From the Schrödinger problem to the Monge-Kantororich problem

C. Léonard

Paris Ouest University

Information Geometry and its Applications III
Leipzig, August 2-6, 2010

Outline

1. Monge, Kantorovich and Schrödinger
   - Optimal transport
   - Schrödinger problem
2. From entropy to transport
   - Freezing the particle system
   - Optimal transport is the frozen limit
3. Some geometry

The Monge transport problem

Take

- $\mathcal{X}$, a measurable space;
- $c : \mathcal{X} \times \mathcal{X} \to [0, \infty)$, a cost function;
- $\mu_0, \mu_1 \in \mathcal{P}(\mathcal{X})$, two probability measures on $\mathcal{X}$.

Consider the measurable maps $T : x \in \mathcal{X} \mapsto y = T(x) \in \mathcal{X}$ such that

$$\mu_1 = T_\# \mu_0 \quad \text{(image measure)}.$$

Monge problem

$$\text{minimize} \quad T \mapsto \int_{\mathcal{X}} c(x, T(x))\, \mu_0(dx) \quad \text{subject to} \quad \mu_1 = T_\# \mu_0. \qquad \text{(M)}$$

Example: $\mathcal{X} = \mathbb{R}^d$, $c(x, y) = \|y - x\|$ or $\|y - x\|^2$.

The Monge-Kantorovich transport problem

We want to transport μ_0 onto μ_1 with a minimal cost.

A *transport plan* is $\pi \in \mathcal{P}(\mathcal{X} \times \mathcal{X})$ such that $\begin{cases} \text{first marginal} & := \pi_0 = \mu_0 \\ \text{second marginal} & := \pi_1 = \mu_1 \end{cases}$

The joint law $\pi = \mathcal{L}(X_0, X_1)$ is a *coupling* of $\mu_0 = \mathcal{L}(X_0)$ and $\mu_1 = \mathcal{L}(X_1)$.

With $X_0, X_1 : \mathcal{X} \times \mathcal{X} \to \mathcal{X}$ the first and second projections:

$$X_{0\#}\pi = \mu_0, \qquad X_{1\#}\pi = \mu_1.$$

Monge-Kantorovich problem

$$\begin{aligned} &\text{minimize} \quad \pi \in \mathcal{P}(\mathcal{X} \times \mathcal{X}) \mapsto \int_{\mathcal{X} \times \mathcal{X}} c(x, y)\, \pi(dxdy) \\ &\text{subject to} \quad X_{0\#}\pi = \mu_0, X_{1\#}\pi = \mu_1. \end{aligned} \tag{MK}$$

It is a relaxed version of the Monge problem.

Disintegration: $\pi(dxdy) = \pi_0(dx)\pi(dy|X_0 = x)$.

Monge corresponds to $\pi(dy|X_0 = x) = \delta_{T(x)}(dy)$.

Wasserstein metric

Take $(\mathcal{X}, d)$ a metric space.
Consider the cost $c(x,y) = d^p(x,y)$, $p = 1$ or 2.
The Wasserstein metric on $\mathcal{P}_p(\mathcal{X})$ is

$$W_p(\mu_0, \mu_1) = \inf \left\{ (\mathrm{MK})_{d^p} \right\}^{1/p}, \quad \mu_0, \mu_1 \in \mathcal{P}(\mathcal{X}).$$

Result: W_p is a metric on $\mathcal{P}_p(\mathcal{X}) := \{\mu \in \mathcal{P}(\mathcal{X}); \int_{\mathcal{X}} d^p(x_o, x)\, \mu(dx) < \infty\}$.

The Schrödinger problem

Take n independent Brownian particles in $\mathcal{X} = \mathbb{R}^d$:

$$X_i(t) = x_i + \sigma W_i(t), \quad 0 \leq t \leq 1, i = 1, \dots, n.$$

- σ^2 is the temperature of the heat bath;
- $W_1, \dots, W_n$ are independent Brownian motions;
- $x_1, \dots, x_n$ are the initial positions.

Suppose that: $L_0^n := \frac{1}{n} \sum_{i=1}^n \delta_{X_i(0)} = \frac{1}{n} \sum_{i=1}^n \delta_{x_i} \underset{n \to \infty}{\to} \mu_0$ in $\mathcal{P}(\mathcal{X})$. Then:

$$L^n := \frac{1}{n} \sum_{i=1}^n \delta_{X_i} \underset{n \to \infty}{\to} R \quad \text{in} \quad \mathcal{P}(\Omega)$$

where $\Omega = \{\text{paths}\} = C([0,1], \mathcal{X})$ and $R := \mathcal{L}(X_0 + \sigma W)$, $W \perp X_0 \sim \mu_0$.

The Schrödinger problem

In particular, with $R_1 = \mu_0 * \mathcal{N}(0, Id)$,

$$L_1^n := \frac{1}{n} \sum_{i=1}^{n} \delta_{X_i(1)} \underset{n\to\infty}{\longrightarrow} R_1 := X_{1\#}R \quad \text{in} \quad \mathcal{P}(\mathcal{X})$$

With $n = \infty$ the system starts from $R_0 = \mu_0$ and ends up at R_1 almost surely.
But n is finite.

Schrödinger's question

Suppose that at time $t = 1$ you observe L_1^n near μ_1, far from R_1.
What is the most likely path $t \in [0, 1] \mapsto L_t^n \in \mathcal{P}(\mathcal{X})$ of your particle system?

This is a large deviation problem.

Schrödinger's answer

Relative entropy: $H(P|R) := \int_\Omega \log(\frac{dP}{dR})\, dP \in [0,\infty]$, $P, R \in \mathcal{P}(\Omega)$.

Theorem (Sanov's theorem)

For $A \in \mathcal{P}(\Omega)$, $\mathbb{P}(L^n \in A) \underset{n\to\infty}{\asymp} \exp\{-n \inf_{P\in A} H(P|R)\}$.

Conditionally on $L_0^n \underset{n\to\infty}{\to} \mu_0$ and $L_1^n \underset{n\to\infty}{\to} \mu_1$,

$$L^n \underset{n\to\infty}{\to} \widehat{P} \quad \text{almost surely in} \quad \mathcal{P}(\Omega)$$

where $\widehat{P}$ is the unique solution of

Schrödinger dynamical problem

$$\begin{array}{ll} \text{minimize} & P \in \mathcal{P}(\Omega) \mapsto H(P|R) \\ \text{subject to} & X_{0\#}P = \mu_0, X_{1\#}P = \mu_1. \end{array} \tag{S}$$

Schrödinger's answer

Denote for all $P \in \mathcal{P}(\Omega)$,

- $P_t = X_{t\#}P \in \mathcal{P}(\mathcal{X})$ the law of position at time t;
- $P_{01} = (X_0, X_1)_\# P \in \mathcal{P}(\mathcal{X} \times \mathcal{X})$ the joint law of initial and final positions;
- $P^{xy} = P(\cdot | X_0 = x, X_1 = y)$ the bridge between x and y;
- $R^\pi(\cdot) := \int_{\mathcal{X}\times\mathcal{X}} R^{xy}(\cdot)\, \pi(dxdy)$.

Tensorization: $H(P|R) = H(P_{01}|R_{01}) + \int_{\mathcal{X}\times\mathcal{X}} H(P^{xy}|R^{xy})\, P_{01}(dxdy)$.
When $P_{01} = \pi \in \mathcal{P}(\mathcal{X} \times \mathcal{X})$ is fixed, this is minimal when $P^{xy} = R^{xy}$ for all $x, y \in \mathcal{X}$. That is for $P = R^\pi$.
The solution of (S) is

$$\widehat{P} = R^{\hat\pi}$$

where $\hat\pi$ is the solution of

Schrödinger problem

$$\begin{array}{ll} \text{minimize} & \pi \in \mathcal{P}(\mathcal{X} \times \mathcal{X}) \mapsto H(\pi|R_{01}) \\ \text{subject to} & X_{0\#}\pi = \mu_0, X_{1\#}\pi = \mu_1. \end{array} \qquad (S_{01})$$

Schilder's theorem

The main idea is to let the temperature $\epsilon = \sigma^2$ tend down to zero.
Let $X^\epsilon(t) = X^\epsilon(0) + \sqrt{\epsilon} W_t$, $0 \le t \le 1$.
For all path $\omega = (\omega_t)_{0 \le t \le 1}$, denote $C(\omega) = \int_0^1 \frac{1}{2} |\dot{\omega}_t|^2 \, dt \in [0, \infty]$.

Theorem (Schilder)

For all $A \subset \Omega$, $\mathbb{P}(X^\epsilon \in A) \underset{\epsilon \to 0}{\asymp} \exp\left\{ -\frac{1}{\epsilon} \inf_{\omega \in A} [I_0(\omega_0) + C(\omega)] \right\}$.

Conditionally on $X_0^\epsilon \underset{\epsilon \to 0}{\to} x$ and $X_1^\epsilon \underset{\epsilon \to 0}{\to} y$,
X^ϵ tends almost surely to the solution of the

Geodesic problem

$$\text{minimize} \quad \omega \in \Omega \mapsto C(\omega) = \int_0^1 \frac{1}{2} |\dot{\omega}_t|^2 \, dt \quad \text{subject to} \quad \omega_0 = x, \omega_1 = y.$$

Solution: γ^{xy}, the constant speed straight line between x and y and

$$c(x, y) := \inf\{C(\omega); \omega_0 = x, \omega_1 = y\} = C(\gamma^{xy}) = \|y - x\|^2/2$$

is the popular quadratic cost.

Schrödinger problem in a cold world

We mix Sanov and Schilder's theorems.

- $R^\epsilon \in \mathcal{P}(\Omega)$ is the law of X^ϵ;
- $X_1^\epsilon, \dots, X_n^\epsilon$ are independent copies of X^ϵ;
- $L^{n,\epsilon} = \frac{1}{n}\sum_{i=1}^n \delta_{X_i^\epsilon}$ is a random element in $\mathcal{P}(\Omega)$.

Theorem (Dawson-Gärtner, L.)

Assume that $\{X^\epsilon; \epsilon > 0\}$ *obeys the LDP:* $R^\epsilon(\cdot) \underset{\epsilon \to 0}{\asymp} \exp\left\{-\frac{1}{\epsilon} C(\cdot)\right\}$. *Then,*

$$\Gamma\text{-}\lim_{\epsilon \to 0} \epsilon H(P|R^\epsilon) = \int_\Omega C\, dP \in [0, \infty]$$

and $\{L^{n,\epsilon}; \epsilon > 0, n \geq 1\}$ *obeys the double index LDP*

$$\mathbb{P}(L^{n,\epsilon} \in A) \underset{\epsilon \to 0, n \to \infty}{\asymp} \exp\left\{-\frac{n}{\epsilon} \inf_{P \in A} \int_\Omega C\, dP\right\}.$$

Idea of the proof

$$\begin{aligned} \epsilon H(P|R^\epsilon) &= \epsilon \sup_f \left\{ \int f\, dP - \log \int e^f\, dR^\epsilon \right\} \\ &= \sup_f \left\{ \epsilon \int f/\epsilon\, dP - \epsilon \log \int e^{f/\epsilon}\, dR^\epsilon \right\} \\ &= \sup_f \left\{ \int f\, dP - \epsilon \log \int e^{f/\epsilon}\, dR^\epsilon \right\} \end{aligned}$$

But $R^\epsilon(\cdot) \underset{\epsilon \to 0}{\asymp} \exp\{-C(\cdot)/\epsilon\}$ gives $\int e^{f/\epsilon}\, dR^\epsilon \underset{\epsilon \to 0}{\asymp} \sup_\Omega \exp\{(f - C)/\epsilon\}$ and

$$\epsilon \log \int e^{f/\epsilon}\, dR^\epsilon \underset{\epsilon \to 0}{\to} \sup_\Omega (f - C).$$

$$\begin{aligned} \epsilon H(P|R^\epsilon) &\underset{\epsilon \to 0}{\rightsquigarrow} \sup_f \left\{ \int f\, dP - \sup(f - C)\, dP \right\} \\ &= \sup_f \left\{ \int C\, dP + \int [(f - C) - \sup(f - C)]\, dP \right\} \\ &= \int C\, dP. \end{aligned}$$

Γ-convergence

Lower semicontinuous envelope: $\mathrm{lsc}(h)(x) := \sup_{V \in \mathcal{V}(x)} \inf_{y \in V} h(y)$.
Γ-limit: $\Gamma\text{-}\lim_{\epsilon \to 0} h^\epsilon(x) := \sup_{V \in \mathcal{V}(x)} \lim_{\epsilon \to 0} \inf_{y \in V} h^\epsilon(y)$.
This implies that there exists a sequence $\{x^\epsilon\}$ such that

$$x^\epsilon \underset{\epsilon \to 0}{\longrightarrow} x; \quad h^\epsilon(x^\epsilon) \underset{\epsilon \to 0}{\longrightarrow} h(x).$$

More,

$$\min h^\epsilon \underset{\epsilon \to 0}{\longrightarrow} \min h; \quad \operatorname{argmin} h^\epsilon \underset{\epsilon \to 0}{\longrightarrow} \operatorname{argmin} h.$$

In particular, $\Gamma\text{-}\lim_{\epsilon \to 0} \epsilon H(P|R^\epsilon) = \int_\Omega C\, dP$ implies that there exists a sequence $\{P^\epsilon\}$ such that

$$P^\epsilon \underset{\epsilon \to 0}{\longrightarrow} P; \quad \epsilon H(P^\epsilon|R^\epsilon) \underset{\epsilon \to 0}{\longrightarrow} \int_\Omega C\, dP.$$

Remark: $\{R^\epsilon\}$ is a family of mutually singular measures.

Contraction principle

I skip the technical details about the initial condition.
Idea: Take R_0 as the reversing measure of the Markov dynamics.
Here R_0 is Lebesgue measure, which is unbounded.
By the contraction principle

$$R^\epsilon(d\omega) \underset{\epsilon\to 0}{\asymp} \exp\left\{-\frac{1}{\epsilon}C(\omega)\right\}$$

gives

$$R^\epsilon_{01}(dxdy) \underset{\epsilon\to 0}{\asymp} \exp\left\{-\frac{1}{\epsilon}c(x,y)\right\}$$

with

$$\begin{aligned} c(x,y) &= \inf\{C(\omega); \omega : \omega_0 = x, \omega_1 = y\} \\ &= \inf\left\{\int_0^1 \frac{1}{2}\|\dot\omega_t\|^2\,dt; \omega : \omega_0 = x, \omega_1 = y\right\} \end{aligned}$$

Γ-convergence is stable under contraction.

Limit of the minimization problems

Theorem (Dynamical problems)

In the space $\mathcal{P}(\Omega)$, we have the following convergence

$$\begin{array}{rllll} (S^\epsilon): & \textit{minimize } P \mapsto \epsilon H(P|R^\epsilon) & \textit{subject to} & P_0 = \mu_0^\epsilon, & P_1 = \mu_1^\epsilon \\ \Gamma\text{-}\lim_\epsilon \downarrow & & & \downarrow & \downarrow \\ (T): & \textit{minimize } P \mapsto \int_\Omega C\, dP & \textit{subject to} & P_0 = \mu_0, & P_1 = \mu_1 \end{array}$$

And by contraction, we obtain the

Theorem (From (S) to (MK))

In the space $\mathcal{P}(\mathcal{X} \times \mathcal{X})$, we have the following convergence

$$\begin{array}{rllll} (S_{01}^\epsilon): & \textit{minimize } \pi \mapsto \epsilon H(\pi|R_{01}^\epsilon) & \textit{subject to} & \pi_0 = \mu_0^\epsilon, & \pi_1 = \mu_1^\epsilon \\ \Gamma\text{-}\lim_\epsilon \downarrow & & & \downarrow & \downarrow \\ (MK): & \textit{minimize } \pi \mapsto \int_{\mathcal{X}\times\mathcal{X}} c\, d\pi & \textit{subject to} & \pi_0 = \mu_0, & \pi_1 = \mu_1 \end{array}$$

where

$$R_{01}^\epsilon(dxdy) \underset{\epsilon \to 0}{\asymp} \exp\left\{-\frac{1}{\epsilon} c(x,y)\right\}.$$

Limit of the minimizers

Theorem (Convergence of the minimizers)

There exists a sequence $(\mu_0^\epsilon, \mu_1^\epsilon) \underset{k\to\infty}{\longrightarrow} (\mu_0, \mu_1)$ *such that*

$$\begin{array}{ccl} \widehat{P}^\epsilon(\cdot) & = & \int_{\mathcal{X}\times\mathcal{X}} R^{\epsilon,xy}(\cdot)\, \hat{\pi}^\epsilon(dxdy); \quad \widehat{P}_0^\epsilon = \hat{\pi}_0^\epsilon = \mu_0^\epsilon, \widehat{P}_1^\epsilon = \hat{\pi}_1^\epsilon = \mu_1^\epsilon \\ \downarrow & & \qquad\qquad\quad \downarrow \qquad \downarrow \\ \widehat{P}(\cdot) & = & \int_{\mathcal{X}\times\mathcal{X}} \delta_{\gamma^{xy}}(\cdot)\, \hat{\pi}(dxdy); \quad \widehat{P}_0 = \hat{\pi}_0 = \mu_0, \widehat{P}_1 = \hat{\pi}_1 = \mu_1 \end{array}$$

Recall: $R^{\epsilon,xy}$ is the bridge of R^ϵ and γ^{xy} is the geodesic between x and y.

This convergence is a tool for approximating some features of geometry by means of probability theory.

Other dynamics

For any $k \geq 1$, let R^k be the law of the random walk in $\mathcal{X} = \mathbb{R}^d$

$$X^k(t) = X^k(0) + \frac{1}{k} \sum_{j=1}^{\lfloor kt \rfloor} Z_j, \quad 0 \leq t \leq 1$$

where $Z_1, \dots, Z_k$ are independent copies of Z. Define

- $C(\omega) = \int_0^1 c(\dot{\omega}_t)\, dt \in [0, \infty], \quad \omega \in \Omega,$
- $c(v) = \sup_{p \in \mathbb{R}^d} \{v \cdot p - \log \mathbb{E} e^{p \cdot Z}\}, \quad v \in \mathbb{R}^d.$

Theorem

1. $\Gamma\text{-}\lim_{k \to \infty} \frac{1}{k} H(P|R^k) = \int_\Omega C\, dP, \quad P \in \mathcal{P}(\Omega),$
2. $\Gamma\text{-}\lim_{k \to \infty} \frac{1}{k} H(\pi|R^k_{01}) = \int_\Omega c(y - x)\, \pi(dxdy), \quad \pi \in \mathcal{P}(\mathcal{X} \times \mathcal{X})$

A geometry on $\mathcal{P}(\mathcal{X})$

McCann, Gangbo, Otto, Villani,...

Definition ("Geodesics" in $\mathcal{P}(\mathcal{X})$)

The flow

$$t \mapsto \mu_t := \widehat{P}_t(\cdot) = \int_{\mathcal{X}\times\mathcal{X}} \delta_{\gamma^{xy}(t)}(\cdot)\,\hat{\pi}(dxdy) \in \mathcal{P}(\mathcal{X})$$

where $\hat{\pi}$ is a solution of (MK) is *a kind of geodesic* from μ_0 to μ_1 in $\mathcal{P}(\mathcal{X})$. We write it

$$\mu_t = [\mu_0, \mu_1]_t, \quad 0 \leq t \leq 1$$

This "geodesic" is called a *displacement interpolation*.
It is approximated by the entropy minimizers

$$t \mapsto \mu_t^\epsilon := \widehat{P}_t^\epsilon(\cdot) = \int_{\mathcal{X}\times\mathcal{X}} R_t^{\epsilon,xy}(\cdot)\,\hat{\pi}^\epsilon(dxdy) \in \mathcal{P}(\mathcal{X}).$$

Remark: (MK) might admit infinitely many solutions. (S_{01}^ϵ) admits a unique solution.
We select the viscosity solution.

Ricci curvature

Active field of research: Lott, Sturm, von Renesse, Villani,...

Theorem

Let $\mathcal{X}$ be a connected Riemannian manifold. Then, $\mathcal{X}$ has a nonnegative Ricci curvature if and only if the relative entropy $H(\cdot|\mathrm{vol})$ is displacement convex.
That is, for any displacement interpolation $(\mu_t)_{0\le t\le 1}$,

$$H(\mu_t|\mathrm{vol}) \le (1-t)H(\mu_0|\mathrm{vol}) + tH(\mu_1|\mathrm{vol}), \forall 0 \le t \le 1.$$

This theory works mainly with the quadratic transport.
This gives a very interesting definition of *lower-bounded-Ricci-curvature* in a general metric space $\mathcal{X}$.

Work in progress

Try to obtain curvature properties of the measure space $(\mathcal{X}, \rho)$ by means of the second derivative of the function

$$t \mapsto H(\widehat{P}_t|\rho)$$

where

- $\widehat{P}$ is the Schrödinger minimizer
- with respect to a general Markov reversible Markov process R with reversing measure ρ.

Take advantage of the convergence of the functions

$$t \mapsto H(\widehat{P}^\epsilon_t|\rho^\epsilon)$$

where

- $\widehat{P}^\epsilon$ is the Schrödinger minimizer
- with respect to a general Markov reversible Markov process R^ϵ with reversing measure $\rho^\epsilon \underset{\epsilon\to 0}{\longrightarrow} \rho$.

Thank you for your attention.