On the Wasserstein-Fisher-Rao metric

François-Xavier Vialard

Abstract

This note gives a summary of the presentation that I gave at the workshop on shape analysis[1]. Based on [CSPV15, CPSV15], we present a generalization of optimal transport to measures that have different total masses. This generalization enjoys most of the properties of standard optimal transport but we will focus on the geometric formulation of the model. We expect this new metric to have interesting applications in imaging.

1 Motivation and a Dynamical Model

In several contexts of applications including imaging, it is natural to consider data that can be represented by densities and these densities might have different masses. Often, optimal transport has been used in these applications (for instance, [HZTA04, AKS15]) since it provides an "easily computable" (at least, an efficient approximation [Cut13]) distance between probability measures that reflects a geometric displacement between them. Therefore, the mass constraint on the densities has to be taken into account and this problem seems to bring renewed interest in the applied mathematics literature [PR13, PR14, FG10, LM13, MRSS15] although this issue has been addressed since Kantorovich [Gui02].

In the following, we describe a dynamical approach to define optimal transport between general non-negative Radon measures. We will present the model only in a smooth setting although it is well defined on the space of Radon measures.

The Benamou-Brenier formulation: In [BB00], the authors formulated the Wasserstein L^2 distance as a convex variational problem, inspired by a fluid dynamic approach. In what follows, M will be a compact manifold without boundary. Let $\rho \in C^\infty(M, \mathbb{R}_+)$ be a positive function, note that all the quantities will be implicitly time dependent. The dynamic formulation of the Wasserstein distance consists in minimizing

$$\mathcal{E}(v) = \frac{1}{2} \int_0^1 \int_M |v(t,x)|^2 \rho(t,x) \, \mathrm{d}x \, \mathrm{d}t \ , \tag{1}$$

subject to the constraints $\dot{\rho} + \nabla \cdot (v\rho) = 0$ and initial condition $\rho(0) = \rho_0$ and final condition $\rho(1) = \rho_1$. Equivalently, following [BB00], a convex reformulation using the momentum $m = \rho v$ reads

$$\mathcal{E}(m) = \frac{1}{2} \int_0^1 \int_M \frac{|m(t,x)|^2}{\rho(t,x)} \, \mathrm{d}x \, \mathrm{d}t \, , \tag{2}$$

subject to the constraints $\dot{\rho} + \nabla \cdot m = 0$ and initial condition $\rho(0) = \rho_0$ and final condition $\rho(1) = \rho_1$. Let us underline that the functional (4) is convex in ρ, m and the constraint is linear.

The Wasserstein-Fisher-Rao metric: The continuity equation enforces the mass conservation property. In view of the optimal transport generalization, this constraint needs to be relaxed, for instance by introducing a source term $\mu \in C^\infty(M, \mathbb{R})$,

$$\dot{\rho} = -\nabla \cdot (v\rho) + \mu \, . \tag{3}$$

[1] Grundsund, July 2015

For a given variation of the density $\dot{\rho}$, there exist a priori many couples (v, μ) that reproduces this variation. Following [TY05], it can be determined via the minimization of a norm of (v, μ) for an arbitrary choice of the norm. The penalization of μ was chosen in [MRSS15] as the L^2 norm but a natural choice is rather the Fisher-Rao metric

$$FR(\mu) = \frac{1}{2}\int_M \frac{\mu(t,x)^2}{\rho(t,x)}\,\mathrm{d}x\,\mathrm{d}t$$

since (1) it is 1-homogeneous with respect to the couple (μ, ρ) and (2) it is parametrization invariant [MBM14]. The first point is important for convex analysis properties in order to define the model on singular measures and the second point is natural from a modeling point of view if one thinks that μ represents a growth term. Thus, the problem becomes:

$$WF(m,\mu) = \frac{1}{2}\int_0^1\int_M \frac{|m(t,x)|^2}{\rho(t,x)}\,dx\, + \frac{1}{2}\int_0^1\int_M \frac{\mu(t,x)^2}{\rho(t,x)}\,\mathrm{d}x\,\mathrm{d}t\,, \tag{4}$$

subject to the constraints $\dot{\rho} + \nabla\cdot m = \mu$ and initial condition $\rho(0) = \rho_0$ and final condition $\rho(1) = \rho_1$. This dynamical formulation enjoys most of the analytical properties of the initial Benamou-Brenier formulation (1) and especially convexity. An important consequence is the existence of optimal paths in the space of time-dependent measures [CSPV15].

2 A Geometric Point of View

Not only analytical properties are conserved but also some interesting geometrical properties of standard optimal transport such as the Riemannian submersion of Otto [Ott01]. Namely, for a fixed reference measure ρ_0, the map $\varphi \mapsto \varphi_*(\rho_0)$ from the group of diffeomorphisms of M with the $L^2(\rho_0)$ metric into the space of densities with the Wasserstein L^2 metric. See the appendix of [KW08] for more details. This property is proved simply by passing from the Eulerian point of view of the formulation (1) to a Lagrangian formulation. In this section, we extend this property to the generalized model.

A cone metric: Let us first discuss informally what happens for a particle of mass $m(t)$ at a spatial position $x(t)$ in a Riemannian manifold (M, g) under the generalized continuity constraint (3). The system reads

$$\begin{cases} \dot{x}(t) = v(x(t)) \\ \dot{m}(t) = \alpha(x(t))m(t) \end{cases} \tag{5}$$

where $\alpha \stackrel{\text{def.}}{=} \frac{\mu}{\rho}$ is the growth rate. The action associated with the action defined in (4) is $\int_0^1 |v(x(t))|^2 m(t) + \frac{\dot{m}(t)^2}{m(t)}\,\mathrm{d}t$. Thus, considering the particle as a point in $M \times \mathbb{R}_+^*$, the Riemannian metric seen by the particle is $mg + \frac{\mathrm{d}m^2}{m}$. Using the change of variable $r = \sqrt{m}$, we get $r^2 g + 4\,\mathrm{d}r^2$ which is known under the name of cone metric in Riemannian geometry. This is a flat metric when M is flat and if $M = \mathbb{R}$, a local isometry with the Euclidean space is given by $(x, m) \mapsto \sqrt{m}e^{ix/2} \in \mathbb{C}$. The distance on $M \times \mathbb{R}_+^*$ is explicit in terms of the distance on M with a Riemannian metric g,

$$\frac{1}{4}d((x_1,m_1),(x_2,m_2))^2 = m_2 + m_1 - 2\sqrt{m_1 m_2}\cos\left(\frac{1}{2}d_M(x_1,x_2)\wedge\pi\right)\,. \tag{6}$$

This implies that mass can appear and disappear at a finite cost. In other words, the cone metric is not complete but adding the vertex of the cone, which represents $M \times \{0\}$, to $M \times \mathbb{R}_+^*$ turns it into a complete metric space.

Note that this distance squared is 1-homogeneous in (m_1, m_2).

A semi-direct product of groups: Going from Eulerian to Lagrangian coordinates in this new model is properly done by introducing a semi-direct product of group that extends the group of diffeomorphisms by introducing an action on mass that can be described as pointwise multiplication with a positive function on M. Working in a smooth context, we define $\Lambda(M) \stackrel{\text{def.}}{=} \{\lambda \in C^\infty(M, \mathbb{R}) \, : \, \lambda > 0\}$. It is a group under pointwise multiplication. We will also denote the same space as $\mathrm{Dens}(M)$ to represent densities, that are smooth and positive L^1 function w.r.t. a reference measure ν. We define the semi-direct product of group between $\mathrm{Diff}(M)$ and $\Lambda(M)$ in order to turn the map π defined by

$$\begin{aligned} &\pi : (\mathrm{Diff}(M) \ltimes_\Psi \Lambda(M)) \times \mathrm{Dens}(M) \mapsto \mathrm{Dens}(M) \\ &\pi\left((\varphi, \lambda), \rho\right) \stackrel{\text{def.}}{=} \varphi \cdot \lambda \, \varphi_* \rho = \varphi_*(\lambda \rho) \end{aligned}$$

into a left-action of the group $\mathrm{Diff}(M) \ltimes_\Psi \Lambda(M)$ on the space of (generalized) densities. The group composition law is defined by:

$$(\varphi_1, \lambda_1) \cdot (\varphi_2, \lambda_2) = (\varphi_1 \circ \varphi_2, (\lambda_1 \circ \varphi_2)\lambda_2) \tag{7}$$

The important result is the following:

Proposition 1. *Let $\rho_0 \in \mathrm{Dens}(M)$ and $\pi_0 : \mathrm{Diff}(M) \ltimes_\Psi \Lambda(M) \mapsto \mathrm{Dens}(M)$ be the map defined by $\pi_0(\varphi, \lambda) \stackrel{\text{def.}}{=} \varphi_*(\lambda \rho_0)$.*

Then, the map π_0 is a Riemannian submersion of the metric $L^2(M, M \times \mathbb{R}_+^*)$ (where $M \times \mathbb{R}_+^*$ is endowed with the cone metric (6)) on the group $\mathrm{Diff}(M) \ltimes_\Psi \Lambda(M)$ to the Wasserstein-Fisher-Rao on the space of generalized densities $\mathrm{Dens}(M)$.

A direct application of this result is the formal computation of the sectional curvature of the Wasserstein-Fisher-Rao in this smooth setting by applying O'Neill's formula, see [CPSV15].

The corresponding Monge formulation: Another important consequence of the L^2 metric on the group is that one can define a Monge formulation of the Wasserstein-Fisher-Rao metric as follows:

$$WF(\rho_0, \rho_1) = \inf_{(\varphi,\lambda)} \left\{ \|(\varphi, \lambda) - (Id, 1)\|_{L^2(\rho_0)} \, : \, \varphi_*(\lambda \rho_0) = \rho_1 \right\} \tag{8}$$

3 The Kantorovich Formulation

From a variational point of view, it is important to derive a relaxation of the Monge formulation. It is of interest to understand first the simple situation when the source and target measures are single Dirac masses and when M is a convex and compact domain in the Euclidean space [CSPV15].

Proposition 2. *Let M be a convex and compact domain in $\mathbb{R}^d$ with the Euclidean metric. Let $m_1 \delta_{x_1}$ and $m_2 \delta_{x_2}$ be two Dirac masses.*

If $\frac{1}{2} d(x_1, x_2) < \pi/2$, there exists a unique geodesic which is $m(t)\delta_{x(t)}$ where $(x(t), m(t))$ is the geodesic in $M \times \mathbb{R}_+^*$ with the cone metric between (x_1, m_1) and (x_2, m_2).

If $\frac{1}{2} d(x_1, x_2) > \pi/2$, there exists a unique geodesic which is $m_1(t)\delta_{x_1} + m_2(t)\delta_{x_2}$ where $m_1(t) = m_1(1-t)^2$ and $m_2(t) = m_2 t^2$ which describe the geodesics between (x_i, m_i) and the vertex of the cone denoted by $\mathcal{S}$ for $i = 1, 2$.

If $\frac{1}{2} d(x_1, x_2) = \pi/2$, there exists an infinite number of geodesics which are interpolations of the two first types defined above.

The important point is that passing to the case of measures the angle of the cone has been (surprisingly) divided by 2. This is because we are not looking for geodesics on $M \times \mathbb{R}_+^*$ but on the space of measures on M. The generalization to any positive Radon measures gives a Kantorovich relaxation: For two given positive Radon measures ρ_1, ρ_2, we define, for $\mathcal{M}(M^2)$ the space of positive Radon measures on M^2,

$$\Gamma(\rho_1, \rho_2) \stackrel{\text{def.}}{=} \left\{ (\gamma_1, \gamma_2) \in \left(\mathcal{M}_+(M^2)\right)^2 \colon (\mathrm{Proj}_1)_* \gamma_1 = \rho_1,\, (\mathrm{Proj}_2)_* \gamma_2 = \rho_2 \right\}, \tag{9}$$

where Proj_1 and Proj_2 denote the projection on the first and second factors of M^2. The variational problem associated with the Wasserstein-Fisher-Rao distance is

$$WF(\rho_1, \rho_2)^2 = \inf_{(\gamma_1, \gamma_2) \in \Gamma(\rho_1, \rho_2)} \int_{M^2} d^2 \left((x, \frac{\mathrm{d}\gamma_1}{\mathrm{d}\gamma}), (y, \frac{\mathrm{d}\gamma_2}{\mathrm{d}\gamma}) \right) \mathrm{d}\gamma(x, y)\,, \tag{10}$$

where d^2 is the square of the cone distance defined in (6) and γ is any measure that dominates ρ_1 and ρ_2. The fact that the integration does not depend on this choice is because of the 1-homogeneity of d^2 in function of the mass. We also state the dual formulation:

Proposition 3. *It holds*

$$WF^2(\rho_0, \rho_1) = \sup_{(\phi, \psi) \in C(M)^2} \int_M \phi(x)\, \mathrm{d}\rho_0 + \int_M \psi(y)\, \mathrm{d}\rho_1 \tag{11}$$

subject to $\forall (x, y) \in M^2$,

$$\begin{cases} \phi(x) \leq 1\,, \quad \psi(y) \leq 1\,, \\ (1 - \phi(x))(1 - \psi(y)) \geq \cos^2\left(|x - y|/2\right) \end{cases} \tag{12}$$

For numerical computation, this formulation can be further reduced with a change of variable given by taking the logarithm of the multiplicative constraint (12).

4 Conclusion

We generalized the Wasserstein L^2 distance to a Riemannian-like metric on the space of densities whose total masses are different. Of important interest for application is that a static formulation is equivalent to the original dynamic one, which reduces the computational time. This Wasserstein-Fisher-Rao distance might be a useful tool in applications: On one hand, it can be seen as a modification of the Fisher-Rao metric that is stable under small spatial deformations and on the other hand as a modification of the Wasserstein metric which does not allow for mass transfer if masses are too far apart (note once again that mass creation and destruction is enabled due to the cone metric).

This natural generalization introduces a cone metric on the product between space and mass. In a smooth setting, it is possible to formally apply O'Neill's formula to obtain the sectional curvature of the space of generalized densities. However, we did not study the global geometry of the space: one expects that, as for the Euclidean cone, the curvature is concentrated at its singularity. We refer to [CSPV15, CPSV15] for more details and generalizations.

After the presentation at the workshop, two important papers [LMS15b, LMS15a] also appeared on the same model motivated by different applications.

References

[AKS15] Martial Agueh, Boualem Khouider, and Louis-Philippe Saumier. Optimal transport for particle image velocimetry. *Communications in Mathematical Sciences*, 13(1):269–296, 2015.

[BB00] J-D. Benamou and Y. Brenier. A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem. *Numerische Mathematik*, 84(3):375–393, 2000.

[CPSV15] L. Chizat, G. Peyré, B. Schmitzer, and F.-X. Vialard. Unbalanced Optimal Transport: Geometry and Kantorovich Formulation. *ArXiv e-prints*, August 2015.

[CSPV15] L. Chizat, B. Schmitzer, G. Peyré, and F.-X. Vialard. An Interpolating Distance between Optimal Transport and Fisher-Rao. *ArXiv e-prints*, June 2015.

[Cut13] Marco Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport. In C.J.C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K.Q. Weinberger, editors, *Advances in Neural Information Processing Systems 26*, pages 2292–2300. Curran Associates, Inc., 2013.

[FG10] A. Figalli and N. Gigli. A new transportation distance between non-negative measures, with applications to gradients flows with dirichlet boundary conditions. *Journal de mathématiques pures et appliquées*, 94(2):107–130, 2010.

[Gui02] K. Guittet. Extended Kantorovich norms: a tool for optimization. Technical report, Tech. Rep. 4402, INRIA, 2002.

[HZTA04] S. Haker, L. Zhu, A. Tannenbaum, and S. Angenent. Optimal mass transport for registration and warping. *International Journal of computer vision*, 60(3):225–240, 2004.

[KW08] B. Khesin and R. Wendt. *The geometry of infinite-dimensional groups*, volume 51. Springer Science & Business Media, 2008.

[LM13] D. Lombardi and E. Maitre. Eulerian models and algorithms for unbalanced optimal transport. <hal-00976501v3>, 2013.

[LMS15a] M. Liero, A. Mielke, and G. Savaré. Optimal Entropy-Transport problems and a new Hellinger-Kantorovich distance between positive measures. *ArXiv e-prints*, August 2015.

[LMS15b] M. Liero, A. Mielke, and G. Savaré. Optimal transport in competition with reaction: the Hellinger-Kantorovich distance and geodesic curves. *ArXiv e-prints*, August 2015.

[MBM14] M.Bauer, M. Bruveris, and P. W. Michor. Uniqueness of the Fisher-Rao metric on the space of smooth densities. arXiv:1411.5577, 2014.

[MRSS15] J. Maas, M. Rumpf, C. Schönlieb, and S. Simon. A generalized model for optimal transport of images including dissipation and density modulation. arXiv:1504.01988, 2015.

[Ott01] Felix Otto. The geometry of dissipative evolution equations: The porous medium equation. *Communications in Partial Differential Equations*, 26(1-2):101–174, 2001.

[PR13] B. Piccoli and F. Rossi. On properties of the Generalized Wasserstein distance. arXiv:1304.7014, 2013.

[PR14] B. Piccoli and F. Rossi. Generalized Wasserstein distance and its application to transport equations with source. *Archive for Rational Mechanics and Analysis*, 211(1):335–358, 2014.

[TY05] A. Trouvé and L. Younes. Metamorphoses through lie group action. *Foundations of Computational Mathematics*, 5(2):173–198, 2005.