Mathematisches Forschungsinstitut Oberwolfach

Report No. 7/2024

DOI: 10.4171/OWR/2024/7

Applications of Optimal Transportation

Organized by
Guillaume Carlier, Paris
Maria Colombo, Lausanne
Virginie Ehrlacher, Marne la Vallée
Daniel Matthes, Garching bei München

4 February – 9 February 2024

Abstract. The mathematical theory of optimal transportation is constantly expanding its range of application, while applications give impulses for new research directions in the field. This workshop was specifically devoted to applications of optimal transportation in the natural sciences, and to the recent developments of the theory that have been motivated by these. The bouquet of current applications includes mathematical models for large-scale air motion, dynamics of plasmas, material design, pattern formation in fluids, collective behaviour in biology, and many more. Related theoretical developments are in the analysis of the Hellinger-Kantorovich metric for modeling reaction–diffusion processes, and in efficient numerical methods for multi-marginal optimal transport, to name two of many examples encountered in this workshop.

Mathematics Subject Classification (2020): 49Q22.

License: Unless otherwise noted, the content of this report is licensed under CC BY SA 4.0.

Introduction by the Organizers

This workshop brought together 48 mathematicians — mainly from France, Italy, and Germany, but also e.g. from USA, Canada, and Singapore — that are working either on current applications of optimal transport in the natural sciences, or on theoretical developments that have been triggered by such applications. This has been the third Oberwolfach workshop with such a focus, following up on the initial event in 2016, and an online edition in 2020. The series is intended to fill the gap in between the strongly theory-oriented conferences like the bi-annual meetings in Pisa and the countless events with narrow focus on data science,

financial mathematics etc. Over time, we could observe a gradual shift in the range of applications. To name just one example: while density function theory and the corresponding developments in multi-marginal optimal transport have been significant topics in the previous editions, more classical questions related to fluid motion on various scales and the related theoretical studies of variants of the Euler equations have played a dominant role this time.

A highlight of the workshop has been the historical overview given by Felix Otto — one of the founding fathers and greatest promoters of the field — on the role of optimal transport in viscous fluid dynamics, and in particular the modeling of multi-phase mixtures by gradient descent, nowadays known as JKO scheme. The presentation by another pioneer in the field, Yann Brenier, was also on multi-phase fluids, specifically electrically charged ones, and a novel related convex optimization problem. Doubtlessly one of the most impressive contributions has been made by David Bourne, who was covering two quite different applications of optimal transport in a single talk: the first was a model for large-scale air motion for use in weather prediction, the second was about formation of grains in steel for use in material science. Applications to climate and to material design were actually recurrent themes during the workshop, e.g. Lauro Silini picked up on the former by analysing the semigeostrophic equations, and Benedikt Wirth continued on the latter by discussing shape optimization for maximal resistance to forces. As final example of many further applications that were presented, we mention the gradient flows studied by Yao Yao, that were motivated by models for collective behaviour of animals, and whose analysis uses profound results on the geometry underlying optimal transportation.

About half of the presented topics were not immediately linked to a physical or biological system, but exhibited recent developments of optimal transport theory and related numerical methods that have been motivated by applications. For example, the first presentation of the workshop, given by Alexander Mielke, featured novel results on the Hellinger-Kantorovich metric, which has been introduced to describe reaction-diffusion systems by means of gradient flows. Another example is the sophisticated and very efficient method for numerical solution of multi-marginal problems presented by Gero Friesecke, with the target at density function theory. Let us finally mention the talks of Quentin Mérigot and Dejan Slepčev on fine properties of the sliced Wasserstein distance, a distance that was born out of the need to find an efficient yet intuitive method to sample from given data in large dimensions.

As organizers, we were extremely pleased with the mathematical quality of the talks, that has been constantly extremely high throughout the entire workshop. The collection of abstracts below nicely reflects this, and we warmly invite the reader to browse through it.

Workshop: Applications of Optimal Transportation

Table of Contents

Alexander Mielke (joint with Vaios Laschos, Matthias Liero, Giuseppe Savaré)
On EVI flows in the (spherical) Hellinger-Kantorovich space 313

François-Xavier Vialard (joint with F. Léger and G. Todeschi)
Synthetic nonnegative cross-curvature: the Wasserstein space 318

Filippo Santambrogio
Estimates on the Brenier map via maximum principle 320

Aymeric Baradat (joint with Anastasiia Hraivoronska, Filippo Santambrogio)
Using Sinkhorn in JKO adds diffusion in the limiting PDE 322

Emanuela Radici (joint with Simone Di Marino, Lorenzo Portinale)
Transport problems with non linear mobilities: a particle approximation result 325

André Schlichting (joint with Daniel Matthes, Eva-Maria Rott, Giuseppe Savaré)
Diffusive transport: geodesics, convexity, and gradient flows with their structure preserving discretizations 327

Jean-David Benamou (joint with Guillaume Chazareix, Marc Hoffmann, Grégoire Loeper, François-Xavier Vialard)
Entropic optimal martingale transport 330

Quentin Mérigot (joint with Guillaume Carlier, Alessio Figalli, Yi Wang)
Sharp comparisons between the sliced Wasserstein and Wasserstein distances 333

Dejan Slepčev (joint with Sangmin Park and Lantian Xu)
Geometry of sliced optimal transport and projected-transport gradient flows 335

Nathael Gozlan (joint with Philippe Choné, Francis Kramarz)
Weak optimal transport with unnormalized kernels 338

Brendan Pass (joint with Luca Nenna)
An ordinary differential equation characterization of optimal transport and variants with linear constraints 340

Simone Di Marino (joint with Rodolfo Assereto, Kristian Bredies, Emanuele Naldi, Silvia Villa)
Wasserstein valued BV maps 342

Hugo Lavenant
Lifting functionals via optimal transport 344

David Bourne
Applications of optimal transport theory in meteorology and materials science 347

Lauro Silini
Solutions of the semigeostrophic equation on a rotating sphere 350

Gero Friesecke
Quantum mechanics, density functional theory, optimal transport, and the curse of dimension 352

Benedikt Wirth
Optimal transport bounds on pattern-forming energies with divergence constraint 356

Yao Yao (joint with José Carrillo, Matias Delgadino, Sabine Hittmeir, Bruno Volzone, Xukai Yan)
Aggregation-diffusion equation: symmetry, uniqueness and non-uniqueness of steady states 360

Andrea Natale (joint with Thomas O. Gallouët, Gabriele Todeschi)
Extrapolation in the Wasserstein space 362

Matthias Erbar (joint with Martin Huesmann, Jonas Jalowy, Bastian Müller)
Optimal transport for stationary point processes 365

Annette Dumas (joint with Filippo Santambrogio)
Minimizing the total variation in a congested setting 368

Inwon Kim
Stefan problem via optimal Skorokhod problem 369

Johannes Zimmer
Thermodynamic gradient flow evolution from particles 370

Maxime Sylvestre (joint with Hugo Malamut)
Disentangling Entropy and Suboptimality in Entropic Optimal Transport 373

Giuseppe Savaré (joint with Emanuele Naldi)
Opial Property in Wasserstein Spaces and Asymptotic Convergence of EVI Flows and of Proximal Point Algorithms 375

Noemi David (joint with T. Debiec, B. Perthame, A.R. Mészáros, F. Santambrogio)
Convergence rate for the incompressible limit of porous medium equations 378

Yann Brenier
Solution of the initial value problem for the gravitational Vlasov-Poisson system in an optimal transport style 381

Abstracts

On EVI flows in the (spherical) Hellinger-Kantorovich space

Alexander Mielke
(joint work with Vaios Laschos, Matthias Liero, Giuseppe Savaré)

Consider a bounded, closed, and convex subset X of $\mathbb{R}^d$ with nonempty interior and denote by $\mathbf{M}(X)$ and $\mathbf{P}(X)$ the set of all (non-negative) measures and all probability measures, respectively. The Hellinger-Kantorovich distance HK is a geodesic distance on $\mathbf{M}(X)$ that describes unbalanced optimal transport, where mass can be created and annihilated, see [LMS16, LMS18]. The same distance was introduced in the independent works [KMV16, GaM17, CP*15, CP*18] under the name "Wasserstein-Fisher-Rao distance". We comment on the naming at the end of this note.

This note summarizes old and recent joint work with V. Laschos, M. Liero, and G. Savaré, see [LiM13, LMS16, LMS18, LaM19] and [LMS23, LaM23].

1. Geodesics for the Hellinger-Kantorovich distance

Using two non-negative numbers α and β we can scale the interaction between the pure Hellinger distance He for $(\alpha,\beta)=(0,4)$ and the Kantorovich-Wasserstein distance W for $(\alpha,\beta)=(1,0)$. We define

$$\mathsf{HK}_{\alpha,\beta}(\mu^0,\mu^1)^2 := \inf\Big\{ \int_0^1 \int_X \big(\alpha|\nabla\xi_s|^2+\beta\xi_s^2\big)\mu_s(\mathrm{d}x)\,\mathrm{d}s \;\Big|\; \mu_0=\mu^0,\ \mu_1=\mu^1,\ s\mapsto(\mu_s,\xi_s) \text{ solves } (\mathrm{gCE})_{\alpha,\beta}\Big\}.$$

Any minimizer $s\mapsto\mu_s$ is called a (constant-speed) geodesic. Here,

$$(\mathrm{gCE})_{\alpha,\beta} \qquad \partial_s\mu_s + \operatorname{div}\big(\alpha\nabla\xi_s\mu_s\big) = \beta\xi_s\mu_s \quad \text{in the sense of distributions.}$$

It is shown in [LMS18, LMS23] that the optimal potential ξ_s is a suitable solution of the Hamilton-Jacobi equation

$$\partial_s\xi_s + \frac{\alpha}{2}|\nabla\xi_s|^2 + \frac{\beta}{2}\xi_s^2 = 0,$$

satisfying $\xi_t = \mathfrak{P}^{\alpha,\beta}_{t-s}(\xi_s)$ for $0<s<t<1$, where

$$\mathfrak{P}^{\alpha,\beta}_\tau(\eta)(x) = \inf\Big\{ \frac{2}{\beta\tau}\Big(1-\frac{\cos\big(\sqrt{\beta/(4\alpha)}|x-y|\big)}{1+\tau\,\beta/2\,\eta(y)}\Big) \;\Big|\; |x-y|\le\pi\sqrt{\frac{\alpha}{\beta}}\Big\}.$$

One can check that the formal limits $\alpha\to0$ and $\beta\to0$ recover the classical distances, namely $\mathsf{HK}_{\alpha,0}=\frac{1}{\sqrt{\alpha}}\mathsf{W}$ and $\mathsf{HK}_{0,\beta}=\frac{2}{\sqrt{\beta}}\mathsf{He}$.

A remarkable fact is that transport can only occur over distances less or equal to the length $\ell_*=\pi/\sqrt{\alpha/\beta}$. Thus, in general a geodesic curve between two measures $\mu_0,\mu_1\in\mathbf{M}(X)$ consists of three pieces:

- mass of μ_0 located further from $\mathrm{sppt}(\mu_1)$ than ℓ_* is annihilated,
- mass of μ_1 located further from $\mathrm{sppt}(\mu_0)$ than ℓ_* is created,

• mass of μ_0 located less than ℓ_* from $\mathrm{sppt}(\mu_1)$ is transported but again with changing mass along the way.

Transport over a distance of exactly ℓ_* is more difficult, in particular non-uniqueness of geodesics may happen. Neglecting this case, the geodesics take the following form. Set $A_0 = \{ x \in \operatorname{sppt}\mu_0 \mid \operatorname{dist}(x, \operatorname{sppt}\mu_1) > \ell_* \}$, $A_1 = \{ x \in \operatorname{sppt}\mu_1 \mid \operatorname{dist}(x, \operatorname{sppt}\mu_0) > \ell_* \}$, and $A_{\mathrm{tra}} = \{ x \in \operatorname{sppt}\mu_0 \mid \operatorname{dist}(x, \operatorname{sppt}\mu_1) < \ell_* \}$, then

$$\mu_s \;=\; (1{-}s)^2 \mu_0|_{A_0} \;+\; \Phi_s^{\alpha,\beta}\big(\mu_0|_{A_{\mathrm{tra}}}\big) \;+\; s^2 \mu_1|_{A_1},$$

where $\Phi_s^{\alpha,\beta}$ is given in terms of a growth factor q_s^2 and a transport map $\mathbf{T}_s$:

$$\Phi_s^{\alpha,\beta}(\mu) = \mathbf{T}_{\#}\big(q_s^2 \mu\big) \text{ with } q_s^2(x) = \Big(1{+}s\,\frac{\beta}{4}\,\xi_0(x)\Big)^2 + \frac{\alpha\beta}{4}\,|\nabla \xi_0(x)|^2$$

$$\text{and } \mathbf{T}_s(x) = x + \sqrt{\frac{4\alpha}{\beta}}\,\mathbf{arctan}\Big(\frac{s\sqrt{\alpha\beta}}{2{+}s\beta\xi_0(x)}\,\nabla\xi_0(x)\Big). \tag{1}$$

2. Geodesic Λ-convexity on $(\mathbf{M}(X), \mathsf{HK}_{\alpha,\beta})$

A function $\mathcal{E} : \mathbf{M}(X) \to \mathbb{R}_\infty$ is called geodesically Λ-convex, where $\Lambda \in \mathbb{R}$, if

$$\mathcal{E}(\mu_s) \le (1{-}s)\mathcal{E}(\mu_0) + s\,\mathcal{E}(\mu_1) - \Lambda \frac{s{-}s^2}{2} \mathsf{HK}_{\alpha,\beta}(\mu_0,\mu_1)^2 \text{ for } s \in [0,1],$$

along all constant speed geodesics. The total mass functional satisfies

$$\mu_s(X) = (1{-}s)\mu_0(X) + s\,\mu_1(X) - \frac{\beta}{2}\,\frac{s{-}s^2}{2}\,\mathsf{HK}_{\alpha,\beta}(\mu_0,\mu_1)^2,$$

i.e. $\mu \mapsto a\mu(X)$ is $(a\beta/2)$-convex for all $a \in \mathbb{R}$.

The geodesic convexity of linear functionals $\mu \mapsto \int_X V(x)\mu(\mathrm{d}x)$ was resolved in [LMS16, Prop. 20] and amounts for smooth functions V in the condition

$$\begin{pmatrix} \alpha \mathrm{D}^2 V(x) + \frac{\beta}{2} V(x) & \sqrt{\alpha\beta/4}\,\nabla V(x) \\ \sqrt{\alpha\beta/4}\,\nabla V(x)^\top & \frac{\beta}{2} V(x) \end{pmatrix} \ge \Lambda\, I_{d+1} \qquad \text{in } \mathbb{R}^{(d+1)\times(d+1)}.$$

For internal energies of the form $\mathcal{E}(\mu) = \int_X E(\rho)\,\mathrm{d}x + E'_\infty \mu^\perp$, where $\mu = \rho\mathcal{L} + \mu^\perp$ with $\mathcal{L} \perp \mu^\perp$, the necessary and sufficient conditions for geodesic Λ-convexity in [LMS23] involve the auxiliary function $N_E(\gamma,\delta) = (\delta/\gamma)^d E\big(\gamma^{2+d}/\delta^d\big)$ and read as follows:

$$\mathcal{E} \text{ is } \Lambda\text{-convex} \iff \begin{cases} (a) & (\gamma,\delta) \mapsto N_E(\gamma,\delta) - \frac{2\Lambda}{\beta}\gamma^2 \text{ is convex,} \\ (b) & \delta \mapsto (d{-}1)\,N_E(\gamma,\delta) \text{ is non-decreasing.} \end{cases}$$

Ignoring all technicalities, the proof relies on rewriting the transport part of the geodesics in the form $\mu_s = \rho_s(x)\mathrm{d}x$ with $\rho_s(\mathbf{T}_s(x)) = \rho_0(x) q_s(x)^2 / \det \mathrm{D}\mathbf{T}_s(x)$. Setting $\gamma_s = \rho_0(x)^{1/2} q_s$ and $\delta_s = \rho_0(x)^{1/2} q_s (\det \mathrm{D}\mathbf{T}_s(x))^{1/d}$ one obtains

$$\mathcal{E}(\mu_s) = \int_{\mathbf{T}_s(A_{\mathrm{tra}})} E(\rho_s(y))\,\mathrm{d}y = \int_{A_{\mathrm{tra}}} E\big(\frac{\rho_0(x) q_s(x)^2}{\det \mathrm{D}\mathbf{T}_s(x)}\big)\,\mathrm{d}x = \int_{A_{\mathrm{tra}}} N_E\big(\gamma_s(x), \delta_s(x)\big)\,\mathrm{d}x.$$

Using the properties (a) and (b) and the explicit representation of q_s and $\mathbf{T}_s$ from (1), the result follows via lengthy computations.

Reducing the conditions to $\gamma \equiv 1$ one obtains the McCann conditions for geodesic convexity in $(\mathbf{M}(X), \mathsf{W})$, and reducing to $\delta \equiv 1$ gives the conditions for geodesic Hellinger convexity. However, the joint condition is stronger, see below.

Examples are functionals with $E(c) = c^m$ with $m \geq 1$ giving $\Lambda = 0$. Of course, for $E(c) = ac$ we obtain $\Lambda = a\beta/2$. The case $E(c) = -c^\theta$ leads to geodesic convexity (with $\Lambda = 0$) if and only if $\theta \in \big[d/(d{+}2), 1/2\big]$, which only has nontrivial solutions for $d = 1$ and 2. Recall that the McCann condition asks for $\theta \geq (d{-}1)/d$, and Hellinger convexity asks for $\theta \leq 1/2$.

For $\beta > 0$, the Boltzmann entropy with $E(c) = c \log c$ is not geodesically Λ convex in $(\mathbf{M}, \mathsf{HK}_{\alpha,\beta})$ for any $\Lambda \in \mathbb{R}$.

3. EVI_λ flows in $(\mathbf{M}(X), \mathsf{HK}_{\alpha,\beta})$ and $(\mathbf{P}(X), \mathsf{SHK}_{\alpha,\beta})$

For metric gradient systems $(M, \mathcal{E}, \mathcal{D})$, where $(M, \mathcal{D})$ is a geodesic space and $\mathcal{E}$ is geodesically Λ-convex, a curve $u : [0, \infty[\to M$ is called an EVI_λ-solution (see [AGS05, MuS20] if it is continuous and satisfies

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{D}(u(t), w)^2 + \frac{\Lambda}{2}\mathcal{D}(u(t), w)^2 \leq \mathcal{E}(w) - \mathcal{E}(u(t)) \quad \text{for all } w \in \mathrm{dom}(\mathcal{E}).$$

The existence theory for EVI solutions was recently extended in [MuS20, MuS24] where the crucial new idea is to use the so-called *local-angle condition* and K-concavity of the squared norm norm $u \mapsto \frac{1}{2}\mathcal{D}(w, u)^2$. This approach was extended in [LaM23] based on previous results in [LaM19] on K-concavity on suitable subsets of $\mathbf{M}(X)$, where measures have densities with upper and lower bounds. A general existence result for EVI_λ flows in $(\mathbf{M}(X), \mathsf{HK}_{\alpha,\beta})$ was obtained, by constructing solutions via the minimizing movement scheme, such that lower and upper bounds for the densities propagate and K-concavity can be exploited at least on finite time horizons. By density and the Λ-contraction property of EVI flows, these results lead to global EVI flows.

The space $(\mathbf{P}(X), \mathsf{SHK}_{\alpha,\beta})$ is geodesic if the spherical HK distance is defined via

$$\mathsf{SHK}_{\alpha,\beta}(\nu_0, \nu_1) := 2 \arcsin\Big(\frac{\sqrt{\beta}}{4}\,\mathsf{HK}_{\alpha,\beta}(\nu_0, \nu_1)\Big) \in [0, \pi/2].$$

This is a consequence of the following scaling relation: $\mathsf{HK}_{\alpha,\beta}(r_0^2\mu_0, r_1^2\mu_1)^2 =$

$$= r_0 r_1 \mathsf{HK}_{\alpha,\beta}(r_0\mu_0, r_1\mu_1)^2 + \frac{4}{\beta}\big[(r_0^2 - r_0 r_1)\mu_0(X) + (r_1^2 - r_0 r_1)\mu_1(X)\big],$$

which was established in [LaM19] and allows us to interpret $(\mathbf{M}(X), \mathsf{HK}_{\alpha,\beta})$ as a metric cone over $(\mathbf{P}(X), \mathsf{SHK}_{\alpha,\beta})$. The geodesics $s \mapsto \nu_s^{\mathsf{SHK}}$ are obtained from the HK geodesics μ^{HK} via $\nu_s = n(s)\mu^{\mathsf{HK}}_{\sigma(s)}$, where the functions $n(s)$ and $\sigma(s)$ are explicitly given in [LaM19].

The EVI-flow theory on $(\mathbf{P}(X), \mathsf{SHK}_{\alpha,\beta})$ works similarly as in the HK case, but has the caveat that there are almost no functionals for which geodesic Λ-convexity in $(\mathbf{P}(X), \mathsf{SHK}_{\alpha,\beta})$ is known. So far, only for the family $\mu \mapsto \int_X (-\rho)^\theta \,\mathrm{d}x$ with $\theta \in \big[d/(d{+}2), 1/2\big]$ geodesic convexity is established in [LaM23].

4. Reaction-diffusion equations versus EVI flows

The corresponding PDE associated with the EVI flows can be obtained by the Otto calculus. The Onsager operators for $\mathsf{HK}_{\alpha,\beta}$ and $\mathsf{SHK}_{\alpha,\beta}$ are given by

$$\mathbb{K}^{\alpha,\beta}_{\mathsf{HK}}(\mu)\xi = -\alpha\,\mathrm{div}(\mu\nabla\xi) + \beta\mu\xi \text{ and } \mathbb{K}^{\alpha,\beta}_{\mathsf{SHK}}(\nu)\xi = -\alpha\,\mathrm{div}(\nu\nabla\xi) + \beta\nu\big(\xi - \textstyle\int_X \xi\,\mathrm{d}\nu\big).$$

Hence for functionals $\mathcal{E}(\mu) = \int_X \big(E(\rho)-\rho V\big)\,\mathrm{d}x$ the induced reaction-diffusion equations take the form

$$\partial_t\rho = \alpha\,\mathrm{div}\,\big(\rho\nabla(E'(\rho)-V)\big) - \beta\rho\,(E'(\rho)-V) + \delta_{\mathsf{SHK}}\beta\rho\int_X \rho\,(E'(\rho)-V)\,\mathrm{d}x,$$

where $\delta_{\mathsf{SHK}} = 0$ for the case HK and $\delta_{\mathsf{SHK}} = 1$ for SHK, where also $\int_X \rho\,\mathrm{d}x \equiv 1$.

Existence of weak solutions for these equations based on minimizing movement schemes was established in a series of papers, see e.g. [GaM17, GLM19, DiC20, Fle21]. The following example provides a case where there are several weak solutions, but there is only one EVI solution, because the EVI theory generates a contractive semiflow using $\Lambda = 0$.

We choose $V \equiv 0$ and $E(c) = -c^{1/2} + \frac{1}{3}\,c^{3/2}$ and the initial datum $\mu(0) = 0$. Clearly, the energy $\mathcal{E}$ is geodesically Λ-convex on $(\mathbf{M}(X), \mathsf{HK}_{1,4})$ for some $\Lambda > 0$, and it is shown in [LaM23] that $\mu(t) = \tanh(t)^2\mathcal{L}$ is the unique EVI solution. However, the induced reaction-diffusion equation reads

$$\partial_t\rho = \Delta\big(\frac{1}{2}\,\rho^{1/2} + \frac{1}{6}\,\rho^{3/2}\big) + 2\big(\rho^{1/2} - \rho^{3/2}\big) \text{ in } X, \qquad \nabla\rho\cdot\mathrm{n} = 0 \text{ on } \partial X.$$

It is easily seen that there are multiple solutions starting with $\mu(0) = 0$, namely

$$\mu^\zeta(t) = 0 \text{ for } t \in [0,T] \quad \text{and} \quad \mu^\zeta(t) = \big(\tanh(t-\zeta)\big)^2 \text{ for } t \geq \zeta.$$

Thus, the usage of the gradient structure $(\mathbf{M}(X), \mathcal{E}, \mathsf{HK}_{1,4})$ and its EVI flow provides a much better theory than the concept of weak solutions.

5. On the names "Hellinger-Kantorovich" and "Wasserstein-Fisher-Rao"

In his dissertation [Hel07] and habilitation [Hel09] thesis, Hellinger introduced integrals of the type $\int_a^b u(t)\frac{\mathrm{d}f_1\mathrm{d}f_2}{\mathrm{d}g}$ for functions $u, f_1, f_2, g, h \in \mathrm{C}^0([a,b])$ where additionally g and h are increasing, lie in $\mathrm{BV}([a,b])$ and satisfy $\big(f_j(t_2)-f_j(t_1)\big)^2 \leq \big(g(t_2)-g(t_1)\big)\big(h(t_2)-h(t_1)\big)$.

Using the modern tool of the Radon-Nikodým derivative, [Kak48, Eqn. (11)] introduces the so-called Hellinger integral $\rho(\mu,\nu) = \int_\Omega \sqrt{\mu(\mathrm{d}\omega)\nu(\mathrm{d}\omega)}$ and defines what is nowadays called the Hellinger distance

$$\mathsf{He}(\mu,\nu) = \big(2-2\rho(\mu,\nu)\big)^{1/2} \in \big[0,\sqrt{2}\,\big] \quad \text{for } \mu,\nu \in \mathbf{P}(\Omega).$$

Note that different normalizations for He are used in the literature.

Since the early 1960s, the name Hellinger distance is consistently used, which can be checked by a search of "Hellinger distance" in MathSciNet which leads to more than 600 hits in abstracts or titles. In particular, Rao's paper [RaV63] introduces the Hellinger integral and distance explicitly by name.

The Fisher-Rao distance was popularized by [Rao45] as geodesic distance for the Fisher information metric. The construction is as follows: For a (finite-dimensional) parameter manifold Θ and a function $\widehat{p} : \Theta \to \mathbf{P}(\Omega)$ consider the parametrized family $\left\{ \mu = \widehat{p}(\cdot, \theta) \in \mathbf{P}(\Omega) \;\middle|\; \theta \subset \Theta \right\}$ of measures. To distinguish objects in this family Fisher introduces the information metric $\mathbb{G}_{\mathrm{Fi}}$ on Θ via

$$\langle \mathbb{G}_{\mathrm{Fi}}(\theta) v_1, v_2 \rangle = \int_\Omega \mathrm{D}_\theta(\log \widehat{p}(\omega, \theta))[v_1] \, \mathrm{D}_\theta(\log \widehat{p}(\omega, \theta))[v_2] \, p(\mathrm{d}\omega, \theta),$$

where v_1, v_2 are tangent vectors in $\mathrm{T}_\theta\Theta$. The Fisher-Rao distance is then the geodesic distance $d_{\mathrm{FR}} : \Theta \times \Theta \to [0, \infty[$ induced by the Riemannian tensor $\mathbb{G}_{\mathrm{Fi}}$.

Taking $\Theta = \mathbf{P}(\Omega)$ itself with $\mu = \widehat{p}(\mu)$ one obtains indeed, as a special case of Fisher's information metric, the infinitesimal metric

$$\langle \mathbb{G}_{\mathsf{He}}(\mu) \nu_1, \nu_2 \rangle = \int_\Omega \frac{\mathrm{d}\nu_1}{\mathrm{d}\mu} \frac{\mathrm{d}\nu_2}{\mathrm{d}\mu} \, \mathrm{d}\mu = \int_\Omega \frac{\mathrm{d}\nu_1 \mathrm{d}\nu_2}{\mathrm{d}\mu},$$

which is again has the form of Hellinger's original integrals $\int_a^b \mathrm{d}f_1 \mathrm{d}f_2 / \mathrm{d}g$, see [BBM16] for an instance of this usage. On $\mathbf{M}(\Omega)$ the induced "Fisher-Rao distance" is (one half of) the Hellinger distance He; however, the associated Fisher-Rao distance on $\mathbf{P}(\Omega)$ is the *Bhattacharya distance* (see [Bha42] and [Rao45, p. 246]) given by $\mathrm{Bh}(\nu_1, \nu_2) = 2 \arcsin\left(\frac{1}{2}\mathsf{He}(\nu_1, \nu_2)\right) \in [0, \pi/2]$, which is also called the *spherical Hellinger distance* in [LaM19]. .

References

[AGS05] L. Ambrosio, N. Gigli, and G. Savaré, *Gradient flows in metric spaces and in the space of probability measures*, Lectures in Mathematics ETH Zürich, Birkhäuser (2005).

[BBM16] M. Bauer, M. Bruveris, and P. W. Michor: *Uniqueness of the Fisher-Rao metric on the space of smooth densities*. Bull. Lond. Math. Soc. **48**:3 (2016) 499–506.

[Bha42] A. Bhattacharya: *On discrimination and divergence*. Proc. Indian Sci. Congress. Part III (1942) 13.

[CP*15] L. Chizat, G. Peyré, B. Schmitzer, and F.-X. Vialard: *An interpolating distance between optimal transport and Fisher–Rao metrics*. Found. Comput. Math. **18**:1 (2015) 1–44.

[CP*18] ———: *Unbalanced optimal transport: geometry and Kantorovich formulation*. J. Funct. Analysis **274**:11 (2018) 3090–3123.

[DiC20] S. Di Marino and L. Chizat: *A tumor growth model of Hele-Shaw type as a gradient flow*. ESAIM Control Optim. Calc. Var. **26**:103 (2020) 1–38.

[Fle21] F. Fleißner: *A minimizing movement approach to a class of scalar reaction-diffusion equations*. ESAIM Control Optim. Calc. Var. **27**:18 (2021) 1–29.

[GaM17] T. O. Gallouët and L. Monsaingeon: *A JKO splitting scheme for Kantovorich-Fisher-Rao gradient flows*. SIAM J. Math. Analysis **49**:2 (2017) 1100–1130.

[GLM19] T. Gallouët, M. Laborde, and L. Monsaingeon: *An unbalanced optimal transport splitting scheme for general advection-reaction-diffusion problems*. ESAIM Control Optim. Calc. Var. **25** (2019) 8/1–32.

[Hel07] E. Hellinger, *Die Orthogonalinvarianten quadratischer Formen von unendlich vielen Variablen*, Ph.D. thesis, Universität Göttingen, 1907.

[Hel09] ———: *Neue Begründung der Theorie quadratischer Formen von unendlichvielen Veränderlichen (in German)*. J. Reine Angew. Math. **136** (1909) 210–271.

[Kak48] S. Kakutani: *On equivalence of infinite product measures*. Annals Math. **49**:1 (1948) 214–224.

[KMV16] S. Kondratyev, L. Monsaingeon, and D. Vorotnikov: *A new optimal transport distance on the space of finite Radon measures*. Adv. Differ. Eqns. **21**:11/12 (2016) 1117–1164.
[LaM19] V. Laschos and A. Mielke: *Geometric properties of cones with applications on the Hellinger–Kantorovich space, and a new distance on the space of probability measures*. J. Funct. Analysis **276**:11 (2019) 3529–3576.
[LaM23] ———: *Evolutionary Variational Inequalites on the Hellinger-Kantorovich and the spherical Hellinger-Kantorovich spaces*. Submitted (2023), arXiv:2207.09815v3.
[LiM13] M. Liero, A. Mielke: *Gradient structures and geodesic convexity for reaction-diffusion systems*. Phil. Trans. Royal Soc. A **371**:2005 (2013) 20120346/1–28.
[LMS16] M. Liero, A. Mielke, and G. Savaré: *Optimal transport in competition with reaction – the Hellinger–Kantorovich distance and geodesic curves*. SIAM J. Math. Analysis **48**:4 (2016) 2869–2911.
[LMS18] ———: *Optimal entropy-transport problems and a new Hellinger–Kantorovich distance between positive measures*. Invent. math. **211** (2018) 969–1117.
[LMS23] ———: *Fine properties of geodesics and geodesic λ-convexity for the Hellinger–Kantorovich distance*. Arch. Rational Mech. Anal. **247**:112 (2023) 1–73.
[MuS20] M. Muratori and G. Savaré: *Gradient flows and evolution variational inequalities in metric spaces. I: structural properties*. J. Funct. Analysis **278**:4 (2020) 108347/1–67.
[MuS24] ———: *Gradient flows and evolution variational inequalities in metric spaces. II: variational convergence and III: generation results*. In preparation (2024).
[Rao45] C. R. Rao: *Information and the accuracy attainable in the estimation of statistical parameters*. Bull. Calcutta Math. Soc. **37** (1945) 81–91.
[RaV63] C. R. Rao and V. S. Varadarajan: *Discrimination of Gaussian processes*. Sankhya: Indian J. Stat. A **25**:3 (1963) 303–330.

Synthetic nonnegative cross-curvature: the Wasserstein space

François-Xavier Vialard
(joint work with F. Léger and G. Todeschi)

Starting from a cost c on a product space $X \times Y$, the Wasserstein space can be seen as a natural lift of c on the space of measures. In this ongoing work, we are interested in an other property of the base space $(X \times Y, c)$ that also holds on the Wasserstein space $(\mathcal{P}(X) \times \mathcal{P}(Y), \mathcal{T}_c)$. This property is called nonnegative cross-curvature and it was introduced by Kim and McCann around 2010 [3] as a strenghtening of the Ma-Trudinger-Wang condition which is a necessary condition for smoothness of optimal transport maps. The Ma-Trudinger-Wang tensor is reformulated as the evaluation of the curvature tensor of the Kim-McCann Riemannian pseudo-metric in orthogonal directions for this metric. Recall that the Kim-McCann metric is defined by

$$g(x,y) = -\begin{pmatrix} 0 & \frac{1}{2}\partial_{yx}c \\ \frac{1}{2}\partial_{xy}c & 0 \end{pmatrix}. \tag{1}$$

The Ma-Trudinger-Wang condition says that the curvature tensor of g is nonnegative on some planes whereas the nonnegative cross-curvature imposes more planes. Although more stringent, this new condition which we refer to nonnegative cross-curvature is more stable under natural geometric properties such as products and submersions as shown by Kim and McCann.

A formal argument on Riemannian manifolds: Consider the case where $X = Y$ and the cost is the distance squared of a Riemannian metric. The starting point of our work is the remark that there is a formal Riemannian submersion between the space of maps $L^2(X, X)$ and the space of measures endowed with the Wasserstein metric. Using a formal computation due to Freed and Groisser, the cross-curvature tensor of the space $L^2(X, X)$ is nonnegative when it is the case for the base space $(X \times X, d^2)$ where d^2 is the Riemannian squared distance. Now, we can apply formally the result of Kim and McCann on Riemannian submersions in this infinite dimensional case to infer that it is also the case for the Wasserstein space.

The rigorous formulation: In fact, we are able to prove the following statement:

> If $(X \times Y, c)$ has synthetic nonnegative cross-curvature, it is also holds for the Wasserstein space.

In order to formulate the statement on the Wasserstein space, we needed a more synthetic formulation on a space of nonnegative cross-curvature which can encompass boundaries and singularities of the space. For instance, Dirac masses in the Wasserstein space are extremal points in the sense of convex geometry. On the space of positive symmetric matrices, rank deficient matrices lie on the boundary. Our definition of synthetic nonnegative cross-curvature (SNCC) is as follows: We say that $(X \times Y, c)$ has synthetic nonnegative cross-curvature (on X, or w.r.t. Y) if for every couple $(x_0, x_1) \in X^2$ and $y \in Y$, there exists a curve $x(t)$ joining x_0 and x_1 such that for all $y' \in Y$, the following function is convex:

$$t \mapsto c(x(t), y) - c(x(t), y') \,. \tag{2}$$

This definition is called synthetic since it does not require any differentiability property of the cost and on the curve. However, in a smooth setting, $x(t)$ is necessarily unique and given by the c-segment for c between x_0 and x_1 w.r.t. the point y, whose definition is recalled below:

$$\partial_y c(x(t), y) = (1-t)\partial_y c(x_0, y) + t\partial_y c(x_1, y) \,, \quad \forall t \in [0, 1] \,. \tag{3}$$

Examples of spaces that satisfy our synthetic notion of nonnegative cross-curvature are

(1) compact manifolds with the distance squared cost that are nonnegatively cross-curved on the injectivity domain of the exponential map,
(2) the space of positive semidefinite matrices (PSD) endowed with the Bures-Wasserstein metric,
(3) the Wasserstein space when the underlying cost has the SNCC property, for instance probability measures on the sphere with the Riemannian distance squared, probability on the space of PSD matrices endowed with the Bure-Wasserstein metric as a cost.

Interestingly, we also study the unbalanced setting of optimal transport for which we have the following result: If the cost on the product of cones (see [1, 2]), then it is also the case for unbalanced optimal transport. Examples are the following:

(1) the base space is the sphere with the Riemannian distance squared,
(2) the base space is the set of unit trace PSD matrices endowed with the Bures angle metric.

We touched upon important potential applications for optimization on measures, gradient flows of first and second-order in time.

References

[1] Liero, Mielke and Savaré, *Optimal entropy-transport problems and a new Hellinger–Kantorovich distance between positive measures*, Inv. Math **211** (2018), 969–2017.
[2] Chizat, Peyré, Schmitzer and Vialard, *Unbalanced optimal transport: dynamic and Kantorovich formulations*, JFA **274** (2018), 3090–3123.
[3] Kim and McCann, *Continuity, curvature, and the general covariance of optimal transportation*, JEMS **12** (2010), 1009–1040.

Estimates on the Brenier map via maximum principle

Filippo Santambrogio

I presented in this talk three results concerning the optimal transport map T between two densities on the Eulidean space, all obtained with the same technique, consisting in differentiating once the Monge-Ampère equation.

1. Lipschitz bounds and displacement in the JKO scheme

The first result concerns the preservation of the Lipschitz bounds of $\log\rho + V$ in the JKO scheme for the Fokker-Planck equation $\partial_t\rho = \Delta\rho + \nabla\cdot(\rho\nabla V)$. Given a time step $\tau > 0$, the JKO scheme consists in finding a sequence ρ_k such that, at each step, the new density ρ_{k+1} is obtained by solving

$$\min_\rho \int \rho(\log\rho + V) + \frac{W_2^2(\rho,\rho_k)}{2\tau}.$$

The optimal ρ is characterized by $\log\rho + V + \frac{\varphi}{\tau} = c$, where φ is the Kantorovich potential for the cost $\frac{1}{2}|x-y|^2$ from ρ itself to ρ_k. The function φ solves the Monge-Ampère equation $\det(I - D^2\varphi) = \rho/\rho_k\circ T$, where T is the optimal transport map given by $T(x) = x - \nabla\varphi(x)$. By looking at the maximum point of $|\nabla\varphi|^2$, using the first and second order optimality conditions and the derivative of the Monge-Ampère equation, one obtains the following result

$$\text{if } D^2V \geq \lambda I \text{ then } (1+\lambda\tau)||\nabla(\log\rho_{k+1} + V)||_{L^\infty} \leq ||\nabla(\log\rho_k + V)||_{L^\infty}.$$

This is also a result on $||T - id||_{L^\infty}$, which will be estimated by $C\tau$. This result was already proven by Lee ([1]) in the case of the torus, and extended to the case of convex domains in collaboration with Ferrari ([2]). The main novelty of [2] consists in proving that, on the Euclidean ball, the maximum of $|\nabla\varphi|^2$ cannot be attained on the boundary when the densities are smooth, and then proceed by approximation.

2. Convergence to equilibrium in a reproduction model

The second result (taken from [3]) concerns the convergence to the equilibrium of a non-linear reproduction model in mathematical biology. This model is a discrete-in-time evolution given by $\rho_{k+1} = \mathcal{T}[\rho_k]$, where the operator $\mathcal{T}$ is given by

$$\mathcal{T}[\rho](x) := e^{-m(x)} \int\int G(x - \frac{x_1 + x_2}{2})\rho(x_1)\rho(x_2)dx_1dx_2,$$

where $x_1, x_2 \in \mathbb{R}$ represent the traits of two parents, and ρ_{k+1} is the distribution of this trait in the next generation, obtained after reproduction, averaging of the two parents' traits, mutation (i.e. convolution with the Gaussian G) and then applying a mortality rate e^{-m}. It is assumed that m is strongly convex, i.e. $m'' \geq \alpha > 0$.

It can be proven that there exists an invariant state F with a number $\lambda > 0$ such that $\mathcal{T}[F] = \lambda F$. Moreover, if ρ is γ-log-concave then $\mathcal{T}[\rho]$ is $(\alpha + \frac{2\gamma}{2\gamma+1})$-log concave, which implies that F is of the form $F = e^{-V}$ with $V'' \geq \beta$ and $\beta = \alpha + \frac{2\beta}{2\beta+1}$. In particular, $\beta > 1/2$. We then look at the evolution ρ_k writing $\rho_k = u_k \cdot F$ and we observe that we have

$$u_{k+1}(x) = \int P_x(x_1, x_2)u_k(x_1)u_k(x_2)dx_1dx_2,$$

where P_x is a probabily density on $\mathbb{R}^2$ proportional to $e^{-W(x,x_1,x_2)}$ where we set $W(x, x_1, x_2) = V(x_1) + V(x_2) + |x - \frac{x_1+x_2}{2}|^2$. We want to prove that u_k converges (and quantify this convergence) to a constant and we use the following estimate

$$|\log(u_{k+1}(x)) - \log(u_{k+1}(x'))| \leq \mathrm{Lip}(\log(u_k))W_{\infty,1}(P_x, P_{x'}),$$

where $W_{\infty,1}$ stands for the W_∞ distance in $\mathbb{R}^2$ coputed according to the norm ℓ_1. In order to estimate the distance $W_{\infty,1}(P_x, P_{x'})$, we consider the optimal Brenier map T between P_x and $P_{x'}$ and estimate $||T - id||$ in a similar way as what done in [2]. Since we need to deal with a particular norm and obtain a sharp result, we adapt the proof or [2] replacing the Euclidean ball with the ℓ_1 ball and obtain,

$$W_{\infty,1}(P_x, P_{x'}) \leq ||T - id||_{L^\infty,1} \leq \frac{|x - x'|}{1 + 2\beta}.$$

The condition $\beta > 1/2$ allows to obtain exponential convergence.

3. Linear growth of the optimal map

The last result, obtained in collaboration with G. Carlier and A. Figalli and still in preparation, concers instead the map T itself and not its displacement $T - id$. The goal being to extend the celebrated Caffarelli's theorem on Lipschitz transport maps between log-concave measures to the case of measures with a different tail at infinity. For this proof, it is useful to have a uniform bound on the ration $|T(x)|/|x|$ in the limit $|x| \to \infty$. We considering the optimal transport map T between two densities of the form $\rho_0 = V^{-d}$ and $\rho_1 = W^{-d}$ (the number d being

the dimension of the space) and set $T(x) = \nabla u(x)$. By looking at the maximum point of $|\nabla u|^2/u$ it is possible to obtain the following result

$$\text{if } V(x) \approx |x|^p, W(x) \approx |x|^q,\ p \leq q \text{ then } |T(x)| \leq C(1 + |x|)$$

for a constant C depending on V and W (and on their precise estimates in terms of the powers $|x|^p$ or $|x|^q$).

References

[1] P. Lee, *On the Jordan-Kinderlehrer-Otto scheme,* J. Math. An. Appl. **429** Issue 1 (2015), 131–142.

[2] V. Ferrari and F. Santambrogio *Lipschitz estimates on the JKO scheme for the Fokker–Planck equation on bounded convex domains*, Appli. Math. Letters, **112** (2021), 106806.

[3] V. Calvez, D. Poyato and F. Santambrogio *Uniform contractivity of the Fisher infinitesimal model with strongly convex selection.*, preprint

Using Sinkhorn in JKO adds diffusion in the limiting PDE

Aymeric Baradat

(joint work with Anastasiia Hraivoronska, Filippo Santambrogio)

The JKO scheme [6], has become in the last decades a very popular numerical scheme to compute solutions to PDEs that have the structure of gradient flows in the Wasserstein space [1, 9].

The purpose of our work is to understand how replacing the Wasserstein distance by its entropic regularization in the JKO scheme – as done when using variants of the Sinkhorn algorithm – affects the limiting PDE. We show that in the scaling where the regularization parameter is proportional to the time step, this replacement amounts to adding linear diffusion.

1. The JKO scheme

Working in the flat torus $\mathbb{T}^d$, $d \in \mathbb{N}^*$, giving ourselves three regular functions $V : \mathbb{T}^d \to \mathbb{R}$, $W : \mathbb{T}^d \to \mathbb{R}$ even, and $f : \mathbb{R}_+ \to \mathbb{R}_+$ convex, as well as an initial datum $\rho_0 \in \mathcal{P}(\mathbb{T}^d)$, a classical instance of a Wasserstein gradient flow in the Wasserstein space $\mathcal{P}(\mathbb{T}^d)$ is the following Cauchy problem:

$$\begin{cases} \partial_t \rho = \operatorname{div}\Big(\rho(\nabla V + \nabla W * \rho)\Big) + \Delta g(\rho), \\ \rho|_{t=0} = \rho_0, \end{cases} \tag{1}$$

where $g : \mathbb{R}_+ \to \mathbb{R}_+$ satisfies the differential relation $g'(s) = sf''(s)$. Indeed, it appears formally as the gradient flow of the functional

$$\mathcal{E} : \rho \in \mathcal{P}(\mathbb{T}^d) \longmapsto \begin{cases} \displaystyle\int \left(V + \frac{1}{2} W * \rho\right) \mathrm{d}\rho + \int f(\rho)\mathrm{d}x & \text{if } \rho \ll \mathrm{d}x \text{ or } f = 0, \\ +\infty & \text{else.} \end{cases} \tag{2}$$

In this context, calling D the 2-Wasserstein distance on $\mathcal{P}(\mathbb{T}^d)$ and given a time step $\tau > 0$, the JKO scheme consists in considering a sequence $(\rho_n^\tau)_{n\in\mathbb{N}}$ satisfying $\rho_0^\tau = \rho_0$ and the following induction relation:

$$\forall n \in \mathbb{N}, \quad \rho_{n+1}^\tau \text{ is a minimizer of } \rho \in \mathcal{P}(\mathbb{T}^d) \longmapsto \frac{D(\rho_n^\tau, \rho)^2}{2\tau} + \mathcal{E}(\rho). \tag{3}$$

Under mild assumptions on V, W and f, such a sequence always exists. Also, the family labelled by τ of piecewise constant curves valued in $\mathcal{P}(\mathbb{T}^d)$ jumping to ρ_n^τ at time τn for all $n \in \mathbb{N}^*$ is relatively compact for the topology of local uniform convergence, and its limit points solve (1) distributionally [9].

2. Regularization of the scheme and main result

A very popular way to compute Wasserstein distances nowadays is to use the Sinkhorn algorithm [4]. It has the advantage of being very fast, but the drawback of computing only approximately the Wasserstein distance, or more precisely, of computing exactly an approximation of the Wasserstein distance. We will call here the Schrödinger cost this approximation. We refer to [8] for various definitions of this cost and its links with statistical mechanical problems (in fact, the definition we use will be specified in (5) below). This cost depends on a parameter $\varepsilon > 0$, and therefore we will denote it by D_ε.

The convergence $D_\varepsilon \to D$ is now well understood [7, 3, 2]. Hence, replacing D by D_ε in (3) should be reasonable when ε is small. However, our discovery is that when ε is taken proportional to the time step τ, then this replacement has the effect of adding linear diffusion to the limiting PDE.

Theorem. *Let $\mathcal{E}$ be defined as in (2) with V and W of class C^2 and with $f : s \in \mathbb{R}_+ \mapsto 0$, $s\log s$ or $s^m/(m-1)$, for some $m > 1$.*

Let $\theta > 0$ and $\rho_0 \in \mathcal{P}(\mathbb{T}^d) \cap L\log L$ be such that $\mathcal{E}(\rho_0) < +\infty$. For any $\tau > 0$, there exists a sequence $(\rho_n^\tau)_{n\in\mathbb{N}}$ satisfying $\rho_0^\tau = \rho_0$ and the induction relation:

$$\forall n \in \mathbb{N}, \quad \rho_{n+1}^\tau \text{ is a minimizer of } \rho \in \mathcal{P}(\mathbb{T}^d) \longmapsto \frac{D_{\theta\tau}(\rho_n^\tau, \rho)^2}{2\tau} + \mathcal{E}(\rho).$$

Moreover, the family labelled by τ of piecewise constant curves jumping to ρ_n^τ at time τn for all $n \in \mathbb{N}^*$ is relatively compact for the topology of local uniform convergence in $\mathcal{P}(\mathbb{T}^d)$, and its limit points are distributional solutions of

$$\begin{cases} \partial_t \rho = \operatorname{div}\Big(\rho(\nabla V + \nabla W * \rho)\Big) + \Delta g(\rho) + \dfrac{\theta}{2}\Delta\rho, \\ \rho|_{t=0} = \rho_0. \end{cases} \tag{4}$$

Remark. *Let us make two remarks about this result.*

- *First, it implies that if we do not want to see any effect of the regularization and recover (1) as the limiting PDE, we have to take at worst $\varepsilon = o(\tau)$ as $\tau \to 0$. Actually, our proof shows that this condition is sufficient, and hence extends a result from* [3].

- *The speed of the Sinkhorn algorithm increases with ε. So when there is linear diffusion in the PDE we want to solve, our result suggests to remove the entropic part of $\mathcal{E}$ and to make it appear through the regularization.*

3. Idea of the proof

The key is to use the most appropriate definition of the Schrödinger cost. Even in Benamou-Brenier formulation, there are two ways to do it: either one adds a Fischer information term in the action to minimize, or one includes linear diffusion in the continuity equation [5]. As we want to see linear diffusion appearing, we choose the second one. That is, for all $\varepsilon > 0$ and $\mu, \nu \in \mathcal{P}(\mathbb{T}^d) \cap L \log L$, we define

$$\frac{D_\varepsilon(\mu,\nu)^2}{2} := \varepsilon \int \mu \log \mu + \inf \left\{ \frac{1}{2} \int_0^1 \int |v|^2 \mathrm{d}\rho \, \mathrm{d}t \;\middle|\; \begin{array}{c} \partial_t \rho + \mathrm{div}(\rho v) = \frac{\varepsilon}{2} \Delta \rho \\ \rho_0 = \mu, \quad \rho_1 = \nu \end{array} \right\}. \tag{5}$$

Duality shows that given $\theta, \tau > 0$ and $\mu \in \mathcal{P}(\mathbb{T}^d) \cap L \log L$, an optimizer of

$$\rho \in \mathcal{P}(\mathbb{T}^d) \longmapsto \frac{D_{\theta\tau}(\mu,\rho)^2}{2\tau} + \mathcal{E}(\rho)$$

must be the value at time τ of a solution of the forward-backward system

$$\begin{cases} \partial_t \rho = \mathrm{div}(\rho \nabla \varphi) + \dfrac{\theta}{2} \Delta \rho, & \rho_0 = \mu, \\ \partial_t \varphi = \dfrac{1}{2} |\nabla \varphi|^2 - \dfrac{\theta}{2} \Delta \varphi, & \varphi_\tau = V + W * \rho_\tau + f'(\rho_\tau). \end{cases}$$

The first line of this system would coincide exactly with our targetted PDE (4) if φ would be $V + W * \rho + f'(\rho)$ for all time. But this identity holds at time τ only, as expressed in the second line. Thus, the proof reduces to showing that φ is well approximated by its value at the time steps. This can be read in the PDE on φ.

4. An open problem

So far, there is no result concerning the crucial point of estimating the rate of convergence of our regularized scheme. Yet, it is worth it to use Sinkhorn in order to improve the computational speed of each step only if the rate of convergence of the scheme is not catastrophically affected by the regularization.

References

[1] L. Ambrosio, N. Gigli, and G. Savaré. *Gradient flows in metric spaces and in the space of probability measures*. Springer, 2006.

[2] A. Baradat and L. Monsaingeon. Small Noise Limit and Convexity for Generalized Incompressible Flows, Schrödinger Problems, and Optimal Transport. *Arch. Ration. Mech. Anal.*, 235(2):1357–1403, 2020.

[3] G. Carlier, V. Duval, G. Peyré, and B. Schmitzer. Convergence of entropic schemes for optimal transport and gradient flows. *SIAM J. Math. Anal.*, 49(2):1385–1418, 2017.

[4] M. Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport. In *Adv. Neur. In.*, pages 2292–2300, 2013.

[5] I. Gentil, C. Léonard, and L. Ripani. About the analogy between optimal transport and minimal entropy. *Ann. Fac. Sci. Toulouse. Série 6*, 3:569–600, 2017.

[6] R. Jordan, D. Kinderlehrer, and F. Otto. The variational formulation of the Fokker–Planck equation. *SIAM J. Math. Anal.*, 29(1):1–17, 1998.
[7] C. Léonard. From the Schrödinger problem to the Monge–Kantorovich problem. *J. Funct. Anal.*, 262(4):1879–1920, 2012.
[8] C. Léonard. A survey of the Schrödinger problem and some of its connections with optimal transport. *Discrete Cont. Dyn-A*, 34(4):1533–1574, 2014.
[9] F. Santambrogio. {Euclidean, metric, and Wasserstein} gradient flows: an overview. *Bull. Math. Sci.*, 7(1):87–154, 2017.

Transport problems with non linear mobilities: a particle approximation result

Emanuela Radici
(joint work with Simone Di Marino, Lorenzo Portinale)

We are interested in evolutive dynamics of the form

$$\partial_t \rho - \nabla \cdot (m(\rho)\mathbf{v}) = 0 \tag{1}$$

which might include local and/or non local effects, in particular prevention of overcrowding. Congested evolutions mathematically translate into considering concave non-linear mobility functions m in expression (1). We further request the support of m to be compact $[0, \rho_{max}]$ so to focus on the more singular situation of hard congested dynamics where the total density of the system cannot go beyond a certain maximal density ρ_{max}.

When the velocity field $\mathbf{v}$ can be written as $\nabla \delta F[\rho]$ for some energy functional F, one can associate to the system a corresponding Wasserstein-like distance that takes into account the nonlinearity. Such distances allow only for a dynamical formulation: for every μ_0, μ_1 probability measures which are absolutely continuous with respect to Lebesgue we write

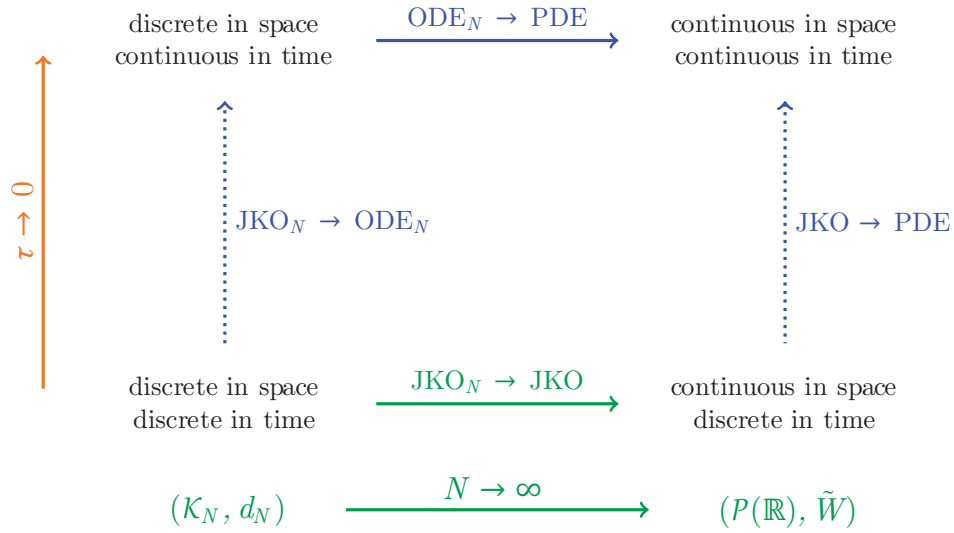
$$\tilde{W}^2(\mu_0, \mu_1) = \inf \left\{ \int_0^1 \int_{\mathbb{R}} |v_t|^2 m(\rho_t)\, dx dt \,:\, \partial_t \rho_t + \nabla \cdot (m(\rho_t) v_t) = 0,\ \rho_{t=0,1} = \mu_{0,1} \right\}$$

while $\tilde{W}$ is $+\infty$ otherwise. These distances were first introduced in [3] and further studied in [2, 1] for monotone concave mobilities m and in [5] for the compactly supported case. However, in this latter case there is not yet an exaustive understanding of the underlying metric, e.g. no characterization of geodesics is available at the moment.

Our goal is to study suitable discretisations of the problems (1) when $F[\rho] = \int_{\mathbb{R}} V d\rho$, both at the level of the generalised optimal transport distances and the associated gradient flows, with particular attention to two types of approximations: in space and in time. We introduce, in the scalar case, a space discretisation in the framework of nonlinear mobilities, adopting a Lagrangian point of view through systems of N-ordered particles. Our discrete settings correspond to finite-dimensional manifolds $(\mathcal{K}_N, d_N)$ where $\mathcal{K}_N$ is the N-dimensional cone of points with ordered components (describing the possible configurations of the moving

N-ordered particles) and $d_N : \mathcal{K}_N \times \mathcal{K}_N \to [0, \infty)$ are suitable mobility-weighted distances on $\mathcal{K}_N$.

In [4] we provide a Γ-convergence result for the associated discrete metrics d_N as $N \to \infty$ towards $\tilde{W}$ in the framework of probability measures on $\mathbb{R}$ and discuss applications to the approximation of one-dimensional conservation laws when $F[\rho] = \int_{\mathbb{R}} V d\rho$ via the so-called generalised minimising movements (JKO schemes) involving metrics d_N and $\tilde{W}$ respectively. Our contribution aim towards the understanding of the discrete-to-continuum time and space *commuting diagram* regarding the one dimensional evolutionary PDEs that can be seen as gradient flows with respect to a non-linear mobility Wasserstein like distance, as described in Figure



With respect to the diagram, this work deals with the bottom arrow for (1) with velocity fields of the form $\mathbf{v} = \nabla V(x)$. The top arrow has been already investigated in the scalar case by several authors, see [6] for the most general result. We aim this work to be the first of a series trying to shed new lights on the interplay between generalised gradient-flow structures, conservation laws, and Wasserstein distances with nonlinear mobilities. The analysis of the vertical arrows in the dyagram being the first step in this direction.

References

[1] P. Cardaliaguet, G. Carlier, B. Nazaret, *Geodesics for a class of distances in the space of probability measures*, Calc. Var., **48** (2013) 395–420.

[2] J.A. Carrillo, S. Lisini, G. Savaré, D. Slepcev, *Nonlinear mobility continuity equations and generalized displacement convexity*, J. Funct. Anal., **258** (2010) 1273–1309.

[3] J. Dolbeault, B. Nazaret, G. Savaré, *A new class of transport distances between measures*, Calc. Var. Partial Differ. Equ. **34**-2 (2009) 193–231.
[4] S. Di Marino, L. Portinale, E. Radici, *Optimal transport with nonlinear mobilities: a deterministic particle approximation result*, to appear on Adv. Calc. Var. (2023).
[5] S. Lisini, A. Marigonda, *On a class of modified Wasserstein distances induced by concave mobility functions defined on bounded intervals*, Manuscripta Math. **133**-1-2 (2010) 197–224.
[6] E. Radici, F. Stra, *Entropy solutions of non-local scalar conservation laws with congestion via deterministic particle method*, SIAM J. Math. Anal., **55**-3 (2022) 2001–2041.

Diffusive transport: geodesics, convexity, and gradient flows with their structure preserving discretizations

André Schlichting

(joint work with Daniel Matthes, Eva-Maria Rott, Giuseppe Savaré)

1. The diffusive transport metric

We consider the space $X := \{\rho \in L^1(\mathbb{S}^1) | \rho \geq 0, \int \rho \, \mathrm{d}x = 1\}$ of probability densities on the circle. On X, we recall the dynamic formulations of the Hellinger distance $\mathbb{H}$ and the L^2-Wasserstein metric $\mathbb{W}$ and propose a new *diffusive transport distance* $\mathbb{D}$ defined by

$$\mathbb{H}(\rho_0, \rho_1)^2 = \inf \left\{ \int_0^1 \int_{\mathbb{S}^1} \frac{w_s^2}{\rho_s} \, \mathrm{d}x \, \mathrm{d}s \,\middle|\, \partial_s \rho_s - w_s = 0 \right\}, \tag{1}$$

$$\mathbb{W}(\rho_0, \rho_1)^2 = \inf \left\{ \int_0^1 \int_{\mathbb{S}^1} \frac{w_s^2}{\rho_s} \, \mathrm{d}x \, \mathrm{d}s \,\middle|\, \partial_s \rho_s + \partial_x w_s = 0 \right\}, \tag{2}$$

$$\mathbb{D}(\rho_0, \rho_1)^2 = \inf \left\{ \int_0^1 \int_{\mathbb{S}^1} \frac{w_s^2}{\rho_s} \, \mathrm{d}x \, \mathrm{d}s \,\middle|\, \partial_s \rho_s - \partial_{xx} w_s = 0 \right\}. \tag{3}$$

In those definitions, the infima are taken over all parametrized pairs $(\rho_s, w_s)_{s\in[0,1]}$ of probability densities ρ_s and Radon measures w_s on $\mathbb{S}^1$, respectively, that connect ρ_0 to ρ_1 by means of suitable zeroth, first and second order continuity equations, respectively, understood in the weak sense.

The space $(X, \mathbb{H})$ is a complete metric space with the L^1-topology, and $(\overline{X}, \mathbb{W})$ is a complete metric space, where the completion $\overline{X}$ is the space of probability measures on $\mathbb{S}^1$, with the narrow topology. A similar results holds for $(\overline{X}, \mathbb{D})$.

Theorem. *$(\overline{X}, \mathbb{D})$ is a complete metric space with the narrow topology and the diffusive transport metric $\mathbb{D}$ satisfies*

$$\big\|\mu_1 - \mu_0\big\|_{(\dot{W}^{2,\infty}(\mathbb{S}^1))'} \leq \mathbb{D}(\rho_0, \rho_1) \leq \frac{2}{-\log \big\|\mu_1 - \mu_0\big\|_{(\dot{H}^1(\mathbb{S}^1))'}}.$$

Besides being a generalization of the Hellinger and Wasserstein distances, the diffusive transport can be also seen as the relaxation of martingale transport. The dynamic formulation of martingale transport recently obtained in [3] is given by (3) with the additional martingale constraint manifesting in the condition $w_s \geq 0$ for

a.e. $s \in [0,1]$. In particular, for marginals in convex order, the geodesics of $\mathbb{D}$ agree with those of martingale transport.

However, for marginals not of convex order, which is the situation on any compact manifold like $\mathbb{S}^1$, geodesics w.r.t $\mathbb{D}$ are currently little understood, since there will be space-time regions for which the diffusive flux is negative $w_s < 0$. At least formally, the geodesic equations for $\mathbb{H}$, $\mathbb{W}$ and $\mathbb{D}$ read, respectively, as

$$\begin{aligned} \partial_s \rho_s - \rho_s\, \psi_s = 0, \quad \partial_s \psi_s + \tfrac{1}{2}\psi_s^2 = 0, \\ \partial_s \rho_s + \partial_x(\rho_s\, \partial_x \psi_s) = 0, \quad \partial_s \psi_s + \tfrac{1}{2}(\partial_x \psi_s)^2 = 0, \\ \partial_s \rho_s - \partial_{xx}(\rho_s\, \partial_{xx} \psi_s) = 0, \quad \partial_s \psi_s + \tfrac{1}{2}(\partial_{xx} \psi_s)^2 = 0. \end{aligned}$$

We recall that the first system is solvable by plain linear interpolation w.r.t. $\sqrt{\rho_s}$, and the second one is solvable in principle by the method of characteristics. Except for very special solution, the third one appears inaccessible to explicit constructions, which leaves also possible static formulations of $\mathbb{D}$ open for the moment.

2. Contractive and gradient flows

Observation 1. *The linear diffusion equation $\partial_t \rho = \partial_{xx} \rho$ induces on X ...*

- *... a contractive flow w.r.t. $\mathbb{H}$,*
- *... a contractive gradient flow w.r.t. $\mathbb{W}$,*
- *... a contractive flow w.r.t. $\mathbb{D}$.*

The contractivity properties are essentially consequences of Jensen's inequality and the fact that linear diffusion is a linear averaging process. The free energy for the gradient flow w.r.t. $\mathbb{W}$ is Boltzmann's entropy functional $\mathcal{H}(\rho) = \int \rho \log \rho \, \mathrm{d}x$. On $\mathbb{R}$ instead of $\mathbb{S}^1$, the linear diffusion equation is actually a gradient flow w.r.t. $\mathbb{D}$ for the negative second moment $-\frac{1}{2}\int x^2 \rho(x)\, \mathrm{d}x$.

Observation 2. *The DLSS equation* [1] $\partial_t \rho = -\partial_{xx}(\rho\, \partial_{xx} \log \rho)$ *is on X ...*

- *... a contractive flow* [4] *w.r.t. $\mathbb{H}$;*
- *... a (non-contractive) gradient flow* [2] *w.r.t. $\mathbb{W}$ for the Fisher information $\mathcal{F}(\rho) = \int \rho\, (\partial_x \log \rho)^2 \, \mathrm{d}x$;*
- *... a (non-contractive) gradient flow* [7] *w.r.t. $\mathbb{D}$ for the entropy $\mathcal{H}$.*

There is apparently no easy explanation for the contractivity in $\mathbb{H}$.

3. Discretization

Consider an equidistant discretization of $\mathbb{S}^1$ of mesh width $\delta > 0$, denote the space of piecewise constant probability densities by X^δ. A mere restriction of the distances $\mathbb{W}$ or $\mathbb{D}$ to X^δ would produce metric spaces with pathological properties. Instead, the definitions of $\mathbb{H}$, $\mathbb{W}$ and $\mathbb{D}$ can be modified to provide adapted distances $\mathbb{H}^\delta$, $\mathbb{W}^\delta$ and $\mathbb{D}^\delta$ on X^δ: replace the derivative(s) in the continuity equations by (forward) difference quotients, and replace the denominator in w_s^2/ρ_s by a suitable mean value of the neighboring densities — simply ρ_k for $\mathbb{H}^\delta$, a two-point average $\mathbf{m}(\rho_k, \rho_{k+1})$ for $\mathbb{W}^\delta$, and a three-point average $\mathbf{M}(\rho_{k-1}, \rho_k, \rho_{k+1})$ for $\mathbb{D}^\delta$.

Observation 3. *The discretization* $\dot\rho_k = (\rho_{k+1} - 2\rho_k + \rho_{k-1})/\delta^2$ *of the linear diffusion equation by central finite differences induces on* X^δ *...*

- *... a contractive flow w.r.t.* $\mathbb{H}^\delta$*,*
- *... a contractive gradient flow w.r.t.* $\mathbb{W}^\delta$ [5, 6]
- *... a contractive flow w.r.t.* $\mathbb{D}^\delta$ [7].

Contractivity follows again by the linear averaging effect of the (discretized) diffusion. For the appropriate mean in the definition of $\mathbb{W}^\delta$, one uses the logarithmic mean $\mathbf{m}(\rho_k, \rho_{k+1}) = (\rho_{k+1} - \rho_k)/\log(\rho_{k+1} - \log \rho_k)$, and in the definition of $\mathbb{D}^\delta$, one uses $\mathbf{M}(\rho_{k-1}, \rho_k, \rho_{k+1}) = \rho_k$.

Observation 4 ([7])**.** *The following discretization of the DLSS equation*

$$\dot\rho_k = \big(F_{k+1} - 2F_k + F_{k-1}\big)/\delta^2, \quad F_\ell = \big(\sqrt{\rho_{\ell+1}\rho_{\ell-1}} - \rho_\ell\big)/\delta^2 \tag{4}$$

induces on X^δ *...*

- *... a contractive flow w.r.t.* $\mathbb{H}^\delta$
- *... a (non-contractive) gradient flow w.r.t.* $\mathbb{W}^\delta$ *for some suitable discrete Fisher information* $\mathcal{F}^\delta$
- *... a (non-contractive) gradient flow w.r.t.* $\mathbb{D}^\delta$ *for the discrete entropy* $\mathcal{H}^\delta$*.*

Differently from Observation 3, we choose $\mathbf{m}(\rho_k, \rho_{k+1}) = \sqrt{\rho_{k+1}\rho_k}$ for $\mathbb{W}^\delta$, and for $\mathbb{D}^\delta$ the 3-stencil mobility

$$\mathbf{M}(\rho_{k-1}, \rho_k, \rho_{k+1}) = \frac{\sqrt{\rho_{k+1}\rho_{k-1}} - \rho_k}{\log\sqrt{\rho_{k+1}\rho_{k-1}} - \log\rho_k}.$$

These choices of mobilities $\mathbf{m}/\mathbf{M}$ appear to be crucial to guarantee the contractivity in $\mathbb{H}^\delta$. Indeed, the proof uses that (4) can be re-formulated as

$$\partial_t\sqrt{\rho_k} = -\frac{u_{k+1} - 2u_k + u_{k-1}}{\delta^2} + \frac{u_k^2}{\sqrt{\rho_k}} \quad \text{with} \quad u_k = \frac{\sqrt{\rho_{k+1}} - 2\sqrt{\rho_k} + \sqrt{\rho_{k-1}}}{\delta^2},$$

which is the discrete analog of the identity $\partial_t\sqrt{\rho} = -\partial_{xxxx}\sqrt{\rho} + \frac{(\partial_{xx}\sqrt{\rho})^2}{\sqrt{\rho}}$.

Formally, the sceme (4) provides a second order approximation of the DLSS equation $\partial_t\rho = -\partial_{xx}(\rho\partial_{xx}\log\rho)$ as can be seen by the expansion of the flux

$$F_\ell = -\frac{2}{\delta^2}\rho_\ell\Big[\exp\big(\log\sqrt{\rho_{\ell-1}\rho_{\ell+1}} - \log\rho_\ell\big) - 1\Big] \approx -\rho\partial_{xx}\log\rho + O(\delta^2).$$

Our main result is about that those heuristics are indeed correct and scheme (4) converges to a weak solution of hte DLSS equation.

Theorem ([7])**.** *Let an initial condition* $\hat\rho \in X$ *be given. For each mesh width* δ*, consider a strictly positive approximation* $\hat\rho^\delta \in X^\delta$ *of* $\hat\rho$*. Then the initial value problem for* (4) *possesses a unique solution* $\rho^\delta : [0,\infty) \to X^\delta$*, and*

$$\rho^\delta \to \rho^* \quad \text{in} \quad L^1_{\mathrm{loc}}\big((0,\infty)\times\mathbb{S}^1\big) \cap C^\alpha\big([0,\infty); (W^{2,\infty}(\mathbb{S}^1))'\big) \quad \text{as } \delta \to 0,$$

where ρ^* *is a weak solution to the DLSS equation.*

The proof heavily uses the properties stated in Observation 4, particularly the contractivity in $\mathbb{H}^\delta$ and the monotonicity of $\mathcal{H}$. The key a priori estimate is

$$-\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{H}(\rho^\delta) \geq \delta \sum_k \left(\frac{\sqrt{\rho_{k+1}} - 2\sqrt{\rho_k} + \sqrt{\rho_{k+1}}}{\delta^2} \right)^2,$$

which provides weak compactness of the $\sqrt{\rho^\delta}$ in $L^2\big((0,\infty); H^2(\mathbb{S}^1)\big)$.

References

[1] B. Derrida, J. L. Lebowitz, E. R. Speer, and H. Spohn. Fluctuations of a stationary nonequilibrium interface. *Phys. Rev. Lett.*, 67:165–168, 1991.

[2] U. Gianazza, G. Savaré, G. Toscani, *The Wasserstein gradient flow of the Fisher information and the quantum drift-diffusion equation*, Arch. Ration. Mech. **194** (2009), 133–220.

[3] M. Huesmann and D. Trevisan. A Benamou-Brenier formulation of martingale optimal transport. *Bernoulli*, 25(4A):2729–2757, 2019.

[4] A. Jüngel, R. Pinnau, *Global nonnegative solutions of a nonlinear fourth-order parabolic equation for quantum systems*, SIAM J. Math. Anal. **32** (2000), 760–777.

[5] J. Maas, *Gradient flows of the entropy for finite Markov chains*, J. Funct. Anal. **261** (2011), 2250–2292.

[6] A. Mielke, *Geodesic convexity of the relative entropy in reversible Markov chains*, Calc. Var. Partial Differ. Equ. **48** (2013), 1–31.

[7] D. Matthes, E.-M. Rott, G. Savaré, A. Schlichting, *A structure preserving discretization for the Derrida-Lebowitz-Speer-Spohn equation based on diffusive transport*, `arXiv:2312.13284` (2023), 36pp.

Entropic optimal martingale transport

Jean-David Benamou

(joint work with Guillaume Chazareix, Marc Hoffmann, Grégoire Loeper, François-Xavier Vialard)

Entropic regularisation is a powerful theoretical and numerical tool for optimal transport problems. We introduce a new concept of "specific relative entropy regularisation" (3) for a prototype stochastic optimal transport problem (1-2) leading, after time discretisation, to an unconstrained multi-marginal entropic (in the classical sense) optimal transport problem. Conversely to "standard" relative entropy regularisation (see [6] [2] for instance) freezing volatility and leaving the control to the drift; "specific" relative entropy induces, through time discretisation, a local in time and space volatility control on the process, hence the title of our contribution. The martingale part is here the optimiser and not simply a constraint or a regularisation. This note is a formal summary, see [3] for a complete description and a numerical study. For applications in mathematical finance see [5] [4]. Our starting point is a class of optimal control problems governed by diffusion processes of the form [7] [5] :

$$\inf_{(\mathbb{P}, b_t, a_t) \in DP(\rho_0, \rho_1)} \mathbb{E}_{\mathbb{P}} \left(\int_0^1 F(b_t(X_t), a_t(X_t))\, dt \right) \tag{1}$$

$$DP(\rho_0,\rho_1) = \left\{ \begin{array}{l} (\mathbb{P}, b_t, a_t) \quad s.t. \\ dX_t = b_t(X_t)\,dt + \sqrt{a_t(X_t)}\,dW_t\,. \\ \mathbb{P}_t := e_{t\#}\,\mathbb{P} \sim X_t, \quad t \in (0,1), \\ e_{0\#}\,\mathbb{P} = \rho_0 \quad e_{1\#}\,\mathbb{P} = \rho_1. \end{array} \right. \tag{2}$$

Here, ρ_0 and ρ_1 are the usual initial/final probabilities of optimal transport with finite second moments, W_t is the Wiener process, X_t the canonical process under $\mathbb{P}$ (a probability over paths in $H^1((0,1),\mathbb{R})$, (b_t, a_t) are time and space dependent deterministic functions acting as drift and diffusion coefficients controlling the process, $\mathbb{E}_{\mathbb{P}}$) is the expectation over $\mathbb{P}$. We also introduced the push-forward notation to disintegrate $\mathbb{P}$ into a curve in time of marginals: $\mathbb{P}_t := e_{t\#}\,\mathbb{P}$, the same notation will be used lated for joint probabilities: $\mathbb{P}_{s,t} := (e_s, e_t)_{\#}\,\mathbb{P}$. The cost function F is convex and has to enforce the positivity of a_t. A simple example is the classical dynamic optimal transport problem: $F = \|b\|^2/2 + \chi_{a_t=0}$. Note that the set of diffusions process $DP(\rho_0,\rho_1)$ is non-empty, containing at least the (ρ_0,ρ_1) Schroedinger bridge. Specific relative entropy between diffusion processes $\mathbb{P}$ and $\overline{\mathbb{P}}$ ([1] for details and references) is defined as ($\mathcal{S}^{\mathcal{I}}$ given in (5)):

$$\mathcal{S}^{\mathcal{H}}(\mathbb{P}|\overline{\mathbb{P}}) := \lim_{h\searrow 0} h\,\mathcal{H}(\mathbb{P}^h|\overline{\mathbb{P}}^h) = \frac{1}{2}\mathbb{E}_{\mathbb{P}}\left(\int_0^1 \mathcal{S}^{\mathcal{I}}(a_t(X_t)|\overline{a}))\,dt\right). \tag{3}$$

where $\mathbb{P}^h = (e_{t_0},\ldots,e_{t_N}))_{\#}\mathbb{P}$ is a time discretisation of $\mathbb{P}$ ($t_i = i\,h$ and $h = 1/N$ a time step) and same for $\overline{\mathbb{P}}^h$, $\mathcal{H}$ is the usual relative entropy. We choose $\overline{\mathbb{P}}$ to be a martingale with initial distribution ρ_0 and volatility $\sqrt{\overline{a}}$ (later possibly local in time and space). Formally, $\mathcal{S}^{\mathcal{I}}$ is explicitly identified using the discrete relative entropy

$$h\,\mathcal{H}(\mathbb{P}^h|\overline{\mathbb{P}}^h) = \frac{1}{2}\mathbb{E}_{\mathbb{P}^h}\left(h\sum_{i=0}^{N-1}\frac{a_i^h(X_i^h)}{\overline{a}} - 1 - \log\left(\frac{a_i^h(X_i^h)}{\overline{a}}\right)\right) + O(h)$$

with $((X_i^h)_{i=0..N}$ being the discrete Markov chain associated to $\mathbb{P}^h$ i.e. $X_i^h \sim P_{t_i}^h$):

$$\left\{ \begin{array}{l} b_i^h(x_i) = \dfrac{1}{h}E_{\mathbb{P}^h}\left(X_{i+1}^h - x_i | X_i^h = x_i\right), \\ a_i^h(x_i) = \dfrac{1}{h}\mathbb{E}_{\mathbb{P}^h}\left((X_{i+1}^h - x_i)^2 | X_i^h = x_i\right). \end{array} \right. \tag{4}$$

Letting $h \searrow 0$ we get in (3):

$$\mathcal{S}^{\mathcal{I}}(a|\overline{a}) = \frac{a}{\overline{a}} - 1 - \log\left(\frac{a}{\overline{a}}\right). \tag{5}$$

This function is strictly convex with minimum at $\overline{a}$, a barrier for vanishing a and strictly increasing but just sub-linearly as $a \to +\infty$.
The "specific" entropy regularisation of (1) is

$$\inf_{(\mathbb{P}, b_t, a_t) \in DP(\rho_0,\rho_1)} \mathbb{E}_{\mathbb{P}}\left(\int_0^1 F(b_t(X_t), a_t(X_t)) + \mathcal{S}^{\mathcal{I}}(a_t(X_t)|\overline{a}))\,dt\right). \tag{6}$$

Well-posedness of this class of problems is discussed in [7] [5] for instance. In view of (3) the natural discretisation of the cost functional is:

$$\mathcal{I}^h(\mathbb{P}^h) = \mathbb{E}_{\mathbb{P}^h}\left(h\sum_{i=0}^{N-1} F(b_i^h(X_i^h), a_i^h(X_i^h))\right) + h\,\mathcal{H}(\mathbb{P}^h|\overline{\mathbb{P}})\,. \tag{7}$$

with (b_i^h, a_i^h), the discrete drift and quadratic variation increments defined in (4). The final modelling step is to relax the diffusion process constraint and solve as an approximation of (6-2):

$$\inf_{\left\{(\mathbb{P}^h),\ s.t.\ \mathbb{P}^h_{t_0} = \rho_0\ and\ \mathbb{P}^h_{t_1} = \rho_1\right\}} \mathcal{I}^h(\mathbb{P}^h). \tag{8}$$

The strict convexity and lower semi continuity is not immediately seen but obtained, as often in optimal transport, using a change of variable in the conditional moments $(b_i^h(X_i^h), a_i^h(X_i^h))$ uncovering a perspective function. Simplifying and abusing notations:

$$\mathbb{P}^h \to \mathbb{E}_{\mathbb{P}^h}\left(F(\frac{1}{h}\mathbb{E}_{\mathbb{P}^h}\left((X_{i+1}^h - X_i^h)^\beta|X_i^h\right))\right) \text{ is replaced by}$$
$$(\mathbb{P}^h_{t_i}(X_i^h), \frac{1}{h}\mathbb{E}_{\mathbb{P}^h_{t_{i+1}}}\left((X_{i+1}^h - X_i^h)^\beta\right)) \to \mathbb{E}_{\mathbb{P}^h_{t_i}}\left(F(\frac{1}{h}\frac{\mathbb{E}_{\mathbb{P}^h_{t_{i+1}}}\left((X_{i+1}^h - X_i^h)^\beta\right)}{\mathbb{P}^h_{t_i}(X_i^h)})\right).$$

The Markovianity of the minimisers is a direct consequence of the structure of the cost and allow to consider the minimisation under this change of variable. We thus obtain the relative entropic regularisation of a multi marginal optimal transport problem involving moments of $\mathbb{P}$. Under further assumptions on F, essentially L^p control of (b, a) and also over higher order conditional moments of $\mathbb{P}^h$, we claim: *i*) that sequences of minimisers $(\mathbb{P}^h)_h$ admit weakly converging subsequences as $h \searrow 0$, *ii*) that the limit $\mathbb{P}^0$ is in $DP(\rho_0, \rho_1)$ (in a weak sense [9]) and *iii*) $\mathbb{P}^0$ is a minimiser of (6). A key point of the proof is to remark that for any Markov chain $\mathbb{P}^h$ (i.e. not necessarily the discretisation of a diffusion process):

$$h\,\mathcal{H}(\mathbb{P}^h|\overline{\mathbb{P}}^h) \geq \frac{1}{2}\mathbb{E}_{\mathbb{P}^h}\left(h\sum_{i=0}^{N-1}\mathcal{S}^{\mathcal{I}}(a_h^i(X_i^h)|\overline{a})\right),$$

giving a control of the h-scaled discrete relative entropy over the specific entropy. The function $\mathcal{S}^{\mathcal{I}}$ (5) seems unfortunately too weak to enforce point *ii*) and stronger assumption on F are needed to meet the criteria in [8] for the convergence of Markov chains towards diffusion processes. After space discretisation and truncation, neglecting the tails of the controlled diffusion, (8) can be solved as in [2] in its dual formulation using Sinkhorn algorithm. Sinkhorn iterates are not explicit and dealt with a newton method. As we now are on a compact space can consider a simple test case: $F(b, a) = \alpha\,\|b\|^2$ ($\alpha >> 1$) strongly penalising the drift (a soft martingale constraint). We use a parabolic scaling for the space grid size $dx^2 = h$. The numerical test (more in [3]) is for marginal data (figure A) not in convex order and a constant reference measure volatility $\overline{a}$ (figure B). The implementation can

be optimised to give a linear time complexity in the number of space-time grid points (see figure G) when $h = 1/N$ goes to 0. The final drift/volatility surfaces (figures F and E) are consistent with F and $\mathcal{S}^{\mathcal{I}}(a|\overline{a})$. Figure H illustrates the specific relative entropy definition (3).

References

[1] J. Backhoff-Veraguas, C. Unterberger, *On the specific relative entropy between martingale diffusions on the line*, Preprint (2023).

[2] J.-D. Benamou, G. Carlier, S. Di Marino, L. Nenna, *An entropy minimization approach to second-order variational mean-field games*, Math. Models Methods Appl. Sci., 29 (2019) 1553-1583.

[3] J.D. Benamou , G. Chazareix, M. Hoffman, G. Loeper, F.-X. Vialard, *Entropic optimal martingale transport*, in preparation (2024).

[4] J.D. Benamou , G. Chazareix, G. Loeper, *Local volatility calibration using entropic martingale transport* in preparation (2024).

[5] I. Guo, G. Loeper, J. Obloj, S. Wang, *Joint Modeling and Calibration of SPX and VIX by Optimal Transport*, SIAM Journal on Financial Mathematics, 13,1 (2022) 1-31.

[6] C. Léonard, *A survey of the Schrödinger problem and some of its connections with optimal transport*, Discrete Contin. Dyn. Syst. 34 (2014) 1533-1574.

[7] X. Tan, N. Touzi, *Optimal transportation under controlled stochastic dynamics*, The Annals of Probability (41) 5 (2013).

[8] D.W. Stroock, S.R.S. Varadhan, *Multidimensional Diffusion Processes*, Classics in Mathematics, Springer (2007).

[9] D. Trevisan, *Well-posedness of multidimensional diffusion processes with weakly differentiable coefficients*, Electronic Journal of Probability 21 (2015).

Sharp comparisons between the sliced Wasserstein and Wasserstein distances

Quentin Mérigot

(joint work with Guillaume Carlier, Alessio Figalli, Yi Wang)

Sliced Wasserstein distances have originally be introduced by Marc Bernot as a computationally inexpensive variant of the optimal transport distances. The sliced Wasserstein distance with exponent $p \geq 1$ between two probability measures μ, ν on $\mathbb{R}^d$ with finite pth moment is defined as an average of (one-dimensional) Wasserstein distances between orthogonal projection of these measures on lines:

$$SW_p(\mu,\nu)^p = \int_{\theta\in\mathcal{S}^{d-1}} W_p(P_{\theta\#}\mu,\nu)^p d\theta.$$

In the previous formula, given a unit vector θ, the projection $P_\theta : \mathbb{R}^d \to \mathbb{R}$ is defined by $P_\theta(x) = x \cdot \theta$. Moreover, we assume that the surface measure on the unit sphere $\mathcal{S}^{d-1}$ is normalized to one. Sliced Wasserstein distances have been used first in the context of texture generation and generally imaging science [3] and have since then emerged as a useful variant of optimal transport distances in many applications. Their statistical and computational advantages make these distances and particularly suited to high-dimensional statistics and machine learning applications. We refer to the PhD thesis [1] and references therein. The interest of

the analysis/PDE side of the optimal transport community to sliced-Wasserstein distances is more recent [2, 4], with the notable exception of [5].

Our aim in this talk is to understand quantitatively the relation between the sliced Wasserstein distance and the standard Wasserstein distance in the case $p = 1$. The first (immediate) remark is that since the projection P_θ is 1-Lipschitz,

$$SW_1(\mu, \nu) \leq W_1(\mu, \nu).$$

Conversely, Bonnotte established an upper bound on the Wasserstein distance from the sliced one, for measures supported in a ball [5]. More precisely, for any $R > 0$, he showed that there exists a constant C_R such that

$$\forall \mu, \nu \in \mathcal{P}(B(0, R)), \quad W_1(\mu, \nu) \leq C_R SW_1(\mu, \nu)^{\frac{1}{d+1}}. \tag{1}$$

A similar Hölder comparison result between distances on probability measures and sliced distances has been established in the much older article [6]. In this article, the authors compare the bounded-Lipschitz distance $\mathrm{d_{bL}}$ between probability measures on a fixed ball, and the maximum sliced distance $\max_\theta \mathrm{d_{bL}}(P_{\theta\#}\mu, P_{\theta\#}\nu)$.

The aim of our work was to see by how much the exponent in the estimation of Bonnotte (1) can be improved. We prove the following: for any $R > 0$, there exists a constant C_R such that

$$\forall \mu, \nu \in \mathcal{P}(B(0, R)), \quad W_1(\mu, \nu) \leq C_R SW_1(\mu, \nu)^{\frac{1}{d}}. \tag{2}$$

Moreover, the exponent $1/d$ is sharp: there exists two family of probability measures $(\mu_\varepsilon)_\varepsilon$ and $(\nu_\varepsilon)_\varepsilon$ supported in a fixed ball for ε small enough, such that

$$W_1(\mu_\varepsilon, \nu_\varepsilon) \geq C\varepsilon \text{ and } SW_1(\mu_\varepsilon, \nu_\varepsilon) \leq C'\varepsilon^d.$$

In addition to this result, we present similar comparisons between general sliced distances which involve average of Wasserstein distances between the projections of μ, ν on k-dimensional linear subspaces of $\mathbb{R}^d$.

References

[1] Nadjahi, K. (2021). Sliced-Wasserstein distance for large-scale machine learning: theory, methodology and extensions (Doctoral dissertation, Institut polytechnique de Paris).

[2] Park, S., Slepčev, D. (2023). Geometry and analytic properties of the sliced Wasserstein space. arXiv preprint arXiv:2311.05134.

[3] Rabin, J., Peyré, G., Delon, J., Bernot, M. (2012). Wasserstein barycenter and its application to texture mixing. In Scale Space and Variational Methods in Computer Vision: Third International Conference, SSVM 2011

[4] Kitagawa, J., Takatsu, A. (2023). Two new families of metrics via optimal transport and barycenter problems. arXiv preprint arXiv:2311.15874.

[5] Bonnotte, N. (2013). Unidimensional and evolution methods for optimal transportation (Doctoral dissertation, Université Paris Sud-Paris XI; Scuola normale superiore (Pise, Italie)).

[6] Hahn, M. G., Quinto, E. T. (1985). Distances between measures from 1-dimensional projections as implied by continuity of the inverse Radon transform. Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete, 70(3), 361–380.

Geometry of sliced optimal transport and projected-transport gradient flows

Dejan Slepčev
(joint work with Sangmin Park and Lantian Xu)

The sliced-Wasserstein (SW) distance, introduced by Rabin, Peyré, Delon, and Bernot [5], compares probability measures on $\mathbb{R}^d$ with finite second moment ($\mathscr{P}_2(\mathbb{R}^d)$) by taking averages of the Wasserstein distances between projections of the measures to 1-dimensional subspaces of $\mathbb{R}^d$. For $\theta \in \mathbb{S}^{d-1}$ let $\pi^\theta : \mathbb{R}^d \to \mathbb{R}$ be the projection to span$\{\theta\}$:

$$\pi^\theta(x) = \theta \cdot x.$$

The sliced Wasserstein distance SW is defined by

$$SW(\mu,\nu) = \left(\frac{1}{|\mathbb{S}^{d-1}|} \int_{\mathbb{S}^{d-1}} W^2(\pi^\theta_\# \mu, \pi^\theta_\# \nu)\, d\theta \right)^{\frac{1}{2}}$$

where $\pi^\theta_\#$ denotes the pushforward of a measure by π^θ, W is the Wasserstein distance, and $d\theta$ denotes the integration with respect to surface measure.

The sliced Wasserstein distance has found a variety of applications in statistics and machine learning due to the fact that it can be approximated well in high dimensions based on finite samples of the measure [2, 3], which is not the case for the Wasserstein metric.

Our goal was to gain a better understanding of the properties of the metric and its associated geometry. We first establish that $(\mathscr{P}_2(\mathbb{R}^d), SW)$ is not a length space, which leads us to investigate sliced Wasserstein length metric ℓ_{SW} (the infimum of the lengths of curves between measures in the SW-space), in addition to the SW metric.

Comparing sliced Wasserstein metric locally with with negative Sobolev norms and the Wasserstein metric. We show that sliced Wasserstein metric exhibits rather different properties near measures with positive densities with respect to the Lebesgue measure and near discrete measures. In particular, consider an absolutely continuous measure μ bounded away from zero and infinity on some bounded open convex domain Ω. For all measures μ, ν which are within constant multiples of the Lebesgue measure restricted to Ω, we show

$$(1)\quad \|\mu-\nu\|_{\dot{H}^{-\frac{d+1}{2}}(\mathbb{R}^d)} \lesssim SW(\mu,\nu) \le \ell_{SW}(\mu,\nu) \lesssim SW(\mu,\nu) \lesssim \|\mu-\nu\|_{\dot{H}^{-\frac{d+1}{2}}(\mathbb{R}^d)},$$

where the rightmost inequality additionally requires ν to coincide with μ near the boundary of Ω. In other words, near μ, SW is equivalent to $\dot{H}^{-\frac{d+1}{2}}$.

On the other hand, for ν near discrete measures, $\mu^n = \sum_{i=1}^n m_i \delta_{x_i}$,

$$(2)\qquad SW(\mu^n,\nu) \le \ell_{SW}(\mu^n,\nu) \le \frac{1}{d} W(\mu^n,\nu) \le (1+o(1)) SW(\mu^n,\nu).$$

In conclusion, near smooth measures SW behaves like a highly negative Sobolev space, in contrast to the Wasserstein metric which for such measures behaves like

the $\dot{H}^{-1}$ norm as shown by Peyre [4], while near discrete measures SW behaves like the Wasserstein distance.

Approximation by discrete measures in sliced Wasserstein length. The results of [2, 3] show that finite-sample estimation of measures with respect to Sliced Wasserstein distance enjoys parametric rate. We remark that finite-sample estimation of measures with respect to maximum mean discrepancy (MMD) also enjoys parametric rate [6, Theorem 3.3]. MMD distance is nothing but the norm in the dual of a reproducing kernel Hilbert space (RKHS). In particular the results of [6] apply to the dual of the Sobolev space H^s with $s > \frac{d}{2}$ (when the spaces embeds in the spaces of Hölder continuous functions and are RKHS). The comparison (1) says that near absolutely continuous measures, SW behaves like $\dot{H}^{-(d+1)/2}$-norm; as the associated norm $\|\cdot\|_{H^{-(d+1)/2}(\mathbb{R}^d)}$ is an MMD, we can formally understand SW to exhibit behaviors like an MMD. Thus MMD parametric estimation can be seen as a tangential or a linearized analogue of the finite sample estimation rates in SW distance.

We establish that finite sample approximation in ℓ_{SW} happens at the parametric rate up to a logarithmic correction, namely that

$$SW(\mu, \mu^n) \leq \ell_{SW}(\mu, \mu^n) \lesssim \sqrt{\frac{\log n}{n}} \text{ with high probability,}$$

where $\mu^n = \frac{1}{n}\sum_{i=1}^n \delta_{X_i}$ with $X_i \overset{i.i.d.}{\sim} \mu$. As mentioned, this is in stark contrast with the Wasserstein distance where the approximation rate scales like $n^{-\frac{1}{d}}$.

Implications to gradient flows. The comparison results on ℓ_{SW} and SW can be used to obtain comparisons for the metric slopes. Given a metric space (X, m), recall that metric slope $|\partial\mathcal{E}|_m$ of a functional $\mathcal{E} : X \to \mathbb{R}$ is defined by

$$|\partial\mathcal{E}|_m(u) = \limsup_{v \xrightarrow{m} u} \frac{[\mathcal{E}(u) - \mathcal{E}(v)]_+}{m(u, v)}. \tag{3}$$

Consider the potential energy $\mathcal{V}(\mu) := \int_{\mathbb{R}^d} V(x)\, d\mu(x)$. When V is smooth and compactly supported, for suitable absolutely continuous $\mu \in \mathscr{P}_2(\mathbb{R}^d)$ it holds that

$$|\partial\mathcal{V}|_{\dot{H}^{(d+1)/2}(\mathbb{R}^d)}(\mu) \lesssim |\partial\mathcal{V}|_{\ell_{SW}}(\mu) \leq |\partial\mathcal{V}|_{SW}(\mu) \lesssim |\partial\mathcal{V}|_{\dot{H}^{(d+1)/2}(\mathbb{R}^d)}(\mu) \tag{4}$$

whereas the slope behaves quite differently at discrete measures, $\mu^n = \sum_{i=1}^n m_i \delta_{x_i}$, namely that

$$|\partial\mathcal{V}|_{SW}(\mu^n) = |\partial\mathcal{V}|_{\ell_{SW}}(\mu^n) = \sqrt{d}\, |\partial\mathcal{V}|_W(\mu^n). \tag{5}$$

Hence $|\partial\mathcal{V}|_{SW}$ (resp. $|\partial\mathcal{V}|_{\ell_{SW}}$) is not lower-semicontinuous in SW (resp. ℓ_{SW}) in general, even when $V \in C_c^\infty(\mathbb{R}^d)$. This implies that the potential energy is not λ-geodesically convex in $(\mathscr{P}_2(\mathbb{R}^d), \ell_{SW})$. Furthermore, the curves of maximal slope in the sliced Wasserstein space starting from discrete measures, after a constant rescaling of time, are the curves of maximal slope in W space. On the other hand, for smooth measures, the curves of maximal slope with respect to the Wasserstein metric are not curves of maximal slope in SW space. We formally

show that SW gradient flow of potential energy is a higher order equation given by a pseudodifferential operator of order d.

Projected Wasserstein gradient flows. To obtain gradient flows of relative entropy that can be approximated well in high dimensions we introduce the projected Wasserstein distance where the space of velocities has been restricted to have low complexity. Namely first consider the local metric with a very small space of tangent velocities: Given a unit vector θ let

$$g_\theta(w,w) = \begin{cases} \int u(x\cdot\theta))^2 d\rho(x) & \text{if } w_\theta(x) = \theta u(\theta\cdot x) \\ \infty & otherwise. \end{cases}$$

Consider the *projected* metric $\overline{g}$ given by

$$\begin{aligned}\overline{g}(v,v) &= \inf\left\{ \int_{\S^{d-1}} g_\theta(w_\theta, w_\theta) d\theta \ : \ v(x) = \int_{\S^{d-1}} w_\theta dS(\theta) \right\} \\ &= \inf\left\{ \int_{\S^{d-1}} g_\theta(w_\theta, w_\theta) d\theta \ : \ v = \vec{R}^* w \right\}.\end{aligned}$$

The gradient flow of relative entropy

$$E(\rho) = \int \log \frac{\rho}{\pi} d\rho$$

(where $\pi \sim e^{-U}$), with respect to $\overline{g}$ is

$$\begin{aligned}&\partial_t \rho + \nabla\cdot(\rho v) = 0 \\ &v = -\int_{\S^{d-1}} \theta \left(\partial_s \ln(R_\theta \rho) + \frac{R_\theta(\rho \nabla U\cdot\theta)}{R_\theta \rho} \right)(x\cdot\theta) d\theta \\ &\quad = -R^* \nabla \ln(R\rho) + R^* \frac{R_\theta(\rho\nabla U\cdot\theta)}{R_\theta\rho}\end{aligned}$$

where for $s\in\mathbb{R}$ the Radon transform

$$R_\theta(s) = \int_{\theta^\perp} f(s\theta + y) dy.$$

The key property of the flow is that the equations are closed for any θ.

While well-posedness of this gradient flow is still open, we discussed that it can be approximated accurately in high dimensions using particle approximations. Numerical experiments in 256d indicate that the flow can be computed in relatively high dimensions and that its output approximates the target measure π well.

References

[1] L. Ambrosio and N. Gigli, and G. Savaré. *Gradient flows in metric spaces and in the space of probability measures.* Lectures in Mathematics ETH Zürich, Birkhäuser Verlag, Basel, second ed, 2008.

[2] T. Manole, S. Balakrishan, and L. Wasserman. *Minimax confidence intervals for the sliced Wasserstein distance* Electronic Journal of Statistics, 16(2022), pp.2252–2345.

[3] K. Nadjahi, A. Durmus, L. Chizat, S. Kolouri, S. Shahrampour, U. Simsekli. *Statistical and Topological Properties of Sliced Probability Divergences* NeurIPS, 33 (2020), pp. 20802–20812.

[4] R. Peyre. *Comparison between W distance and $\dot{H}^{-1}$ norm, and localization of Wassesrstein distance* ESAIM:COCV, 24(2018), pp.1489-1501.
[5] J. Rabin, G. Peyré, J. Delon, and M. Bernot. *Wasserstein barycenter and its application to texture mixing* in Scale Space and Variational Methods in Computer Vision, A. M. Bruckstein, B.M. ter Haar Romeny, A. M. Bronstein, and M.M. Bronstein, eds., Berlin, Heidelberg, 2012, Springer Berlin Heidelberg, pp. 435–446.
[6] B. Sriperumbudur *On the optimal estimation of probability measures in weak and strong topologies* Bernoulli, 22 (2016), pp.1839–1893.

Weak optimal transport with unnormalized kernels

Nathael Gozlan
(joint work with Philippe Choné, Francis Kramarz)

Given two probability measures μ, ν on two compact Polish spaces $\mathcal{X}, \mathcal{Y}$ and a cost function $c : \mathcal{X} \times \mathcal{P}(\mathcal{Y}) \to \mathbb{R}$, where $\mathcal{P}(\mathcal{Y})$ denotes the set of all probability measures on $\mathcal{Y}$, the Weak Optimal Transport (WOT) problem is defined by

$$\mathcal{T}_c(\mu, \nu) = \inf_p \int c(x, p^x)\, \mu(dx)$$

where the infimum runs over all *probability* kernels $p = (p^x)_{x\in\mathcal{X}}$ such that $\nu(dy) = \int p^x(dy)\, \mu(dx)$.

This definition was first introduced in [6] and [1]. The WOT framework encompasses the classical Monge-Kantorovich Optimal Transport problem (which corresponds to a cost function c linear with respect to p), as well as several well studied variants of this problem such as the Entropy Regularized Optimal Transport or the Martingale Optimal Transport probems. General tools are available for studying WOT problems, such as a Kantorovich type duality formula or a cyclical monotonicity criterion for optimality ; see [6, 1, 2]. WOT also founds specific applications in the field of concentration of measure [8, 7, 6, 5]. Other applications are discussed in [3].

In [4] is introduced a new variant of the WOT problem involving unnormalized kernels (WOTUK problem). More precisely, given a cost function $c : \mathcal{X} \times \mathcal{M}(\mathcal{Y}) \to \mathbb{R}$, where $\mathcal{M}(\mathcal{Y})$ denotes the set of all finite non-negative measures on $\mathcal{Y}$, one considers

$$\mathcal{I}_c(\mu, \nu) = \inf_q \int c(x, q^x)\, \mu(dx) \tag{1}$$

where the infimum now runs over all *non-negative* kernels $q = (q^x)_{x\in\mathcal{X}}$ such that $\nu(dy) = \int q^x(dy)\, \mu(dx)$. Let us emphasize that, in this definition, $q^x(\mathcal{Y})$ is not always 1.

One of the main results of [4] is the following theorem which gives primal attainment and a Kantorovich type duality formula for the WOTUK problem.

Theorem. *Assume that*

(*A*) *c is lower bounded, convex w.r.t its second variable and jointly l.s.c*

(B) for all $x \in \mathcal{X}$ and $m \in \mathcal{M}(\mathcal{Y}) \setminus \{0\}$,

$$c'_\infty(x,m) := \lim_{\lambda\to\infty} \frac{c(x,\lambda m)}{\lambda} = +\infty.$$

Then, there exists a non-negative kernel q achieving equality in (1). *Moreover, it holds*

$$\mathcal{I}_c(\mu,\nu) = \sup_{f\in\mathcal{C}(\mathcal{Y})} \left\{ \int K_c f \, d\mu - \int f \, d\nu \right\}, \tag{2}$$

where $\mathcal{C}(\mathcal{Y})$ denotes the space of continuous functions on $\mathcal{Y}$ and

$$K_c f(x) = \inf_{m\in\mathcal{M}(\mathcal{Y})} \left\{ \int f \, dm + c(x,m) \right\}, \qquad x \in \mathcal{X}.$$

The duality formula (2) is actually true under less restrictive conditions than Assumption (B) (see [4, Theorem 4.3] for more details).

If Assumption (B) is not in force, the WOTUK problem can have no solution. To palliate this problem, a notion of weak solution is introduced in [4]. A probability measure π on $\mathcal{X}\times\mathcal{Y}$ with second marginal equal to ν is called a weak solution if it is a limit point of a sequence $\pi_n(dxdy) = \mu(dx)q_n(dy)$, with $(q_n)_{n\geq 1}$ a minimizing sequence for the problem (1). Under a suitable assumption on the cost function c, it can be shown that weak solutions are minimizers of the following functional

$$\bar{I}_c[\pi] = \int c\left(x, \frac{d\pi_1^{ac}}{d\mu}(x)\pi_x\right) d\mu(x) + \int c'_\infty(x,\pi_x)\, d\pi_1^s(x)$$

among probability measures π on $\mathcal{X}\times\mathcal{Y}$ with second marginal equal to ν and first marginal π_1 such that $\pi_1(\mathrm{Support}(\mu)) = 1$ and having the Lebesgue decomposition $\pi_1 = \pi^{ac} + \pi_1^s$ into absolutely continuous and singular parts with respect to μ. See [4, Theorem 3.7] for a precise statement.

As a byproduct of the Kantorovich type duality (2), a generalization of Strassen's characterization of the convex order [9] is obtained in [4]. It reads as follows: two compactly supported probability measures μ,ν on $\mathbb{R}^d$ are such that

$$\int f \, d\mu \leq \int f \, d\nu$$

for all $f : \mathbb{R}^d \to \mathbb{R}$ convex and positively 1 homogenous, if and only if there exists a non-negative kernel $q = (q^x)_{x\in\mathbb{R}^d}$ such that $\nu(dy) = \int q^x(dy)\,\mu(dx)$ and

$$\int y\, q^x(dy) = x,$$

for μ almost every $x \in \mathbb{R}^d$.

References

[1] J.-J. Alibert, G. Bouchitté, and T. Champion, *A new class of costs for optimal transport planning*, Eur. J. Appl. Math. **30** (2019), no. 6, 1229–1263.

[2] J. Backhoff-Veraguas, M. Beiglböck, and G. Pammer, *Existence, duality, and cyclical monotonicity for weak transport costs*, Calc. Var. Partial Differ. Equ. **58** (2019), no. 6, 28 (English), Id/No 203.

[3] J. Backhoff-Veraguas and G. Pammer, *Applications of weak transport theory*, Bernoulli **28** (2022), no. 1, 370–394.

[4] P. Choné, N. Gozlan, and F. Kramarz, *Weak Optimal Transport with Unnormalized Kernels*, SIAM J. Math. Anal. 55, No. 6, 6039–6092.

[5] N. Gozlan, C. Roberto, P.-M. Samson, Y. Shu, and P. Tetali, *Characterization of a class of weak transport-entropy inequalities on the line*, Ann. Inst. Henri Poincaré Probab. Stat. **54** (2018), no. 3, 1667–1693.

[6] N. Gozlan, C. Roberto, P.-M. Samson, and P. Tetali, *Kantorovich duality for general transport costs and applications*, J. Funct. Anal. **273** (2017), no. 11, 3327–3405.

[7] K. Marton, *Bounding $\overline{d}$-distance by informational divergence: a method to prove measure concentration*, Ann. Probab. **24** (1996), no. 2, 857–866.

[8] K. Marton, *A measure concentration inequality for contracting Markov chains*, Geom. Funct. Anal. **6** (1996), no. 3, 556–571.

[9] V. Strassen, *The existence of probability measures with given marginals*, Ann. Math. Statist. **36** (1965), 423–439.

An ordinary differential equation characterization of optimal transport and variants with linear constraints

Brendan Pass

(joint work with Luca Nenna)

Given marginal probability measures μ_i supported on compact sets $X_i \subseteq \mathbb{R}^n$, for $i = 1, 2, ...m$ and a cost function $c(x_1, ..x_m)$, the multi-marginal optimal transport problem is to minimize

$$\inf_{\gamma \in \Gamma(\mu_1,...,\mu_m)} \int_{X_1 \times \cdots \times X_m} c(x_1, ..., x_m) d\gamma.$$

Here $\Gamma(\mu_1, ..., \mu_m)$ represents the set of joint measures on the product space $X_1 \times ... \times X_m$ whose marginals are the μ_i. This problem is very challenging both theoretically, as well as computationally, because of the large state space.

The contribution described in this extended abstract is to characterize solutions, when each of the marginals μ_i is a discrete probability measure, via ordinary differential equations (ODEs). In fact, we deal with an entropically regularized version of the problem, which is, for a small $\eta > 0$, to minimize

$$\inf_{\gamma \in \Gamma(\mu_1,...,\mu_m)} \int_{X^1 \times \cdots \times X^m} c(x_1, ..., x_m) d\gamma + \eta H_{\otimes_{i=1}^n \mu_i}(\gamma), \tag{1}$$

where

$$H_{\otimes_{i=1}^n \mu_i}(\gamma) = \int_{X^1 \times \cdots \times X^n} \frac{d\gamma}{d(\otimes_{i=1}^n \mu_i)} \log\Big(\frac{d\gamma}{d(\otimes_{i=1}^n \mu_i)}\Big) d(\otimes_{i=1}^n \mu_i)$$

is the relative entropy (taken to be $+\infty$ is γ is not absolutely continuous with respect to the product measure $\otimes_{i=1}^n \mu_i$ of the marginals). It is well known that solutions to the regularized problem converge to solutions of the unregularized one as $\eta \to 0$. We will describe two different ODE characterizations; in the first, the entropic regularization is not entirely essential, but serves to make the solutions of the ODE smooth. In the the second project, the regularization is absolutely essential, as characterizing solutions via an ODE through this approach would be impossible without it.

In the first project [1], joint work with Luca Nenna, we consider pairwise costs of the form

$$c(x_1, ..., x_m) = \sum_{i \neq j} c_{ij}(x_i, x_j). \tag{2}$$

We introduce the one parameter family of costs c_ϵ defined by:

$$c(x_1, ..., x_m) = \sum_{i=2}^{m} c_{1i} c(x_1, x_i) + \epsilon \sum_{i,j>1 i \neq j} c_{ij}(x_i, x_j).$$

When $\epsilon = 1$, we of course have $c_\epsilon = c$. On the other hand, for $\epsilon = 0$, we have a cost function $c(x_1, ..., x_m) = \sum_{i=2}^m c_{1i}(x_1, x_i)$ which includes only pairwise interactions between x_1 and the other variables. For this reason, the $\epsilon = 0$ problem may be solved by solving the $m-1$ two marginal optimal transport problems between μ_1 and each μ_i, with cost functions c_{1i}. This is much less complex than solving the full multi-marginal problem (the complexity of solving m two marginal problems scales linearly in m whereas for the multi-marginal problem it scales exponentially). We then show that the evolution of the solution in ϵ can be characterized by a well posed ODE. This yields a new computational scheme; the problem with the original cost (2) can be solved by using the two marginals solutions as the initial condition at $\epsilon = 0$ and then solving the ODE (using an explicit Euler scheme, or, since solutions are smooth, a higher order Runge-Kutta method) up to $\epsilon = 1$.

The second project, joint work in progress with Nenna as well as Joshua Hiew, extends this approach to completely general (rather than only pairwise) cost functions [2]. We now take $c_\epsilon = \epsilon c$, so that the initial, $\epsilon = 0$ cost is $c_0 = 0$. In this case, the initial solution is even simpler than in the pairwise case above; we are now simply minimizing the entropy over $\Gamma(\mu_1, ..., \mu_m)$, and the unique minimizer is well known to be product measure. We again show that the the evolution of the solution can be characterized by a well-posed ODE. The resulting numerical scheme is of interest even in the two marginal case, where it yields an interpolation between product measure and solutions to optimal transport problems, and extends naturally to problems with additional linear constraints (including, for instance, martingale optimal transport).

Numerical simulations suggest that using the ODE method is roughly comparable to the Sinkhorn method (the standard method for solving regularized OT problems) in speed and accuracy. It has the additional advantage of yielding the minimizer γ_ϵ for each $\epsilon \in [0, 1]$, an object of independent interest. Computing this

curve of measures via the Sinkhorn would require a separate calculation for each ϵ, and would therefore be much more computationally cumbersome.

We also demonstrate the the formulation of the ODE allows one to easily calculate derivatives of the optimal cost $C(\epsilon)$ (that is, the value of the infimum in (1) with c replaced by c_ϵ) at $\epsilon = 0$. The zeroth and first order terms, $C(0) = \eta H_{\otimes_{i=1}^n \mu_i}(\gamma_0)$ and $C'(0) = \int_{X^1 \times \cdots \times X^m} c(x_1, ..., x_m) d\gamma_0$, are easily determined from the optimal coupling $\gamma_0 = \otimes_{i=1}^n \mu_i$. The ODE formulation allows one to calculate hire order derivatives; we illustrate this by providing a formula for $C''(0)$ when $m = 2$.

References

[1] Luca Nenna and Brendan Pass. An ODE characterisation of multi-marginal optimal transport. Preprint available at https://arxiv.org/abs/2212.12492.

[2] Joshua Zoen-Git Hiew, Luca Nenna and Brendan Pass. An ordinary differential equation for entropic optimal transport and its linearly constrained variants. In preparation.

Wasserstein valued BV maps

Simone Di Marino

(joint work with Rodolfo Assereto, Kristian Bredies, Emanuele Naldi, Silvia Villa)

In image processing, a TV regularizer is often used for the retrieval of an image given a corrupted version of it, expecially in the cases where we want to preserve sharp edges. In diffusion MRI, for every pixel a measure on the sphere (orientation distribution function) is given, that can be identified with a measure on $\mathbb{S}^2$. This gives rise to a natural object which is $\mu : D \to \mathcal{P}(\mathbb{S}^2)$. A natural procedure for recovering the *true* μ given some corrupted data $\hat{\mu}$ is the optimization problem

$$\operatorname{argmin} \left\{ \int_D W_1(\mu_x, \hat{\mu}_x)^2 \, dx + TV(\mu) \right\}, \tag{1}$$

where the first term is a fidelity term (with respect to W_1, the 1-Wasserstein distance), while the second one is a regularizer. In practice, there have been attempts to define the term $TV(\mu)$, stemming from a generalization to the classical dual formulation of total variation for functions of bounded variation in euclidean spaces [3]. Our goal is to give a list of other possible and natural definitions which will turn out to be equivalent. In the sequel we will consider $\mu : D \to \mathcal{P}(Y)$ where $D \subseteq \mathbb{R}^d$ and $Y \subseteq \mathbb{R}^n$.

(i) Primal formulation by relaxing cartoon-like functions. Inspired by a general approach for metric space valued BV functions [1] we consider finite range functions $\mathcal{C} = \{\mu \ : \ \mu = \sum_i \mu^i \chi_{A_i}\}$ where we can use a *geometric* BV energy inspired by the perimeter: $\mathcal{E}(\mu) := \sum_{i,j} \mathcal{H}^{d-1}(\bar{A}_i \cap \bar{A}_j) W_1(\mu^j, \mu^i)$. We can then define

$$TV_c(\mu) := \inf\{\liminf_{n\to\infty} \mathcal{E}(\mu_n) \ : \ \mu_n \to \mu \text{ in } L^1(W_1), \mu_n \in \mathcal{C}\}.$$

(ii) Primal formulation by relaxing smooth functions: whenever $\mu_x = \rho(x,y)\,dy$ and ρ is at least twice differentiable in both variables we can consider $TV_s^*(\mu) := \int_D \|\nabla_x \rho(x,\cdot)\|_{KR}\,dx$, where $\|\cdot\|_{KR}$ is a Katorovich-Rubenstein norm on the space of zero mean vector valued functions. We define:

$$TV_s(\mu) := \inf\{\liminf_{n\to\infty} TV_s^*(\mu_n) \ : \ \mu_n \to \mu \text{ in } L^1(W_1)\}$$

(iii) Beckmann formulation (which is explored in [2] in the context of functional lifting). Given a norm $\|\cdot\|$ on $n \times d$ matrices, we define:

$$TV_w(\mu) := \sup\{\|\sigma\|(D\times Y) \ : \ \sigma \in \mathcal{M}(D\times Y;\mathbb{R}^{n\times d}),\ \nabla_x \mu + \mathrm{div}_y(\sigma) = 0\}$$

(iv) Dual formulation, namely the original one studied in [3]. Given another norm $\|\cdot\|_*$ on $n \times d$ matrices, we define:

$$TV_d(\mu) := \sup\left\{\int_D \int_Y \mathrm{div}_x(\phi)\,d\mu_x(y)\,dx \ : \ \|\|\nabla_y \phi(x,y)\|_*\|_{L^\infty(Y)} \le 1,\ \ \forall x \in D\right\},$$

where the supremum is taken among all smooth functions $\phi : D \times Y \to \mathbb{R}^d$ which are uniformly Lipschitz with respect to y.

We can prove that all the formulations produce the same BV space, and moreover $TV_s(\mu) = TV_w(\mu) = TV_d(\mu)$ whenever there is compatibility of the norms, namely $\|\cdot\|_*$ (in definition (iv)) should be the dual norm of $\|\cdot\|$ (in definition (iii)), while for $\eta \in \mathcal{M}(Y;\mathbb{R}^d)$ such that $\eta(Y) = 0$ we have

$$\|\eta\|_{KR} = \sup\left\{\int_Y \langle \psi,\, d\eta\rangle \ : \ \psi \in \mathrm{Lip}(Y;\mathbb{R}^d),\ \ \|\nabla\psi(y)\|_* \le 1 \ \ \forall y \in Y\right\}.$$

To have equality also with (i) necessarily $\|\cdot\|$ should be the nuclear norm of matrices, that is the 1-Shatten norm. Consequently, $\|\cdot\|_*$ is the operator norm, so that $\|\cdot\|_{KR}$ is the Kanthorovich-Rubenstein norm in duality with the 1-Lipschitz functions in the metric sense (so the test functions are $\psi \in \mathrm{Lip}(Y;\mathbb{R}^d)$ such that $d_{\mathbb{R}^d}(\psi(y),\psi(y')) \le d_{\mathbb{R}^n}(y,y)$). We generalize these results to the unbalanced case. In this setting, we need to change the norm on the Lipschitz space: all the equivalences are quite robust, and the only formulation that really changes is the Beckmann one.

There are a lot of interesting future directions: for example, understanding the structure theorem for BV functions in this case (in particular for the Beckmann problem) or investigating the role of the norms in the optimization problem (1).

References

[1] L. Ambrosio, *Metric space valued functions of bounded variation*, Annali della Scuola Normale Superiore di Pisa-Classe di Scienze, **32**, no. 3 (1990), 439-478.

[2] H. Lavenant, *Lifting functionals defined on maps to measure-valued maps via optimal transport*, arXiv preprint arXiv:2309.02260. (2023).

[3] T. Vogt, J. Lellmann, *Measure-valued variational models with applications to diffusion-weighted imaging*, Journal of Mathematical Imaging and Vision. **60**, (2018) 1482–1502.

Lifting functionals via optimal transport

Hugo Lavenant

We investigate the problem of the lifting if functionals defined on maps to measure-valued maps. A simple example of lifting would be the following: given $f : Y \to [0, +\infty]$, we can always lift it as a *linear* function $\mathcal{T}_f$ defined on $\mathcal{P}(Y)$ the space of probability measures over Y. Indeed for a probability measure μ we set

$$\mathcal{T}_f : \mu \mapsto \int_Y f(y) \, \mathrm{d}\mu(y).$$

We now look at E a functional defined on *maps* from a space X into a space Y that we want to lift in $\mathcal{T}_E$ a functional defined on measure-valued maps μ from X into $\mathcal{P}(Y)$. If E is a functional of zero order then we can extend the previous lifting:

$$E : (u : X \to Y) \mapsto \int_X f(x, u(x)) \, \mathrm{d}x$$

is lifted in

$$\mathcal{T}_E : (u : X \to \mathcal{P}(Y)) \mapsto \int_X \int_Y f(x, y) \, \mathrm{d}\mu_x(y) \, \mathrm{d}x.$$

Difficulty (and interest!) arises when E is for instance the Dirichlet energy, or another functional depending on the *derivatives* of the map. There has been several liftings already proposed and studied, see e.g. [2, 4, 8]. Here the goal is to give a characterization of the possible liftings.

1. The case of the action

The case which is already well understood is the one of the action of a curve, when $X = [0, 1]$ is a segment of $\mathbb{R}$. We restrict to $Y = \mathbb{R}^d$ to be a Euclidean space. For the p-action, which reads

$$E(u) = \int_0^1 |\dot{u}_t|^p \, \mathrm{d}t,$$

there is one canonical lifting which has three equivalent formulations [6, 1].

The lifted energy is the action of the curve of measures $(\mu_t)_{t\in[0,1]}$ in the p-Wasserstein space:

$$\mathcal{T}_E(\mu) = \int_0^1 |\dot{\mu}_t|_{W_p}^p \, \mathrm{d}t,$$

being $|\mu_t|_{W_p}$ the metric derivative of the curve in the space $\mathcal{P}_p(\mathbb{R}^d)$ endowed with the p-Wasserstein distance W_p. This action has a "Eulerian" representation:

$$\mathcal{T}_E(\mu) = \min_{v:[0,1]\times\mathbb{R}^d\to\mathbb{R}^d} \left\{ \int_0^1 \int_{\mathbb{R}^d} |v(t, y)|^p \, \mathrm{d}\mu_t(y) \, \mathrm{d}t \ : \ \partial_t\mu + \mathrm{div}(\mu v) = 0 \right\},$$

where the infimum is taken among all velocity fields such that the continuity equation $\partial_t\mu + \mathrm{div}(\mu v) = 0$ is satisfied, that is, represents the motion of mass $(\mu_t)_t$

evolving in time. The last representation is the "Lagrangian" one: with e_t the evalluation map at time t it reads

$$\mathcal{T}_E(\mu) = \min_{Q\in\mathcal{P}(H^1([0,1],\mathbb{R}^d))} \left\{ \int_{H^1([0,1],\mathbb{R}^d)} \int_0^1 |\dot u_t|^p \,\mathrm{d}t \,\mathrm{d}Q(u) \ : \ \forall t,\, e_t\#Q = \mu_t \right\}.$$

Here minimization is done over Q a probability measure over curves whose superposition represents μ. All these expressions coincide, and moreover for the minimal v and Q, we have that Q-a.e. curve u satisfies the ODE $\dot u_t = v(t, u_t)$.

2. Functionals of order one

We move on to the case where $X = \Omega$ is an open bounded subset of $\mathbb{R}^q$ while $Y = \mathbb{R}^d$. We look at

$$E(u) = \int_\Omega W(\nabla u(x))\,\mathrm{d}x,$$

where W is a non-negative, convex and coercive function on $d \times q$ matrices. In the case W grows linearly at infinity we need to account for the singular part of the distributional derivative of u as usual, but we omit this subtlety here for the sake of the exposition. We want to mimic the case of the action. We will not develop the metric point of view on the lifting, referring to [4] in the case E is the Dirichlet energy and to Simone Di Marino's talk in the present workshop when E is a bounded variation norm. The *Eulerian* lifting reads

$$\hat{\mathcal{T}}_E(\mu) = \min_{v:\Omega\times\mathbb{R}^d\to\mathbb{R}^{d\times q}} \left\{ \int_\Omega \int_{\mathbb{R}^d} W(v(x,y))\,\mathrm{d}\mu_x(y)\,\mathrm{d}x \ : \ \nabla_x\mu + \mathrm{div}_y(\mu v) = 0 \right\}.$$

Here $v(x, y) \in \mathbb{R}^{d\times q}$ is the "density of Jacobian matrix" at the point (x, y). On the other hand the Lagrangian lifting reads: with $L^0(\Omega, \mathbb{R}^d)$ the space of measurable functions from Ω to $\mathbb{R}^d$,

$$\mathcal{T}_E(\mu) = \min_{Q\in\mathcal{P}(L^0(\Omega,\mathbb{R}^d))} \left\{ \int_{L^0(\Omega,\mathbb{R}^d)} E(u)\,\mathrm{d}Q(u) \ : \ \text{“}\forall x,\, e_x\#Q = \mu_x\text{”} \right\}$$

The condition $\forall x,\, e_x\#Q = \mu_x$ is written between "$\cdot$" because a special care has to be given to define it, as all functions are defined only for a.e. x. We refer to [5] and not discuss the issue here.

Contrary to the case of the action of a curve, generically there holds

$$\mathcal{T}_E \neq \hat{\mathcal{T}}_E.$$

The reason is that, if there were equality, then for the optimal Q and v, we should have that Q-a.e. map u satisfies $\nabla u(x) = v(x, u(x))$. However this equation is in general not solvable: not every matrix field is a field of Jacobian matrices.

Our results, stated and proved in [5], aim at characterizing these liftings as the optimal ones in some sense. For a map u, let us write μ_u for the measure-valued map sending x to $\delta_{u(x)}$: this is the canonical embedding of a map u as a measure-valued map. We proved the following.

(1) The Lagrangian lifting $\mathcal{T}_E$ is the largest lifting which is convex, lower semi-continuous and which satisfies the lifting identity $\mathcal{T}_E(\mu_u) = E(u)$ for any map u.
(2) The Eulerian lifting $\hat{\mathcal{T}}_E$ is the largest lifting which is convex, lower semi-continuous, subadditive, increasing, inner regular and which satisfies the lifting identity $\hat{\mathcal{T}}_E(\mu_u) = E(u)$ for any map u.

As an immediate corollary we deduce $\hat{\mathcal{T}}_E \leq \mathcal{T}_E$. Here the topology we put on measure-valued maps is simply the topology of narrow convergence on the product space $\Omega \times \mathbb{R}^d$, by identifying a measure-valued map with the "fubinization" it induces on the product space. For the Eulerian lifting, by subadditive, increasing and inner regular we refer here to the localized version of the functional, see [3] for more details. The result for the Lagrangian lifting, which is actually interpreted as an optimal transport problem with an infinity of marginals, could be valid for much more general functionals E. It generalizes the result of [7] characterizing the value of the (classical) optimal transport problem as a suitable convex relaxation.

3. An open question

Consider a functional of second order for a map $u : \Omega \subset \mathbb{R}^q \to \mathbb{R}^d$:

$$E(u) = \int_\Omega H(\nabla^2 u(x))\,\mathrm{d}x,$$

being H a convex function over $d \times q \times q$ tensors. *What is the largest convex, lower semi-continuous and* subadditive *functional which lifts E in this case?* That would correspond to the "Eulerian" lifting for functionals of second order.

References

[1] L. Ambrosio, N. Gigli, and G. Savaré. *Gradient flows: in metric spaces and in the space of probability measures*. Springer, 2008.
[2] Y. Brenier. Extended Monge-Kantorovich Theory. *Lecture Notes in Mathematics*, Springer-Verlag, 2003, pp. 91-122.
[3] G. Dal Maso. *An introduction to Γ-convergence*. Springer, 2012.
[4] H. Lavenant. Harmonic mappings valued in the Wasserstein space. *Journal of Functional Analysis* 277.3 (2019): p. 688-785.
[5] H. Lavenant. Lifting functionals defined on maps to measure-valued maps via optimal transport. *Arxiv Preprint Arxiv:2309.02260* (2023).
[6] S. Lisini. Characterization of absolutely continuous curves in Wasserstein spaces. *Calculus of variations and partial differential equations* 28.1 (2007): 85-120.
[7] G. Savaré and G. E. Sodini. A simple relaxation approach to duality for Optimal Transport problems in completely regular spaces. *Journal of Convex Analysis* 29.1 (2022): 1-12.
[8] T. Vogt, R. Haase, D. Bednarski and J. Lellmann. On the connection between dynamical optimal transport and functional lifting. *arXiv preprint arXiv:2007.02587* (2020).

Applications of optimal transport theory in meteorology and materials science

David Bourne

In this talk I present two very different applications of optimal transport, one in meteorology and one in materials science, united by common tools from semi-discrete optimal transport theory with exotic costs.

1. Application in meteorology

The first half of the talk is about the results in the papers [4], [5], [13]. The semi-geostrophic (SG) equations are a simplified model of atmospheric flows and frontogenesis. There has been a lot of interest in these equations in the optimal transport community ever since the seminal paper of [2] on the 3D incompressible SG equations. In this talk we focus on the 3D *compressible* SG equations in geostrophic coordinates. The main equation is a nonlocal continuity equation of the form

$$\partial_t \alpha_t + \operatorname{div}(\alpha_t v[\alpha_t]) = 0, \tag{1}$$

where $\alpha : [0,T] \to \mathcal{P}(Y)$, $t \mapsto \alpha_t$, is a curve of probability measures, called the *potential vorticity*, and $Y = \mathbb{R}^2 \times (0,\infty)$ is geostrophic space. The physical space, where the fluid lies, is a compact set $X \subset \mathbb{R}^3$ of volume 1. The velocity $v[\alpha_t] : Y \to \mathbb{R}^3$ in the continuity equation is defined via an optimal transport problem with unknown source measure as follows:

$$v[\alpha_t] := J(\mathrm{id} - (T[\alpha_t])^{-1}), \qquad \text{where } J = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

and $T[\alpha_t] : X \to Y$ is an optimal transport map from physical space to geostrophic space. To be precise, $T[\alpha_t]$ is the optimal map for transporting the source measure $\sigma \in \mathcal{P}_{\mathrm{ac}}(X)$ to $\alpha_t \in \mathcal{P}(Y)$ with respect to the 'exotic' cost $c : X \times Y \to \mathbb{R}$ given by

$$c(x,y) = \frac{1}{y_3}\left(\tfrac{1}{2}(x_1 - y_1)^2 + \tfrac{1}{2}(x_2 - y_2)^2 + g x_3\right),$$

where g is the acceleration due to gravity.[1] The source measure σ (corresponding to the density of the fluid) is an unknown of the problem, and it is defined by the following convex optimisation problem:

$$\sigma[\alpha_t] = \arg\min\left\{ \mathcal{T}_c(\rho, \alpha_t) + \int_X \rho^\gamma(x)\,\mathrm{d}x : \rho \in \mathcal{P}_{\mathrm{ac}}(X) \right\},$$

where $\gamma \in (1,2)$ and $\mathcal{T}_c(\rho,\alpha_t)$ is the cost of transporting ρ to α_t with respect to the cost c. The cost c corresponds to the kinetic and gravitational potential energy of the fluid. The terms $\int_X \rho^\gamma\,\mathrm{d}x$ corresponds to the internal energy of the fluid. While the physical meaning of the potential vorticity α_t is perhaps not very apparent, the temperature of the fluid and its geostrophic velocity can be read

[1] The other physical constants have been suppressed throughout.

off from the components of the optimal transport map $T[\alpha_t]$, as described in [10], where the model was introduced.

The main result of [4] is the existence of weak solutions of the compressible SG equations, for the case where the initial data α_0 is a probability measure. This extends the existence result of [10] for the case of L^p initial data.

The proof is based on a particle method, where the PDE (1) is reduced to a system of ODEs by approximating the initial data by a discrete measure. This extends ideas introduced in [5] from the incompressible case to the compressible case. The proof relies on a characterisation of the unknown source measure σ in terms of a dual problem, along with recent regularity results for semi-discrete optimal transport theory from [11] and [14], as well as the observation that our 'exotic' cost satisfies the non-negative cross-curvature condition.

The next step in our research programme is to extend the numerical results for the 2D incompressible SG equations from [13] and [3] to the compressible case.

2. Application in materials science

The second half of the talk is about the results in the papers [6], [7], [8], in collaboration with the industrial partner Tata Steel Research & Development.

The microstructure of metals and foams is often modelled by engineers using generalised Voronoi diagrams. These are used for example as representative volume elements for computational homogenisation. The challenge is to generate realistic geometric models with prescribed statistical properties, such as the distribution of the volumes and shapes of the cells.

Following the influential work of [1], there has been a lot of research in the microstructure modelling community on representing the microstructure of polycrystalline materials, such as steel, using *anisotropic power diagrams*, which are a class of tessellations defined as follows. Let $X \subset \mathbb{R}^d$ be the domain to be tessellated, n be the number of cells in the tessellation, $Y = (y_1, \dots, y_n) \in X^n$, $W = (w_1, \dots, w_n) \in \mathbb{R}^n$, and $\Lambda = (A_1, \dots, A_n) \in (\mathbb{R}^{d\times d})^n$ be symmetric positive definite matrices, which represent the anisotropy of the cells. The *anisotropic power diagram* (APD) or *anisotropic Laguerre tessellation* generated by the triples $\{(y_i, w_i, A_i)\}_{i=1}^n$ is the partition $\{L_i\}_{i=1}^n$ of X defined by

$$L_i(Y, W, \Lambda) = \left\{ x \in X : \|x - y_i\|_{A_i}^2 - w_i \le \|x - y_j\|_{A_j}^2 - w_j \ \forall\, j \right\},$$

where $\|\cdot\|_{A_i}$ denotes the A_i-norm, $\|x\|_{A_i} = (x^T A_i x)^{1/2}$ for all $x \in \mathbb{R}^d$.

First we consider the *isotropic* case $A_i = I$ for all i. In the computational geometry and optimal transport literature it has been known for a long time how to generate Laguerre tessellations with cells of given volumes, with the state-of-the-art algorithm being the damped Newton method of [14]. In [6] and [7] we transferred this knowledge to the microstructure modelling community, where we applied it to generate synthetic polycrystalline microstructures.

More recently, in [8], we consider the general case of *anisotropic* APDs, where we develop a fast algorithm for generating APDs with cells of prescribed volumes. Our approach uses semi-discrete optimal transport theory with 'exotic' anisotropic

cost $c : \mathbb{R}^d \times Y \to \mathbb{R}$ given by $c(x, y_i) = \|x - y_i\|^2_{A_i}$. The challenge here is mainly one of implementation, namely, how to efficiently compute c-transforms, i.e., how to compute APDs. In [8] we overcome this difficulty with a fast GPU implementation using the KeOps library [9]. We present runtime tests and examples of how to generate synthetic microstructures and fit APDs to EBSD data.

Going forward, we hope to improve the speed of the algorithm further using adaptive meshing to compute the APDs, using ideas from the optimal transport community introduced in [12].

References

[1] A. Alpers, A. Brieden, P. Gritzmann, A. Lyckegaard & H.F. Poulsen, *Generalized balanced power diagrams for 3D representations of polycrystals*, Philosophical Magazine **95** (2015), 1016-1028.

[2] J.D. Benamou & Y. Brenier, *Weak existence for the semigeostrophic equations formulated as a coupled Monge-Ampère/transport P problem*, SIAM Journal on Applied Mathematics **58** (1998), 1450–1461.

[3] J.D. Benamou, C.J. Cotter & H. Malamut, *Entropic optimal transport solutions of the semigeostrophic equations*, Journal of Computational Physics **500** (2024), 112745.

[4] D.P. Bourne, C.P. Egan, T. Lavier & B. Pelloni, *Semi-Discrete optimal transport techniques for the compressible semi-geostrophic equations*, in preparation.

[5] D.P. Bourne, C.P. Egan, B. Pelloni & M. Wilkinson, *Semi-discrete optimal transport methods for the semi-geostrophic equations*, Calculus of Variations and Partial Differential Equations **61** (2022), 39.

[6] D.P. Bourne, P.J.J. Kok, S.M. Roper & W.D.T. Spanjer, *Laguerre tessellations and polycrystalline microstructures: a fast algorithm for generating grains of given volumes*, Philosophical Magazine **100** (2020), 2677–2707.

[7] D.P. Bourne, M. Pearce & S.M. Roper, *Geometric modelling of polycrystalline materials: Laguerre tessellations and periodic semi-discrete optimal transport*, Mechanics Research Communications **127** (2023), 104023.

[8] M. Buze, J. Feydy, S.M. Roper, K. Sedighiani & D.P. Bourne *Anisotropic power diagrams for polycrystal modelling: Efficient generation of curved grains via optimal transport*, in preparation.

[9] B. Charlier, J. Feydy, J.A. Glaunès, F.-D. Collin & G. Durif, *Kernel Operations on the GPU, with Autodiff, without Memory Overflows*, Journal of Machine Learning Research **22** (2021), 1–6.

[10] M.J.P. Cullen & H. Maroofi, *The fully compressible semi-geostrophic system from meteorology*, Archive for Rational Mechanics and Analysis **167** (2003), 309–336.

[11] F. de Gournay, J. Kahn & L. Lebrat, *Differentiation and regularity of semi-discrete optimal transport with respect to the parameters of the discrete measure*, Numerische Mathematik **141** (2019), 429–453.

[12] L. Dieci & J.D. Walsh III, *The boundary method for semi-discrete optimal transport partitions and Wasserstein distance computation*, Journal of Computational and Applied Mathematics **353** (2019), 318–344.

[13] C.P. Egan, D.P. Bourne, C.J. Cotter, M.J.P. Cullen, B. Pelloni, S.M. Roper & M. Wilkinson, *A new implementation of the geometric method for solving the Eady slice equations*, Journal of Computational Physics **469** (2022), 111542.

[14] J. Kitagawa, Q. Mérigot & B. Thibert, *Convergence of a Newton algorithm for semi-discrete optimal transport*, Journal of the European Mathematical Society **21** (2019), 2603–2651.

Solutions of the semigeostrophic equation on a rotating sphere

Lauro Silini

The study of the formation of transported interfaces (frontogenesis) is an important area of research in meteorology, and Hoskin's semigeostrophic equation is believed to be one of the most effective models. It consists in considering transported in the inviscid Navier-Stokes equation only the wind that balances the Coriolis term:

$$\begin{cases} (\partial_t + u_t \cdot \nabla) u_t^G + k(u - u_t^G)^\perp = 0, \\ \operatorname{div}(u) = 0. \end{cases} \tag{1}$$

Here u_t, p_t, k, and u_t^G are the velocity of the incompressible fluid, the pressure, the Coriolis force, and $u_t^G := k^{-1}\nabla^\perp p_t$. The most fascinating feature of this equation was first sensed by Hoskins, and further enriched by Cullen: in the periodic case with $k \equiv$ constant, the change of variables $y = \nabla p_t + x$ leads to a dual reformulation

$$\begin{cases} \partial_t \rho_t + \operatorname{div}\big(\rho_t U_t\big) = 0, \\ U_t = \big(Id - \nabla P_t^*\big)^\perp, \\ \det(D^2 P_t^*) = \rho_t. \end{cases} \tag{2}$$

that appears as a fully non-linear version of the Euler vorticity equation

$$\begin{cases} \partial_t \omega_t + \operatorname{div}\big(\omega_t v_t\big) = 0, \\ v_t = -\nabla \psi, \\ \Delta \psi = \omega. \end{cases} \tag{3}$$

where the Laplace is replaced with a Monge-Ampère equation. Optimal transport theory allowed the construction of a solution for the dual system, see Benamou and Brenier. The challenging task of converting this solution in the original Eulerian coordinates was achieved by Ambrosio, Colombo, De Philippis, and Figalli who established the global-in-time existence and uniqueness of weak solutions for (1) in the flat case with constant k. This was made possible by the regularity theory for bounded Monge-Ampère equations, as developed by De Philippis and Figalli, and further improved by Savin and the latter in the following sharp form

$$0 < \lambda \leq \det(D^2 Q) \leq \Lambda \Rightarrow D^2 Q \in L^{1+\varepsilon}, \quad \varepsilon = \varepsilon(n, \lambda, \Lambda, n) > 0.$$

Local-in-time existence and uniqueness of smooth solutions was then solved in the same framework by Loeper with sharp higher order regularity estimates on the Monge-Amère equation. The same result with varying and smooth enough $k > 0$ was proven later by Cheng, Cullen, and Feldman, via an approximation argument in Lagrangian coordinates.

Semigeostrophic equations are a very fascinating and not yet completely understood subject. In this talk we address the absence of dual reformulations within a physically meaningful context. Specifically, we investigate the well-posedness of (1) on curved domains. The main result presented is the local-in-time existence and uniqueness of solutions in subdomains of a rotating sphere:

Theorem (S. 2022)**.** *Let Ω be an open, smooth, and simply connected subset of the sphere S^2 such that $\bar{\Omega}$ is contained either in the upper or in the lower open hemisphere. Let $\nabla p_0 \in H^s(\Omega, \mathbb{R}^2)$, $s \geq 4$, and suppose that there exists $\mu_0 < 1$ such that the uniform ellipticity condition*

$$\mathcal{Q}_0 := Id + D^2 p_0 - \nabla p_0 \otimes \nabla \ln(k) - \nabla \ln(k) \otimes \nabla p_0 \geq (1 - \mu_0) Id > 0, \tag{4}$$

is satisfied in Ω. Then, there exists $t^* > 0$ such that for all $0 < t' < t^*$ there exists a unique pair

$$\nabla p_t \in C^1(0, t'; C^{s-3,\alpha}(\Omega, \mathbb{R}^2)) \cap L^\infty(0, t'; H^s(\Omega, \mathbb{R}^2)),$$

and

$$u_t = -\nabla^\perp \psi_t \in C(0, t'; C^{s-2,\alpha}(\Omega, \mathbb{R}^2)) \cap L^\infty(0, t'; H^s(\Omega, \mathbb{R}^2)),$$

solving Equation (1) *in $[0, t'] \times \Omega$ with $\nabla p_t|_{t=0} = \nabla p_0$, and u_t tangent to $\partial\Omega$.*

The proof is robust and overcomes the absence of a dual reformulation on the sphere, holding true in general bounded and conformally flat domains with nowhere vanishing and possibly varying k.

We finally present some questions and open problems that naturally arise from the previous part of the talk. In particular: *Is it possible to extend Theorem to include the whole sphere as domain, and not only a closed subdomain strictly contained in an hemisphere?* This problem is interesting from a physical standpoint because it involves the well-posedness of the semigeostrophic equation in a more accurate context, which, to my knowledge, remains an open question even for meteorologists. The major difficulty relies on the singular nature of the geostrophic wind $u_t^G = k^{-1}\nabla^\perp p_t$: being the Coriolis force k proportional to the latitude, it is not clear if the equation makes sense at all in proximity of the equator. A possible toy model could be the upper half plane $\{x^2 > 0\}$ with degenerate weight $k = \frac{1}{x^2}$.

For what concerns the existence of weak solutions in the curved setting, one of the major difficulty relies in the absence of a clear dual reformulation that allows the application of optimal transport theory. In the flat periodic setting the map

$$y = \nabla P_t(x) = x + \nabla p_t,$$

can be seen as an optimal transport map if P_t is convex, in virtue of Brenier Theorem. A direct computation shows that the transported density $\rho_t := (\nabla P_t)_\# dx$ solves the transport equation

$$\partial_t \rho_t + \operatorname{div}(\rho_t (\nabla P_t^* - y)^\perp) = 0.$$

On a sphere the direct parallelism with the Riemannian generalization given by McCann implies that the map should be in the form

$$y = \exp_x(\nabla p_t(x)),$$

with p_t c-convex, with respect to the cost function $c(x, y) = d^2(x, y)/2$. When deducing the associates differential equations for transported density

$$\rho_t := (\exp_x(\nabla p_t(x)))_\# d \operatorname{vol}_{S^2}$$

even in the simple case $k \equiv 1$ however, the curvature of the underlying space enters into play, generating in the orthogonal direction of the pressure gradient additional terms in the forms of Jacobi fields

$$\partial_t \rho_t + \mathrm{div}(\rho_t \nabla^\perp p^c) = E(u_t, \nabla p_t, \kappa),$$

where $E(u_t, \nabla p_t, \kappa)$ denotes the extra term, and $\kappa > 0$ the curvature of S^2. Two natural and interesting questions are :
Formally one can check that $E \to 0$ as $\kappa \to 0$ (at least point-wise). Is it possible to prove stability properties of the solutions in terms of $\kappa \ll 1$?
Is there another cost function $c = c(x, y)$ on S^2 for which the transport map c-$\exp_x(\nabla p_t(x))$ gives a continuity equation for the associated density that does not involve u_t?
Because u_t and ∇p_t are much smaller in size compared to the Earth's radius (around 10 meters/second versus approximately 6 millions meters), the fact that $E(u_t, \nabla p_t, 1/R)$ is also small is a fundamental aspect of the problem. Solving the dual problem on the sphere when we set E to zero is a significant and challenging first step in this direction.

Quantum mechanics, density functional theory, optimal transport, and the curse of dimension

Gero Friesecke

Abstract. I survey the state of the art regarding the application of optimal transport to density functional theory, covering both analysis and numerics.

The area of (electronic) density functional theory, in which optimal transport ideas have been very fruitful, is a re-formulation of quantum mechanics (for electrons). I begin by describing, *formally*, how one gets from quantum mechanics to OT.

Quantum mechanics. The fundamental governing equation for electrons, called electronic Schroedinger equation, was first formulated not by Schroedinger (who only covered the case of 1 electron) but by Dirac in 1929. This equation, "in principle", predicts the chemically specific behaviour of molecules, such as emission/adsorption spectra, binding energies, equilibrium geometries, interatomic forces. In the Born-Oppenheimer approximation, for stationary states, and N electrons, one needs to solve

$$\Big(-\tfrac{1}{2}\Delta + \sum_{1\le i<j\le N} \frac{1}{|r_i - r_j|} + \sum_{i=1}^{N} \mathrm{v}(r_i)\Big)\Psi = E\Psi$$

for $\Psi \in L^2\big((\mathbb{R}^3 \times \mathbb{Z}_2)^N; \mathbb{C}\big)$ subject to the constraints of antisymmetry and normalization,

$$\Psi(z_1, .., z_i, .., z_j, .., z_N) = -\Psi(z_1, .., z_j, .., z_i, ..z_N), \quad ||\Psi||_{L^2} = 1.$$

Here $\mathrm{v} : \mathbb{R}^3 \to \mathbb{R}$ is the external potential exerted by the molecule's atomic nuclei onto the electrons, $v(r) = -\sum_{\alpha=1}^{M} \frac{Z_\alpha}{|r - R_\alpha|}$.

The N-point probability density of positions is

$$\gamma_\Psi(r_1,...,r_N) = \sum_{s_1,...,s_N\in\mathbb{Z}_2} |\Psi(r_1,s_1,...,r_N,s_N)|^2 \quad \text{(Born formula)}.$$

A key collective variable is the electron density

$$\rho = \sum_{i=1}^N \pi_{i\sharp}\,\gamma_\Psi \ (= N\pi_{1\sharp}\,\gamma_\Psi = ... = N\pi_{N\sharp}\,\gamma_\Psi).$$

Curse of dimension → DFT. The problem with the N-electron Schroedinger equation is that it suffers from the curse of dimension. Discretizing each coordinate of $\mathbb{R}^{3N}$ by 10 gridpoints gives, for a single CO_2 molecule ($N = 22$), a totally unfeasible 10^{66} gridpoint values. DFT overcomes this by approximating the Schroedinger equation by system of equations – or equivalently a variational principle – based on the single-particle density. Under the above discretization, one is left with a mere 10^3 gridpoint values, independently of N.

An exact re-formulation goes as follows. The electronic Schroedinger equation is the optimality condition of the variational principle

$$\text{stationarize } T[\Psi] + V_{ee}[\Psi] + V_{ext}[\Psi] \ \text{ s/to } \ ||\Psi||_{L^2}^2 = 1, \ \Psi \text{ antisymm.}$$

Here $T[\Psi] = \int_{\mathbb{R}^{3N}} \sum_{s\in(\mathbb{Z}_2)^N} \sum_{i=1}^N |\nabla_{r_i}\Psi|^2$ (kinetic energy),

$V_{ee}[\Psi] = \int_{\mathbb{R}^{3N}} \sum_{s\in(\mathbb{Z}_2)^N} \sum_{1\le i<j\le N} \frac{1}{|r_i-r_j|}|\Psi|^2$ (interaction energy),

$V_{ext}[\Psi] = \int_{\mathbb{R}^{3N}} \sum_{s\in(\mathbb{Z}_2)^N} \sum_{i=1}^N \mathrm{v}(r_i)|\Psi|^2$ (external potential energy).

The physicist M.Levy introduced the following constrained-search formulation of the above variational priniple:

$$\begin{aligned} E_0 &= \min_{\Psi\,:\,||\Psi||_{L^2}^2=1} \Big(T[\Psi] + V_{ee}[\Psi] + V_{ext}[\Psi]\Big) \\ &= \min_\rho \ \min_{\Psi\,:\,\pi_{i\sharp}\gamma_\Psi=\rho\,\forall i} \Big(T[\Psi] + V_{ee}[\Psi] + V_{ext}[\Psi]\Big) \\ &= \min_\rho \ \Big(\underbrace{\min_{\Psi\,:\,\pi_{i\sharp}\gamma_\Psi=\rho\,\forall i} \big(T[\Psi] + V_{ee}[\Psi]\big)}_{=:F_{LL}[\rho] \text{ universal part}} + \underbrace{\int_{\mathbb{R}^3} \mathrm{v}(r)\rho(r)dr}_{\text{chemically specific part}} \Big) \end{aligned}$$

The outer min is just over densities. But the universal functional still involves a minimization over the high-dimensional space of wavefunctions, and must be approximated to obtain practical methods.

Density scaling → OT. For any given density, consider its dilation $\rho_\lambda(r) = \lambda^3\rho(\lambda r)$, $\lambda > 0$. An elementary calculation shows that

$$F_{LL}[\rho_\lambda] = \min_{\pi_{i\sharp}\gamma_\Psi=\rho/N\,\forall i} \Big(\lambda^2 T[\Psi] + \lambda V_{ee}[\Psi]\Big).$$

It follows that at high density ($\lambda \gg 1$) kinetic enery dominates,

$$F_{LL}[\rho_\lambda] \;\approx\; \lambda^2 \min_{\pi_{i\sharp}\gamma_\Psi = \rho/N \;\forall i} T[\Psi],$$

whereas at low density ($\lambda \ll 1$) potential energy dominates,

$$F_{LL}[\rho_\lambda] \;\approx\; \lambda \min_{\pi_{i\sharp}\gamma_\Psi = \rho/N \;\forall i} V_{ee}[\Psi].$$

Since $V_{ee}[\Psi] = \int_{\mathbb{R}^{3N}} \sum_{1\le i<j\le N} \frac{1}{|r_i - r_j|}\, d\gamma_\Psi(r_1, ..., r_N)$ this is OT (with Coulomb cost). The minimum on the right, i.e. the optimal cost as a functional of the marginal density, is called the SCE functional $V_{ee}^{SCE}[\rho]$. It is the Coulomb analogue of the Wasserstein distance. The above considerations mean that when ρ solves the constrained-search problem with F_{LL} replaced by V_{SCE}, then the optimal plan γ for OT with Coulomb cost approximates the quantum wavefunction squared, and this approxiation is (formally) asymptotically correct in the low density limit.

Analytical work. A great deal of analytical work has been done on this problem. Here is an incomplete list of important contributions (for a much longer list and precise references see the review [1]).

* Seidl '99: introduces model (in the physics lit.); found the exact solution in 1D
* Cotar, F, Klueppelberg 2011-13; Buttazzo, Gori-Giorgi, de Pascale 2012: rigorous formulation, interpretation as OT
* Colombo, DiMarino, DePascale 2015: proof of exact sol'n in 1D
* Cotar, F, Klueppelberg 2018: rigorous justif. of low-density limit as Gamma conv.
* Colombo, DiMarino, Stra 2022: rigorous results on next-order correction.

Numerical work. Numerically, OT with Coulomb cost for N particles still suffers from the curse of dimension, just like the Schroedinger equation it approximates. Many authors have tackled the problem with different approaches (including Chen, F. and Mendl 2014, Benamou, Carlier and Nenna 2015, Khoo and Ying 2019, Alfonsi, Coyaud and Ehrlacher 2021, Nenna and Pass 2022 (see the review [1]). A very efficient approach delivering highly accurate solutions (the latter being important in electronic structure) was recently developed by F, Schulz and Voegler [4] and F and Penka [2]. This approach rigorously overcomes the curse of dimension with a sparse ansatz similar to, but different from, the Monge ansatz. More precisely one has the following theorem on discretized multi-marginal OT

$$\min \langle c, \gamma \rangle \text{ over } \gamma \,:\, X_1 \times ... \times X_N \to \mathbb{R} \;\; \text{s/to} \;\; \gamma \ge 0, \;\; \pi_{i\sharp}\gamma = \mu_i.$$

Theorem [4, 2]**.** If $|X_k| = \ell_k$, for any cost the discrete multi-marginal OT problem has an optimizer with

$$|\mathrm{supp}\,\gamma| \le \sum_{\mathrm{k}=1}^{\mathrm{N}} \ell_{\mathrm{k}}$$

(instead of $|\mathrm{supp}\,\gamma| \le \sum_{\mathrm{k}=1}^{\mathrm{N}} \ell_{\mathrm{k}}$ as needed for general plans).

This breaks the curse of dimension in terms of the number of nonzero parameters needed, as this number grows only linearly instead of exponentially with the number N of particles/marginals.

An algorithm – termed *Genetic column generation* – which finds the optimal plan was introduced in [4] for the (symmetric) Coulomb problem and extended in [2] to general (non-symmetric) multi-marginal problems. Numerically, on the benchmark problem of the homogeneous electron gas in one dimension where the exact plan is known, the algorithm has been demonstrated to find the exact ground state extremely efficiently, in polynomial computational time with respect to the particle number N [4]. The idea of the algorithm is as follows: Instead of full discrete OT, which has $\ell_1 \cdot ... \cdot \ell_N$ DOF's, iteratively solve OT on a small subset,

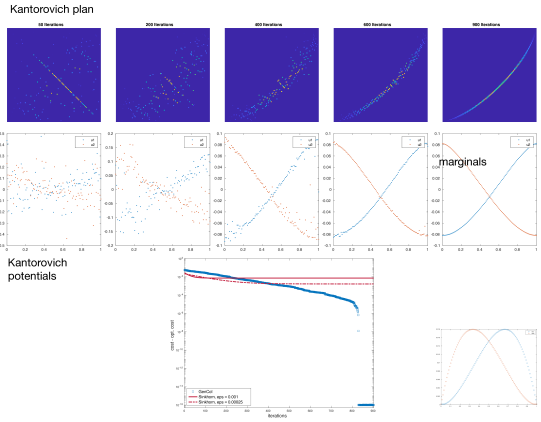
$$\min \langle c, \gamma \rangle_\Omega \text{ over } \gamma : \Omega \subset X_1 \times ... \times X_N \to \mathbb{R} \text{ s/to } \gamma \geq 0, \ \pi_{i\sharp}\gamma = \mu_i,$$

with

$$|\Omega| \leq \beta \, (\ell_1 + ... + \ell_N),$$

and update the subset. Here β is a hyperparameter (taken to be 3 in our simulations).

The following didactic example, taken from [2], shows how GenCol solves classical OT with quadratic cost, exhibiting exponential convergence to the unique optimal plan.



The algorithm builds on classical column generation going back to Dantzig and Wolfe. But there one never discards any configuration, so that the domain grows with each iteration, and one uses a deterministic search rule ("pricing problem") which is known to be NP-hard.

The updating step of Ω is partially inspired by ML protocols in adversarial learning, with the correspondence

proposal step	$\sim$	SGD step for generator
acceptance step	$\sim$	learning from a critic (the current dual, as in WGANs).

In particular, the algorithm contains initial stochasticity through the initally poor critic, which is automatically tuned down (as in GANs, and unlike SGD).

For $N = 2$ one has the following global convergence result; the main proof idea is due to my PhD student Maximilian Penka.

Theorem [3]**.** Let X, Y discrete with $|X| = \ell_1$, $|Y| = \ell_2$. For any cost, any marginals, any $\beta \geq 2$, and any feasible initial subset $\Omega \subset X \times Y$ with $|\Omega| \leq \beta(\ell_1 + \ell_2)$, GenCol converges with probability 1 to a global minimizer.

This result is quite remarkable because GenCol maintains the nonconvex constraint

$$||\gamma||_{\ell^0} \leq \beta(\ell_1 + \ell_2)$$

yet does not get stuck in a local minimizer subject to the local update $\Omega \rightsquigarrow \Omega \cup \{r'\}$. The proof can be generalized to N marginals when children are searched for in the (in practice too large) search space differing from a parent in $N-1$ components, but breaks down when children are searched for in the (in practice efficient) search space differing from a parent in only 1 component.

An application. OT with Coulomb cost is only an accurate DFT model in the low-density limit. This regime is of interest for certain physical systems (electron gas; nanowires) but not for molecules. To build realistic DFT models for real molecules with the help of OT, one needs to interpolate between the high and low density limits. Recent advances, providing e.g. a remarably accurate binding energy curve for the pyridine-pyridine dimer $(C_5NH_5)_2$ (which contains 84 electrons), are discussed in [5].

References

[1] G. Friesecke, A. Gerolin, P. Gori-Giorgi, *The strong-interaction limit of density functional theory*, Chapter in the book 'Density functional theory', E. Cancès and G. Friesecke, Springer (2023), http://arxiv.org/abs/2202.09760 (2022)

[2] G. Friesecke and M. Penka, *The GenCol algorithm for high-dimensional optimal transport: general formulation and application to barycenters and Wasserstein splines*, SIAM J. Math. Data Sci. 5 (4) (2023), 899-919

[3] G. Friesecke and M. Penka, *Convergence proof for the GenCol algorithm in the case of two-marginal optimal transport*, arxiv:2303.07137 (2023)

[4] G. Friesecke, A. S. Schulz, D. Vögler, *Genetic column generation: Fast computation of high-dimensional multi-marginal optimal transport problems*, SIAM J. Sci. Comp. 44 No. 3 (2022), A1632-A1654

[5] S. Vuckovic, A. Gerolin, T.J. Daas, H. Bahmann, G. Friesecke, P. Gori-Giorgi, *Density functionals based on the mathematical structure of the strong-interaction limit of DFT*, Wiley Interdisciplinary Reviews: Computational Molecular Science Vol. 13 (2), 2022

Optimal transport bounds on pattern-forming energies with divergence constraint

Benedikt Wirth

So-called energy scaling laws represent a basic approach to understand the patterns forming during minimization of singularly perturbed energies. If the energies involve a divergence constraint, then a simple and elegant argument for the derivation of the lower bounds is sometimes possible using convex duality arguments for

optimal transportation. Below we exemplify this for the setting of compliance minimization under a uniaxial load (we will stay at the level of an informal description, but it is not much work to make everything rigorous).

Compliance minimization is the task to optimize the shape of an elastic body $\mathcal{O} \subset \Omega$ within some fixed domain of interest $\Omega \subset \mathbb{R}^d$, $d = 2, 3$, such that it withstands well a given applied mechanical load (a surface stress) $f : \partial\Omega \to \mathbb{R}^d$ and at the same time uses little material. A simple measure for the mechanical weakness of $\mathcal{O}$ under the load f is given by the compliance, the total elastic energy stored within $\mathcal{O}$ as soon as the load is applied. Assuming for simplicity the elasticity tensor of the employed material to be the identity, the compliance is given by the squared L^2-norm of the equilibrium stress field, which in turn is the minimizer of that squared L^2-norm among all admissible stress fields σ,

$$\mathrm{comp}(\mathcal{O}) = \min_{\sigma \text{ admissible}} \int_{\mathcal{O}} |\sigma|^2 \, \mathrm{d}x.$$

Admissible stress fields are those that conserve angular and linear momentum and are consistent with the boundary loads, thus σ is admissible if and only if

$$\sigma \in \Sigma = \{\sigma \in L^2(\Omega; \mathbb{R}^{d\times d}_{\mathrm{sym}}) \,|\, \mathrm{div}\sigma = 0 \text{ in } \Omega, \sigma n = f \text{ on } \partial\Omega, \sigma = 0 \text{ on } \Omega \setminus \mathcal{O}\},$$

where n is the unit outward normal on $\partial\Omega$. One now aims to minimize the energy

$$\mathcal{J}(\mathcal{O}) = \mathrm{comp}(\mathcal{O}) + \mathrm{vol}(\mathcal{O}).$$

Unfortunately, this energy is not lower semicontinuous so that minimizers do not exist. Instead, along a minimizing sequence, microstructure will form. As a remedy one can add the perimeter of the shape as regularization with a positive weight ε,

$$\mathcal{J}^\varepsilon(\mathcal{O}) = \mathcal{J}(\mathcal{O}) + \varepsilon \mathrm{per}(\mathcal{O}),$$

which then can be shown to have minimizers.

For small ε (modelling for instance bone), very fine, complex structures will be optimal. In this regime one may try to obtain some understanding of the forming structures by proving an energy scaling law, an estimate of the form

$$ch(\varepsilon) \le \Delta\mathcal{J}(\mathcal{O}_{\mathrm{opt}}) = \mathcal{J}^\varepsilon(\mathcal{O}_{\mathrm{opt}}) - \inf \mathcal{J} \le Ch(\varepsilon) \tag{1}$$

for the optimal shape $\mathcal{O}_{\mathrm{opt}}$, two constants $C \ge c > 0$, and some function h. $\Delta\mathcal{J}(\mathcal{O}_{\mathrm{opt}})$ represents the excess cost paid for introducing the regularization. The relevance of (1) is: If one finds a shape $\mathcal{O}$ with $\Delta\mathcal{J}(\mathcal{O}) \le Ch(\varepsilon)$ (which is how one proves the upper bound), then this shape already has optimal cost $\Delta\mathcal{J}$ up to a constant factor, since the lower bound rules out any stronger improvement. Thus, the found shapes give an indication of how near-optimal shapes may look like.

The analysis of energy scaling laws really got started in the 90s with work by Müller, Kohn and others on martensite patterns, but meanwhile many different patterns have been analysed this way. For the example of compliance minimization under a uniaxial load $f = Fe_3 \otimes e_3 n$ on $\Omega = [0, \ell)^2 \times [0, 1]$ with $F \in \mathbb{R}$, $e_3 = (0, 0, 1)^T$, and n the unit outward normal, one obtains the following.

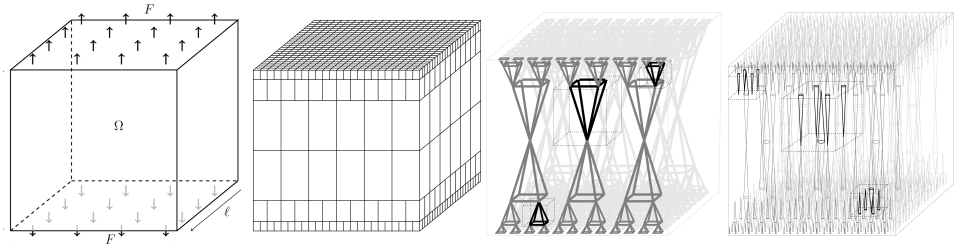


FIGURE 1. Left to right: Setting, subdivision of Ω into elementary cells, construction sketch for $F \le \frac{1}{2}$, sketch for $F > \frac{1}{2}$.

Theorem ([1]). *Under some conditions on ℓ, $ch(\varepsilon,F) \le \Delta\mathcal{J}(\mathcal{O}_{opt}) \le Ch(\varepsilon,F)$*

$$\textit{for } h(\varepsilon,F) = \begin{cases} \varepsilon & \textit{if } F \le \varepsilon^{\frac{1}{2}}, \\ F^{\frac{2}{3}}\varepsilon^{\frac{2}{3}} & \textit{if } \varepsilon^{\frac{1}{2}} < F \le \frac{1}{2}, \\ (1-F)|\log(1-F)|^{\frac{1}{3}}\varepsilon^{\frac{2}{3}} & \textit{if } \frac{1}{2} < F \ \&\ \varepsilon^{\frac{2}{3}} \le (1-F)|\log(1-F)|^{-\frac{1}{3}}, \\ (1-F)^2 & \textit{if } (1-F)\,|\log(1-F)|^{-\frac{1}{3}} < \varepsilon^{\frac{2}{3}}. \end{cases}$$

In addition to ε there is another small parameter, F or $(1-F)$, so that the energy scaling exhibits different regimes, depending on the relative size of these parameters. (Note: For $F \ge 1$ even $\mathcal{J}$ has a minimizer, $\mathcal{O} = \Omega$.) The constructed $\mathcal{O}$'s to prove the upper bound are illustrated in Fig. 1 (except for the last regime for which $\mathcal{O}_{\rm opt} = \Omega$): Near the boundary the structure needs to be fine to support the load, but towards the center the structure coarsens to save perimeter. This is done by stacking levels consisting of many copies of an elementary cell, where the elementary cell width halves from level to level (the aspect ratio changes as well). For $F \le \frac{1}{2}$ the elementary cell contains thin struts along the edges of a pyramid, while for $F > \frac{1}{2}$ it is a material block with a (not quite) conical hole at the centre and four holes of similar shape but half width arranged around the centre.

Before sketching the lower bound proof via optimal transport, let us mention that, using exactly the same type of constructions for the upper and techniques for the lower bound, also other patterns can be analysed, for instance the ones of the intermediate state in type-I superconductors [2] (the underlying singularly perturbed energy is closely related to $\mathcal{J}^\varepsilon$, and [1] actually follows the analysis of [2]) or the ones found in so-called branched transport or urban planning [3].

We just sketch the proof of the lower bound for the second regime, setting $\ell = 1$ for notational simplicity. We argue by contradiction, so assume $\Delta\mathcal{J}(\mathcal{O}_{\rm opt}) \ll (\varepsilon F)^{2/3}$ and abbreviate $\mathcal{O}_t$ to be the cross-section of $\mathcal{O}_{\rm opt}$ at height t.

Step 1. We bound volume/perimeter of a typical cross-section (using $\inf \mathcal{J} = 2F$):

$$\Delta\mathcal{J}(\mathcal{O}_{\rm opt}) \ge \int_0^1 \underbrace{\int_{\mathcal{O}_t} \sigma_{33}^2 \,\mathrm{d}x + \mathrm{vol}\mathcal{O}_t - 2F}_{=A} + \underbrace{\varepsilon \mathrm{per}\mathcal{O}_t}_{=B} \,\mathrm{d}t.$$

If $A+B$ were bigger than $(\varepsilon F)^{2/3}$ for, say, half the cross-sections, our assumption on $\Delta\mathcal{J}(\mathcal{O}_{\rm opt})$ would be violated, so A and B are both small for at least half the cross-sections. Now notice that due to force balance, the total vertical force in cross-section t must equal the total force applied at the bottom, thus $\int_{\mathcal{O}_t}\sigma_{33}\,\mathrm{d}x = F$. This then leads to

$$(\varepsilon F)^{2/3} \gtrsim A \geq \frac{1}{\mathrm{vol}\mathcal{O}_t}\left(\int_{\mathcal{O}_t}\sigma_{33}\,\mathrm{d}x\right)^2 + \mathrm{vol}\mathcal{O}_t - 2F = \frac{F^2}{\mathrm{vol}\mathcal{O}_t} + \mathrm{vol}\mathcal{O}_t - 2F,$$

which implies $\mathrm{vol}\mathcal{O}_t \approx F$ up to a small error, as well as

$$(\varepsilon F)^{2/3} \gtrsim A = \int_{\mathcal{O}_t}(\sigma_{33}-1)^2\mathrm{d}x,$$

which implies $\sigma_{33} \approx 1_{\mathcal{O}_t}$, the characteristic function of $\mathcal{O}_t$, up to a small error.

Step 2. We now know $\mathrm{vol}\mathcal{O}_t$ and have an upper bound on $\mathrm{per}\mathcal{O}_t$. This implies that the typical width of $\mathcal{O}_t$ is no smaller than $\mathrm{vol}\mathcal{O}_t/\mathrm{per}\mathcal{O}_t \gtrsim (\varepsilon F)^{1/3} =: w$. In other words, $\mathcal{O}_t$ must contain $N \sim \mathrm{vol}\mathcal{O}_t/w^2$ balls of diameter comparable to w.

Step 3. Abbreviating $\omega = (\sigma_{31},\sigma_{32})$, we finally find

$$\Delta\mathcal{J}(\mathcal{O}_{\rm opt}) \geq \int_0^t\int_{\mathcal{O}_s}|\omega|^2\,\mathrm{d}x + \underbrace{\int_{\mathcal{O}_s}\sigma_{33}^2\mathrm{d}x + \mathrm{vol}\mathcal{O}_s - 2F}_{\geq 0}\,\mathrm{d}s \geq \frac{1}{\int_0^t \mathrm{vol}\mathcal{O}_s\,\mathrm{d}s}\left(\int_0^t\int_{\mathcal{O}_s}|\omega|\,\mathrm{d}x\,\mathrm{d}s\right)^2.$$

The fraction scales like $1/F$, and the term in parenthesis equals the Benamou–Brenier formulation of the Wasserstein-1 distance $W_1(\sigma_{33}|_{s=0},\sigma_{33}|_{s=t})$ between σ_{33} at height 0 and height t. Indeed, ω is the momentum associated with the transport of σ_{33}, since they satisfy the continuity equation

$$0 = \mathrm{div}\sigma_3 = \partial_t\sigma_{33} + \mathrm{div}_{x_1,x_2}\omega.$$

Now $\sigma_{33}|_{s=0} = F\mathcal{L}$ and $\sigma_{33}|_{s=0} \approx 1_{\mathcal{O}_t}\mathcal{L}$ for the Lebesgue measure $\mathcal{L}$ so that one can easily exploit Kantorovich–Rubinstein duality to find a lower bound on the Wasserstein-1 distance (knowing that $\mathcal{O}_t$ contains N balls of diameter w). This finally yields the desired lower bound on $\Delta\mathcal{J}(\mathcal{O}_{\rm opt})$.

References

[1] J. Potthoff and B. Wirth, *Optimal fine-scale structures in compliance minimization for a uniaxial load in three space dimensions*, ESAIM: Control, Optimisation and Calculus of Variations **28** (2022), 27.

[2] R. Choksi, S. Conti, R.V. Kohn, F. Otto, *Ground state energy scaling laws during the onset and destruction of the intermediate state in a type I superconductor*, Communications on Pure and Applied Mathematics **61**(5) (2008), 595–626.

[3] A. Brancolini and B. Wirth. *Optimal energy scaling for micropatterns in transport networks*, SIAM Journal on Mathematical Analysis **49**(1) (2016), 311–359.

Aggregation-diffusion equation: symmetry, uniqueness and non-uniqueness of steady states

Yao Yao

(joint work with José Carrillo, Matias Delgadino, Sabine Hittmeir, Bruno Volzone, Xukai Yan)

The aggregation-diffusion equation describes the mean-field limit of an interacting particle system driven by local repulsion and pairwise local attraction, and it arises in various models in mathematical biology and physics. The equation reads

$$\partial_t \rho = \Delta \rho^m + \nabla \cdot (\rho \nabla (W * \rho)), \quad x \in \mathbb{R}^d, t \geq 0. \tag{1}$$

See [1, 5] for surveys on this equation regarding the modeling, analysis and numerical aspects. We assume the interaction potential W is be radially symmetric and *attractive*, that is, W is differentiable in $\mathbb{R}^d \setminus \{0\}$ and satisfies $W'(r) > 0$ for all $r > 0$, where r is the radial variable.

When $m = 1$, the linear diffusion term corresponds to the Brownian motion of the particles. One can also change linear diffusion into a degenerate diffusion term $\Delta \rho^m$ with $m > 1$, which models the anti-overcrowding mechanism between individuals. In particular, when $m \geq 1$ and $W = \mathcal{N}$ is the Newtonian potential in $\mathbb{R}^d$, the equation becomes the Patlak-Keller-Segel equation which models collective motion of cells driven by chemotaxis, and it also arises in the study of gravitational collapse models.

In the analysis of (1), the following *free energy functional* plays an important role:

$$\mathcal{E}[\rho] = \frac{1}{m-1} \int_{\mathbb{R}^d} \rho^m(x)\, dx + \frac{1}{2} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \rho(x) W(x-y) \rho(y)\, dxdy, \tag{2}$$

where the first integral on the right hand side is called entropy, and the second integral is called interaction energy. (Note that when $m = 1$, the entropy integral becomes $\int_{\mathbb{R}^d} \rho \log \rho dx$.) A simple computation yields that the $E[\rho]$ is non-increasing along a solution. In fact, the solution $\rho(\cdot, t)$ is formally a gradient flow of $\mathcal{E}[\rho]$ in the metric space endowed by 2-Wasserstein metric, but rigorously justifying this gradient flow requires some convexity assumptions on W.

Using the monotonicity of the free energy functional, the global well-posedness v.s. finite-time blow-up question for (1) has been well understood. However, in the cases where solutions are known to exist globally, the long-time behavior of solutions remain unclear in many cases. And to understand the long time dynamics, a necessary step is to identify all the steady states.

In a joint work [3] with Carrillo, Hittmeir and Volzone, we study the stationary solutions to (1) for all attracting kernels W that is no more singular than Newtonian potential, and prove that every $L^1 \cap L^\infty$ stationary solution must be radially symmetric. This is done by combining Steiner symmetrization techniques with some a priori regularity estimates on stationary solutions. As an application of this symmetry result, we show that it leads to asymptotic convergence towards

(a translation of) the global minimizer as $t \to \infty$ for the 2D Patlak-Keller-Segel equation with degenerate diffusion.

Once the radial symmetry (up to a translation) of all steady states is known, the next natural question is whether there is a unique radial steady state for each given mass. Previously, uniqueness in the diffusion-dominated regime was known for Newtonian and other Riesz potentials, and also for convex potentials due to displacement convexity along the geodesic in 2-Wasserstein metric. But for a generic attractive potential, the only results in the literature were done for the $m = 2$ case, under some additional assumption on the regularity of the potential at the origin.

Together with Delgadino and Yan [4], we show that $m = 2$ is indeed the threshold separating uniqueness and non-uniqueness for a generic attractive potential. Namely, we obtain the following:

- Let $m \geq 2$ and $W \in C^1(\mathbb{R}^d \setminus \{0\})$ be an attractive potential with $W'(r) \leq Cr^{-d-1+\delta}$ for some $\delta > 0$ for all $r \in (0, 1)$. Then there is at most one steady state for (1) (up to a translation) for any given mass.
- Let $1 < m < 2$. There exists a smooth attractive kernel W which gives an infinite sequence of radially decreasing steady states of (1) with the same mass.

The proof of the uniqueness result is based a natural idea: since steady states are (formally) critical points of the free energy functional, suppose we are able to construct a smooth curve conecting them such that the energy along this curve is strictly convex, it immediately yields that there cannot be more than one critical points.

Of course, the main question is how to find an interpolation curve along which the energy is convex, if it exists at all. Note that the most common interpolations such as linear interpolation or Wasserstein geodesics all fail to be convex for a general attractive potential W, as they require some additional convexity properties of W. We introduce a novel interpolation curve between any two radially decreasing functions, such that the interaction energy along this curve is convex for any attractive potential, and the entropy along this curve is convex if and only if $m \geq 2$.

Open questions. Now that the uniqueness/nonuniqueness question has been understood for a generic attractive poential for $m > 1$, the following open question remains:

(1) When $m = 1$, for a given mass, are stationary solutions unique for a generic attractive interaction potential W? Note that they are known to be radially decreasing by [3], but both our uniqueness and non-uniqueness proofs in [4] fail in the $m = 1$ case.
(2) When $m > 2$, as $t \to \infty$, does the solution to (1) converge to the unique stationary solution with the same mass and center of mass as the initial data? Here the difficulty is to show that mass cannot escape to infinity as $t \to \infty$. For large m, some progress has been made in [6].

(3) If W is such that there is no blow-up and no steady state, as $t \to \infty$, does all solutions dissipate with the heat equation (for $m = 1$) / porous medium equation (for $m > 1$) scaling? For the $m = 1$ case, there has been some recent progress in [2] when W is bounded, but this question remains open when W is growing very slowly at infinity.

References

[1] J. Carrillo, K. Craig and Y. Yao, Aggregation-diffusion equations: dynamics, asymptotics, and singular limits, Active Particles, Volume 2, 65-108, Birkhäuser Basel, 2019.

[2] J. Carrillo, D. Gómez-Castro, Y. Yao and C. Zeng, *Asymptotic simplification of Aggregation-Diffusion equations towards the heat kernel* , Arch. Ration Mech. Anal., 247 (2023), 11.

[3] J. A. Carrillo, S. Hittmeir, B. Volzone, and Y. Yao, *Nonlinear aggregation-diffusion equations: radial symmetry and long time asymptotics*, Invent. Math **218** (2019), 889–977.

[4] M. Delgadino, X. Yan and Y. Yao, *Uniqueness and nonuniqueness of steady states of aggregation-diffusion equations*, Comm. Pure Appl. Math., **75** (2022), 3–59.

[5] D. Gómez-Castro, Beginner's guide to Aggregation-Diffusion Equations, preprint, arXiv:2309.13713, 2023.

[6] Ruiwen Shu, *Equilibration of aggregation-diffusion equations with weak interaction forces*, SIAM J. Math. Anal. **53** (2021), 6562–6616.

Extrapolation in the Wasserstein space

Andrea Natale

(joint work with Thomas O. Gallouët, Gabriele Todeschi)

Given a metric space (X, d), a (globally minimizing) geodesic on X defined on $[t_0, t_1] \subset \mathbb{R}$, with $t_1 > t_0$, is a curve $x : s \in [t_0, t_1] \to x(s) \in X$ verifying

$$d(x(s_0), x(s_1)) = \frac{|s_1 - s_0|}{|t_1 - t_0|} d(x(t_0), x(t_1)) \,, \tag{1}$$

for all $s_0, s_1 \in [t_0, t_1]$. Given two points $x_0, x_1 \in X$ and $t > 1$, consider the following minimization problem

$$\inf_{x \in X} \left\{ \frac{d^2(x, x_1)}{2(t-1)} - \frac{d^2(x, x_0)}{2t} \right\}. \tag{2}$$

By triangular and Young's inequality, we always have that such infimum is larger than $-d^2(x_0, x_1)/2$. Moreover, if there exists a geodesic $x : s \in [0, t] \to x(s) \in X$ such that $x(0) = x_0$ and $x(1) = x_1$ then this lower bound is attained by $x(t)$, which is therefore a minimizer. In general, problem (2) gives a variational definition of geodesic extrapolation even if no geodesic connecting x_0 to x_1 on the interval $[0, 1]$ may be extended up to time $t > 1$.

We consider a specific instance of problem (2) where (X, d) is $\mathcal{P}_2(\mathbb{R}^d)$, the set of probability measures with finite second moments, equipped with the L^2-Wasserstein distance W_2. This latter is defined as follows: for any $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$

$$W_2^2(\mu, \nu) = \inf \left\{ \int |x - y|^2 \mathrm{d}\gamma(x, y) \ : \ \gamma \in \Gamma(\mu, \nu) \right\} \tag{3}$$

where $\Gamma(\mu,\nu) \subset \mathcal{P}_2(\mathbb{R}^d \times \mathbb{R}^d)$ is the set of couplings having μ and ν as first and second marginal, respectively. The problem we consider is therefore given by:

$$(\mathcal{P}) \qquad \inf_{\mu \in \mathcal{P}_2(\mathbb{R}^d)} \left\{ \frac{W_2^2(\mu,\nu_1)}{2(t-1)} - \frac{W_2^2(\mu,\nu_0)}{2t} \right\} .$$

By the derivation above, problem $(\mathcal{P})$ provides a natural notion for geodesic extrapolation in the Wasserstein space. It was introduced in [2] for the construction of a second order time discretization of Wasserstein gradient flows. Here we show that this problem admits a convex formulation which can be derived via Toland duality, and that it can be also further reformulated as a weak optimal transport problem.

1. Geodesics in the Wasserstein space

Consider the optimal transport problem (3) from ν_0 to ν_1. If, for example, ν_0 is absolutely continuous then Brenier's theorem states that there exists a unique solution γ^* to this problem which is furthermore induced by a transport map $\nabla u : \mathbb{R}^d \to \mathbb{R}^d$, where $u : \mathbb{R}^d \to \mathbb{R}$ is a convex function usually called the Brenier potential for the transport from ν_0 to ν_1, i.e.

$$\gamma^* = (\nabla u, \mathrm{Id})_{\#}\nu_0 .$$

Furthermore, in this case, there exists a uniquely defined geodesic on the interval $[0,1]$ connecting ν_0 to ν_1, which is given by

$$(4) \qquad \nu(s) = ((1-s)\mathrm{Id} + s\nabla u)_{\#}\nu_0, , \quad \forall\, s \in [0,1] .$$

This curve can be extended up to $s = t > 1$ while staying a length-minimizing geodesic if and only if $((1-t)\mathrm{Id}+t\nabla u)$ is an optimal transport map, or equivalently

$$x \mapsto u(x) - \frac{t-1}{t}\frac{|x|^2}{2} \quad \text{is convex.}$$

However, this condition is not verified in general, since we may only expect u to be convex (and not strongly convex) which corresponds to the fact that the particle trajectories induced by (4) may cross precisely at $s = 1$.

2. Toland duality

Toland's duality concerns the minimization of the difference of two convex, proper and lower semi-continuous function $F, G : V \to (\infty, \infty]$, where V is a normed vector space. Specifically, we have the equivalence

$$\inf_{x \in V} \{F(x) - G(x)\} = \inf_{p \in V^*} \{G^*(p) - F^*(p)\}$$

where V^* is the topological dual of V and F^* and G^* are the Legendre transforms of F and G respectively. The idea of using Toland duality to deal with differences of Wasserstein distances stems from the work of Carlier [1]. In our case, F and G are replaced by the maps

$$\mu \in \mathcal{P}_2(\mathbb{R}^d) \mapsto \frac{W_2^2(\mu,\nu_1)}{2(t-1)} \quad \text{and} \quad \mu \in \mathcal{P}_2(\mathbb{R}^d) \mapsto \frac{W_2^2(\mu,\nu_0)}{2t}$$

Then, at least formally, one can check that the resulting dual problem is given by

$$(\mathcal{P}^*) \qquad \inf\left\{\int u^* \mathrm{d}\nu_1 + \int u \mathrm{d}\nu_0 \ : \ u - \frac{t-1}{t}\frac{|\cdot|^2}{2} \text{ is convex}\right\}.$$

Remarkably, and differently from $(\mathcal{P})$, this is a convex optimization problem in the usual sense. Also, we observe that requiring u to be only convex, one recovers the usual dual formulation of the optimal transport problem from ν_0 to ν_1 and in particular, any Brenier potential u is a solution and, if ν_1 is a.c., $\nabla u^*_{\#}\nu_1 = \nu_0$. In general, however, if u is a solution to $(\mathcal{P}^*)$ one only has that $\overline{\nu}_0 = \nabla u^*_{\#}\nu_1$ is domininated in the convex order by ν_0, i.e.

$$\int \varphi \, \mathrm{d}\overline{\nu}_0 \leq \int \varphi \, \mathrm{d}\nu_0 \tag{5}$$

for all convex function $\varphi : \mathbb{R}^d \to \mathbb{R}$. Moreover, we show that the (unique) solution to problem $(\mathcal{P})$ can be written as

$$\nu_t = ((1-t)\mathrm{Id} + t\nabla u)_{\#}\overline{\nu}_0 \,,$$

at least in the case where $\overline{\nu}_0$ is absolutely continuous. Since u is strongly convex this means that the measure $\overline{\nu}_0$, defined via the solution of problem $(\mathcal{P}^*)$ is such that the geodesic from $\overline{\nu}_0$ to ν_1 on the time interval $[0,1]$ can be extended up to time t and the resulting extension is precisely the solution to $(\mathcal{P})$.

3. Weak optimal transport formulation

The convex order relation between $\overline{\nu}_0 = \nabla u^*_{\#}\nu_1$ and ν_0, with u solving $(\mathcal{P}^*)$, can be exploited to derive an equivalent formulation of problem $(\mathcal{P})$ which fits in the framework of weak (and in particular barycentric) optimal transport, a generalization of optimal transport introduced in [3]. This reads as follows:

$$(\mathcal{B}) \quad \inf_{\gamma \in \Gamma(\nu_0,\nu_1)} \int |tx_1 - (t-1)\mathrm{bary}(\gamma_{x_1})|^2 \mathrm{d}\nu_1(x_1) \,, \quad \mathrm{bary}(\gamma_{x_1}) = \int x_0 \mathrm{d}\gamma_{x_1}(x_0) \,.$$

By Strassen's theorem, condition (5) implies the existence of a coupling $\theta \in \Gamma(\overline{\nu}_0, \nu_0)$ which is the law of a martingale, i.e. $\mathrm{d}\theta(x,y) = \mathrm{d}\theta_x(y)\mathrm{d}\overline{\nu}_0(x)$ and

$$\int y \, \mathrm{d}\theta_x(y) = x \,, \qquad \text{for } \overline{\nu}_0\text{-a.e. } x \in \mathbb{R}^d.$$

The link bewteen problem $(\mathcal{P}^*)$ and $(\mathcal{B})$ is the following: u solves $(\mathcal{P}^*)$ if and only if the coupling $\pi \in \Gamma(\nu_0,\nu_1)$ defined by

$$\mathrm{d}\pi(x_0,x_1) = \mathrm{d}\theta_{\nabla u^*(x_1)}(x_0)\mathrm{d}\nu_1(x_1) \tag{6}$$

solves $(\mathcal{B})$. Note that this gives a characterization of minimizers of problem $(\mathcal{B})$ as the composition of a martingale and a suffiently smooth transport plan. Such a characterization can also be derived as a slight modification of a result in [4]. Our proof shows that this can be alternatively obtained as a consequence of Strassen theorem and Toland duality.

References

[1] G. Carlier. *Remarks on Toland's duality, convexity constraint and optimal transport.* Pacific Journal of Optimization 4.3 (2008): 423–432.

[2] T. Gallouët, A. Natale and G. Todeschi. *From geodesic extrapolation to a variational BDF2 scheme for Wasserstein gradient flows*, Math. of Comp. (to appear), arXiv preprint arXiv:2209.14622 (2022).

[3] N. Gozlan, C. Roberto, P.-M. Samson and P. Tetali. *Kantorovich duality for general transport costs and applications*, J. Funct. Anal. 273 (2017) 3327–3405.

[4] N. Gozlan and N. Juillet. *On a mixture of Brenier and Strassen theorems.* Proceedings of the London Mathematical Society 120.3 (2020) 434–463.

Optimal transport for stationary point processes

Matthias Erbar

(joint work with Martin Huesmann, Jonas Jalowy, Bastian Müller)

We develop a theory of optimal transport for stationary random measures with a focus on stationary point processes and construct a family of distances on the set of stationary random measures. These induce a natural notion of interpolation between two stationary random measures along a shortest curve connecting them. In the setting of stationary point processes we leverage this transport distance to give a geometric interpretation for the evolution of infinite particle systems with stationary distribution. Namely, we characterise the evolution of infinitely many Brownian motions as the gradient flow of the specific relative entropy w.r.t. the Poisson point process. Further, we establish displacement convexity of the specific relative entropy along optimal interpolations of point processes and establish a stationary analogue of the HWI inequality, relating specific relative entropy, transport distance, and a specific relative Fisher information.

We consider random measures $\xi^\bullet$, i.e. random variables with values in the space $\mathcal{M}(\mathbb{R}^d)$ of locally finite measures on $\mathbb{R}^d$. $\xi^\bullet$ is a point process if it almost surely takes values in the $\mathbb{N}$. The distribution of a random measure is an element of $\mathcal{P}\big(\mathcal{M}(\mathbb{R}^d)\big)$ the set of probability measures over $\mathcal{M}(\mathbb{R}^d)$. $\mathbb{R}^d$ naturally acts on $\mathcal{M}(\mathbb{R}^d)$ by shift of the support and we say that $\mathsf{P} \in \mathcal{P}\big(\mathcal{M}(\mathbb{R}^d)\big)$ is *stationary* if it conicides with its image measure under the shift operation by any vector $x \in \mathbb{R}^d$. Stationarity of the distribution of a random measure is implied by the following stronger property. A random measure $\xi^\bullet : (\Omega, \mathcal{F}, \mathbb{P}) \to \mathcal{M}(\mathbb{R}^d)$ is called *invariant* if the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ admits a measurable flow, i.e. a family of measurable mappings $\mathbb{R}^d \times \Omega \ni (x, \omega) \mapsto \theta_x\omega \in \Omega$ with $\theta_0 = \mathsf{id}$ and $\theta_x \circ \theta_y = \theta_{x+y}$ for all $x, y \in \mathbb{R}^d$, such that $\mathbb{P}$ is invariant under θ and for all $x \in \mathbb{R}^d$, $\omega \in \Omega$, we have $\xi^\omega(\cdot) = \xi^{\theta_x\omega}(\cdot - x)$.

On the space $\mathcal{M}(\mathbb{R}^d)$ we introduce a cost function that measures the asymptotic transport cost per volume:

$$c(\xi, \eta) = \inf_{\mathsf{q} \in \mathsf{cpl}(\xi,\eta)} \limsup_{n\to\infty} \frac{1}{n^d} \int_{\Lambda_n \times \mathbb{R}^d} |x - y|^p \mathsf{q}(dx, dy) \ , \tag{1}$$

where $\mathsf{cpl}(\xi,\eta)$ denotes the set of all couplings $\mathsf{q} \in \mathcal{M}(\mathbb{R}^d \times \mathbb{R}^d)$ of ξ and η and $\Lambda_n = [-n/2, n/2]^d$ denotes the box of side length n centered at the origin.

Given $\mathsf{P}_0, \mathsf{P}_1 \in \mathsf{P}\big(\mathcal{M}(\mathbb{R}^d)\big)$ two distributions of stationary random measures with unit intensity we consider the transport problem

$$\mathsf{C}(\mathsf{P}_0,\mathsf{P}_1) = \inf_{\mathsf{Q}\in\mathsf{Cpl}_s(\mathsf{P}_0,\mathsf{P}_1)} \int c(\xi,\eta)\mathsf{Q}(d\xi,d\eta) = \inf_{(\xi^\bullet,\eta^\bullet)} \mathbb{E}\big[c(\xi^\bullet,\eta^\bullet)\big]\,, \tag{2}$$

where the first infimum is taken over the set of stationary couplings of P_0 and P_1 and the second infimum is taken over all pairs of jointly invariant random measures $(\xi^\bullet,\eta^\bullet)$ with distributions P_0 and P_1 respectively.

Note that (2) is a two layer optimisation problem. The first layer of optimisation is on the level of the coupling of the distributions, the second layer is on the level of the spatial coupling of the realisations of the random measures in the transport problem defining c. Moreover, it is an optimal transport problem with an additional probabilistic constraint, namely stationarity.

Alternatively, the value of the transport problem can be represented in terms of a cost function for pairs jointly invariant random measures with given joint distribution that was analysed by Sturm and Huesmann [1, 2]:

$$\mathsf{cost}(\xi^\bullet,\eta^\bullet) := \inf_{\mathsf{q}^\bullet\in\mathsf{cpl}_e(\xi^\bullet,\eta^\bullet)} \mathbb{E}\Big[\int_{\Lambda_1\times\mathbb{R}^d} |x-y|^p \mathsf{q}^\bullet(dx,dy)\Big]\,,$$

where the infimum runs over all *invariant* couplings $\mathsf{q}^\bullet$ of $\xi^\bullet,\eta^\bullet$. Then the following representation holds:

$$\mathsf{C}(\mathsf{P}_0,\mathsf{P}_1) = \inf_{(\xi^\bullet,\eta^\bullet)} \mathsf{cost}(\xi^\bullet,\eta^\bullet). \tag{3}$$

where the infimum runs over all jointly invariant random measures $(\xi^\bullet,\eta^\bullet)$ such that $\xi^\bullet \sim \mathsf{P}_0, \eta^\bullet \sim \mathsf{P}_1$. Moreover, if $\mathsf{C}(\mathsf{P}_0,\mathsf{P}_1) < \infty$, there exists an optimal pair $(\mathsf{Q},\mathsf{q}^\bullet)$ of a coupling Q of $\mathsf{P}_0,\mathsf{P}_1$ and $\mathsf{q}^\bullet$ achieving the infimum.

We fix $p \geq 1$ in (1) and put $\mathsf{W}_p := \mathsf{C}^{\frac{1}{p}}$.

Theorem. *W_p defines a geodesic extended distance on the space of stationary distributions $\mathcal{P}_s(\mathcal{M}(\mathbb{R}^d))$ with unit intensity.*

For P_0 and P_1 with $\mathsf{W}_p(\mathsf{P}_0,\mathsf{P}_1) < \infty$ and an optimal pair $(\mathsf{Q},\mathsf{q}^\bullet)$ the geodesic $(\mathsf{P}_a)_{a\in[0,1]}$ is given as follows. Let $\xi_a^\bullet := (\mathsf{geo}_a)_\# \mathsf{q}^\bullet$ be the random measure interpolating the points of $\xi_0^\bullet \sim \mathsf{P}_0$ and $\xi_1^\bullet \sim \mathsf{P}_1$ along straight lines, i.e. $\mathsf{geo}_a(x,y) = x + a(y-x)$, and put $\mathsf{P}_a = \mathsf{law}(\xi_a^\bullet)$.

The novel geometry induced by the transport distance W_2 can be used to study functionals on stationary point processes and infinite particle dynamics.

We consider the *specific relative entropy* of a stationary point process P with respect to the Poisson point process Poi given by

$$\mathcal{E}(\mathsf{P}) := \limsup_{n\to\infty} \frac{1}{n^d} \mathsf{Ent}(\mathsf{P}_{\Lambda_n}|\mathsf{Poi}_{\Lambda_n})\,,$$

where P_{Λ_n} denotes the restriction of P to Λ_n. Our second main result is

Theorem. *The specific entropy $\mathcal{E}$ is convex along W_2-geodesics, i.e. for any W_2-geodesic $(\mathsf{P}_a)_{a\in[0,1]}$ we have*

$$\mathcal{E}(\mathsf{P}_a) \le (1-a)\mathcal{E}(\mathsf{P}_0) + a\mathcal{E}(\mathsf{P}_1) \ .$$

This can be seen as the natural analogue in the context of stationary point processes of McCann's seminal result that the relative entropy wr.t. Lebesgue measure is convex along Wasserstein geodesics.

Similarly, we obtain an analogue for stationary point processes of the celebrated HWI inequality by Otto and Villani which relates entropy, Wasserstein distance, and Fisher information. We define the *specific (relative) Fisher information* of a stationary point process P with respect to Poi by

$$\mathcal{I}(\mathsf{P}) := \limsup_{n\to\infty} \frac{1}{n^d} I(\mathsf{P}_{\Lambda_n}|\mathsf{Poi}_{\Lambda_n}) \ . \tag{4}$$

i.e. as the large volume limit of the normalized relative Fisher information of P on boxes. Then we have the following

Theorem. *For any stationary point processes $\mathsf{P}_0, \mathsf{P}_1$ with $\mathsf{W}_2(\mathsf{P}_0,\mathsf{P}_1) < \infty$, $\mathcal{E}(\mathsf{P}_0) < \infty$ we have*

$$\mathcal{E}(\mathsf{P}_0) - \mathcal{E}(\mathsf{P}_1) \le \mathsf{W}_2(\mathsf{P}_0,\mathsf{P}_1)\sqrt{\mathcal{I}(\mathsf{P}_0)}. \tag{5}$$

In fact, the convexity of the specific entropy in Theorem is a consequence of our analysis of the gradient flow of $\mathcal{E}$ w.r.t. to the geometry induced by W_2. For a stationary point process P, let P_t be the point process obtained by evolving each point of P by an independent Brownian motion for time t.

Theorem. *$(\mathsf{P}_t)_t$ is the gradient flow of $\mathcal{E}$ wr.t. W_2 in the sense of the following* Evolution Variational Inequality*: For any stationary point processes R with $\mathsf{W}_2(\mathsf{P},\mathsf{R}) < \infty$ we have*

$$\mathsf{W}_2^2(\mathsf{P}_t,\mathsf{R}) - \mathsf{W}_2^2(\mathsf{P},\mathsf{R}) \le 2t\big[\mathcal{E}(\mathsf{R}) - \mathcal{E}(\mathsf{P}_t)\big] \ .$$

This result can be seen as the analogue for stationary point processes of the celebrated result by Jordan, Kinderlehrer and Otto that the heat flow is the Wasserstein gradient flow of the Boltzmann entropy. Among several possible characterisations of gradient flows in metric spaces, Evolution Variational Inequalities are among the strongest entailing numerous consequences for the evolution. One is the geodesic convexity of the functional $\mathcal{E}$. Furthermore, we obtain e.g. that for any P with $\mathsf{W}_2(\mathsf{P},\mathsf{Poi}) < \infty$, $\mathcal{E}(\mathsf{P}) < \infty$ we that P_t converges to Poi as $t \to \infty$ in the strong sense that $\mathcal{E}(\mathsf{P}_t) \to 0$. An example is the shifted lattice in $d \ge 3$ where each point is perturbed by an i.i.d. random variable which is uniform on a small ball.

References

[1] M. Huesmann, K.Th. Sturm, *Optimal transport from Lebesgue to Poisson*, Annals of Probability, 41(4):2426–2478, (2013).

[2] M. Huesmann, *Optimal transport between random measures*, Annales de l'Institut Henri Poincaré, Probabilités et Statistiques, 52:196–23, (2016).

[3] M. Erbar, M. Huesmann, J. Jalowy, B. Müller, *Optimal transport of stationary point processes: Metric structure, gradient flow and convexity of the specific entropy*, arXiv:2304.11145 (2023).

Minimizing the total variation in a congested setting

Annette Dumas
(joint work with Filippo Santambrogio)

The model I presented is motivated by the study of the dynamics of the population inside a city. For example, inhabitants can either chose to live in the center of the city or in the suburbs. Their trajectory can be seen as a piecewise constant curve γ that associates time $t \in [0,T]$ to a place in the city Ω. Each agent would like to minimize the following cost:

$$\min_{\substack{\gamma\in \mathrm{BV}([0,T],\Omega)\\ \gamma(0)=x_0}} \underbrace{S(\gamma)}_{\text{number of jumps}} + \int_0^T \underbrace{\frac{dI}{dm}(e_t\sharp Q)(\gamma(t))}_{I \text{ admits a first variation}} dt + \int_0^T \underbrace{F(\gamma(t), e_t\sharp Q)}_{\mathcal{C}(\Omega,\mathcal{P}(\Omega))} dt + \underbrace{\varphi(\gamma(T))}_{\text{penalization at time } T}$$

where $Q \in \mathcal{P}(\mathrm{BV}([0,T],\Omega))$ is a probability measure over the curves and e_t the evaluation at time t.

Since we consider that each player is negligible, but their strategy depends on the distribution of the other's $e_t\sharp Q$, this model is some class of Mean Field Games model whose theory was introduced simultaneously by Lasry and Lions ([1, 2]) and Caines, Huang and Malhamé ([3]). In particular, this model is a Variational Mean Field Games whose idea was introduced by Santambrogio in his lectures notes [4].

In our case, we would like to find a measure $\bar{Q}$ such that it is minimizes

$$\mathcal{U}_{\bar{Q}}(Q) := \int S(\gamma) dQ(\gamma) + \int I(e_t\sharp Q)dt + \int\int F(\gamma(t), e_t\sharp\bar{Q})dt\ dQ(\gamma). \tag{1}$$

Such a measure exists by restricting to a compact subset Γ of probability measures and applying Kakutani's theorem in order to find a fixed point.

Considering measures $Q \in \mathcal{P}(\mathrm{BV}([0,T],\Omega))$ over the curves corresponds to the Lagrangian point of view. Now, we switch to the Eulerian point of view where we consider curves of measures such as $\rho(t) = e_t\sharp Q \in \mathcal{P}(\Omega)$. A result in Optimal Transport says that when the transport cost is trivial, then the whole cost is equal to half of the total variation of the difference between the two measures:

$$\inf_{\pi\in\Pi(\mu,\nu)} \int_{\Omega\times\Omega} \mathbb{1}_{x\neq y} d\pi(x,y) = \frac{1}{2}\|\mu-\nu\|_{TV}\,. \tag{2}$$

Thanks to this remark, minimizing (1) is equivalent to minimizing

$$\min_{\substack{\rho\in E,\rho\geq 0\\ \forall t\in[0,T],\int_\Omega \rho(t,x)dx=1}} \int_0^T \int_\Omega \frac{1}{2}|\dot{\rho}| + V\rho + f(\rho)dxdt + \psi_0(\rho(0)) + \psi_T(\rho(T)), \tag{3}$$

in the sense that the minimas are equal and if $\bar{\rho}$ minimizes (3), then one can construct a measure $\bar{Q}$ by (2) such that $e_t\sharp\bar{Q} = \bar{\rho}(t)$ and $\bar{Q}$ minimizes (1). One may notice that we have considered the problem in a purely variational way, which means that $F = 0$. Moreover, the function $I(\rho) = V\rho + f(\rho)$ is the sum of a linear term and an α-convex term.

This motivates the main result of the talk, which is the Lipschitz regularity in time valued in $L^2(\Omega)$ of the unique solution $\bar{\rho}$ to (3) for which I presented an idea of the proof. The Lipschitz regularity still holds on an infinite time horizon where we consider $\rho\colon [0, +\infty[\to L^2(\Omega)$ and we add a discount factor in the integrand of (3).

When it comes to regularity in space, it can occur when we impose Dirichlet boundary conditions in time. Actually, the solution $\bar{\rho}$ inherits the regularity of V, $m_0(x)$ and $m_T(x)$ if m_0 and m_T are the boundary conditions at time 0 and T.

Numerical simulations can be carried out thanks the proximal gradient method which is presented in [5]. Several examples were displayed in the case where $\Omega = [0, T]$ for different forms of the given V: $V(t, x) = \cos(t - x)$, $V(t, x) = (x - x_0(t))^2$, etc. Generally, the solution $\bar{\rho}(t)$ either equals $c - V(t, \cdot)$, or it is constant. We noticed that with Dirichlet boundary conditions, the solution can appear non-constant in space where it should be.

It is possible to apply the model to a two-population model where we consider the densities ρ_1 and ρ_2. The problem is not necessarily purely variational since we define $I_i(\rho_i) = f(\rho_i)$ and $F_i(\rho_1, \rho_2) = V(\rho_1, \rho_2)\rho_i$, but the equivalence result still holds. Thanks to Banach fixed point theorem, a solution $(\bar{\rho}_1, \bar{\rho}_2)$ exists and can be computed by applying the algorithm several times in order to perceive the aspect of the solution.

References

[1] J.-M. Lasry, P.-L. Lions, *Jeux à champ moyen. I - Le cas stationnaire*, C. R. Acad. Sci. Paris, Ser. I 343 (2006)

[2] J.-M. Lasry, P.-L. Lions, *Jeux à champ moyen. II - Horizon fini et contrôle optimal,* C. R. Acad. Sci. Paris, Ser. I 343 (2006)

[3] M. Huang, R. P. Malhamé and P. E. Caines, *Large Population Stochastic Dynamic Games: Closed-Loop Mckean Vlasov Systems and the Nash Certainty Equivalence Principle*, Communications in Information and Systems, Vol. 6, No. 3, 2006, pp. 221-252.

[4] F. Santambrogio, Lecture notes on Variational Mean Field Games, *Mean Field Games – Cetraro, Italy, 2019*, Cardaliaguet and Porretta (Eds), Springer, C.I.M.E. Foundation Subseries.

[5] A. Beck, *First-Order Methods in Optimization*. Society for Industrial and Applied Mathematics, 2017.

Stefan problem via optimal Skorokhod problem

Inwon Kim

In the talk we discuss particle-oriented formulation of Stefan problem, both for the stable problem which describes melting of the ice into water, and for the unstable problem which describes freezing of the supercooled water into the ice. While

this formulation itself is interesting for both problems, it is particularly helpful to understand the unstable problem, in that our formulation yields global existence (and partial stability) of solutions for the problem: previously such existence was only known in one space dimension.

Our approach builds on an optimization problem, where the stopping time of brownian motion is optimized to minimize the given Lagrangian cost, with a given initial and final-time distribution of the brownian motion.This problem (optimal Skorokhod problem) can be viewed as a variant of optimal transport problem, as discussed in [1]. Their it is shown that the optimal stopping time is characterized as an exit time of a space-time domain. This suggests potential connections between the optimal Skorokhod problem and interface motions created by Brownian particles. In the talk we describe how one can characterize both types of Stefan problems when one add another layer on the optimal Skorokhod problem, where now one optimize over the final-time distribution of the brownian motion, under a upper bound constraint. Formally the optimization problem describes the motion of Brownian particles with stopping time that tries to minimize the kinetic energy during their motion, but also under the capacity constraint for where they are allowed to stop. This interpretation coincides with the formal description of interacting particle systems that leads to the Stefan problem. The talk is based on the works [3] and [2].

References

[1] M. Beiglboeck, A. Cox and M. Huesmann, *Optimal transport and Skorokhod embedding,* Inventiones Mathematicae **208**, No. 2(2017) pp.327–400.
[2] S. Choi, I.C. Kim and Y.H. Kim, *Existence for the Supercooled problem in general dimensions,* arXiv:2402.17154
[3] I.C. Kim and Y.H. Kim, *The Stefan problem and Free Targets of Optimal Brownian Martingale Transport* to appear in Annals of Applied Probability

Thermodynamic gradient flow evolution from particles

Johannes Zimmer

In this talk, we consider diffusive scalar equations and discuss methods to determine coefficients of their evolution operator or the entire operator itself from data, here particle models. Thus, a prerequisite is that the equation under consideration is the hydrodynamic limit of a particle model.

We study equations of the form

$$\partial_t \rho(x,t) = \mathbb{K}(\rho) D\mathcal{S}(\rho),$$

where $\mathbb{K}$ is the evolution operator, acting on the derivative of the entropy $\mathcal{S}$. Examples include linear and nonlinear diffusion (as scaling limit of Brownian particles and the zero-range process, for example).

The description of the evolution of *finitely many* particles can be shown, at least formally, to be given by a stochastic differential equation, namely

$$\partial_t \rho = \mathbb{K}(\rho) D\mathcal{S}(\rho) + \epsilon \sqrt{2\mathbb{K}(\rho)} \dot{W}_{x,t},$$

where ϵ is related to the number of particles and $\dot{W}_{x,t}$ is space-time white noise. For example, for the zero-range process, which has a nonlinear diffusion equation $\partial_t \rho(x,t) = \Delta(\Phi(\rho(x,t))$ as hydrodynamic limit, this equation can be written in more explicit form as

$$\partial_t \rho_L = \frac{1}{2}\Delta(\Phi(\rho_L)) + \frac{1}{\sqrt{L^d}}\mathrm{div}\left(\sqrt{2m(\rho_L)}\dot{W}_{x,t}\right).$$

The talk concerns the question how to compute the mobility m from particle data, and more generally the evolution operator $\mathbb{K}$, first in situations where $\mathcal{S}$ is known and then in one situation where the entropy/free energy is unknown. To explain the idea, we describe the analogous situation for a stochastic ODE (see [1] for this analogy and more details):

$$\mathrm{d}X = f\,\mathrm{d}t + \sqrt{\sigma}\,\mathrm{d}W.$$

It is a classic result that σ can be determined from trajectories X by computing the quadratic variation,

$$\lim_{h\searrow 0}\frac{1}{h}\mathbb{E}\left[\left[X(t_0+h) - X(t_0)\right]^2\right] = \sigma.$$

The benefit for us in this formula is that the left-hand side can be determined in computer simulations.

In this talk, we show that this idea can be extended to the setting of diffusive partial differential equations. Details are described in the papers [1, 4, 3]. At the core, the idea is to study fluctuations, by considering the stochastic process $Y^L(t,x) := \sqrt{L^d}\,(\rho_L(t,x) - \rho(t,x))$ (where ρ_L is the empirical measure of finitely many particles and ρ is the limit $L\to\infty$; d is the dimension). It is convenient to consider a weak formulation, $Y^L_\gamma(t) := \sqrt{L^d}\,\langle \gamma, \rho_L(t,\cdot) - \rho(t,\cdot)\rangle$, with test function γ. Then $Y^L_\gamma(t) = \langle Y, \gamma\rangle + O(1/L^d) =: Y_\gamma + O(1/L^d)$, with

$$\mathrm{d}Y_\gamma(t) = \frac{1}{2}\left\langle \Delta\gamma, \Phi'(\rho(t,.))Y(t,\cdot)\right\rangle \mathrm{d}t - \left\langle \nabla\gamma, \sqrt{2m(\rho(t,\cdot))}\,\mathrm{d}W_{x,t}\right\rangle,$$

see [2]. The noise term contains the quantities of interest (the mobility m and in fact the associated Wasserstein operator in a weak form). As in the case of the ODE, this expression can be computed using the quadratic variation. Indeed, one can show that [1]

$$m(\rho(t_0,x_0)) \simeq \frac{\lim_{h\searrow 0}\frac{1}{h}\mathbb{E}\left[\left[Y_\gamma(t_0+h) - Y_\gamma(t_0)\right]^2\right]}{2\left\langle \nabla\gamma, \nabla\gamma\right\rangle}.$$

For the zero-range process, it is possible to compute the mobility m explicitly in the case of quadratic jump rates, and one finds [1] an expression in terms of modified Bessel functions I_j, $\rho\left(\frac{m}{2}\right) = \sqrt{m}\cdot\frac{I_1(2\sqrt{m})}{I_0(2\sqrt{m})}$, which defines m implicitly.

One can show that with his approach, one can compute diffusivities even in situations where standard approaches based on the mean square displacement are not applicable, as tagged particles move too slowly (standard mean square displacement methods require the mean square displacement to scale $\sim t$, while the

mean square displacement for the symmetric exclusion process in one dimension scales $\sim \sqrt{t}$).

One can see that in this manner, it is possible to compute the entire operator $\mathbb{K}$ in this way, hence, if the entropy $\mathcal{S}$ is known. This can be achieved by considering the weak formulation of the gradient flow $\partial_t \rho = \mathbb{K}(\rho)F$, (that is, $F = D\mathcal{S}$), where $\{\gamma_a\}$ is a basis of functions parametrised by a,

$$\begin{aligned} \rho(x,t) &\approx \textstyle\sum_a \rho_a(t)\gamma_a(x) \\ F(x,t) &\approx \textstyle\sum_a F_a(t)\gamma_a(x) \\ \textstyle\sum_a \langle \gamma_a, \gamma_b\rangle \partial_t \rho_a &\approx \textstyle\sum_a \langle \mathbb{K}(\rho)\gamma_a, \gamma_b\rangle F_a \text{ for all } b. \end{aligned}$$

Similar as in the case of the mobility, one can show [4]

$$\langle \mathbb{K}(\rho)\gamma_a, \gamma_b\rangle = \lim_{h\searrow 0} \frac{1}{2h}\mathbb{E}\Big[\left(Y_{\gamma_a}(t_0+h) - Y_{\gamma_a}(t_0)\right)\cdot \left(Y_{\gamma_b}(t_0+h) - Y_{\gamma_b}(t_0)\right)\Big].$$

For example, it is possible to determine in this way the evolution of the hydrodynamic limit of the zero-range process.

We remark that it is possible to apply this framework also in situations where the entropy is not known. An example is Arrhenius diffusion, which appears in surface absorption, chemical reactions, or vacancies and intersticials in solids. Here jump rates are given by $p(x \to y) = d\,\eta(x)\,(1-\eta(y))\,e^{-\beta U(x)}$ (jump frequency d, energy $U(x) = U_0 + \sum_{\xi\neq x} J(x-\xi)\eta(\xi)$, with binding energy U_0 and interaction J), where it is possible to learn the free energy as driving functional in the case of short-range interaction [3]. In the case of long-range interaction the hydrodynamic limit is known [5],

$$\partial_t \rho = -\mathrm{div}\,(m(\rho)\nabla D\mathcal{S}) =: \mathbb{K}(\rho)D\mathcal{S}(\rho)$$

with mobility $m\,(\rho) = D\rho(1-\rho)e^{-J*\rho}$, $D = de^{-U_0}$ and free energy

$$S(\rho) = \int \frac{1}{2}\rho\,(J*\rho)\,\mathrm{d}x - \int \left[\rho\ln\rho + (1-\rho)\ln(1-\rho)\right]\mathrm{d}x\,.$$

The difference to short-range interaction is that under long-range, fluctuations around averages become independent, and law of large number applies.

References

[1] Peter Embacher, Nicolas Dirr, Johannes Zimmer, and Celia Reina. Computing diffusivities from particle models out of equilibrium. *Proc. A.*, 474(2212):20170694, 21, 2018.

[2] P. A. Ferrari, E. Presutti, and M. E. Vares. Non equilibrium fluctuations for a zero range process. *Ann. Inst. H. Poincaré Probab. Statist.*, 24(2):237–268, 1988.

[3] Shenglin Huang, Zequn He, Nicolas Dirr, Johannes Zimmer, and Celia Reina. Statistical-Physics-Informed Neural Networks (Stat-PINNs): A machine learning strategy for coarse-graining dissipative dynamics. http://arxiv.org/abs/2312.05810, 2023.

[4] Xiaoguai Li, Nicolas Dirr, Peter Embacher, Johannes Zimmer, and Celia Reina. Harnessing fluctuations to discover dissipative evolution equations. *J. Mech. Phys. Solids*, 131:240–251, 2019.

[5] D. G. Vlachos and M. A. Katsoulakis. Derivation and validation of mesoscopic theories for diffusion of interacting molecules. *Phys. Rev. Lett.*, 85:3898–3901, 2000.

Disentangling Entropy and Suboptimality in Entropic Optimal Transport

Maxime Sylvestre
(joint work with Hugo Malamut)

In this work we study the regularized optimal transport problem for a cost $c \in \mathcal{C}^2(\mathbb{R}^d \times \mathbb{R}^d, \mathbb{R})$

$$\inf_{\gamma \in \Pi(\mu_0,\mu_1)} \int c d\gamma + \varepsilon H(\gamma | \mu_0 \otimes \mu_1) \tag{1}$$

where the infimum is taken over all measures $\gamma \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$ with marginals μ_0, μ_1 which are two absolutely continuous probability measures with finite moment of ordre two. Here $H(.|.)$ is the relative entropy also known as Kullback-Leibler divergence. We focus on the case where μ_0, μ_1 have finite entropy with respect to $\mathcal{H}^d$, the Lebesgue measure, that is $H(\mu_i \mid \mathcal{H}^d) < \infty$ and have finite order two moments. In that case the minimizer γ_ε is the same as if $\mu_0 \otimes \mu_1$ is replaced by the Lebesgue measure $\mathcal{H}^{2d}$ in the entropy term, and hence we consider

$$OT_\varepsilon := \inf_{\gamma \in \Pi(\mu_0,\mu_1)} \int c d\gamma + \varepsilon H(\gamma | \mathcal{H}^{2d}). \tag{εEOT}$$

Note that $\varepsilon = 0$ yields the classical optimal transport problem. In this case the minimizer need not to be unique and γ_0 will denote any of them.

The regularized problem (εEOT) is a smoothed version of the optimal transport problem it is thus natural to ask its properties as $\varepsilon \to 0$. It is well known [6] that the problem (εEOT) Γ-converges to the optimal transport problem OT_0 . Here in trying to understand the convergence of the entropic optimal transport problem, we are interested in disentangling the role of the entropy $H(\gamma_\varepsilon \mid \mathcal{H}^{2d})$ and the role of the suboptimality $\int c d\gamma_\varepsilon - \int c d\gamma_0$. The suboptimality is of interest in itself because it is a faster converging approximation of OT_0. And the entropy term is of interest because is lower bounds $W_2(\gamma_\varepsilon, \gamma_0)$, the Wasserstein distance between the entropic optimal transport plan γ_ε and any optimal transport plan γ_0.

The disentangling relies on two key elements. First the known convergence rates of the value of the regularized problem. Indeed under a compactness of the support assumption [2, 5] on μ_0, μ_1, we have

$$OT_\varepsilon = (c, \gamma_\varepsilon) + \varepsilon H(\gamma_\varepsilon \mid \mathcal{H}^{2d}) \leq OT_0 - \frac{d}{2}\varepsilon \ln(\varepsilon) + O(\varepsilon). \tag{2}$$

Also for the quadratic cost and under the stronger assumption of finite Fisher information for the marginals, [3, claim 4.1], we have

$$OT_\varepsilon = OT_0 - \frac{d}{2}\varepsilon ln(2\pi\varepsilon) + \varepsilon H_m + o(\varepsilon) \tag{3}$$

where $H_m := \frac{1}{2}[H(\mu_0 \mid \mathcal{H}^d) + H(\mu_1 \mid \mathcal{H}^d)]$.

The second key element is the lower bound of the entropy by quadratic distance to lower dimensional measure. This can be stated in the following way. Let γ be

a probability measure on $\mathbb{R}^{2d}$ with finite entropy with respect to $\mathcal{H}^d$. Let γ_0 be a probability measure on $\mathbb{R}^{2d}$ supported on a k-dimensional subspace of $\mathbb{R}^{2d}$. Then

$$H(\gamma \mid \mathcal{H}^d) \geq \frac{2d-k}{2} \ln(W_2^2(\gamma, \gamma_0)) + C_{\gamma_0}.$$

In fact the suboptimality enjoys a similar property

$$H(\gamma_\varepsilon \mid \mathcal{H}^d) \geq \frac{2d-d}{2} \ln\left(\int c d\gamma_\varepsilon - \int c d\gamma_0\right) + C_{\gamma_0}.$$

By combining the two elements above we get the following three results of increasing generality.

First if the marginals have finite Fisher information and are compactly supported then we are able to retrieve a taylor expansion for the suboptimality and the entropy term. It is remarkable that the suboptimality is always $\frac{d}{2}\varepsilon$ at the first order.

Theorem. *Suppose that the cost is quadratic, that is $c(x,y) = \frac{1}{2}\|x-y\|^2$. Further assume that $I(\mu_i) < \infty$ and $Supp(\mu_i)$ compact. Then*

$$H(\gamma_\varepsilon \mid \mathcal{H}^{2d}) = -\frac{d}{2} ln(2\pi\varepsilon) + H_m - \frac{d}{2} + o(1) \tag{4}$$

and

$$(c, \gamma_\varepsilon) = OT_0 + \frac{d}{2}\varepsilon + o(\varepsilon) \tag{5}$$

Secondly if we relax the finite Fisher information and the compact support assumptions then in the case of the quadratic cost we still recover the first order of magnitude. However the precise expansion is unclear.

Theorem. *Suppose that the cost is quadratic, that is $c(x,y) = \frac{1}{2}\|x-y\|^2$. Further assume that μ_i have finite moment of order $2+\delta$ then*

(6)

$$(c, \gamma_\varepsilon) = OT_0 + \Theta(\varepsilon), \quad H(\gamma_\varepsilon \mid \mathcal{H}^{2d}) = -\frac{d}{2}\ln(\varepsilon) + O(1), \quad \sqrt{\varepsilon} = O(W_2(\gamma_\varepsilon, \gamma_0))$$

In the special case where the Monge map ∇f associated to the optimal transport plan γ_0 is Lipschitz then

$$W_2(\gamma_\varepsilon, \gamma_0) = \Theta(\sqrt{\varepsilon}) \tag{7}$$

Note that the lower bound on the Wasserstein distance between γ_ε and the unique optimal transport plan γ_0 is tight in the sense that if the Monge map is Lipschitz then the Wasserstein distance is exactly of order $\sqrt{\varepsilon}$.

Finally we introduce a similar result for infinitesimally twisted costs under the stronger assumption of compactly supported marginals. We recall that a function $c : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ is infinitesimally twisted if it is twice differentiable $\nabla^2_{xy} c$ is invertible everywhere. This property of the cost allows to locally retrieve the structure of the optimal transport plan γ_0 that we have in the quadratic case [1]. And we conclude by localizing the argument of the previous theorem.

Theorem. *Suppose that the cost is $\mathcal{C}^2$ and infinitesimally twisted. Further assume that μ_i is compactly supported then*
(8)

$$(c, \gamma_\varepsilon) = OT_0 + \Theta(\varepsilon), \quad H(\gamma_\varepsilon \mid \mathcal{H}^{2d}) = -\frac{d}{2}\ln(\varepsilon) + O(1), \quad \sqrt{\varepsilon} = O(W_2(\gamma_\varepsilon, \gamma_0))$$

References

[1] McCann, Robert J. and Pass, Brendan and Warren, Micah. Rectifiability of Optimal Transportation Plans.Canadian Journal of Mathematics, 64, 4, 924–934, 2012.

[2] Carlier, Guillaume and Pegon, Paul and Tamanini, Luca. Convergence rate of general entropic optimal transport costs. Calc. Var. 62, 116, 2023.

[3] Erbar, Matthias and Maas, Jan and Renger, Michiel. From large deviations to Wasserstein gradient flows in multiple dimensions. Electronic Communications in Probability, 20, 1–12, 2015.

[4] Hugo Malamut, Maxime Sylvestre. Convergence Rates of the Regularized Optimal Transport: Disentangling Suboptimality and Entropy. 2306.06940, arXiv, 2023.

[5] Eckstein, Stephan and Nutz, Marcel. Convergence Rates for Regularized Optimal Transport via Quantization. Mathematics of Operations Research, 49, 2, 1223–1240, 2024.

[6] Carlier, Guillaume and Duval, Vincent and Peyré, Gabriel and Schmitzer, Bernhard. Convergence of entropic schemes for optimal transport and gradient flows. SIAM Journal on Mathematical Analysis, 49, 2, 1385–1418, 2017.

Opial Property in Wasserstein Spaces and Asymptotic Convergence of EVI Flows and of Proximal Point Algorithms

Giuseppe Savaré

(joint work with Emanuele Naldi)

According to a nice result by Opial [9], weak convergence in a separable Hilbert space $(H, |\cdot|)$ admits a purely metric characterization: if $(x_n)_{n\in\mathbb{N}}$ is a sequence in H weakly converging to $x \in H$, then

$$|y - x|^2 + \liminf_{n\to\infty} |x_n - x|^2 \le \liminf_{n\to\infty} |x_n - y|^2. \tag{1}$$

The *proof* can be easily obtained by passing to the limit in the identity

$$|x_n - y|^2 = |x_n - x|^2 + |x - y|^2 + 2\langle x_n - x, x - y\rangle,$$

using the fact that $\lim_{n\to\infty}\langle x_n - x, x - y\rangle = 0$ by weak convergence. (1) shows in particular that the weak limit x of a sequence $(x_n)_{n\in\mathbb{N}}$ is the unique strict minimizer of the function

$$L(y) := \liminf_{n\to\infty} |x_n - y|.$$

Opial property has been further extended and studied in more general Banach spaces (see e.g. [10]) and has many interesting applications: we just quote here the approximation of a fixed point of a non-expansive map defined in a closed and convex subset of H [9], the asymptotic convergence of gradient flows of convex and lower semicontinuous functions $\varphi : H \to (-\infty, +\infty]$ to a minimizer of φ [2], and the convergence of the Proximal Point Algorithm [4, 5, 11].

In the paper [8] we study the extension of the Opial Lemma to the metric space $(\mathcal{P}_2(H), W_2)$ of Borel probability measures on H with finite quadratic moments, endowed with the Kantorovich-Rubinstein-Wasserstein distance W_2 (see e.g. [13, 1, 12]) and we derive similar applications.

A first question concerns the appropriate definition of a suitable weak topology in $\mathcal{P}_2(H)$, which enjoys at least some of the most useful properties of weak convergence in Hilbert spaces:

(1) bounded sequences admit weakly convergent subsequences,
(2) the distance function from a given element is a weakly lower semicontinuous map,
(3) the scalar product is sequentially continuous w.r.t. strong/weak convergence of its factors,
(4) weakly convergent sequences are bounded,
(5) strongly closed convex sets are also weakly closed.

Recalling that the topology of $\mathcal{P}_2(H)$ can be equivalently characterized as the initial topology induced by the family of real functions $F_\zeta : \mathcal{P}_2(H) \to \mathbb{R}$ (i.e. the coarsest topology that makes all the functions F_ζ) where

$$F_\zeta : \mu \to \int_H \zeta \, \mathrm{d}\mu, \quad \zeta \in \mathrm{C}(H), \ \sup_{x\in H} \frac{\zeta(x)}{1+|x|^2} < \infty,$$

we thus define the weak topology in $\mathcal{P}_2(H)$ as the initial topology $\sigma(\mathcal{P}_2(H), \mathrm{C}_2^w(H))$ induced by F_ζ as ζ varies in the set

$$\mathrm{C}_2^w(H) := \Big\{ \zeta : H \to \mathbb{R} \text{ is seq. weakly continuous}, \ \lim_{|x|\to\infty} \frac{\zeta(x)}{1+|x|^2} = 0 \Big\},$$

and we call $\mathcal{P}_2^w(H)$ the corresponding topological space $\big(\mathcal{P}_2(H), \sigma(\mathcal{P}_2(H), \mathrm{C}_2^w(H))\big)$. In this way, $\mathcal{P}_2^w(H)$ inherits the weak* topology of a subset of the dual of the Banach space $\mathrm{C}_2^w(H)$. If H is finite dimensional then $\mathrm{C}_2^w(H)$ contains all the continuous functions $\zeta : H \to \mathbb{R}$ such that $\lim_{|x|\to\infty} \frac{\zeta(x)}{1+|x|^2} = 0$.

The sequence of measures $\mu_n = (1-\frac{1}{n})\delta_0 + \frac{1}{n}\delta_{\sqrt{n}}$ in $\mathcal{P}_2(\mathbb{R})$ is a simple example of weakly converging sequence to $\mu = \delta_0$ in $\mathcal{P}_2^w(\mathbb{R})$ which does not converge in $(\mathcal{P}_2(\mathbb{R}), W_2)$ since $W_2(\mu_n, \mu) \equiv 1$.

We show that the weak topology of $\mathcal{P}_2^w(H)$ satisfies all the previous properties (1),... (5). In particular every lower semicontinuous geodesically convex function [6] $\phi : \mathcal{P}_2(H) \to (-\infty, +\infty]$ is also sequentially lower semicontinuous w.r.t. the weak topology $\sigma(\mathcal{P}_2(H), \mathrm{C}_2^w(H))$. As a byproduct, for every $\mu_0 \in \mathcal{P}_2(H)$ and $\tau > 0$ the Proximal Point Algorithm in $\mathcal{P}_2(H)$ (also known as JKO [3] or Minimizing Movement scheme [1])

$$\mu_\tau^k \quad \text{minimizes} \quad \mu \mapsto \frac{1}{2\tau} W_2^2(\mu, \mu_\tau^{k-1}) + \phi(\mu) \tag{2}$$

has always a solution $(\mu_\tau^k)_{k\in\mathbb{N}}$.

It turns out that the Opial property holds in $\mathcal{P}_2^w(H)$ with the same structure of (1).

Theorem ([8]). *If $(\mu_n)_{n\in\mathbb{N}}$ is a sequence weakly converging to μ in $\mathcal{P}_2^w(H)$ Then*

$$W_2^2(\nu,\mu)+\liminf_{n\to\infty} W_2^2(\mu_n,\mu)\le \liminf_{n\to\infty} W_2^2(\mu_n,\nu) \qquad \text{for every } \nu\in\mathcal{P}_2(H).$$

Applications to the asymptotic convergence of the gradient flows of a lower semicontinuous and geodesically convex functional $\phi:\mathcal{P}_2(H)\to(-\infty,+\infty]$ can then be easily derived by the same strategy of [2], by using the metric characterization of a solution $\mu:(0,\infty)\to D(\phi)$ of the gradient flow of ϕ in $\mathcal{P}_2(H)$ in terms of Evolution Variational Inequalities [1, 7]

$$\text{(EVI)} \qquad \frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}W_2^2(\mu_t,\sigma)\le\phi(\sigma)-\phi(\mu_t) \quad \text{a.e. in } (0,\infty), \quad \text{for every } \sigma\in D(\phi).$$

Theorem ([8]). *Let $\phi:\mathcal{P}_2(H)\to(-\infty,+\infty]$ be a proper, l.s.c. and geodesically convex functional and let $\mu:(0,+\infty)\to\mathcal{P}_2(H)$ be a locally Lipschitz curve satisfying* (EVI). *Then* $\operatorname{argmin}\phi\neq\emptyset$ *if and only if the curve $(\mu_t)_{t\ge1}$ is bounded in $\mathcal{P}_2(H)$; in this case there exists a minimizer μ of ϕ such that $\mu_t\to\mu$ in $\mathcal{P}_2^w(H)$ as $t\to+\infty$.*

An analogous result holds for the convergence of the Proximal Point Algorithm (2) (here we use the discrete estimates of [1] for (2) assuming convexity along generalized geodesics)

Theorem ([8]). *Let $\phi:\mathcal{P}_2(H)\to(-\infty,+\infty]$ be proper, lower semicontinuous and convex along generalized geodesics and let $(\mu_\tau^k)_{k\in\mathbb{N}}$ be a solution to the PPA algorithm* (2). *Then* $\operatorname{argmin}\phi\neq\emptyset$ *if and only if $(\mu_\tau^k)_{k\in\mathbb{N}}$ is bounded in $\mathcal{P}_2(H)$. If this is the case, there exists the limit $\mu:=\lim_{k\to\infty}\mu_\tau^k$ in $\mathcal{P}_2^w(H)$ and μ is a minimizer of ϕ.*

References

[1] L. Ambrosio, N. Gigli, and G. Savaré. *Gradient flows in metric spaces and in the space of probability measures*. Lectures in Mathematics ETH Zürich. Birkhäuser Verlag, Basel, second edition, 2008.

[2] R.E. Bruck, Jr. Asymptotic convergence of nonlinear contraction semigroups in Hilbert space. *J. Functional Analysis*, 18:15–26, 1975.

[3] R. Jordan, D. Kinderlehrer, and F. Otto. The variational formulation of the Fokker-Planck equation. *SIAM J. Math. Anal.*, 29(1):1–17, 1998.

[4] B. Martinet. Régularisation d'inéquations variationnelles par approximations successives. *Rev. Française Informat. Recherche Opérationnelle*, 4(Sér. R-3):154–158, 1970.

[5] B. Martinet. Détermination approchée d'un point fixe d'une application pseudo-contractante. Cas de l'application prox. *C. R. Acad. Sci. Paris Sér. A-B*, 274:A163–A165, 1972.

[6] R.J. McCann. A convexity principle for interacting gases. *Adv. Math.*, 128(1):153–179, 1997.

[7] M. Muratori and G. Savaré. Gradient flows and evolution variational inequalities in metric spaces. I: Structural properties. *J. Funct. Anal.*, 278(4):108347, 67, 2020.

[8] E. Naldi, G. Savaré. Weak topology and Opial property in Wasserstein spaces, with applications to gradient flows and proximal point algorithms of geodesically convex functionals *Atti Accad. Naz. Lincei Rend. Lincei Mat. Appl.*, 32:725–750, 2021.

[9] Z. Opial. Weak convergence of the sequence of successive approximations for nonexpansive mappings. *Bull. Amer. Math. Soc.*, 73:591–597, 1967.

[10] S. Prus. Banach spaces with the uniform Opial property. *Nonlinear Anal.*, 18(8):697–704, 1992.

[11] R.T. Rockafellar. Monotone operators and the proximal point algorithm. *SIAM J. Control Optim.*, 14(5):877–898, 1976.
[12] F. Santambrogio. *Optimal transport for applied mathematicians*, volume 87 of *Progress in Nonlinear Differential Equations and their Applications*. Birkhäuser/Springer, Cham, 2015. Calculus of variations, PDEs, and modeling.
[13] C. Villani. *Optimal transport*, volume 338 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 2009. Old and new.

Convergence rate for the incompressible limit of porous medium equations

Noemi David
(joint work with T. Debiec, B. Perthame, A.R. Mészáros, F. Santambrogio)

We consider the following porous medium equation with convective effects

$$\partial \varrho_m = \Delta \varrho_m^m + \nabla \cdot (\varrho_m \nabla V), \qquad m > 1. \tag{1}$$

Our goal is to infer quantitative estimates on the rate of convergence of the solution towards its so-called *incompressible* or *stiff-pressure* limit when $m \to \infty$.

1. Motivations

Nonlinear partial differential equations such as (1) have been employed in a variety of applications such as, for instance, the description of tissue growth and pedestrian motion. The density of individuals, $\varrho_m(x,t)$, evolves under the effect of an external drift $V : \mathbb{R}^d \to \mathbb{R}$ and Darcy's law. Indeed, the velocity field of the continuity equation is $-\nabla p_m - \nabla V$ where the pressure satisfies the law

$$p_m = P_m(\varrho_m) := \frac{m}{m-1} \varrho_m^{m-1}, \qquad m > 1.$$

We also treat the case of the so-called *singular pressure law*, namely

$$p_\varepsilon = P_\varepsilon(\varrho_\varepsilon) := \varepsilon \frac{\varrho_\varepsilon}{1 - \varrho_\varepsilon}, \qquad \varepsilon > 0. \tag{2}$$

In the limit $m \to \infty$ or $\varepsilon \to 0$, both laws lead to the graph relation

$$p_\infty (1 - \varrho_\infty) = 0, \qquad 0 \le \varrho_\infty \le 1.$$

It is well known that equation (1) possesses a gradient flow structure with respect to the 2-Wasserstein distance, and the associated energy functional is

$$E_m(\varrho) = \int_{\mathbb{R}^d} \varrho V \mathrm{d}x + \frac{1}{m-1} \int_{\mathbb{R}^d} \varrho^m \mathrm{d}x.$$

In [2] the authors prove that as $m \to \infty$ the gradient flow solution, ϱ_m, converges to ϱ_∞, which is the gradient flow associated to the limit energy functional

$$E_\infty(\varrho) = \begin{cases} \int_{\mathbb{R}^d} \varrho V \mathrm{d}x, & \text{for } \varrho \le 1 \text{ a.e.,} \\ +\infty, & \text{otherwise.} \end{cases} \tag{3}$$

The gradient flow associated to (3) has been extensively studied due to its importance in applications, in particular, to its equivalence to the crowd motion model proposed in the seminal work [4]. In [2] the authors also show that ϱ_∞ is a solution of a free boundary problem of Hele-Shaw type, which is also widely used to describe tumour growth, see for instance [5].

Using the JKO scheme, [2, Theorem 4.2] provides the first result on the convergence rate for ϱ_m as $m \to \infty$, namely,

$$\sup_{t\in[0,T]} W_2(\varrho_m(t), \varrho_\infty(t)) \leq \frac{C(T)}{m^{1/24}}.$$

The result is obtained assuming the existence of $\lambda \in \mathbb{R}$ such that $D^2V \geq \lambda I_d$.

2. Results on the convergence rate

2.1. **Result in a negative Sobolev norm.** In [3], in collaboration with Tomasz Debiec and Benoît Perthame, we show that it is possible to find a faster polynomial rate in the $\dot{H}^{-1}$-norm. We assume that the potential satisfies

$$D^2V - \frac{\Delta V}{2} I_d \geq \lambda I_d, \quad \text{for some } \lambda \in \mathbb{R}. \tag{4}$$

Theorem. *Let ϱ_m be the weak solution of* (1) *endowed with initial data ϱ^{in} with compact support in $\mathbb{R}^d$. There exists a unique $\varrho_\infty \in C(0,T;\dot{H}^{-1}(\mathbb{R}^d))$ such that*

$$\sup_{t\in[0,T]} \|\varrho_m(t) - \varrho_\infty(t)\|_{\dot{H}^{-1}(\mathbb{R}^d)} \leq \frac{C(T)}{\sqrt{m}}.$$

2.2. **Idea of the proof.** Unlike for the 2-Wasserstein distance, the drift part in equation (1) is not a gradient flow with respect to the $\dot{H}^{-1}$-norm. However, this choice allows us to account also for additional (linear) reaction terms, $\varrho G(x,t)$. Our strategy relies on computing the evolution of the $\dot{H}^{-1}$-norm of the difference between ϱ_{m_1} and ϱ_{m_2}, solutions to equation (1) with $m_1 < m_2$. The estimation of the porous medium part of the equation gives the polynomial rate, while the drift and growth part can be controlled by the distance itself upon using integration by parts and Young's and Sobolev's inequalities. Our estimate leads to a differential inequality of the form

$$\frac{\mathrm{d}}{\mathrm{d}t}\|\varrho_{m_1} - \varrho_{m_2}\|^2_{\dot{H}^{-1}} \leq \frac{C}{m_1} + \frac{C}{m_2} - \lambda\|\varrho_{m_1} - \varrho_{m_2}\|^2_{\dot{H}^{-1}},$$

and we conclude by using Gronwall's lemma and sending $m_2 \to \infty$. The compact support assumption on the initial data is needed to ensure that the pressure, p_m, is uniformly bounded in $L^\infty(0,T;L^\infty(\mathbb{R}^d))$. This bound is technically necessary to treat the porous medium term which gives the rate.

We could also recover the same polynomial rate for the singular pressure law (2), namely $\sqrt{\varepsilon}$ as $\varepsilon \to 0$. However, in this case, the compact support assumption is no longer needed since the uniform (in ε) bound $\varrho_\varepsilon < 1$ is already available thanks to the singularity of the law itself.

2.3. **Result in the 2-Wasserstein distance.** In a work in progress in collaboration with Alpár R. Mészáros and Filippo Santambrogio, we are able to further improve the result in the 2-Wasserstein distance. We adopt the same strategy used in [3] as, instead of discretizing in time, we compute $W_2^2(\varrho_{m_1}, \varrho_{m_2})$ along the flow and then use Gronwall's lemma. We assume that the potential V satisfies

$$\alpha I_d \leq D^2 V \leq \beta I_d, \quad \text{for some } \ 0 < \alpha \leq \beta.$$

Theorem. *Let ϱ_m be the gradient flow solution of* (1)*. There exists a unique $\varrho_\infty \in C(0,\infty;\mathcal{P}_2(\mathbb{R}^d))$ such that*

$$\sup_{t\in[0,\infty)} W_2(\varrho_m(t), \varrho_\infty(t)) \leq \frac{C}{\sqrt{m}}. \tag{5}$$

2.4. **Idea of the proof.** Computing the time derivative of the square of the distance and using the Monge-Ampère equation to estimate the nonlinear diffusion part, we find the following inequality

$$\frac{\mathrm{d}}{\mathrm{d}t} W_2^2(\varrho_{m_1}, \varrho_{m_2}) \leq \frac{C}{m_1} + \frac{C}{m_2} - \alpha W_2^2(\varrho_{m_1}, \varrho_{m_2}).$$

The fact that the rate is *global* in time is a consequence of the contractivity property of the 2-Wasserstein distance for strictly convex potentials [1]. If α is non-positive the result holds locally in time. This highlights a clear advantage in using W_2, since to have a global rate in the $\dot{H}^{-1}$-norm one would need $\lambda > 0$ in (4), which is, for instance, not possible for $d = 2$.

Furthermore, we show that the assumption on the support of the initial data imposed in Theorem 2.1 is not necessary. In fact, instead of using an L^∞-control on the pressure, p_m, we are able to prove that a higher integrability of ϱ_m is sufficient to close the computation – in particular, a bound in L^{2m-1} which is available thanks to the control of the Fisher information given by the gradient flow structure of the equation. Once again, this technical argument is not needed for the singular law. Since we already know that $\varrho_\varepsilon < 1$, the argument turns out to be even simpler in that this implies that we can directly compute the time derivative of $W_2^2(\varrho_\varepsilon, \varrho_\infty)$ rather than considering two different sequences of solutions.

3. Open problems

Despite these recent advances, several questions remain open. In particular:

- is this estimate on the rate optimal?
- which is the convergence rate of the pressure sequence, p_m, as $m \to \infty$?
- can we include pressure-dependent reaction rates, $\varrho G(p)$, in equation (1)?

References

[1] L. Ambrogio, N. Gigli, G. Savaré, *Gradient flows in metric spaces and in the space of probability measures*, Second. Birkhäuser Verlag, Basel (2008)

[2] D. Alexander, I. Kim, Y. Yao, *Quasi-static evolution and congested crowd transport*, Nonlinearity, **27(4)**, (2014) 823–858.

[3] N. David, T. Debiec, B. Perthame, *Convergence rate for the incompressible limit of nonlinear diffusion-advection equations*, Ann. Inst. H. Poincaré (C) Anal. Non Linéaire, **40(3)** (2022) 511–529
[4] B. Maury, A. Roudneff-Chupin, F. Santambrogio, *A Macroscopic Crowd Motion Model of Gradient Flow Type*, Math. Models Methods Appl. Sci, **20(10)** (2010), 1787–1821.
[5] B. Perthame, F. Quirós, J.L. Vázquez, *The Hele-Shaw asymptotics for mechanical models of tumor growth*,Archive for Rational Mechanics and Analysis, **212(1)**, (2014) 93–127.

Solution of the initial value problem for the gravitational Vlasov-Poisson system in an optimal transport style

Yann Brenier

In a joint work with Ivan Moyano [10], we have introduced an optimal transport (OT) strategy to solve the initial value problem (IVP) for the (gravitational) Vlasov-Poisson (VP) system:

$$\partial_t f(t,x,\xi)+\nabla_x\cdot(\xi f(t,x,\xi))+\nabla_\xi\cdot(E(t,x)f(t,x,\xi))=0,\quad (t,x,\xi)\in[0,T]\times\mathbb{T}^d\times\mathbb{R}^d,$$

$$E=-\nabla\psi(t,x)\in\mathbb{R}^d,\quad \epsilon\nabla\cdot E(t,x)=1-\int_{\xi\in\mathbb{R}^d} f(t,x,\xi)$$

with initial condition $f_0\geq 0$ at $t=0$, normalized so that $\int_{x,\xi} f_0(x,\xi)=1$.

We rely on the following ingredients: 1) a multi-phase formulation of the VP system (as was done by E. Grenier and Y.B in the 90's [11], [6], [7] and renewed by A. Baradat in the recent years [2]);

2) a reformulation as a convex optimization problem in OT "Benamou-Brenier" style [3], which is a priori not suitable for the IVP, going back to G. Loeper [12] in the monophasic case and Y.B. [6], [7] in the case $\epsilon=0$, and closely related to the theory of variational mean-field games (as in J.-D. Benamou, G. Carlier [4], J.-D. Benamou, G. Carlier F. Santambrogio [5]);

3) the introduction of an augmented Lagrangian to input the initial boundary conditions, which, unfortunately, no longer corresponds to a convex optimisation problem, and the use of weak duality to restaure convexity, as was done by Y.B. [8] in 2018 for the incompressible Euler equations. See also closely related recent contributions by D. Vorotnikov [13], Y.B. (for Einstein's equations in vacuum) [9], A. Acharya and al. [1].

Step I. We first write the Vlasov-Poisson written as a "multi-phase" (pressureless, potential) Euler-Poisson system

$$\partial_t c+\nabla\cdot q=0,\quad c=c(t,x,a)\geq 0,\quad q=q(t,x,a)\in\mathbb{R}^d,\quad t\in[0,T],\quad x\in\mathbb{T}^d,\quad a\in\mathcal{A}$$

(a is the phase label, $\mathcal{A}$ may be discrete or continuous),

$$q=cv,\quad v=v(t,x,a)=\nabla\theta(t,x,a)\in\mathbb{R}^d,$$

$$\partial_t(cv)+\nabla\cdot(cv\otimes v)=cE,$$

$$E=-\nabla\psi(t,x)\in\mathbb{R}^d,\quad 1-\epsilon\nabla\cdot E(t,x)=\int_{a\in\mathcal{A}} c(t,x,a),$$

so that $f(t,x,\xi)=\int_a c(t,x,a)\delta(\xi-v(t,x,a))$ (formally) solves the VP system.

Notice that the limit $\epsilon = 0$ corresponds to the *relaxed* incompressible Euler equations (as in [7]).

Step 2. We observe that the multi-phase equations are just the optimality conditions for the *convex* OT problem, in Benamou-Brenier style (as done by Y.B. 97, 99, for $\epsilon = 0$, and G. Loeper 06 for $\epsilon > 0$ and just one phase):

$$\inf_{c,q,E} \int_{(t,x,a)\in[0,T]\times\mathbb{T}^d\times\mathcal{A}} \frac{|q|^2}{2c} + \int_{t,x} \frac{\epsilon|E|^2}{2},$$

subject to

$$\partial_t c + \nabla\cdot q = 0, \quad c = c(t,x,a) \geq 0, \quad q = q(t,x,a) \in \mathbb{R}^d,$$

$$1 - \epsilon\nabla\cdot E(t,x) = \int_a c(t,x,a), \quad c \text{ being prescribed at t} = 0, \mathrm{T}.$$

Equivalently, we have the saddle-point formulation:

$$\inf_{c,q,E} \sup_{\theta,\psi} BT(\theta) + \int_{t,x} \psi + \epsilon E\cdot\nabla\psi + \frac{\epsilon|E|^2}{2} + \int_{t,x,a} \frac{|q|^2}{2c} - (\partial_t\theta + \psi)c - q\cdot\nabla\theta$$

where $BT(\theta) = \int_{x,a} c_T(x,a)\theta(T,x,a) - c_0(x,a)\theta(0,x,a)$. This leads to the optimality conditions: $q = c\nabla\theta$, $-\frac{|q|^2}{2c^2} - \partial_t\theta - \psi = 0$, $E = -\nabla\psi$.

Step 3. We rewrite the optimality equation

$$-\frac{|q|^2}{2c^2} - \partial_t\theta - \psi = 0 \quad \Rightarrow \quad \partial_t q + \nabla\cdot\frac{q\otimes q}{c} = cE$$

using that $q = c\nabla\theta$, $E = -\nabla\psi$, $\partial_t c + \nabla\cdot q = 0$.

The crucial idea (borrowed from [8]) is now to *add* to the Lagrangian the extra term

$$\int_{t,x,a} -\partial_t A\cdot q - \nabla A\cdot(\frac{q\otimes q}{c}) - cA\cdot E + \int_{x,a} A(T,x,a)\cdot q(T,x,a) - A(0,x,a)\cdot q(0,x,a)$$

with a new Lagrange multiplier $A(t,x,a) \in \mathbb{R}^d$. This allows us to input the new time boundary conditions:

$\theta(T,x,a) = 0$, $A(T,x,a) = 0$, $c(t,x,a)$ and $q(t,x,a)$ being prescribed at $t = 0$. So, we have obtained the new saddle-point problem

$$I(c_0,q_0) = \inf_{c,q,E} \sup_{\theta,A,\psi} NBT(\theta,A) + \int_{(t,x)\in Q} \psi + \epsilon E\cdot\nabla\psi + \frac{\epsilon|E|^2}{2}$$

$$+ \int_{(t,x,a)\in Q'} \frac{|q|^2}{2c} - (\partial_t\theta + \psi)c - q\cdot\nabla\theta - \partial_t A\cdot q - \nabla A\cdot(\frac{q\otimes q}{c}) - cA\cdot E$$

where the new boundary term reads

$$NBT(\theta,A) = -\int_{(x,a)\in D\times\mathcal{A}} c_0(x,a)\theta(0,x,a) + q_0(x,a)\cdot A(0,x,a)$$

and test functions θ and A must vanish at time $t = T$.

This "augmented Lagrangian" strategy allows us to input initial condition q_0, together with c_0, while data c_T is no longer needed.

By exchanging the sup and the inf, we obtain the "dual" problem

$$I(c_0, q_0) = \inf_{c,q,E} \sup_{\theta,A,\psi} (\cdot) \ \geq \ \sup_{\theta,A,\psi} \inf_{c,q,E} (\cdot) \ = J(c_0, q_0).$$

which is a *concave* maximization problem in (θ, A, ψ), due to the linearity of $(\cdot)$ with respect to (θ, A, ψ). Observe that we have a priori no more than the "weak" duality property $I(c_0, q_0) \ \geq \ J(c_0, q_0)$. Indeed, by augmenting the Lagrangian, we have destroyed the convex structure of the original problem because of the nonlinear term $q \otimes q/c$ and, therefore, a duality gap cannot be excluded. Anyway, we get the following result [10]:

Theorem (Y.B., I. Moyano 2022). Let $(c^s > 0, q^s, E^s)$ be a smooth solution of MP-VP on $[0, T] \times D \times \mathcal{A}$ of form

$$q^s = c^s v^s, \quad v^s(t, x, a) = \nabla \theta^s(t, x, a), \quad E^s(t, x) = -\nabla \psi^s(t, x)$$

$$\text{s.t. } (\nabla v^s + (\nabla v^s)^T)(t, x, a) < \frac{\mathbb{I}_d}{T - t}, \ \ \forall (t, x, a) \in [0, T[\times \mathbb{T}^d \times \mathcal{A}$$

and, for all fixed t, x and for all vector W in $\mathbb{R}^d$, there is a nonnegative mesure λ on $\mathcal{A}$ (depending on t, x and W) s.t. $W = \int_{a \in \mathcal{A}} v^s(t, x, a) \lambda(da)$.

(We say that v^s is weakly "absorbing").

Then there is no duality gap and the dual problem admits the solution

$$A(t, x, a) = (t - T) v^s(t, x, a), \quad \psi(t, x) = \partial_t((T - t)\psi^s(t, x)),$$

$$\theta(t, x, a) = (t - T)(\psi^s(t, x) - \frac{|v^s(t, x, a)|^2}{2})$$

so that the IVP can be solved by space-time concave maximization.

References

[1] A. Acharya, A. Sengupta, arxXiv 2306.10616.

[2] A. Baradat, *Nonlinear instability in Vlasov type equations around rough velocity profiles,* Ann. IHP, Anal. Non Linéaire 37 (2020) 489-547.

[3] J.-D. Benamou, Y. Brenier *A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem,* Numer. Math. 84 (2000) 375-393.

[4] J.-D. Benamou, G. Carlier, *Augmented Lagrangian methods for transport optimization, mean field games and degenerate elliptic equations,* JOTA 167 (2015) 1-26.

[5] J.-D. Benamou, G. Carlier, F. Santambrogio, *Variational Mean Field Games,* Active Particles, Volume 1, Springer (2017) 141-171.

[6] Y. Brenier, *A homogenized model for vortex sheets,* ARMA 138 (1997) 319-353.

[7] Y. Brenier, Brenier, Yann, *Minimal geodesics on groups of volume-preserving maps and generalized solutions of the Euler equations,* CPAM 52 (1999) 411-452.

[8] Y. Brenier, *The initial value problem for the Euler equations of incompressible fluids viewed as a concave maximization problem,* Commun. Math. Phys. 364 (2018) 579-605.

[9] Y. Brenier, *Une interprétation variationnelle de la relativité générale dans le vide en termes de transport optimal,* Comptes Rendus Acad. Sc. Math. 360 (2022) 25-33.

[10] Y. Brenier, I. Moyano, *Relaxed solutions for incompressible inviscid flows: a variational and gravitational approximation to the initial value problem,* Phil. Trans. Royal Soc. A (2022).

[11] E. Grenier, *Oscillations in quasineutral plasmas,* Commun. PDE. 21 (1996) 363-394.
[12] G. Loeper, *The reconstruction problem for the Euler-Poisson system in cosmology,* ARMA 179 (2006) 153-216.
[13] D. Vorotnikov, *Partial differential equations with quadratic nonlinearities viewed as matrix-valued optimal ballistic transport problems,* ARMA 243 (2023) 1653-1698.

Reporter: Maxime Sylvestre

Participants

Prof. Dr. Luigi Ambrosio
Scuola Normale Superiore
Piazza dei Cavalieri, 7
56126 Pisa
ITALY

Aymeric Baradat
Institut Camille Jordan
Université Claude Bernard Lyon 1
43 blvd. du 11 novembre 1918
69622 Villeurbanne Cedex
FRANCE

Dr. Jean-David Benamou
INRIA Paris
75012 Paris
FRANCE

Dr. David Bourne
Heriot-Watt University
Edinburgh EH14 4AS
UNITED KINGDOM

Prof. Dr. Yann Brenier
CNRS, Laboratoire de mathématiques
d'Orsay (LMO), Université Paris-Saclay
Bâtiment 307, rue Michel Magat
91405 Orsay
FRANCE

Dr. Maria Bruna
Centre for Mathematical Sciences
Department of Applied Mathematics and
Theoretical Physics
University of Cambridge
Wilberforce Road
CB3 0WA Cambridge
UNITED KINGDOM

Prof. Dr. Giuseppe Buttazzo
Dipartimento di Matematica
Universita di Pisa
Largo Bruno Pontecorvo, 5
56127 Pisa
ITALY

Dr. Clément Cancès
Equipe projet RAPSODI
INRIA Université de Lille
40, Avenue Halley
59650 Villeneuve d'Ascq
FRANCE

Prof. Dr. Guillaume Carlier
CEREMADE
Université Paris Dauphine
Place du Marechal de Lattre de Tassigny
75775 Paris Cedex 16
FRANCE

Lenaic Chizat
Mathematics institute
EPFL Lausanne
Station 8
1015 Lausanne
SWITZERLAND

Prof. Dr. Maria Colombo
Mathematics institute
EPFL Lausanne
Station 8
Lausanne 1015
SWITZERLAND

Prof. Dr. Codina Cotar
Department of Statistics
University College London
Gower Street
London WC1E 6BT
UNITED KINGDOM

Dr. Noemi David
Institut Camille Jordan, Université
Claude Bernard Lyon 1
43 bd du 11 novembre 1918
69622 Villeurbanne Cedex
FRANCE

Prof. Dr. Simone Di Marino
Università di Genova
Via Dodecaneso, 35
15146 Genova
ITALY

Annette Dumas
Institut Camille Jordan
Université Claude Bernard Lyon 1
43 blvd. du 11 novembre 1918
69622 Villeurbanne Cedex
FRANCE

Dr. Geneviève Dusson
Faculté des Sciences et Techniques
Laboratoire Mathématiques de Besancon
Université de Franche-Comte
16, route de Gray
25030 Besançon Cedex
FRANCE

Prof. Dr. Virginie Ehrlacher
CERMICS - ENPC
Bât. Coriolis B 312
Cité Descartes, Champs-sur-Marne
6 et 8 Avenue Blaise Pascal
77455 Marne-la-Vallée Cedex 2
FRANCE

Dr. Katharina Eichinger
Centre de Mathématiques Appliqées
(CMAP)
Ecole Polytechnique
Route de Saclay
91128 Palaiseau
FRANCE

Prof. Dr. Matthias Erbar
Fakultät für Mathematik
Universität Bielefeld
Postfach 100131
33501 Bielefeld
GERMANY

Dr. Max Fathi
Laboratoire Jacques-Louis Lions
Batiment Sophie Germain
8, Place Aurélie Nemours
75013 Paris Cedex
FRANCE

Prof. Dr. Gero Friesecke
Fakultät für Mathematik
Technische Universität München
Boltzmannstraße 3
85748 Garching bei München
GERMANY

Dr. Thomas Gallouët
Laboratoire de Mathématiques
Université Paris Sud (Paris XI)
Batiment 425
91405 Orsay Cedex
FRANCE

Dr. Nathael Gozlan
Université Paris Cité
MAP 5
45 rue des Saints Pères
P.O. Box 75006
Paris Cedex
FRANCE

Prof. Dr. Inwon C. Kim
Department of Mathematics
University of California at Los Angeles
405 Hilgard Avenue
Los Angeles, CA 90095-1555
UNITED STATES

Dr. Vaios Laschos
Weierstraß-Institut für
Angewandte Analysis und Stochastik
Mohrenstraße 39
10117 Berlin
GERMANY

Dr. Hugo Lavenant
Department of Decision Sciences
Universita Bocconi
Via Sarfatti 25
20100 Milano
ITALY

Hugo Malamut
Paris Dauphine
Place du Marechal DeLattre de Tassigny
75016 Paris
FRANCE

Prof. Dr. Daniel Matthes
Department of Mathematics
School of Computation, Information and
Technology (SoCIT)
Technische Universität München
Boltzmannstraße 3
85747 Garching bei München
GERMANY

Prof. Dr. Robert J. McCann
Department of Mathematics
University of Toronto
40 St George Street
Toronto ON M5S 2E4
CANADA

Quentin Mérigot
Université Paris-Saclay
Laboratoire de Mathématiques d'Orsay
Batiment 307
Rue Michel Magat
91405 Orsay
FRANCE

Prof. Dr. Alexander Mielke
Weierstraß-Institut für
Angewandte Analysis und Stochastik
Mohrenstraße 39
10117 Berlin
GERMANY

Dr. Andrea Natale
Inria, Université de Lille
59000 Lille Cedex
FRANCE

Dr. Luca Nenna
Laboratoire de Mathématiques
Université Paris Sud (Paris XI)
Batiment 425
91405 Orsay Cedex
FRANCE

Prof. Dr. Felix Otto
Max-Planck-Institut für Mathematik
in den Naturwissenschaften
Inselstraße 22 - 26
04103 Leipzig
GERMANY

Prof. Dr. Brendan Pass
Department of Mathematical Sciences
University of Alberta
Edmonton AB T6G 2R3
CANADA

Dr. Paul Pegon
CEREMADE
Université Paris Dauphine
Place du Marechal de Lattre de
Tassigny
75775 Paris Cedex 16
FRANCE

Dr. Emanuela Radici
Dipartimento di Matematica
Universita dell'Aquila
via Vetoio - Loc. Coppito
67100 L'Aquila
ITALY

Prof. Dr. Filippo Santambrogio
Institut Camille Jordan
Université Claude Bernard Lyon 1
43 blvd. du 11 novembre 1918
69622 Villeurbanne Cedex
FRANCE

Prof. Giuseppe Savaré
Department of Decision Sciences
Bocconi University
Via Roentgen, 1
20136 Milano
ITALY

Prof. Dr. André Schlichting
Institut für Analysis und Numerik
Mathematics Münster
Orléans-Ring 10
48149 Münster
GERMANY

JProf. Dr. Markus Schmidtchen
Fakultät Mathematik
Technische Universität Dresden
Zellescher Weg 25
01069 Dresden
GERMANY

Lauro Silini
Departement Mathematik
ETH-Zentrum
Rämistr. 101
8092 Zürich
SWITZERLAND

Prof. Dr. Dejan Slepčev
Department of Mathematical Sciences
Carnegie Mellon University
5000 Forbes Avenue
Pittsburgh, PA 15213-3890
UNITED STATES

Maxime Sylvestre
Université Paris Dauphine
Place du Marechal DeLattre de Tassigny
75016 Paris
FRANCE

Dr. Gabriele Todeschi
CERMICS - ENPC
Cite Descartes, Champs-sur-Marne
6 et 8 avenue Blaise Pascal
77455 Marne-la-Vallée Cedex 2
FRANCE

Anton Ullrich
Max-Planck-Institut für Mathematik
in den Naturwissenschaften
Inselstr. 22 - 26
04103 Leipzig
GERMANY

Prof. Dr. Francois-Xavier Vialard
Laboratoire d'Informatique Gaspard
Monge
Université Gustave Eiffel
UMR 8049
77454 Marne-la-Vallée
FRANCE

Prof. Dr. Benedikt Wirth
Fachbereich Mathematik und Informatik
Universität Münster
Einsteinstraße 62
48149 Münster
GERMANY

Prof. Dr. Yao Yao
Department of Mathematics
National University of Singapore
S17, 10 Lower Kent Ridge Road
Singapore 119076
SINGAPORE

Prof. Dr. Johannes Zimmer
Fakultät für Mathematik (M6)
School of Computation, Information and
Technology
TU München
Boltzmannstr. 3
85748 Garching bei München
GERMANY