Journal of the Egyptian Mathematical Society (2011) **19**, 95–96

Egyptian Mathematical Society

Journal of the Egyptian Mathematical Society

www.etms-eg.org
www.elsevier.com/locate/joems

SHORT COMMUNICATION

Optimal transport: Monge meets Riemann and Fourier

Cédric Villani

ENS Lyon & Institut Henri Poincaré, 11 rue Pierre et Marie Curie, 75231 Paris Cedex 05, France

Available online 5 December 2011

KEYWORDS

Optimal transport;
Sobolev inequality;
Curvature;
Heat equation

Abstract This short contribution summarizes a talk given on May 5, 2010, in Cairo, describing some unexpected links between the Monge problem of optimal transport, the Riemann curvature and the heat equation.

For a start, let us recall the notion of push-forward, or change of variables. If $\mu(dx)$ and $\nu(dy)$ are two (probability) measures, and $y = T(x)$ is a measurable map, then $\nu = T_\#\mu$ if for any measurable set B, one has $\mu[T^{-1}(B)] = \nu[B]$, or equivalently, for any bounded measurable function φ, one has $\int \varphi \circ T d\mu = \int \varphi d(T_\#\mu)$. If $\mu(dx) = f(x)dx$ and $\nu(dy) = g(y)dy$ in $\mathbb{R}^n$, and T is 1-to-1, the equation is $f(x) = g(T(x))|\det(d\,T)(x)|$.

The *Monge–Kantorovich problem* is stated as follows: given two probability measures $\mu(dx)$ and $\nu(dy)$, and a cost function $c(x,y)$, look at the variational problem

$$\inf_{T_\#\mu=\nu} \int_c (x, T(x))d\mu(x).$$

In words, one wants to *transport material at lowest cost, the initial and final distributions of mass being given*. In probabilistic words, we are searching for a coupling of two random variables U and V, such that the law of each is given, and we wish to minimize the expected value of $c(U,V)$. For instance if $c(x,y) = |x-y|^2$, we are trying to maximize the correlation $\mathbb{E}\langle U, V\rangle$.

It can be shown (Brenier, Rachev, Rüschendorf) that the optimal coupling takes the form $T = \nabla\Phi$, where Φ is convex. This is a *monotone change of variables*: the Jacobian matrix of T (= the Hessian of Φ) has nonnegative eigenvalues.

Monotone changes of variables are powerful: to illustrate this, here is a short proof of the isoperimetric inequality, which is a variation of an argument by Gromov. Define $|\Omega| = \mathcal{L}^n[\Omega]$, $|\partial\Omega| = \mathcal{H}^{n-1}[\partial\Omega]$, and make a change of variables from Ω to B: so $y = T(x) \in B = B(0,1)$. Assume that (i) T pushes uniform measure forward to uniform measure; (ii) dT has nonnegative eigenvalues at each point. So $f(x) = 1/|\Omega|$, $g(y) = 1/|B|$, so $\det(dT) = |B|/|\Omega|$. Then we write

$$\left(\frac{|B|}{|\Omega|}\right)^{\frac{1}{n}} = (det dT)^{\frac{1}{n}} = \left(\prod_{i=1}^n \lambda_i\right)^{\frac{1}{n}} \leqslant \frac{\sum_{i=1}^n \lambda_i}{n} = \frac{\nabla\cdot T}{n},$$

where the $\lambda_i = \lambda_i(x)$ are the eigenvalues of the Jacobian matrix. Integration over Ω and Stokes formula yield

$$|\Omega| \times \left(\frac{|B|}{|\Omega|}\right)^{\frac{1}{n}} \leqslant \int_\Omega \frac{\nabla\cdot T}{n} = \frac{1}{n}\int_{\partial\Omega} T\cdot\nu \leqslant \frac{|\partial\Omega|}{n},$$

where $\nu = \nu(x)$ is the outer unit normal, and the last inequality follows from the fact that T is valued in the unit ball and ν has unit norm.

E-mail address: villani@math.univ-lyon1.fr

Peer review under responsibility of Egyptian Mathematical Society.
doi:10.1016/j.joems.2011.09.007

ELSEVIER **Production and hosting by Elsevier**

As a variation, let us prove the Sobolev inequality in $\mathbb{R}^n$, following an argument by Cordero–Nazaret–Villani (2004). Take $p \in (1,n)$, then we wish to prove

$$\|u\|_{L^{p^\star}(\mathbb{R}^n)} \leqslant S_n(p)\|\nabla u\|_{L^p(\mathbb{R}^n)} \qquad p^\star = \frac{np}{n-p}.$$

Without loss of generality we may assume that $u \geqslant 0$ and $\int u^{p^\star} = 1$. Then the Sobolev inequality becomes just $0 < K \leqslant \|\nabla u\|_{L^p(\mathbb{R}^n)}$.

Let us pick an arbitrary probability density g, and introduce a monotone change of variables $T : u^{p^\star}dx \to g(y)dy$; so the Jacobian identity is $g(T(x)) = \frac{u(x)^{p^\star}}{det(dT(x))}$. Then we compute, assuming we can integrate by parts: $\int g^{1-\frac{1}{n}} = \int g(y)^{-\frac{1}{n}}g(y)dy = \int g(T(x))^{-\frac{1}{n}}u^{p^\star}(x)dx = \int (det dT(x))^{\frac{1}{n}}(u^{p^\star})^{1-\frac{1}{n}}(x)dx \leqslant \frac{1}{n}\int(\nabla\cdot T(x))(u^{p^\star(1-\frac{1}{n})})(x)dx = -\frac{p^\star}{n}\left(1-\frac{1}{n}\right)\int u^{p^\star(1-\frac{1}{n})-1}\nabla u\cdot Tdx$.

A bit of algebra with the exponents shows that the latter is $= -\frac{p^\star}{n}\left(1-\frac{1}{n}\right)\int u^{p^\star/p'}\nabla u\cdot Tdx$ where $1/p + 1/p' = 1$. At this point we apply Hölder's inequality, to obtain an upper bound

$$\frac{p^\star}{n}\left(1-\frac{1}{n}\right)\left(\int u(x)^{p^\star}\mid T(x)|^{p'}\right)^{\frac{1}{p'}}\left(\int \mid\nabla u|^p\right)^{\frac{1}{p}}$$
$$=\frac{p^\star}{n}\left(1-\frac{1}{n}\right)\left(\int g(y)\mid y|^{p'}\right)^{\frac{1}{p'}}\left(\int\mid\nabla u|^p\right)^{\frac{1}{p}}.$$

As a conclusion we have shown

$$\int g^{1-\frac{1}{n}} \leqslant \frac{p^\star}{n}\left(1-\frac{1}{n}\right)\left(\int g(y)\mid y|^{p'}\right)^{\frac{1}{p'}}\left(\int\mid\nabla u|^p\right)^{\frac{1}{p}},$$

and obviously this solves the problem since g is arbitrary and fixed (independently of u).

Spectacular developments of this method were achieved by Figalli, Maggi and Pratelli in relation to the so-called quantitative Wulff isoperimetry.

Other unexpected applications of optimal transport have flourished in the past few decades: incompressible fluid mechanics (Brenier); invariant measures for Lagrangian systems (Mather); semi-geostrophic equations (Cullen); weak solutions of Monge-Ampère (Brenier, Caffarelli); Boltzmann equation (Tanaka); collapse of sandpiles (Prigozhin); design of reflectors and lenses (Oliker, Wang); image matching/warping (Tannenbaum); modelling of irrigation basins (Santambrogio); reconstruction of early Universe (Frisch); etc.

I will mention a particularly striking one, which makes the connection between the Monge problem and the Fourier equation, thus reuniting in mathematics these two mathematicians who were very close in real life.

In 1998, Jordan, Kinderlehrer and Otto discovered a deep link between the heat/Fourier equation $\partial_t\rho = \Delta\rho$, the Boltzmann H functional $H(\rho) = \int \rho\log\rho$, and the optimal transport cost functional $C(\mu,\nu) = \inf_{T_\#\mu=\nu}\int d(x,T(x))^2\mu(dx)$. The link appears when the base space is $\mathbb{R}^n$ or, say, a compact Riemannian manifold (M,g). It provides a way to solve the Fourier equation by a Monge-based scheme, which is an unorthodox *gradient flow* scheme. One way to present this is to discretize in time; and from time t to time $t + \Delta t$, given $\rho(t)$, search for $\rho(t + \Delta t)$ as the *minimizer* of $H(\rho) + \frac{C(\rho(t),\rho)}{2\Delta t}$.

As $\Delta t \to 0$, this evolution gives the heat equation. Note that *by construction* the entropy $-H = -\int\rho\log\rho$ increases with time (which we know by other means for the heat equation, of course).

This observation was the starting point of unexpected developments relating optimal transport and the Riemannian curvature. To understand them, let us introduce an *interpolation along optimal transport*: the interpolation μ_t between μ_0 and μ_1 is obtained by *stopping each geodesic at time t* in the transport process: $T_t(x)$ is the trajectory from $T_0(x) = x$ to $T_1(x) = T(x)$, and $\mu_t = (T_t)_\#\mu_0$. The path $(\mu_t)_{0\leqslant t\leqslant 1}$ is then a a *geodesic* in the space of probability measures, when the geometric structure is given by the distance $\sqrt{C}(\mu,\nu)$.

We have seen that H always goes down along the gradient flow, which is the heat equation. But now what is the behavior of H along this interpolation?

Recall the definition of sectional curvature. A possible definition is as follows. Let u, $v \in T_xM$ be orthogonal unit vectors, then the sectional curvature $\kappa(u,v)$ at x along the plane generated by u, v measures the divergence of geodesics, w.r.t. to Euclidean geometry: $d(\exp_x tu, \exp_x tv) = \sqrt{2}t\left(1-\frac{\kappa}{12}t^2 + O(t^4)\right)$. Then Ricci curvature is, up to a constant, the "average sectional curvature": If $(e,e_2,\ldots,e_n)$ is an orthonormal basis of T_xM, then $Ric(e) := \sum_{j=2}^n\kappa(e,e_j)$; this extends to a quadratic form, which can be expressed in terms of second derivatives of the metric g.

The relation discovered as a consequence of works by Otto–Villani, Cordero–McCann–Schmuckenschläger, Lott–Villani, Sturm, is that *the Ricci curvature is* $\geqslant 0$ if and only if $H(\mu_t) = \int\rho_t\log\rho_t dvol$ is a *convex* function of t along any interpolation along optimal transport.

This discovery has been the basis of the development of synthetic theory of Ricci curvature bounds, in a way that complements the synthetic theory of sectional curvature bounds by Cartan–Alexandrov–Toponogov.

All this story is told, with many details and hundreds of references, in my reference textbook *Optimal transport, old and new* (Springer, 2008).