Some geometric consequences of the Schrödinger problem

Christian Léonard

Modal-X. Université Paris Ouest.
Bât. G, 200 av. de la République. F-92001 Nanterre, France
christian.leonard@u-paris10.fr

Abstract. This note presents a short review of the Schrödinger problem and of the first steps that might lead to interesting consequences in terms of geometry. We stress the analogies between this entropy minimization problem and the renowned optimal transport problem, in search for a theory of lower bounded curvature for metric spaces, including discrete graphs.

Keywords: Schrödinger problem, entropic interpolations, optimal transport, displacement interpolations, lower bounded curvature of metric spaces, Lott-Sturm-Villani theory.

Introduction

This note presents a short review of the Schrödinger problem and of the first steps that might lead to interesting consequences in terms of geometry. It doesn't contain any new result, but is aimed at introducing this entropy minimization problem to the community of geometric sciences of information.

We briefly describe Schrödinger's problem, see [12] for a recent review. It is very similar to an optimal transport problem. Several analogies with the Lott-Sturm-Villani theory about lower bounded curvature on geodesic spaces, which has been thoroughly investigated recently with great success, will be emphasized. The results are presented in the setting of a Riemannian manifold.

As a conclusion, some arguments are put forward that advocate for replacing the optimal transport problem by the Schrödinger problem when seeking for a theory of lower bounded curvature on discrete graphs.

For any measurable space Y, $\mathrm{M}(Y)$ is the set of all positive measures and $\mathrm{P}(Y)$ is the subset of all probability measures on Y.

1 Optimal transport

Let $\mathcal{X}$ be some state space equipped with a σ-field so that we can consider measures on $\mathcal{X}$ and $\mathcal{X}^2$. The Monge problem amounts to find a mapping $T : \mathcal{X} \to \mathcal{X}$ that solves the minimizing problem

$$\int_{\mathcal{X}} c(x, Tx)\, \mu_0(dx) \to \min; \qquad T : \mathcal{X} \to \mathcal{X} \text{ such that } T_{\#}\mu_0 = \mu_1, \tag{1}$$

where $c : \mathcal{X}^2 \to [0, \infty)$ is a given measurable function, $\mu_0, \mu_1 \in \mathrm{P}(\mathcal{X})$ are prescribed probability measures on $\mathcal{X}$ and $T_\# \mu_0(dy) := \mu_0(T^{-1}(dy))$ is the image of μ_0 by the measurable mapping T. One interprets $c(x, y)$ as the cost for transporting a unit mass from $x \in \mathcal{X}$ to $y \in \mathcal{X}$. Hence the integral $\int_{\mathcal{X}} c(x, Tx)\, \mu_0(dx)$ represents the global cost for transporting the mass profile $\mu_0 \in \mathrm{P}(\mathcal{X})$ onto $T_\# \mu_0 \in \mathrm{P}(\mathcal{X})$ by means of the transport mapping T. A solution of the Monge problem is a mapping T from $\mathcal{X}$ to $\mathcal{X}$ which transports the mass distribution μ_0 onto the target mass distribution μ_1 at a minimal cost.

The most efficient way to solve Monge's problem is to consider the following relaxed version which was introduced by Kantorovich during the 40's. The Monge-Kantorovich problem is

$$\int_{\mathcal{X}^2} c(x, y)\, \pi(dxdy) \to \min; \qquad \pi \in \mathrm{P}(\mathcal{X}^2) : \pi_0 = \mu_0, \pi_1 = \mu_1, \tag{2}$$

where $\pi_0(dx) := \pi(dx \times \mathcal{X})$ and $\pi_1(dy) := \pi(\mathcal{X} \times dy)$ are the marginals of the joint distribution π on the product space $\mathcal{X}^2$. It consists of finding a coupling $\pi \in \mathrm{P}(\mathcal{X}^2)$ of the mass distributions μ_0 and μ_1 which minimizes the average cost $\int_{\mathcal{X}^2} c(x, y)\, \pi(dxdy)$. Considering the so-called deterministic coupling $\pi^T(dxdy) := \mu_0(dx)\delta_{Tx}(dy)$ of μ_0 and $T_\# \mu_0$ where δ_y stands for the Dirac measure at y, we see that (2) extends (1): if (2) admits π^T as a solution, then T solves (1). In contrast with the highly nonlinear problem (1), (2) enters the well-understood class of convex minimization problems. The interest of the Monge-Kantorovich problem goes over its tight relation with the Monge transport problem. It is a source of fertile connections. For instance, it leads to the definition of many useful distances on the set $\mathrm{P}(\mathcal{X})$ of probability measures. Other connections are sometimes more surprising at first sight. We shall invoke below a few links with the geometric notion of curvature.
A key reference for the optimal transport theory is Villani's textbook [19].

2 Schrödinger problem

In 1931, that is ten years before Kantorovich discovered (2), Schrödinger [16,17] addressed a new statistical physics problem motivated by its amazing similarity with several aspects of the time reversal symmetry in quantum mechanics. In modern terms, the Schrödinger problem is expressed as follows

$$H(\pi|\rho) \to \min; \pi \in \mathrm{P}(\mathcal{X}^2) : \pi_0 = \mu_0, \pi_1 = \mu_1, \tag{3}$$

where $\rho \in \mathrm{M}(\mathcal{X}^2)$ is some reference positive measure on the product space $\mathcal{X}^2$ and

$$H(p|r) := \int_Y \log(dp/dr)\, dp \in (-\infty, \infty], \quad p, r \in \mathrm{P}(Y)$$

denotes the relative entropy of the probability measure p on Y with respect to the reference positive measure r on the same space Y and it is understood that $H(p|r) = \infty$ when p is not absolutely continuous with respect to r. For a survey of basic results about the Schrödinger problem, see [12].

Schrödinger considers a large collection of N independent particles living in a Riemannian manifold $\mathcal{X}$ (he takes $\mathcal{X} = \mathbb{R}^n$, but we need a manifold for further developments) and moving according to a Brownian motion. It is supposed that at time $t = 0$ they are spatially distributed according to a profile $\mu_0 \in \mathrm{P}(\mathcal{X})$ and that at time $t = 1$, we observe that they are distributed according to a profile $\mu_1 \in \mathrm{P}(\mathcal{X})$ which is far away from the expected configuration. One asks what is the most likely behavior of the whole system of particles which performs this very unlikely event. It is a large deviation problem which is solved by means of Sanov's theorem (see Föllmer's lecture notes [7] for the first rigorous derivation of Schrödinger's problem) and leads to (3) with the reference measure

$$\rho^\epsilon(dxdy) = \mathrm{vol}(dx)\,(2\pi\epsilon)^{-n/2} \exp\left(-\frac{d(x,y)^2}{2\epsilon}\right) \mathrm{vol}(dy)$$

where d is the Riemannian distance. It is the joint law of the endpoint position (X_0, X_1) of a Brownian motion $(X_t)_{0\le t\le 1}$ on the unit time interval $[0,1]$ with variance ϵ which starts at time $t = 0$ uniformly at random according to the volume measure on $\mathcal{X}$ (this process is reversible). For any probability π on $\mathcal{X}^2$ with a finite entropy we easily see that

$$\lim_{\epsilon\to 0^+} \epsilon H(\pi|\rho^\epsilon) = \int_{\mathcal{X}^2} \frac{d(x,y)^2}{2}\,\pi(dxdy)$$

which is an average cost as in (2) with respect to the so-called quadratic cost

$$c(x,y) = d(x,y)^2/2. \tag{4}$$

Therefore the Monge-Kantorovich problem with the quadratic cost

$$\int_{\mathcal{X}^2} \frac{d(x,y)^2}{2}\,\pi(dxdy) \to \min; \qquad \pi \in \mathrm{P}(\mathcal{X}^2) : \pi_0 = \mu_0, \pi_1 = \mu_1, \tag{5}$$

appears as the limit as the fluctuation parameter ϵ tends to zero of a family of Schrödinger problems associated with a Gaussian reference measure. This is a specific instance of a general phenomenon that has been discovered by Mikami [14] and explored in detail in [11].

3 Dynamical analogues

Many aspects of the static problems (2) and (3) are easier to clarify by means of their dynamical analogues. To keep things easy, we stick to the quadratic cost (4) on a Riemannian manifold.

Notation We introduce some useful notation. The path space on $\mathcal{X}$ is denoted by $\Omega \subset \mathcal{X}^{[0,1]}$. The canonical process $(X_t)_{t\in[0,1]}$ is defined for each $t \in [0,1]$ and $\omega \in \Omega$ by $X_t(\omega) = \omega_t \in \mathcal{X}$. For any $Q \in \mathrm{M}(\Omega)$ and $0 \le t \le 1$, we denote $Q_t := (X_t)_\# Q := Q(X_t \in \cdot) \in \mathrm{M}(\mathcal{X})$ the law of X_t under Q. We denote the endpoint distribution $Q_{01}(dxdy) := Q(X_0 \in dx, X_1 \in dy) \in \mathrm{M}(\mathcal{X}^2)$ and use the probabilistic notation E_P for $\int_\Omega dP$.

Displacement interpolations We introduce the dynamical analogue of (5). It consists of minimizing the average kinetic action

$$E_P \int_{[0,1]} |\dot{X}_t|^2_{X_t}/2\, dt \to \min; \quad P \in \mathrm{P}(\Omega) : P_0 = \mu_0, P_1 = \mu_1 \tag{6}$$

under the constraint that the initial and final marginals P_0 and P_1 of P are equal to the prescribed probability measures μ_0 and $\mu_1 \in \mathrm{P}(\mathcal{X})$ on $\mathcal{X}$.

Suppose for simplicity that there is a unique solution P to this problem. Then P has the form $P(\cdot) = \int_{\mathcal{X}^2} \delta_{\gamma^{xy}}(\cdot)\, \pi(dxdy)$ where $\delta_{\gamma^{xy}}$ is the Dirac mass at γ^{xy}: the unique geodesic between x and y, and its endpoint projection $P_{01} = \pi \in \mathrm{P}(\mathcal{X}^2)$ is the unique solution of the optimal transport problem (5).

Definition 1. *The* displacement interpolation *between μ_0 and μ_1 is the flow of marginals $[\mu_0, \mu_1] := (P_t)_{0\le t\le 1}$ of the solution P of (6).*

This notion has been introduced by McCann in his PhD Thesis [13]. It is the basis of the development of the theory of lower bounds for the Ricci curvature of geodesic spaces, see the textbooks [1,19].

Entropic interpolations Now, we introduce the dynamical analogue of (3). It consists of minimizing the relative entropy

$$H(P|R) := E_P \log(dP/dR) \to \min; \quad P \in \mathrm{P}(\Omega) : P_0 = \mu_0, P_1 = \mu_1 \tag{7}$$

with respect to the reference path measure $R \in \mathrm{M}(\Omega)$ under the same marginal constraints as in (6). Following Schrödinger, if we choose R to be the law of the reversible Brownian motion on $\mathcal{X}$, we obtain with Girsanov's theory that

$$H(P|R) = H(P_0|\mathrm{vol}) + E_P \int_{[0,1]} |v_t^P(X_t)|^2_{X_t}/2\, dt$$

where vol denotes the volume measure and v_t^P is the Nelson forward velocity field of the diffusion law P, [15]. When $\mathcal{X} = \mathbb{R}^n$, denoting $E_P[\cdot|\cdot]$ the conditional expectation,

$$v_t^P(x) = \lim_{h\to 0, h>0} \frac{1}{h} E_P[X_{t+h} - X_t \mid X_t = x]. \tag{8}$$

Definition 2. *The* entropic interpolation *between μ_0 and μ_1 is the flow of marginals $[\mu_0, \mu_1]^R := (P_t)_{0\le t\le 1}$ of the unique solution P of (7).*

If $P \in \mathrm{P}(\Omega)$ solves the dynamical problem (7), then $P_{01} \in \mathrm{P}(\mathcal{X}^2)$ solves the static problem

$$H(\pi|R_{01}) \to \min; \pi \in \mathrm{P}(\mathcal{X}^2) : \pi_0 = \mu_0, \pi_1 = \mu_1, \tag{9}$$

where the reference measure $R_{01} \in \mathrm{M}(\mathcal{X}^2)$ is the endpoint projection of the reference path measure $R \in \mathrm{M}(\Omega)$.

Slowing down As already seen in the static case, the analogy between (6) and (7) is not only formal. Considering the slowed down process $R^\epsilon = (X^\epsilon)_\# R$ which is the law of $X^\epsilon_t = X_{\epsilon t}, 0 \le t \le 1$, it is known that

$$\epsilon H(P|R^\epsilon) \to \min; \quad P \in \mathrm{P}(\Omega) : P_0 = \mu_0, P_1 = \mu_1$$

Γ-converges to (6), see [11]. In particular, the entropic interpolation $[\mu_0, \mu_1]^{R^\epsilon}$ is a smooth approximation of the displacement interpolation $[\mu_0, \mu_1]$.

This kind of convergence also holds for optimal L^1-transport on graphs [9] and Finsler manifolds (instead of optimal L^2-transport on a Riemannian manifold) where diffusion processes must be replaced by random processes with jumps (work in progress).

4 Dynamics of the interpolations

Unlike entropic interpolations, displacement interpolations lack regularity. Already known results about the dynamics of the displacement interpolations in the so-called RCD geodesic spaces with a Ricci curvature bounded from below can be found in [8]. Understanding the dynamics of entropic interpolations could be a first step (before letting ϵ tend to zero) to recover such results.

Dynamics of the displacement interpolations A *formal* representation of the displacement interpolation is given by

$$\dot{X}_t = \nabla\psi(t, X_t), \quad P\text{-a.s.}$$

where P is a solution of (6), ψ is the viscosity solution of the Hamilton-Jacobi equation

$$\begin{cases} \partial_t\psi + |\nabla\psi|^2/2 = 0 \\ \psi_{t=1} = \psi_1 \end{cases} \tag{10}$$

and ψ_1 is in accordance with the endpoint data μ_0 and μ_1. Note that

$$\ddot{X}_t = \nabla[\partial_t\psi + |\nabla\psi|^2/2](t, X_t) = 0, \quad P\text{-a.s.} \tag{11}$$

fitting the standard geodesic picture.

Dynamics of the entropic interpolations Similarly, a *rigorous* representation of the entropic interpolation is given by

$$v^P_t = \nabla\psi(t, X_t), \quad P\text{-a.s.}$$

where v^P is defined at (8), P is the solution of (7) and ψ is the classical solution of the Hamilton-Jacobi-Bellman equation

$$\begin{cases} \partial_t\psi + \Delta\psi/2 + |\nabla\psi|^2/2 = 0, \\ \psi_{t=1} = \psi_1. \end{cases} \tag{12}$$

Iterating time derivations in the spirit of (8) in both directions of time allows to define a relevant notion of stochastic acceleration a^P, see for instance [15,5]. We obtain the following analogue of (11)

$$a_t^P = \frac{1}{2}\nabla[\partial_t\psi+\Delta\psi/2+|\nabla\psi|^2/2]+\frac{1}{2}\nabla[-\partial_t\varphi+\Delta\varphi/2+|\nabla\varphi|^2/2](t,X_t)=0, \quad P\text{-a.s.}$$

where φ solves some HJB equation $\begin{cases} -\partial_t\varphi+\Delta\varphi/2+|\nabla\varphi|^2/2=0 \\ \varphi_{t=0}=\varphi_0 \end{cases}$ in the other direction of time.

5 Interpolations are sensitive to Ricci curvature

On a Riemannian manifold $\mathcal{X}$, one says that the Ricci curvature is bounded below by some constant $\kappa \in \mathbb{R}$, when

$$\mathrm{Ric}_x(v,v) \geq \kappa g_x(v,v), \quad \forall (x,v) \in \mathrm{T}\mathcal{X}$$

where Ric is the Ricci tensor and g is the Riemannian metric defined on the tangent bundle $\mathrm{T}\mathcal{X}$.

Displacement interpolations Ten years ago, Sturm and von Renesse [18] have discovered that this lower bound holds if and only if along any displacement interpolation $(\mu_t)_{0\leq t\leq 1}$, the entropy

$$t \in [0,1] \mapsto H(\mu_t|\mathrm{vol}) \in (-\infty,\infty]$$

is κ-convex with respect to W_2, i.e.

$$H(\mu_t|\mathrm{vol}) \leq (1-t)H(\mu_0|\mathrm{vol}) + tH(\mu_1|\mathrm{vol}) - \kappa\frac{t(1-t)}{2}W_2^2(\mu_0,\mu_1), \quad \forall t \in [0,1]. \tag{13}$$

The Wasserstein distance W_2 of order 2 is defined by means of the quadratic optimal transport problem by $W_2^2(\mu_0,\mu_1) := \inf(5)$. It plays the role of a Riemannian distance on the set $\mathrm{P}_2(\mathcal{X}) := \left\{\mu \in \mathrm{P}(\mathcal{X}); \int_{\mathcal{X}} d(x_o,x)^2\,\mu(dx) < \infty\right\}$ of all probability measures on $\mathcal{X}$ with a finite second moment. Accordingly, the displacement interpolations are similar to geodesics. Unfortunately $(\mathrm{P}_2(\mathcal{X}),W_2)$ is not a Riemannian manifold and the displacement interpolations are not regular enough to be differentiable in time. In particular, the expected equivalent local statement

$$\frac{d^2}{dt^2}H(\mu_t|\mathrm{vol}) \geq \kappa W_2^2(\mu_0,\mu_1), \ \forall 0 \leq t \leq 1 \tag{14}$$

of the convex inequality (13) is meaningless.

However, this remarkable result of Sturm and von Renesse was the basic step for developing the Lott-Sturm-Villani theory of lower bounded Ricci curvature of geodesic spaces, see [19]. The program of this theory is to extend the notion of lower bounded Ricci curvature from Riemannian manifolds to geodesic spaces (a special class of metric spaces) by taking advantage of the *almost* Riemannian structure of $(\mathrm{P}_2(\mathcal{X}), W_2)$ and in particular of the dynamical properties the corresponding almost geodesics: the displacement interpolations. The heuristic formula obtained with Otto's heuristic calculus, see [19, Ch. 15], for the second derivative of the entropy along a displacement interpolations (μ_t) is

$$\frac{d^2}{dt^2} H(\mu_t|\mathrm{vol}) = \Gamma_2(\psi_t), \quad 0 \le t \le 1, \tag{15}$$

where ψ solves the Hamilton-Jacobi equation (10). We see that it formally implies (14) under the Γ_2-criterion

$$\Gamma_2(\psi) \ge \kappa g(\nabla\psi, \nabla\psi), \quad \forall\psi$$

where the Bakry-Émery operator Γ_2 is given by

$$\Gamma_2(\psi) = \mathrm{Ric}(\nabla\psi) + \sum_{i,j} (\partial_i \partial_j \psi)^2.$$

Entropic interpolations As an interesting consequence of the dynamical properties of the entropic interpolations, we obtain in [10] that along any entropic interpolation $(\mu_t)_{0\le t\le 1}$ on a Riemannian manifold, we have

$$\frac{d^2}{dt^2} H(\mu_t|\mathrm{vol}) = \frac{1}{2} \left\{ \Gamma_2(\varphi_t) + \Gamma_2(\psi_t) \right\}$$

where φ and ψ are the solutions of the above HJB equations (12) in both directions of time. This formula is a rigorous (in the sense that the second derivative is well defined) analogue of the heuristic identity (15).

Conclusion

As a conclusion we sketch a research program and cite a few recent publications related to the Schrödinger problem in the domains of numerical analysis and engineering sciences.

A research program

In view of the analogies between the optimal transport problem and the Schrödinger problem on a Riemannian manifold, one can hope that the program of the Lott-Sturm-Villani theory can be transferred successfully from geodesic spaces to a larger class of metric spaces. As a guideline, one should consider the Schrödinger problem as the basic "geodesic" problem instead of the Monge-Kantorovich problem. We see several advantages to this strategy:

1. Unlike the displacement interpolations, the entropic interpolations are regular enough for their second derivative in time to be considered without any trouble.

2. Slowing down the reference process, which might be a diffusion process on a RCD space (see [8]) or a random walk on a graph (see [9,10]), one retrieves displacement interpolations as limits of entropic interpolations.
3. As shown in [9], the entropic interpolations are well defined on discrete graphs. They also lead to natural displacement interpolations. Remark that discrete graphs are not geodesic and as a consequence, are ruled out by the Lott-Sturm-Villani approach.

This program remains to be investigated ...

Recent literature

A recent resurgence of the use of the Schrödinger problem arises in applied and numerical sciences. In [6], the Schrödinger problem is solved using the Sinkhorn algorithm. This appears to be very competitive with respect to other optimal transport solvers because of its simplicity, parallelism and convergence speed (at the expense of an extra smoothing).

A notion of interpolation quite similar to the entropic interpolation might be defined by means of entropic barycenters as introduced in [2]. It would be interesting to investigate their curvature properties.

Motivated both by engineering problems and theoretical physics, in the spirit of [14] the recent papers [3,4] look at the entropic interpolations with a stochastic control viewpoint.

References

1. L. Ambrosio, N. Gigli, and G. Savaré. *Gradient flows in metric spaces and in the space of probability measures.* Lectures in Mathematics ETH Zürich. Birkhäuser, Basel, second edition, 2008.
2. J-D. Benamou, G. Carlier, M. Cuturi, L. Nenna, and G. Peyré. Iterative Bregman projections for regularized transportation problems. *SIAM J. Sci. Comput.*, 37(2):A1111–A1138, 2015.
3. Y. Chen, T. Georgiou, and M. Pavon. On the relation between optimal transport and schrödinger bridges: A stochastic control viewpoint. Preprint, arXiv:1412.4430.
4. Y. Chen, T. Georgiou, and M. Pavon. Optimal transport over a linear dynamical system. Preprint, arXiv:1502.01265.
5. K.L. Chung and J.-C. Zambrini. *Introduction to Random Time and Quantum Randomness.* World Scientific, 2008.
6. M. Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport. NIPS 2013, arXiv:1306.0895.
7. H. Föllmer. *Random fields and diffusion processes, in École d'été de Probabilités de Saint-Flour XV-XVII-1985-87*, volume 1362 of *Lecture Notes in Mathematics.* Springer, Berlin, 1988.
8. N. Gigli. Nonsmooth differential geometry. An approach tailored for spaces with Ricci curvature bounded from below. Preprint. http://cvgmt.sns.it/paper/2468/.
9. C. Léonard. Lazy random walks and optimal transport on graphs. Preprint. arXiv:1308.0226, to appear in Ann. Probab.

10. C. Léonard. On the convexity of the entropy along entropic interpolations. Preprint, arXiv:1310.1274.
11. C. Léonard. From the Schrödinger problem to the Monge-Kantorovich problem. *J. Funct. Anal.*, 262(1879-1920), 2012.
12. 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.
13. R. McCann. *A convexity theory for interacting gases and equilibrium crystals*. PhD thesis, Princeton Univ., 1994.
14. T. Mikami. Monge's problem with a quadratic cost by the zero-noise limit of h-path processes. *Probab. Theory Relat. Fields*, 129:245–260, 2004.
15. E. Nelson. *Dynamical theories of Brownian motion*. Princeton University Press, 1967.
16. E. Schrödinger. Über die Umkehrung der Naturgesetze. *Sitzungsberichte Preuss. Akad. Wiss. Berlin. Phys. Math.*, 144:144–153, 1931.
17. E. Schrödinger. Sur la théorie relativiste de l'électron et l'interprétation de la mécanique quantique. *Ann. Inst. H. Poincaré*, 2:269–310, 1932. `http://archive.numdam.org/ARCHIVE/AIHP/`.
18. K-T. Sturm and M-K. von Renesse. Transport inequalities, gradient estimates, entropy, and Ricci curvature. *Comm. Pure Appl. Math.*, 58(7):923–940, 2005.
19. C. Villani. *Optimal Transport. Old and New*, volume 338 of *Grundlehren der mathematischen Wissenschaften*. Springer, 2009.