t=0} = \rho_0,} \tag{FP}$$

where $B : \mathbb{R}^d \to \mathbb{R}^d$ is a Borel $\boldsymbol{\lambda}$-monotone operator, $\boldsymbol{\lambda} \in \mathbb{R}$, i.e.

$$\langle B(x) - B(y), x - y\rangle \geq \boldsymbol{\lambda}\,\big|x - y\big|^2 \quad \text{for every } x, y \in \mathbb{R}^d.$$

$(\rho_t)_{t\geq 0} \subset \mathscr{P}(\mathbb{R}^d)$ is a *weakly continuous* family of probability measures satisfying the equation (FP) **in the sense of distributions**

$$\int_0^{+\infty}\int_{\mathbb{R}^d}\Big(\partial_t\zeta + \Delta\zeta - B\cdot\nabla\zeta\Big)\,\mathrm{d}\rho_t\mathrm{d}t = 0 \quad \forall\,\zeta \in C_c^\infty(\mathbb{R}^d \times (0,+\infty)),$$

and the initial condition $\lim_{t\downarrow 0}\rho_t = \rho_0$ in $\mathscr{P}(\mathbb{R}^d)$.

- $\boldsymbol{\lambda} = 0$: B is monotone.
  $\boldsymbol{\lambda} > 0$: B is *strongly* monotone.
  $\boldsymbol{\lambda} < 0$: B is a *Lipschitz perturbation* of a monotone operator.
  In the *gradient case* $B = \nabla V$:
  B is **$\boldsymbol{\lambda}$-monotone** iff V is **$\boldsymbol{\lambda}$-convex**, i.e. $\mathrm{D}^2 V \geq \boldsymbol{\lambda}\mathrm{I}$.
- B is locally bounded

Measure-valued solutions to Fokker-Planck equations

We consider nonnegative measure-valued solutions to the Fokker-Planck equation

$$\boxed{\partial_t\rho - \Delta\rho - \nabla\cdot(\rho B) = 0, \quad \rho\big|_{t=0} = \rho_0,} \tag{FP}$$

where $B:\mathbb{R}^d\to\mathbb{R}^d$ is a Borel $\boldsymbol{\lambda}$-monotone operator, $\boldsymbol{\lambda}\in\mathbb{R}$, i.e.

$$\langle B(x)-B(y), x-y\rangle \geq \boldsymbol{\lambda}\left|x-y\right|^2 \quad \text{for every } x,y\in\mathbb{R}^d.$$

$(\rho_t)_{t\geq 0}\subset\mathscr{P}(\mathbb{R}^d)$ is a *weakly continuous* family of probability measures satisfying the equation (FP) **in the sense of distributions**

$$\int_0^{+\infty}\int_{\mathbb{R}^d}\Big(\partial_t\zeta + \Delta\zeta - B\cdot\nabla\zeta\Big)\,\mathrm{d}\rho_t\mathrm{d}t = 0 \quad \forall\,\zeta\in C_c^\infty(\mathbb{R}^d\times(0,+\infty)),$$

and the initial condition $\lim_{t\downarrow 0}\rho_t = \rho_0$ in $\mathscr{P}(\mathbb{R}^d)$.

- $\boldsymbol{\lambda}=0$: B is monotone.
  $\boldsymbol{\lambda}>0$: B is *strongly* monotone.
  $\boldsymbol{\lambda}<0$: B is a *Lipschitz perturbation* of a monotone operator.
  In the *gradient case* $B=\nabla V$:
  B is **$\boldsymbol{\lambda}$-monotone** iff V is **$\boldsymbol{\lambda}$-convex**, i.e. $\mathrm{D}^2V\geq\boldsymbol{\lambda}\mathrm{I}$.
- B is locally bounded
- No growth restrictions are imposed on B.

References

Equations of this type are the subject of several papers in *a very general situation*

- the Laplacian is replaced by a second order elliptic operator with variable coefficients
- B is just locally bounded with some control on $\langle B(x), x\rangle$.

References

Equations of this type are the subject of several papers in *a very general situation*

- the Laplacian is replaced by a second order elliptic operator with variable coefficients
- B is just locally bounded with some control on $\langle B(x), x \rangle$.

Existence Bogachev-DaPrato-Rockner, 2008
On parabolic equations for measures,
Commun. Partial Differ. Equations

Uniqueness Bogachev-DaPrato-Rockner-Stannat, 2007
Uniqueness of solutions to weak parabolic equations for measures,
Bulletin of the London Mathematical Society

Regularity Bogachev-Krylov-Rockner, 2001
On regularity of transition probabilities and invariant measures of singular diffusions under minimal conditions.,
Commun. Partial Differ. Equations.

References

Equations of this type are the subject of several papers in *a very general situation*

- the Laplacian is replaced by a second order elliptic operator with variable coefficients
- B is just locally bounded with some control on $\langle B(x), x\rangle$.

Existence Bogachev-DaPrato-Rockner, 2008
On parabolic equations for measures,
Commun. Partial Differ. Equations

Uniqueness Bogachev-DaPrato-Rockner-Stannat, 2007
Uniqueness of solutions to weak parabolic equations for measures,
Bulletin of the London Mathematical Society

Regularity Bogachev-Krylov-Rockner, 2001
On regularity of transition probabilities and invariant measures of singular diffusions under minimal conditions.,
Commun. Partial Differ. Equations.

In particular, it has been shown that ρ_t **is absolutely continuous** with respect to the Lebesgue measure for $\mathscr{L}^1$-a.e. t.

Main problem

Here we want to study the stability properties of the distributional solutions of

$$\partial_t \rho - \Delta\rho - \nabla \cdot (\rho B) = 0 \qquad \text{(FP)}$$

in terms of the *transportation costs*

$$\mathcal{C}_h(\boldsymbol{\rho}^1, \boldsymbol{\rho}^2) := \inf \left\{ \int_{\mathbb{R}^d \times \mathbb{R}^d} h(|\mathbf{x}_1 - \mathbf{x}_2|) \, \mathrm{d}\boldsymbol{\rho}(\mathbf{x}_1, \mathbf{x}_2) : \boldsymbol{\rho} \in \mathscr{P}(\mathbb{R}^d \times \mathbb{R}^d), \right.$$
$$\left. \boldsymbol{\rho} \text{ is a coupling between } \boldsymbol{\rho}^1 \text{ and } \boldsymbol{\rho}^2 \right\}.$$

Main problem

Here we want to study the stability properties of the distributional solutions of

$$\partial_t \rho - \Delta \rho - \nabla \cdot (\rho B) = 0 \tag{FP}$$

in terms of the *transportation costs*

$$\mathcal{C}_h(\boldsymbol{\rho}^1, \boldsymbol{\rho}^2) := \inf \Big\{ \int_{\mathbb{R}^d \times \mathbb{R}^d} h(|\mathbf{x}_1 - \mathbf{x}_2|) \,\mathrm{d}\boldsymbol{\rho}(\mathbf{x}_1, \mathbf{x}_2) : \boldsymbol{\rho} \in \mathscr{P}(\mathbb{R}^d \times \mathbb{R}^d), \\ \boldsymbol{\rho} \text{ is a coupling between } \boldsymbol{\rho}^1 \text{ and } \boldsymbol{\rho}^2 \Big\}.$$

where

$h : [0, +\infty) \to [0, +\infty)$ *is a continuous and* **non-decreasing** *function.*

Main problem

Here we want to study the stability properties of the distributional solutions of

$$\partial_t \rho - \Delta\rho - \nabla\cdot(\rho B) = 0 \tag{FP}$$

in terms of the *transportation costs*

$$\mathcal{C}_h(\boldsymbol{\rho}^1, \boldsymbol{\rho}^2) := \inf\Big\{ \int_{\mathbb{R}^d\times\mathbb{R}^d} h(|\mathbf{x}_1 - \mathbf{x}_2|)\,\mathrm{d}\boldsymbol{\rho}(\mathbf{x}_1, \mathbf{x}_2) : \boldsymbol{\rho} \in \mathscr{P}(\mathbb{R}^d \times \mathbb{R}^d),$$
$$\boldsymbol{\rho} \text{ is a coupling between } \boldsymbol{\rho}^1 \text{ and } \boldsymbol{\rho}^2 \Big\}.$$

where

$h : [0,+\infty) \to [0,+\infty)$ *is a continuous and* **non-decreasing** *function.*

Problem

If $\boldsymbol{\rho}^1, \boldsymbol{\rho}^2$ are two distributional solutions to (FP) find suitable estimates of

$$\mathcal{C}_h(\boldsymbol{\rho}^1_{\mathbf{t}}, \boldsymbol{\rho}^2_{\mathbf{t}}) \textit{ in terms of } \mathcal{C}_h(\boldsymbol{\rho}^1_{\mathbf{0}}, \boldsymbol{\rho}^2_{\mathbf{0}}).$$

Outline

The Wasserstein approach to Fokker-Planck in the gradient case

When

$B = \nabla V$ is the **gradient** of a **λ**-convex function $V : \mathbb{R}^d \to \mathbb{R}$

then (FP) can be considered as the **gradient flow** of the relative entropy functional

$$\begin{aligned} \mathrm{Ent}_\gamma(\rho) &:= \int_{\mathbb{R}^d} u(x) \log u(x)\,\mathrm{d}x + \int_{\mathbb{R}^d} V(x)\,\mathrm{d}\rho(x) && \rho = u\mathscr{L}^d \\ &= \int_{\mathbb{R}^d} \frac{\mathrm{d}\rho}{\mathrm{d}\gamma} \log\Big(\frac{\mathrm{d}\rho}{\mathrm{d}\gamma}\Big)\,\mathrm{d}\gamma && \gamma := \mathrm{e}^{-V}\mathscr{L}^d \end{aligned}$$

The Wasserstein approach to Fokker-Planck in the gradient case

When

$$\boxed{B = \nabla V \text{ is the **gradient** of a } \lambda\text{-convex function } V : \mathbb{R}^d \to \mathbb{R}}$$

then (FP) can be considered as the **gradient flow** of the relative entropy functional

$$\begin{aligned} \mathrm{Ent}_\gamma(\rho) &:= \int_{\mathbb{R}^d} u(x) \log u(x)\,\mathrm{d}x + \int_{\mathbb{R}^d} V(x)\,\mathrm{d}\rho(x) \qquad && \rho = u\mathscr{L}^d \\ &= \int_{\mathbb{R}^d} \frac{\mathrm{d}\rho}{\mathrm{d}\gamma} \log\Big(\frac{\mathrm{d}\rho}{\mathrm{d}\gamma}\Big)\,\mathrm{d}\gamma && \gamma := \mathrm{e}^{-V}\mathscr{L}^d \end{aligned}$$

in the space $\mathscr{P}_2(\mathbb{R}^d)$ of probability measures with finite quadratic moments endowed with the so called L^2-Wasserstein distance $\mathrm{W}_2(\cdot,\cdot)$

$$\mathrm{W}_2^2(\rho^1,\rho^2) := \mathcal{C}_h(\rho^1,\rho^2), \quad h(r) := r^2.$$

The Wasserstein approach to Fokker-Planck in the gradient case

When

$B = \nabla V$ is the **gradient** of a λ-convex function $V : \mathbb{R}^d \to \mathbb{R}$

then (FP) can be considered as the **gradient flow** of the relative entropy functional

$$\begin{aligned}\mathrm{Ent}_\gamma(\rho) &:= \int_{\mathbb{R}^d} u(x) \log u(x)\,\mathrm{d}x + \int_{\mathbb{R}^d} V(x)\,\mathrm{d}\rho(x) \qquad && \rho = u\mathscr{L}^d \\ &= \int_{\mathbb{R}^d} \frac{\mathrm{d}\rho}{\mathrm{d}\gamma} \log\Big(\frac{\mathrm{d}\rho}{\mathrm{d}\gamma}\Big)\,\mathrm{d}\gamma && \gamma := \mathrm{e}^{-V}\mathscr{L}^d\end{aligned}$$

in the space $\mathscr{P}_2(\mathbb{R}^d)$ of probability measures with finite quadratic moments endowed with the so called L^2-Wasserstein distance $\mathrm{W}_2(\cdot,\cdot)$

$$\mathrm{W}_2^2(\rho^1,\rho^2) := \mathcal{C}_h(\rho^1,\rho^2), \quad h(r) := r^2.$$

This remarkable interpretation found by Jordan-Kinderlehrer-Otto (1998) gave rise to a series of studies on the relationships between certain classes of diffusion equations and distances between probability measures induced by optimal transport problems.
General overviews can be found in the books Villani 2003-2009, Ambrosio-Gigli-S. 2005.

Wasserstein distance (I): the point of view of optimal transport

Wasserstein space:

$$\mathscr{P}_2(\mathbb{R}^d) := \Big\{\rho \text{ Borel probability measures on } \mathbb{R}^d \text{ s.t. } \int_{\mathbb{R}^d} |x|^2 \,\mathrm{d}\rho(x) < +\infty\Big\}.$$

Wasserstein distance (I): the point of view of optimal transport

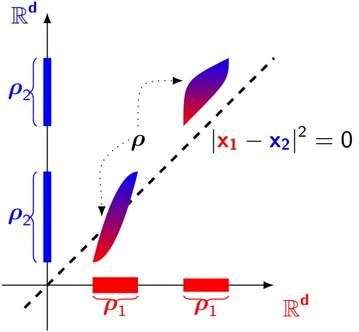
Wasserstein space:

$$\mathscr{P}_2(\mathbb{R}^d) := \Big\{\rho \text{ Borel probability measures on } \mathbb{R}^d \text{ s.t. } \int_{\mathbb{R}^d} |x|^2 \,\mathrm{d}\rho(x) < +\infty\Big\}.$$

Couplings between $\rho_1, \rho_2 \in \mathscr{P}_2(\mathbb{R}^d)$: measures $\rho \in \mathscr{P}_2(\mathbb{R}^d \times \mathbb{R}^d)$ whose marginals are ρ_1, ρ_2, i.e.

$$\rho(A \times \mathbb{R}^d) = \rho_1(A), \quad \rho(\mathbb{R}^d \times B) = \rho_2(B)$$

for every Borel sets $A, B \subset \mathbb{R}^d$.
$\Gamma(\rho_1, \rho_2)$ is the collection of all couplings of ρ_1 and ρ_2.



Wasserstein distance (I): the point of view of optimal transport

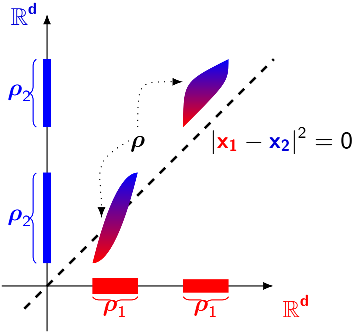
Wasserstein space:

$$\mathscr{P}_2(\mathbb{R}^d) := \Big\{\rho \text{ Borel probability measures on } \mathbb{R}^d \text{ s.t. } \int_{\mathbb{R}^d} |x|^2 \, \mathrm{d}\rho(x) < +\infty\Big\}.$$

Couplings between $\boldsymbol{\rho_1}, \boldsymbol{\rho_2} \in \mathscr{P}_2(\mathbb{R}^d)$: measures $\boldsymbol{\rho} \in \mathscr{P}_2(\mathbb{R}^{\mathbf{d}} \times \mathbb{R}^{\mathbf{d}})$ whose marginals are $\boldsymbol{\rho_1}, \boldsymbol{\rho_2}$, i.e.

$$\boldsymbol{\rho}(\mathbf{A}\times\mathbb{R}^d) = \boldsymbol{\rho_1}(\mathbf{A}), \quad \boldsymbol{\rho}(\mathbb{R}^d\times\mathbf{B}) = \boldsymbol{\rho_2}(\mathbf{B})$$

for every Borel sets $\mathbf{A}, \mathbf{B} \subset \mathbb{R}^d$.
$\Gamma(\boldsymbol{\rho_1}, \boldsymbol{\rho_2})$ is the collection of all couplings of $\boldsymbol{\rho_1}$ and $\boldsymbol{\rho_2}$.



Wasserstein distance:

$$\boxed{\mathrm{W}_2^2(\boldsymbol{\rho_1}, \boldsymbol{\rho_2}) := \min\Big\{\int_{\mathbb{R}^{\mathbf{d}}\times\mathbb{R}^{\mathbf{d}}} |\mathbf{x_1} - \mathbf{x_2}|^2 \, \mathrm{d}\boldsymbol{\rho}(\mathbf{x_1}, \mathbf{x_2}) : \boldsymbol{\rho} \in \Gamma(\boldsymbol{\rho_1}, \boldsymbol{\rho_2})\Big\}}$$

“Soft” properties

- $\big(\mathscr{P}_2(\mathbb{R}^d), \mathrm{W}_2\big)$ is a **complete and separable** metric space.

"Soft" properties

- $(\mathscr{P}_2(\mathbb{R}^d), W_2)$ is a **complete and separable** metric space.
- Convergence with respect to W_2 $\Leftrightarrow$ **Weak** convergence + convergence of the **quadratic moments.**

"Soft" properties

- $(\mathscr{P}_2(\mathbb{R}^d), \mathrm{W}_2)$ is a **complete and separable** metric space.
- Convergence with respect to W_2 $\Leftrightarrow$ **Weak** convergence + convergence of the **quadratic moments.**
- **Convexity** (but linear segments are not geodesics!)

"Soft" properties

- $(\mathscr{P}_2(\mathbb{R}^d), \mathrm{W}_2)$ is a **complete and separable** metric space.
- Convergence with respect to W_2 $\Leftrightarrow$ **Weak** convergence + convergence of the **quadratic moments.**
- **Convexity** (but linear segments are not geodesics!)
- Existence of (constant speed, minimizing) **geodesics** connecting arbitrary measures $\rho_0, \rho_1 \in \mathscr{P}_2(\mathbb{R}^d)$: they are curves $\rho : s \in [0,1] \mapsto \rho_s$ s.t.

$$\mathrm{W}_2(\rho_0, \rho_1) = \mathcal{L}_0^1[\rho], \quad \mathrm{W}_2(\rho_{s_0}, \rho_{s_1}) = |s_1 - s_0|\, \mathrm{W}_2(\rho_0, \rho_1), \quad 0 \le s_0 \le s_1 \le 1.$$

Wasserstein distance (II): the Brenier dynamical approach

In $\mathbb{R}^d$ consider a moving family $\rho_s \in \mathscr{P}_2(\mathbb{R}^d)$, $s \in [0,1]$, satisfying the **continuity equation**

$$\boxed{\partial_s \rho + \operatorname{div}(\rho\, \mathbf{v}) = 0; \quad \mathbf{v} \text{ is the } \textbf{velocity vector} \text{ associated to } \rho}$$

Wasserstein distance (II): the Brenier dynamical approach

In $\mathbb{R}^d$ consider a moving family $\rho_s \in \mathscr{P}_2(\mathbb{R}^d)$, $s \in [0,1]$, satisfying the **continuity equation**

$$\partial_s \rho + \operatorname{div}(\rho\, \mathbf{v}) = 0; \quad \mathbf{v} \text{ is the velocity vector associated to } \rho$$

The **scalar velocity** at time s is given by

$$\mathcal{V}_s[\rho_s] := \|\mathbf{v_s}\|_{L^2(\rho_s;\mathbb{R}^d)} = \Big(\int_{\mathbb{R}^d} |\mathbf{v_s(x)}|^2 \,\mathrm{d}\rho_s(x)\Big)^{1/2}$$

Wasserstein distance (II): the Brenier dynamical approach

In $\mathbb{R}^d$ consider a moving family $\rho_s \in \mathscr{P}_2(\mathbb{R}^d)$, $s \in [0,1]$, satisfying the **continuity equation**

$$\boxed{\partial_s \rho + \operatorname{div}(\rho\, \mathbf{v}) = 0; \quad \mathbf{v} \text{ is the velocity vector associated to } \rho}$$

The **scalar velocity** at time s is given by

$$\mathcal{V}_s[\rho_s] := \|\mathbf{v_s}\|_{L^2(\rho_s;\mathbb{R}^d)} = \Big(\int_{\mathbb{R}^d} |\mathbf{v_s(x)}|^2 \,\mathrm{d}\rho_s(x)\Big)^{1/2}$$

The length of the curve ρ between s_0 and s_1

$$\mathcal{L}_{s_0}^{s_1}[\rho] := \int_{s_0}^{s_1} \mathcal{V}_s[\rho]\,\mathrm{d}s = \int_{s_0}^{s_1} \Big(\int_{\mathbb{R}^d} |\mathbf{v_s(x)}|^2 \,\mathrm{d}\rho_s(x)\Big)^{1/2} \mathrm{d}s$$

Wasserstein distance (II): the Brenier dynamical approach

In $\mathbb{R}^d$ consider a moving family $\rho_s \in \mathscr{P}_2(\mathbb{R}^d)$, $s \in [0,1]$, satisfying the **continuity equation**

$$\partial_s \rho + \operatorname{div}(\rho\,\mathbf{v}) = 0; \quad \mathbf{v} \text{ is the } \textbf{velocity vector} \text{ associated to } \rho$$

The **scalar velocity** at time s is given by

$$\mathcal{V}_s[\rho_s] := \|\mathbf{v_s}\|_{L^2(\rho_s;\mathbb{R}^d)} = \Big(\int_{\mathbb{R}^d} |\mathbf{v_s(x)}|^2 \,\mathrm{d}\rho_s(x)\Big)^{1/2}$$

The length of the curve ρ between s_0 and s_1

$$\mathcal{L}_{s_0}^{s_1}[\rho] := \int_{s_0}^{s_1} \mathcal{V}_s[\rho]\,\mathrm{d}s = \int_{s_0}^{s_1} \Big(\int_{\mathbb{R}^d} |\mathbf{v_s(x)}|^2 \,\mathrm{d}\rho_s(x)\Big)^{1/2} \mathrm{d}s$$

Wasserstein distance W_2 between ρ_0 and ρ_1:

$$\mathrm{W}_2(\rho_0,\rho_1) := \min\Big\{\mathcal{L}_0^1[\rho] : \rho\big|_{s=0} = \rho_0,\ \rho\big|_{s=1} = \rho_1\Big\}.$$

Otto's approach to Gradient flows through Wasserstein distance

$$\begin{cases} \partial_t \rho + \operatorname{div}\left(\rho\, \mathbf{v}\right) = 0 & \textbf{(Continuity equation)} \\ \mathbf{v} = -\mathrm{D}\left(\dfrac{\boldsymbol{\delta\Phi}}{\boldsymbol{\delta\rho}}\right) & \textbf{(Nonlinear condition)} \\ \rho\big|_{t=0} = \rho_0 & \rho_0 \in \mathscr{P}_2(\mathbb{R}^d). \end{cases}$$

Otto's approach to Gradient flows through Wasserstein distance

$$\begin{cases} \partial_t \rho + \operatorname{div}(\rho\, \mathbf{v}) = 0 & \textbf{(Continuity equation)} \\ \mathbf{v} = -\mathrm{D}\Big(\dfrac{\boldsymbol{\delta\Phi}}{\boldsymbol{\delta\rho}}\Big) & \textbf{(Nonlinear condition)} \\ \rho\big|_{t=0} = \rho_0 & \rho_0 \in \mathscr{P}_2(\mathbb{R}^d). \end{cases}$$

Here $\Phi : \mathscr{P}_2(\mathbb{R}^d) \to \mathbb{R}$ is a suitable integral functional and $\dfrac{\boldsymbol{\delta\Phi}}{\boldsymbol{\delta\rho}}$ is its Euler-Lagrange first variation

$$\Phi(\rho) := \int L(x, u, \mathrm{D}u)\, dx, \quad \frac{\boldsymbol{\delta\Phi}}{\boldsymbol{\delta\rho}} = \mathbf{L_u(x, u, Du) - div\, L_{Du}(x, u, Du)}, \quad \rho = u\mathscr{L}^d.$$

Otto's approach to Gradient flows through Wasserstein distance

$$
\begin{cases}
\partial_t \rho + \operatorname{div}\left(\rho\, \mathbf{v}\right) = 0 & \textbf{(Continuity equation)} \\
\mathbf{v} = -\mathrm{D}\left(\dfrac{\delta \Phi}{\delta \rho}\right) & \textbf{(Nonlinear condition)} \\
\rho\big|_{t=0} = \rho_0 & \rho_0 \in \mathscr{P}_2(\mathbb{R}^d).
\end{cases}
$$

Here $\Phi : \mathscr{P}_2(\mathbb{R}^d) \to \mathbb{R}$ is a suitable integral functional and $\dfrac{\delta \Phi}{\delta \rho}$ is its Euler-Lagrange first variation

$$
\Phi(\rho) := \int L(x, u, \mathrm{D}u)\, dx, \quad \frac{\delta \Phi}{\delta \rho} = \mathbf{L_u(x, u, Du) - \operatorname{div} L_{Du}(x, u, Du)}, \quad \rho = u\mathscr{L}^d.
$$

In the case of the **relative entropy functional**

$$
\Phi(\rho) := \mathrm{Ent}_\gamma(\rho) = \int u \log u \,\mathrm{d}x + \int V(x)\,\mathrm{d}\rho, \qquad \frac{\delta \Phi}{\delta \rho} = \log u + V(x) + 1
$$

$$
\partial_t \rho = \operatorname{div}\left(\rho \mathrm{D}\,(\mathbf{\log u + V})\right) = \Delta \rho + \operatorname{div}(\rho \mathrm{D} V).
$$

Otto's approach to Gradient flows through Wasserstein distance

$$\begin{cases} \partial_t\rho + \mathrm{div}\,(\rho\,\mathbf{v}) = 0 & \textbf{(Continuity equation)} \\ \mathbf{v} = -\mathrm{D}\Big(\dfrac{\delta\Phi}{\delta\rho}\Big) & \textbf{(Nonlinear condition)} \\ \rho|_{t=0} = \rho_0 & \rho_0 \in \mathscr{P}_2(\mathbb{R}^d). \end{cases}$$

Here $\Phi : \mathscr{P}_2(\mathbb{R}^d) \to \mathbb{R}$ is a suitable integral functional and $\dfrac{\delta\Phi}{\delta\rho}$ is its Euler-Lagrange first variation

$$\Phi(\rho) := \int L(x,u,\mathrm{D}u)\,dx, \quad \frac{\delta\Phi}{\delta\rho} = \mathbf{L_u(x,u,\mathrm{D}u) - \mathrm{div}\,L_{\mathrm{D}u}(x,u,\mathrm{D}u)}, \quad \rho = u\mathscr{L}^d.$$

In the case of the **relative entropy functional**

$$\Phi(\rho) := \mathrm{Ent}_\gamma(\rho) = \int u\log u\,\mathrm{d}x + \int V(x)\,\mathrm{d}\rho, \qquad \frac{\delta\Phi}{\delta\rho} = \log u + V(x) + 1$$

$$\partial_t\rho = \mathrm{div}\,\big(\rho\mathrm{D}\,(\mathbf{\log u + V})\big) = \Delta\rho + \mathrm{div}(\rho\mathrm{D}V).$$

[Otto, with Jordan and Kinderlehrer '97-'00] showed in many interesting cases that such kind of equations can be interpreted as

> the **"gradient flow"** of Φ with respect to the **Wasserstein distance** W_2 in $\mathscr{P}_2(\mathbb{R}^d)$.

A new geometric insight

One of the strengths of the Wasserstein approach is a new geometric insight (developed by OTTO, 2001) in the evolution process:

A new geometric insight

One of the strengths of the Wasserstein approach is a new geometric insight (developed by Otto, 2001) in the evolution process: in the case of (FP)

> *the λ-convexity of the potential V reflects a λ-convexity property (also called* **displacement λ-convexity***) of the functional Ent_γ along the geodesics of $\mathscr{P}_2(\mathbb{R}^d)$.*

A new geometric insight

One of the strengths of the Wasserstein approach is a new geometric insight (developed by OTTO, 2001) in the evolution process: in the case of (FP)

> *the λ-convexity of the potential V reflects a λ-convexity property (also called* **displacement λ-convexity***) of the functional Ent_γ along the geodesics of $\mathscr{P}_2(\mathbb{R}^d)$.*

This nice feature, discovered by R.J. MCCANN, 1997, suggests that

> *one can adapt some typical basic existence, approximation, and regularity results for* **gradient flows of convex functionals in Euclidean spaces** *to the measure-theoretic setting of $\mathscr{P}_2(\mathbb{R}^d)$.*

A new geometric insight

One of the strengths of the Wasserstein approach is a new geometric insight (developed by OTTO, 2001) in the evolution process: in the case of (FP)

> *the λ-convexity of the potential V reflects a λ-convexity property (also called* **displacement λ-convexity***) of the functional Ent_γ along the geodesics of $\mathscr{P}_2(\mathbb{R}^d)$.*

This nice feature, discovered by R.J. MCCANN, 1997, suggests that

> *one can adapt some typical basic existence, approximation, and regularity results for* **gradient flows of convex functionals in Euclidean spaces** *to the measure-theoretic setting of $\mathscr{P}_2(\mathbb{R}^d)$.*

This program has been carried out and among the most interesting estimates, it provides the λ-contraction property

$$\boxed{\mathrm{W}_2(\rho^1_t, \rho^2_t) \le \mathrm{e}^{-\lambda t}\, \mathrm{W}_2(\rho^1_0, \rho^2_0) \qquad \text{for every } t \ge 0,}$$

where ρ^1, ρ^2 are the solutions to Fokker-Planck starting from the initial data $\rho^1_0, \rho^2_0 \in \mathscr{P}_2(\mathbb{R}^d)$.

A new geometric insight

One of the strengths of the Wasserstein approach is a new geometric insight (developed by Otto, 2001) in the evolution process: in the case of (FP)

> *the λ-convexity of the potential V reflects a λ-convexity property (also called* **displacement λ-convexity***) of the functional Ent_γ along the geodesics of $\mathscr{P}_2(\mathbb{R}^d)$.*

This nice feature, discovered by R.J. McCann, 1997, suggests that

> *one can adapt some typical basic existence, approximation, and regularity results for* **gradient flows of convex functionals in Euclidean spaces** *to the measure-theoretic setting of $\mathscr{P}_2(\mathbb{R}^d)$.*

This program has been carried out and among the most interesting estimates, it provides the λ-contraction property

$$\boxed{\mathrm{W}_2(\rho^1_t, \rho^2_t) \le \mathrm{e}^{-\lambda t}\, \mathrm{W}_2(\rho^1_0, \rho^2_0) \qquad \text{for every } t \ge 0,}$$

where ρ^1, ρ^2 are the solutions to Fokker-Planck starting from the initial data $\rho^1_0, \rho^2_0 \in \mathscr{P}_2(\mathbb{R}^d)$.

Basic ingredients:

1. Derivative of the **squared distance function** along locally Lipschitz curves.

A new geometric insight

One of the strengths of the Wasserstein approach is a new geometric insight (developed by Otto, 2001) in the evolution process: in the case of (FP)

> *the λ-convexity of the potential V reflects a λ-convexity property (also called* **displacement λ-convexity***) of the functional Ent_γ along the geodesics of $\mathscr{P}_2(\mathbb{R}^d)$.*

This nice feature, discovered by R.J. McCann, 1997, suggests that

> *one can adapt some typical basic existence, approximation, and regularity results for* **gradient flows of convex functionals in Euclidean spaces** *to the measure-theoretic setting of $\mathscr{P}_2(\mathbb{R}^d)$.*

This program has been carried out and among the most interesting estimates, it provides the λ-contraction property

$$\mathrm{W}_2(\rho^1_t, \rho^2_t) \le e^{-\lambda t}\, \mathrm{W}_2(\rho^1_0, \rho^2_0) \qquad \text{for every } t \ge 0,$$

where ρ^1, ρ^2 are the solutions to Fokker-Planck starting from the initial data $\rho^1_0, \rho^2_0 \in \mathscr{P}_2(\mathbb{R}^d)$.

Basic ingredients:

1. Derivative of the **squared distance function** along locally Lipschitz curves.
2. **Displacement convexity** of the relative entropy functional Ent_γ.

A new geometric insight

One of the strengths of the Wasserstein approach is a new geometric insight (developed by OTTO, 2001) in the evolution process: in the case of (FP)

> *the λ-convexity of the potential V reflects a λ-convexity property (also called* **displacement λ-convexity***) of the functional Ent_γ along the geodesics of $\mathscr{P}_2(\mathbb{R}^d)$.*

This nice feature, discovered by R.J. MCCANN, 1997, suggests that

> *one can adapt some typical basic existence, approximation, and regularity results for* **gradient flows of convex functionals in Euclidean spaces** *to the measure-theoretic setting of $\mathscr{P}_2(\mathbb{R}^d)$.*

This program has been carried out and among the most interesting estimates, it provides the λ-contraction property

$$\boxed{\mathrm{W}_2(\rho^1_t, \rho^2_t) \le \mathrm{e}^{-\lambda t}\, \mathrm{W}_2(\rho^1_0, \rho^2_0) \qquad \text{for every } t \ge 0,}$$

where ρ^1, ρ^2 are the solutions to Fokker-Planck starting from the initial data $\rho^1_0, \rho^2_0 \in \mathscr{P}_2(\mathbb{R}^d)$.

Basic ingredients:

1. Derivative of the **squared distance function** along locally Lipschitz curves.
2. **Displacement convexity** of the relative entropy functional Ent_γ.
3. **Subgradient inequality** for the logarithmic gradient $\nabla\rho/\rho + \nabla V$

Outline

Derivative of the squared distance function

The first tool is the formula which evaluates the derivative of the squared Wasserstein distance from a fixed measure σ along the locally Lipschitz curve ρ in $\mathscr{P}_2(\mathbb{R}^d)$.

Moreover

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{1}{2}W_2^2(\rho_t,\sigma) = \int_{\mathbb{R}^d\times\mathbb{R}^d} \langle \mathbf{v}_t(x), x-y\rangle \,\mathrm{d}\boldsymbol{\rho}_t(x,y) \quad \textit{for } \mathscr{L}^1\textit{-a.e. } t>0$$

where $\boldsymbol{\rho}_t$ is an optimal coupling between ρ_t and σ.

Derivative of the squared distance function

The first tool is the formula which evaluates the derivative of the squared Wasserstein distance from a fixed measure σ along the locally Lipschitz curve ρ in $\mathscr{P}_2(\mathbb{R}^d)$.

Theorem

Let $\rho : (0, +\infty) \to \mathscr{P}_2(\mathbb{R}^d)$ be a locally Lipschitz curve. Then for $\mathscr{L}^1$-a.e. $t > 0$ there exists a (ρ_t-essentially) unique Borel map $\mathbf{v}_t \in L^2(\rho_t; \mathbb{R}^d)$ such that

$$\partial_t \rho_t + \mathrm{div}(\rho_t \, \mathbf{v}_t) = 0 \quad \text{in } \mathbb{R}^d \times (0, +\infty)$$

and

$$\lim_{h \to 0} \frac{\mathrm{W}_2(\rho_{t+h}, \rho_t)}{|h|} = \Big(\int_{\mathbb{R}^d} |\mathbf{v}_t|^2 \, \mathrm{d}\rho_t \Big)^{1/2} \in L^\infty_{\mathrm{loc}}(0, +\infty).$$

$\mathbf{v}_t$ *belongs to* $\mathrm{Tan}_{\rho_t} \mathscr{P}_2(\mathbb{R}^d)$, *the closure in* $L^2(\rho_t; \mathbb{R}^d)$ *of the* **space generated by the gradients of $C^\infty_c(\mathbb{R}^d)$ functions.**

Derivative of the squared distance function

The first tool is the formula which evaluates the derivative of the squared Wasserstein distance from a fixed measure σ along the locally Lipschitz curve ρ in $\mathscr{P}_2(\mathbb{R}^d)$.

Theorem

Let $\rho : (0,+\infty) \to \mathscr{P}_2(\mathbb{R}^d)$ be a locally Lipschitz curve. Then for $\mathscr{L}^1$-a.e. $t > 0$ there exists a (ρ_t-essentially) unique Borel map $\mathbf{v}_t \in L^2(\rho_t;\mathbb{R}^d)$ such that

$$\partial_t \rho_t + \mathrm{div}(\rho_t\,\mathbf{v}_t) = 0 \quad \text{in } \mathbb{R}^d \times (0,+\infty)$$

and

$$\lim_{h\to 0} \frac{\mathrm{W}_2(\rho_{t+h},\rho_t)}{|h|} = \Big(\int_{\mathbb{R}^d} |\mathbf{v}_t|^2\,\mathrm{d}\rho_t\Big)^{1/2} \in L^\infty_{\mathrm{loc}}(0,+\infty).$$

$\mathbf{v}_t$ belongs to $\mathrm{Tan}_{\rho_t}\mathscr{P}_2(\mathbb{R}^d)$, the closure in $L^2(\rho_t;\mathbb{R}^d)$ of the **space generated by the gradients of $C^\infty_c(\mathbb{R}^d)$ functions.** *Moreover*

$$\boxed{\frac{\mathrm{d}}{\mathrm{d}t}\frac{1}{2}W_2^2(\rho_t,\sigma) = \int_{\mathbb{R}^d\times\mathbb{R}^d} \langle \mathbf{v}_t(x), x-y\rangle\,\mathrm{d}\boldsymbol{\rho}_t(x,y) \quad \text{for } \mathscr{L}^1\text{-a.e. } t>0}$$

where $\boldsymbol{\rho}_t$ is an optimal coupling between ρ_t and σ.

Displacement convexity

Definition (Displacement λ-convex, McCann '97)

A functional $\Phi : \mathscr{P}_2(\mathbb{R}^d) \to (-\infty, +\infty]$ is **displacement convex** if each couple of measures $\rho_0, \rho_1 \in D(\Phi)$ can be connected by a **geodesic** (i.e. displacement interpolation) $\rho_s,\ s \in [0,1]$ such that

$$\Phi(\rho_s) \leq (1-s)\Phi(\rho_0) + s\Phi(\rho_1) \qquad \forall\, s \in (0,1).$$

Displacement convexity

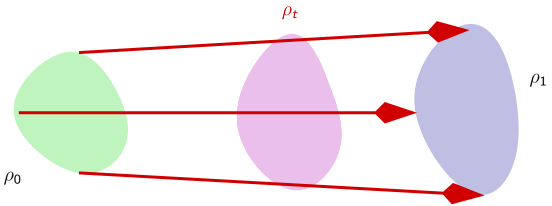
Definition (Displacement λ-convex, McCann '97)

A functional $\Phi : \mathscr{P}_2(\mathbb{R}^d) \to (-\infty, +\infty]$ is **displacement convex** if each couple of measures $\rho_0, \rho_1 \in D(\Phi)$ can be connected by a **geodesic** (i.e. displacement interpolation) $\rho_s,\ s \in [0,1]$ such that

$$\Phi(\rho_s) \leq (1-s)\Phi(\rho_0) + s\Phi(\rho_1) \qquad \forall\, s \in (0,1).$$

The geodesic connecting ρ_0 to ρ_1 can be represented as

$$\rho_s := \Big((1-s)\pi^1 + s\pi^2\Big)_{\#}\boldsymbol{\rho} \qquad \text{where } \boldsymbol{\rho} \text{ is an optimal coupling}$$



Displacement convexity

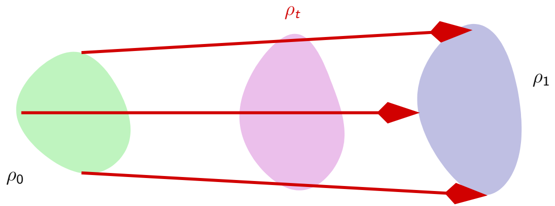
Definition (Displacement λ-convex, McCann '97)

A functional $\Phi : \mathscr{P}_2(\mathbb{R}^d) \to (-\infty, +\infty]$ is **$\boldsymbol{\lambda}$-displacement convex** if each couple of measures $\rho_0, \rho_1 \in D(\Phi)$ can be connected by a **geodesic** (i.e. displacement interpolation) $\rho_s,\ s \in [0,1]$ such that

$$\Phi(\rho_s) \le (1-s)\Phi(\rho_0) + s\Phi(\rho_1) - \frac{1}{2}\boldsymbol{\lambda}\, s(1-s)\mathrm{W}_2^2(\rho_0,\rho_1) \quad \forall\, s \in (0,1).$$

The geodesic connecting ρ_0 to ρ_1 can be represented as

$$\rho_s := \left((1-s)\pi^1 + s\pi^2\right)_{\#}\boldsymbol{\rho} \qquad \text{where } \boldsymbol{\rho} \text{ is an optimal coupling}$$



Displacement convexity of the relative entropy functional

Consider

$$\begin{aligned}\Phi(\rho) = \mathrm{Ent}_\gamma(\rho) &:= \int_{\mathbb{R}^d} u(x)\log u(x)\,\mathrm{d}x + \int_{\mathbb{R}^d} V(x)\,\mathrm{d}\rho(x) && \rho = u\mathscr{L}^d\\ &= \int_{\mathbb{R}^d} \frac{\mathrm{d}\rho}{\mathrm{d}\gamma}\log\left(\frac{\mathrm{d}\rho}{\mathrm{d}\gamma}\right)\mathrm{d}\gamma && \gamma := \mathrm{e}^{-V}\mathscr{L}^d\end{aligned}$$

Displacement convexity of the relative entropy functional

Consider

$$\Phi(\rho) = \mathrm{Ent}_\gamma(\rho) := \int_{\mathbb{R}^d} u(x) \log u(x) \,\mathrm{d}x + \int_{\mathbb{R}^d} V(x) \,\mathrm{d}\rho(x) \qquad \rho = u\mathscr{L}^d$$

$$= \int_{\mathbb{R}^d} \frac{\mathrm{d}\rho}{\mathrm{d}\gamma} \log\left(\frac{\mathrm{d}\rho}{\mathrm{d}\gamma}\right) \mathrm{d}\gamma \qquad \gamma := \mathrm{e}^{-V}\mathscr{L}^d$$

Theorem (McCann '97)

The relative entropy functional $\mathrm{Ent}_\gamma(\cdot)$ *is* $\boldsymbol{\lambda}$*-displacement convex iff*

the potential V **is** $\boldsymbol{\lambda}$**- convex**, *i.e.* $D^2 V(x)\, \xi \cdot \xi \geq \boldsymbol{\lambda} |\xi|^2 \qquad \forall x, \xi \in \mathbb{R}^d$.

Displacement convexity of the relative entropy functional

Consider

$$\Phi(\rho) = \mathrm{Ent}_\gamma(\rho) := \int_{\mathbb{R}^d} u(x)\log u(x)\,\mathrm{d}x + \int_{\mathbb{R}^d} V(x)\,\mathrm{d}\rho(x) \qquad \rho = u\mathscr{L}^d$$

$$= \int_{\mathbb{R}^d} \frac{\mathrm{d}\rho}{\mathrm{d}\gamma}\log\Big(\frac{\mathrm{d}\rho}{\mathrm{d}\gamma}\Big)\,\mathrm{d}\gamma \qquad \gamma := \mathrm{e}^{-V}\mathscr{L}^d$$

Theorem (McCann '97)

The relative entropy functional $\mathrm{Ent}_\gamma(\cdot)$ *is* $\boldsymbol{\lambda}$*-displacement convex iff*

the potential V **is** $\boldsymbol{\lambda}$**- convex**, *i.e.* $D^2 V(x)\,\boldsymbol{\xi}\cdot\boldsymbol{\xi} \geq \boldsymbol{\lambda}|\boldsymbol{\xi}|^2 \qquad \forall\, x, \boldsymbol{\xi} \in \mathbb{R}^d.$

When $\boldsymbol{\lambda} = 0$ one can give an equivalent characterization in terms of $\gamma = \mathrm{e}^{-V}\mathscr{L}^d$ which holds even in infinite dimensional Hilbert spaces H:

Theorem (Ambrosio-Gigli-S. '05)

The relative entropy functional Ent_γ *is displacement convex iff* γ **is log-concave**

$$\gamma((1-s)A + sB) \geq \gamma(A)^{1-s}\gamma(B)^s \quad \forall\, A, B \text{ open subset of } H, \quad s \in [0,1].$$

Subgradient characterization of the (FP)-velocity field

The last ingredient is the "subgradient" property of the vector field $\mathbf{v}$ given by the "logarithmic gradient" of ρ

$$\mathbf{v} = \frac{\nabla\rho}{\rho} + \nabla V$$

if and only if there exists a (unique) vector field $\mathbf{v} \in \mathrm{Tan}_\rho \mathscr{P}_2(\mathbb{R}^d) \subset L^2(\rho; \mathbb{R}^d)$ *such that*

$$\boxed{\int_{\mathbb{R}^d \times \mathbb{R}^d} \langle \mathbf{v}(x), y - x \rangle \, \mathrm{d}\boldsymbol{\rho}(x,y) \le \mathrm{Ent}_\gamma(\sigma) - \mathrm{Ent}_\gamma(\rho) - \frac{1}{2}\boldsymbol{\lambda}\, \mathrm{W}_2^2(\rho,\sigma)}$$

for every $\sigma \in \mathscr{P}_2(\mathbb{R}^d)$; $\boldsymbol{\rho}$ *is the optimal coupling between* ρ *and* σ.
In this case

$$\rho\mathbf{v} = \nabla\rho + \rho\nabla V.$$

Subgradient characterization of the (FP)-velocity field

The last ingredient is the "subgradient" property of the vector field $\mathbf{v}$ given by the "logarithmic gradient" of ρ

$$\mathbf{v} = \frac{\nabla \rho}{\rho} + \nabla V$$

Theorem (Ambrosio-Gigli-S. '05)

A measure $\rho \in \mathscr{P}_2(\mathbb{R}^d)$ has a Lebesgue density $u \in W^{1,1}_{\mathrm{loc}}(\mathbb{R}^d)$ with

$$\frac{\nabla u}{u} + \nabla V \in L^2(\rho; \mathbb{R}^d) \quad \textit{i.e.} \quad \mathrm{I}_\gamma(\rho) := \int_{\mathbb{R}^d} \Big| \frac{\nabla u}{u} + \nabla V \Big|^2 \mathrm{d}\rho(x) < +\infty$$

Subgradient characterization of the (FP)-velocity field

The last ingredient is the "subgradient" property of the vector field $\mathbf{v}$ given by the "logarithmic gradient" of ρ

$$\mathbf{v} = \frac{\nabla \rho}{\rho} + \nabla V$$

Theorem (Ambrosio-Gigli-S. '05)

A measure $\rho \in \mathscr{P}_2(\mathbb{R}^d)$ has a Lebesgue density $u \in W^{1,1}_{\mathrm{loc}}(\mathbb{R}^d)$ with

$$\frac{\nabla u}{u} + \nabla V \in L^2(\rho; \mathbb{R}^d) \quad \textit{i.e.} \quad \mathrm{I}_\gamma(\rho) := \int_{\mathbb{R}^d} \Big|\frac{\nabla u}{u} + \nabla V\Big|^2 \mathrm{d}\rho(x) < +\infty$$

if and only if there exists a (unique) vector field $\mathbf{v} \in \mathrm{Tan}_\rho \mathscr{P}_2(\mathbb{R}^d) \subset L^2(\rho; \mathbb{R}^d)$ such that

$$\boxed{\int_{\mathbb{R}^d \times \mathbb{R}^d} \langle \mathbf{v}(x), y - x\rangle \,\mathrm{d}\boldsymbol{\rho}(x,y) \le \mathrm{Ent}_\gamma(\sigma) - \mathrm{Ent}_\gamma(\rho) - \frac{1}{2}\boldsymbol{\lambda}\, \mathrm{W}_2^2(\rho,\sigma)}$$

for every $\sigma \in \mathscr{P}_2(\mathbb{R}^d)$; $\boldsymbol{\rho}$ is the optimal coupling between ρ and σ.

Subgradient characterization of the (FP)-velocity field

The last ingredient is the "subgradient" property of the vector field $\mathbf{v}$ given by the "logarithmic gradient" of ρ

$$\mathbf{v} = \frac{\nabla\rho}{\rho} + \nabla V$$

Theorem (Ambrosio-Gigli-S. '05)

A measure $\rho \in \mathscr{P}_2(\mathbb{R}^d)$ has a Lebesgue density $u \in W^{1,1}_{\rm loc}(\mathbb{R}^d)$ with

$$\frac{\nabla u}{u} + \nabla V \in L^2(\rho;\mathbb{R}^d) \quad \textit{i.e.} \quad \mathrm{I}_\gamma(\rho) := \int_{\mathbb{R}^d} \Big|\frac{\nabla u}{u} + \nabla V\Big|^2 \mathrm{d}\rho(x) < +\infty$$

if and only if there exists a (unique) vector field $\mathbf{v} \in \mathrm{Tan}_\rho\mathscr{P}_2(\mathbb{R}^d) \subset L^2(\rho;\mathbb{R}^d)$ such that

$$\boxed{\int_{\mathbb{R}^d\times\mathbb{R}^d} \langle \mathbf{v}(x), y-x\rangle \,\mathrm{d}\boldsymbol{\rho}(x,y) \le \mathrm{Ent}_\gamma(\sigma) - \mathrm{Ent}_\gamma(\rho) - \frac{1}{2}\boldsymbol{\lambda}\,\mathrm{W}_2^2(\rho,\sigma)}$$

for every $\sigma \in \mathscr{P}_2(\mathbb{R}^d)$; $\boldsymbol{\rho}$ is the optimal coupling between ρ and σ.
In this case

$$\rho\mathbf{v} = \nabla\rho + \rho\nabla V.$$

Subgradient characterization of the (FP)-velocity field

The last ingredient is the "subgradient" property of the vector field $\mathbf{v}$ given by the "logarithmic gradient" of ρ

$$\mathbf{v} = \frac{\nabla \rho}{\rho} + \nabla V$$

Theorem (Ambrosio-Gigli-S. '05)

A measure $\rho \in \mathscr{P}_2(\mathbb{R}^d)$ has a Lebesgue density $u \in W^{1,1}_{\rm loc}(\mathbb{R}^d)$ with

$$\frac{\nabla u}{u} + \nabla V \in L^2(\rho; \mathbb{R}^d) \quad \text{i.e.} \quad \mathrm{I}_\gamma(\rho) := \int_{\mathbb{R}^d} \Big| \frac{\nabla u}{u} + \nabla V \Big|^2 \mathrm{d}\rho(x) < +\infty$$

if and only if there exists a (unique) vector field $\mathbf{v} \in \mathrm{Tan}_\rho \mathscr{P}_2(\mathbb{R}^d) \subset L^2(\rho; \mathbb{R}^d)$ such that

$$\boxed{\int_{\mathbb{R}^d \times \mathbb{R}^d} \langle \mathbf{v}(x), y - x \rangle \, \mathrm{d}\boldsymbol{\rho}(x, y) \le \mathrm{Ent}_\gamma(\sigma) - \mathrm{Ent}_\gamma(\rho) - \frac{1}{2} \boldsymbol{\lambda} \, \mathrm{W}_2^2(\rho, \sigma)}$$

for every $\sigma \in \mathscr{P}_2(\mathbb{R}^d)$; $\boldsymbol{\rho}$ is the optimal coupling between ρ and σ. In this case

$$\rho \mathbf{v} = \nabla \rho + \rho \nabla V.$$

The functional I_γ is called **relative Fisher information**.

Wasserstein solutions to (FP)

If $\rho : (0, +\infty) \to \mathscr{P}_2(\mathbb{R}^d)$ is a locally Lipschitz curve solving the Fokker-Planck equation

$$\partial_t \rho_t + \mathrm{div}(\rho_t \mathbf{v}_t) = 0, \quad -\mathbf{v}_t = \Big(\frac{\nabla \rho_t}{\rho_t} + \nabla V\Big), \tag{FP}$$

and $\boldsymbol{\rho}_t$ is an optimal coupling between ρ_t and $\sigma \in \mathscr{P}_2(\mathbb{R}^d)$ we have

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{1}{2}W_2^2(\rho_t, \sigma) = \int_{\mathbb{R}^d \times \mathbb{R}^d} \langle -\mathbf{v}_t(x), y - x\rangle \,\mathrm{d}\boldsymbol{\rho}_t(x, y)$$

Wasserstein solutions to (FP)

If $\rho : (0, +\infty) \to \mathscr{P}_2(\mathbb{R}^d)$ is a locally Lipschitz curve solving the Fokker-Planck equation

$$\partial_t \rho_t + \mathrm{div}(\rho_t \mathbf{v}_t) = 0, \quad -\mathbf{v}_t = \Big(\frac{\nabla \rho_t}{\rho_t} + \nabla V\Big), \tag{FP}$$

and $\boldsymbol{\rho}_t$ is an optimal coupling between ρ_t and $\sigma \in \mathscr{P}_2(\mathbb{R}^d)$ we have

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\frac{1}{2}W_2^2(\rho_t, \sigma) &= \int_{\mathbb{R}^d \times \mathbb{R}^d} \langle -\mathbf{v}_t(x), y - x\rangle \,\mathrm{d}\boldsymbol{\rho}_t(x, y) \\ &\le \mathrm{Ent}_\gamma(\sigma) - \mathrm{Ent}_\gamma(\rho) - \frac{1}{2}\boldsymbol{\lambda}\, \mathrm{W}_2^2(\rho, \sigma)\end{aligned}$$

Wasserstein solutions to (FP)

If $\rho : (0,+\infty) \to \mathscr{P}_2(\mathbb{R}^d)$ is a locally Lipschitz curve solving the Fokker-Planck equation

$$\partial_t \rho_t + \mathrm{div}(\rho_t \mathbf{v}_t) = 0, \quad -\mathbf{v}_t = \Big(\frac{\nabla \rho_t}{\rho_t} + \nabla V\Big), \tag{FP}$$

and $\boldsymbol{\rho}_t$ is an optimal coupling between ρ_t and $\sigma \in \mathscr{P}_2(\mathbb{R}^d)$ we have

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\frac{1}{2}W_2^2(\rho_t,\sigma) &= \int_{\mathbb{R}^d\times\mathbb{R}^d} \langle -\mathbf{v}_t(x), y-x\rangle \,\mathrm{d}\boldsymbol{\rho}_t(x,y) \\ &\le \mathrm{Ent}_\gamma(\sigma) - \mathrm{Ent}_\gamma(\rho) - \frac{1}{2}\boldsymbol{\lambda}\,\mathrm{W}_2^2(\rho,\sigma)\end{aligned}$$

Definition

A *locally Lipschitz* curve $\rho : (0,+\infty) \to \mathscr{P}_2(\mathbb{R}^d)$ is a "Wasserstein" solution to (FP) if it satisfies the Evolution Variational Inequality

$$\boxed{\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\mathrm{W}_2^2(\rho_t,\sigma) \le \mathrm{Ent}_\gamma(\sigma) - \mathrm{Ent}_\gamma(\rho_t) - \frac{1}{2}\boldsymbol{\lambda}\,\mathrm{W}_2^2(\rho_t,\sigma)} \tag{EVI}$$

for every $\sigma \in \mathscr{P}_2(\mathbb{R}^d)$.

Proof of contraction

Let ρ^1, ρ^2 be locally Lipschitz curves solving (EVI) ($\lambda = 0$)

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\mathrm{W}_2^2(\rho_t,\sigma) \le \mathrm{Ent}_\gamma(\sigma) - \mathrm{Ent}_\gamma(\rho_t). \qquad \text{(EVI)}$$

Proof of contraction

Let ρ^1, ρ^2 be locally Lipschitz curves solving (EVI) ($\lambda = 0$)

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\mathrm{W}_2^2(\rho_t, \sigma) \leq \mathrm{Ent}_\gamma(\sigma) - \mathrm{Ent}_\gamma(\rho_t). \tag{EVI}$$

Doubling variables, we write (EVI) for ρ^1_r choosing $\sigma := \rho^2_s$

$$\frac{1}{2}\frac{\partial}{\partial r}\mathrm{W}_2^2(\rho^1_r, \rho^2_s) \leq \mathrm{Ent}_\gamma(\rho^2_s) - \mathrm{Ent}_\gamma(\rho^1_r)$$

Proof of contraction

Let $\boldsymbol{\rho}^1, \boldsymbol{\rho}^2$ be locally Lipschitz curves solving (EVI) ($\boldsymbol{\lambda} = 0$)

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t} \mathrm{W}_2^2(\rho_t, \sigma) \leq \mathrm{Ent}_\gamma(\sigma) - \mathrm{Ent}_\gamma(\rho_t). \qquad \text{(EVI)}$$

Doubling variables, we write (EVI) for $\boldsymbol{\rho}^1_{\mathsf{r}}$ choosing $\sigma := \boldsymbol{\rho}^2_{\mathsf{s}}$ and the corresponding (EVI) for $\boldsymbol{\rho}^2_{\mathsf{s}}$ choosing $\sigma := \boldsymbol{\rho}^1_{\mathsf{r}}$:

$$\frac{1}{2}\frac{\partial}{\partial r} \mathrm{W}_2^2(\boldsymbol{\rho}^1_{\mathsf{r}}, \boldsymbol{\rho}^2_{\mathsf{s}}) \leq \mathrm{Ent}_\gamma(\boldsymbol{\rho}^2_{\mathsf{s}}) - \mathrm{Ent}_\gamma(\boldsymbol{\rho}^1_{\mathsf{r}})$$

$$\frac{1}{2}\frac{\partial}{\partial s} \mathrm{W}_2^2(\boldsymbol{\rho}^1_{\mathsf{r}}, \boldsymbol{\rho}^2_{\mathsf{s}}) \leq \mathrm{Ent}_\gamma(\boldsymbol{\rho}^1_{\mathsf{r}}) - \mathrm{Ent}_\gamma(\boldsymbol{\rho}^2_{\mathsf{s}})$$

Proof of contraction

Let ρ^1, ρ^2 be locally Lipschitz curves solving (EVI) ($\lambda = 0$)

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\mathrm{W}_2^2(\rho_t, \sigma) \le \mathrm{Ent}_\gamma(\sigma) - \mathrm{Ent}_\gamma(\rho_t). \qquad \text{(EVI)}$$

Doubling variables, we write (EVI) for ρ^1_r choosing $\sigma := \rho^2_s$ and the corresponding (EVI) for ρ^2_s choosing $\sigma := \rho^1_r$:

$$\frac{1}{2}\frac{\partial}{\partial r}\mathrm{W}_2^2(\rho^1_r, \rho^2_s) \le \mathrm{Ent}_\gamma(\rho^2_s) - \mathrm{Ent}_\gamma(\rho^1_r)$$
$$\frac{1}{2}\frac{\partial}{\partial s}\mathrm{W}_2^2(\rho^1_r, \rho^2_s) \le \mathrm{Ent}_\gamma(\rho^1_r) - \mathrm{Ent}_\gamma(\rho^2_s)$$

Summing up

$$\frac{1}{2}\frac{\partial}{\partial r}\mathrm{W}_2^2(\rho^1_r, \rho^2_s) + \frac{1}{2}\frac{\partial}{\partial s}\mathrm{W}_2^2(\rho^1_r, \rho^2_s) \le 0$$

Proof of contraction

Let ρ^1, ρ^2 be locally Lipschitz curves solving (EVI) ($\lambda = 0$)

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\mathrm{W}_2^2(\rho_t, \sigma) \leq \mathrm{Ent}_\gamma(\sigma) - \mathrm{Ent}_\gamma(\rho_t). \qquad \text{(EVI)}$$

Doubling variables, we write (EVI) for ρ^1_r choosing $\sigma := \rho^2_s$ and the corresponding (EVI) for ρ^2_s choosing $\sigma := \rho^1_r$:

$$\frac{1}{2}\frac{\partial}{\partial r}\mathrm{W}_2^2(\rho^1_r, \rho^2_s) \leq \mathrm{Ent}_\gamma(\rho^2_s) - \mathrm{Ent}_\gamma(\rho^1_r)$$

$$\frac{1}{2}\frac{\partial}{\partial s}\mathrm{W}_2^2(\rho^1_r, \rho^2_s) \leq \mathrm{Ent}_\gamma(\rho^1_r) - \mathrm{Ent}_\gamma(\rho^2_s)$$

Summing up

$$\frac{1}{2}\frac{\partial}{\partial r}\mathrm{W}_2^2(\rho^1_r, \rho^2_s) + \frac{1}{2}\frac{\partial}{\partial s}\mathrm{W}_2^2(\rho^1_r, \rho^2_s) \leq 0$$

and evaluating the inequality at $t = r = s$

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\mathrm{W}_2^2(\rho^1_t, \rho^2_t) = \frac{1}{2}\Big(\frac{\partial}{\partial r}\mathrm{W}_2^2(\rho^1_r, \rho^2_s) + \frac{\partial}{\partial s}\mathrm{W}_2^2(\rho^1_r, \rho^2_s)\Big)_{t=r=s} \leq 0$$

A general existence result

Here H is a separable Hilbert space.

Theorem (Ambrosio-Gigli-S. '05, S. '07)

If $\Phi : \mathscr{P}_2(H) \to (-\infty, +\infty]$ *is a proper, lower semicontinuous, and displacement* $\boldsymbol{\lambda}$*-convex functional then for every initial measure* $\rho_0 \in \overline{D(\Phi)} \subset \mathscr{P}_2(\mathbb{R}^d)$ *there exists a unique locally Lipschitz solution to*

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\mathrm{W}_2^2(\rho_t, \sigma) \le \mathrm{Ent}_\gamma(\sigma) - \mathrm{Ent}_\gamma(\rho_t) - \frac{1}{2}\boldsymbol{\lambda}\,\mathrm{W}_2^2(\rho_t, \sigma) \quad \forall \sigma \in \mathscr{P}_2(H) \quad \text{(EVI)}$$

with $\lim_{t\downarrow 0} \rho_t = \rho_0$.

A general existence result

Here H is a separable Hilbert space.

Theorem (Ambrosio-Gigli-S. '05, S. '07)

If $\Phi : \mathscr{P}_2(H) \to (-\infty, +\infty]$ *is a proper, lower semicontinuous, and displacement* $\boldsymbol{\lambda}$*-convex functional then for every initial measure* $\rho_0 \in \overline{D(\Phi)} \subset \mathscr{P}_2(\mathbb{R}^d)$ *there exists a unique locally Lipschitz solution to*

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\mathrm{W}_2^2(\rho_t, \sigma) \le \mathrm{Ent}_\gamma(\sigma) - \mathrm{Ent}_\gamma(\rho_t) - \frac{1}{2}\boldsymbol{\lambda}\,\mathrm{W}_2^2(\rho_t, \sigma) \quad \forall \sigma \in \mathscr{P}_2(H) \quad \text{(EVI)}$$

with $\lim_{t\downarrow 0} \rho_t = \rho_0$.
The map $\rho_0 \mapsto S_t(\rho_0) := \rho_t$ *defines a continuous semigroup of* $\boldsymbol{\lambda}$*-contractions in* $\overline{D(\Phi)}$.

A general existence result

Here H is a separable Hilbert space.

Theorem (Ambrosio-Gigli-S. '05, S. '07)

If $\Phi : \mathscr{P}_2(H) \to (-\infty, +\infty]$ is a proper, lower semicontinuous, and displacement $\boldsymbol{\lambda}$-convex functional then for every initial measure $\rho_0 \in \overline{D(\Phi)} \subset \mathscr{P}_2(\mathbb{R}^d)$ there exists a unique locally Lipschitz solution to

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\mathrm{W}_2^2(\rho_t, \sigma) \le \mathrm{Ent}_\gamma(\sigma) - \mathrm{Ent}_\gamma(\rho_t) - \frac{1}{2}\boldsymbol{\lambda}\,\mathrm{W}_2^2(\rho_t, \sigma) \quad \forall \sigma \in \mathscr{P}_2(H) \quad \text{(EVI)}$$

with $\lim_{t\downarrow 0} \rho_t = \rho_0$.
The map $\rho_0 \mapsto S_t(\rho_0) := \rho_t$ defines a continuous semigroup of $\boldsymbol{\lambda}$-contractions in $\overline{D(\Phi)}$.

In the case $H = \mathbb{R}^d$ when $\Phi = \mathrm{Ent}_\gamma$, ρ is a distributional solution to

$$\partial_t\rho - \Delta\rho - \mathrm{div}(\rho\nabla V) = 0 \quad \text{(FP)}$$

and satisfies the regularization property

$$\mathrm{I}_\gamma(\rho_t) = \int_{\mathbb{R}^d} \Big|\frac{\nabla\rho_t}{\rho_t} + \nabla V\Big|^2 \mathrm{d}\rho_t \le C\,t^{-2}\,\mathrm{e}^{2\boldsymbol{\lambda} t}$$

Stability: an example in infinite dimension

Let H be an infinite dimensional Hilbert space and $\gamma \in \mathscr{P}(H)$ be a **log-concave** non degenerate measure.

Take a family of finite n-dimensional spaces H^n with a log-concave measure $\gamma_n = e^{-V^n} \cdot \mathscr{L}^n|_{H^n} \in \mathscr{P}(H^n)$ and suppose that

$$\boxed{\gamma^n \rightharpoonup \gamma \quad \text{weakly in } \mathscr{P}(H).}$$

Stability: an example in infinite dimension

Let H be an infinite dimensional Hilbert space and $\gamma \in \mathscr{P}(H)$ be a **log-concave** non degenerate measure.

Take a family of finite n-dimensional spaces H^n with a log-concave measure $\gamma_n = e^{-V^n} \cdot \mathscr{L}^n|_{H^n} \in \mathscr{P}(H^n)$ and suppose that

$$\boxed{\gamma^n \rightharpoonup \gamma \quad \text{weakly in } \mathscr{P}(H).}$$

Typical example: $\gamma^n = (\pi^n)_\# \gamma = \mathrm{e}^{-V^n} \mathscr{L}^n|_{H^n}$ are the marginals of γ with respect to the finite dimensional orthogonal projections $\pi^n : H \to H^n \subset H$.

Stability: an example in infinite dimension

Let H be an infinite dimensional Hilbert space and $\gamma \in \mathscr{P}(H)$ be a **log-concave** non degenerate measure.

Take a family of finite n-dimensional spaces H^n with a log-concave measure $\gamma_n = e^{-V^n} \cdot \mathscr{L}^n|_{H^n} \in \mathscr{P}(H^n)$ and suppose that

$$\boxed{\gamma^n \rightharpoonup \gamma \quad \text{weakly in } \mathscr{P}(H).}$$

Typical example: $\gamma^n = (\pi^n)_\# \gamma = \mathrm{e}^{-V^n} \mathscr{L}^n|_{H^n}$ are the marginals of γ with respect to the finite dimensional orthogonal projections $\pi^n : H \to H^n \subset H$.

On each finite dimensional space H^n we are solving

$$\partial_t \rho_t^n - \Delta \rho_t^n - \mathrm{div}(\rho_t^n \nabla V^n) = 0 \quad \rho^n|_{t=0} = \rho_0^n = (\pi^n)_\# \rho_0.$$

Stability: an example in infinite dimension

Let H be an infinite dimensional Hilbert space and $\gamma \in \mathscr{P}(H)$ be a **log-concave** non degenerate measure.

Take a family of finite n-dimensional spaces H^n with a log-concave measure $\gamma_n = e^{-V^n} \cdot \mathscr{L}^n\big|_{H^n} \in \mathscr{P}(H^n)$ and suppose that

$$\boxed{\gamma^n \rightharpoonup \gamma \quad \text{weakly in } \mathscr{P}(H).}$$

Typical example: $\gamma^n = (\pi^n)_\# \gamma = \mathrm{e}^{-V^n} \mathscr{L}^n\big|_{H^n}$ are the marginals of γ with respect to the finite dimensional orthogonal projections $\pi^n : H \to H^n \subset H$.

On each finite dimensional space H^n we are solving

$$\partial_t \rho_t^n - \Delta \rho_t^n - \mathrm{div}(\rho_t^n \nabla V^n) = 0 \quad \rho^n\big|_{t=0} = \rho_0^n = (\pi^n)_\# \rho_0.$$

Theorem [Ambrosio-S.-Zambotti]

If $\rho_0^n \to \rho_0$ in $\mathscr{P}_2(H)$ then $\rho_t^n \to \rho_t$ in $\mathscr{P}_2(H)$, and ρ_t is the unique solution of the gradient flow generated by Ent_γ in $\mathscr{P}_2(X)$ starting from ρ_0.

Stability: an example in infinite dimension

Let H be an infinite dimensional Hilbert space and $\gamma \in \mathscr{P}(H)$ be a **log-concave** non degenerate measure.

Take a family of finite n-dimensional spaces H^n with a log-concave measure $\gamma_n = e^{-V^n} \cdot \mathscr{L}^n|_{H^n} \in \mathscr{P}(H^n)$ and suppose that

$$\boxed{\gamma^n \rightharpoonup \gamma \quad \text{weakly in } \mathscr{P}(H).}$$

Typical example: $\gamma^n = (\pi^n)_\# \gamma = \mathrm{e}^{-V^n} \mathscr{L}^n|_{H^n}$ are the marginals of γ with respect to the finite dimensional orthogonal projections $\pi^n : H \to H^n \subset H$.

On each finite dimensional space H^n we are solving

$$\partial_t \rho_t^n - \Delta \rho_t^n - \mathrm{div}(\rho_t^n \nabla V^n) = 0 \quad \rho^n|_{t=0} = \rho_0^n = (\pi^n)_\# \rho_0.$$

Theorem [Ambrosio-S.-Zambotti]

If $\rho_0^n \to \rho_0$ in $\mathscr{P}_2(H)$ then $\rho_t^n \to \rho_t$ in $\mathscr{P}_2(H)$, and ρ_t is the unique solution of the gradient flow generated by Ent_γ in $\mathscr{P}_2(X)$ starting from ρ_0.

Application:

The Dirichlet form $\int_H \|\nabla u\|_H^2 \, \mathrm{d}\gamma$ is **closable.**

Outline

Goals

- Prove contraction for **general λ-monotone drifts** which are not necessarily gradients.

Goals

- Prove contraction for **general λ-monotone drifts** which are not necessarily gradients.
- Extend the estimates to **more general transportation costs**

Goals

- Prove contraction for **general λ-monotone drifts** which are not necessarily gradients.
- Extend the estimates to **more general transportation costs**
- Avoid any time-regularity assumption (like Lipschitz continuity) and use just the **notion of distributional solution**.

Goals

- Prove contraction for **general λ-monotone drifts** which are not necessarily gradients.
- Extend the estimates to **more general transportation costs**
- Avoid any time-regularity assumption (like Lipschitz continuity) and use just the **notion of distributional solution**.

Difficulties:

- Lack of entropy and metric formulation

Goals

- Prove contraction for **general λ-monotone drifts** which are not necessarily gradients.
- Extend the estimates to **more general transportation costs**
- Avoid any time-regularity assumption (like Lipschitz continuity) and use just the **notion of distributional solution**.

Difficulties:

- Lack of entropy and metric formulation
- The velocity field

$$\mathbf{v}_t := -\frac{\nabla \rho_t}{\rho_t} - B \notin \mathrm{Tan}_{\rho_t}\mathscr{P}_2(\mathbb{R}^d)$$

is not the natural “Wasserstein velocity”, since B is not a gradient

Goals

- Prove contraction for **general λ-monotone drifts** which are not necessarily gradients.
- Extend the estimates to **more general transportation costs**
- Avoid any time-regularity assumption (like Lipschitz continuity) and use just the **notion of distributional solution**.

Difficulties:

- Lack of entropy and metric formulation
- The velocity field

$$\mathbf{v}_t := -\frac{\nabla \rho_t}{\rho_t} - B \notin \mathrm{Tan}_{\rho_t} \mathscr{P}_2(\mathbb{R}^d)$$

  is not the natural "Wasserstein velocity", since B is not a gradient
- We cannot directly differentiate the Wasserstein distance.

General transportation costs

We consider a function

$$h : [0, +\infty) \to [0, +\infty) \textit{ continuous and } \textbf{non-decreasing},\ h(0) = 0,$$

and an associated cost function $c(\mathbf{x_1}, \mathbf{x_2}) := h\big(|\mathbf{x_1} - \mathbf{x_2}|\big)$.

General transportation costs

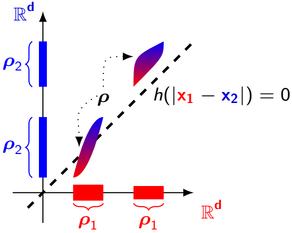
We consider a function

$h : [0,+\infty) \to [0,+\infty)$ *continuous and* **non-decreasing**, $h(0) = 0$,

and an associated cost function $c(\mathbf{x_1},\mathbf{x_2}) := h\big(|\mathbf{x_1}-\mathbf{x_2}|\big)$. We then introduce the cost functional

$$\mathcal{C}_h(\boldsymbol{\rho}^1,\boldsymbol{\rho}^2) := \inf\Big\{ \int_{\mathbb{R}^d\times\mathbb{R}^d} h\big(|\mathbf{x_1}-\mathbf{x_2}|\big)\,\mathrm{d}\boldsymbol{\rho}(\mathbf{x_1},\mathbf{x_2}) : \boldsymbol{\rho}\in\mathscr{P}(\mathbb{R}^d\times\mathbb{R}^d),$$

$$\boldsymbol{\rho} \text{ is a coupling between } \boldsymbol{\rho}^1 \text{ and } \boldsymbol{\rho}^2\Big\}.$$



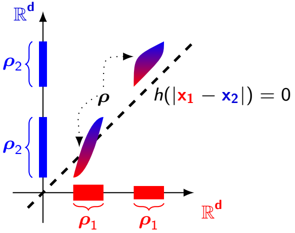
General transportation costs

We consider a function

$$h : [0, +\infty) \to [0, +\infty) \textit{ continuous and } \textbf{non-decreasing}, \ h(0) = 0,$$

and an associated cost function $c(\mathbf{x}_1, \mathbf{x}_2) := h\big(|\mathbf{x}_1 - \mathbf{x}_2|\big)$. We then introduce the cost functional

$$\mathcal{C}_h(\boldsymbol{\rho}^1, \boldsymbol{\rho}^2) := \inf \Big\{ \int_{\mathbb{R}^d \times \mathbb{R}^d} h\big(|\mathbf{x}_1 - \mathbf{x}_2|\big)\, \mathrm{d}\boldsymbol{\rho}(\mathbf{x}_1, \mathbf{x}_2) : \boldsymbol{\rho} \in \mathscr{P}(\mathbb{R}^d \times \mathbb{R}^d),$$
$$\boldsymbol{\rho} \text{ is a coupling between } \boldsymbol{\rho}^1 \text{ and } \boldsymbol{\rho}^2 \Big\}.$$



Interesting cases:

- $h(r) = r^p$, $\mathcal{C}_h(\rho^1, \rho^2)$ induces the p-Wasserstein distance W_p^p

General transportation costs

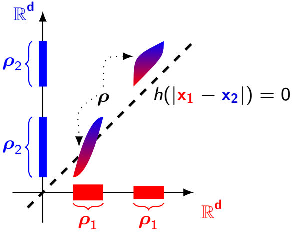
We consider a function

$$h : [0, +\infty) \to [0, +\infty) \textit{ continuous and } \textbf{non-decreasing}, \; h(0) = 0,$$

and an associated cost function $c(\mathbf{x}_1, \mathbf{x}_2) := h\big(|\mathbf{x}_1 - \mathbf{x}_2|\big)$. We then introduce the cost functional

$$\mathcal{C}_h(\boldsymbol{\rho}^1, \boldsymbol{\rho}^2) := \inf \Big\{ \int_{\mathbb{R}^d \times \mathbb{R}^d} h\big(|\mathbf{x}_1 - \mathbf{x}_2|\big) \, \mathrm{d}\boldsymbol{\rho}(\mathbf{x}_1, \mathbf{x}_2) : \boldsymbol{\rho} \in \mathscr{P}(\mathbb{R}^d \times \mathbb{R}^d),$$

$$\boldsymbol{\rho} \text{ is a coupling between } \boldsymbol{\rho}^1 \text{ and } \boldsymbol{\rho}^2 \Big\}.$$



Interesting cases:

- $h(r) = r^p$, $\mathcal{C}_h(\rho^1, \rho^2)$ induces the p-Wasserstein distance W_p^p
- h bounded, concave, strictly increasing: $\mathcal{C}_h(\rho^1, \rho^2)$ is a metric on $\mathscr{P}(\mathbb{R}^d)$ inducing weak convergence.

Main result

Assume that $B : \mathbb{R}^d \to \mathbb{R}^d$ is a Borel monotone drift

then

$$\mathcal{C}_{h_{\lambda_t}}(\rho^1_t, \rho^2_t) \le \mathcal{C}_h(\rho^1_0, \rho^2_0) \quad \textit{for every } t \ge 0.$$

In the p-homogeneous case when $h(r) := r^p$ *we have*

$$\mathrm{W}_p(\rho^1_t, \rho^2_t) \le \mathrm{e}^{-\lambda t}\mathrm{W}_p(\rho^1_0, \rho^2_0).$$

Main result

Assume that $B : \mathbb{R}^d \to \mathbb{R}^d$ is a Borel monotone drift

Theorem (Natile-Peletier-S. '10)

If ρ^1, ρ^2 are two distributional solutions to (FP)

$$\partial_t \rho - \Delta\rho - \nabla \cdot (\rho B) = 0 \tag{FP}$$

satisfying the summability condition

$$\int_{t_0}^{t_1} \int_{\mathbb{R}^d} |B(x) \qquad | \, \mathrm{d}\rho_t(x) \, \mathrm{d}t < +\infty \quad \textit{for every } 0 < t_0 < t_1 < +\infty,$$

Main result

Assume that $B : \mathbb{R}^d \to \mathbb{R}^d$ is a Borel monotone drift

Theorem (Natile-Peletier-S. '10)

If ρ^1, ρ^2 are two distributional solutions to (FP)

$$\partial_t \rho - \Delta \rho - \nabla \cdot (\rho B) = 0 \tag{FP}$$

satisfying the summability condition

$$\int_{t_0}^{t_1} \int_{\mathbb{R}^d} |B(x) \qquad | \, \mathrm{d}\rho_t(x) \, \mathrm{d}t < +\infty \quad \textit{for every } 0 < t_0 < t_1 < +\infty,$$

then

$$\mathcal{C}_h \quad (\rho^1_t, \rho^2_t) \le \mathcal{C}_h(\rho^1_0, \rho^2_0) \quad \textit{for every } t \ge 0.$$

Main result

Assume that $B : \mathbb{R}^d \to \mathbb{R}^d$ is a Borel monotone drift

Theorem (Natile-Peletier-S. '10)

If ρ^1, ρ^2 are two distributional solutions to (FP)

$$\partial_t \rho - \Delta\rho - \nabla \cdot (\rho B) = 0 \tag{FP}$$

satisfying the summability condition

$$\int_{t_0}^{t_1} \int_{\mathbb{R}^d} |B(x)| \, \mathrm{d}\rho_t(x)\,\mathrm{d}t < +\infty \quad \textit{for every } 0 < t_0 < t_1 < +\infty,$$

then

$$\mathcal{C}_h(\rho^1_t, \rho^2_t) \le \mathcal{C}_h(\rho^1_0, \rho^2_0) \quad \textit{for every } t \ge 0.$$

In the p-homogeneous case when $h(r) := r^p$ we have

$$W_p(\rho^1_t, \rho^2_t) \le W_p(\rho^1_0, \rho^2_0).$$

Main result

Assume that $B : \mathbb{R}^d \to \mathbb{R}^d$ is a Borel $\boldsymbol{\lambda}$-monotone drift and denote by $h_{\boldsymbol{\lambda} t}$ the rescaled cost

$$h_{\boldsymbol{\lambda} t}(r) := h(r\,\mathrm{e}^{\boldsymbol{\lambda} t}).$$

Theorem (Natile-Peletier-S. '10)

If ρ^1, ρ^2 are two distributional solutions to (FP)

$$\partial_t \rho - \Delta\rho - \nabla\cdot(\rho B) = 0 \tag{FP}$$

satisfying the summability condition

$$\int_{t_0}^{t_1}\int_{\mathbb{R}^d} |B(x) \qquad | \,\mathrm{d}\rho_t(x)\,\mathrm{d}t < +\infty \quad \textit{for every } 0 < t_0 < t_1 < +\infty,$$

then

$$\mathcal{C}_{h_{\boldsymbol{\lambda} t}}(\boldsymbol{\rho}^1_{\mathbf{t}}, \boldsymbol{\rho}^2_{\mathbf{t}}) \le \mathcal{C}_h(\boldsymbol{\rho}^1_{\mathbf{0}}, \boldsymbol{\rho}^2_{\mathbf{0}}) \quad \textit{for every } t \ge 0.$$

In the p-homogeneous case when $h(r) := r^p$ we have

$$\mathrm{W}_p(\boldsymbol{\rho}^1_{\mathbf{t}}, \boldsymbol{\rho}^2_{\mathbf{t}}) \le \qquad \mathrm{W}_p(\boldsymbol{\rho}^1_{\mathbf{0}}, \boldsymbol{\rho}^2_{\mathbf{0}}).$$

Main result

Assume that $B : \mathbb{R}^d \to \mathbb{R}^d$ is a Borel λ-monotone drift and denote by $h_{\lambda t}$ the rescaled cost

$$h_{\lambda t}(r) := h(r\,\mathrm{e}^{\lambda t}).$$

Theorem (Natile-Peletier-S. '10)

If ρ^1, ρ^2 are two distributional solutions to (FP)

$$\partial_t \rho - \Delta\rho - \nabla\cdot(\rho B) = 0 \qquad \text{(FP)}$$

satisfying the summability condition

$$\int_{t_0}^{t_1}\int_{\mathbb{R}^d} |B(x) - \lambda\, x|\,\mathrm{d}\rho_t(x)\,\mathrm{d}t < +\infty \quad \textit{for every } 0 < t_0 < t_1 < +\infty,$$

then

$$\mathcal{C}_{h_{\lambda t}}(\rho^1_t, \rho^2_t) \le \mathcal{C}_h(\rho^1_0, \rho^2_0) \quad \textit{for every } t \ge 0.$$

In the p-homogeneous case when $h(r) := r^p$ we have

$$\mathrm{W}_p(\rho^1_t, \rho^2_t) \le \qquad \mathrm{W}_p(\rho^1_0, \rho^2_0).$$

Main result

Assume that $B : \mathbb{R}^d \to \mathbb{R}^d$ is a Borel $\boldsymbol{\lambda}$-monotone drift and denote by $h_{\boldsymbol{\lambda} t}$ the rescaled cost

$$h_{\boldsymbol{\lambda} t}(r) := h(r\,\mathrm{e}^{\boldsymbol{\lambda} t}).$$

Theorem (Natile-Peletier-S. '10)

If $\boldsymbol{\rho}^1, \boldsymbol{\rho}^2$ are two distributional solutions to (FP)

$$\partial_t \rho - \Delta\rho - \nabla\cdot(\rho B) = 0 \qquad \text{(FP)}$$

satisfying the summability condition

$$\int_{t_0}^{t_1}\int_{\mathbb{R}^d} |B(x) - \boldsymbol{\lambda}\,x|\,\mathrm{d}\rho_t(x)\,\mathrm{d}t < +\infty \quad \textit{for every } 0 < t_0 < t_1 < +\infty,$$

then

$$\mathcal{C}_{h_{\boldsymbol{\lambda} t}}(\boldsymbol{\rho}^1_{\mathbf{t}}, \boldsymbol{\rho}^2_{\mathbf{t}}) \le \mathcal{C}_h(\boldsymbol{\rho}^1_{\mathbf{0}}, \boldsymbol{\rho}^2_{\mathbf{0}}) \quad \textit{for every } t \ge 0.$$

In the p-homogeneous case when $h(r) := r^p$ we have

$$\mathrm{W}_p(\boldsymbol{\rho}^1_{\mathbf{t}}, \boldsymbol{\rho}^2_{\mathbf{t}}) \le \mathrm{e}^{-\boldsymbol{\lambda} t}\mathrm{W}_p(\boldsymbol{\rho}^1_{\mathbf{0}}, \boldsymbol{\rho}^2_{\mathbf{0}}).$$

Arguments

- **Dual Kantorovich formulation** of the cost functional $\mathcal{C}_h(\cdot, \cdot)$.

Arguments

- **Dual Kantorovich formulation** of the cost functional $\mathcal{C}_h(\cdot, \cdot)$.
- Smooth solutions of the **backward Kolmogorov equation**.

Arguments

- **Dual Kantorovich formulation** of the cost functional $\mathcal{C}_h(\cdot,\cdot)$.
- Smooth solutions of the **backward Kolmogorov equation**.
- **4 "regularization" steps** (for the cost, the Kantorovich potentials, the drift, and the λ-monotonicity)

Arguments

- **Dual Kantorovich formulation** of the cost functional $\mathcal{C}_h(\cdot, \cdot)$.
- Smooth solutions of the **backward Kolmogorov equation**.
- **4 "regularization" steps** (for the cost, the Kantorovich potentials, the drift, and the λ-monotonicity)
- **Variable-doubling** technique and **maximum principle** to get refined comparison results for smooth solutions of the backward equation.

Kantorovich duality

Our argument strongly relies on the dual Kantorovich representation of $\mathcal{C}_h(\cdot,\cdot)$

$$\mathcal{C}_h(\rho^1,\rho^2) = \sup\Big\{ \int_{\mathbb{R}^d} \psi^1(\mathsf{x}_1)\,\mathrm{d}\rho^1(\mathsf{x}_1) + \int_{\mathbb{R}^d} \psi^2(\mathsf{x}_2)\,\mathrm{d}\rho^2(\mathsf{x}_2) : \psi^1,\psi^2 \in C_\mathrm{b}(\mathbb{R}^d),\quad \psi^1(\mathsf{x}_1)+\psi^2(\mathsf{x}_2) \le h(|\mathsf{x}_1-\mathsf{x}_2|)\Big\}$$

Kantorovich duality

Our argument strongly relies on the dual Kantorovich representation of $\mathcal{C}_h(\cdot,\cdot)$

$$\mathcal{C}_h(\rho^1,\rho^2) = \sup\Big\{ \int_{\mathbb{R}^d} \psi^1(\mathsf{x}_1)\,\mathrm{d}\rho^1(\mathsf{x}_1) + \int_{\mathbb{R}^d} \psi^2(\mathsf{x}_2)\,\mathrm{d}\rho^2(\mathsf{x}_2) : \psi^1,\psi^2 \in C_\mathrm{b}(\mathbb{R}^d),\quad \psi^1(\mathsf{x}_1)+\psi^2(\mathsf{x}_2) \le h(|\mathsf{x}_1-\mathsf{x}_2|)\Big\}$$

Kantorovich duality reduces the estimate of the cost $\mathcal{C}_h(\rho^1_\mathsf{T},\rho^2_\mathsf{T})$ of two solutions of (FP) at a certain final time T to the estimate of

$$\Sigma(\psi^1,\psi^2;T) := \int_{\mathbb{R}^d} \psi^1\,\mathrm{d}\rho^1_\mathsf{T} + \int_{\mathbb{R}^d} \psi^2\,\mathrm{d}\rho^2_\mathsf{T}$$

for an arbitrary pair of functions ψ^1,ψ^2 satisfying the constraint

$$\psi^1(\mathsf{x}_1)+\psi^2(\mathsf{x}_2) \le h(|\mathsf{x}_1-\mathsf{x}_2|) \quad \text{for every } \mathsf{x}_1,\mathsf{x}_2 \in \mathbb{R}^d.$$

Backward Kolmogorov equation

Assume for a moment that B is **bounded and smooth**. In order to estimate

$$\Sigma(\psi^1, \psi^2; T) := \int_{\mathbb{R}^d} \psi^1 \, \mathrm{d}\rho^1_T + \int_{\mathbb{R}^d} \psi^2 \, \mathrm{d}\rho^2_T, \quad \psi^1(x_1) + \psi^2(x_2) \le h(|x_1 - x_2|),$$

we can **solve the final-value problem for the adjoint backward Kolomogorov equation**

$$\partial_t \psi^{1,2} + \Delta \psi^{1,2} - B \cdot \nabla \psi^{1,2} = 0 \quad \text{in } \mathbb{R}^d \times (0, T), \quad \psi^{1,2}(\cdot, T) := \psi^{1,2} \quad \text{(BKE)}$$

Backward Kolmogorov equation

Assume for a moment that B is **bounded and smooth**. In order to estimate

$$\Sigma(\psi^1,\psi^2;T) := \int_{\mathbb{R}^d} \psi^1 \,\mathrm{d}\boldsymbol{\rho}^1_{\mathsf{T}} + \int_{\mathbb{R}^d} \psi^2 \,\mathrm{d}\boldsymbol{\rho}^2_{\mathsf{T}}, \quad \psi^1(\mathsf{x}_1) + \psi^2(\mathsf{x}_2) \le h(|\mathsf{x}_1 - \mathsf{x}_2|),$$

we can **solve the final-value problem for the adjoint backward Kolomogorov equation**

$$\partial_t \psi^{1,2} + \Delta\psi^{1,2} - B\cdot\nabla\psi^{1,2} = 0 \quad \text{in } \mathbb{R}^d\times(0,T), \quad \psi^{1,2}(\cdot,T) := \psi^{1,2} \quad \text{(BKE)}$$

If $\psi^{1,2}$ belong to $C^{2,1}_{\mathrm{b}}(\mathbb{R}^d\times[0,T])$, the distributional formulation (FP) yields

$$\boxed{\Sigma(\psi^1_{\mathsf{T}},\psi^2_{\mathsf{T}};T) = \Sigma(\psi^1_0,\psi^2_0;0) = \int_{\mathbb{R}^d} \psi^1_0 \,\mathrm{d}\boldsymbol{\rho}^1_0 + \int_{\mathbb{R}^d} \psi^2_0 \,\mathrm{d}\boldsymbol{\rho}^2_0}$$

Backward Kolmogorov equation

Assume for a moment that B is **bounded and smooth**. In order to estimate

$$\Sigma(\psi^1, \psi^2; T) := \int_{\mathbb{R}^d} \psi^1 \, \mathrm{d}\rho_T^1 + \int_{\mathbb{R}^d} \psi^2 \, \mathrm{d}\rho_T^2, \quad \psi^1(x_1) + \psi^2(x_2) \le h(|x_1 - x_2|),$$

we can **solve the final-value problem for the adjoint backward Kolomogorov equation**

$$\partial_t \psi^{1,2} + \Delta \psi^{1,2} - B \cdot \nabla \psi^{1,2} = 0 \quad \text{in } \mathbb{R}^d \times (0, T), \quad \psi^{1,2}(\cdot, T) := \psi^{1,2} \quad \text{(BKE)}$$

If $\psi^{1,2}$ belong to $C^{2,1}_{\mathrm{b}}(\mathbb{R}^d \times [0, T])$, the distributional formulation (FP) yields

$$\boxed{\Sigma(\psi_T^1, \psi_T^2; T) = \Sigma(\psi_0^1, \psi_0^2; 0) = \int_{\mathbb{R}^d} \psi_0^1 \, \mathrm{d}\rho_0^1 + \int_{\mathbb{R}^d} \psi_0^2 \, \mathrm{d}\rho_0^2}$$

If we could prove that ψ_0^1, ψ_0^2 satisfy the constraint

$$\psi_0^1(x_1) + \psi_0^2(x_2) \le h(|x_1 - x_2|),$$

Backward Kolmogorov equation

Assume for a moment that B is **bounded and smooth**. In order to estimate

$$\Sigma(\psi^1,\psi^2;T) := \int_{\mathbb{R}^d}\psi^1\,\mathrm{d}\rho^1_T + \int_{\mathbb{R}^d}\psi^2\,\mathrm{d}\rho^2_T,\quad \psi^1(x_1)+\psi^2(x_2)\le h(|x_1-x_2|),$$

we can **solve the final-value problem for the adjoint backward Kolomogorov equation**

$$\partial_t\psi^{1,2}+\Delta\psi^{1,2}-B\cdot\nabla\psi^{1,2}=0\quad\text{in }\mathbb{R}^d\times(0,T),\quad \psi^{1,2}(\cdot,T):=\psi^{1,2}\quad\text{(BKE)}$$

If $\psi^{1,2}$ belong to $C^{2,1}_{\mathrm{b}}(\mathbb{R}^d\times[0,T])$, the distributional formulation (FP) yields

$$\boxed{\Sigma(\psi^1_T,\psi^2_T;T)=\Sigma(\psi^1_0,\psi^2_0;0)=\int_{\mathbb{R}^d}\psi^1_0\,\mathrm{d}\rho^1_0+\int_{\mathbb{R}^d}\psi^2_0\,\mathrm{d}\rho^2_0}$$

If we could prove that ψ^1_0,ψ^2_0 satisfy the constraint

$$\psi^1_0(x_1)+\psi^2_0(x_2)\le h(|x_1-x_2|),$$

then we would get the *uniform bound in term of the initial cost*

$$\Sigma(\psi^1_T,\psi^2_T;T)=\Sigma(\psi^1_0,\psi^2_0;0)=\int_{\mathbb{R}^d}\psi^1_0\,\mathrm{d}\rho^1_0+\int_{\mathbb{R}^d}\psi^2_0\,\mathrm{d}\rho^2_0\le\mathcal{C}_h(\rho^1_0,\rho^2_0).$$

4 regularization steps

- **Regularization of h**: we can assume that

$$h \text{ is } C^1, \text{ Lipschitz}, \lim_{r\uparrow+\infty} h(r) = +\infty, \text{ and } \mathcal{C}_h(\rho_0^1, \rho_0^2) < +\infty$$

4 regularization steps

- **Regularization of h**: we can assume that

$$h \text{ is } C^1, \text{ Lipschitz, } \lim_{r\uparrow+\infty} h(r) = +\infty, \text{ and } \mathcal{C}_h(\rho_0^1, \rho_0^2) < +\infty$$

Yosida regularization, convolution, Γ-convergence for the cost functional

4 regularization steps

- **Regularization of h**: we can assume that

$$h \text{ is } C^1, \text{ Lipschitz}, \lim_{r\uparrow+\infty} h(r) = +\infty, \text{ and } \mathcal{C}_h(\rho_0^1, \rho_0^2) < +\infty$$

Yosida regularization, convolution, Γ-convergence for the cost functional

- ψ^1, ψ^2 can be assumed **smooth with compact support**:

$$\mathcal{C}_h(\rho^1, \rho^2) = \sup\Big\{ \int_{\mathbb{R}^d} \psi^1(\mathsf{x}_1)\,\mathrm{d}\rho^1(\mathsf{x}_1) + \int_{\mathbb{R}^d} \psi^2(\mathsf{x}_2)\,\mathrm{d}\rho^2(\mathsf{x}_2) : \psi^1, \psi^2 \in C_c^\infty(\mathbb{R}^d), \quad \psi^1(\mathsf{x}_1) + \psi^2(\mathsf{x}_2) \le h(|\mathsf{x}_1 - \mathsf{x}_2|)\Big\}$$

4 regularization steps

- **Regularization of** h: we can assume that

$$h \text{ is } C^1, \text{ Lipschitz}, \lim_{r\uparrow+\infty} h(r) = +\infty, \text{ and } \mathcal{C}_h(\rho_0^1, \rho_0^2) < +\infty$$

Yosida regularization, convolution, Γ-convergence for the cost functional

- ψ^1, ψ^2 can be assumed **smooth with compact support**:

$$\mathcal{C}_h(\rho^1, \rho^2) = \sup \Big\{ \int_{\mathbb{R}^d} \psi^1(\mathsf{x}_1)\,\mathrm{d}\rho^1(\mathsf{x}_1) + \int_{\mathbb{R}^d} \psi^2(\mathsf{x}_2)\,\mathrm{d}\rho^2(\mathsf{x}_2) : \psi^1, \psi^2 \in C_\mathrm{c}^\infty(\mathbb{R}^d), \quad \psi^1(\mathsf{x}_1) + \psi^2(\mathsf{x}_2) \le h(|\mathsf{x}_1 - \mathsf{x}_2|) \Big\}$$

- **Rescaling argument**: it is sufficient to consider the case $\lambda = 0$.

4 regularization steps

- **Regularization of h**: we can assume that

$$h \text{ is } C^1, \text{ Lipschitz}, \lim_{r\uparrow+\infty} h(r) = +\infty, \text{ and } \mathcal{C}_h(\rho_0^1, \rho_0^2) < +\infty$$

Yosida regularization, convolution, Γ-convergence for the cost functional

- ψ^1, ψ^2 can be assumed **smooth with compact support**:

$$\mathcal{C}_h(\rho^1, \rho^2) = \sup\Big\{ \int_{\mathbb{R}^d} \psi^1(x_1)\,\mathrm{d}\rho^1(x_1) + \int_{\mathbb{R}^d} \psi^2(x_2)\,\mathrm{d}\rho^2(x_2) : \psi^1, \psi^2 \in C_c^\infty(\mathbb{R}^d), \quad \psi^1(x_1) + \psi^2(x_2) \le h(|x_1 - x_2|)\Big\}$$

- **Rescaling argument**: it is sufficient to consider the case $\lambda = 0$.
  The time-rescaled measures

$$\mathrm{t}(s) := \frac{1}{2\lambda}\log(1 + 2\lambda s), \quad \tilde\rho_s^i(A) := \rho_{\mathrm{t}(s)}(\mathrm{e}^{-\lambda \mathrm{t}(s)} A)$$

solve a (FP)-equation with a monotone (time dependent) drift.

4 regularization steps

- **Regularization of h**: we can assume that

$$h \text{ is } C^1, \text{ Lipschitz, } \lim_{r\uparrow+\infty} h(r) = +\infty, \text{ and } \mathcal{C}_h(\rho_0^1, \rho_0^2) < +\infty$$

Yosida regularization, convolution, Γ-convergence for the cost functional

- ψ^1, ψ^2 can be assumed **smooth with compact support**:

$$\mathcal{C}_h(\rho^1, \rho^2) = \sup \Big\{ \int_{\mathbb{R}^d} \psi^1(x_1)\,\mathrm{d}\rho^1(x_1) + \int_{\mathbb{R}^d} \psi^2(x_2)\,\mathrm{d}\rho^2(x_2) : \psi^1, \psi^2 \in C_c^\infty(\mathbb{R}^d), \quad \psi^1(x_1) + \psi^2(x_2) \le h(|x_1 - x_2|) \Big\}$$

- **Rescaling argument**: it is sufficient to consider the case $\lambda = 0$.
  The time-rescaled measures

$$\mathrm{t}(s) := \frac{1}{2\lambda} \log(1 + 2\lambda s), \quad \tilde{\rho}_s^i(A) := \rho_{\mathrm{t}(s)}(\mathrm{e}^{-\lambda \mathrm{t}(s)} A)$$

  solve a (FP)-equation with a monotone (time dependent) drift.

- B can be assumed **bounded and smooth.**

4 regularization steps

- **Regularization of h**: we can assume that

$$h \text{ is } C^1, \text{ Lipschitz, } \lim_{r\uparrow+\infty} h(r) = +\infty, \text{ and } \mathcal{C}_h(\rho_0^1, \rho_0^2) < +\infty$$

Yosida regularization, convolution, Γ-convergence for the cost functional

- ψ^1, ψ^2 can be assumed **smooth with compact support**:

$$\mathcal{C}_h(\rho^1, \rho^2) = \sup\Big\{ \int_{\mathbb{R}^d} \psi^1(x_1)\,\mathrm{d}\rho^1(x_1) + \int_{\mathbb{R}^d} \psi^2(x_2)\,\mathrm{d}\rho^2(x_2) : \\ \psi^1, \psi^2 \in C_c^\infty(\mathbb{R}^d), \quad \psi^1(x_1) + \psi^2(x_2) \le h(|x_1 - x_2|)\Big\}$$

- **Rescaling argument**: it is sufficient to consider the case $\lambda = 0$.
The time-rescaled measures

$$\mathrm{t}(s) := \frac{1}{2\lambda}\log(1+2\lambda s), \quad \tilde\rho^i_s(A) := \rho_{\mathrm{t}(s)}(\mathrm{e}^{-\lambda \mathrm{t}(s)} A)$$

solve a (FP)-equation with a monotone (time dependent) drift.

- B can be assumed **bounded and smooth.**
Combine the Fitzpatrick-Phelps approximation of B by bounded montone maps with convolution.

A comparison result for the backward equation

Theorem

Let

- $h : [0, +\infty) \to \mathbb{R}$ *be* C^1 *non-decreasing function,*

A comparison result for the backward equation

Theorem

Let

- $h : [0, +\infty) \longrightarrow \mathbb{R}$ *be* C^1 *non-decreasing function,*
- $B_t : \mathbb{R}^d \longrightarrow \mathbb{R}^d$ *be a smooth family of bounded and monotone operators*

A comparison result for the backward equation

Theorem

Let

- $h : [0, +\infty) \to \mathbb{R}$ *be* C^1 *non-decreasing function,*
- $B_t : \mathbb{R}^d \to \mathbb{R}^d$ *be a smooth family of bounded and monotone operators*
- $\psi^1, \psi^2 \in C^{2,1}_{\mathrm{b}}(\mathbb{R}^d \times [0, T])$ *be solutions of the "backward" inequality*

$$\partial_t \psi + \Delta \psi - B_t \cdot \nabla \psi \geq 0 \quad \text{in } \mathbb{R}^d \times [0, T]$$

such that

$$\psi^1(\mathsf{x}_1, T) + \psi^2(\mathsf{x}_2, T) \leq h(|\mathsf{x}_1 - \mathsf{x}_2|) \quad \text{for every } \mathsf{x}_1, \mathsf{x}_2 \in \mathbb{R}^d.$$

A comparison result for the backward equation

Theorem

Let

- $h : [0, +\infty) \to \mathbb{R}$ *be* C^1 *non-decreasing function,*
- $B_t : \mathbb{R}^d \to \mathbb{R}^d$ *be a smooth family of bounded and monotone operators*
- $\psi^1, \psi^2 \in C^{2,1}_{\mathrm{b}}(\mathbb{R}^d \times [0, T])$ *be solutions of the "backward" inequality*

$$\partial_t \psi + \Delta \psi - B_t \cdot \nabla \psi \geq 0 \quad \text{in } \mathbb{R}^d \times [0, T]$$

such that

$$\psi^1(\mathbf{x_1}, T) + \psi^2(\mathbf{x_2}, T) \leq h(|\mathbf{x_1} - \mathbf{x_2}|) \quad \text{for every } \mathbf{x_1}, \mathbf{x_2} \in \mathbb{R}^d.$$

Then

$$\boxed{\psi^1(\mathbf{x_1}, 0) + \psi^2(\mathbf{x_2}, 0) \leq h(|\mathbf{x_1} - \mathbf{x_2}|) \quad \text{for every } \mathbf{x_1}, \mathbf{x_2} \in \mathbb{R}^d.}$$

The maximum-principle argument

If ψ^1, ψ^2 satisfy the *strict* inequality $\boxed{\partial_t\psi + L\psi > 0}$, $L\psi := \Delta\psi - B\cdot\nabla\psi$,
then the function

$$f(\mathbf{x}_1, \mathbf{x}_2, t) := \psi^1(\mathbf{x}_1, s) + \psi^2(\mathbf{x}_2, t) - h(|\mathbf{x}_1 - \mathbf{x}_2|)$$

cannot attain a (local) maximum in a point $(\bar{\mathbf{x}}_1, \bar{\mathbf{x}}_2, \bar{t})$ with $\bar{t} < T$.

The maximum-principle argument

If ψ^1, ψ^2 satisfy the *strict* inequality $\boxed{\partial_t\psi + L\psi > 0}$, $L\psi := \Delta\psi - B\cdot\nabla\psi$, then the function

$$f(\mathbf{x_1},\mathbf{x_2},t) := \psi^1(\mathbf{x_1},s) + \psi^2(\mathbf{x_2},t) - h(|\mathbf{x_1}-\mathbf{x_2}|)$$

cannot attain a (local) maximum in a point $(\bar{\mathbf{x}}_1,\bar{\mathbf{x}}_2,\bar t)$ with $\bar t < T$.

By contradiction, if $(\bar{\mathbf{x}}_1,\bar{\mathbf{x}}_2,\bar t)$ is a **local maximizer of f** with $\bar t < T$ then

$$\partial_t f(\bar{\mathbf{x}}_1,\bar{\mathbf{x}}_2,\bar t) \le 0, \quad \nabla_{\mathbf{x_1}} f(\bar{\mathbf{x}}_1,\bar{\mathbf{x}}_2,\bar t) = 0, \quad \nabla_{\mathbf{x_2}} f(\bar{\mathbf{x}}_1,\bar{\mathbf{x}}_2,\bar t) = 0;$$

The maximum-principle argument

If ψ^1, ψ^2 satisfy the *strict* inequality $\boxed{\partial_t\psi + L\psi > 0}$, $L\psi := \Delta\psi - B\cdot\nabla\psi$,
then the function

$$f(\mathbf{x}_1, \mathbf{x}_2, t) := \psi^1(\mathbf{x}_1, s) + \psi^2(\mathbf{x}_2, t) - h(|\mathbf{x}_1 - \mathbf{x}_2|)$$

cannot attain a (local) maximum in a point $(\bar{\mathbf{x}}_1, \bar{\mathbf{x}}_2, \bar{t})$ with $\bar{t} < T$.
By contradiction, if $(\bar{\mathbf{x}}_1, \bar{\mathbf{x}}_2, \bar{t})$ is a **local maximizer of f** with $\bar{t} < T$ then

$$\partial_t f(\bar{\mathbf{x}}_1, \bar{\mathbf{x}}_2, \bar{t}) \le 0, \quad \nabla_{\mathbf{x}_1} f(\bar{\mathbf{x}}_1, \bar{\mathbf{x}}_2, \bar{t}) = 0, \quad \nabla_{\mathbf{x}_2} f(\bar{\mathbf{x}}_1, \bar{\mathbf{x}}_2, \bar{t}) = 0;$$

$$\boxed{\partial_t\psi^1(\bar{\mathbf{x}}_1, \bar{t}) + \partial_t\psi^2(\bar{\mathbf{x}}_2, \bar{t}) \le 0}$$

The maximum-principle argument

If ψ^1, ψ^2 satisfy the *strict* inequality $\boxed{\partial_t\psi + L\psi > 0}$, $L\psi := \Delta\psi - B\cdot\nabla\psi$, then the function

$$f(\mathbf{x}_1, \mathbf{x}_2, t) := \psi^1(\mathbf{x}_1, s) + \psi^2(\mathbf{x}_2, t) - h(|\mathbf{x}_1 - \mathbf{x}_2|)$$

cannot attain a (local) maximum in a point $(\bar{\mathbf{x}}_1, \bar{\mathbf{x}}_2, \bar{t})$ with $\bar{t} < T$.
By contradiction, if $(\bar{\mathbf{x}}_1, \bar{\mathbf{x}}_2, \bar{t})$ is a **local maximizer of f** with $\bar{t} < T$ then

$$\partial_t f(\bar{\mathbf{x}}_1, \bar{\mathbf{x}}_2, \bar{t}) \le 0, \quad \nabla_{\mathbf{x}_1} f(\bar{\mathbf{x}}_1, \bar{\mathbf{x}}_2, \bar{t}) = 0, \quad \nabla_{\mathbf{x}_2} f(\bar{\mathbf{x}}_1, \bar{\mathbf{x}}_2, \bar{t}) = 0;$$

$$\boxed{\partial_t\psi^1(\bar{\mathbf{x}}_1, \bar{t}) + \partial_t\psi^2(\bar{\mathbf{x}}_2, \bar{t}) \le 0}$$

$$\nabla_{\mathbf{x}_1}\psi^1(\bar{\mathbf{x}}_1, \bar{t}) = \bar{h}\,(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2) = -\nabla_{\mathbf{x}_2}\psi^2(\bar{\mathbf{x}}_2, \bar{t}), \quad \bar{h} := \frac{h'(|\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2|)}{|\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2|} \ge 0$$

The maximum-principle argument

If ψ^1, ψ^2 satisfy the *strict* inequality $\boxed{\partial_t\psi + L\psi > 0}$, $L\psi := \Delta\psi - B\cdot\nabla\psi$, then the function

$$f(\mathbf{x}_1, \mathbf{x}_2, t) := \psi^1(\mathbf{x}_1, s) + \psi^2(\mathbf{x}_2, t) - h(|\mathbf{x}_1 - \mathbf{x}_2|)$$

cannot attain a (local) maximum in a point $(\bar{\mathbf{x}}_1, \bar{\mathbf{x}}_2, \bar{t})$ with $\bar{t} < T$.
By contradiction, if $(\bar{\mathbf{x}}_1, \bar{\mathbf{x}}_2, \bar{t})$ is a **local maximizer of f** with $\bar{t} < T$ then

$$\partial_t f(\bar{\mathbf{x}}_1, \bar{\mathbf{x}}_2, \bar{t}) \le 0, \quad \nabla_{\mathbf{x}_1} f(\bar{\mathbf{x}}_1, \bar{\mathbf{x}}_2, \bar{t}) = 0, \quad \nabla_{\mathbf{x}_2} f(\bar{\mathbf{x}}_1, \bar{\mathbf{x}}_2, \bar{t}) = 0;$$

$$\boxed{\partial_t\psi^1(\bar{\mathbf{x}}_1, \bar{t}) + \partial_t\psi^2(\bar{\mathbf{x}}_2, \bar{t}) \le 0}$$

$$\nabla_{\mathbf{x}_1}\psi^1(\bar{\mathbf{x}}_1, \bar{t}) = \bar{h}\,(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2) = -\nabla_{\mathbf{x}_2}\psi^2(\bar{\mathbf{x}}_2, \bar{t}), \quad \bar{h} := \frac{h'(|\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2|)}{|\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2|} \ge 0$$

It follows that

$$\boxed{B(\bar{\mathbf{x}}_1)\cdot\nabla_{\mathbf{x}_1}\psi^1(\bar{\mathbf{x}}_1, \bar{t}) + B(\bar{\mathbf{x}}_2)\cdot\nabla_{\mathbf{x}_2}\psi^2(\bar{\mathbf{x}}_2, \bar{t}) = \bar{h}\,\big(B(\bar{\mathbf{x}}_1) - B(\bar{\mathbf{x}}_2)\big)\cdot(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2) \ge 0}$$

The maximum-principle argument

If ψ^1, ψ^2 satisfy the *strict* inequality $\boxed{\partial_t\psi + L\psi > 0}$, $L\psi := \Delta\psi - B\cdot\nabla\psi$, then the function

$$f(\mathbf{x}_1, \mathbf{x}_2, t) := \psi^1(\mathbf{x}_1, s) + \psi^2(\mathbf{x}_2, t) - h(|\mathbf{x}_1 - \mathbf{x}_2|)$$

cannot attain a (local) maximum in a point $(\bar{\mathbf{x}}_1, \bar{\mathbf{x}}_2, \bar{t})$ with $\bar{t} < T$.

By contradiction, if $(\bar{\mathbf{x}}_1, \bar{\mathbf{x}}_2, \bar{t})$ is a **local maximizer of f** with $\bar{t} < T$ then

$$\partial_t f(\bar{\mathbf{x}}_1, \bar{\mathbf{x}}_2, \bar{t}) \le 0, \quad \nabla_{\mathbf{x}_1} f(\bar{\mathbf{x}}_1, \bar{\mathbf{x}}_2, \bar{t}) = 0, \quad \nabla_{\mathbf{x}_2} f(\bar{\mathbf{x}}_1, \bar{\mathbf{x}}_2, \bar{t}) = 0;$$

$$\boxed{\partial_t\psi^1(\bar{\mathbf{x}}_1, \bar{t}) + \partial_t\psi^2(\bar{\mathbf{x}}_2, \bar{t}) \le 0}$$

$$\nabla_{\mathbf{x}_1}\psi^1(\bar{\mathbf{x}}_1, \bar{t}) = \bar{h}\,(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2) = -\nabla_{\mathbf{x}_2}\psi^2(\bar{\mathbf{x}}_2, \bar{t}), \quad \bar{h} := \frac{h'(|\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2|)}{|\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2|} \ge 0$$

It follows that

$$\boxed{B(\bar{\mathbf{x}}_1)\cdot\nabla_{\mathbf{x}_1}\psi^1(\bar{\mathbf{x}}_1, \bar{t}) + B(\bar{\mathbf{x}}_2)\cdot\nabla_{\mathbf{x}_2}\psi^2(\bar{\mathbf{x}}_2, \bar{t}) = \bar{h}\left(B(\bar{\mathbf{x}}_1) - B(\bar{\mathbf{x}}_2)\right)\cdot(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2) \ge 0}$$

On the other hand, the function

$$\mathbb{R}^d \ni z \mapsto \psi^1(\bar{\mathbf{x}}_1 + z, \bar{t}) + \psi^2(\bar{\mathbf{x}}_2 + z, \bar{t}) - H(\bar{\mathbf{x}}_1, \bar{\mathbf{x}}_2) = f(\bar{\mathbf{x}}_1 + z, \bar{\mathbf{x}}_2 + z, \bar{t})$$

has a local maximum at $z = 0$ so that

The maximum-principle argument

If ψ^1, ψ^2 satisfy the *strict* inequality $\boxed{\partial_t\psi + L\psi > 0}$, $L\psi := \Delta\psi - B\cdot\nabla\psi$, then the function

$$f(\mathsf{x}_1, \mathsf{x}_2, t) := \psi^1(\mathsf{x}_1, s) + \psi^2(\mathsf{x}_2, t) - h(|\mathsf{x}_1 - \mathsf{x}_2|)$$

cannot attain a (local) maximum in a point $(\bar{\mathsf{x}}_1, \bar{\mathsf{x}}_2, \bar t)$ with $\bar t < T$.

By contradiction, if $(\bar{\mathsf{x}}_1, \bar{\mathsf{x}}_2, \bar t)$ is a **local maximizer of f** with $\bar t < T$ then

$$\partial_t f(\bar{\mathsf{x}}_1, \bar{\mathsf{x}}_2, \bar t) \le 0, \quad \nabla_{\mathsf{x}_1} f(\bar{\mathsf{x}}_1, \bar{\mathsf{x}}_2, \bar t) = 0, \quad \nabla_{\mathsf{x}_2} f(\bar{\mathsf{x}}_1, \bar{\mathsf{x}}_2, \bar t) = 0;$$

$$\boxed{\partial_t\psi^1(\bar{\mathsf{x}}_1, \bar t) + \partial_t\psi^2(\bar{\mathsf{x}}_2, \bar t) \le 0}$$

$$\nabla_{\mathsf{x}_1}\psi^1(\bar{\mathsf{x}}_1, \bar t) = \bar h\,(\bar{\mathsf{x}}_1 - \bar{\mathsf{x}}_2) = -\nabla_{\mathsf{x}_2}\psi^2(\bar{\mathsf{x}}_2, \bar t), \quad \bar h := \frac{h'(|\bar{\mathsf{x}}_1 - \bar{\mathsf{x}}_2|)}{|\bar{\mathsf{x}}_1 - \bar{\mathsf{x}}_2|} \ge 0$$

It follows that

$$\boxed{B(\bar{\mathsf{x}}_1)\cdot\nabla_{\mathsf{x}_1}\psi^1(\bar{\mathsf{x}}_1, \bar t) + B(\bar{\mathsf{x}}_2)\cdot\nabla_{\mathsf{x}_2}\psi^2(\bar{\mathsf{x}}_2, \bar t) = \bar h\,\big(B(\bar{\mathsf{x}}_1) - B(\bar{\mathsf{x}}_2)\big)\cdot(\bar{\mathsf{x}}_1 - \bar{\mathsf{x}}_2) \ge 0}$$

On the other hand, the function

$$\mathbb{R}^d \ni z \mapsto \psi^1(\bar{\mathsf{x}}_1 + z, \bar t) + \psi^2(\bar{\mathsf{x}}_2 + z, \bar t) - H(\bar{\mathsf{x}}_1, \bar{\mathsf{x}}_2) = f(\bar{\mathsf{x}}_1 + z, \bar{\mathsf{x}}_2 + z, \bar t)$$

has a local maximum at $z = 0$ so that $\boxed{\Delta_{\mathsf{x}_1}\psi^1(\bar{\mathsf{x}}_1, \bar t) + \Delta_{\mathsf{x}_2}\psi^2(\bar{\mathsf{x}}_2, \bar t) \le 0.}$

The maximum-principle argument

If ψ^1, ψ^2 satisfy the *strict* inequality $\boxed{\partial_t \psi + L\psi > 0}$, $L\psi := \Delta\psi - B\cdot\nabla\psi$, then the function

$$f(\mathbf{x}_1, \mathbf{x}_2, t) := \psi^1(\mathbf{x}_1, s) + \psi^2(\mathbf{x}_2, t) - h(|\mathbf{x}_1 - \mathbf{x}_2|)$$

cannot attain a (local) maximum in a point $(\bar{\mathbf{x}}_1, \bar{\mathbf{x}}_2, \bar{t})$ with $\bar{t} < T$.

By contradiction, if $(\bar{\mathbf{x}}_1, \bar{\mathbf{x}}_2, \bar{t})$ is a **local maximizer of f** with $\bar{t} < T$ then

$$\partial_t f(\bar{\mathbf{x}}_1, \bar{\mathbf{x}}_2, \bar{t}) \le 0, \quad \nabla_{\mathbf{x}_1} f(\bar{\mathbf{x}}_1, \bar{\mathbf{x}}_2, \bar{t}) = 0, \quad \nabla_{\mathbf{x}_2} f(\bar{\mathbf{x}}_1, \bar{\mathbf{x}}_2, \bar{t}) = 0;$$

$$\boxed{\partial_t \psi^1(\bar{\mathbf{x}}_1, \bar{t}) + \partial_t \psi^2(\bar{\mathbf{x}}_2, \bar{t}) \le 0}$$

$$\nabla_{\mathbf{x}_1} \psi^1(\bar{\mathbf{x}}_1, \bar{t}) = \bar{h}\,(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2) = -\nabla_{\mathbf{x}_2} \psi^2(\bar{\mathbf{x}}_2, \bar{t}), \quad \bar{h} := \frac{h'(|\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2|)}{|\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2|} \ge 0$$

It follows that

$$\boxed{B(\bar{\mathbf{x}}_1)\cdot\nabla_{\mathbf{x}_1}\psi^1(\bar{\mathbf{x}}_1, \bar{t}) + B(\bar{\mathbf{x}}_2)\cdot\nabla_{\mathbf{x}_2}\psi^2(\bar{\mathbf{x}}_2, \bar{t}) = \bar{h}\,\big(B(\bar{\mathbf{x}}_1) - B(\bar{\mathbf{x}}_2)\big)\cdot(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2) \ge 0}$$

On the other hand, the function

$$\mathbb{R}^d \ni z \mapsto \psi^1(\bar{\mathbf{x}}_1 + z, \bar{t}) + \psi^2(\bar{\mathbf{x}}_2 + z, \bar{t}) - H(\bar{\mathbf{x}}_1, \bar{\mathbf{x}}_2) = f(\bar{\mathbf{x}}_1 + z, \bar{\mathbf{x}}_2 + z, \bar{t})$$

has a local maximum at $z = 0$ so that $\boxed{\Delta_{\mathbf{x}_1}\psi^1(\bar{\mathbf{x}}_1, \bar{t}) + \Delta_{\mathbf{x}_2}\psi^2(\bar{\mathbf{x}}_2, \bar{t}) \le 0.}$

Combining the previous inequality we obtain the contradiction

$$(\partial_t \psi^1 + L\psi^1)(\bar{\mathbf{x}}_1, \bar{t}) + (\partial_t \psi^2 + L\psi^2)(\bar{\mathbf{x}}_2, \bar{t}) \le 0.$$

Open problems

- Find a “natural” Wasserstein-metric notion of solution.
- Prove suitable regularization estimates
- Prove general Trotter-Kato approximation formula for the solution
- Extend the theory to operators defined in a bounded domain
- Extend the theory to the infinite-dimensional case.
- ....