Mathematisches Forschungsinstitut Oberwolfach

Report No. 57/2023

DOI: 10.4171/OWR/2023/57

Variational Methods for Evolution

Organized by
Franca Hoffmann, Pasadena
Alexander Mielke, Berlin
Mark Peletier, Eindhoven
Dejan Slepčev, Pittsburgh

5 – 8 December, 2023.

Abstract. Variational principles for evolutionary systems arise in many settings, both in those describing the physical world and in man-made algorithms for data science and optimization tasks. Variational principles are available for Hamiltonian systems in classical mechanics, gradient flows for dissipative systems, as well as in time-incremental minimization techniques for more general evolutionary problems. Additional challenges arise via the interplay of two or more functionals (e.g. a free energy and a dissipation potential), thus encompassing a large variety of applications in the modeling of materials and fluids, in biology, and in multi-agent systems.

Variational principles and associated evolutions are also at the core of the modern approaches to machine learning tasks, since many of them are posed as minimizing an objective functional that models the problem. The discrete and random nature of these problems and the need for accurate computation in high dimension present a set of challenges that require new mathematical insights. Variational methods for evolution allow for the usage of the rich toolbox provided by the calculus of variations, metric-space geometry, partial differential equations, and other branches of applied analysis.

The variational methods for evolution have seen a rapid growth over the last two decades. This workshop continued the successful line of meetings (2011, 2014, 2017, and 2020), while evolving and branching into new directions. We have brought together a wide scope of mathematical researchers from calculus of variations, partial differential equations, numerical analysis, and stochastics, as well as researchers from data science and machine learning, to exchange ideas, foster interaction, develop new avenues, and generally bring these communities closer together.

Mathematics Subject Classification (2020): 35-XX, 37Nxx, 49-XX, 62Gxx, 68T07, 82Dxx.

Introduction by the Organizers

The workshop *Variational Methods for Evolution*, organized by Franca Hoffmann (Caltech), Alexander Mielke (Berlin), Mark Peletier (Eindhoven), and Dejan Slepčev (Pittsburgh) brought together researchers with variety of backgrounds and from a geographically diverse set of academic institutions.

Hamiltonian systems, gradient systems, or mixtures of these two extreme types are almost ubiquitous in applications. They have been considered in connection with many real-world models such as fluid dynamics, phase transitions, thin films, quantum models, nonlinear diffusion and transport problems, chemical reactions, rate-independent phenomena, material modeling, and many others. Variational approaches to such evolutionary systems provide a powerful set of tools and methods, and the past years have seen impressive growth of this area, with the development of generalized gradient flows in Banach spaces and gradient flows in metric spaces, the characterisation of a very wide range of systems as variational evolutions, the study of the interplay between energy landscape and the dissipation geometry, the connections to stochastic particle systems, and many others.

These variational-evolution methods have recently found new applications in the rapidly developing field of data science. Many of the models of data science are variational in nature: to formulate a machine learning task one often creates an objective functional that describes the desired properties of the solution sought and then minimizes the functional. Many of these involve minimization over the probability measures and function spaces, whose minimization is closely connected to variational evolutions of the relevant functionals. The discrete nature, randomness, and high-dimensionality of the data create challenges that call for new mathematical approaches.

For instance, one task is to utilize the geometry of the data distribution carried by the available random samples. This leads to questions about evolutions on graphs and their many nodes limits. The desire for high-dimensional computations leads to questions about geometries for gradient flows that can be estimated accurately in high dimensions, and are robust to noise. Mean-field limits of neural networks (including deep ones) show promising connections to PDE and evolutionary problems. Likewise, sampling problems and generative models of learning have evolutionary descriptions that raise important questions.

In this workshop we sought to bring together mathematicians studying variational evolutions with researchers from the data science community for a stimulating exchange of ideas. We invited a selected group of experts and young researchers from both communities to work together to recognize the common mathematical structures, formulate the most important mathematical questions, and exchange ideas. Many participants said towards the end of the workshop that they had found the mix of topics particularly motivating; it is clear that this aim of bringing people together from different areas of mathematics was successful in creating a productive scientific meeting.

Another aim of the workshop was to offer the chance to many young and talented researchers that have started in this promising area, to get exposed to broad set of relevant ideas and have scientific discussions with the leaders in the field. Again, this seems to have been successful, based on the observation that a number of young researchers are now in contact with more established members of this community, and various plans for follow-up visits and research activities have already been made.

The workshop was purely on-site and in-person, despite two disruptive events: (i) a large snow storm in South East Germany on the Saturday evening before complicated the arrival of several participants and (ii) a strike involving the Deutsche Bahn on Friday forced many participants to invest time in rearranging their departure. Thanks to collective efforts alternative transport was arranged and most participants were able to stay on Friday. Despite the adversity, the workshop had an excellent atmosphere, featured exciting talks and lively scientific discussions; the participants were uniformly positive about the event when they left.

Acknowledgement: The MFO and the workshop organizers would like to thank the National Science Foundation for supporting the participation of junior researchers in the workshop by the grant DMS-2230648, "US Junior Oberwolfach Fellows". Moreover, the MFO and the workshop organizers would like to thank the Simons Foundation for supporting Bharat K. Sriperumbudur in the "Simons Visiting Professors" program at the MFO.

Workshop: Variational Methods for Evolution

Table of Contents

Abstracts

Entropic interpolations are geodesics

Christian Léonard
(joint work with Marc Arnaudon, Giovanni Conforti)

Entropic interpolation. Let $\Omega = \{\text{paths}\}$ be the set of all continuous paths from the time interval $[0,T]$ to the state space $\mathbb{R}^n$ and denote $\mathrm{P}(\Omega)$ and $\mathrm{M}(\Omega)$ the sets of all probability measures and all positive measures on Ω. We choose as a reference path measure $R \in \mathrm{M}(\Omega)$ the law of the solution of the following stochastic differential equation on $\mathbb{R}^n$:

$$\begin{cases} dX_t = -a\nabla U(X_t)/2\, dt + \sqrt{a} dB_t, & 0 < t \le T, \\ X_0 \sim m := \exp(-U)\,\mathrm{Leb}, & t = 0, \end{cases}$$

where X_t is the random position at time t, $a > 0$ is a positive number, B is a standard Brownian motion, $U : \mathbb{R}^n \to \mathbb{R}$ is a scalar potential and Leb stands for Lebesgue measure. Not only m is an invariant measure, but also R is reversible: X and $X^* : s \in [0,T] \mapsto X^*_s := X_{T-s}$, are statistically indistinguishable.

The relative entropy of $P \in \mathrm{P}(\Omega)$ with respect to R is

$$H(P|R) := \int_\Omega \log\left(\frac{dP}{dR}\right) dP$$

and the Schrödinger problem is

$$\inf H(P|R); \quad P \in \mathrm{P}(\Omega) : P_0 = \alpha, P_T = \beta$$

where $P_0, P_T \in \mathrm{P}(\mathbb{R}^n)$ are the initial and final marginals of P and $\alpha, \beta \in \mathrm{P}(\mathbb{R}^n)$ are prescribed. As a strictly convex problem, it admits a unique solution Q (if any) which is called the *Schrödinger bridge* between α and β and whose time marginal flow is called the *entropic interpolation* between α and β. This problem was addressed by Schrödinger in 1931 [9,10]. For a review see [7] for instance, and for its applications to computational optimal transport see [8].

Bridges. Any Schrödinger bridge Q is a mixture of bridges $R^{ab}(\cdot) := \mathbb{P}(X \in \cdot \mid X_0 = a, X_T = b)$ of R, that is: $Q(\cdot) = \int_{\mathbb{R}^n \times \mathbb{R}^n} R^{ab}(\cdot)\, Q_{0T}(dadb)$ where Q_{0T} is the joint law under Q of the endpoint positions. One can extend the above definition of entropic interpolation to the time marginal flow of any bridge R^{ab}. The results below remain unchanged provided we restrict our attention to the open time interval $(0,T)$.

Principle of least action. Any Schrödinger bridge inherits the Markov property from R. Restricting, without loss of generality, the Schrödinger problem to Markov path measures P allows to write

$$H(P|R) = \mathcal{F}(\mu_T) - \mathcal{F}(\mu_0) + \int_{[0,T]\times\mathbb{R}^n} \frac{|v^{\mathrm{cu}}|^2 + |v^{\mathrm{os}}|^2}{2a}(t,x)\,\mu_t(dx)dt,$$

where $(\mu_t)_{0\le t\le T}$ is the time marginal flow of P, $\mathcal{F}(\gamma) = H(\gamma|m)/2$, $\gamma \in \mathrm{P}(\mathbb{R}^n)$ plays the role of a free energy functional, the current velocity field v^{cu} satisfies the continuity equation

$$\partial_t \mu_t + \nabla\cdot(v^{\mathrm{cu}}\mu) = 0,$$

and the osmotic velocity field is given by the time reversal formula

$$v^{\mathrm{os}}(t,x) = a\nabla \log \sqrt{\frac{d\mu_t}{dm}}(x).$$

The idea of the proofs of these expressions dates back to Föllmer [6]. The main difficulty in the present setting is to give sense to the above expressions while the only assumption $H(P|R) < \infty$ does not imply much regularity. This is done in [3].

One can also show that the current velocity field v^{cu} of the Schrödinger bridge is a gradient field (in some weak sense). Restricting, without loss of generality, our attention to such Markov path measures P, the above continuity equation permits us to interpret v^{cu}_t as the tangent vector $\dot\mu_t$ at μ_t of the marginal flow $(\mu_s)_{0\le s\le T}$, in the sense of the Otto-Wasserstein geometry, see [11, Ch. 15].

On the other hand, plugging the time reversal formula into the expression of $H(P|R)$, noting that as regards the Schrödinger problem $\mu_0 = \alpha$ and $\mu_T = \beta$ are prescribed, and multiplying by a, we arrive at the following least action principle

$$\inf \mathcal{A}(\mu); \quad \mu := (\mu_t)_{0\le t\le T} : \mu_0 = \alpha, \mu_T = \beta,$$

with

$$\mathcal{A}(\mu) = \int_0^T \left(\|\dot\mu_t\|^2_{\mu_t}/2 + a^2\mathcal{I}(\mu_t)\right) dt$$

where $\|\dot\mu_t\|^2_{\mu_t} = \int_{\mathbb{R}^n} |v^{\mathrm{cu}}_t|^2\, d\mu_t$ is the Otto-Wasserstein squared norm of the tangent vector $v^{\mathrm{cu}}_t = \dot\mu_t$ at μ_t, and

$$\mathcal{I}(\gamma) := \int_{\mathbb{R}^n} \frac{1}{2}\left|\nabla \log\sqrt{\frac{d\gamma}{dm}}\right|^2 d\gamma$$

is the Fisher information of $\gamma \in \mathrm{P}(\mathbb{R}^n)$ with respect to m.

Newton's equation. The action $\mathcal{A}$ is analogous to a usual classical mechanical action on a Riemannian manifold M with Lagrangian $\mathcal{L}(\dot\gamma,\gamma) = \|\dot\gamma\|^2_\gamma/2 + a^2\mathcal{I}(\gamma)$ instead of the classical Lagrangian $L(q,v) := |v|^2_q/2 - V(q)$. Since L gives rise to the Newton equation:

$$\ddot x_t = -\mathrm{grad}_{x_t} V,$$

where $\ddot{x}_t = \nabla_{\dot{x}_t}\dot{x}_t$ is the acceleration of the trajectory $t \mapsto x_t$, one can show similarly [4,5] that the entropic interpolation μ solves the Newton equation

$$\ddot{\mu}_t = a^2 \mathrm{grad}^{\mathrm{OW}}_{\mu_t}\mathcal{I}$$

with respect to the Otto-Wasserstein geometry. The main issue when extending the results of [4] to those of [5] is to overcome the lack of regularity under the weak assumption $H(P|R) < \infty$. In the special case (to keep the writing easy) where R is the reversible Brownian path measure (i.e. $U = 0$), we have

$$\mathrm{grad}^{\mathrm{OW}}_{\gamma}\mathcal{I} = \nabla Q_\gamma \quad \text{where} \quad Q_\gamma := -\frac{\Delta}{2}\sqrt{\gamma}\Big/\sqrt{\gamma}$$

is the quantum potential and γ also stands for the density of the measure γ.

Geodesic in spacetime. On the other hand, it is known since Cartan's article [2] in 1923 that any solution of Newton's equation $\ddot{x}_t = -\mathrm{grad}_{x_t}V$, is such that (t, x_t) is a geodesic in the curved spacetime $\mathbb{R} \times M$ with some Riemann curvature tensor built on the original curvature tensor of M, plus an additional curvature tensor built with the Hessian of the potential V. Similarly, it is proved in [1] that the same property holds for entropic interpolations in a curved spacetime $\mathbb{R} \times \mathrm{P}(\mathbb{R}^n)$ whose curvature tensor is the sum of a curvature coming from the Otto-Wasserstein geometry and a curvature tensor built with the Hessian of $\mathcal{I}$. In the above special case where $U = 0$,

$$\mathrm{Hess}^{\mathrm{OW}}_{\gamma}\,\mathcal{I}(\nabla\theta, \nabla\theta) = \int_{\mathbb{R}^n} \Big(\mathrm{Hess}\, Q_\gamma(\nabla\theta, \nabla\theta) + \Big|\frac{\Delta}{2}\nabla\theta + \nabla \log\sqrt{\gamma}\cdot\nabla\theta\Big|^2\Big)\, d\gamma.$$

References

[1] M. Arnaudon and C. Léonard. Brownian bridges seen as geodesics. In preparation.

[2] É. Cartan. Sur les variétés à connexion affine et la théorie de la relativité généralisée (première partie). *Ann. Sci. Éc. Norm. Supér., 3ème série*, **40**:325–412, 1923.

[3] P. Cattiaux, G. Conforti, I. Gentil, and C. Léonard. Time reversal of diffusion processes under a finite entropy condition. *Annales de l'Institut Henri Poincaré Probab. Statist.*, **59**(4):1844–1881, 2023.

[4] G. Conforti. A second order equation for Schrödinger bridges with applications to the hot gas experiment and entropic transportation cost. *Probability Theory and Related Fields*, **174**(1):1–47, 2019.

[5] G. Conforti and C. Léonard. Entropic interpolations solve some Newton equation. In preparation.

[6] H. Föllmer. Time reversal on Wiener space. In *Stochastic Processes - Mathematics and Physics*, volume 1158 of *Lecture Notes in Math.*, pages 119–129. Springer, Berlin, 1986.

[7] C. Léonard. A survey of the Schrödinger problem and some of its connections with optimal transport. *Discrete Contin. Dyn. Syst. A*, **34**(4):1533–1574, 2014.

[8] G. Peyré and M. Cuturi. Computational optimal transport. https://optimaltransport.github.io/book/ , ArXiv:1803.00567, 2018.

[9] E. Schrödinger. Über die Umkehrung der Naturgesetze. *Sitzungsberichte Preuss. Akad. Wiss. Berlin. Phys. Math.*, **144**:144–153, 1931.

[10] E. Schrödinger. Sur la théorie relativiste de l'électron et l'interprétation de la mécanique quantique. *Ann. Inst. H. Poincaré*, **2**:269–310, 1932.

[11] C. Villani. *Optimal Transport. Old and New*, volume 338 of *Grundlehren der mathematischen Wissenschaften*. Springer, 2009.

Discrete-to-continuum limits of graph-based gradient flows

Yves van Gennip

(joint work with Yoshikazu Giga, Jun Okamoto, Samuel Mercer)

We are interested in discrete-to-continuum limits of graph-based gradient flows. Such flows are of interest to image analysis, machine learning, and other graph-based problems. A well-known example is the gradient flow based on the Allen–Cahn (or Ginzburg–Landau) functional for image segmentation, as proposed by Bertozzi and Flenner [2]. This treats the image segmentation problem as a graph classification problem, to be solved by minimizing

$$\frac{1}{2}\sum_{i,j\in V}\omega_{ij}(u_i-u_j)^2+\frac{1}{\varepsilon}\sum_{i\in V}W(u_i)+\frac{1}{2}\sum_{i\in Z}\mu_i(u_i-f_i)^2$$

over real-valued functions u defined on the node set V of a graph with edge weights ω_{ij}. The double-well potential $W:\mathbb{R}\to\mathbb{R}$ has wells at 0 and 1, f contains *a priori* known labels on a subset $Z\subset V$ and ε and μ_i are parameters to be chosen. Such a minimization problem can be tackled by computing a gradient flow.

By establishing a continuum limit, *consistency* of the method in the limit $|V|\to\infty$ is shown. In the main part of this talk we discuss explicit interpolation methods on a periodic grid to establish the continuum limit for total variation flow and for one-dimensional Allen–Cahn flow. A key ingredient in the proofs of these results is the variational inequality formulation for gradient flows.

In the latter part of the talk we present ideas from ongoing research into semigroup methods for establishing convergence of gradient flows if Γ-convergence of the underlying functionals is known. These methods can be applied to flows on random geometric graphs, but (at the moment) still demand stronger convexity requirements on the functionals than the variational-inequality-based results.

The main part of this talk is based on the work in [4]. Given a Hilbert space $(H,\|\cdot\|)$ and $\lambda\in\mathbb{R}$, a function $\Phi:H\to\mathbb{R}\cup\{+\infty\}$ is called geodesically λ-convex if $\Phi(\cdot)-\frac{\lambda}{2}\|\cdot\|^2$ is convex. In this case, a gradient flow of Φ with respect to $\|\cdot\|$ is defined to be a locally absolutely continuous curve $u:(0,\infty)\to H$ which satisfies, for almost all $t>0$ and for all v in the domain of Φ, the evolution variational inequality

$$\frac{1}{2}\frac{d}{dt}\|u(t)-v\|^2+\frac{\lambda}{2}\|u(t)-v\|^2\leq\Phi(v)-\Phi(u(t)).$$

From the perspective of generalisability this formulation is interesting, because by replacing the norms $\|u(t)-v\|$ with a general distance $d(u(t),v)$, it allows us to define gradient flows on (non-normed) metric spaces as well[1]. For the purposes of this talk, it suffices to restrict ourselves to the Hilbert space setting. In this setting, the evolution variational inequality is equivalent to the, perhaps more commonly used, differential-inclusion-based gradient flow definition:

$$u'(t)\in-\partial\left(\Phi(u(t))-\frac{\lambda}{2}\|u(t)\|^2\right)-\lambda u(t).$$

[1]In which case we also need to define geodesic λ-convexity in this generalised setting.

Here ∂ denotes the subdifferential.

We wish to compare gradient flows on graphs with continuum gradient flows. The evolution variational inequality allows us to do so, once we have determined a way to embed these flows into the same space. Thus we consider a sequence of Hilbert spaces $(H_n, \|\cdot\|_n)$ indexed by a parameter $n \in \mathbb{N}$, which in our setting are going to be spaces of functions defined on the node set (of size n) of a graph. We require an embedding $i_n : H_n \to H$ and a corresponding 'projection' $p_n : H \to H_n$ such that $p_n \circ i_n$ is the identity map on H_n. We assume moreover that the maps i_n are isometries and that the maps p_n are 1-Lipschitz continuous.

Given functions $\Phi_n : H_n \to \mathbb{R} \cup \{+\infty\}$, we require each Φ_n as well as Φ to be geodesically λ-convex for some $\lambda \leq 0$ (the same λ for each function), lower semicontinuous, not identically equal to $+\infty$ and locally bounded below at some point in their domains (not necessarily the same point for each function). By Ambrosio *et al.* [1, Theorem 4.0.4], these conditions guarantee the unique existence of gradient flows of Φ_n and Φ for given initial conditions in the closure of the domains of the respective functions. We assume that these closures of the domains are equal to the whole spaces H_n and H, for Φ_n and Φ, respectively.

Writing u_n for the gradient flow of Φ_n with initial condition $u_n^0 \in H_n$ and u for the gradient flow of Φ with initial condition $u^0 \in H$, the evolution variational inequality allows us, in [4], to prove the following two theorems.

Theorem 1. *Assume the following three conditions are satisfied, for all $n \in \mathbb{N}$, all $v \in H_n$ and all $w \in H$: $\Phi(i_n v) \leq \Phi_n(v)$, $\Phi_n(p_n w) \leq \Phi(w)$, and*

$$\|v - p_n w\|^2 + \|i_n p_n w - w\|^2 = \|i_n v - w\|^2. \tag{1}$$

Then $i_n u_n$ is the gradient flow of Φ with initial condition $i_n u_n^0$ and, for all $t > 0$,

$$\|i_n u_n(t) - u(t)\|^2 \leq e^{-2\lambda t} \|i_n u_n^0 - u^0\|^2.$$

Theorem 2. *Assume that the equality in (1) is satisfied and the following hold.*

(a) *For all $w \in H$, $\limsup_{n\to\infty} \Phi_n(p_n w) \leq \Phi(w)$.*
(b) *There exist $T > 0$, $\delta > 0$, and a nonnegative function $\Psi : H \to \mathbb{R} \cup \{+\infty\}$ such that $\Psi(u(\cdot)) \in L^1(0,T)$ and, for all $w \in H$ and n large enough $\Phi_n(p_n w) \leq \Psi(w)$.*
(c) *For all $t \in [0,T]$ and n large enough, $\Phi\big(i_n u_n(t)\big) \leq \Phi_n\big(u_n(t)\big) + o(1)$.*

If $i_n u_n^0 \to u^0$, then

$$\lim_{n\to\infty} \sup_{t\in[0,T]} \|i_n u_n(t) - u(t)\| = 0.$$

We apply these theorems in a setting in which H_n is the space of real-valued functions on the node set V_n of the graph we obtain by discretising the flat d-dimensional torus by a regular grid with n^d nodes.

Theorem 1 can be applied to the total variation flow, with

$$\Phi_n(u) = \frac{1}{2} \sum_{z\in V_n} \sum_{\substack{\tilde{z}\in V_n \\ \tilde{z}\sim z}} n^{1-d} |u(z) - u(\tilde{z})|, \qquad \Phi(u) = \int_{\mathbb{T}^d} |Du|_{\ell^1}.$$

The embedding maps i_n are constructed via piecewise-constant embedding and the 'projection' maps p_n by averaging over grid cells.

We apply Theorem 2 to the one-dimensional (i.e. $d = 1$) Allen–Cahn flow with

$$\Phi_n(t) = \frac{1}{4} \sum_{z \in V_n} \sum_{\substack{\tilde{z} \in V_n \\ \tilde{z} \sim z}} n^{2-d} \big(u(z) - u(\tilde{z})\big)^2 + \sum_{z \in V_n} n^{-d} W(u(z)),$$

$$\Phi(u) = \frac{1}{2} \int_{\mathbb{T}^d} |\nabla u(x)|^2 \, dx + \int_{\mathbb{T}^d} W(u(x)) \, dx.$$

In this case the embedding operator i_n is given by linear interpolation with p_n a corresponding 'orthogonal projection' map. Since these i_n are not an isometries, Theorem 2 cannot be applied directly. Instead, the inner product structure on H_n is adapted such that the i_n become isometries and it is shown that the resulting gradient flows of Φ_n do not differ much from the original gradient flows of Φ_n, for large n.

In the final part of the talk, which is based on work with Samuel Mercer which is currently in preparation, we wish to prove convergence of gradient flows in cases where Γ-convergence of the functions Φ_n is known, in the tradition of Sandier and Serfaty's work [5]. Moreover, we wish to be able to apply this in settings such as random geometric graphs in which no regular grid is available and interpolation techniques for embedding discrete functions into continuum function spaces requires more attention. Preliminary results, based on semigroup techniques and an extension of theorem by Brezis and Pazy [3, Theorem 3.1] to sequences of Banach spaces suggest this is possible, yet potentially at the cost of requiring stronger convexity of the functionals than geodesic λ-convexity.

References

[1] L. Ambrosio, N. Gigli, 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] A. L. Bertozzi, A. Flenner, *Diffuse interface models on graphs for classification of high dimensional data*, Multiscale Modeling & Simulation **10**(3) (2012), 1090–1118.

[3] H. Brézis, A. Pazy, *Convergence and approximation of semigroups of nonlinear operators in Banach spaces*, Journal of Functional Analysis **9** (1972), 63–74.

[4] Y. Giga, Y. van Gennip, J. Okamoto, *Graph gradient flows: from discrete to continuum*, Preprint arXiv:2211.03384 (2022).

[5] E. Sandier, S. Serfaty, *Gamma-convergence of gradient flows with applications to Ginzburg–Landau*, Communications on Pure and Applied Mathematics **57** (2004), 1627–1672.

Thermodynamic limit of stochastic particle systems via EDP convergence

André Schlichting
(joint work with Chun Yin Lam)

We consider N particles on a lattice of L sites. The state of the system is given by the occupation numbers $\eta = (\eta_x)_{x\in 1,\dots,L}$ of each site, and the time evolution is modelled by a continuous-time Markov process $(\eta(t))_{t\geq 0}$ on the set $V^{N,L} := \{\eta \in \mathbb{N}_0^L : \sum_{x=1}^L \eta_x = N\}$ of configurations with a fixed number N of particles. The process is characterized by the generator

$$\mathcal{L}f(\eta) = \frac{1}{L-1}\sum_{x=1}^{L}\sum_{y=1}^{L} K(\eta_x,\eta_y)\big(f(\eta^{x,y}) - f(\eta)\big)\ , \qquad f \in C(V^{N,L})\ , \tag{1}$$

where $\eta^{x,y}$ denotes the configuration obtained from η after one particle jumps from x to y.

This is a particular model class of stochastic particle systems (SPS), which has been introduced in [3] under the name *misanthrope processes*. The process is irreducible and has a unique canonical stationary measure $\pi^{N,L}$ on $V^{N,L}$.

We consider condensation as a phase separation phenomenon in the *thermodynamic limit* $N, L \to \infty$ with $N/L \to \rho \geq 0$: If the particle density ρ exceeds a critical value ρ_c, the system phase separates into a homogeneous bulk and a condensate, where a finite fraction of particles accumulates on a vanishing volume fraction of sites. Mathematically, we say that an SPS with canonical measures $(\pi^{N,L})$ exhibits *condensation* (in the thermodynamic limit) if the single-site marginals converge narrowly

$$\pi_{L,N}[\eta_x \in .] \Rightarrow \nu^\rho\ , \quad \text{a measure on } \mathbb{N}_0 \text{ with } \sum\nolimits_{k\geq 0} k\nu^\rho(k) < \rho = \lim N/L\,.$$

The condensation threshold mass is denoted by $\rho_c \in [0,\infty]$, the largest ρ for which no condensation occurs. Condensation, i.e. $\rho_c < \infty$, in SPS of type (1) has been studied extensively (see e.g. [2] and references therein) and particular models include zero-range processes [8] with bounded kernels of the form

$$K(k,l) = u(k) = 1 + \frac{b}{k^\gamma} \quad \text{with parameters } b > 0, \gamma \in (0,1]\ , \tag{2}$$

or various models with product kernels $K(k,l) = k^\lambda(d+l^\lambda)$ for parameters $d, \lambda > 0$.

Spatially homogeneous SPS with these kernels are known to have stationary measures of product form

$$\pi^{N,L}(\eta) = \frac{1}{Z_{N,L}}\prod_{x=1}^{L} Q_{\eta_x} \quad \text{with normalization} \quad Z_{N,L} = \sum_{\eta\in V^{N,L}}\prod_{x=1}^{L} Q_{\eta_x}\ , \tag{3}$$

and stationary weights $Q : \mathbb{N}_0 \to (0,\infty)$, playing the role of the chemical potential.

The condensation transition has been established rigorously in the thermodynamic limit for zero-range processes of type (2) (see e.g. [8]), where the condensate consists only of a single site.

The configuration $\eta \in V^{N,L}$ of a mean-field SPS can be characterized by the *empirical cluster distribution*

$$F_k^L(\eta) := \frac{1}{L}\sum_{x=1}^{L} \delta_{k,\eta_x} \in [0,1] \ , \quad k \geq 0 \ . \tag{4}$$

In the thermodynamic limit these observables converge under quite general conditions, forming the basis of a mesoscopic description of the dynamics. The law of large numbers was obtained in [7]: Let the process $(\eta(t))_{t\geq 0}$ be given by the generator (1) for a kernel with at most linear growth, i.e.

$$0 \leq K(k,l-1) \leq C_K\, k\, l \qquad \text{for } k,l \geq 1 \ , \tag{K$_1$}$$

If $F_k^L(\eta(0)) \to c_k(0)$ satisfies some suitable tightness assumptions, then $c_k^L(\eta(0)) \to c_k(t)$ converges weakly in the thermodynamic limit $L, N \to \infty$, $N/L \to \rho \geq 0$ to the cluster concentrations $c_k(t)$ solving the (deterministic) *mean-field rate equations*

$$\begin{aligned} \dot{c}_k \ &= \ \sum_{l\geq 1} K(l,k-1)c_l c_{k-1} - \sum_{l\geq 1} K(k,l-1)c_k c_{l-1} \\ &\quad - \sum_{l\geq 1} K(l,k)c_l c_k + \sum_{l\geq 1} K(k+1,l-1)c_{k+1}c_{l-1} \ , \qquad \text{for } k \geq 0 \ . \end{aligned} \tag{EDG}$$

Note that the deterministic set of equations (EDG) can formally be obtained from (5) by mass-action kinetics, and describe the time evolution of concentrations of finite clusters, i.e. the bulk of the system, on a *mesoscopic scale*. This description, also known as *exchange-driven growth* [1]. Basic mathematical properties regarding the well-posedness and the longtime behavior of the EDG model in the form of (EDG) are investigated in [4, 6, 15].

Although the exchange-driven growth process is not necessarily realized by chemical kinematics, it is convenient to be interpreted as a reaction network of the form

$$\{k-1\} + \{l\} \underset{K(k,l-1)}{\overset{K(l,k-1)}{\rightleftharpoons}} \{k\} + \{l-1\} \ , \qquad \text{for} \quad k,l \geq 1 \ . \tag{5}$$

Hereby, clusters of integer size $k \geq 1$ are denoted by $\{k\}$ and the variable $\{0\}$ represents empty volume. The kernel $\big(K(k,l-1)\big)_{k,l\geq 1}$ encodes the rate of the exchange of a single monomer from a cluster of size k to a cluster of size $l-1$. Note that no mass is created or destroyed in the reaction.

In the present work, we lift the law of large numbers result to statement on the convergence of gradient structures related to the large deviation rate functional of the stochastic particle system. For the description and convergence, we use the recent framework of gradient flows in continuity equation format established in [13, 14]. The law of the empirical cluster distribution (4) of the SPS, denoted with $\mathbb{C}^{N,L} \in \mathcal{P}(\mathcal{P}(\mathbb{N}_0))$ is associated with a discrete continuity equation encoding the two conserved quantities of the system given in a suitable weak form of

$$\partial_t \mathbb{C}_t^{N,L} + \widehat{\operatorname{div}}\, \mathbb{J}_t^{N,L} = 0, \tag{6}$$

where the flux is a measure in $\mathbb{J}^{N,L} \in \mathcal{M}(\mathcal{P}(\mathbb{N}_0) \times \mathbb{N} \times \mathbb{N}_0)$ and $\widehat{\mathrm{div}}$ is the adjoint operator to the discrete gradient $\widehat{\nabla} f(c)(k,l-1) = f(c^{k,l-1}) - f(c)$ with $c^{k,l-1} = c + \frac{1}{L}\gamma^{k,l-1}$ and $\gamma^{k,l-1} = e_{k-1} + e_l - e_k - e_{l-1}$. For the specific absolutely continuous flux $\mathrm{d}\overline{\mathbb{J}}_t^{N,L}(c,k,l-1) = \frac{L^2}{L-1} c_k \big(c_{l-1} - \frac{\delta_{k,l-1}}{L}\big) K(k,l-1)\mathrm{d}\mathbb{C}_t^{N,L}(c)$, the solution of (6) is exactly the forward Kolmogorov equation for the SPS-generator (1). Under the detailed balance condition

$$\text{(BDA)} \qquad \frac{K(k,l-1)}{K(l,k-1)} = \frac{K(k,0)\,K(1,l-1)}{K(l,0)\,K(1,k-1)}$$

the system has a also the formulation as a generalized gradient flow by the theory developed in [11], which amounts to the fact, that the rate function $\mathbb{L}^{N,L}(\mathbb{C}^{N,L}, \mathbb{J}^{N,L})$ takes the form

$$\text{(7)} \quad \mathbb{F}^{N,L}(\mathbb{C}_t^{N,L})\Big|_{t=0}^{T} + \int_0^T \Big[\mathcal{R}^{N,L}\big(\mathbb{C}_t^{N,L}, \mathbb{J}_t^{N,L}\big) + \mathcal{R}^{N,L^*}\big(\mathbb{C}_t^{N,L}, -\overline{\nabla} D\mathcal{F}^{N,L}(\mathbb{C}_t^{N,L})\big)\Big]\mathrm{d}t,$$

where the free energy $\mathbb{F}$ is the relative entropy with respect to the equilibrium cluster distribution $F_\sharp^L \pi^{N,L}$ of the SPS from (3) and the functional $\mathcal{R}$ and $\mathcal{R}^*$ are dual dissipation functionals of cosh-type, which are typical for jump processes [9, 11, 13, 14]. In the form (7), a passage to the thermodynamic limit $N, L \to \infty$ of the gradient structure with $N/L \to \rho \geq 0$ is possible via the notion of *EDP-convergence*, also called *evolutionary* Γ*-convergence* [5, 10, 12, 14]. The strategy is to exploit suitable compactness for curves $(\mathbb{C}^{N,L}, \mathbb{J}^{N,L})$ solving (6) such that along converging subsequences $(\mathbb{C}^{N,L}, \mathbb{J}^{N,L}) \rightharpoonup (\mathbb{C}, \mathbb{J})$ the following Γ-lim inf statement holds

$$\liminf_{N/L \to \rho} \mathbb{L}^{N,L}(\mathbb{C}^{N,L}, \mathbb{J}^{N,L}) \geq \mathbb{L}^{\rho \wedge \rho_c}(\mathbb{C}, \mathbb{J}).$$

The limit functional has a density with respect to the limit measure $\mathbb{C}$ and one arrives at the diagram:

$$\begin{array}{ccccc} (\mathbb{C}^{N,L}, \overline{\mathbb{J}}^{N,L}) \text{ solves (6)} & \Longleftrightarrow & \mathbb{L}^{N,L}(\mathbb{C}^{N,L}, \mathbb{J}^{N,L}) = 0 & \xrightarrow[N/L \to \rho]{\text{EDP}} & \mathbb{L}^{\rho \wedge \rho_c}(\mathbb{C}, \mathbb{J}) = 0 \\ \text{Evolution of the law of} \uparrow & & & & \Big\Downarrow \text{implies via superposition} \\ \text{cluster distribution } (F_k^L(\eta(t)))_{t \geq 0} & & \mathsf{L}^{\rho \wedge \rho_c}(c, j(c)) = 0 & \Longleftrightarrow & c \text{ solves (EDG).} \end{array}$$

In the EDP convergence statement the choice of the topology is crucial and we equip the space $\mathcal{P}_{<\infty}(\mathbb{N}_0)$ with the distance

$$\text{(8)} \quad d_{\mathrm{Ex}}(\mu^0, \mu^1) = |\mathrm{T}(\mu^0) - \mathrm{T}(\mu^1)|_{\ell^1(\mathbb{N})} \quad \text{with} \quad \mathrm{T}_k(\mu) = \sum_{l=k}^{\infty} \mu_l \quad \textit{tail distribution.}$$

The EDP convergence statement is then formulated in the topology $(\mathcal{P}_{<\infty}, d_{\mathrm{Ex}})$ and for the sake of brevity, we state only the Γ-convergence for the free energy.

Theorem (Γ-convergence of free energy)**.** *In the thermodynamic limit* $\frac{N}{L} \to \rho$*, the free energy* Γ *converges*

$$\mathbb{F}^{N,L}(\mathbb{C}^{N,L}) \xrightarrow{\Gamma} \int \mathcal{H}(c|\nu^{\rho \wedge \rho_c})\mathrm{d}\mathbb{C}^{\rho} \qquad \textit{in the narrow topology on } (\mathcal{P}_{<\infty}, d_{\mathrm{Ex}})\,.$$

References

[1] E. Ben-Naim and P. L. Krapivsky. Exchange-driven growth. *Phys. Rev. E*, 68(3):031104, 2003.
[2] P. Chleboun and S. Grosskinsky. Condensation in stochastic particle systems with stationary product measures. *Journal of Statistical Physics*, 154(1):432–465, 2014.
[3] C. Cocozza-Thivent. Processus des misanthropes. *Z Wahrscheinlichkeit*, 70(4):509–523, 1985.
[4] C. Eichenberg and A. Schlichting. Self-similar behavior of the exchange-driven growth model with product kernel. *Comm. Partial Differential Equations*, 46(3):498–546, 2021.
[5] M. Erbar, M. Fathi, V. Laschos, and A. Schlichting. Gradient flow structure for McKean-Vlasov equations on discrete spaces. *Discrete Contin. Dyn. Syst.*, 36(12):6799–6833, 2016.
[6] E. Esenturk. Mathematical theory of exchange-driven growth. *Nonlinearity*, 31(7):3460, jun 2018.
[7] S. Grosskinsky and W. Jatuviriyapornchai. Derivation of mean-field equations for stochastic particle systems. *Stochastic Processes and their Applications*, 129(4):1455–1475, Apr. 2019.
[8] S. Großkinsky, G. M. Schütz, and H. Spohn. Condensation in the zero range process: stationary and dynamical properties. *Journal of statistical physics*, 113:389–410, 2003.
[9] J. Hoeksema and O. Tse. Generalized gradient structures for measure-valued population dynamics and their large-population limit. *Calculus of Variations and Partial Differential Equations*, 62(5), May 2023.
[10] M. Liero, A. Mielke, M. A. Peletier, and D. R. M. Renger. On microscopic origins of generalized gradient structures. *Discrete and Continuous Dynamical Systems - S*, 10(1), 2017.
[11] A. Mielke, M. A. Peletier, and D. M. Renger. On the relation between gradient flows and the large-deviation principle, with applications to markov chains and diffusion. *Potential Analysis*, 41:1293–1327, 2014.
[12] A. Mielke. On evolutionary Γ-Convergence for gradient systems. In A. Muntean, J. Rademacher, and A. Zagaris, editors, *Macroscopic and large scale phenomena: Coarse graining, mean field limits and ergodicity*, pages 187–249. Springer International Publishing, Cham, 2016.
[13] M. A. Peletier, R. Rossi, G. Savaré, and O. Tse. Jump processes as generalized gradient flows. *Calculus of Variations and Partial Differential Equations*, 61(1):1–85, 2022.
[14] M. A. Peletier and A. Schlichting. Cosh gradient systems and tilting. *Nonlinear Analysis*, 231:113094, 2023.
[15] A. Schlichting. The exchange-driven growth model: basic properties and longtime behavior. *J. Nonlinear Sci.*, 30(3):793–830, 2020.

Normal form and the Cauchy problem for cross-diffusive mixtures

Katharina Hopf

(joint work with Pierre-Étienne Druet, Ansgar Jüngel)

Irreversible physical processes compatible with the second law of thermodynamics can be modelled using the Onsager approach, which is based on a formal gradient-flow ansatz in the dual form

$$\dot{u} = -\mathcal{K}(u)D\mathcal{H}(u). \tag{1}$$

Here, $u = u(t)$ denotes the state, $\dot{u} = \frac{\mathrm{d}}{\mathrm{d}t}u$, $\mathcal{H}$ a differentiable driving functional and $\mathcal{K}$ the Onsager operator, a symmetric and positive semi-definite linear operator, whose symmetry property reflects the Onsager reciprocal relations. We are interested in diffusive processes formally obtained by choosing $u = u(t, x) \in \mathcal{O}$,

$t > 0, x \in \mathbb{T}^d$, for a convex domain $\mathcal{O} \subset \mathbb{R}^n$, $\mathcal{H}(u) = \int_{\mathbb{T}^d} h(u)\,dx$ with $h : \mathcal{O} \to \mathbb{R}$ smooth and strongly convex such that $\mathbb{H} := D^2 h > 0$ in $\mathcal{O}$. The Onsager operator is assumed to take the form $\mathcal{K}(u)\xi = -\nabla \cdot (\mathbb{M}(u)\nabla\xi)$ with $\mathbb{M}(u) \in \mathbb{R}^{n\times n}$ symmetric and positive semidefinite. Inserting these choices into (1) gives the quasi-linear second-order system

$$\partial_t u = \nabla \cdot (\mathbb{A}(u)\nabla u), \qquad \mathbb{A}(u) = \mathbb{M}(u)\mathbb{H}(u). \tag{2}$$

While the matrix $\mathbb{A}(u)$ need not be symmetric, the positive definiteness of $\mathbb{H}(u)$ and the positive semi-definiteness of $\mathbb{M}(u)$ ensure that it is diagonalisable over $\mathbb{R}$ and all its eigenvalues are non-negative. If in addition $\operatorname{rank}\mathbb{A}(u) = n$, the PDE system (2) is parabolic in the sense of Petrovskii, rendering the Cauchy problem locally well-posed for sufficiently regular data.

The present note is motivated by an application in population dynamics determined by the choice

$$h(u) = \sum_{i=1}^{n} \frac{1}{\lambda_i} u_i(\log u_i - 1), \qquad \mathbb{M}_{ij}(u) = u_i \mathbb{B}_{ij} \lambda_j u_j, \quad i,j \in \{1,\dots,n\}, \tag{3}$$

with $\mathcal{O} = (0,\infty)^n$ and where $\mathbb{B} = (\mathbb{B}_{ij}) \in \mathbb{R}^{n\times n}$, $\lambda = (\lambda_i) \in (0,\infty)^n$ are such that the product $\mathbb{B}\mathbb{D}(\lambda)$ is symmetric positive semidefinite, $\mathbb{D}(\lambda) := \operatorname{diag}(\lambda_1,\dots,\lambda_n)$. Thus, in this application, $\mathbb{A}_{ij}(u) = u_i\mathbb{B}_{ij}$, and the system is no longer parabolic if $\operatorname{rank}\mathbb{B} < n$. We are thus faced with a borderline case, where local well-posedness cannot directly be inferred from classical literature, but might still be expected given the non-negativity of all eigenvalues of $\mathbb{A}(u)$. To gain insights in the Cauchy problem, it is necessary to better understand the structure of the system. In the context of fluid dynamics a systematic procedure has been developed by Kawashima and Shizuta [2] for quasi-linear second-order systems with an entropy structure, who introduced a normal form, i.e. a change of the dependent variables that brings the PDE system in the form of a composite symmetric hyperbolic–parabolic system. The classical theory on normal forms strongly relies on a null-space/range invariance property of the matrix associated with the diffusive effects, which is not satisfied in the above model (with $\operatorname{rank}\mathbb{B} < n$) because $\operatorname{range}\mathbb{M}(u) = \operatorname{range}\mathbb{D}(u)\mathbb{B}$ depends on the state u. Nevertheless, in the specific example considered above, explicit calculations detailed in [1] allow us to identify a change of variables $u \mapsto w$ that brings system (2) in the form of a symmetric hyperbolic–parabolic system

$$\begin{aligned}
\mathbb{A}_0^{\mathrm{I}}(w)\partial_t w_{\mathrm{I}} + \sum_{\nu=1}^{d} \mathbb{A}_1^{\mathrm{I}}(w, \partial_{x_\nu} w_{\mathrm{II}})\,\partial_{x_\nu} w_{\mathrm{I}} &= f^{\mathrm{I}}(w, \nabla w_{\mathrm{II}}),\\
\mathbb{A}_0^{\mathrm{II}}\partial_t w_{\mathrm{II}} - \nabla \cdot \Big(\mathbb{A}_*^{\mathrm{II}}(w)\nabla w_{\mathrm{II}}\Big) &= 0,
\end{aligned}$$

where the matrices $\mathbb{A}_0^{\mathrm{I}}(w) \in \mathbb{R}^{(n-r)\times(n-r)}$, $\mathbb{A}_0^{\mathrm{II}}, \mathbb{A}_*^{\mathrm{II}}(w) \in \mathbb{R}^{r\times r}$, $r := \operatorname{rank}\mathbb{B}$, are symmetric positive definite, and $\mathbb{A}_1^{\mathrm{I}}(w, \partial_{x_\nu} w_{\mathrm{II}}) \in \mathbb{R}^{(n-r)\times(n-r)}$ is symmetric. At this point, established methods for symmetric hyperbolic and symmetric parabolic systems can be applied separately to the respective subsystem in order to construct

short-time classical solutions emanating from initial data in $H^s(\mathbb{T}^d)$, $s > \frac{d}{2} + 1$, that are positive componentwise.

Finally, consider more generally system (2) with $\operatorname{rank} \mathbb{A}(u) = r < n$. The following question arises naturally: under which conditions can it be recast in a normal form that ensures local well-posedness for smooth data?

References

[1] P.-É. Druet, K. Hopf, and A. Jüngel, *Hyperbolic-parabolic normal form and local classical solutions for cross-diffusion systems with incomplete diffusion*, Comm. Partial Differential Equations **48** (2023), 863–894.

[2] S. Kawashima and Y. Shizuta, *On the normal form of the symmetric hyperbolic-parabolic systems associated with the conservation laws*, Tohoku Math. J. **40** (1988), 449–464.

The large-data limit of the MBO scheme for data clustering

Tim Laux

(joint work with Jona Lelmi)

The MBO scheme is an efficient algorithm for data clustering, the task of partitioning a given dataset into several meaningful clusters. Vaguely speaking, a clustering is considered meaningful if all elements in a given cluster are similar to each other while they differ from those in others. Quantitatively, this is often interpreted as finding minimal cuts in an associated graph. However, nonlinear methods (like finding minimal graph cuts) have the disadvantage of being computationally inefficient, sometimes even giving rise to NP-hard problems. On the other hand, there are plenty linear algorithms, such as k-Means, which find some clustering, but cannot resolve the possibly nonlinear structure of the data set without suitable pre-processing of the data. The MBO scheme mediates between those two extreme cases: One merely solves a linear problem and then applies a pointwise nonlinearity which is computationally trivial. Therefore, it is as performant as a linear method but is not blind to nonlinear effects in the data structure. In this talk, I present the first rigorous analysis of this scheme in the large-data limit.

Given a point cloud $X = \{x_1, \dots, x_N\} \subset \mathbb{R}^d$, we encode a clustering by a partition $X = \Omega_1 \cup \dots \cup \Omega_P$ for some $P \in \mathbb{N}$. Equipping the set X with a graph structure (for example by setting $x \sim y$ if and only if $|x-y| < \varepsilon$ for some fixed scale $\varepsilon > 0$), one can exploit the (say, random walk) graph Laplacian Δ to understand the geometry of the data set X.

The MBO scheme improves an initial guess (for example given by k-Means or a random assignment) by alternating between linear diffusion and pointwise thresholding. More precisely, given an (artificial) time-step size and an initial clustering $\Omega_1^{(0)} \cup \dots \cup \Omega_P^{(0)}$, for $\ell = 1, 2, \dots$, compute

(1) Diffusion: $u_i^{(\ell)} := e^{-h\Delta} \chi_{\Omega_i^{(\ell-1)}} \quad (1 \le i \le P)$,

(2) Thresholding: $\Omega_i^{(\ell)} := \big\{ x \in X \colon u_i^{(\ell)}(x) = \max_{1 \le j \le P} u_j^{(\ell)}(x) \big\} \quad (1 \le i \le P)$,

until some stopping criterion is met, such as only few points changing their label from Step $L-1$ to Step L. The partition in the last step then is the proposed clustering $X = \Omega_1^{(L)} \cup \ldots \cup \ldots \Omega_P^{(L)}$.

The starting point of the analysis is that each iteration of the MBO scheme can be viewed as one step of minimizing movements for the thresholding energy on the similarity graph of the dataset, i.e., writing $\Omega = (\Omega_1, \ldots, \Omega_P)$, the combination of (1) and (2) is equivalent to

$$\Omega^{(\ell)} \in \arg\min_{\Omega} \Big\{ E_h^{N,\varepsilon}(\Omega) + \frac{1}{2h} \big(d_h^{N,\varepsilon}\big(\Omega, \Omega^{(\ell-1)}\big)\big)^2 \Big\},$$

where the energy $E_h^{N,\varepsilon}$ is defined on partitions $\Omega = (\Omega_1, \ldots, \Omega_P)$ of X via

$$E_h^{N,\varepsilon}(\Omega) := \frac{1}{\sqrt{h}} \sum_{1 \le i < j \le P} \langle \chi_{\Omega_j}, e^{-h\Delta} \chi_{\Omega_j} \rangle \tag{3}$$

with $\langle \cdot, \cdot \rangle$ a suitable scalar product on functions on X that makes the graph Laplacian $\Delta = \Delta_{N,\varepsilon}$ self-adjoint, and $d_h^{N,\varepsilon}$ is a suitable distance function on partitions.

It is then natural to think that outcomes $\Omega^{(L)}$ of the MBO scheme are (local) minimizers of this energy. In [2], we prove that for large data sets the algorithm is consistent with the original task of finding minimal cuts in the sense that these (local) minimizers converge to (local) minimizers of the optimal partition problem given by the continuum limit of the minimal cut problem.

More precisely, we employ the so-called *manifold assumption* postulating that the points $(x_n)_n$ are independent samples of some probability measure $\mu = \rho \operatorname{Vol}_M$, where (M, g) is a closed k-dimensional submanifold of the high-dimensional feature space $\mathbb{R}^d$ and $\rho \colon M \to (0, \infty)$ is a smooth function. Then, the first result in [2] establishes the large-data limit of the energies for fixed time-step size h.

Theorem. *Under the manifold assumption, as the sample size N goes to infinity, almost surely and in a suitable scaling regime for the length scale $\varepsilon_N \to 0$, we have $E_h^{N,\varepsilon_N} \to E_h$ in the sense of Γ-convergence w.r.t. the weak TL^2-topology.*

Here, the continuum energy E_h is defined on relaxed partitions, i.e., maps $u \colon M \to [0,1]^P$ such that $\sum_i u_i = 1$, and is of the form

$$E_h(u) = \frac{1}{\sqrt{h}} \sum_{1 \le i < j \le P} \int_M u_i e^{-h\Delta_{\rho^2}} u_j \, \rho^2 d\operatorname{Vol}_M$$

(modulo some constants), where $\Delta_{\rho^2} f = -\frac{1}{\rho^2} \nabla \cdot (\rho^2 \nabla f)$ is the natural Laplacian on the weighted manifold (M, g, ρ^2).

The main ingredient for this result is the following natural fact that the diffusion equation upgrades weak to strong convergence.

Proposition. *In the situation of the theorem above, for any $t > 0$, we have*

$$u_N \rightharpoonup u \text{ weakly in } TL^2 \quad \Longrightarrow \quad e^{-t\Delta_{N,\varepsilon_N}} u_N \to e^{-t\Delta_{\rho^2}} u \text{ strongly in } TL^2. \tag{4}$$

Indeed, the proposition implies that the Γ-convergence in the theorem is in fact continuous convergence: Every weakly converging sequence is a recovery sequence

for its limit. Furthermore, the proposition even implies that the whole minimizing movements functional (3) Γ-converges and hence we answer positively a question of Bertozzi:

Theorem. *Under the assumption of the above theorem, the iterates of the MBO scheme on the graph converge to the corresponding iterates of the MBO scheme on the weighted data manifold.*

The proposition is shown in [2] via the stability principle of gradient flows and exploiting the fact that the diffusion equation on the weighted manifold (M, g, ρ^2) is well-behaved so that the chain rule holds.

Finally, in the limit of vanishing time-step size, the problem converges to the desired optimal partition problem.

Theorem. *As $h \downarrow 0$, we have $E_h \to E$ in the sense of Γ-convergence w.r.t. the L^1 topology, where the sharp-interface energy is the following weighted optimal partition energy*

$$E(u) = \sum_{1 \le i < j \le P} \int_{\partial^* \Omega_i \cap \partial^* \Omega_j} \rho^2 \, d\mathcal{H}^{k-1} \quad \text{if } u = (\chi_{\Omega_1}, \dots, \chi_{\Omega_P}) \tag{5}$$

and $E(u) = +\infty$ otherwise.

This confirms that the MBO scheme indeed places small cuts in regions of low data density. The work [2] is the first result on the large-data limit of the MBO scheme and still the only one valid for more than two clusters. In the case of two clusters, however, one can use the theory of viscosity solutions to get a more precise understanding of the dynamics, see [1]. This is a crucial next step as the non-convex energy (5) has many local minimizers. Understanding the effective behavior of the dynamics of the scheme gives insight into the path taken by the scheme in the energy landscape and therefore the selection of local minimizers.

The main ingredients for this analysis are (i) a new abstract convergence result for arbitrary discrete structures based on quantitative estimates for heat operators and (ii) the derivation of these estimates in the setting of random geometric graphs.

Overall, the results in [1] roughly state that the following.

Theorem. *Under the manifold assumption and in the joint limit $N \to \infty$, $\varepsilon \to 0$, $h \to 0$, in a suitable scaling regime, the MBO scheme for two clusters converges to the viscosity solution of mean curvature flow in the weighted data manifold (M, g, ρ^2), satisfying the level set equation*

$$\partial_t u = \frac{1}{\rho^2} \nabla \cdot \Big(\rho^2 \frac{\nabla u}{|\nabla u|} \Big) = \nabla \cdot \Big(\frac{\nabla u}{|\nabla u|} \Big) + \frac{\nabla u}{|\nabla u|} \cdot \nabla \log \rho^2.$$

Formally, this means that the limit is a solution to the geometric evolution equation

$$V = -H - \nu \cdot \nabla \log \rho^2, \tag{6}$$

which shows that the evolution is driven by both surface tension and data density. Naturally, this flow is the L^2-gradient flow of (the two-phase version of) the energy (5) which wants to straighten the cut and move it to low-density regions.

Remarkably, this proof also applies in case of a frequency cut-off, i.e., when replacing the diffusion semigroup $e^{-h\Delta}$ with the computationally much simpler projected version $e^{-h\Delta}P_{\langle\psi_1,\dots,\psi_K\rangle}$, where ψ_k denotes the k-th eigenfunction of $\Delta_{N,\varepsilon}$. The lower bound for the frequency cut-off which still guarantees convergence to (6) is of the form $K \gtrsim (\log N)^q$ for some (explicit) exponent $q > 0$.

References

[1] T. Laux and J. Lelmi, *Large data limit of the MBO scheme for data clustering: Convergence of the Dynamics*, To appear in J. Mach. Learn. Res. arXiv:2209.05837.

[2] T. Laux and J. Lelmi, *Large data limit of the MBO scheme for data clustering: Γ-convergence of the energies*, Preprint. arXiv:2112.06737.

Globally Lipschitz transport maps

Max Fathi

(joint work with Dan Mikulincer, Yair Shenfeld)

This talk presented some results of [1] on existence of globally Lipschitz transport maps between probability measures, including in the Riemannian setting, as well as some conjectures on global Lipschitz regularity for optimal transport maps. An extended abstract on these results previously appeared in [2]

References

[1] M. Fathi, D. Mikulincer and Y. Shenfeld, Transportation onto log-Lipschitz perturbations. *Arxiv preprint* 2023.

[2] M. Gordina, T. Kumagai, L. Saloff-Coste and K.-Th. Sturm, Heat Kernels, Stochastic Processes and Functional Inequalities. *Oberwolfach Rep.* 19 (2022), no. 4, pp. 2811–2893

On convergence of the fully discrete JKO scheme

Anastasiia Hraivoronska

(joint work with Filippo Santambrogio)

We study the convergence of the JKO scheme discretized on a regular lattice, motivated by application to developing numerical schemes. The JKO scheme introduced in [1] proved to be a powerful tool for analysis of evolutionary equations with gradient structure in the space of probability measures $\mathcal{P}(\Omega)$ endowed with the L^2-Wasserstein distance W_2. We recall that it is an iterative scheme that for a given energy functional $\mathcal{F} : \mathcal{P}(\Omega) \to \mathbb{R} \cup \{+\infty\}$, initial datum $\rho_0 \in \mathcal{P}(\Omega)$, and a time step $\tau > 0$ produces a sequence of probability measures $\{\rho_k^\tau\}$ as

$$\text{(JKO)} \qquad \rho_{k+1}^\tau \in \arg\min_{\rho\in\mathcal{P}(\Omega)} \Big\{\mathcal{F}(\rho) + \frac{1}{2\tau}W_2^2(\rho, \rho_k^\tau)\Big\}.$$

The theory initiated in [1,2] and further developed in [3] allows to prove under appropriate assumptions on $\mathcal{F}$ that the sequence of minimizers from (JKO) converges to a solution of

$$\partial_t \rho - \operatorname{div}\Big(\rho \nabla \frac{\delta \mathcal{F}}{\delta \rho}\Big) = 0 \qquad (0,T)\times\Omega, \tag{1}$$

with non-flux boundary condition on $\partial\Omega$. In this work, we focus on the energy functionals including internal and potential energy:

$$\mathcal{F}(\rho) = \begin{cases} \int_\Omega f\Big(\frac{d\rho}{d\mathcal{L}^d}\Big) d\mathcal{L}^d + \int_\Omega V d\rho, & \rho \ll \mathcal{L}^d, \\ +\infty, & \text{otherwise.} \end{cases} \tag{2}$$

The JKO scheme is a natural time discretization that preserves structural features of the equation such as conservation of mass and energy dissipation, as well as some properties of solutions of the corresponding PDEs. It is tempting to come up with a numerical scheme based on (JKO) that enjoys similar properties. The challenging part is dealing with the Wasserstein distance term. Existing approaches to this problem include the methods exploiting the Benamou-Brenier dynamic formulation [4], using entropic regularisation and Sinkhorn algorithm [5, 6], and semi-discrete approaches [7]. We explore a new approach based on discretizing the JKO scheme on a regular lattice.

Let $\Omega \subset \mathbb{R}^d$ be a bounded domain discretized with $\mathcal{T}^h := h\mathbb{Z}^d \cap \Omega$. We notice that any probability measure on $\mathcal{T}^h$ can be represented as

$$\rho^h = \sum_{x\in\mathcal{T}^h} \rho^h_x \delta_x, \qquad \sum_{x\in\mathcal{T}^h} \rho^h_x = 1.$$

We call *the fully discrete JKO scheme*, the iterative scheme

$$\rho^{h,\tau}_{k+1} \in \arg\min_{\rho\in\mathcal{P}(\mathcal{T}^h)} \Big\{ \mathcal{F}_h(\rho^h) + \frac{1}{2\tau} W_2^2(\rho^h, \rho^{h,\tau}_k) \Big\}. \tag{JKO$_h$}$$

The problem we want to address is convergence of the sequence of minimizers in (JKO$_h$) to a solution of (1) in a joint limit $h \to 0$ and $\tau \to 0$. The first question is what is an appropriate relation between h and τ. We illustrate the importance of this relation for convergence on a toy example with the potential energy.

Example (Movement driven by a potential)**.** Let the energy functional include only potential energy with $V \in C^{1,1}(\mathbb{R}^d)$:

$$\mathcal{F}_h(\rho^h) = \sum_{x\in\mathcal{T}^h} V(x)\rho^h_x.$$

In this case, it is reasonable to consider separately the movement of the Dirac masses $\rho^h_0(x)\delta_x$ for $x \in \operatorname{spt}(\rho^h_0)$, because they move independently in absence of diffusion. Consider the movement of δ_{x_0}, $x_0 \in \mathcal{T}^h$. If x_1 is the minimizer of $V(x) + |x - x_0|^2/2\tau$ restricted to $\mathcal{T}^h$, then

$$V(x_0) \geq V(x_1) + \frac{|x_1 - x_0|^2}{2\tau},$$

which implies $h/\tau \le 2\|\nabla V\|_{\mathrm{Lip}}$. We see that if asymptotically $h/\tau > 2\|\nabla V\|_{\mathrm{Lip}}$, then we cannot expect convergence to the continuous solution, because every subsequent minimizer is equal to x_0 and the discrete evolution is "frozen".

Moreover, one can derive that the accumulated error between minimizers $\{x_k\}$ restricted to $\mathcal{T}^h$ and minimizers on the full space $\{\overline{x}_k\}$ for $T = k\tau$ is bounded as

$$|x_k - \overline{x}_k| \le Ckh = CT\frac{h}{\tau}.$$

Therefore, the convergence holds only if $h/\tau \to 0$.

Now we turn to a more interesting case of (JKO$_h$) with the internal energy

$$\mathcal{F}_h(\rho^h) = \sum_{x\in\mathcal{T}^h} f\Big(\frac{\rho^h_x}{h^d}\Big)h^d,$$

with convex and differentiable f such that f' is monotone. Let $\{\rho_k^{h,\tau}\}_{k=0,\dots,N}$ be a sequence of minimizers of (JKO$_h$) and $T = N\tau$. The goal is to prove convergence of $\{\rho_k^{h,\tau}\}_{k=0,\dots,N}$ to a solution of (1). Our strategy is to show that there exists a limit curve $[0,T] \ni t \mapsto \rho_t$ which is a solution of (1) in the EDI sense. This means that there exists a velocity field v such that (ρ, v) satisfies the continuity equation

$$\partial_t\rho + \operatorname{div}(\rho v) = 0 \qquad \text{on } (0,T)\times\Omega$$

and the energy-dissipation inequality (EDI) holds true

$$\text{(3)}\quad \mathcal{F}(\rho_T) - \mathcal{F}(\rho_0) + \frac{1}{2}\int_0^T \Big\{ \int_\Omega |v_t|^2 d\rho_t + \int_\Omega |\nabla \ell(u_t)|^2 d\mathcal{L}^d \Big\} dt \le 0, \quad \rho_t = u_t\mathcal{L}^d,$$

where ℓ related to the energy density f in the following way: $\sqrt{s}f''(s) = \ell'(s)$ and $\ell(0) = 0$.

For the standard JKO scheme, the analogous convergence result is proven using the variational interpolation [3, Chapter 3]. The idea is to prove the inequality

$$\text{(4)}\qquad \mathcal{F}(\rho^\tau_{k+1}) - \mathcal{F}(\rho^\tau_k) + \frac{W_2^2(\rho^\tau_k, \rho^\tau_{k+1})}{2\tau} + \int_0^\tau \frac{W_2^2(\rho^\tau_k, \rho^\tau_r)}{2r^2} dr \le 0,$$

where ρ^τ_r is variational interpolant between ρ^τ_k and ρ^τ_{k+1}. Combining (4) with the lower bound on the Wasserstein distance with a slope of the energy

$$\text{(5)}\qquad \frac{1}{\tau} W_2(\rho^\tau_{k+1}, \rho^\tau_k) \ge \operatorname{Slope}\mathcal{F}(\rho^\tau_{k+1}),$$

one gets a sharp inequality which is convenient to pass to the limit to recover (3).

The crucial step in the discrete setting is to find an appropriate replacement for (5). Note that we cannot use the metric slope, because it blows up as $h \to 0$. Instead of the slope, we use the discrete Fisher information defined as

$$\mathcal{S}_h(\rho^h) := \frac{1}{4}\sum_{x\in\mathcal{T}^h}\sum_{y\sim x} \frac{|\ell(u^h_y) - \ell(u^h_x)|^2}{h^2} h^d, \qquad u^h = \frac{\rho^h}{h^d}.$$

Second, we do not expect the discrete counterpart of (5) to hold exactly. An intuitive reason for that is that we know that we expect the inequality to fail if h/τ does not tend to 0.

The lower bound on the Wasserstein distance with the discrete Fisher information we find for the fully discrete case is presented in the following lemma.

Lemma. Let $\rho_0^h \in \mathcal{P}(\mathcal{T}^h)$ be given and $\rho^{\tau,h}$ is the minimizer of (JKO_h). Then

$$\frac{1}{\tau^2} W_2^2(\rho^{\tau,h}, \rho_0^h) \geq \left(1 - \frac{h}{2\tau}\right) \mathcal{S}_h(\rho^{\tau,h}) - \frac{dh}{2\tau}.$$

This abstract presents ideas on convergence of the fully discrete JKO scheme. There are plenty of related questions that have to be explored, in particular: extend the result to different energies such as interaction energy and energy appearing in crowd motion models, where f becomes a constraint $\rho \leq 1$; establish the rate of convergence; and develop the numerical algorithm.

References

[1] R. Jordan, D. Kinderlehrer, and F. Otto, *The variational formulation of the Fokker–Planck equation*, SIAM journal on mathematical analysis **29(1)** (1998), 1–17.

[2] F. Otto, *The geometry of dissipative evolution equations: the porous medium equation*, Comm. Partial Differential Equations **26(1-2)** (2001), 101–174.

[3] 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 (2008).

[4] J.-D. Benamou, G. Carlier, and M. Laborde, *An augmented Lagrangian approach to Wasserstein gradient flows and applications*, ESAIM: Proceedings and surveys **54** (2016), 1–17.

[5] G. Peyré, *Entropic approximation of Wasserstein gradient flows*, SIAM Journal on Imaging Sciences **8(4)** (2015), 2323–2351.

[6] M. Cuturi, *Sinkhorn distances: Lightspeed computation of optimal transport*, Advances in neural information processing systems **26** (2013), 120–140.

[7] J. Kitagawa, Q. Mérigot, and B. Thibert *Convergence of a Newton algorithm for semi-discrete optimal transport*, Journal of the European Mathematical Society **21(9)** (2019), 2603–2651.

Stability in Gagliardo-Nirenberg-Sobolev inequalities.

Nikita Simonov

(joint work with Matteo Bonforte, Jean Dolbeault and Bruno Nazaret)

In some functional inequalities, best constants and minimizers are known. The next question is stability: suppose that a function "almost attains the equality", in which sense it is close to one of the minimizers? We will address a recent result on the quantitative stability of a subfamily of Gagliardo-Nirengerg-Sobolev. The approach is based on the entropy method for the fast diffusion equation and allows us to obtain completely constructive estimates.

We consider the family of Gagliardo-Nirenberg-Sobolev inequalities given by

$$\|\nabla f\|_2^\theta \, \|f\|_{p+1}^{1-\theta} \geq \mathcal{C}_{\mathrm{GNS}}(p) \, \|f\|_{2p} \quad \forall\, f \in \mathcal{H}_p(\mathbb{R}^d)\,, \tag{1}$$

for simplicity we focus on the case $d \geq 3$, but $d = 1, 2$ can be treated (see [1]). The invariance of (1) under dilations determines the exponent

$$\theta = \frac{d\,(p-1)}{\big(d+2-p\,(d-2)\big)\,p}\,, \text{where} \quad p \in (1, p^\star] \quad \text{and} \quad p^\star := \tfrac{d}{d-2}\,.$$

The space $\mathcal{H}_p(\mathbb{R}^d)$ is defined as the completion of $C_c^\infty(\mathbb{R}^d)$, with respect to the norm $f \mapsto (1-\theta)\|f\|_{p+1} + \theta\|\nabla f\|_2$, where $\|f\|_q = \Big(\int_{\mathbb{R}^d} |f|^q dx\Big)^{1/q}$ for any $q > 1$. In the limit case where $p = p^\star$, for which $\theta = 1$, we are left with the space $\mathcal{H}_{p^\star}(\mathbb{R}^d) := \Big\{ f \in \mathrm{L}^{2\,p^\star}(\mathbb{R}^d) \,:\, |\nabla f| \in \mathrm{L}^2(\mathbb{R}^d) \Big\}$ and the Sobolev's inequality

$$\|\nabla f\|_2 \geq \mathsf{S}_d\,\|f\|_{2\,p^\star} \quad \forall\, f \in \mathcal{H}_{p^\star}(\mathbb{R}^d)\,. \tag{2}$$

Optimality in both (1) and (2) is achieved on the manifold of the *Aubin-Talenti functions* (see, for instance [2] and [3])

$$\mathfrak{M} := \Big\{ g_{\lambda,\mu,y} \,:\, (\lambda,\mu,y) \in (0,+\infty) \times \mathbb{R} \times \mathbb{R}^d \Big\}$$

$$\text{where} \quad \mathsf{g}(x) = \Big(1 + |x|^2\Big)^{-\frac{1}{p-1}} \quad \forall\, x \in \mathbb{R}^d\,,$$

and $g_{\lambda,\mu,y}(x) := \lambda^{\frac{d}{2p}}\,\mu^{\frac{1}{2p}}\,\mathsf{g}\big(\lambda\,(x-y)\big)$ with the convention $\mu^q = |\mu|^{q-1}\,\mu$ if $\mu < 0$. We can rewrite inequalities (1) and (2) in the form of a *positive, non-scale-invariant* functional which we shall call the *deficit functional*

$$\delta[f] := (p-1)^2\,\|\nabla f\|_2^2 + 4\,\tfrac{d-p\,(d-2)}{p+1}\,\|f\|_{p+1}^{p+1} - \mathcal{K}_{\mathrm{GNS}}\,\|f\|_{2\,p}^{2\,p\,\gamma} \geq 0$$

with $\gamma = \frac{d+2-p\,(d-2)}{d-p\,(d-4)}$ and $\mathcal{K}_{\mathrm{GNS}}$ chosen so that $\delta[\mathsf{g}] = 0$. Up to a scaling, the fact that $\delta[f] \geq 0$ is equivalent to (1) and (2) with optimal constants. In particular $\mathcal{K}_{\mathrm{GNS}}$ can be computed in terms of $\mathcal{C}_{\mathrm{GNS}}$.

Let us explain how *fast diffusion equation* enter into play. In *self-similar variables*, the fast diffusion equation, posed on $\mathbb{R}^d$, $d \geq 3$, with exponent $m \in [m_1, 1)$ and $m_1 := 1 - 1/d$, is

$$\frac{\partial v}{\partial t} + \nabla \cdot \Big(v\,\nabla v^{m-1}\Big) = 2\,\nabla \cdot (x\,v)\,, \quad v(t=0,\cdot) = v_0\,. \tag{FDE}$$

By applying this flow to the *relative entropy* (see [2])

$$\mathcal{F}[v] := \frac{1}{m-1}\int_{\mathbb{R}^d} \Big(v^m - \mathcal{B}^m - m\,\mathcal{B}^{m-1}\,(v-\mathcal{B})\Big)\,dx$$

$$\text{where} \quad \mathcal{B}(x) := \big(1+|x|^2\big)^{\frac{1}{m-1}}\,,$$

we have $\frac{d}{dt}\mathcal{F}[v(t,\cdot)] = -\,\mathcal{I}[v(t,\cdot)]$ where the *relative Fisher information* functional $\mathcal{I}$ defined by

$$\mathcal{I}[v] := \frac{m}{1-m}\int_{\mathbb{R}^d} v\,\big|\nabla v^{m-1} - \nabla \mathcal{B}^{m-1}\big|^2 dx\,.$$

It is a key step to recognise that we are dealing with the same quantities as in the variational approach. With

$$p = \frac{1}{2\,m-1} \quad \Longleftrightarrow \quad m = \frac{p+1}{2\,p}\,, \quad v = f^{2p}\,, \quad \mathcal{B} = \mathsf{g}^{2p}$$

and in particular with the condition $1 < p \le d/(d-2)$, $d \ge 3$, which is equivalent to $m_1 \le m < 1$. Indeed, it turns out that , as observed in [2],

$$\frac{p+1}{p-1}\,\delta[f] = \mathcal{I}[v] - 4\,\mathcal{F}[v] \ge 0 \tag{EEP}$$

for $v = |f|^{2\,p}$. Inequality (EEP) is called *entropy-entopy production* inequality and its optimal constant is 4. In particular, one of the main observations of [2] is that inequalities (1) and (2) are equivalent to (EEP). At the same time, by applying (EEP) and Gronwall's lemma, we get

$$\mathcal{F}[v(t,\cdot)] \le \mathcal{F}[v_0]\,e^{-4t} \quad \forall\, t \ge 0 \tag{3}$$

if v solves (FDE) . Here the main observation is that the exponential decay estimate with factor 4 in (3) is equivalent to the optimal constant in (EEP), see [1]. In the same spirit, if we are able to obtain a better convergence rate than the one in (3) (for instance under some moment condition), this would translate into an improved entropy-entropy prodcution inequality and, therefore, into a stability result.

Our overall strategy is now to prove that under some moment conditions on v, we can improve the decay with the rate $\mathcal{F}[v(t,\cdot)] \le \mathcal{F}[v_0]\,e^{-(4+\zeta)t}$ for all $t \ge 0$ using the properties of (FDE). In a word, we look for improved decay rates of the entropy in order to establish an *improved entropy - entropy production inequality.* Details are given in [1, Chapter 2]. Why is it that we can expect to obtain an improved decay rate of $\mathcal{F}[v(t,\cdot)]$? This can be obtained by a careful analysis of the *asymptotic time layer* regime (that is, as $t \to +\infty$). It is of standard knowledge, see for instance [4], that solutions to (FDE) converge to $\mathcal{B}$ in strong topologies. Hence, it makes sense to consider the Taylor expansions of the entropy and the Fisher information around $\mathcal{B}$. This expansion give us two quadrativ forms defined by

$$\mathsf{F}[h] = \lim_{\varepsilon\to 0} \varepsilon^{-2}\,\mathcal{F}\big[\mathcal{B} + \varepsilon\,\mathcal{B}^{2-m}\,h\big] \quad \text{and} \quad \mathsf{I}[h] = \lim_{\varepsilon\to 0} \varepsilon^{-2}\,\mathcal{I}\big[\mathcal{B} + \varepsilon\,\mathcal{B}^{2-m}\,h\big]\,.$$

By a *Hardy-Poincaré inequality* detailed in [1, Chapter 2], we have

$$\mathsf{I}[h] \ge \Lambda\,\mathsf{F}[h]$$

with $\Lambda = 4$ if $\int_{\mathbb{R}^d} h\,\mathcal{B}^{2-m}dx = 0$ and $\Lambda = 4\,\big(1 + d\,(m-m_1)\big)$ if, additionally, we assume that $\int_{\mathbb{R}^d} x\,h\,\mathcal{B}^{2-m}dx = 0$. In other words, the optimal decay rate of $\mathcal{F}[v(t,\cdot)]$ is characterized in the asymptotic time layer regime as $t \to +\infty$ by the spectral gap $\Lambda = 4$. Under the additional moment condition on the center of mass, we obtain $\zeta = \Lambda - 4 > 0$ if $m > m_1$. Recall that $m > m_1$ means $p < d/(d-2)$ and covers the whole subcritical range of inequality (1), inequality (2) can also be treated but the method is more involved. Altogether, we have an improved decay rate on an *asymptotic time layer* $[T_\star, +\infty)$, that has been explored in [4]

and subsequent papers. An important feature is that the estimates on Λ are explicit but require strong regularity conditions, i.e., $(1-\varepsilon)\mathcal{B}(x) \le v(t,x) \le (1+\varepsilon)\mathcal{B}(x)$ for all $x \in \mathbb{R}^d$ and $t \ge T_\star$. This condition is ensured only if the initial datum v_0 satisfies the following moment condition (see [1, Chapters 3 and 7]) $\sup_{R>0} R^{\frac{2}{1-m}-d} \int_{|x|>R} v_0(x)dx \le A < \infty$.

Once an improved decay rate is obtained in the asymptotic time layer, by using a nonlinear nonlinear generalization of the *carré du champ method* of D. Bakry and M. Emery, we are able to obtain an imporved decay rate in the *initial time layer* $[0, T_\star]$, which is also explicit. Combining the two layers, we are able to obtain the *improved* entropy-entropy production inequality $\mathcal{I}[v] \ge (4+\zeta)\,\mathcal{F}[v]$ for a functions v which satisfy the above moment conditions.

In terms of the variational language introduced in the beginning, we can say that for $d \ge 3$ and $1 < p < p^\star$, for any $f \in \mathrm{L}^{2p}(\mathbb{R}^d)$ with $\nabla f \in \mathrm{L}^2(\mathbb{R}^d)$ such that $A := \sup_{r>0} r^{\frac{d-p\,(d-4)}{p-1}} \int_{|x|>r} |f|^{2p}\,dx < \infty$ we have the estimate

$$\delta[f] \ge \kappa \inf_{\varphi\in\mathfrak{M}} \int_{\mathbb{R}^d} \left|(p-1)\,\nabla f + f^p\,\nabla\varphi^{1-p}\right|^2 dx$$

for some explicit positive constant κ which depends only on d, p, $\|f\|_{2p}$, A, and takes positive values on $\mathfrak{M}$. In the case $p = p^\star$ the above result still holds true under a stonger moment assumption.

References

[1] M. Bonforte, J. Dolbeault, B. Nazaret, and N. Simonov, *Stability in Gagliardo-Nirenberg-Sobolev inequalities: Flows, regularity and the entropy method*, and to appear in *Memoirs of the AMS*, (2023).

[2] M. Del Pino and J. Dolbeault, *Best constants for Gagliardo-Nirenberg inequalities and applications to nonlinear diffusions*, J. Math. Pures Appl. (9), 81 (2002), pp. 847–875.

[3] G. Bianchi and H. Egnell, *A note on the Sobolev inequality*, J. Funct. Anal., 100 (1991), pp. 18–24.

[4] A. Blanchet, M. Bonforte, J. Dolbeault, G. Grillo, J. L. Vázquez, *Asymptotics of the fast diffusion equation via entropy estimates*, Archive for Rational Mechanics and Analysis 191 (2009), 2, pp. 347-385.

Minimal Acceleration for the Multi-Dimensional Isentropic Euler Equation

Michael Westdickenberg

We consider the multi-dimensional isentropic Euler equations

$$\left.\begin{aligned} \partial_t \varrho + \nabla\cdot(\varrho\mathbf{u}) &= 0 \\ \partial_t(\varrho\mathbf{u}) + \nabla\cdot(\varrho\mathbf{u}\otimes\mathbf{u}) + \nabla\pi &= 0 \end{aligned}\right\} \quad \text{in } [0,\infty)\times\mathbf{R}^d, \tag{1}$$

$$(\varrho,\mathbf{u})(0,\cdot) =: (\bar\varrho,\bar{\mathbf{u}}) \quad \text{initial data.}$$

This system expresses local conservation of mass and momentum. To close system (1) one needs to specify the pressure. We consider polytropic gases, for which

$$\pi(t,\cdot) = P\big(r(t,\cdot)\big)\,\mathcal{L}^d \quad \text{for all } t \in [0,\infty),$$

where $U(r) := \kappa r^\gamma$ with constants $\kappa > 0$ and $\gamma > 1$ and

$$P(r) = U'(r)r - U(r) \quad \text{for } r \geq 0.$$

Here r is the Radon-Nikodym derivative of ϱ w.r.t. the Lebesgue measure $\mathcal{L}^d$.

Smooth solutions $(\varrho, \mathbf{u})$ of (1) satisfy the additional conservation law

$$\partial_t\Big(\tfrac{1}{2}\varrho|\mathbf{u}|^2 + U(\varrho)\Big) + \nabla\cdot\bigg(\Big(\tfrac{1}{2}\varrho|\mathbf{u}|^2 + U'(\varrho)\varrho\Big)\mathbf{u}\bigg) = 0, \tag{2}$$

which expresses local conservation of total energy

$$E(\varrho,\mathbf{u}) := \tfrac{1}{2}\varrho|\mathbf{u}|^2 + U(\varrho),$$

which is the sum of kinetic and internal energy. Since solutions of (1) may become discontinuous in finite time, solutions must be considered in the weak sense and energy conservation (2) must be relaxed to an $\leq$ inequality.

Global existence of weak solutions to (1) is still an open problem in several space dimension. A useful relaxation with guaranteed existence is the notion of dissipative solutions, introduced by [1]. Dissipative solutions are defined as tuples of $(\varrho, \mathbf{m})$ and defect measures $\mathbf{R}, \phi$ that satisfy the continuity equation and

$$\partial_t\mathbf{m} + \nabla\cdot\left(\frac{\mathbf{m}\otimes\mathbf{m}}{\varrho} + P(\varrho)\mathbf{1}\right) + \boxed{\nabla\cdot(\mathbf{R}+\phi\mathbf{1})} = 0,$$

$$\frac{d}{dt}\int_{\mathbf{R}^d}\left(\tfrac{1}{2}\varrho|\mathbf{u}|^2 + U(\varrho) + \boxed{\frac{1}{2}\mathrm{tr}(\mathbf{R}) + \frac{1}{\gamma-1}\phi}\right)(t,dx) \leq 0.$$

Here $\mathbf{R}$, ϕ are measures taking values in the symmetric, positive semidefinite matrices and the non-negative numbers, which form closed convex cones. Dissipative solutions become weak solutions of (1) iff the defect measures vanish.

The construction of infinitely many weak solutions to (1), pioneered by De Lellis-Székelyhidi [2], starts from so-called subsolutions, which can be interpreted as dissipative solutions with defect measures nonvanishing in open sets. Superimposing over $(\varrho, \mathbf{m})$ highly oscillatory waves, one can then remove the discrepancy between dissipative and weak solutions. To the extent that abstract arguments like the Baire category theorem are used to ensure the convergence of the iterative procedure, this result is based on the axiom of choice.

In contrast, our goal is to construct dissipative solutions to the isentropic Euler equations (1) that minimizes the defect measures from the start. It may very well be possible that for certain configurations, such as Kevin-Helmholtz instabilities, nonvanishing defect measures must occur. Indeed, since no viscosity is present, oscillatory features may persist at arbitrarily small length scales. In such cases, the best one can hope for is to construct solutions that are as close to being a weak solution as possible. One can speculate that in regions where defect measures do

not vanish a variant of the De Lellis-Székelyhidi method could be used to repair the dissipate solution to become a (or infinitely many) weak solution(s).

In order to construct dissipative solutions with minimal defect measures we consider the acceleration functional, defined as

$$|\mathbf{m}'|(t) = \int_{\mathbf{R}^d} \operatorname{tr}\big(\mathbb{U}(t, dx)\big) \quad \text{for a.e. } t \in [0, \infty), \tag{3}$$

with momentum flux

$$\mathbb{U} := \left(\frac{\mathbf{m} \otimes \mathbf{m}}{\varrho} + P(\varrho)\mathbf{1} \right) + \boxed{\mathbf{R} + \phi\mathbf{1}}.$$

As the notation suggests, (3) can be understood as the metric derivative of the momentum curve $t \mapsto \mathbf{m}(t, \cdot)$ with respect to the dual Lipschitz norm, which is the natural topology for the momentum field, given its finiteness in total variation due to the energy bound. Notice that (3) is nonnegative because the defect measures $\mathbf{R}$ and ϕ are in closed convex cones. Since minimizing (3) for all times amounts to a multi-objective optimization problem, which typically does not have minimizers, we instead look for Pareto-optimal solutions, i.e., for minimal elements with respect to a suitable quasi-order defined in terms of comparing the acceleration (3) of different dissipative solutions at all times. A quasi-order is a binary relation that is reflexive and transitive, but not necessarily antisymmetric. If this quasi-order is compatible with a topology, one can use the following result by Wallace [3]:

Theorem (Wallace). Suppose that X is a nonempty compact set with a quasi-order R such that the set of predecessors $P(x)$ of x is closed for every $x \in X$. Then X has a minimal element, i.e., an element $m \in X$ with the property that,

$$\text{if } y \in X \text{ and } m \text{ can be compared at all, then } (m, y) \in R.$$

This result can be applied with X the set of dissipative solutions of (1) to given initial data, and with the quasi-order defined in terms of the acceleration functional (3). A suitable topology can be chosen as weak* convergence of Young measures. Note that Wallace's existence result constructs minimal elements starting from totally ordered subsets of X, which exist because of the Hausdorff maximal principle. Ultimately, it it therefore again an application of the axiom of choice.

References

[1] D. Breit, E. Feireisl, and M. Hofmanová, *Solution semiflow to the isentropic Euler system*, Arch. Ration. Mech. Anal. **235** (2020), 167–194.

[2] C. De Lellis and L. Székelyhidi, Jr., *On admissibility criteria for weak solutions of the Euler equations*, Arch. Ration. Mech. Anal. **195** (2010), 225–260.

[3] A. D. Wallace, *A fixed-point theorem*, Bull. Amer. Math. Soc. **51** (1945), 413–416.

Transport meets variational inference

Nikolas Nüsken
(joint work with Francisco Vargas, Shreyas Padhy and Denis Blessing)

Given probability measures $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$ and a fixed terminal time $T > 0$, our objective is to (algorithmically) construct vector fields $a : [0,T] \times \mathbb{R}^d \to \mathbb{R}^d$ with appropriate growth and regularity properties, such that the diffusion process

$$\mathrm{d}X_t = a_t(X_t)\,\mathrm{d}t + \mathrm{d}W_t, \qquad X_0 \sim \mu \tag{1}$$

transports μ to ν, that is, $X_T \sim \nu$. More precisely, we aim to construct functionals $\mathcal{L} : a \mapsto \mathbb{R}_{\geq 0}$ whose minimisers provide solutions to the stated transport problem. Clearly, neither a nor $\mathcal{L}$ will be unique without imposing further constraints. Building on a parameterisation $\theta \mapsto a_\theta$, typically in terms of neural networks, such functionals allow us to approximate transporting diffusions of the form (1) by applying gradient-descent type algorithms to $\theta \mapsto \mathcal{L}(a_\theta)$.

In the recent preprint [1], we propose a framework based on augmenting (1) to forward and reverse time diffusions,

$$\mathrm{d}X_t = a_t(X_t)\,\mathrm{d}t + \overrightarrow{\mathrm{d}}W_t, \qquad X_0 \sim \mu, \tag{2a}$$

$$\mathrm{d}X_t = b_t(X_t)\,\mathrm{d}t + \overleftarrow{\mathrm{d}}W_t, \qquad X_T \sim \nu, \tag{2b}$$

where $\overrightarrow{\mathrm{d}}$ and $\overleftarrow{\mathrm{d}}$ denote forward and backward Itô integrals, respectively. [1] The diffusions (2a) and (2b) induce path measures $\overrightarrow{\mathbb{P}}^{\mu,a}, \overleftarrow{\mathbb{P}}^{\nu,b} \in \mathcal{P}(C([0,T];\mathbb{R}^d))$, and we consider mappings of the form

$$(a,b) \mapsto D(\overrightarrow{\mathbb{P}}^{\mu,a} | \overleftarrow{\mathbb{P}}^{\nu,b}), \tag{3}$$

where D is a divergence (meaning that $D(\mathbb{Q}|\mathbb{P}) \geq 0$ for all $\mathbb{Q}, \mathbb{P} \in \mathcal{P}(C([0,T];\mathbb{R}^d))$, with equality if and only if $\mathbb{P} = \mathbb{Q}$), for example the Kullback-Leibler divergence

$$D_{KL}(\mathbb{Q}|\mathbb{P}) = \mathbb{E}_{X\sim\mathbb{Q}}\left[\log \frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}(X)\right].$$

The simple but key observation is that $D(\overrightarrow{\mathbb{P}}^{\mu,a} | \overleftarrow{\mathbb{P}}^{\nu,b}) = 0$ if and only if the pair (a,b) produces diffusions that transport μ to ν (and back), and therefore modifications of (3), such as imposing further constraints on a and b, allow us to approach the transport problem stated at the beginning. In [1], we thereby recover entropic interpolations, stochastic optimal control problems, as well as the recently introduced score matching and action matching objectives from machine learning. We also develop a novel loss functional for the Bayesian sampling problem,

$$\phi \mapsto \mathbb{E}\left[\int_0^T |\nabla \log \pi_t(X_t)|^2\,\mathrm{d}t + \frac{1}{\sqrt{2}}\int_0^T (\nabla \log \pi_t - \nabla\phi_t)(X_t)\cdot \overleftarrow{\mathrm{d}}W_t - \log \pi_T(X_T)\right].$$

[1] The notions of stochastic integration in (1) and (2a) are the same; we use $\overrightarrow{\mathrm{d}}$ in (2a) to promote the symmetry of the framework.

In the above, $(\pi_t)_{0\leq t\leq T} \subset \mathcal{P}(\mathbb{R}^d)$ is a fixed curve of probability measures, and at optimality, the diffusion driven by $a = \nabla\phi^*$ reproduces these time-marginal laws. Motivated by excellent numerical results and relationships to the Crooks and Jarzinksy identities from statistical physics, future work will aim at a deeper understanding of this nonstandard control functional (nonstandard because of the backward Itô integral), and extensions to kinetic diffusions.

References

[1] F. Vargas, S. Padhy, D. Blessing, N. Nüsken: *Transport meets Variational Inference: Controlled Monte Carlo Diffusions*, arxiv: 2307.01050.

Gradient flow characaterisation of the heat flow with Dirichlet boundary conditions

Matthias Erbar
(joint work with Giulia Meglioli)

In a bounded domain $\Omega \subset \mathbb{R}^d$ we consider the porous medium equation

$$\begin{cases} \partial_t \rho = \Delta \rho^\alpha & \text{in } (0,\infty)\times\Omega; \\ \rho(0,\cdot) = \rho_0 & \text{in } \Omega \\ \rho = \lambda & \text{on} (0,\infty)\times\partial\Omega\,, \end{cases} \tag{1}$$

with constant Dirichlet boundary condition $\lambda \in (0,\infty)$ and $\alpha \geq 1$. Our goal is to give a variational characterisation of solutions in terms of gradient flows in the space of measures. While a large number of results characterising various evlutionary PDEs with Neumann boundary conditions as gradient flow w.r.t. the Wasserstein distance is available, little is known to date concerning other types of boundary conditions. Figalli and Gigli [1] have introduced a variant of the Wasserstein distance allowing for change of mass by letting the boundary $\partial\Omega$ act as a reservoir. For μ,ν positive measures in $\mathcal{M}_p(\Omega) := \{\mu \in \mathcal{M}_+(\Omega) : \int_\Omega d(\cdot,\partial\Omega)^p d\mu < \infty\}$ they define a distance $Wb_p(\mu,\nu)$ via

$$Wb_p(\mu,\nu)^p := \inf_{\gamma\in Adm(\mu,\nu)} \int_{\overline{\Omega}\times\overline{\Omega}} |x-y|^p d\gamma(x,y)\,,$$

where $Adm(\mu,\nu)$ is the set of admissible transport plans and consists of all $\gamma \in \mathcal{M}_+(\overline{\Omega}\times\overline{\Omega})$ such that $\pi^1_\#\gamma|_\Omega = \mu$ and $\pi^2_\#\gamma|_\Omega = \nu$.

Figalli and Gigli [1] for the heat flow ($\alpha = 1$) and later Kim, Koo, and Seo [2] for the porous medium equation ($\alpha > 1$) showed the following. Consider the internal energy functional

$$E_\alpha(\mu) = \begin{cases} \int_\Omega U_\alpha(\rho)dx\,, & \mu = \rho\mathsf{Leb}|_\Omega\,, \\ +\infty\,, & \text{else}\,, \end{cases},$$

with

$$U_\alpha(s) = \begin{cases} s\big[\log s - \log\lambda - 1\big] + \lambda\,, & \alpha = 1, \lambda > 0\,, \\ \frac{s}{\alpha-1}\big[s^{\alpha-1} - \alpha\lambda^{\alpha-1}\big] + \lambda^\alpha\,, & \alpha > 1, \lambda \geq 0\,. \end{cases}$$

Then we have

Theorem 1 ([1,2])**.** *Solutions of the JKO-scheme*

$$\rho^\tau_{n+1} = \operatorname*{argmin}_\rho E_\alpha(\rho) + \frac{1}{2\tau} Wb_2(\rho, \rho_n)^2$$

converge in Wb_2 as the time step τ goes to zero to weak solutions $(\rho_t)_t$ of (1), i.e. $t \mapsto \rho^{\alpha-1/2} - \lambda^{\alpha-1/2}$ belongs to $L^2_{\mathsf{loc}}\big([0,\infty), H^1_0(\Omega)\big)$ and it holds

$$\int_\Omega \phi(\rho_t - \rho_s) = \int_s^t \int_\Omega \Delta\phi \rho_r^\alpha dr \quad \forall \phi \in C_c^\infty(\Omega), s < t\,.$$

This is strong evidence that (1) should be regarded as the gradient flow of E_α with respect to Wb_2. We also mention the work of Profeta and Sturm [3] who give a description of the heat flow with boundary condition $\lambda = 0$ as a contraction of a larger auxiliary system of positive and negative densities which can be characterised as a gradient flow.

We show that the porous medium equation can indeed be characterised as the gradient flow of E_α w.r.t. Wb_2 in the sense of curves of maximal slope. To this end, we first give a dynamic characterisation of the transport distance Wb_2. We denote by $\mathcal{CE}^\Omega_T$ the set of all pairs (μ, v) of time-dependent measures and vectorfields such that

(i) $[0,T] \ni t \mapsto \mu_t \in \mathcal{M}_2(\Omega)$ is vaguely continuous,
(ii) $\int_0^T \int_A |v_t| d\mu_t dt < \infty$ for all compact $A \subset \Omega$,
(iii) the continuity equation holds in the following sense:

$$\frac{d}{dt}\int \phi d\mu_t = \int \nabla\phi v_t d\mu_t \quad \forall \phi \in C_c^\infty(\Omega)\,.$$

Note that the choice of test functions in the continuity equation above allows for transport to and from the boundary. We obtain the following characterisation of absolutely continuous curves w.r.t. the distance Wb_p in terms of solutions to the continuity equation.

Theorem 2. *A curve $(\mu_t)_{t\in[0,T]}$ in $(\mathcal{M}_p(\Omega), Wb_p)$ is absolutely continuous if and only if there exists a Borel family $(v_t)_t$ of vector fields with*

$$\int_0^T \int |v_t|^p d\mu_t dt < \infty\,,$$

such that $(\mu, v) \in \mathcal{CE}^\Omega_T$. In this case, the family of vector fields with minimal L^p-norm satisfies $|\mu'|(t) = ||v_t||_{L^p(\mu_t)}$ for a.e. $t \in [0,T]$, where $|\mu'|$ denotes the metric derivative w.r.t. Wb_p.

As an immediate consequence we obtain a dynamic characterisation of the distance Wb_p in the spirit of the Benamou-Brenier formula for the Wasserstein distance.

Corollary 1 (Benamou-Brenier formula)**.** *For $\mu_0, \mu_1 \in \mathcal{M}_p(\Omega)$ we have*

$$Wb_p(\mu_0, \mu_1) = \inf \left\{ \int_0^1 \int |v_t|^p d\mu_t dt \right\}$$

where the infimum is taken over all pairs $(\mu, v) \in \mathcal{CE}_1$ connecting μ_0 and μ_1.

We define the energy dissipation functional $\mathcal{I}_\alpha : \mathcal{M}_2(\Omega) \to [0, +\infty]$ as follows:

$$\mathcal{I}_\alpha(\mu) := \begin{cases} C(\alpha) \int_\Omega \left| \nabla \left(\rho^{\alpha - 1/2} \right) \right|^2 & \text{if } \mu = \rho \mathsf{Leb}|_\Omega \text{ and } \rho^{\alpha-1/2} - \lambda^{\alpha-1/2} \in H_0^1(\Omega), \\ +\infty & \text{otherwise}, \end{cases}$$

for a numerical constant $C(\alpha)$. Note that the boundary condition λ for the density of ρ of μ is encoded in finiteness of $\mathcal{I}_\alpha(\mu)$. We then obtain the following variational characterisation of the porous medium equation (1) with Dirichlet boundary condition λ.

Theorem 3. *For any curve $(\mu_t)_{t \in [0,T]}$ in $(\mathcal{M}_2(\Omega), Wb_2)$ with $E_\alpha(\mu_0) < +\infty$ we have*

$$\mathcal{L}_T(\mu) := E_\alpha(\mu_T) - E_\alpha(\mu_0) + \frac{1}{2} \int_0^T \left[|\mu'|^2(r) + \mathcal{I}_\alpha(\mu_r) \right] dr \geq 0 \,.$$

Moreover, $\mathcal{L}_T(\mu_t) = 0$ if and only if $\mu_t = \rho_t \mathsf{Leb}|_\Omega$ with (ρ_t) a weak solution to the porous medium equation (1)*.*

In the framework of gradient flows in metric spaces the claim that $\mathcal{L}_T(\mu) \geq 0$ states that $\mathcal{I}_\alpha$ is a strong upper gradient of the functional E_α. The second claim states that weak solutions to (1) are precisely the curves of maximal slope w.r.t. this strong upper gradient. Note that the Dirichlet boundary condition is encoded through finiteness of $\mathcal{L}$ though the appearance of $\mathcal{I}_\alpha$. This is consistent with the observation that the De Giorgi functional $\mathcal{L}$ of a gradient flow PDE is strongly related with the path level large deviation rate functional of an underlying particle dynamics. In boundary driven particle systems leading to a macroscopic limit described by a PDE with Dirichlet boundary conditions, the rate function is typically infinite unless the boundary condition is satisfied for positive times.

We conclude by noting that the dissipation functional can be related to the metric slope of E_α w.r.t. Wb_2.

Proposition 1. *For any $\mu \in \mathcal{M}_2(\Omega)$ we have*

$$\mathcal{I}_\alpha(\mu) \leq |\nabla^- E_\alpha|(\mu) := \limsup_{\nu \to \mu} \frac{\left(E_\alpha(\mu) - E_\alpha(\nu) \right)^+}{Wb_2(\mu, \nu)}$$

In particular, this shows that finiteness of the metric slope $|\nabla^- E_\alpha|(\mu)$ implies that $\mu = \rho \mathsf{Leb}|_\Omega$ and ρ satisfies the Dirichlet boundary condition. Moreover, we note that by abstract results for gradient flows in metric spaces together with the last Proposition allow us to recover Theorem 1, i.e. the convergence of the JKO scheme to a a weak solution, from the variational characterisation in Theorem 3.

References

[1] A. Figalli, N. Gilgi, *A new transportation distance between non-negative measures, with applications to gradients flows with Dirichlet boundary conditions*, J. Math. Pures Appl. **94** (2010), 107–130.

[2] D. Kim, D. Koo, G. Seo, *A gradient flow for the porous medium equations with Dirichlet boundary conditions*, arXiv:2212.06092

[3] A. Profeta, K.T. Sturm, *Heat flow with Dirichlet boundary conditions via optimal transport and gluing of metric measure spaces* , Calc. Var. PDE **59**(117) (2020).

Poincaré and Logarithmic Sobolev Inequalities for Brownian Motion with Sticky-Reflecting Boundary Diffusion

Max von Renesse

(joint work with Marie Bormann and Feng-Yu Wang)

Brownian motion on domains with sticky-reflecting boundary diffusion appears naturally as a microscopic model for heat flow in solids with surface coating or in interacting particle systems with sticky-reflecting zero-range interaction. A rigorous construction of the such processes can be given efficiently via Dirichlet forms, where both the invariant measure and the energy form are mixtures from corresponding bulk and boundary contributions [5, 6]. The question of the speed of convergence to equilibrium arises naturally. For convergence in quadratic mean this question was addressed in a previous work [10], where we derived upper bounds for the Poincaré constant under strict positivity assumptions on the Ricci curvature of the manifold and the second fundamental form of the boundary. The central method is an interpolation in the decomposition of the total variance into partial variances. The latter can then be estimated by the bulk energy through (variants of) the Steklov eigenvalue problem. In positive curvature one can get explicit quantitative bounds from application of the Reilly formula to the corresponding minimizers.

The talk presents new work [2] which extends the previous estimates in two ways. First we extend the interpolation method to the case of general curvature bounds. Instead of the Reilly formula the main tool in this case is based on integration by parts with a properly chosen test function of specific boundary behaviour and controlled energy contribution in the interior. As a side result we obtain new explicit estimates for the Steklov eigenvalue in this case. The second extension gives also bounds for the logarithmic Sobolev constant, where a similar type of interpolation in the decomposition of the entropy of mixtures is used. As another side result we obtain new explicit estimates for the norm of the boundary trace operator for Sobolev functions and a corresponding boundary trace logarithmic Sobolev inequality which was studied before in the special case of Euclidan balls by Beckner [1].

References

[1] William Beckner, *Sharp Sobolev inequalities on the sphere and the Moser-Trudinger inequality*, Ann. of Math. (2) **138** (1993), no. 1, 213–242. MR 1230930

[2] Marie Bormann, Max von Renesse, and Feng-Yu Wang, *Functional Inequalities for Brownian Motion with Sticky-Reflecting Boundary Diffusion*, arXiv:2401.00206 (2023)

[3] Djalil Chafai and Florent Malrieu, *On fine properties of mixtures with respect to concentration of measure and Sobolev type inequalities.* Ann. Inst.H.Poincare, Probab.Statist.46 No. 1 72–96,2010.

[4] José F. Escobar, *The geometry of the first non-zero Stekloff eigenvalue*, J. Funct. Anal. **150** (1997), no. 2, 544–556. MR 1479552

[5] Torben Fattler, Martin Grothaus, and Robert Voßhall, *Construction and analysis of a sticky reflected distorted Brownian motion*, Ann. Inst. Henri Poincaré Probab. Stat. **52** (2016), no. 2, 735–762. MR 3498008

[6] Martin Grothaus and Robert Voßhall, *Stochastic differential equations with sticky reflection and boundary diffusion*, Electron. J. Probab. **22** (2017), Paper No. 7, 37. MR 3613700

[7] Georg Menz and Andre Schlichting, *Poincare and logarithmic Sobolev inequalities by decomposition of the energy landscape.* Ann. Probab. 42, No. 5, 1809–1884 (2014).

[8] Robert C. Reilly, *Applications of the Hessian operator in a Riemannian manifold*, Indiana Univ. Math. J. **26** (1977), no. 3, 459–472. MR 474149

[9] Feng-Yu Wang, *Estimates of the first Neumann eigenvalue and the log-Sobolev constant on non-convex manifolds.* Math. Nachr. 280, No. 12, 1431–1439 (2007).

[10] Victor Marx, Vitalii Konarovskyi, and Max von Renesse, *Spectral gap estimates for Brownian motion on domains with sticky-reflecting boundary diffusion*, arXiv:2106.00080 (2021)

Regularized Stein Variational Gradient Flow

Bharath K. Sriperumbudur

(joint work with Ye He, Krishnakumar Balasubramanian, Jianfeng Lu)

Given a potential function $V : \mathbb{R}^d \to \mathbb{R}$, the sampling problem involves generating samples from the density

$$\pi(x) := Z^{-1} e^{-V(x)}, \qquad \text{with} \qquad Z := \int_{\mathbb{R}^d} e^{-V(x)}\, dx$$

being the normalization constant, which is typically assumed to be unknown or hard to compute. The task of sampling arises in several fields of applied mathematics, including Bayesian statistics and machine learning in the context of numerical integration. There are two widely-used approaches for sampling: (i) diffusion-based *randomized* algorithms, which are based on discretizations of certain diffusion processes, and (ii) particle-based *deterministic* algorithms, which are discretizations of certain *approximate* gradient flows. A central idea connecting the two approaches is the seminal work [1] which provides a variational interpretation of the Langevin diffusion as the Wasserstein Gradient Flow (WGF),

$$\partial_t \mu_t = \nabla \cdot \left(\mu_t \, \nabla_{W_2} F(\mu_t)\right) = \nabla \cdot \left(\mu_t \, \nabla \log \frac{\mu_t}{\pi}\right),$$

where the term $\nabla_{W_2} F(\mu_t) = \nabla \log \frac{\mu_t}{\pi}$ is the Wasserstein gradient of the relative entropy functional (also called as the Kullback–Leibler divergence), defined by

$$F(\mu_t) = \mathrm{KL}(\mu_t|\pi) := \int_{\mathbb{R}^d} \log \frac{\mu_t(x)}{\pi(x)} \mu_t(x) dx,$$

evaluated at μ_t. This leads to the idea that sampling could be viewed as *optimization on the space of densities/measures*.

The Wasserstein gradient of the relative entropy, i.e., $\nabla \log \frac{\mu_t}{\pi}$ is related to the *Stein operator* by noting that, for any $v \in L_2^d(\mu_t)$,

$$\begin{aligned}
&\langle \nabla_{W_2} \mathrm{KL}(\mu_t|\pi), v \rangle_{L_2^d(\mu_t)} \\
&= \left\langle \nabla \log \frac{\mu_t}{\pi}, v \right\rangle_{L_2^d(\mu_t)} = \langle \nabla \log \mu_t, v \rangle_{L_2^d(\mu_t)} - \langle \nabla \log \pi, v \rangle_{L_2^d(\mu_t)} \\
&= -\int_{\mathbb{R}^d} \left(\nabla \cdot v + \langle \nabla \log \pi, v \rangle_2 \right) \mu_t(x)\, dx =: -\int_{\mathbb{R}^d} \mathcal{S}_\pi v \, d\mu_t,
\end{aligned}$$

where $\mathcal{S}_\pi$ is called the *Stein operator* and $L_2^d(\mu_t) := \{f = (f_1, \ldots, f_d), f_i \in L_2(\mu_t), \forall i : \sum_{i=1}^d \|f_i\|^2_{L_2(\mu_t)} < \infty\}$. Since

$$\mathrm{KL}\left((I + hv)_{\#}\mu_t|\pi\right) = \mathrm{KL}(\mu_t|\pi) + h\langle \nabla_{W_2} \mathrm{KL}(\mu_t|\pi), v \rangle_{L_2^d(\mu_t)} + o(h),$$

we have

$$\nabla_{W_2} \mathrm{KL}(\mu_t|\pi) = -\arg \inf_{\|v\|_{L_2^d(\mu_t)} \le 1} \mathrm{KL}\left((I + hv)_{\#}\mu_t|\pi\right) = \arg \sup_{\|v\|_{L_2^d(\mu_t)} \le 1} \int_{\mathbb{R}^d} \mathcal{S}_\pi v \, d\mu_t.$$

Recently, in the machine learning community, the Stein Variational Gradient Descent (SVGD) [2,3] is proposed as a deterministic space-time discretization—in contrast to the Langevin diffusion which is a randomized space-time discretization of WGF—of the Stein Variational Gradient Flow (SVGF) [4] defined as

$$\partial_t \mu_t = \nabla \cdot \left(\mu_t \, \mathcal{T}_{k,\mu_t} \nabla \log \frac{\mu_t}{\pi} \right),$$

where $\mathcal{T}_{k,\mu} : L_2^d(\mu) \to L_2^d(\mu)$ is the integral operator defined as $\mathcal{T}_{k,\mu} f(x) = \int_{\mathbb{R}^d} k(x,y) f(y) \mu(y) dy$ for any function $f \in L_2^d(\mu)$, and $k : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ is the reproducing kernel (r.k.) of a reproducing kernel Hilbert space, $\mathcal{H}$. By defining $\mathrm{id}_\mu : \mathcal{H}^d \to L_2^d(\mu)$, $f \mapsto f$ as the inclusion operator, it can be shown that $\mathcal{T}_{k,\mu} = \mathrm{id}_\mu \mathrm{id}_\mu^*$, which yields $\mathcal{T}_{k,\mu_t} \nabla \log \frac{\mu_t}{\pi} = \mathrm{id}_{\mu_t} \mathrm{id}_{\mu_t}^* \nabla \log \frac{\mu_t}{\pi}$, where

$$-\mathrm{id}_{\mu_t}^* \nabla \log \frac{\mu_t}{\pi} = \arg \sup_{\|v\|_{\mathcal{H}^d} \le 1} \int_{\mathbb{R}^d} \mathcal{S}_\pi v \, d\mu_t.$$

SVGD given by

$$x_i^{(n+1)} = x_i^{(n)} - \frac{h}{N} \sum_{j=1}^N k(x_i^{(n)}, x_j^{(n)}) \nabla V(x_j^{(n)}) - \nabla k(x_i^{(n)}, x_j^{(n)}), \; i = 1, \ldots, N$$

is an interactive particle system (unlike Langevin diffusion), where N is the number of particles, $h > 0$ is the step-size, and n is the time index. However, SVGD (which

is based on SVGF) only provides a discretization of a *constant-order approximation* to WGF due to the presence of the kernel integral operator in its vector field. Indeed, if $\mathrm{supp}(\mu_t) = \mathbb{R}^d$ and k is a bounded continuous translation invariant *characteristic* kernel [5] on $\mathbb{R}^d$ (e.g., Gaussian and Laplacian kernels), then

$$\begin{aligned}\|\mathcal{T}_{k,\mu_t} - I\|_{\mathrm{op}} &= \sup\{\|\mathcal{T}_{k,\mu_t} f - f\|_{L_2^d(\mu_t)} : \|f\|_{L_2^d(\mu_t)} = 1\} \geq \|\mathcal{T}_{k,\mu_t}\mathbf{1} - \mathbf{1}\|_{L_2^d(\mu_t)} \\ &\geq \|1 - \textstyle\int k(\cdot, x)\mu_t(x)\, dx\|_{L_2(\mu_t)} > 0,\end{aligned}$$

where $\mathbf{1} = (1, \overset{d}{\ldots}, 1)^\top$.

To overcome the above issue with the SVGF, we propose the Regularized Stein Variational Gradient Flow (R-SVGF) [6] where the vector field is obtained as

$$-\arg \sup_{(1-\nu)\|v\|^2_{L_2^d(\mu_t)} + \nu\|v\|^2_{\mathcal{H}^d} \leq 1} \int_{\mathbb{R}^d} \mathcal{S}_\pi v \, d\mu_t, \tag{1}$$

where $0 \leq \nu \leq 1$ interpolates between WGF and SVGF. Clearly, $\nu = 0$ corresponds to the vector field in WGF while $\nu = 1$ yields that of SVGF. The vector field in (1) can be shown to be $((1-\nu)\mathcal{T}_{k,\mu_t} + \nu I)^{-1}\mathcal{T}_{k,\mu_t}\nabla \log(\mu_t/\pi)$ when seen as an element of $L_2^d(\mu_t)$, which satisfies

$$\|((1-\nu)\mathcal{T}_{k,\mu_t} + \nu I)^{-1}\mathcal{T}_{k,\mu_t}\nabla \log(\mu_t/\pi) - \nabla \log(\mu_t/\pi)\|_{L_2^d(\mu_t)} \to 0 \quad \text{as} \quad \nu \to 0$$

if $\nabla \log(\mu_t/\pi) \in \overline{\mathrm{Ran}(\mathcal{T}_{k,\mu_t})}$. Additionally, if $\nabla \log(\mu_t/\pi)$ is sufficiently smooth, i.e., there exists $\gamma \in \left(0, \frac{1}{2}\right]$ such that $\nabla \log(\mu_t/\pi) = \mathcal{T}^\gamma_{k,\mu_t} h$, for some $h \in L_2^d(\mu_t)$, then

$$\|((1-\nu)\mathcal{T}_{k,\mu_t} + \nu I)^{-1}\mathcal{T}_{k,\mu_t}\nabla \log(\mu_t/\pi) - \nabla \log(\mu_t/\pi)\|_{L_2^d(\mu_t)} = O(\nu^{2\gamma}) \quad \text{as} \quad \nu \to 0.$$

In other words, $((1-\nu)\mathcal{T}_{k,\mu_t} + \nu I)^{-1}\mathcal{T}_{k,\mu_t}\nabla \log(\mu_t/\pi)$ is a good approximation to $\nabla \log(\mu_t/\pi)$ for small ν. With this motivation, the corresponding gradient flow

$$\partial_t \mu_t = \nabla \cdot \left(\mu_t \ \left((1-\nu)\mathcal{T}_{k,\mu_t} + \nu I\right)^{-1} \mathcal{T}_{k,\mu_t} \left(\nabla \log \frac{\mu_t}{\pi} \right) \right), \tag{2}$$

is referred to as R-SVGF, where $\nu \in (0, 1]$. Clearly, R-SVGF interpolates between WGF and SVGF. The key advantage is that (2), which approximates WGF, can be discretized to yield a deterministic interacting particle system (similar to that of SVGD but with modifications involving the inverse of regularized Gram matrix), R-SVGD:

$$\bar{x}_{n+1} = \bar{x}_n - h_{n+1}\mathbf{K}_n^{-1} \left(\frac{1}{N} K_n (L_n \nabla V) - \frac{1}{N} \sum_{j=1}^{N} L_n \nabla k(x_j^{(n)}, \cdot) \right)$$

where $(h_n)_{n=1}^\infty$ is the sequence of step-sizes, $\bar{x}_n = [x_1^{(n)}, \cdots, x_N^{(n)}]^T$,

$$\mathbf{K}_n := \left(\frac{(1-\nu_{n+1})}{N} K_n + \nu_{n+1} I_N \right)$$

with K_n being the Gram matrix, $(K_n)_{ij} = k(x_i^{(n)}, x_j^{(n)})$ for all $i, j \in \{1, \ldots, N\}$, and $L_n f := [f(x_1^{(n)}), \ldots, f(x_N^{(n)})]^\top$ for $f : \mathbb{R}^d \to \mathbb{R}^N$.

Our contributions in this work [6] are as follows:

(1) For the R-SVGF, we provide rates of convergence to the target density, π in two cases: (i) in the Fisher Information metric under no further assumptions on π and (ii) in the relative entropy under an LSI (log Sobolev inequality) assumption on π. We also establish similar results for the time-discretized R-SVGF.
(2) We characterize the existence, uniqueness, and stability of the solutions to the R-SVGF in the mean-field limit.
(3) We provide preliminary numerical experiments demonstrating the superiority of R-SVGD over SVGD in estimating certain functionals involving π based on their respective particle approximations.

References

[1] R. Jordan, D. Kinderlehrer, and F. Otto, *The variational formulation of the Fokker–Planck equation*, SIAM Journal on Mathematical Analysis **29** (1998), 1–17.
[2] Q. Liu, and D. Wang, *Stein variational gradient descent: A general purpose Bayesian inference algorithm*, Advances in Neural Information Processing Systems **29** (2016).
[3] Q. Liu, *Stein variational gradient descent as gradient flow*, Advances in Neural Information Processing Systems **30** (2017).
[4] J. Lu, Y. Lu, and J. Nolen, *Scaling limit of the Stein variational gradient descent: The mean field regime*, SIAM Journal on Mathematical Analysis **51** (2019), 648–671.
[5] B. K. Sriperumbudur, A. Gretton, K. Fukumizu, B. Schölkopf, and G. Lanckriet, *Hilbert space embeddings and metrics on probability measures*, Journal of Machine Learning Research **11** (2010), 1517–1561.
[6] Y. He, K. Balasubramanian, J. Lu, and B. K. Sriperumbudur, *Regularized Stein variational gradient flow*, (2022). `https://arxiv.org/pdf/2211.07861`

The superposition principle for BV curves of measures

Riccarda Rossi
(joint work with Stefano Almi, Giuseppe Savaré)

The evolution equations of diffusive type whose gradient-flow structure, in the space of probability measures metrized by the Wasserstein distance, was unveiled by Jordan, Kinderleher & Otto more than 25 years ago, all share a common structure. The cornerstone of such structure is the continuity equation

$$\partial_t \mu + \mathrm{div}(\boldsymbol{\nu}) = 0 \qquad \text{in } (0,T)\times\mathbb{R}^d, \tag{1a}$$

where $\mu = (\mu_t)_{t\in(0,T)}$ is a Borel family of probability measures on $(0,T)$, and the flux measure $\boldsymbol{\nu}$ disintegrates into a family of measures absolutely continuous with respect to μ_t, namely

$$\boldsymbol{\nu}_t = \boldsymbol{v_t}\mu_t \qquad \text{for } \mathcal{L}\text{-a.a. } t\in(0,T). \tag{1b}$$

The vector field $\boldsymbol{v} : (0,T)\times\mathbb{R}^d \to \mathbb{R}^d$ is usually referred to as *velocity field*; equation (1a) is understood in the sense of contributions. The central role of (1) in the variational approach to diffusion has motivated a thorough study of its properties. In particular, we mention the deep results in [1, Chap. 8], where

(i) it was proved that *The continuity equation characterizes absolutely continuous curves of measures with values in Wasserstein spaces.* More precisely, it was shown in [1, Thm. 8.3.1] that, in the case $p > 1$, for any given curve of probability measures (with finite pth-moment) $\mu : [0,T] \to \mathcal{P}_p(\mathbb{R}^d)$, p-absolutely continuous w.r.t. to the Wasserstein metric W_p with (W_p)-metric derivative $|\mu'| \in \mathrm{L}^p(0,T)$, there exists a velocity field $\boldsymbol{v} : (0,T) \times \mathbb{R}^d \to \mathbb{R}^d$ such that $\boldsymbol{v}_t \in \mathrm{L}^p(\mathbb{R}^d;\mu_t)$ for $\mathcal{L}$-a.a. $t \in (0,T)$, the pair $(\mu,\boldsymbol{v})$ solves the continuity equation (1), and the velocity field satisfies the 'optimality condition'

$$\|\boldsymbol{v}_t\|_{\mathrm{L}^p(\mathbb{R}^d;\mu_t)} \le |\mu'|(t) = \lim_{h\to 0} \frac{W_p(\mu_t,\mu_{t+h})}{|h|} \qquad \text{for } \mathcal{L}\text{-a.a. } t \in (0,T). \tag{2}$$

Conversely, in [1, Thm. 8.3.1] it was also proved that for any solution $(\mu,\boldsymbol{v})$ of the continuity equation, the curve $(0,T) \ni t \mapsto \mu_t \in \mathcal{P}_p(\mathbb{R}^d)$ is p-absolutely continuous and (2) holds as an equality.

(ii) *A probabilistic representation of solutions of the continuity equation via the superposition principle* was provided.

We have extended the above results to curves of measures with values in $W_1(\mathbb{R}^d)$, that are just with bounded variation as functions of time. Simple examples show that it is not to be expected that, with a curve $\mu \in \mathrm{BV}([0,T];W_1(\mathbb{R}^d))$, a flux measure $\boldsymbol{\nu}$ *absolutely continuous* w.r.t. μ as in (1b) may be associated. We have thus focused on the investigation of (1a) *per se*, understanding as solution of (1a) a pair $(\mu,\boldsymbol{\nu})$ such that

- μ is a finite positive Borel measure on $(0,T)\times\mathbb{R}^d$;
- the flux measure $\boldsymbol{\nu}$ has finite variation on $(0,T)\times\mathbb{R}^d$;
- $(\mu,\boldsymbol{\nu})$ solve (1a) in the distributional sense.

Hence, we have proved the following analogue of [1, Thm. 8.3.1], namely that *For any $\mu \in \mathrm{BV}([0,T];\mathcal{P}_1(\mathbb{R}^d))$ there exists a Borel measure $\boldsymbol{\nu} \in \mathcal{M}([0,T]\times\mathbb{R}^d;\mathbb{R}^d)$ solving the continuity equation* (1a) *in the sense above specified, such that, moreover,*

$$|\boldsymbol{\nu}|([0,T]\times\mathbb{R}^d) = \mathrm{Var}_{W_1}(\mu;[0,T]) \tag{3}$$

(which is the counterpart to (2)*), and the singular part of $\boldsymbol{\nu}$ w.r.t. μ, $\boldsymbol{\nu}^\perp$, is minimal in a suitable sense. Conversely, let $\mu \in \mathcal{M}^+([0,T]\times\mathbb{R}^d)$ and $\boldsymbol{\nu} \in \mathcal{M}([0,T]\times\mathbb{R}^d;\mathbb{R}^d)$ solve the continuity equation* (1a) *in the* BV *sense. Then, μ disintegrates w.r.t. the Lebesgue measure $\mathcal{L}$ on $(0,T)$, i.e. $\mu = \int_0^T \mu_t \,\mathrm{d}t$, such that the curve $(0,T) \ni t \mapsto \mu_t$ is in* $\mathrm{BV}([0,T];\mathcal{P}_1(\mathbb{R}^d))$*, and* (3) *holds.*

We have also provided a 'BV' counterpart to the superposition principle, by

- associating with (1a) a continuity equation in an augmented state space,
- resorting to the superposition principle for the 'augmented continuity equation'
- obtaining therefrom a probabilistic representation for the measures μ and $\boldsymbol{\nu}$ in terms of trajectories with values in the extended phase space.

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, 2^{nd} edition, Birkhäuser Verlag, Basel, 2008.

Graph limits for nonlocal interaction PDEs

Antonio Esposito

(joint work with Georg Heinze, Francesco Patacchini, André Schlichting, Dejan Slepčev)

The study of evolution equations on graphs and networks has been receiving increasing interest in view of possible applications in several real-world phenomena where individuals interact if they are interconnected in specific ways. In social networks, for example, one can model the spread of opinions, or behaviours, by assigning probabilities for individuals to adopt certain attitudes based on their neighbours' choices. This is useful to model polarisation and formation of echo chambers, cf. for example [1]. Another possible application concerns transportation networks, where the flux from one vertex to a connected one depends on some scalar quantities at the neighbour vertices, see e.g. [7]. Graphs are also used in applications to data science, as they are indeed a suitable mathematical structure to classify and represent data by studying clustering, as in [6, 8] and the references therein. In [5], we introduce nonlocal dynamics relevant to detecting local concentrations in networks. The class of partial differential equations (PDEs) we consider can be specified through three elements: a *nonlocal continuity equation*, an *upwind flux interpolation*, and a *constitutive relation* for a nonlocal velocity. The nonlocal continuity equation is concerned with the time-evolution of a probability measure $\rho_t \in \mathcal{P}(\mathbb{R}^d)$, for $t \in [0,T]$, where mass located at a vertex $x \in \mathbb{R}^d$ can be nonlocally transported to $y \in \mathbb{R}^d$ along a channel with capacity, referred to as weight, given by an edge weight function $\eta : \mathbb{R}^d \times \mathbb{R}^d \setminus \{x = y\} \to [0, \infty)$. The nonlocal continuity equation on a time interval $[0,T]$ is of the form

$$\partial_t \rho_t + \overline{\mathrm{div}} j_t = 0, \qquad \text{with} \quad \overline{\mathrm{div}} j_t(dx) = \int_{\mathbb{R}^d \setminus \{x\}} \eta(x,y) dj_t(x,y), \tag{1a}$$

where the flux is a time-dependent antisymmetric measure, $j_t \in \mathcal{M}(G)$, on the set $G = \{(x,y) \in \mathbb{R}^{2d} \setminus \{x = y\} : \eta(x,y) > 0\}$.

The relation constituting the flux depends on a σ-finite absolutely continuous measure $\mu \in \mathcal{M}^+(\mathbb{R}^d)$, wherein μ acts as an abstract notion of vertices of a graph. More precisely, we associate to a nonlocal time-dependent velocity field $v_t : G \to \mathbb{R}$ the induced flux by using an upwind interpolation as follows

$$dj_t(x,y) = v_t(x,y)_+ d(\rho \otimes \mu)(x,y) - v_t(x,y)_- d(\mu \otimes \rho)(x,y). \tag{1b}$$

Here, for $a \in \mathbb{R}$, we denote with $a_+ = \max\{a, 0\}$ and $a_- = \max\{-a, 0\}$ the positive and negative part, respectively. Intuitively, the support of μ defines the underlying

set of vertices, i.e. $V = \operatorname{supp}\mu$. In particular, any finite graph can be represented by choosing $\mu = \mu^N = \delta_{x_i}/N$, for $x_1, x_2, \dots, x_N \in \mathbb{R}^d$.

The last element is the identification of the velocity field in terms of a symmetric interaction potential $K : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ and a potential $P : \mathbb{R}^d \to \mathbb{R}$ by

$$v_t(x,y) = -\overline{\nabla}K * \rho_t(x,y) - \overline{\nabla}P(x,y), \tag{1c}$$

where the nonlocal gradient is defined by $\overline{\nabla}f(x,y) := f(y) - f(x)$.

In [5] we show that system (1) is a *Finslerian* gradient flow of the interaction energy

$$\mathcal{E}(\rho) = \frac{1}{2}\iint_{\mathbb{R}^{2d}} K(x,y)d\rho(y)d\rho(x) + \int_{\mathbb{R}^d} P(x)d\rho(x), \tag{2}$$

with respect to a nonlocal Wasserstein quasi-metric based on the upwind interpolation. In this framework we show existence of weak solutions, curves of maximal slope with respect to a specific strong upper gradient, and estabilish a discrete-to-continuum limit as the number of vertices n goes to ∞, so called *graph limit*. Different types of flux interpolations are considered in [4].

An intriguing problem is to understand the limiting behaviour of weak solutions to (1) as the graph structure localises, i.e. the range of connection between vertices decreases, while the weight of each connecting edge increases, so called *graph-to-local limit*. One expects to approximate weak solutions of the more standard nonlocal interaction equation on $\mathbb{R}^d$. However, the intrinsic geometry of the graph impacts the limiting gradient structure of the equation. Accordingly, the main goal of [3] is to provide a rigorous proof of the local limit of the system (1) along a sequence of edge weight functions $\eta^\varepsilon : \mathbb{R}^d \times \mathbb{R}^d \setminus \{x = y\} \to [0,\infty)$ defined by

$$\eta^\varepsilon(x,y) := \frac{1}{\varepsilon^{d+2}}\vartheta\left(\frac{x+y}{2}, \frac{x-y}{\varepsilon}\right), \tag{3}$$

in terms of a reference connectivity $\vartheta : \mathbb{R}^d \times \mathbb{R}^d \setminus \{0\} \to [0,\infty)$ satisfying suitable assumptions. The scaling in (3) leads to the local evolution

$$\partial_t\rho_t = \operatorname{div}(\rho_t \mathbb{T}(\nabla K * \rho_t + \nabla P)), \tag{$\mathsf{NLIE}_\mathbb{T}$}$$

where the *tensor* $\mathbb{T} : \mathbb{R}^d \to \mathbb{R}^{d\times d}$ depends on the nonlocal structure encoded through the reference measure μ and the connectivity ϑ.

Following a heuristic argument based on several approximations and smoothness assumptions, which are not a priori satisfied by solutions to (1) and ($\mathsf{NLIE}_\mathbb{T}$), one can show the link between the two equations. This is made rigorous in [3] by using a variational framework, allowing to handle measure-valued solutions.

An interesting byproduct of this result is the link between Finslerian and Riemannian gradient flows. More precisely, (1) is shown to be a gradient flow of the nonlocal interaction energy in the infinite-dimensional Finsler manifold of probability measures endowed with a nonlocal upwind transportation quasi-metric, $\mathcal{T}$, peculiar of the upwind interpolation (1b). Due to the loss of symmetry the underlying structure of $\mathcal{P}(\mathbb{R}^d)$ does not have the formal Riemannian structure, but Finslerian instead. On the other hand, following [9], we establish a chain-rule inequality for the nonlocal interaction energy in a 2-Wasserstein space defined over

$\mathbb{R}^d_{\mathbb{T}}$, which is $\mathbb{R}^d$ endowed with a metric induced by $\mathbb{T}^{-1}$. Upon considering the corresponding *Wasserstein scalar product* on the tangent space of $\mathcal{P}_2(\mathbb{R}^d_{\mathbb{T}})$, at some probability measure with bounded second moment, one can notice the underlying Riemannian structure, thereby making the connection between the weak and variational formulations of ($\mathsf{NLIE}_{\mathbb{T}}$). We stress that not only do we connect the graph and tensorized local gradient structures using the notion of curves of maximal slope for gradient flows after De Giorgi, but, upon identifying weak solutions of ($\mathsf{NLIE}_{\mathbb{T}}$) with curves of maximal slopes, we also obtain an existence result for ($\mathsf{NLIE}_{\mathbb{T}}$) via stability of gradient flows. This is indeed another interesting property of the graph, as it represents a valuable space-discretisation for the PDE under study, working in any dimension, in addition to other methods, e.g. particle approximations and tessellations. Indeed, our result can be also seen as a deterministic approximation of ($\mathsf{NLIE}_{\mathbb{T}}$). The results in [3] are extended to the multi-species case in [2].

References

[1] F. Baumann, P. Lorenz-Spreen, I. M. Sokolov, and M. Starnini, *Modeling echo chambers and polarization dynamics in social networks*, Phys. Rev. Lett. **124** (2020).

[2] A. Esposito, G. Heinze, J. - F. Pietschmann, and A. Schlichting, *Graph-to-local limit for a multi-species nonlocal cross-interaction system*, Proceedings in Applied Mathematics & Mechanics (2023).

[3] A. Esposito, G. Heinze, and A. Schlichting, *Graph-to-local limit for the nonlocal interaction equation*, Preprint arXiv:2306.03475 (2023).

[4] A. Esposito, F. S. Patacchini, and A. Schlichting, *On a class of nonlocal continuity equations on graphs*, Eur. J. Appl. Math. (2023), 1–18.

[5] A. Esposito, F. S. Patacchini, A. Schlichting, and D. Slepčev, *Nonlocal-interaction equation on graphs: gradient flow structure and continuum limit*, Arch. Ration. Mech. Anal. **240 (2)** (2021), 699–760.

[6] N. García Trillos and D. Slepčev, *A variational approach to the consistency of spectral clustering*, Appl. Comput. Harmon. Anal. **45(2)** (2018), 239–281.

[7] H. Koike, H. Takayasu, and M. Takayasu, *Diffusion-localization transition point of gravity type transport model on regular ring lattices and bethe lattices*, Journal of Statistical Physics **186 (3)** (2022).

[8] T. Laux and J. Lelmi, *Large data limit of the mbo scheme for data clustering: convergence of the dynamics*, J. Mach. Learn. Res. **24(344)** (2023), 1–49.

[9] S. Lisini, *Nonlinear diffusion equations with variable coefficients as gradient flows in Wasserstein spaces*, ESAIM Control Optim. Calc. Var. **15(3)** (2009), 712–740.

An adversarial mean curvature flow

Leon Bungert

(joint work with Tim Laux, Kerrek Stinson)

In this talk we discuss how mean curvature flows appear in the context of training adversarially robust classifiers in machine learning. Such classifiers can be obtained using an algorithm called adversarial training. To set the scene, let $\Omega \subset \mathbb{R}^d$ denote an open and bounded set and let $\mu \in \mathcal{P}(\Omega \times \{0,1\})$ be a probability measure, modeling the distribution of training data with the labels 0 and 1 in Ω. Adversarial

training finds a binary classifier A in the Borel subset $\mathfrak{B}(\Omega)$ of Ω by solving the following robust optimization problem

$$\inf_{A\in\mathfrak{B}(\Omega)} \mathbb{E}_{(x,y)\sim\mu}\left[\sup_{\tilde{x}\in B_\varepsilon(x)} |1_A(\tilde{x})-y|\right], \tag{1}$$

parameterized by the so-called adversarial budget $\varepsilon > 0$.

$$\mathbb{E}_{(x,y)\sim\mu}\left[\sup_{\tilde{x}\in B_\varepsilon(x)} |1_A(\tilde{x})-y|\right] = \mathbb{E}_{(x,y)\sim\mu}\left[|1_A(x)-y|\right] + \varepsilon\,\mathrm{Per}_\varepsilon(A;\mu), \tag{2}$$

where $\mathrm{Per}_\varepsilon(A;\mu)$ denotes a nonlocal perimeter functional. This rewriting, the asymptotic results in [3], and the fact that $\mathrm{Per}_\varepsilon(\cdot;\mu)$ Gamma-converges to a local perimeter as $\varepsilon\to 0$ [2] suggest that (1) can be interpreted as time discretization of mean curvature flow, where ε acts both as time step and non-locality scale of the perimeter functional. To make this connection rigorous, we consider the following iteration for $k\in\mathbb{N}_0$:

$$A_0 \in \operatorname*{arg\,min}_{A\in\mathfrak{B}(\Omega)} \mathbb{E}_{(x,y)\sim\mu}\left[|1_A(x)-y|\right] \tag{3a}$$

$$A_{k+1} \in \operatorname*{arg\,min}_{A\in\mathfrak{B}(\Omega)} \int_\Omega |1_A(x)-1_{A_k}(x)|\,\mathrm{dist}(x,\partial A_k)\,\mathrm{d}\varrho(x) + \varepsilon\,\mathrm{Per}_\varepsilon(A;\mu). \tag{3b}$$

Here the set A_0 in (3a) is a so-called Bayes classifier which acts as initial condition. Starting from there, the iteration (3b) performs adversarial training using the label distribution from the previous classifier A_k and modifying (2) by means of the distance function $\mathrm{dist}(\cdot,\partial A_k)$ to the decision boundary of the previous classifier. The probability measure $\varrho\in\mathcal{P}(\Omega)$ in (3b) is the first marginal of μ, that is $\varrho := \mu(\cdot\times\{0,1\})$. The presence of the distance function is necessary to obtain the correct normal velocity for mean curvature flow [7]. Since the minimization problem in (3b) does not have unique solutions, we take the approach of [5] to select a solution using a strongly convex minimization problem. For this we replace (3b) by $A_{k+1} := T_\varepsilon(A_k)$, where the operator $T_\varepsilon:\mathfrak{B}(\Omega)\to\mathfrak{B}(\Omega)$ is defined as

$$\begin{aligned} T_\varepsilon(A) &:= \{u^*\le 0\} \qquad \text{where } u^* \text{ solves}\\ u^* &:= \operatorname*{arg\,min}_{u\in L^2(\Omega)} \frac{1}{2}\int_\Omega |u(x)-\mathrm{sdist}(x,A)|^2\,\mathrm{d}\varrho(x) + \varepsilon\,\mathrm{TV}_\varepsilon(u;\mu). \end{aligned} \tag{4}$$

Here $\mathrm{TV}_\varepsilon(u;\mu) := \int_{\mathbb{R}}\mathrm{Per}_\varepsilon(\{u\ge t\};\mu)\,\mathrm{d}t$ denotes a total variation functional and $\mathrm{sdist}(\cdot,A_k) := \mathrm{dist}(\cdot,A_k) - \mathrm{dist}(\cdot,\mathbb{R}^d\setminus A_k)$ is the signed distance function of the set A_k. It can indeed be shown that $A_{k+1} := T_\varepsilon(A_k)$ is a solution of the minimization problem in (3b).

The goal is to prove that (4) is a monotone and consistent scheme for the weighted mean curvature flow $t\mapsto A(t)$ with normal velocity

$$v(t) := -\frac{1}{\varrho}\,\mathrm{div}\left(\varrho\,\nu_{A(t)}\right), \tag{5}$$

where $\nu_{A(t)}$ is the outer unit normal to the boundary $\partial A(t)$. The abstract results of [6] then imply that (3) converges to a so-called barrier solution of the weighted mean curvature flow as $\varepsilon \to 0$.

Monotonicity of (4) in the sense of set inclusion (i.e., $A \subset B$ implies $T_\varepsilon(A) \subset T_\varepsilon(B)$) is a straightforward consequence of a comparison principle for the minimization problem in (4).

To verify consistency, one works with smooth super- / subflows, i.e., smooth evolutions of smooth sets $t \mapsto A(t)$ which move strictly faster / slower than mean curvature flow. Consistency of T_ε then means that for $\varepsilon > 0$ small enough it holds $T_\varepsilon(A(t)) \supset A(t+\varepsilon)$, meaning that the superflow also moves strictly faster than the scheme (and vice versa for subflows).

To show this we follow the strategy developed in [4] and utilize a superflow $t \mapsto A(t)$ in order to construct a supersolution of the minimization problem in (4). The signed distance function $d(t,x) := \operatorname{sdist}(x, A(t))$ of a smooth superflow satisfies the partial differential inequality

$$\partial_t d(t,x) > \frac{1}{\varrho(x)} \operatorname{div}\left(\varrho(x)\nabla d(t,x)\right), \quad x \in \partial A(t). \tag{6}$$

Considering the rescaled function $v_\varepsilon(x) := \psi(d(t+\varepsilon,x))$ with an appropriately chosen function $\psi : \mathbb{R} \to \mathbb{R}$ that satisfies $\psi(s) \geq s$ for all $s \in \mathbb{R}$ and $\psi(s) = s$ for $|s|$ small, one then gets for small $\varepsilon > 0$ that:

$$v_\varepsilon(x) - d(t,x) > \frac{1}{\varrho(x)} \operatorname{div}\left(\varrho(x)\frac{\nabla v_\varepsilon(x)}{|\nabla v_\varepsilon(x)|}\right). \tag{7}$$

The main ingredient for proving consistency is a careful analysis of the subdifferential of the total variation $u \mapsto \mathrm{TV}_\varepsilon(u;\mu)$ and the proof that for functions u with non-vanishing gradient and $\varepsilon \to 0$ it is consistent with the 1-Laplace operator $-\operatorname{div}\left(\varrho\frac{\nabla u}{|\nabla u|}\right)$. Applying this expansion to the function $u = v_\varepsilon$ yields that on a neighborhood of the interface $\partial A(t)$ we have

$$\left(v_\varepsilon(x) - d(t,x)\right)\varrho(x) + \varepsilon p(x) \geq 0 \tag{8}$$

for a subgradient $p \in \partial\,\mathrm{TV}_\varepsilon(v_\varepsilon;\mu)$. Inequality (8) together with a careful analysis of boundary conditions imply that v_ε is a supersolution of the problem (4) which implies $T_\varepsilon(A(t)) \supset \{v_\varepsilon \leq 0\} = \{d(\cdot, t+\varepsilon) \leq 0\} = A(t+\varepsilon)$ and hence consistency.

These results have interesting theoretical and practical implications: First, the minimizing movement scheme (3) can be modified to build discrete approximations to mean curvature flow on grids. The corresponding discrete perimeter functional was already investigated in [2]. Following a similar approach as in the continuum setting, we expect that such schemes are also monotone and consistent with mean curvature flow, yielding a novel discretization method. Second, on the applied side, the schemes (3b) or (4) give rise to novel adversarial training methods that involve the distance function to the decision boundary. The interpretation of (3b) is that—in contrast to (2)—the adversarial budget ε is replaced by a data dependent budget $\frac{\varepsilon}{\operatorname{dist}(x,\partial A_k)}$ which becomes large for points x which are close to the decision boundary and small for points far away.

References

[1] L. Bungert, N. García Trillos, R. Murray, *The geometry of adversarial training in binary classification*, Information and Inference: A Journal of the IMA, **12(2)** 2023, 921–968.

[2] L. Bungert, K. Stinson, *Gamma-convergence of a nonlocal perimeter arising in adversarial machine learning*, arXiv preprint arXiv:2211.15223, 2022.

[3] N. García Trillos, R. Murray, *Adversarial classification: Necessary conditions and geometric flows*, Journal of Machine Learning Research, **23(187)** 2022, 1–38.

[4] A. Chambolle, M. Novaga, *Approximation of the anisotropic mean curvature flow*, Mathematical Models and Methods in Applied Sciences, **17(06)** 2007, 833–844.

[5] A. Chambolle, *An algorithm for mean curvature motion*, Interfaces and free Boundaries, **6(2)** 2004, 195–218.

[6] G. Bellettini, M. Novaga, *Minimal barriers for geometric evolutions*, Journal of Differential Equations, **139(1)** 1997, 76–103.

[7] F. Almgren, J. E. Taylor, L. Wang, *Curvature-driven flows: a variational approach*, SIAM Journal on Control and Optimization, **31(2)** 1993, 387–438.

Minimizing movements along families of energies and dissipations

Andrea Braides

This is a short user's guide for the analysis of evolution for families of energies and dissipations, and its connection with Γ-convergence. For applications I have found it convenient to use a flexible version of the minimizing-movement approach as in the following definition.

Definition (see e.g. [5]). Given sequences $\varepsilon = \varepsilon_k \to 0$ and $\tau = \tau_k \to 0$ of positive numbers, X a topological space, $F_\varepsilon \colon X \to \mathbb{R} \cup \{+\infty\}$ and $D_\varepsilon \colon X \times X \to [0, +\infty]$, a *minimizing movement along F_ε with dissipations D_ε at scale τ* with initial data $u_0^\varepsilon \to u_0$ is a function $u \colon [0, +\infty) \to X$ such that $u(t)$ is the pointwise limit of $u^k(t)$ for all $t \geq 0$, and $u^k(t) = u^k_{\lfloor t/\tau_k \rfloor}$, where $u_0^k = u_0^{\varepsilon_k}$ and u_i^k is a minimizer of $v \mapsto F_{\varepsilon_k}(v) + \frac{1}{\tau_k} F_{\varepsilon_k}(v, u^k_{i-1})$.

We note that the usual conditions ensuring the existence of a minimizing movement for a single functional and dissipation (see [2]) allow to prove the existence of a minimizing movement along F_ε with dissipations D_ε. In particular this is achieved when (X, d) is a complete metric space, F_ε are lower semicontinuous and coercive, and $D_\varepsilon(u, v) = \frac{1}{2} d^2(u, v)$ (in which case we use the shorthand $D_\varepsilon = \frac{1}{2} d^2$). However that is an extremely abstract result and must be coupled with some characterization of u. A way to characterize u is in terms of curves of maximal slope.

Theorem (commutativity in terms of curves of maximal slope). *Let (X, d) be a complete metric space, $D_\varepsilon = \frac{1}{2} d^2$, F_ε a family of functionals, and F_0 a functional on X such that the following property holds.*

(H) *if $v_\varepsilon \to v$ is such that $\sup_\varepsilon (F_\varepsilon(v_\varepsilon) + |\partial F_\varepsilon|(v_\varepsilon)) < +\infty$, then we have $\lim_{\varepsilon \to 0} F_\varepsilon(v_\varepsilon) = F_0(v)$ and $\liminf_{\varepsilon \to 0} |\partial F_\varepsilon|(v_\varepsilon) \geq |\partial F_0|(v)$.*

Then every minimizing movement along F_ε is a curve of maximal slope for F_0.

We note that, while implied by convexity (and valid also under more general assumptions), condition (H) is unlikely to hold when we have many local minima

v_ε, for which $|\partial F_\varepsilon|(v_\varepsilon) = 0$. Indeed, if such families are dense, we obtain that $|\partial F_0|$ is identically 0, and F_0 is a constant.

Link with Γ-convergence. If $|\partial F_\varepsilon|$ are equibounded in a neighbourhood of v then $F_0(v)$ coincides with the Γ-limit of F_ε at v. An example by M. Solci shows that this equality in general may fail at all points even if $|\partial F|$ is everywhere finite (see [5]). Here and below we give as understood that Γ-limits are computer with respect to the topology of X.

We now consider the minimizing-movement scheme in terms of convergence of minimum problems, which are compatible with Γ-convergence.

Theorem (extreme regimes). *Let (X, d) be a complete metric space, let F_ε be a equi-coercive family, $D_\varepsilon = \frac{1}{2}d^2$, and $u_\varepsilon^0 \to u^0$, and let u be a minimizing movement along F_ε with dissipations D_ε at scale τ with initial data $u_0^\varepsilon \to u_0$. Then*

(i) *there exists $\underline{\tau}_\varepsilon$ such that if $\tau_k \le \underline{\tau}_{\varepsilon_k}$ then any such minimizing movement u is a limit of minimizing movements for F_{ε_k} with initial datum $u_0^{\varepsilon_k}$;*

(ii) *there exists $\overline{\tau}_\varepsilon$ such that, if $\tau_k \ge \overline{\tau}_{\varepsilon_k}$ and F_0 is the Γ-limit of F_{ε_k}, any such minimizing movement u is a minimizing movements for F_0 with initial datum u_0.*

Critical scales. If the minimizing movements in the two extreme cases described by items (i) and (ii) above do not coincide, then there exist one or more *critical scales* at which we have a "change of regime". The simplest such case is when the domain of F_ε is a discrete space, in which the only possible minimizing movements in regime (i) are constant (*pinning*). In this case there exists a minimal scale $\tau = \tau_\varepsilon$ for which the evolution is not trivial for some initial datum (*depinning regime*). Conversely, if the minimizing movements in cases (i) and (ii) coincide, it is not clear if all possible u are characterized by (ii) (or (i)).

The following result states that in the convex case minimizing movements are independent of the scale (see [3]).

Theorem (the convex case) *Let (X, d) be a complete metric space, let F_ε be a equi-coercive family of convex energies and $D_\varepsilon = \frac{1}{2}d^2$, and $u_\varepsilon^0 \to u^0$. Let F_0 be the Γ-limit of F_{ε_k}. Then every minimizing movement u along F_ε with dissipations D_ε at scale τ with initial data $u_0^\varepsilon \to u_0$ is a minimizing movement u for F_0 with initial datum u_0, and is also a limit of minimizing movements for F_{ε_k} with initial data $u_0^{\varepsilon_k}$.*

This theorem states in a sense that the convex case is 'trivial' since the limit is the same at all scales. Nevertheless it may be useful to characterize the limits of gradient flows of convex energies through the study of their discrete-in-time approximations obtained by solving the Euler-Lagrange equations of the incremental problems. As such it has been applied for example to prove the convergence of non-local gradient flows to standard parabolic equations [1], and of gradient flows of double-porosity models to parabolic equations with memory [4].

References

[1] R. Alicandro, N. Ansini, A. Braides, A. Piatnitski, and A. Tribuzio. *A Variational Theory of Convolution-Type Functionals*, Springer-Verlag, Berlin, 2023.

[2] L. Ambrosio, N. Gigli, and G. Savaré. *Gradient Flows: in Metric Spaces and in the Space of Probability Measures.* Birkhäuser, Basel, 2005.

[3] A. Braides. *Local Minimization, Variational Evolution and* Γ*-convergence.* Lecture Notes in Math. **2094**, Springer-Verlag, Berlin, 2014.

[4] A. Braides, V. Chiadò Piat, and A. Piatnitski. *Homogenization of discrete high-contrast energies.* SIAM J. Math. Anal. **47** (2015), 3064–3091.

[5] A. Braides and M. Solci. *Geometric Flows on Planar Lattices.* Birkhäuser, Basel, 2021.

Diffusive transport: geodesics, convexity, and gradient flows

Daniel Matthes

(joint work with Eva-Maria Rott, André Schlichting, Giuseppe Savaré)

1. The diffusive transport metric

On the space $X := \{\rho \in L^1(\mathbb{S}^1) | \rho \geq 0, \int \rho\, dx = 1\}$ of probability densities on the circle, introduce the Hellinger distance $\mathbb{H}$, the L^2-Wasserstein metric $\mathbb{W}$, and the diffusive transport distance $\mathbb{D}$, respectively, by

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

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

$$\mathbb{D}(\rho_0,\rho_1)^2 = \inf\left\{ \int_0^1 \int_{\mathbb{S}^1} \frac{w_s^2}{\rho_s}\, dx\, ds \,\middle|\, \partial_s\rho_s - \partial_{xx} w_s = 0 \right\},$$

where 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 the (generalized) continuity equation. It is known that $(X, \mathbb{H})$ is a complete metric space with the L^1-topology, and that $(\bar{X}, \mathbb{W})$ is a complete metric space, where the completion $\bar{X}$ is the space of probability measures on $\mathbb{S}^1$, with the narrow topology. We show:

Theorem 1. *$(\bar{X}, \mathbb{D})$ is a complete metric space with the narrow topology. More precisely:*

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

Geodesics w.r.t $\mathbb{D}$ are currently little understood. Formally, the geodesic equations for $\mathbb{H}$, $\mathbb{W}$ and $\mathbb{D}$ read, respectively, as follows:

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

While 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, the third one appears inaccessible to explicit solution.

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 potential for the gradient flow w.r.t. $\mathbb{W}$ is Boltzmann's entropy functional $\mathcal{H}(\rho) = \int \rho \log \rho \, dx$.

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

- *... a contractive flow w.r.t. $\mathbb{H}$ [2],*
- *... a (non-contractive) gradient flow w.r.t. $\mathbb{W}$ [1],*
- *... a (non-contractive) gradient flow w.r.t. $\mathbb{D}$ [5].*

There is apparently no easy explanation for the contractivity in $\mathbb{H}$. The potentials for the gradient flows w.r.t. $\mathbb{W}$ and $\mathbb{D}$ are, respectively, the Fisher information $\mathcal{F}(\rho) = \int \rho\, (\partial_x \log \rho)^2 \, dx$ and the entropy $\mathcal{H}$.

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 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 $\mathbf{m}(\rho_{k-1/2}, \rho_{k+1/2})$ 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. $\mathbb{W}^\delta$ [3, 4]*
- *... a contractive flow w.r.t. $\mathbb{D}^\delta$ [5].*

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 $\mathbf{m}(\rho_{\kappa-1/2},\rho_{\kappa+1/2}) = (\rho_{\kappa+1/2}-\rho_{\kappa-1/2})/\log(\rho_{\kappa+1/2}-\log\rho_{\kappa-1/2})$, and in the definition of $\mathbb{D}^\delta$, one uses $\mathbf{M}(\rho_{k-1},\rho_k,\rho_{k+1}) = \rho_k$.

Observation 4 ([5]). *The following discretization of the DLSS equation*

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

induces on X^δ ...

- *... a contractive flow w.r.t. $\mathbb{H}^\delta$*
- *... a (non-contractive) gradient flow w.r.t. $\mathbb{W}^\delta$*
- *... a (non-contractive) gradient flow w.r.t. $\mathbb{D}^\delta$.*

Differently from Observation 3, we choose $\mathbf{m}(\rho_{\kappa-1/2},\rho_{\kappa+1/2}) = \sqrt{\rho_{\kappa+1/2}\rho_{\kappa-1/2}}$ for $\mathbb{W}^\delta$, and for $\mathbb{D}^\delta$:

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

These choices of $\mathbf{m}/\mathbf{M}$ appear to be crucial to guarantee the contractivity in $\mathbb{H}^\delta$. Indeed, the proof uses that (1) can be re-formulated as

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

Our main result is about the convergence of the scheme (1).

Theorem 2 ([5]). *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 (1) possesses a unique solution $\rho^\delta : [0,\infty) \to X^\delta$, and*

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

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

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

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

References

[1] U. Gianazza, G. Savaré, G. Toscani, *The Wasserstein gradient flow of the Fisher information and the quantum drift-diffusion equation*, Archive for rational mechanics and analysis **194** (2009), 133–220.

[2] A. Jüngel, R. Pinnau, *Global nonnegative solutions of a nonlinear fourth-order parabolic equation for quantum systems*, SIAM Journal on Mathematical Analysis **32** (2000), 760–777.

[3] J. Maas, *Gradient flows of the entropy for finite Markov chains*, Journal of Functional Analysis **261** (2011), 2250–2292.

[4] A. Mielke, *Geodesic convexity of the relative entropy in reversible Markov chains*, Calculus of Variations and Partial Differential Equations **48** (2013), 1–31.
[5] 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), 36p.

Damage in viscoelastic materials at finite strains

Marita Thomas
(joint work with Manuel Friedrich, Martin Kružík, and Riccarda Rossi)

This contribution reports on an ongoing work in progress dedicated to the mathematical analysis of a model for the evolution of damage in viscoelastic materials with physical and geometrical nonlinearities and under the influence of dynamic effects due to the propagation of elastic waves.

1. Challenges related to dynamic effects at finite strains

As has been already observed in existing literature, cf. e.g., [1,2], one major challenge in this setting is the correct treatment of the axiom of material frame indifference ensuring that a model is independent of orthogonal rotations of the chosen coordinate system. Firstly, this requires static material frame indifference, i.e., $W(QF) = W(F)$ for all $(d\times d)$-matrices $F \in \mathrm{GL}^+(d)$ and $Q \in \mathrm{SO}(d)$, for a hyperelastic material with a stored elastic energy density $W : \mathrm{GL}^+(d) \to \mathbb{R}$. Yet, due to the presence of dynamic effects, this static condition is not enough to ensure the independence of the model of orthogonal rotations. Additionally, also dynamic material frame indifference is required, cf. [1], i.e., $V(QF; \partial_t(QF)) = V(F; \partial_t F)$ for all sufficiently smooth maps $F : [0,T] \to \mathrm{GL}^+(d)$ and $Q : [0,T] \to \mathrm{SO}(d)$, and where $V : \mathrm{GL}^+(d) \times \mathbb{R}^{d\times d} \to [0,\infty]$ denotes a dissipation potential to account for viscous effects of Kelvin-Voigt-type rheology. A simple, suitable choice is given by

$$V(F;\dot F) = \frac{1}{2}\mathbb{V}(F)\partial_t(F^\top F) : \partial_t(F^\top F)\,, \tag{1}$$

with $\mathbb{V}(F)G : G \geq c_{\mathbb{V}}|G|^2$ for all $F, G \in \mathbb{R}^{d\times d}$ and with a constant $c_{\mathbb{V}} > 0$. Suppose now that certain a priori estimates result in a bound on the corresponding integral functional, i.e. that $\mathcal{V}(\nabla\varphi;\nabla\dot\varphi) := \int_\Omega V(\nabla\varphi;\nabla\dot\varphi)\,\mathrm{d}x \leq C$. Then (1) directly results in the bound $\frac{c_{\mathbb{V}}}{2}\|\dot F^\top F + F^\top \dot F\|_{L^2}^2 \leq C$, but it does not provide a separate estimate on the partial time derivative $\dot F = \partial_t F = \nabla\dot\varphi$. In turn, this can be achieved thanks to generalized Korn's inequalities [3,4] of the form

$$\|\nabla\dot\varphi(t)\|_{L^2} \leq C_{\mathrm{K}}\|\nabla\dot\varphi^\top(t)\nabla\varphi(t) + \nabla\varphi^\top(t)\nabla\dot\varphi(t)\|_{L^2} \tag{2}$$

with a constant $C_{\mathrm{K}} > 0$. Yet, (2) to be valid requires that, firstly, $\nabla\varphi(t) \in C^0(\Omega;\mathbb{R}^{d\times d})$ with $\|\nabla\varphi(t)\|_\infty \leq C$, and secondly, that

$$\det\nabla\varphi(t) \geq c > 0 \quad \text{on } \Omega, \text{ uniformly for all } t \in [0,T]. \tag{3}$$

The first condition can be achieved by adding a higher order gradient term to the energy density in the spirit of second grade non-simple materials, i.e., a term

$$\mathcal{H}(\varphi) := \int_\Omega H(\nabla^2\varphi)\,\mathrm{d}x \tag{4}$$

will ensure the required regularity. Secondly, the term (4), given that $W + H$ additionally satisfies a growth estimate of the form

$$W(F) + H(G) \geq c_W\big(|F|^s + \frac{1}{(\det F)^q}\big) + c_H|G|^p \tag{5}$$

for all $F \in \mathrm{GL}^+(d)$ and $G \in \mathbb{R}^{d\times d\times d}$ with fixed constants $c_W, c_H > 0$ and fixed exponents $p > d$, $q \geq \frac{pd}{p-d}$, $s > 1$, will also provide condition (3) thanks to a result by Healey and Krömer [5]. The above considerations motivate the structure of the stored elastic energy density and of the viscoelastic dissipation potential.

2. The damage model

The effects of an evolving damage process on the elastic behavior of a body with reference configuration $\Omega \subset \mathbb{R}^d$ are further modeled with the aid of a damage variable $z : [0,T] \times \Omega \to [0,1]$, where $z(t,x) = 1$ means that the material is undamaged and $z(t,x) = 0$ that the material is maximally damaged in the material point $x \in \Omega$ at time $t \in [0,T]$. The energy functional is of the form

$$\mathcal{E}(t,z,\varphi) := \begin{cases} \int_\Omega \big(E_1(z,\varphi,\nabla\varphi,\nabla^2\varphi) - \langle \ell(t),\varphi\rangle + E_2(z,\nabla z)\big)\,\mathrm{d}x \\ \qquad \text{if } E_1(z,\varphi,\nabla\varphi,\nabla^2\varphi) - \langle \ell(t),\varphi\rangle + E_2(z,\nabla z) \in L^1(\Omega)\,, \\ \infty \quad \text{otherwise,} \end{cases} \tag{6}$$

with $E_1(z,\varphi,\nabla\varphi,\nabla^2\varphi) := W(z,\nabla\varphi) + H(z,\nabla^2\varphi)\,,\ E_2(z,\nabla z) := \frac{1}{2}|\nabla z|^2 + \phi(z)\,.$ Here, the energy term E_2 serves as a regularization for the damage variable and the function ϕ is chosen such that $z \in [0,1]$ can be ensured for a solution of the problem. The densities W and H are assumed to be suitably smooth, equipped with suitable analytical and physically reasonable growth properties, e.g., in the line of (5) and to ensure that a decrease of the damage variable (corresponding to an increase of damage) leads to a decrease of the stored elastic energy and, hence, the elastic stresses. A further ingredient to the model is the dissipation potential for the damage variable, which is assumed to be of the form

$$\mathcal{R}(\dot z) := \int_\Omega \big(R_1(\dot z) + R_2(\dot z) + I_{(-\infty,0]}(\dot z)\big)\mathrm{d}x \tag{7}$$

with $R_1(\dot z) := a_1|\dot z|$, $R_2(\dot z) := \frac{a_2}{2}|v|^2$, with constants $a_1, a_2 > 0$, and the indicator function $I_{(-\infty,0]}$, i.e., $I_{(-\infty,0]}(\dot z) = 0$ if $\dot z \in (-\infty,0]$ and $I_{(-\infty,0]}(\dot z) = \infty$ otherwise. The presence of two convex but non-smooth terms, the rate-independent dissipation R_1 and the indicator function $I_{(-\infty,0]}$ to prevent healing of damage, leads to an evolution law for the damage variable in terms of a subdifferential inclusion

$$a_1\mathrm{Sign}(\dot z) + a_2\dot z + \partial I_{(-\infty,0]}(\dot z) + \phi'(z) + \mathrm{D}_z W(z,\nabla\varphi) - \Delta z \ni 0\,, \tag{8}$$

where Sign and $\partial I_{(-\infty,0]}$ denote the subdifferentials of the absolute value function and the indicator function in the sense of convex analysis.

3. Existence of weak solutions and improved results

By means of a staggered time-discrete scheme one can prove the existence of weak solutions (z, φ), which are defined by the following three ingredients:

1. Weak formulation of the momentum balance:

$$\begin{aligned}\langle \ddot{\varphi}(t), \eta\rangle_{W^{2,p}} + \int_\Omega \big(\mathrm{D}_{\dot F} V(z(t), \nabla\varphi(t); \nabla\dot\varphi(t)) + \mathrm{D}_F W(z(t), \nabla\varphi(t))\big) : \nabla\eta \,\mathrm{d}x \\ + \int_\Omega \mathrm{D}_G H(z(t), \nabla^2\varphi(t)) : \nabla^2\eta \,\mathrm{d}x = \langle \ell(t), \eta\rangle_{W^{2,p}}\end{aligned} \tag{9a}$$

for almost all $t \in (0,T)$ and for all $\eta \in W^{2,p}(\Omega;\mathbb{R}^d) \cap H^1_0(\Omega;\mathbb{R}^d)$, together with

$$\min_{(t,x)\in[0,T]\times\overline{\Omega}} \det(\nabla\varphi(t)) > 0 \quad \text{and} \quad \varphi(t)|_{\partial\Omega}(t,\cdot) = \mathrm{Id} \quad \text{for all } t \in [0,T].$$

2. Damage flow rule in terms of a one-sided variational inequality, cf. also [6]:

$$\begin{aligned}\int_\Omega \big(a_1 + a_2\dot z(t) + \phi'(z(t)) + \mathrm{D}_z W(z(t), \nabla\varphi(t)) + \mathrm{D}_z H(z(t), \nabla^2\varphi(t))\big)\zeta \,\mathrm{d}x \\ + \int_\Omega \nabla z(t) \cdot \nabla\zeta \,\mathrm{d}x \geq 0\end{aligned} \tag{9b}$$

for almost all $t \in (0,T)$ and for all $\zeta \in H^1(\Omega) \cap L^\infty(\Omega)$ with $\zeta \leq 0$ a.e. in Ω.

3. Upper energy-dissipation estimate:

$$\begin{aligned}&\mathcal{E}(t, \varphi(t), z(t)) + \mathcal{K}(\dot\varphi(t)) \\ &\qquad + 2\int_s^t \big(\mathcal{V}(z(r), \varphi(r); \dot\varphi(r)) + \mathcal{R}_2(\dot z(r))\big)\,\mathrm{d}r + \int_s^t \mathcal{R}_1(\dot z(r))\,\mathrm{d}r \\ &\leq \mathcal{E}(s, \varphi(s), z(s)) + \mathcal{K}(\dot\varphi(s)) + \int_s^t \partial_t\mathcal{E}(r, \varphi(r), z(r))\,\mathrm{d}r\,,\end{aligned} \tag{9c}$$

for all $t \in [0,T]$ and almost all $s \in [0,t)$ and where $\mathcal{K}(\dot\varphi) := \int_\Omega \frac{\rho}{2}|\dot\varphi|^2 \,\mathrm{d}x$ denotes the kinetic energy with a constant mass density $\rho > 0$.

Under the additional assumption that $H : (z, G) \mapsto H(z, G)$ is convex, it can be further shown that the inequality (9c) improves to an equality. Moreover, if the regularization term H does not depend on z, then one obtains $z \in H^1(0,T; H^1(\Omega))$ and (9b) can be replaced by the subdifferential inclusion (8).

References

[1] S.S. Antman, *Physically unacceptable viscous stresses*, Z. Angew. Math. Phys. **49** (1998), 980–988.

[2] A. Mielke, T. Roubíček, *Thermoviscoelasticity in Kelvin-Voigt rheology at large strains*, Arch. Ration. Mech. Anal. **238** (2020), 1–45.

[3] P. Neff, *On Korn's first inequality with non-constant coefficients*, Proc. R. Soc. Edinb. **132A** (2002), 221–243.

[4] W. Pompe, *Korn's First Inequality with variable coefficients and its generalization*, Comment. Math. Univ. Carol. **44** (2003), 57–70.

[5] T.J. Healey, S. Krömer, *Injective weak solutions in second-gradient nonlinear elasticity*, ESAIM Control. Optim. Calc. Var. **15** (2009), 863–871.
[6] Ch. Heinemann, Ch. Kraus, *Existence of weak solutions for Cahn-Hilliard systems coupled with elasticity and damage*, Adv. Math. Sci. Appl. **21** (2011), 321–359.

Gradient flow: yes or no?

Jan Maas
(joint work with Morris Brooks)

Let $X \in \Gamma(TM)$ be a vector field on a smooth manifold M and let $f : M \to \mathbb{R}$ be a smooth function. Does there exist a Riemannian metric g on M such that the evolution equation $\dot{u} = X(u)$ is the gradient flow equation for f with respect to the metric g? In order words, using standard index notation, does there exist a metric $g_{\alpha\beta}$ such that $g_{\alpha\beta}X^\beta = -D_\alpha f$?

Some assumptions are clearly needed:

Firstly, Df should be zero at every stationary point of the evolution: the vector field X and the co-vectorfield Df should have the same set of zeroes.

Secondly, X and $-Df$ should "agree on the sign" outside the set where they vanish: for all $x \in M$ with $X(x) \neq 0$, they should satisfy $X^\alpha D_\alpha f(x) < 0$ (since $X^\alpha D_\alpha f = -g_{\alpha\beta}X^\alpha X^\beta$ and g is positive definite.) This requirement reflects the fact that f should decrease along the evolution.

Thirdly, at every point x where $X(x) = 0$, one should have that $D_\alpha D_\gamma f = \bar{g}_{\alpha\beta}D_\gamma X^\beta$, for some scalar product $\bar{g}_{\alpha\beta}$ on T_xM. This somewhat less obvious condition is obtained by differentiating the equation $D_\alpha f = g_{\alpha\beta}X^\beta$ at x, using the assumption that $X(x) = 0$.

Our main result asserts that these conditions are not only necessary, but also sufficient, under mild regularity conditions.

Theorem 1. *Let $f : M \to \mathbb{R}$ be a function and $X^\alpha \in \Gamma(TM)$ be a vector field. We assume that f and X are real-analytic (in some coordinate chart). Suppose further that Df has a unique zero, $\bar{x} \in M$, at which f attains its minimum. Then there exists a Riemannian metric $g_{\alpha\beta} \in \Gamma(T^*M \otimes T^*M)$ satisfying*

$$\nabla_\beta f = g_{\alpha\beta}X^\alpha,$$

if and only if the following conditions hold:

(1) *$D_{X^\alpha} f(x) < 0$ for all $x \in M$ with $x \neq \bar{x}$;*
(2) *$X^\alpha|_{\bar{x}} = 0$;*
(3) *The linear map $\Lambda := D_\alpha X^\beta|_{\bar{x}} : T_{\bar{x}}M \to T_{\bar{x}}M$ is positive and symmetric with respect to the Hessian scalar product $h_{\alpha\beta} := D_\alpha D_\beta f|_{\bar{x}}$ on $T_{\bar{x}}M$.*

In fact, [3] contains a more general version of this result, in which Df is replaced by an arbitrary co-vector field Y. We also prove a variant of this result in which X and Y are of class C^{k+1} for $k \geq 0$. In this case, the metric g is of class C^k. The case $k = 0$ was proved earlier in [2].

The existence of a metric with the desired properties is easy to prove outside the set of critical points; see, e.g., [1]. The nontrivial part of the proof is to establish

the existence of a smooth metric in a neighbourhood of every point where X vanishes. This is done using a power series construction by an iterative argument, in which each iterative step involves the solution of a certain tensor equation.

As an application of Theorem 1 we solve a problem that arose in joint work with Carlen on gradient flow formulations of Lindblad equations, which describe the time-evolution of open quantum systems. It was shown earlier [4, 9] that Lindblad equations with a certain symmetry condition (*GNS-detailed balance*) can be formulated as gradient flow equation for the quantum relative entropy. The notion of GNS-detailed balance is one among several quantum generalisations of the notion of detailed balance for classical Markov chains. Subsequently, a different notion of detailed balance (*BKM-detailed balance*) was shown to be necessary for the existence of an entropic gradient flow structure for Lindblad equations [5]. However, as the notion of BKM-detailed balance is strictly weaker than the notion of GNS-detailed balance, there was a gap between the necessary and sufficient conditions above. As a consequence of Theorem 1, we close this gap: the notion of BKM-detailed balance is also sufficient for the existence of an entropic gradient flow structure. This result provides a quantum analogue of earlier work on the sufficiency of detailed balance [6] in gradient flow structures for classical Markov chains [7, 8].

References

[1] T. Bárta, R. Chill, and E. Fašangová. Every ordinary differential equation with a strict Lyapunov function is a gradient system. *Monatsh. Math.*, 166(1):57–72, 2012.
[2] M. Bílý. Transformations of ODEs into gradient systems in stationary points. Bachelor thesis, Charles University in Prague, 2014.
[3] M. Brooks and J. Maas, Characterisation of gradient flows for a given functional, arXiv:2209.11149, 2022.
[4] E. A. Carlen and J. Maas. Gradient flow and entropy inequalities for quantum Markov semigroups with detailed balance. *J. Funct. Anal.*, 273(5):1810–1869, 2017.
[5] E. A. Carlen and J. Maas. Non-commutative Calculus, Optimal Transport and Functional Inequalities in Dissipative Quantum Systems. *J. Stat. Phys.*, 178(2):319–378, 2020.
[6] H. Dietert. Characterisation of gradient flows on finite state Markov chains. *Electron. Commun. Probab.*, 20:no. 29, 8, 2015.
[7] J. Maas. Gradient flows of the entropy for finite Markov chains. *J. Funct. Anal.*, 261(8):2250–2292, 2011.
[8] A. Mielke. A gradient structure for reaction-diffusion systems and for energy-drift-diffusion systems. *Nonlinearity*, 24(4):1329–1346, 2011.
[9] M. Mittnenzweig and A. Mielke. An entropic gradient structure for Lindblad equations and couplings of quantum systems to macroscopic models. *J. Stat. Phys.*, 167(2):205–233, 2017.

Shape Optimisation for nonlocal anisotropic energies

Lucia Scardia
(joint work with R. Cristoferi, M.G. Mora)

In this work we consider shape optimisation problems for sets of prescribed mass, where the driving energy functional is nonlocal and *anisotropic*. More precisely, for a given mass $m > 0$, we are interested in the minimisation of the energy functional

$$\mathcal{I}(\Omega) = \int_\Omega \int_\Omega \left(W(x-y) + \frac{1}{2}|x-y|^2 \right) dxdy, \tag{1}$$

over the class of sets with mass m,

$$\mathcal{A}_m = \left\{ \Omega \subset \mathbb{R}^d : \ \Omega \ \text{measurable}, \ |\Omega| = m \right\},$$

for $d = 2, 3$. In (1), the interaction potential W is defined for $x \neq 0$ as

$$W(x) = \begin{cases} -\log|x| + \kappa\left(\dfrac{x}{|x|}\right) & \text{if } d = 2, \\ \dfrac{1}{|x|}\kappa\left(\dfrac{x}{|x|}\right) & \text{if } d = 3, \end{cases} \tag{2}$$

and $W(0) = +\infty$. For the profile $\kappa : \mathbb{S}^{d-1} \to \mathbb{R}$ we require that it is even, and that both W and $\widehat{W}$ are continuous on $\mathbb{S}^{d-1}$. Additionally, if $d = 3$, κ is assumed to be strictly positive on $\mathbb{S}^{d-1}$. The potential W is an anisotropic extension of the classical, radially symmetric Coulomb potential, which corresponds to the special case of a constant profile κ. The anisotropy is fully encoded in the profile κ, which introduces an additional dependence on the directions of interaction.

The energy $\mathcal{I}$ is the sum of two competing terms: an attractive, quadratic interaction, that dominates at large distances, and a repulsive, Coulomb-like interaction, driven by the anisotropic potential W. The additional positivity requirement for κ in the three-dimensional case is there to preserve the repulsive nature of W; this is not needed for $d = 2$ since κ is bounded, and hence at short range the repulsive nature of $-\log|\cdot|$ is not affected by the additional anisotropy κ.

1. Main result

Our main result is the characterisation of the minimiser of $\mathcal{I}$ in the class of sets $\mathcal{A}_m$, for any mass $m > 0$. This is done under the sole assumption that the Fourier transform $\widehat{W}$ of the potential W on the sphere $\mathbb{S}^{d-1}$ is nonnegative.

In fact we have two main results, depending on whether $\widehat{W}$ is strictly positive or not. In the first case we show that above a given threshold for the mass the unique minimiser of $\mathcal{I}$ is a d-dimensional ellipsoid. Uniqueness has to be intended up to translations, since the functional $\mathcal{I}$ is translation-invariant.

In the case of degeneracy of $\widehat{W}$, instead, we have the following dichotomy: either there exists a threshold value for the mass as in the case above, or the minimiser is an ellipsoid for any positive value of the mass.

The occurrence of one or the other possibility is related to the minimisation problem for the energy (1) in the wider class of measures (rather than sets) with prescribed mass (see [2, 3, 5]).

2. Method of proof

For the proof of existence, we consider the relaxed energy

$$\mathcal{I}(\rho) = \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \left(W(x-y) + \frac{1}{2}|x-y|^2 \right) \rho(x)\rho(y)\, dxdy, \tag{3}$$

which extends (1) to the class of densities

$$\mathcal{A}_{m,1} = \left\{ \rho \in L^1(\mathbb{R}^d) \cap L^\infty(\mathbb{R}^d) : \ \|\rho\|_{L^1} = m,\ 0 \le \rho \le 1 \text{ a.e.} \right\}. \tag{4}$$

It was proved in [1] that a set $\Omega \in \mathcal{A}_m$ is a minimiser of (1) if and only if its characteristic function $\chi_\Omega \in \mathcal{A}_{m,1}$ is a minimiser of the relaxed energy (3), and the same holds true in our case. Since for small mass m the minimising densities are not the characteristic functions of a set, our original problem on sets can only have a solution for large enough mass.

For large mass we then (equivalently) study the problem on densities, for which existence and compact support of minimisers can be proved by standard arguments. Uniqueness, up to translations, follows by the sign condition on the Fourier transform of W, which implies that the energy $\mathcal{I}$ is strictly convex (on measures with barycentre at the origin). Strict convexity of the energy, in its turn, guarantees that the minimiser can be characterised as the only solution of the Euler-Lagrange optimality conditions.

Motivated by the results in [2,3] and [5] we look for a candidate ellipsoid $E \subset \mathbb{R}^d$ centred at the origin, with $|E| = m$, such that its characteristic function χ_E satisfies the Euler-Lagrange conditions

$$(W * \chi_E)(x) + m\frac{|x|^2}{2} = \lambda \qquad \text{if } x \in \partial E, \tag{5}$$

$$(W * \chi_E)(x) + m\frac{|x|^2}{2} \le \lambda \qquad \text{if } x \in E^\circ, \tag{6}$$

$$(W * \chi_E)(x) + m\frac{|x|^2}{2} \ge \lambda \qquad \text{if } x \in \mathbb{R}^d \setminus E, \tag{7}$$

for a constant $\lambda \in \mathbb{R}$. To evaluate the potential of a generic ellipsoid E we use the representation of the potential in Fourier form proved in [5, 6] for $d = 2, 3$. Following [5, 6] one can see that condition (7) is automatically satisfied by any solution E of (5)–(6).

The key idea to solve (5)–(6) is to rewrite (5) as the stationarity condition for an auxiliary scalar function f defined on symmetric and positive definite matrices M (encoding the information on the semi-axes and orientation of E), under the determinant constraint $\det M = \frac{m^2}{|B|^2}$ (encoding the mass constraint $|E| = m$). One of the main advantages of this alternative formulation is that (6) corresponds to a condition on the sign of the Lagrange multiplier associated to the constraint.

The strategy is then to first show that the auxiliary minimisation problem for f obtained by replacing the equality constraint for the determinant with the unilateral condition $\det M \geq \frac{m^2}{|B|^2}$ admits a solution. As a final step we show that this solution in fact satisfies the equality constraint. This immediately gives the required sign condition for the multiplier, and concludes the proof of (5)–(6).

2.1. **Motivation and comparison with the radially symmetric case.** The problem we consider can be interpreted as a first shape optimisation result for nonlocal anisotropic energies with competing attractive and repulsive terms.

The isotropic counterpart of this problem is well-studied. The closest analogue to our energy $\mathcal{I}$ is the energy considered in [1, 4], namely

$$\mathcal{E}(\Omega) = \int_\Omega \int_\Omega K(x-y)\,dxdy, \quad \Omega \in \mathcal{A}_m, \tag{8}$$

where K is a power-law potential of the form

$$K(x) = \frac{|x|^q}{q} - \frac{|x|^p}{p}, \quad -d < p < 0, \quad q > 0. \tag{9}$$

In the special case of Coulomb repulsion and quadratic attraction there is a threshold for the mass, given by the volume of the unit ball B, such that the energy $\mathcal{E}$ admits no minimiser if $m < |B|$, while for $m \geq |B|$ the minimiser of $\mathcal{E}$ is a ball of mass m. While this is similar to our main result, the corresponding proofs are substantially different. In particular, the radial symmetry of the interactions in $\mathcal{E}$ allows immediately to identify a (unique) ball as the candidate minimiser and greatly simplifies the proof.

Another important class of isotropic attractive/repulsive energies is given by

$$\mathcal{E}_{\mathrm{P}}(\Omega) = \int_\Omega \int_\Omega \frac{1}{d-2}\frac{1}{|x-y|^{2-d}}\,dxdy + \mathrm{Per}(\Omega), \quad \Omega \in \mathcal{A}_m,$$

where Per denotes the classical perimeter. The energies $\mathcal{E}_{\mathrm{P}}$ have been first introduced by Gamow in his liquid drop model and widely studied since.

Considering an anisotropic analogue of $\mathcal{E}_{\mathrm{P}}$ is a very natural direction of investigation, but this is not a direction we will pursue in this work.

References

[1] A. Burchard, R. Choksi, I. Topaloglu, *Nonlocal shape optimization via interactions of attractive and repulsive potentials*, Indiana Univ. Math. J. **67** (2018), 375–395.

[2] J.A. Carrillo, R. Shu, *Global minimizers of a large class of anisotropic attractive-repulsive interaction energies in 2D*, Comm. Pure Appl. Math. **77** (2024), 1353–1404.

[3] J.A. Carrillo, R. Shu, *Minimizers of 3D anisotropic interaction energies*, Preprint *arXiv:2206.140546*.

[4] R. Frank, E. Lieb, *Proof of spherical flocking based on quantitative rearrangement inequalities*, Ann. Sc. Norm. Super. Pisa Cl. Sci. **22**(3) (2021), 1241–1263.

[5] J. Mateu, M.G. Mora, L. Rondi, L. Scardia, J. Verdera, *Explicit minimisers for anisotropic Coulomb energies in 3D*, Adv. Math. **434** (2023), no.109333, 28pp.

[6] M.G. Mora, *Nonlocal anisotropic interactions of Coulomb type*, Preprint *arXiv:2305.14474*.

Existence and uniqueness in law for some doubly nonlinear SPDEs

Ulisse Stefanelli
(joint work with Carlo Orrieri, Luca Scarpa)

Assume to be given a stochastic basis $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ satisfying the usual conditions, a bounded Lipschitz domain $\mathcal{O} \subset \mathbb{R}^d$, and a cylindrical Wiener process W on $L^2(\mathcal{O}) =: H$. The progressive σ-algebra associated with $(\mathcal{F}_t)_{t\geq 0}$ is indicated by $\mathcal{P}$. We are interested in the doubly nonlinear parabolic SPDE

$$\mathrm{d}u = (\partial_t u^d)\mathrm{d}t + G\,\mathrm{d}W, \tag{1a}$$

$$\alpha(\partial_t u^d) - \Delta u + \beta(u) = f(u). \tag{1b}$$

Relation (1a) entails that u^d is the absolutely continuous part of the Ito process u. The Hilbert-Schmidt operator $G \in \mathcal{L}^2(H, H)$ is given. Moreover, α, $\beta \in C^0(\mathbb{R})$ are nondecreasing, with β bounded, $\alpha^{-1} \in C^{0,\eta}(\mathbb{R})$ for some $\eta \in (0, 1)$, and $r \in \mathbb{R} \mapsto \alpha^{-1}(r) - \kappa r$ bounded for some $\kappa > 0$. This in particular entails that α is nondegenerate: there exists $c > 0$ such that $\alpha(r)r \geq cr^2 - 1/c$ for all $r \in \mathbb{R}$. Eventually, we ask $f \in W^{1,\infty}(\mathbb{R})$.

Equation (1b) is posed in the space-time cylinder $\mathcal{O} \times (0, \infty)$ and is complemented with boundary and initial conditions

$$u = 0 \quad \text{on } \partial\mathcal{O} \times (0, \infty), \tag{1c}$$

$$u(\cdot, 0) = u^0 \quad \text{in } \mathcal{O}. \tag{1d}$$

In particular, the Laplacian $-\Delta$ in (1b) is seen as an unbounded, linear, selfadjoint operator in H with domain $D(-\Delta) := H^2(\mathcal{O}) \cap H^1_0(\mathcal{O})$. Via spectral decomposition, one classically defines the powers $(-\Delta)^\sigma$ and the corresponding domains $D((-\Delta)^\sigma)$ for any $\sigma > 0$. Eventually, we ask G to have $\ker G = \{0\}$ and to commute with $-\Delta$. Problem (1) is a concrete example for the abstract theory developed in [4,5], where indeed the above assumptions are somewhat generalized.

In the deterministic case $G \equiv 0$, a general well-posedness theory covering (1) has been obtained by Akagi [2], see also Colli & Visintin [3] for the unperturbed case of $f \equiv 0$. Solutions to the deterministic problem are unique for β Lipschitz continuous and α strongly monotone. In case α is not strongly monotone, uniqueness may fail, also for $\beta = 0$, see [1].

In the stochastic case $G \not\equiv 0$, we are able to give two distinct results, relating to two different regularity setting for the initial datum u^0 and the noise G. Correspondingly, we consider two different type of solutions to (1), both of probabilistic-weak type. For more regular data, we focus on analytically strong solutions. In the less regular setting, we resort to Friedrichs weak solutions instead. We record our findings in the two theorems below.

Theorem 1 (More regular setting)**.** *Let $u^0 \in H^1_0(\mathcal{O})$ and $G \in \mathcal{L}^2(H, H^1(\mathcal{O}))$. Then, there exists an* analytically strong solution *to* (1)*, namely,*

$$(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P}, W, u, u^d)$$

where $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ is a filtered probability space, W is a cylindrical Wiener process on H, and, for all $T > 0$, $u \in L^2_{\mathcal{P}}(\Omega; C^0([0,T]; H) \cap L^\infty(0,T; H^1_0(\mathcal{O})) \cap L^2(0,T; H^2(\mathcal{O})))$, and $u^d \in L^2_{\mathcal{P}}(\Omega; H^1(0,T; H))$ solve

$$u(t) = u^0 + \int_0^t \partial_t u^d(s)\,\mathrm{d}s + \int_0^t G\,\mathrm{d}W(s) \quad \text{a.e. in } \mathcal{O},\ \forall t \geq 0,\ \mathbb{P}-a.s.,$$
$$\alpha(\partial_t u^d) - \Delta u + \beta(u) = f(u) \quad \text{a.e. in} \quad \mathcal{O}\times(0,+\infty),\ \mathbb{P}-a.s.$$

In addition, if α is strongly monotone and β is Lipschitz continuous we have that analytically strong solutions are unique and $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ and W can be a-priori chosen.

This result is proved in [5] by means of an approximation procedure: for all $\lambda > 0$ one solves the approximate problem

$$\mathrm{d}u_\lambda = (\partial_t u^d_\lambda)\mathrm{d}t + G\,\mathrm{d}W, \tag{2a}$$
$$\lambda\partial_t u^d + \alpha_\lambda(\partial_t u^d) + B_\lambda u = f(u), \tag{2b}$$

along with the boundary and initial conditions (1c)-(1d). In (2b), B_λ is the Yosida approximation of $-\Delta+\beta$ at level $\lambda > 0$: for all $u \in H$ we define $B_\lambda u = (u - v_\lambda)/\lambda$, where $v_\lambda \in H^2(\mathcal{O}) \cap H^1_0(\mathcal{O})$ is the unique solution to $v_\lambda - \lambda\Delta v_\lambda + \lambda\beta(v_\lambda) = u$ a.e. in $\mathcal{O}$. The well-posedness of (2) for all $\lambda > 0$ follows from the Lipschitz continuity of $(\lambda\,\mathrm{id} + \alpha_\lambda)^{-1}$, B_λ, and f. One derives λ-independent estimates on u_λ and u^d_λ, extracts suitably converging subsequences, and passes to the limit $\lambda \to 0$ in (2a)-(2b) obtaining an analytically strong solution. Pathwise uniqueness in case α is strongly monotone and β is Lipschitz is straightforward. Based on such uniqueness, the possibility of a-priori fixing $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ and W follows by classical arguments.

Before moving on, let us remark that (1) can be equivalently rewritten as

$$\mathrm{d}u - \kappa\Delta u\,\mathrm{d}t + \kappa\beta(u)\,\mathrm{d}t = \kappa f(u)\,\mathrm{d}t + C(u)\,\mathrm{d}t + G\,\mathrm{d}W$$

where the operator $C : H^2(\mathcal{O}) \cap H^1_0(\mathcal{O})$ is defined as

$$C(u) = \alpha^{-1}(f(u) + \Delta u - \beta(u)) - \kappa(f(u) + \Delta u - \beta(u)).$$

Note that the range of C is bounded in H. For all $\lambda > 0$, we also define $C_\lambda(u) = \alpha^{-1}(f(u) - B_\lambda u) - \kappa(f(u) - B_\lambda u)$ for all $u \in H$.

Theorem 2 (Less regular setting)**.** *Let $u^0 \in H$ and $G(H) \subset D((-\Delta)^\sigma)$ for some σ with*

$$0 < \sigma < \min\left\{\frac{\eta}{4-2\eta}, \frac{1}{6}\right\}. \tag{3}$$

Then, for any sequence $\lambda_n \to 0$ there exists a Friedrichs-weak solution *to (1), namely, $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P}, W, u, y)$ where $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ is a filtered probability space satisfying the usual conditions, W is a cylindrical Wiener process on H, and, for all $T > 0$, $u \in L^2_{\mathcal{P}}(\Omega; C^0([0,T]; H) \cap L^2(0,T; H^1_0(\mathcal{O})))$, and $y \in L^\infty_{\mathcal{P}}(\Omega \times (0,T); H)$ such that there exists a not relabeled subsequence λ_n, a sequence of data*

$(u_n^0, G_n) \in H_0^1(\mathcal{O}) \times \mathcal{L}^2(H, H_0^1(\mathcal{O}))$, and a sequence of analytically strong solutions to

$$\mathrm{d}u_n - \kappa\Delta u_n \,\mathrm{d}t + \kappa\beta(u_n)\,\mathrm{d}t = \kappa f(u_n) + C_{\lambda_n}(u_n)\,\mathrm{d}t + G_n\,\mathrm{d}W$$

with boundary condition (1c) *and initial condition $u_n(\cdot, 0) = u_n^0$ such that, for all $T > 0$,*

$$u_n^0 \to u^0 \quad \text{in } H, \quad \lambda_n^{-1/2-\sigma}\|G_n - G\|_{\mathcal{L}^2(H,H)} \to 0,$$

$$u_n \overset{*}{\rightharpoonup} u \quad \text{in } L^2_{\mathcal{P}}(\Omega; L^\infty(0,T;H) \cap L^2(0,T;H_0^1(\mathcal{O}))),$$

$$u_n \to u \quad \text{in } L^2(0,T;H) \quad \mathbb{P}\text{-a.s.}, \quad C_{\lambda_n}(u_n) \overset{*}{\rightharpoonup} y \quad \text{in } L^\infty_{\mathcal{P}}(\Omega \times (0,T); H).$$

Two such Friedrichs-weak solutions u_1 and u_2 with the same $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$, W, and initial datum u^0 coincide in law in $C^0([0,T];H)$, namely,

$$\mathbb{E}[g(u_1(t))] = \mathbb{E}[g(u_2(t))] \quad \forall g \in C_b(H),\ \forall t \geq 0, \tag{4}$$

where $C_b(H)$ are the bounded continuous functions on H and $\mathbb{E}$ denotes the expectation w.r.t. $\mathbb{P}$.

This result is proved in [4], as a subcase of a more general abstract theory. The existence statement follows by an approximation argument via analytically strong solutions. The uniqueness-in-law statement results from the analysis of the associated Kolmogorov equation. For given $\gamma > 0$ and $g \in C_b(H)$ one considers

$$\gamma\varphi(h) - \frac{1}{2}\mathrm{Tr}(G^*GD^2\varphi(h)) = g(h) + \int_{\mathcal{O}} (\Delta h - \beta(h) + f(h))D\varphi(h)\,\mathrm{d}x,$$

to be solved for all $h \in H^2(\mathcal{O}) \cap H_0^1(\mathcal{O})$. Assuming $\varphi \in C_b^2(H)$ to solve such Kolmogorov equation, an application of the Ito formula to $\varphi \circ u$ on the time interval $[0,t]$ and a limit for $t \to \infty$ formally entail that

$$\int_0^\infty \mathrm{e}^{-\gamma s}\mathbb{E}[g(u(s))]\,\mathrm{d}s = \varphi(u^0).$$

By establishing the latter for any Friedrichs-weak solution, as $\gamma > 0$ is arbitrary, basic properties of the Laplace transform and the a.s. continuity of $g \circ u$ imply (4). In order to make the above argument rigorous, one has to argue at the approximate λ_n level, where the corresponding Kolmogorov equation is solved by φ_{λ_n}. Then, one proves estimates on φ_{λ_n} and its derivatives, applies the Ito formula to $\varphi_{\lambda_n} \circ u_n$, and passes to the limit. In the process, the qualification (3) on σ is used.

References

[1] G. Akagi, *On some doubly nonlinear parabolic equations*, GAKUTO International Series, Mathematical Sciences and Applications, Gakko-Tosho, vol. 32 (2010), pp. 239–254.

[2] G. Akagi, *Doubly nonlinear evolution equations with non-monotone perturbations in reflexive Banach spaces*, J. Evol. Equ. **11** (2011),1:1–41.

[3] P. Colli, A. Visintin, *On a class of doubly nonlinear evolution equations* Comm. Partial Differential Equations **15** (1990), 5:737–756.

[4] C. Orrieri, L. Scarpa, U. Stefanelli, *Stability by noise for a class of doubly nonlinear stochastic evolution equations*, in preparation (2024).

[5] L. Scarpa, U. Stefanelli, *Doubly nonlinear stochastic evolution equations II*, Stoch. Partial Differ. Equ. Anal. Comput. **11** (2023), 1:307–347. *Correction*, ibid. **11** (2023), 4:1740–1743.

Quantum computing with Rydberg-atom quantum processors: A personal journey

Oliver Tse

(joint work with Robert de Keijzer, Servaas Kokkelmans, Luke Visser)

Recent advancements in Rydberg atoms along with the spectacular degree of experimental control of state-of-the-art platforms have made it possible to realize quantum gates with high fidelity, thereby drawing the advent of universal quantum computers closer to reality. Yet, quantum computing is still in the so-called noisy intermediate-scale quantum (NISQ) era, where the number of available error-free qubits is modest, and quantum algorithms have yet to outperform their classical counterparts in practice.

This talk introduces and reports on progress in the following research topics:

Variational Quantum Optimal Control (VQOC). The development of hybrid and near-term quantum algorithms, such as Variational Quantum Eigensolvers (VQEs) based on digital quantum circuits, has been progressing at an enormous pace to allow for quantum advantage in the NISQ era. This development, however, has mostly been independent of the developments in quantum computing hardware, where the physical control of qubits in Rydberg systems is governed by inherently analog laser pulses. In this talk, we introduce the VQOC framework, which brings together recent progress in the understanding and control of Rydberg platforms and the well-developed theory of quantum optimal control, and show applications of VQOC on examples related to the electronic structure problem.

Learning quantum channels. The state of a closed quantum system evolves under the Schrödinger equation, where the reversible evolution of the state is propagated from initial time by an action of a unitary operator. However, realistic quantum systems are open, i.e. they interact with their environment, resulting in non-reversible evolutions, described by quantum semigroups on density matrices. To simulate an open quantum system using an *ideal* quantum computer, which is intrinsically closed, thus requires one to model an open quantum system with a closed one. We do this by invoking the Stinespring dilation theorem, allowing us to learn a target quantum semigroup by approximating equivalent unitary evolutions on an extended system. We further report on an experimentally feasible method to extrapolate the quantum evolution at later times using only data from the first few time steps.

Towards understanding noisy qubits. Noise on a controlled quantum system is generally introduced via the non-reversible Lindblad equation. This equation describes the average state of the system via the density matrix. One way of deriving this Lindblad equation is by taking a sample average of states evolving under the stochastic Schrödinger equation (SSE) driven by white noise. However, white noise, where all noise frequencies contribute equally in the power spectral density, is not a realistic noise profile as lower frequencies commonly dominate the

spectrum. For this reason, we provide analytical solutions to the full fidelity distribution for important cases of the SSE driven by more realistic noise. This allows for predictions of the mean, variance, and higher-order moments of the fidelities of these qubits, which can be of value when deciding on the allowed noise levels for future quantum computing systems, e.g. deciding what quality of control systems to procure. Furthermore, these methods will prove to be integral in the optimal control of qubit states under (classical) control system noise.

References

[1] R. de Keijzer, S. Kokkelmans and O. Tse, *Pulse-based variational quantum optimal control for hybrid quantum computing*, Quantum **7** (2023), 908.
[2] R. de Keijzer, S. Kokkelmans, O. Tse and L. Visser, *Variational method for learning quantum channels via Stinespring dilation on neutral atom systems*, arXiv:2309.10593.

Discrete to continuous crystalline curvature flows

Antonin Chambolle
(joint work with Daniele DeGennaro, Massimiliano Morini)

The talk described a work in progress, where we investigate a fully discrete version of the Almgren-Taylor-Wang / Luckhaus-Sturzenhecker scheme [1, 7] for building mean curvature flows. This scheme, after some rewriting, can be described as follows: given a set E^0, and d_{E^0} the *signed distance function* to its boundary, we solve in $\mathbb{R}^d$, for $h > 0$ small and each $n \geq 1$:

$$\begin{cases} -h\mathrm{div}\, z^n + u^n = d_{E^{n-1}}\,, \\ |z^n| \leq 1, \quad z^n \cdot Du^n = |Du^n| \end{cases} \tag{1}$$

which is formally the Euler-Lagrange equation of

$$\min_u \int |Du| + \frac{1}{2h}\int (u - d_{E^{n-1}})^2 dx$$

(yet this energy is infinite in the whole space). We let then $E^n = \{x : u^n(x) \leq 0\}$. By translational invariance and comparison, u^n is trivially 1-Lipschitz (since $d_{E^{n-1}}$ is), in particular the second condition in (1) reads $z \in \partial|\cdot|(\nabla d)$ a.e. in $\mathbb{R}^d$ (the subgradient of the Euclidean norm). One also deduces that $d_{E^n} \geq u^n$ in $\{u^n > 0\}$, and $d_{E^n} \leq u^n$ in $\{u^n < 0\}$. Hence,

$$\frac{d^n - d^{n-1}}{h} \geq \mathrm{div}\, z^n$$

out of E^n. Getting some control on d^n in time and $\mathrm{div}\, z^n$ in space allows then to pass to the limit and deduce the existence of $E \subset \mathbb{R}^d \times [0, \infty)$ (the Hausdorff limit of $\bigcup_{n\geq 0} E^n \times \{nh\}$) a closed set such that

$$\partial_t d \geq \mathrm{div}\, z \quad \text{in} \quad \mathcal{D}'((\mathbb{R}^d \times (0,\infty)) \setminus E), \qquad z \in \partial|\cdot|(\nabla_x d) \text{ a.e.} \tag{2}$$

with $d(x,t) = \mathrm{dist}(x, E(t))$ for all x, t. Reasoning with the complement, one finds a similar equation for $A \subset E$, the complement of the Hausdorff limit of $\bigcup_{n\geq 0}(\mathbb{R}^d \setminus E^n) \times \{nh\}$).

This equation, which holds in the distributional or measure sense, is seen to hold also in the viscosity sense [5, 6] and hence characterizes the mean curvature flow (with a possible, but exceptional, fattening of the set $E \setminus A$), as shown in [8]. An important step in proving the convergence is an estimate of the solution of (1) with the right-hand side replaced with $|x|$, first computed in [3], this is crucial to estimate the variation of d^n in time as well as $\mathrm{div}\, z^n$ from above where $d^n > 0$.

Now, in [4, 6], it is also shown that the same scheme (and the same proof) can be applied to build and characterize anisotropic, or crystalline flows. Sticking to the simpler case of [6], and given φ a convex norm (with possibly polyhedral level sets) we replace Du above with $\varphi(Du)$, the distance with the φ° distance ($\varphi^\circ(x) = \sup\{x\cdot\nu : \varphi(\nu) \leq 1\}$), the condition on z^n in (1) with $z^n \in \partial\varphi(\nabla u^n)$ and end up with a distributional definition of a well posed *crystalline* mean curvature flow (see [6] for a comparison result which guarantees the uniqueness, in general, of the limit—when A is the interior of E). In this case, the motion is still described by (2) yet the second condition is $z \in \partial\varphi(\nabla_x d)$ and d is the φ°-distance to E.

In this work, with M. Morini (Parma) and our student D. DeGennaro (Ceremade), we propose a to solve a fully discrete equation, which reads (for $h, \varepsilon > 0$, small time and space steps)

$$\begin{cases} h(D^* z^n)_i + u_i^n = d_i^{n-1} & \text{for } i \in \varepsilon\mathbb{Z}^d \\ |z_{i,j}^n| \leq \beta_{\frac{j-i}{h}}\,, \quad z_{i,j}^n(u_j - u_i) = \beta_{\frac{j-i}{h}}|u_j - u_i| \end{cases}$$

where $D : \mathbb{R}^{\mathbb{Z}^d} \to \mathbb{R}^{\mathbb{Z}^d\times\mathbb{Z}^d}$ is defined by $(Du)_{i,j} = (u_j - u_i)/h$, D^* is its adjoint (for the standard scalar product), and β_k, $k \in \mathbb{Z}^d$, is a finitely supported family of positive weights (positive at least on a basis of $\mathbb{Z}^d$). One then sends $h, \varepsilon \to 0$. In case $\varepsilon \ll h$, but more interestingly and somewhat surprisingly in case $\varepsilon = h$, one may then adapt the techniques above to show again the convergence to (2), for the crystalline anisotropy $\varphi(p) = \sum_{k\in\mathbb{Z}^d} \beta_k |k \cdot p|$. Interestingly, in case $\varepsilon = h$, we may define d^n from u^n with a sort of interpolation scheme (defined by suitable inf/sup convolutions with the distance φ°), so that the limiting evolution is precisely given by (2) without any drift, contrarily to the dicrete scheme introduced previously in [2], which was the starting point for our study, and where a rounding occurs at each step which accumulates in the limit.

References

[1] F. Almgren, J. E. Taylor, and L.-H. Wang. Curvature-driven flows: a variational approach. *SIAM J. Control Optim.*, 31(2):387–438, 1993.

[2] Andrea Braides, Maria Stella Gelli, and Matteo Novaga. Motion and pinning of discrete interfaces. *Arch. Ration. Mech. Anal.*, 195(2):469–498, 2010.

[3] Vicent Caselles and Antonin Chambolle. Anisotropic curvature-driven flow of convex sets. *Nonlinear Anal.*, 65(8):1547–1577, 2006.

[4] Antonin Chambolle, Massimiliano Morini, Matteo Novaga, and Marcello Ponsiglione. Existence and uniqueness for anisotropic and crystalline mean curvature flows. *J. Amer. Math. Soc.*, 32(3):779–824, 2019.
[5] Antonin Chambolle, Massimiliano Morini, Matteo Novaga, and Marcello Ponsiglione. Generalized crystalline evolutions as limits of flows with smooth anisotropies. *Anal. PDE*, 12(3):789–813, 2019.
[6] Antonin Chambolle, Massimiliano Morini, and Marcello Ponsiglione. Existence and uniqueness for a crystalline mean curvature flow. *Comm. Pure Appl. Math.*, 70(6):1084–1114, 2017.
[7] Stephan Luckhaus and Thomas Sturzenhecker. Implicit time discretization for the mean curvature flow equation. *Calc. Var. Partial Differential Equations*, 3(2):253–271, 1995.
[8] Halil Mete Soner. Motion of a set by the curvature of its boundary. *J. Differential Equations*, 101(2):313–372, 1993.

Coarse-graining to GENERIC

Johannes Zimmer
(joint work with Alexander Mielke, Mark A. Peletier)

We study a model system on two different scales, called microscopic and macroscopic, with the aim of deriving the macroscopic description from the microscopic one.

The *macroscopic* description is governed by a thermodynamic formulation in form of GENERIC, the *General Equation for the Non-Equilibrium Reversible-Irreversible Coupling* [3], sometimes also called metriplectic evolution. It takes the form

$$\dot{y} = \mathbb{J}(y)D\mathcal{E}(y) + \mathbb{K}(y)D\mathcal{S}(y); \tag{1}$$

here $y \in \mathbf{Y}$ is the macroscopic variable defined on a state space $\mathbf{Y}$, $\mathbb{J}\colon \mathbf{Y}^* \to \mathbf{Y}$ is a symplectic operator, $\mathbb{K}\colon \mathbf{Y}^* \to \mathbf{Y}$ is a positive semidefinite operator, $\mathcal{E}, \mathcal{S}\colon \mathbf{Y} \to \mathbb{R}$ are the energy and entropy functionals and D denotes a derivative. The structure of GENERIC immediately implies that the energy $\mathcal{E}$ is constant along trajectories and the entropy $\mathcal{S}$ is non-decreasing. Thus, GENERIC ensures thermodynamic consistency, and very different thermodynamic systems can be put in this framework [2].

The aim is to derive (1) as macroscopic description of a microscopic model, i.e., give a mathematically rigorous coarse-graining procedure. We choose a classic microscopic system, which has been investigated in detail [1], though to our knowledge not in connection with GENERIC. The model is purely Hamiltonian, to reflect the description of atoms and molecules by Newtonian mechanics. It consists of a finite-dimensional Hamiltonian system (System $\mathcal{A}$) coupled to an infinite-dimensional heat bath (System $\mathcal{B}$) via a coupling $\mathcal{C}$. Both System $\mathcal{B}$ and the coupling are linear, while system $\mathcal{A}$ can be nonlinear. Specifically, with $z \in \mathbf{Z}$ denoting the state of system $\mathcal{A}$ and $\eta \in \mathbf{H}$ denoting the state of system $\mathcal{B}$, the total Hamiltonian of the microscopic system is

$$\mathcal{H}_{\text{total}}(z, \eta) = \mathcal{H}_{\mathcal{A}}(z) + \mathcal{H}_{\mathcal{B}}(\eta) + \mathcal{H}_{\mathcal{C}}(z, \eta), \tag{2}$$

where the Hamiltonian for the heat bath is

$$\mathcal{H}_{\mathcal{B}}(\eta) = \frac{1}{2}\|\eta\|_{\mathbf{H}}^2 \quad \text{for all } \eta \in \mathbf{H},$$

and the coupling is described by the Hamiltonian

$$\mathcal{H}_{\mathcal{C}} = (\mathbb{A}z|\mathbb{P}\eta)_{\mathbf{H}};$$

here $(\cdot|\cdot)_{\mathbf{H}}$ is the inner product in $\mathbf{H}$, $\mathbb{A}\colon \mathbf{Z} \to \mathbf{H}$ is a linear embedding operator and $\mathbb{P}$ is an orthogonal projection operator discussed later. Below, we write $\mathbb{B} := \mathbb{P}\mathbb{A}$.

One can show that, due to the linearity of the heat bath, the corresponding evolution satisfies for $t \in \mathbb{R}$

$$\dot{z}(t) = \mathbb{J}_{\mathcal{A}}(D\mathcal{H}_{\mathcal{A}}(z(t)) + \mathbb{B}^*\eta(t)), \tag{3a}$$

$$\eta(t) = \mathrm{e}^{\mathbb{J}_{\mathcal{B}}t}\left(\eta(0) + \mathbb{B}z(0)\right) - \mathbb{B}z(t) + \int_0^t \mathrm{e}^{\mathbb{J}_{\mathcal{B}}(t-s)}\mathbb{B}\dot{z}(s)\,\mathrm{d}s, \tag{3b}$$

where $\mathbb{J}_{\mathcal{A}}$ and $\mathbb{J}_{\mathcal{B}}$ are the sympletic operators associated with $\mathcal{H}_{\mathcal{A}}$ and $\mathcal{H}_{\mathcal{B}}$.

So far, we have not specified the initial condition of the heat bath $\mathcal{B}$. While the coupled microscopic system evolves deterministically, we introduce randomness through the initial data $\eta(0)$ for the heat bath. Then (3b) involves a memory term, namely the time integral in the right, and a stochastic term stemming from the initial data. One can show that one can rephrase this equation as a generalized (i.e., non-Markovian) Langevin equation.

Such non-Markovian equations are often encountered in a Mori-Zwanzig reduction procedure, where suitable projections of an infinite-dimensional microscopic system are considered [6]. Often, when dealing with non-Markovian systems, one tries to rephrase them as Markovian ones by augmenting the state space.

The GENERIC equation (1) is Markovian (in the sense that it is a nonlinear semigroup, hence in particular local, i.e., memoryless in time). Abstractly, due to the presence of the dissipative term $\mathbb{K}(y)D\mathcal{S}(y)$, one expects GENERIC systems to be given by a (nonlinear) contraction semigroup. The evolution given by (3), however, corresponds to a unitary group on $\mathbf{Z} \times \mathbf{H}$. To link these two, we use the concept of compressions. *Compressions* can be seen as the 'inverse' of dilations; the theory of dilations provides an embedding of a given contraction semigroup $\mathrm{e}^{-t\mathbb{D}}$ defined on a Hilbert space $\mathbf{W}$ into a strongly continuous unitary group $\mathrm{e}^{t\mathbb{J}}$ defined on a bigger space such that the restriction of the unitary group to $\mathbf{W}$ agrees with the given contraction semigroup,

$$\mathbb{P}\mathrm{e}^{t\mathbb{J}}\big|_{\mathbf{W}} = \begin{cases} \mathrm{e}^{-t\mathbb{D}} & t \geq 0 \\ \mathrm{e}^{t\mathbb{D}^*} & t \leq 0 \end{cases} \qquad \text{on } \mathbf{W}, \tag{4}$$

where $\mathbb{P}$ is the orthogonal projection on $\mathbf{W}$. A minimal dilation is unique up to Hilbert space isomorphism [5, Chapter 1].

The existence of a compression means that given the unitary group $\mathrm{e}^{t\mathbb{J}}$, a subspace $\mathbf{W}$ and an orthogonal projection $\mathbb{P}$ onto $\mathbf{W}$ and a contraction semigroup $\mathrm{e}^{-t\mathbb{D}}$ exist such that (4) holds. For existence of compressions we refer to [4, Section 5].

The coarse-graining to go from the microscopic evolution (3), i.e., the unitary group associated with (2), to a GENERIC evolution, i.e., a macroscopic evolution which can be cast in the general form (1), that involves as key step finding a (finite-dimensional) compression subspace $\mathbf{Y}$ of $\mathbf{Z} \times \mathbf{H}$ such that the compression of the unitary group on $\mathbf{Y}$ is defined. It is natural that the compression subspace $\mathbf{Y}$ contains $\mathbf{Z}$.

Here a compression of the heat bath can be interpreted as reduction to observables, as elements of $\mathbf{W}$. For the coupled system, a compression subspace is given by $\mathbf{Z} \times \mathbf{W}$, where $\mathbf{W} = \mathbb{P}\mathbf{H}$. Thus observables are (z, w) with $w = \mathbb{P}\eta$. A GENERIC evolution can be formulated if this state space is augmented by a suitably defined energy e. The GENERIC form has the following structure. For the contraction semigroup $\mathbb{D}$, we consider the split in symmetric and skew-symmetric parts,

$$\mathbb{D} = \mathbb{D}_{\mathrm{sym}} + \mathbb{D}_{\mathrm{skw}} \quad \text{with } \mathbb{D}_{\mathrm{sym}} = \frac{1}{2}(\mathbb{D}+\mathbb{D}^*) \ \text{ and } \ \mathbb{D}_{\mathrm{skw}} := \frac{1}{2}(\mathbb{D}-\mathbb{D}^*).$$

Then the symplectic operator $\mathbb{J}$ of the GENERIC evolution (1) in $y := (z, w, e)$ involves $\mathbb{J}_{\mathcal{A}}$ and $\mathbb{D}_{\mathrm{skw}}$, while $\mathbb{D}_{\mathrm{sym}}$ enters the positive semi-definite operator $\mathbb{K}$ of GENERIC. The random initial data of the heat bath enters the entropy S and the dissipative operator $\mathbb{K}$. In passing from the stochastic microscopic system via compression to GENERIC, an interim stage is a stochastic version of GENERIC, which includes a noise term Σ. One can show that a fluctuation-dissipation statement linking $\mathbb{K}$ and Σ holds.

References

[1] Vojkan Jakšić and Claude-Alain Pillet. Ergodic properties of classical dissipative systems. I. *Acta Math.*, 181(2):245–282, 1998.

[2] Alexander Mielke. Formulation of thermoelastic dissipative material behavior using GENERIC. *Contin. Mech. Thermodyn.*, 23(3):233–256, 2011.

[3] Hans Christian Öttinger. *Beyond equilibrium thermodynamics*. Wiley Online Library, 2005.

[4] Giorgio Picci. Application of stochastic realization theory to a fundamental problem of statistical physics. In *Modelling, identification and robust control (Stockholm, 1985)*, pages 211–258. North-Holland, Amsterdam, 1986.

[5] Béla Sz.-Nagy, Ciprian Foias, Hari Bercovici, and László Kérchy. *Harmonic analysis of operators on Hilbert space*. Universitext. Springer, New York, enlarged edition, 2010.

[6] Robert Zwanzig. *Nonequilibrium statistical mechanics*. Oxford University Press, New York, 2001.

Analytic properties of the sliced Wasserstein distance

Sangmin Park
(joint work with Dejan Slepčev)

Given two probability measures $\mu, \nu \in \mathscr{P}_2(\mathbb{R}^d) := \{\mu \in \mathscr{P}(\mathbb{R}^d) : \int |x|^2 d\mu(x) < \infty\}$, recall that the 2-Wasserstein distance W_2 between them is defined as follows:

$$W_2(\mu,\nu) := \inf_{\gamma\in\Gamma(\mu,\nu)} \left(\int_{\mathbb{R}^d\times\mathbb{R}^d} |x-y|^2 \, d\gamma(x,y)\right)^{1/2}$$
$$\text{where } \Gamma(\mu,\nu) = \left\{\gamma \in \mathscr{P}(\mathbb{R}^d\times\mathbb{R}^d) : \pi^1_\#\gamma = \mu,\, \pi^2_\#\gamma = \nu\right\}.$$

The sliced Wasserstein distance, introduced by Rabin, Peyré, Delon, and Bernot [4], compares probability measures on $\mathbb{R}^d$ by taking averages of the Wasserstein distances between projections of the measures to each 1-dimensional subspaces of $\mathbb{R}^d$. To be more precise, for each $\theta \in \mathbb{S}^{d-1}$ define the projection $\pi^\theta : \mathbb{R}^d \to \mathbb{R}$ by

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

The 2-sliced Wasserstein distance SW_2 is defined by

$$SW_2(\mu,\nu) = \left(\frac{1}{|\mathbb{S}^{d-1}|}\int_{\mathbb{S}^{d-1}} W_2^2(\pi^\theta_\#\mu, \pi^\theta_\#\nu)\, d\theta\right)^{\frac{1}{2}}$$

where $_\#$ denotes the pushforward of a measure.

Thanks to its lower sample and computational complexities relative to the Wasserstein distance especially in high dimensions, the sliced Wasserstein distance has recently expanded its applications in statistics and machine learning as a tool to compare measures and construct paths in spaces of measures.

In this talk we presented a number of analytic properties of the SW_2 and the sliced Wasserstein length metric ℓ_{SW}, defined as the infimum of the lengths of curves between measures in the SW_2-space. Moreover, we discussed their implications on the sliced Wasserstein gradient flows and statistical estimation rates in the metrics.

Comparison of sliced Wasserstein metric with negative Sobolev norms and Wasserstein metric. To understand the metric properties of the sliced Wasserstein distance, we establish the comparison theorems of SW_2 with negative Sobolev norms near absolutely continuous measures and comparisons of SW_2 with the Wasserstein metric W_2 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)
$$\|\mu-\nu\|_{\dot{H}^{-(d+1)/2}(\mathbb{R}^d)} \lesssim SW_2(\mu,\nu) \le \ell_{SW}(\mu,\nu) \lesssim SW_2(\mu,\nu) \lesssim \|\mu-\nu\|_{\dot{H}^{-(d+1)/2}(\mathbb{R}^d)},$$

where the rightmost inequality additionally requires ν to coincide with μ near the boundary of Ω. In other words, near μ, SW_2 is equivalent to $\dot{H}^{-(d+1)/2}$.

On the other hand, we show that

$$SW_2(\mu^n,\nu) \le \ell_{SW}(\mu^n,\nu) \le \frac{1}{d} W_2(\mu^n,\nu) \le (1+o(1)) SW_2(\mu^n,\nu) \tag{2}$$

for ν near discrete measures of the form $\mu^n = \sum_{i=1}^n m_i \delta_{x_i}$.

These two results provide interesting insights about the SW_2 measure. Near smooth measures it 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 noted by Peyre [3], while near discrete measures SW_2 behaves like the Wasserstein distance.

Approximation by discrete measures in sliced Wasserstein length. It is known that finite-sample estimation of measures with respect to maximum mean discrepancy (MMD) also enjoys parametric rate [5, 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 [5] 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_2 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_2 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_2 distance. Indeed, Manole, Balakrishnan, and Wasserman [2, Proposition 4] have shown that a finite random sample (i.e. the empirical measure of the set of n random points) of a probability measure on $\mathbb{R}^d$ estimates the measure in the sliced Wasserstein distance at a parametric rate, $\frac{1}{\sqrt{n}}$, for a large class of measures.

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

$$SW_2(\mu,\mu^n) \le \ell_{SW}(\mu,\mu^n) \lesssim \sqrt{\frac{\log n}{n}} \quad \text{with high probability,}$$

where $\mu^n = \frac{1}{n}\sum_{i=1}^n \delta_{X_i}$ with $X_i \overset{i.i.d.}{\sim} \mu$. This is in stark contrast with the Wasserstein distance where the approximation rate is poor in high dimensions and scales like $n^{-\frac{1}{d}}$.

Implications on gradient flows. The comparison results on ℓ_{SW}, SW_2 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

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

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

$$|\partial \mathcal{V}|_{SW_2}(\mu^n) = |\partial \mathcal{V}|_{\ell_{SW}}(\mu^n) = \sqrt{d}\,|\partial \mathcal{V}|_W(\mu^n). \tag{5}$$

Hence $|\partial \mathcal{V}|_{SW_2}$ (resp. $|\partial \mathcal{V}|_{\ell_{SW}}$) is not lower-semicontinuous in SW_2 (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 $(\mathscr{P}_2(\mathbb{R}^d), \ell_{SW})$. Consequently, the curves of maximal slope in the Wasserstein space starting from discrete measures with finite number of particles, after a constant rescaling of time, is the curve of maximal slope in SW_2 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_2 space. We formally show that SW_2 gradient flow of potential energy is a higher order equation given by a pseudodifferential operator of order d, which is consistent with the rigorous results (4).

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] R. Peyre. *Comparison between W_2 distance and $\dot{H}^{-1}$ norm, and localization of Wassesrstein distance* ESAIM:COCV, 24(2018), pp.1489–1501.

[4] 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.

[5] B. Sriperumbudur *On the optimal estimation of probability measures in weak and strong topologies* Bernoulli, 22 (2016), pp.1839–1893.

Entropic propagation of chaos for population dynamics

Jasper Hoeksema

Interacting particle systems where particles can be created and deleted form the backbone of several models in ecology, with the particular example of the Bolker-Pacala-Dieckmann-Law (BPDL) model [1], which was originally introduced to study the evolution of and pattern formation in populations of plants, but which turned out to accurately describe models involving mutation of traits.

Various methods exist to derive mean-field limits for these systems, but these sometimes require stringent assumptions on the interactions and use weak notions of convergence. In this talk, we discuss past [6] and current work to alleviate both these restrictions for weakly interacting birth/death processes where, using techniques inspired by convergence of gradient flows for interacting particle systems [3–5], we prove entropic propagation of chaos for the BPDL model.

We model the particles as a collection of points $\{X_t^{n,1}, \dots, X_t^{n,N_t}\}$ in a compact Polish space $\mathcal{X}$, where the parameter n will control the order of the number of particles in the system. We are interested in convergence in a suitable sense of the rescaled empirical measure $\nu_t^n \in \Gamma := \mathcal{M}(\mathcal{X})$ given by

$$\nu_t^n := \frac{1}{n} \sum_{i=1}^{N_t} \delta_{X_t^{n,i}}.$$

Formally, ν_t^n is a measure-valued jump process with generator

$$Q_n F = n \int_{\mathcal{X}} \left(F(\nu + \tfrac{1}{n}\delta_x) - F(\nu) \right) \chi_\nu^+(dx) + n \int_{\mathcal{X}} \left(F(\nu - \tfrac{1}{n}\delta_x) - F(\nu) \right) \chi_\nu^-(dx)$$

with $F \in C_c(\Gamma)$ and the measure-dependent birth/death rates $\chi_\nu^\pm$, which in the case of the BPDL model looks like

$$\chi_\nu^+(dx) = \left(\int_{\mathcal{X}} m(x,y)\nu(dy) \right) \gamma(dx), \qquad \chi_\nu^-(dx) = \left(\int_{\mathcal{X}} c(x,y)\nu(dy) \right) \nu(dx)$$

where m, c are the mutation and competition kernels, and $\gamma \in \Gamma$ is some reference measure.

It is the corresponding forward Kolmogorov equation that is our object of study. It describes the law of ν_t^n, and for a path of measures $(\mathsf{P}_t^n)_{t\in[0,T]}$ over some time horizon $[0,T]$ it satisfies

$$\partial_t \mathsf{P}_t^n = Q_n^* \mathsf{P}_t^n.$$

After a Taylor expansion of $Q_n F$, with F a suitable cylindrical function, one can expect that under suitable conditions

$$\lim_{n\to\infty} Q_n F = Q_\infty F = \int_{\mathcal{X}} (\nabla_\Gamma F)(\nu, x) V_\nu(dx),$$

where $V_\nu(dx) := \chi_\nu^+(dx) - \chi_\nu^-(dx)$, and would surmise that $\lim_{n\to\infty} \mathsf{P}_t^n = \mathsf{P}_t$ narrowly, where P_t satisfies the corresponding Liouville equation with velocity field V. In particular, if $\mathsf{P}_0^n \to \mathsf{P}_0 := \delta_{\bar\nu_t}$ then one would expect that

$$\mathsf{P}_t^n \to \delta_{\bar\nu_t} \tag{1}$$

where $\bar\nu_t$ satisfies the mean-field equation

$$\partial_t \bar\nu_t = V_{\bar\nu_t}.$$

The convergence (1) is known as *propagation of chaos*, and implies narrow convergence of the corresponding correlation functions. However, we are interested in a strictly stronger notion, called *entropic propagation of chaos*, which implies vanishing relative entropy. For suitable bounds on $\bar\nu_t$ the latter can be shown to be equivalent to the statement

$$\lim_{n\to\infty} \frac{1}{n} \mathcal{E}\mathrm{nt}(\mathsf{P}_0^n | \Pi_n) = \int_\Gamma \mathcal{E}\mathrm{nt}(\nu|\gamma) \mathsf{P}_0(d\nu) \Longrightarrow \lim_{n\to\infty} \frac{1}{n} \mathcal{E}\mathrm{nt}(\mathsf{P}_t^n | \Pi_n) = \int_\Gamma \mathcal{E}\mathrm{nt}(\nu|\gamma) \mathsf{P}_\mathrm{t}(d\nu)$$

for all $t \in [0,T]$, with Π_n a rescaled Poisson point measure induced by γ.

In our work we prove this property, using large deviation techniques for the rescaled entropies, lower semicontinuity of the entropy dissipation, and the Sandier-Serfaty approach [2] to obtain convergence.

A key tool is the fact that the rescaled entropy dissipation, in the reversible setting where $m = c$ and $c(x, x) = 0$ for all $x \in \mathcal{X}$, reduces to (with a slight abuse of notation)

$$\mathcal{K}^n(\mathsf{P}) = 2\mathcal{E}\mathrm{nt}\left(\mathsf{P}(d\nu)\chi^+_\nu(dx)\middle|\mathsf{P}\left(d(\nu+\tfrac{1}{n})\right)\chi^-_{\nu+\frac{1}{n}}(dx)\right),$$

and is related to the convergence of the associated gradient flow structures as shown in [6]. In current work we extend this to the irreversible setting.

References

[1] B. Bolker, S.W. Pacala, *Using Moment Equations to Understand Stochastically Driven Spatial Pattern Formation in Ecological Systems*, Theoret. Population Biol. **52(3)** (1997), 179–197

[2] E. Sandier, S. Serfaty, *Gamma-convergence of gradient flows with applications to Ginzburg-Landau*, Comm. Pure Appl. Math., **57(12)** (2004), 1627–1672

[3] M. Erbar, *A gradient flow approach to the Boltzmann equation*, Journal of the European Mathematical Society (2016)

[4] M. Erbar, M. Fathi, V. Laschos, A. Schlichting, *Gradient flow structure for McKean-Vlasov equations on discrete spaces*, Discrete Contin. Dyn. Syst., **36(12)** (2016), 6799–6833

[5] J. Maas, A. Mielke, *Modeling of chemical reaction systems with detailed balance using gradient structures*. J. Stat. Phys., **181(6)** (2020), 2257–2303

[6] J. Hoeksema, O. Tse, *Generalized gradient structures for measure-valued population dynamics and their large-population limit*, arXiv preprint (2022)

Reporter: Sangmin Park

Participants

Prof. Dr. Andrea Braides
SISSA
and Dipartimento di Matematica
Università di Roma "Tor Vergata"
via Bonomea 265
34133 Trieste
ITALY

Prof. Dr. Yann Brenier
CNRS & LMO, Université Paris-Saclay
Building 307, rue Michel Magat
91405 Orsay Cedex
FRANCE

Prof. Dr. Leon Bungert
Institut für Mathematik
Universität Würzburg
Emil-Fischer-Str. 40
97074 Würzburg
GERMANY

Prof. Dr. Antonin Chambolle
CEREMADE, CNRS, Université Paris Dauphine
PSL Research University
Place de Lattre de Tassigny
75775 Paris Cedex 16
FRANCE

Prof. Dr. Matthias Erbar
Fakultät für Mathematik
Universität Bielefeld
Postfach 100131
33501 Bielefeld
GERMANY

Dr. Antonio Esposito
Mathematical Institute
Oxford University
Andrew Wiles Building
Woodstock Road
Oxford OX2 6GG
UNITED KINGDOM

Dr. Max Fathi
Laboratoire Jacques-Louis Lions
Batiment Sophie Germain
8, Place Aurélie Nemours
75013 Paris Cedex
FRANCE

Prof. Dr. Julian Fischer
Institute of Science and Technology Austria(IST Austria)
Am Campus 1
3400 Klosterneuburg
AUSTRIA

Dr. Jasper Hoeksema
Department of Mathematics and Computer Science
Eindhoven University of Technology
MetaForum 5.060
P.O. Box 513
5600 MB Eindhoven
NETHERLANDS

Dr. Franca Hoffmann
California Institute of Technology
1200 E California Blvd
Pasadena, CA 91125
UNITED STATES

Dr. Katharina Hopf
Weierstraß-Institut für Angewandte Analysis und Stochastik
Mohrenstraße 39
10117 Berlin
GERMANY

Dr. Anastasiia Hraivoronska
Institut Camille Jordan
Université Claude Bernard Lyon 1
43 blvd. du 11 novembre 1918
69622 Villeurbanne Cedex
FRANCE

Prof. Dr. Ansgar Jüngel
Institut für Analysis und Scientific Computing
Technische Universität Wien
Wiedner Hauptstrasse 8 - 10
1040 Wien
AUSTRIA

Prof. Dr. Tim Laux
Fakultät für Mathematik
Universität Regensburg
Universitätsstr. 31
93053 Regensburg
GERMANY

Dr. Jona Lelmi
Department of Mathematics
UCLA
405 Hilgard Ave.
Los Angeles, CA 90095-1555
UNITED STATES

Prof. Dr. Christian Leonard
Université Paris-Nanterre
Bâtiment Maurice Allais (bât. G)
200, Avenue de la République
92000 Nanterre
FRANCE

Prof. Dr. Jan Maas
IST Austria
Am Campus 1
3400 Klosterneuburg
AUSTRIA

Prof. Dr. Daniel Matthes
Department of Mathematics / M8
School of Computation, Information and Technology (SoCIT)
Technische Universität München
Boltzmannstraße 3
85747 Garching bei München
GERMANY

Prof. Dr. Alexander Mielke
Weierstraß-Institut für Angewandte Analysis und Stochastik
Mohrenstraße 39
10117 Berlin
GERMANY

Dr. Nikolas Nüsken
Department of Mathematics
King's College London
Strand
London WC2R 2LS
UNITED KINGDOM

Sangmin Park
Department of Mathematical Sciences
Carnegie Mellon University
Pittsburgh, PA 15213-3890
UNITED STATES

Prof. Dr. Mark A. Peletier
Department of Mathematics and Computer Science
Eindhoven University of Technology
MetaForum 5.062
P.O. Box 513
5600 MB Eindhoven
NETHERLANDS

Dr. D.R. Michiel Renger
Zentrum Mathematik
TU München
Boltzmannstr. 3
85748 Garching bei München
GERMANY

Prof. Dr. Riccarda Rossi
Department of Mechanical and Industrial Engineering
University of Brescia
Via Branze 38
25133 Brescia
ITALY

Prof. Dr. Tomas Roubicek
Mathematical Institute
Charles University
Sokolovska 83
186 75 Praha 8
CZECH REPUBLIC

Prof. Dr. Giuseppe Savaré
Department of Decision Sciences
Bocconi University
Via Roentgen, 1
20136 Milano
ITALY

Dr. Lucia Scardia
Department of Mathematics
Heriot-Watt University
Riccarton
Edinburgh EH14 4AS
UNITED KINGDOM

Prof. Dr. André Schlichting
Institut für Analysis und Numerik
Mathematics Münster
Orléans-Ring 10
48149 Münster
GERMANY

Dr. Nikita Simonov
Sorbonne Université
Laboratoire Jacque-Louis Lions
4, place Jussieu
75005 Paris Cedex 05
FRANCE

Prof. Dr. Dejan Slepčev
Department of Mathematical Sciences
Carnegie Mellon University
5000 Forbes Avenue
Pittsburgh, PA 15213-3890
UNITED STATES

Prof. Dr. Bharath Sriperumbudur
Department of Statistics
Pennsylvania State University
314 Thomas Building
University Park, PA 16802
UNITED STATES

Prof. Dr. Ulisse Stefanelli
Fakultät für Mathematik
Universität Wien
Oskar-Morgenstern-Platz 1
1090 Wien
AUSTRIA

Dr. Artur Stephan
Weierstraß-Institut für
Angewandte Analysis und Stochastik
Mohrenstraße 39
10117 Berlin
GERMANY

Prof. Dr. Marita Thomas
Freie Universität Berlin und
Weierstraß-Institut für
Angewandte Analysis und Stochastik
Arnimallee 9
14195 Berlin
GERMANY

Dr. Oliver Tse
Department of Mathematics and
Computer Science
Eindhoven University of Technology
P.O. Box 513
5600 MB Eindhoven
NETHERLANDS

Dr. Yves van Gennip
Delft Institute of Applied Mathematics
Delft University of Technology
Building 36
2628 CD Delft
NETHERLANDS

Prof. Dr. Max von Renesse
Mathematisches Institut
Universität Leipzig
Augustusplatz 10
04109 Leipzig
GERMANY

Prof. Dr. Michael Westdickenberg
RWTH Aachen University
Templergraben 55
52062 Aachen
GERMANY

Prof. Dr. Johannes Zimmer
Fakultät für Mathematik (M6)
School of Computation, Information
and Technology
TU München
Boltzmannstr. 3
85748 Garching bei München
GERMANY