Measures are fun

Introduction to the Wasserstein distance and its geometry

Averil Prost (LMI, INSA Rouen Normandie)

April 2nd, 2024
LITIS doctoral seminar

LMI - EA 3226

Table of Contents

What is a measure

An object modelling the presence of mass.

- continuous: fluids, earth, light... Typically the Lebesgue measure.

What is a measure

An object modelling the presence of mass.

- continuous: fluids, earth, light... Typically the Lebesgue measure.
- discrete: people, histograms, money... Typically Dirac masses.

What is a measure

An object modelling the presence of mass.

- continuous: fluids, earth, light... Typically the Lebesgue measure.
- discrete: people, histograms, money... Typically Dirac masses.

Def 1 – Probability measure
Let $(\Omega; \Sigma)$ be a measurable space. A probability measure μ is an application from Σ to $[0,1]$ such that $\mu(\Omega) = 1$ and for any disjoint sequence $(A_i)_{i\in\mathbb{N}}$,

$$\mu\left(\bigsqcup_{i\in\mathbb{N}} A_i\right) = \sum_{i\in\mathbb{N}} \mu(A_i).$$

Why are we interested...

Applications in

- transport (LMI): limits from peaton models to fluid models

Why are we interested...

Applications in

- transport (LMI): limits from peaton models to fluid models
- statistics, machine learning: distributions, representation of images

Why are we interested...

Applications in

- transport (LMI): limits from peaton models to fluid models
- statistics, machine learning: distributions, representation of images
- mean field games: interaction of populations

Why are we interested... ...and what is the problem

Applications in

- transport (LMI): limits from peaton models to fluid models
- statistics, machine learning: distributions, representation of images
- mean field games: interaction of populations

How to compute the distance between to measures?

Why are we interested... ...and what is the problem

Applications in

- transport (LMI): limits from peaton models to fluid models
- statistics, machine learning: distributions, representation of images
- mean field games: interaction of populations

How to compute the distance between to measures?

With the total variation

$$|\mu - \nu|_{\mathsf{TV}} = \sup_{\substack{\mathcal{A} \in \text{ countable measurable} \\ \text{partitions of } \Omega}} \sum_{A \in \mathcal{A}} |\mu(A) - \nu(A)|,$$

we would have $|\delta_0 - \delta_t|_{\mathsf{TV}} = 2$ regardless of how close t is to 0.

Why are we interested... ...and what is the problem

Applications in

- transport (LMI): limits from peaton models to fluid models
- statistics, machine learning: distributions, representation of images
- mean field games: interaction of populations

How to compute the distance between to measures?

With the total variation

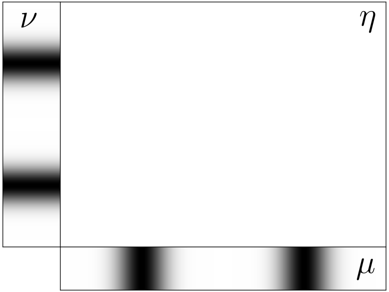
$$|\mu - \nu|_{\mathsf{TV}} = \sup_{\substack{\mathcal{A} \in \text{countable measurable} \\ \text{partitions of } \Omega}} \sum_{A \in \mathcal{A}} |\mu(A) - \nu(A)|,$$

we would have $|\delta_0 - \delta_t|_{\mathsf{TV}} = 2$ regardless of how close t is to 0. **Not physical.**

Table of Contents

Introducing the Wasserstein distance

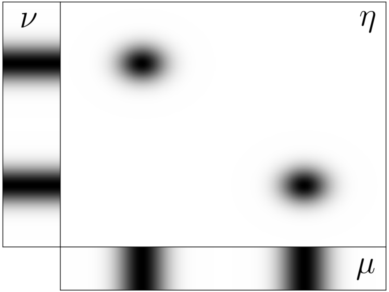
Let $\mu, \nu \in \mathscr{P}(\Omega)$ be two probability measures,



Introducing the Wasserstein distance

Let $\mu, \nu \in \mathscr{P}(\Omega)$ be two probability measures, and denote the set of transport plans by

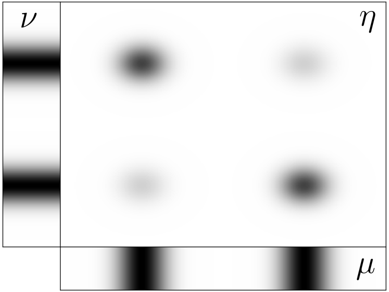
$$\Gamma(\mu,\nu) := \left\{\eta \in \mathscr{P}((\Omega)^2) \;\middle|\; \pi_x\#\eta = \mu,\ \pi_y\#\eta = \nu\right\},$$



Introducing the Wasserstein distance

Let $\mu, \nu \in \mathscr{P}(\Omega)$ be two probability measures, and denote the set of transport plans by

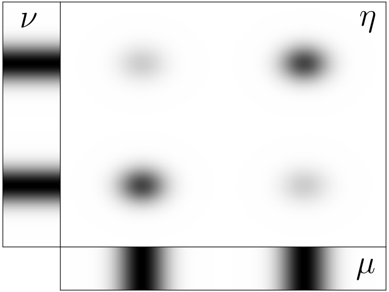
$$\Gamma(\mu,\nu) := \left\{\eta \in \mathscr{P}((\Omega)^2) \;\middle|\; \pi_x \# \eta = \mu,\, \pi_y \# \eta = \nu\right\},$$



Introducing the Wasserstein distance

Let $\mu, \nu \in \mathscr{P}(\Omega)$ be two probability measures, and denote the set of transport plans by

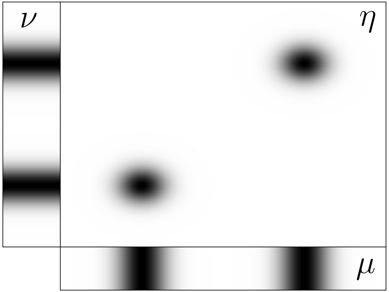
$$\Gamma(\mu, \nu) := \left\{\eta \in \mathscr{P}((\Omega)^2) \;\middle|\; \pi_x \# \eta = \mu,\ \pi_y \# \eta = \nu\right\},$$



Introducing the Wasserstein distance

Let $\mu, \nu \in \mathscr{P}(\Omega)$ be two probability measures, and denote the set of transport plans by

$$\Gamma(\mu, \nu) := \left\{\eta \in \mathscr{P}((\Omega)^2) \mid \pi_x \# \eta = \mu,\ \pi_y \# \eta = \nu\right\},$$



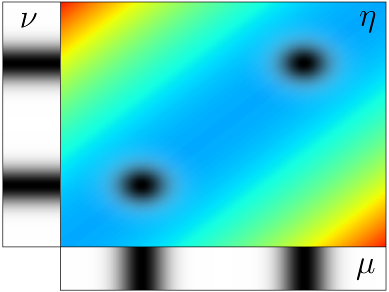
Introducing the Wasserstein distance

Let $\mu, \nu \in \mathscr{P}(\Omega)$ be two probability measures, and denote the set of transport plans by

$$\Gamma(\mu,\nu) := \left\{\eta \in \mathscr{P}((\Omega)^2) \;\middle|\; \pi_x \# \eta = \mu,\ \pi_y \# \eta = \nu\right\},$$

the squared Wasserstein distance by

$$d^2_{\mathcal{W}}(\mu,\nu) := \inf_{\eta \in \Gamma(\mu,\nu)} \int_{(x,y)} |x-y|^2 \, d\eta(x,y).$$



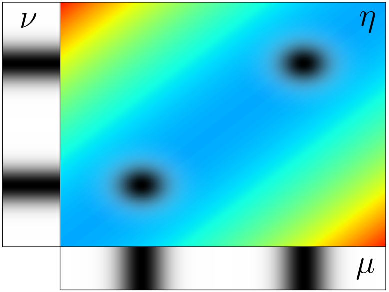
Introducing the Wasserstein distance

Let $\mu, \nu \in \mathscr{P}(\Omega)$ be two probability measures, and denote the set of transport plans by

$$\Gamma(\mu,\nu) := \left\{\eta \in \mathscr{P}((\Omega)^2) \;\middle|\; \pi_x\#\eta = \mu,\ \pi_y\#\eta = \nu\right\},$$

the squared Wasserstein distance by

$$d_{\mathcal{W}}^2(\mu,\nu) := \inf_{\eta\in\Gamma(\mu,\nu)} \int_{(x,y)} |x-y|^2 \, d\eta(x,y).$$



Def 2 – [San15] We call **Wasserstein space** the set $\mathscr{P}_2(\Omega)$ or measures μ such that $d_{\mathcal{W}}(\mu,\delta_0)$ is finite, endowed with the distance $d_{\mathcal{W}}$.

What does it look like

The Wasserstein space is

- a metric space, that is complete.

What does it look like

The Wasserstein space is

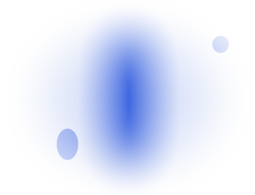
- a metric space, that is complete.
- Infinite-dimensional. If Ω is compact, then $\mathscr{P}_2(\Omega)$ is compact too: otherwise, *not even locally compact*.

What does it look like

The Wasserstein space is

- a metric space, that is complete.
- Infinite-dimensional. If Ω is compact, then $\mathscr{P}_2(\Omega)$ is compact too: otherwise, *not even locally compact*.
- Geodesic: between any couple $(\mu_0, \mu_1) \in (\mathscr{P}_2(\Omega))^2$, there exists *at least* one curve $s \mapsto \mu_s$ such that

$$d_{\mathcal{W}}(\mu_s, \mu_t) = |s - t| \, d_{\mathcal{W}}(\mu_0, \mu_1).$$



What does it look like

The Wasserstein space is

- a metric space, that is complete.
- Infinite-dimensional. If Ω is compact, then $\mathscr{P}_2(\Omega)$ is compact too: otherwise, *not even locally compact*.
- Geodesic: between any couple $(\mu_0, \mu_1) \in (\mathscr{P}_2(\Omega))^2$, there exists *at least* one curve $s \mapsto \mu_s$ such that

$$d_{\mathcal{W}}(\mu_s, \mu_t) = |s - t|\, d_{\mathcal{W}}(\mu_0, \mu_1).$$

What does it look like

The Wasserstein space is

- a metric space, that is complete.
- Infinite-dimensional. If Ω is compact, then $\mathscr{P}_2(\Omega)$ is compact too: otherwise, *not even locally compact*.
- Geodesic: between any couple $(\mu_0, \mu_1) \in (\mathscr{P}_2(\Omega))^2$, there exists *at least* one curve $s \mapsto \mu_s$ such that

$$d_{\mathcal{W}}(\mu_s, \mu_t) = |s - t|\, d_{\mathcal{W}}(\mu_0, \mu_1).$$

What does it look like

The Wasserstein space is

- a metric space, that is complete.
- Infinite-dimensional. If Ω is compact, then $\mathscr{P}_2(\Omega)$ is compact too: otherwise, *not even locally compact*.
- Geodesic: between any couple $(\mu_0, \mu_1) \in (\mathscr{P}_2(\Omega))^2$, there exists *at least* one curve $s \mapsto \mu_s$ such that

$$d_{\mathcal{W}}(\mu_s, \mu_t) = |s - t| \, d_{\mathcal{W}}(\mu_0, \mu_1).$$

What does it look like

The Wasserstein space is

- a metric space, that is complete.
- Infinite-dimensional. If Ω is compact, then $\mathscr{P}_2(\Omega)$ is compact too: otherwise, *not even locally compact*.
- Geodesic: between any couple $(\mu_0, \mu_1) \in (\mathscr{P}_2(\Omega))^2$, there exists *at least* one curve $s \mapsto \mu_s$ such that

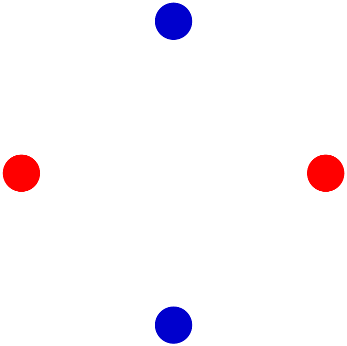
$$d_{\mathcal{W}}(\mu_s, \mu_t) = |s - t|\, d_{\mathcal{W}}(\mu_0, \mu_1).$$

What does it look like

The Wasserstein space is

- a metric space, that is complete.
- Infinite-dimensional. If Ω is compact, then $\mathscr{P}_2(\Omega)$ is compact too: otherwise, *not even locally compact*.
- Geodesic: between any couple $(\mu_0, \mu_1) \in (\mathscr{P}_2(\Omega))^2$, there exists *at least* one curve $s \mapsto \mu_s$ such that

$$d_{\mathcal{W}}(\mu_s, \mu_t) = |s - t|\, d_{\mathcal{W}}(\mu_0, \mu_1).$$

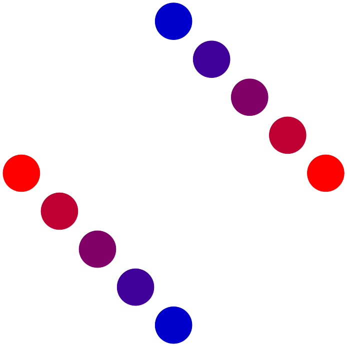


What does it look like

The Wasserstein space is

- a metric space, that is complete.
- Infinite-dimensional. If Ω is compact, then $\mathscr{P}_2(\Omega)$ is compact too: otherwise, *not even locally compact*.
- Geodesic: between any couple $(\mu_0, \mu_1) \in (\mathscr{P}_2(\Omega))^2$, there exists *at least* one curve $s \mapsto \mu_s$ such that

$$d_{\mathcal{W}}(\mu_s, \mu_t) = |s - t| \, d_{\mathcal{W}}(\mu_0, \mu_1).$$



What does it look like

The Wasserstein space is

- a metric space, that is complete.
- Infinite-dimensional. If Ω is compact, then $\mathscr{P}_2(\Omega)$ is compact too: otherwise, *not even locally compact*.
- Geodesic: between any couple $(\mu_0, \mu_1) \in (\mathscr{P}_2(\Omega))^2$, there exists *at least* one curve $s \mapsto \mu_s$ such that

$$d_{\mathcal{W}}(\mu_s, \mu_t) = |s - t| \, d_{\mathcal{W}}(\mu_0, \mu_1).$$

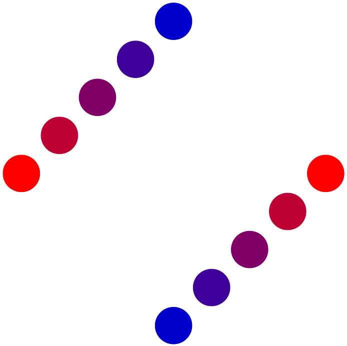


Table of Contents

Application: continuity equations

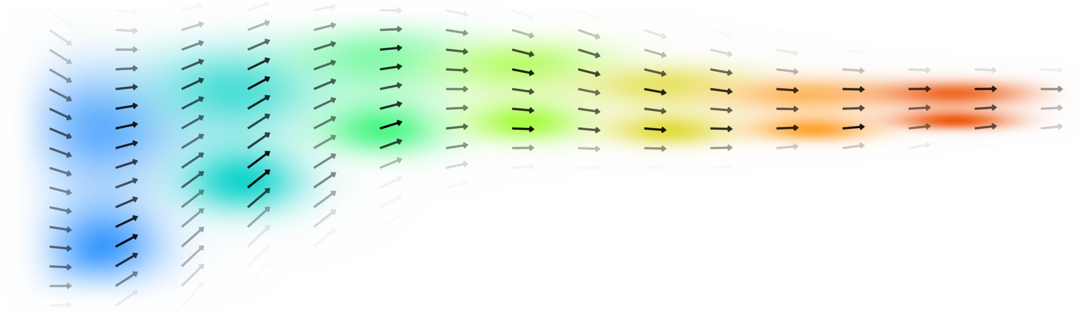
Def 3 – Fokker-Planck equation Given an initial measure $\nu \in \mathscr{P}_2(\Omega)$, find a curve $(\mu_t)_{t\in[0,T]}$ that satisfies $\mu_0 = \nu$ and solves in the sense of distributions

$$\partial_t \mu_t + \text{div}\,(f(t,x,\mu_t)\#\mu_t) = 0 \quad \forall t \in (0,T). \tag{1}$$

Application: continuity equations

Def 3 – Fokker-Planck equation Given an initial measure $\nu \in \mathscr{P}_2(\Omega)$, find a curve $(\mu_t)_{t\in[0,T]}$ that satisfies $\mu_0 = \nu$ and solves in the sense of distributions

$$\partial_t \mu_t + \operatorname{div}\left(f(t,x,\mu_t)\#\mu_t\right) = 0 \quad \forall t \in (0,T). \tag{1}$$



The Cauchy-Lipschitz framework

Theorem 1 – Well-posedness [BF23] Assume that $f : [0,T] \times \Omega \times \mathscr{P}_2(\Omega) \to \mathbb{R}^d$ is bounded and Lipschitz-continuous in all its variables (with respect to $d_{\mathcal{W}}$ for the measure variable). Then there exists an unique solution to (1), that is given by

$$\mu_t = \Phi_t^\mu \# \nu, \quad \text{where} \quad \frac{d}{dt}\Phi_t^\mu(x) = f\left(t, \Phi_t^\mu(x), \mu_t\right) \quad \text{and} \quad \Phi_0^\mu(x) = x.$$

The Cauchy-Lipschitz framework

Theorem 1 – Well-posedness [BF23] Assume that $f : [0,T] \times \Omega \times \mathscr{P}_2(\Omega) \to \mathbb{R}^d$ is bounded and Lipschitz-continuous in all its variables (with respect to $d_{\mathcal{W}}$ for the measure variable). Then there exists an unique solution to (1), that is given by

$$\mu_t = \Phi_t^\mu \# \nu, \quad \text{where} \quad \frac{d}{dt}\Phi_t^\mu(x) = f\left(t, \Phi_t^\mu(x), \mu_t\right) \quad \text{and} \quad \Phi_0^\mu(x) = x.$$

For instance,

$$f(t,x,\mu) := g(t,x) + \int_{y\in\Omega} \varphi(t,x,y)d\mu(y)$$

where $g : [0,T] \times \Omega \to \mathbb{R}^d$ and $\varphi \in \mathcal{C}_c([0,T] \times \Omega \times \Omega; \mathbb{R}^d)$ are Lipschitz in all variables.

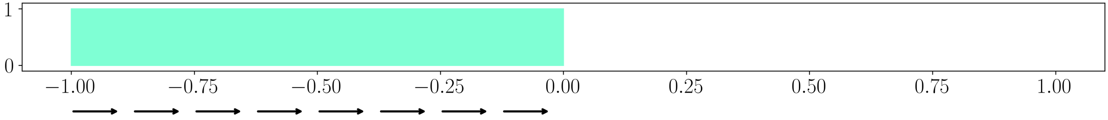
Example (1/2)

Consider the vector field f given by $f(x) = 1$ is $x < 0$, and $f(x) = 0$ if $x \geqslant 0$.

Example (1/2)

Consider the vector field f given by $f(x) = 1$ is $x < 0$, and $f(x) = 0$ if $x \geqslant 0$. Let $\nu = \mathcal{L}_{[-1,0]}$ the Lebesgue measure on the segment, and define

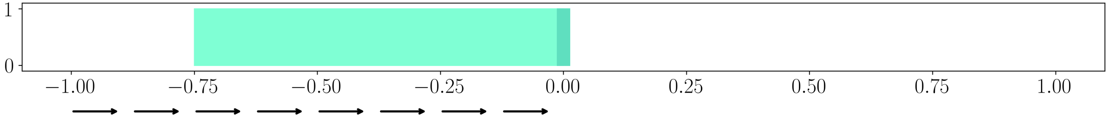
$$\mu_t = \mathcal{L}_{[(-1+t)_+,0]} + \min(t,1)\delta_0.$$



Example (1/2)

Consider the vector field f given by $f(x) = 1$ is $x < 0$, and $f(x) = 0$ if $x \geqslant 0$. Let $\nu = \mathcal{L}_{[-1,0]}$ the Lebesgue measure on the segment, and define

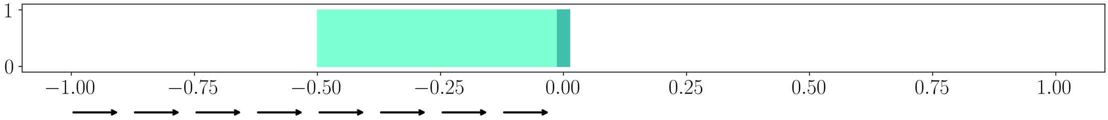
$$\mu_t = \mathcal{L}_{[(-1+t)_+,0]} + \min(t,1)\delta_0.$$



Example (1/2)

Consider the vector field f given by $f(x) = 1$ is $x < 0$, and $f(x) = 0$ if $x \geqslant 0$. Let $\nu = \mathcal{L}_{[-1,0]}$ the Lebesgue measure on the segment, and define

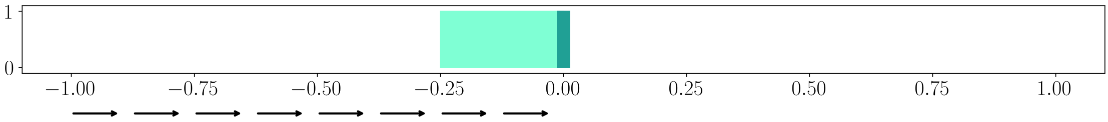
$$\mu_t = \mathcal{L}_{[(-1+t)_+,0]} + \min(t,1)\delta_0.$$



Example (1/2)

Consider the vector field f given by $f(x) = 1$ is $x < 0$, and $f(x) = 0$ if $x \geqslant 0$. Let $\nu = \mathcal{L}_{[-1,0]}$ the Lebesgue measure on the segment, and define

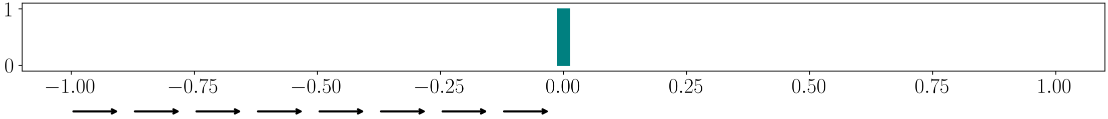
$$\mu_t = \mathcal{L}_{[(-1+t)_+,0]} + \min(t,1)\delta_0.$$



Example (1/2)

Consider the vector field f given by $f(x) = 1$ is $x < 0$, and $f(x) = 0$ if $x \geqslant 0$. Let $\nu = \mathcal{L}_{[-1,0]}$ the Lebesgue measure on the segment, and define

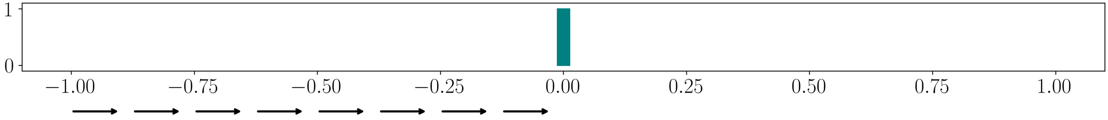
$$\mu_t = \mathcal{L}_{[(-1+t)_+,0]} + \min(t,1)\delta_0.$$



Example (1/2)

Consider the vector field f given by $f(x) = 1$ is $x < 0$, and $f(x) = 0$ if $x \geqslant 0$. Let $\nu = \mathcal{L}_{[-1,0]}$ the Lebesgue measure on the segment, and define

$$\mu_t = \mathcal{L}_{[(-1+t)_+,0]} + \min(t,1)\delta_0.$$

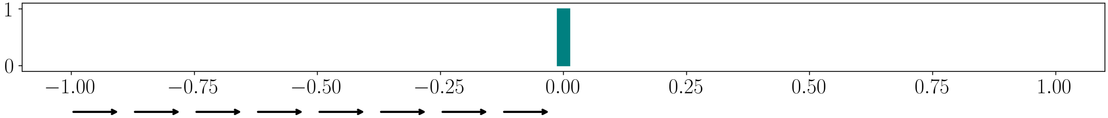


On the one hand, for any $\varphi \in \mathcal{C}^1_c(\mathbb{R};\mathbb{R})$, there holds

Example (1/2)

Consider the vector field f given by $f(x) = 1$ is $x < 0$, and $f(x) = 0$ if $x \geqslant 0$. Let $\nu = \mathcal{L}_{[-1,0]}$ the Lebesgue measure on the segment, and define

$$\mu_t = \mathcal{L}_{[(-1+t)_+,0]} + \min(t,1)\delta_0.$$



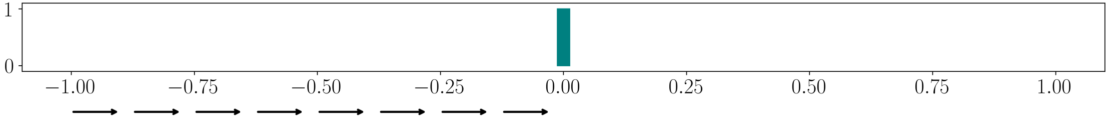
On the one hand, for any $\varphi \in \mathcal{C}^1_c(\mathbb{R};\mathbb{R})$, there holds

$$\int_{x\in\mathbb{R}} \langle \nabla\varphi(x), f(x)\rangle \, d\mu_t(x)$$

Example (1/2)

Consider the vector field f given by $f(x) = 1$ is $x < 0$, and $f(x) = 0$ if $x \geqslant 0$. Let $\nu = \mathcal{L}_{[-1,0]}$ the Lebesgue measure on the segment, and define

$$\mu_t = \mathcal{L}_{[(-1+t)_+,0]} + \min(t,1)\delta_0.$$



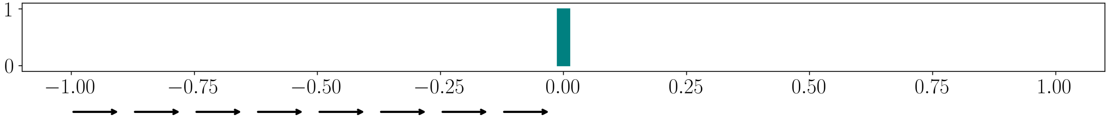
On the one hand, for any $\varphi \in \mathcal{C}_c^1(\mathbb{R};\mathbb{R})$, there holds

$$\int_{x\in\mathbb{R}} \langle \nabla\varphi(x), f(x)\rangle \, d\mu_t(x) = \int_{x=(-1+t)_+}^{0} \varphi'(x)dx$$

Example (1/2)

Consider the vector field f given by $f(x) = 1$ is $x < 0$, and $f(x) = 0$ if $x \geqslant 0$. Let $\nu = \mathcal{L}_{[-1,0]}$ the Lebesgue measure on the segment, and define

$$\mu_t = \mathcal{L}_{[(-1+t)_+,0]} + \min(t,1)\delta_0.$$



On the one hand, for any $\varphi \in \mathcal{C}^1_c(\mathbb{R};\mathbb{R})$, there holds

$$\int_{x\in\mathbb{R}} \langle \nabla\varphi(x), f(x)\rangle \, d\mu_t(x) = \int_{x=(-1+t)_+}^{0} \varphi'(x)dx = \begin{cases} \varphi(0) - \varphi(-1+t) & \text{if } t \in [0,1), \\ 0 & t \geqslant 1. \end{cases}$$

Example (2/2)

On the other hand,

$$\int_{x\in\Omega} \varphi(x) d\mu_t(x) = \min(t,1)\varphi(0) + \int_{s=(-1+t)_+}^{0} \varphi(x)dx,$$

Example (2/2)

On the other hand,

$$\int_{x\in\Omega}\varphi(x)d\mu_t(x)=\min(t,1)\varphi(0)+\int_{s=(-1+t)_+}^{0}\varphi(x)dx,$$

thus

$$\partial_t\int_{x\in\Omega}\varphi(x)d\mu_t(x)=\begin{cases}\varphi(0)-\varphi(-1+t) & t\in[0,1),\\ 0 & t\geqslant 1.\end{cases}$$

Example (2/2)

On the other hand,

$$\int_{x\in\Omega} \varphi(x) d\mu_t(x) = \min(t,1)\varphi(0) + \int_{s=(-1+t)_+}^{0} \varphi(x) dx,$$

thus

$$\partial_t \int_{x\in\Omega} \varphi(x) d\mu_t(x) = \begin{cases} \varphi(0) - \varphi(-1+t) & t \in [0,1), \\ 0 & t \geqslant 1. \end{cases}$$

Equality holds, and $(\mu_t)_{t\in[0,T]}$ is a solution of the continuity equation.

Example (2/2)

On the other hand,

$$\int_{x\in\Omega}\varphi(x)d\mu_t(x)=\min(t,1)\varphi(0)+\int_{s=(-1+t)_+}^{0}\varphi(x)dx,$$

thus

$$\partial_t\int_{x\in\Omega}\varphi(x)d\mu_t(x)=\begin{cases}\varphi(0)-\varphi(-1+t) & t\in[0,1),\\ 0 & t\geqslant 1.\end{cases}$$

Equality holds, and $(\mu_t)_{t\in[0,T]}$ is a solution of the continuity equation.

- Not in L^1 nor in any L^p!

Example (2/2)

On the other hand,

$$\int_{x\in\Omega} \varphi(x) d\mu_t(x) = \min(t,1)\varphi(0) + \int_{s=(-1+t)_+}^{0} \varphi(x) dx,$$

thus

$$\partial_t \int_{x\in\Omega} \varphi(x) d\mu_t(x) = \begin{cases} \varphi(0) - \varphi(-1+t) & t \in [0,1), \\ 0 & t \geqslant 1. \end{cases}$$

Equality holds, and $(\mu_t)_{t\in[0,T]}$ is a solution of the continuity equation.

- Not in L^1 nor in any L^p!
- Here f is not Lipschitz-continuous, actually no available theory in this case.

Table of Contents

The Hodge-Helmholtz decomposition

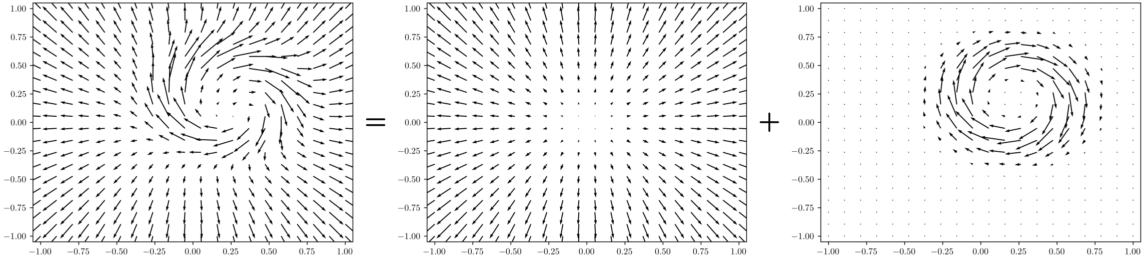
Theorem 2 – HH decomposition [Lad87] Let $f \in L^2(\mathbb{R}^d;\mathbb{R}^d)$. There exists two uniquely defined vector fields $g, h \in L^2(\mathbb{R}^d;\mathbb{R}^d)$ such that

$$f = g + h, \qquad g \simeq \nabla\varphi, \qquad \text{div}\,(h) \simeq 0. \tag{2}$$

The Hodge-Helmholtz decomposition

Theorem 2 – HH decomposition [Lad87] Let $f \in L^2(\mathbb{R}^d; \mathbb{R}^d)$. There exists two uniquely defined vector fields $g, h \in L^2(\mathbb{R}^d; \mathbb{R}^d)$ such that

$$f = g + h, \qquad g \simeq \nabla\varphi, \qquad \text{div}\,(h) \simeq 0. \tag{2}$$



What would it be for any measure?

Fundamental property of divergence-free fields: **their flow let the Lebesgue measure invariant**.

What would it be for any measure?

Fundamental property of divergence-free fields: **their flow let the Lebesgue measure invariant**.

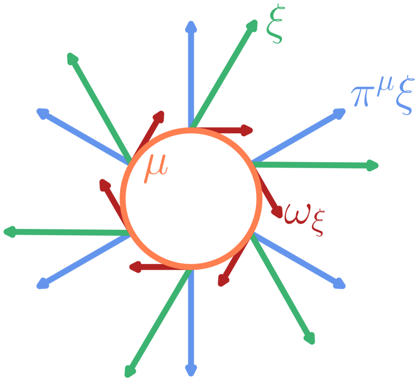
This can be extended to any measure, and to fields that put different probability over different directions.

What would it be for any measure?

Fundamental property of divergence-free fields: **their flow let the Lebesgue measure invariant**.

This can be extended to any measure, and to fields that put different probability over different directions.

Theorem 3 – Hodge decomposition Any field $\xi \in \mathscr{P}_2(\mathrm{T}\Omega)$ decomposes in a "tangent" component akin to a gradient, and a "divergence-free" component.



Thank you!

[BF23] Benoît Bonnet and Hélène Frankowska.
Caratheodory Theory and A Priori Estimates for Continuity Inclusions in the Space of Probability Measures, May 2023.
Preprint (arXiv:2302.00963).

[Lad87] Ol'ga A. Ladyženskaja.
The Mathematical Theory of Viscous Incompressible Flow.
Number 2 in Mathematics and Its Applications. Gordon and Breach, 1987.

[San15] Filippo Santambrogio.
Optimal Transport for Applied Mathematicians, volume 87 of *Progress in Nonlinear Differential Equations and Their Applications*.
Springer International Publishing, Cham, 2015.