On some relations between Optimal Transport and Stochastic Geometric Mechanics

Banff, December 2018

Ana Bela Cruzeiro
Dep. Mathematics IST and
Grupo de Física-Matemática Univ. Lisboa

Based on joint works with M. Arnaudon, X. Chen, J.-C. Zambrini

Extensions of (some) results, with applications to fluids: with T. Ratiu, C. Léonard

Recalling the Monge-Kantorovich (MK) optimal transport problem (flat case)

$$\inf_{\pi\in\Pi(\mu,\sigma)} \frac{1}{2}\int_{\mathbb{R}^d\times\mathbb{R}^d} \|x-y\|^2 d\pi(x,y)$$

where $\Pi(\mu,\sigma) = \{$ joint distributions s.t. the marginals along x and y coordinates are μ and σ resp. $\}$.
Since “the cost"

$$\|x-y\|^2 = \inf_X \int_0^1 \|\dot{X}(t)\|^2 dt$$

$X \in \{C([0,1];\mathbb{R}^d) : X(0)=x, X(1)=y\}$, the MK problem is equivalent to

$$\inf_\pi \frac{1}{2}\int_{\mathbb{R}^d\times\mathbb{R}^d}\Big(\int\int_0^1 \|\dot{X}(t)\|^2 dt\ dP_{x,y}\Big)d\pi(x,y)$$

or, by desintegration, to

$$\inf_P \int\int_0^1 \frac{1}{2}\|\dot{X}(t)\|^2 dt\ dP$$

with P prob. measure on $C^1([0,1];\mathbb{R}^d)$ with marginals μ and σ at $t=0, t=1$ and $P_{x,y}$ the one with initial and final marginals δ_x and δ_y.

Eulerian (control) version of this problem:

$$\inf_v \frac{1}{2} \int \int_0^1 \|v(t, x^v(t))\|^2 dtdP$$

v continuous, $\dot{x}^v(t) = v(t, x^v(t))$ a.s., law of $x^v(0) = \mu$, law of $x^v(1) = \sigma$.
If $d\mu = \rho_0 dx, d\sigma = \rho_1 dx$ given, law of $x^v(t) = \rho_t dx$, then

$$\frac{\partial \rho}{\partial t} = -\text{div}\ (\rho v)$$

(continuity equation)

If ψ solves the Hamilton-Jacobi equation

$$\frac{\partial \psi}{\partial t} = -\frac{1}{2}\|\nabla \psi\|^2$$

and $\frac{\partial \rho}{\partial t} = -\text{div}\,(\rho \nabla \psi)$, with $\rho(0, x) = \rho_0(x), \rho(1, x) = \rho_1(x)$,

then $v = \nabla \psi$ solves the problem. Moreover

$$\frac{\partial v}{\partial t} = -(v \cdot \nabla)v$$

On $C([0,1];\mathbb{R}^d)$ consider the laws (Q) of diffusions

$$dX(t) = \sqrt{\epsilon}\, dW(t) + Y(t)dt, \qquad \text{law of } \; X(0) = dx$$

Entropy functional of Q w.r.t. P:

$$H(Q;P) = \int \log\Big(\frac{dQ}{dP}\Big) dQ$$

Choose for P the law of the Wiener process

$$H(Q;P) = H(Q_0;P_0) + \frac{1}{2}\int\int_0^1 \|Y(t)\|^2 dt dQ,$$

where Q_0, P_0 are the marginals of Q and P at time 0. This is a consequence of Girsanov's Theorem.

Schrödinger problem: minimise the entropy functional, subject to given Q_0 and Q_1.

If $Y(t) = v(t, X(t))$, the density ρ_t of X at time t satisfies the Fokker-Planck equation

$$\frac{\partial \rho}{\partial t} = -\text{div}\,(\rho v) + \frac{\epsilon}{2}\Delta\rho$$

If, moreover, $v = \nabla\varphi$ satisfies Hamilton-Jacobi-Bellman equation,

$$\frac{\partial \varphi}{\partial t} = -\frac{1}{2}\|\nabla\varphi\|^2 + \frac{\epsilon}{2}\Delta\varphi$$

then $v = \nabla\varphi$. So

$$\frac{\partial v}{\partial t} = -(v\cdot\nabla)v + \frac{\epsilon}{2}\Delta v$$

(Burgers equation)

Remarks:

1. With the change of variable $v = \nabla\varphi$, $\varphi = -\log\eta$, $\frac{\partial\eta}{\partial t} = \frac{\epsilon}{2}\Delta\eta$ (heat equation)

2. When $\epsilon \to 0$ formally Schrödinger problem converges to the optimal transport problem (C. Léonard, etc)
(delicate problem)

G Lie group with

$< >$ right (left) invariant metric

∇ Levi-Civita connection

e identity element

$\mathcal{G} \simeq T_e(G)$ Lie algebra

$\{H_i\}$ o.n. basis of $\mathcal{G}$, right invariant, (we suppose finite dimensional), $\nabla_{H_i} H_i = 0$

Brownian motion on G with generator $=$ Laplace-Beltrami operator:

$$dg^0(t) = T_e R_{g^0(t)}\left(\sum_i H_i \circ dW^i(t)\right) = T_e R_{g^0(t)}\left(\sum_i H_i \cdot dW^i(t)\right)$$

where $T_a R_{g(t)} : T_a G \to T_{ag(t)} G$ is the differential of the right translation $R_{g(t)}(x) := xg(t), \forall x \in G$ at the point $x = a \in G$.

General diffusion processes:

$$dg(t) = T_e R_{g(t)} \Big(\sum_i H_i \circ dW^i(t) + u(t)dt \Big)$$

$$T_e R_{g(t)} u(t) = \frac{D^\nabla g(t)}{dt}$$

where, by definition, D^∇ is the (mean) derivative defined as: for $\xi(t) = \int_0^t T_{0 \leftarrow s} \circ dg(s)$, where $T_{\cdot \leftarrow 0} : T_{g(0)}G \to T_{g(\cdot)}G$ is the stochastic parallel transport associated to ∇, define

$$\frac{D\xi(t)}{dt} = \lim_{\epsilon \to 0} \frac{1}{\epsilon} E\Big[\xi(t+\epsilon) - \xi(t) | \mathcal{P}_t \Big]$$

and

$$\frac{D^\nabla g(t)}{dt} := T_{t \leftarrow 0} \frac{D\xi(t)}{dt}$$

From Girsanov's Theorem the law Q of g on $C([0,1];G)$ is absolutely continuous w.r.t. the law P of g^0 (same initial distributions) with density given by

$$\frac{dQ}{dP} = \exp\{\int_0^1 \sum_i < T_g R_{g^{-1}} \frac{D^\nabla g(t)}{dt}, H_i dW^i(t) >$$

$$-\frac{1}{2}\int_0^1 \| T_g R_{g^{-1}} \frac{D^\nabla g(t)}{dt} \|^2 dt\}$$

Therefore the entropy is

$$H(Q;P) = \frac{1}{2}\int \int_0^1 \| T_g R_{g^{-1}} \frac{D^\nabla g(t)}{dt} \|^2 dt dQ$$

Denote by α the invariant measure on G, P_x law of the Brownian motion g^0 on G starting at x, $P_t(x, \alpha(dy))$ its transition semigroup.

$$\tilde{P}_\mu = \int_G P_x \mu(dx)$$

Assume μ and σ prob. measures, to be abs. cont. w.r.t. α.

$\mathcal{M}(\mu, \sigma) := \{\pi$ prob. measures on $G \times G : \pi(g^0(0) \in \cdot) = \mu(\cdot), \pi(g^0(1) \in \cdot) = \sigma(\cdot), \pi$ abs. cont. w.r.t. $\mu \otimes P_1\}$

For $\pi \in \mathcal{M}(\mu, \sigma)$, define

$$P_\pi(d\omega) = \int_{G \times G} \pi(dx, dy) P(d\omega | 0, x; 1, y)$$

prob. measure on $C([0, 1]; G)$.

By Csiszär's Theorem, if $\exists \pi_0 \in \mathcal{M}(\mu, \sigma) : H(\pi_0, \mu \otimes P_1) < \infty$, then $\exists^1 Q$ on $C([0,1]; G]$ attaining the

$$\inf_{Q:\ Q((g(0),g(1) \in \cdot) \in \mathcal{M}(\mu,\sigma)} H(Q; \tilde{P}_\mu)$$

with $Q = P_\pi$, π attaining

$$\inf_\pi H(\pi; \mu \otimes P_1)$$

Moreover Q is the law of a Markov process.

Combining with Girsanov's result, we have

$$H(Q; \tilde{P}_\mu) = \frac{1}{2} \int \int_0^1 \| T_g R_{g^{-1}} \frac{D^\nabla g(t)}{dt} \|^2 dt dQ$$

$$:= A[g]$$

Stochastic Euler-Poincaré reduction theorem

Theorem. The G-valued semi-martingale

$$dg(t) = T_e R_{g(t)} \Big(\sum_i H_i \circ dW^i(t) + u(t)dt \Big)$$

is a critical point of A if and only if the time-dependent vector field $u(\cdot)$ satisfies the equation,

$$\frac{d}{dt} u(t) = -ad^*_{u(t)} u(t) - K(u(t)),$$

where
for $u \in \mathcal{G}$, $ad^*_u : \mathcal{G} \to \mathcal{G}$ is the adjoint of $ad_u : \mathcal{G} \to \mathcal{G}$ with respect to the metric $< >$,
and the operator $K : \mathcal{G} \to \mathcal{G}$ is defined as

$$K(u) = -\frac{1}{2} \sum_i \big(\nabla_{H_i} \nabla_{H_i} u + R(u, H_i) H_i \big) = -\frac{1}{2} \square(u)$$

(de Rham-Hodge Laplacian)

(Left) variations:

$$\partial_L A[\xi(\cdot)] = \frac{d}{d\varepsilon}|_{\varepsilon=0} A[\xi_{\varepsilon,v}(\cdot)]$$

with $\xi_{\varepsilon,v}(t) = e_{\varepsilon,v}(t)\xi(t)$

$$\begin{cases} \frac{d}{dt} e_{\varepsilon,v}(t) = \varepsilon T_e R_{e_{\varepsilon,v}(t)} \dot{v}(t), \\ e_{\varepsilon,v}(0) = e \end{cases} \tag{1}$$

for $v(\cdot) \in C^1([0,T];\mathcal{G})$, $v(0) = v(T) = 0$

We differentiate in the direction of v ($\frac{d}{d\epsilon}\Big|_{\varepsilon=0} e_{\varepsilon,v} = v$)

"Proof" of the Theorem:

For $g_\varepsilon(t) = e_{\varepsilon,v}(t)g(t)$, by Itô's formula ,

$$\begin{aligned} dg_\varepsilon(t) = & \sum_i T_e R_{g_\varepsilon(t)} Ad_{e_\varepsilon^{-1}(t)} H_i \circ dW_t^i \\ & + T_e R_{g_\varepsilon(t)} \Big(Ad_{e_\varepsilon^{-1}(t)} (u(t)) + T_{e_\varepsilon(t)} R_{e_\varepsilon^{-1}(t)} \dot{e}_\varepsilon(t) \Big) dt, \end{aligned} \tag{2}$$

(no contraction term)

We have $T_{e_\varepsilon(t)} R_{e_\varepsilon^{-1}(t)} \dot{e}_\varepsilon(t) = \varepsilon \dot{v}(t)$, $\frac{d}{d\varepsilon}\big|_{\varepsilon=0} Ad_{e_\varepsilon^{-1}(t)} u(t) = -ad_{v(t)} u(t)$ and $\frac{d}{d\varepsilon}\big|_{\varepsilon=0} Ad_{e_\varepsilon^{-1}(t) H_i} = ad_{v(t)} H_i$

$$
\begin{aligned}
&\frac{dA[g_\varepsilon(\cdot)]}{d\varepsilon}\Big|_{\varepsilon=0} = \mathbb{E}\int_0^1 < \frac{d}{d\varepsilon}(T_{g_\varepsilon(t)}R_{g_\varepsilon^{-1}(t)}\frac{D^\nabla g_\varepsilon(t)}{dt})\Big|_{\varepsilon=0}, u(t) > dt \\
&= \int_0^1 < u(t), \dot{v}(t) - ad_{(u(t))}v(t) \\
&+ \frac{1}{2}\sum_i (\nabla_{ad_{v(t)}H_i}H_i + \nabla_{H_i}(ad_{v(t)}H_i)) > dt \\
&= \int_0^1 < -\dot{u}(t) - ad^*_{u(t)}u(t) - K(u(t)), v(t) > dt
\end{aligned}
\tag{3}
$$

Remark : When $H_i = 0$ we recover the deterministic Euler-Poncaré reduction theorem.

The change of variables $u(t) = \nabla\varphi(t)$ gives the Hamilton-Jacobi-Bellman equation

$$\frac{\partial\varphi}{\partial t} = -\frac{1}{2}\|\nabla\varphi\|^2 + \frac{1}{2}\Delta_{LB}(\varphi)$$

Extra remarks:

In the assumptions of Csiszär's theorem ($\exists\pi_0 \in \mathcal{M}(\mu,\sigma) : H(\pi_0, \mu\otimes P_1) < \infty$), if π_0 is abs. cont. w.r.t. $\mu\otimes P_1$, then the minimiser is of the form

$$\pi(dx, dy) = \psi(x)\phi(y)\alpha(dx)P_1(x, \alpha(dy))$$

and

$$\frac{d\mu}{d\alpha} = \psi P_1\phi, \quad \frac{d\nu}{d\alpha} = \phi P_1^*\psi \quad a.e.$$

which allows to transform initial and final conditions μ, σ in ϕ, ψ, initial and final conditions for the equations satisfied by the drift and the one of its time reversed.

- M. Arnaudon, X. Chen, A. B. C., *Stochastic Euler-Poincaré reduction*, J. Math. Physics (2014)
- M. Arnaudon, A.B.C., C. Léonard, J.C. Zambrini, *An entropic interpolation problem for incompressible viscid fluids*, arXiv:1704.02126.
  (*infinite dimensional entropic approach*)
- X. Chen, A. B. C., T. S. Ratiu, *Constrained and stochastic variational principles for dissipative equations with advected quantities* arxiv:1506.05024
  (*extension of stochastic geometric mechanics results*)
- A.B.C., Liming Wu, J.C. Zambrini, *Bernstein processes associated with a Markov process*, Trends Math., Birkhäuser Boston (2000)
- C. Léonard, *A survey of the Schrödinger problem and some of its connections with optimal transport*, Discrete and Cont. Dyn. Systems (2014)
  (*general results on the entropy approach*)