

Variational Heuristics for Optimal Transportation Maps on Compact Manifolds*

Ph. Delanoë[†]

Abstract

Variational derivation of the expression of the solution of Monge's problem posed on compact manifolds (possibly with boundary), assuming all data are smooth, the solution is a diffeomorphism and the cost function satisfies a generating type condition.

1 Introduction

Let X be a compact connected manifold (all objects are C^∞ unless otherwise stated). We assume that, either X has no boundary, or it is a domain contained in some larger manifold. Let Y be a manifold diffeomorphic to X and μ (resp. ν) an everywhere positive probability measure on X (resp. Y). We say that a map $\varphi : X \rightarrow Y$ pushes μ to ν , and write $\varphi\#\mu = \nu$, if the following equality¹:

$$(1) \quad \int_Y h \, d\nu = \int_X (h \circ \varphi) \, d\mu$$

holds for each function $h : Y \rightarrow \mathbb{R}$. Furthermore, let $c : \Omega \subset X \times Y \rightarrow \mathbb{R}$ be a function defined in a domain² Ω projecting onto X (resp. Y) by the canonical first (resp. second) projection p_X (resp. p_Y). In this setting, Monge's problem consists in minimizing the functional

$$(2) \quad \mathcal{C}(\varphi) = \int_X c(x, \varphi(x)) \, d\mu$$

among the Borel maps $\varphi : X \rightarrow Y$ which satisfy the constraint:

$$(3) \quad \varphi\#\mu = \nu .$$

Optimal transportation is the theory designed for tackling such problems (see [Vil09] and references therein). In general, even though the present data are smooth, one can only solve a relaxed version of Monge's problem due to Kantorovich; but if the cost function c satisfies a weak form of the generating condition stated below, Monge's problem becomes solvable [Vil09, Theorem 5.30]

*2000 MSC: 58E99; 58D05; 58A10; **Key-words:** Monge's problem, generating cost function, c -exponential, Euler equation, Helmholtz lemma

[†]Supported by the CNRS

¹here, we may allow φ and h to be merely Borel

²we make no smoothness assumption on its boundary

(a fact first observed by Brenier [Bre91, p.404]). Further restrictions on the cost function c have been made to analyze the smoothness of the optimal map [Vil09, Chap.12] among which, in first place, a generating type condition (like the one below), referred to as condition (A1) [MTW05, T-W07] or as twist condition [K-M09], which suffices to derive the shape of the optimal map, provided it is a diffeomorphism. Working in \mathbb{R}^n , Evans, followed by Trudinger, found a tricky variational device to do that derivation [Eva97, Tru06, Tru07]; they use pointwise Lagrange multipliers to cope with the constraint (3) which, bearing on diffeomorphisms, becomes a (pointwise) Jacobian equation. Clearly, a way of writing down directly an Euler equation for the total cost functional (2) would be preferable. Urbas took a step in this direction in case the measure μ (resp. ν) is just the standard Lebesgue measure of \mathbb{R}^n restricted to the domain X (resp. Y) and the function $c(x, y)$ factors through a function of the single variable $x - y$. He wrote the Euler equation of the total cost functional pulled back to the diffeomorphisms of X which preserve the Lebesgue measure and the points located close enough to the boundary ∂X ; recalling that any such diffeomorphism must be the flow of a divergence free vector field compactly supported in X , he readily inferred the desired characterization of the optimal map [Urb98, Section 3]. Our present aim is to show that a similar, equally simple, method applies in a general setting. We obtain probably the straightest heuristical way of getting the optimal map expression (see Section 4 below).

First of all, we must specify a further requirement on the cost function c . For each $(x_0, y_0) \in X \times Y$, our assumption on the domain Ω implies that the counter-sets:

$$\Omega_X(y_0) := \{x \in X, (x, y_0) \in \Omega\}, \quad \Omega_Y(x_0) := \{y \in Y, (x_0, y) \in \Omega\},$$

are non-empty; we will consider the maps

$$F_{y_0} : x \in \Omega_X(y_0) \rightarrow (-d_y c)(x, y_0) \in T_{y_0}^* Y,$$

$$F_{x_0} : y \in \Omega_Y(x_0) \rightarrow (-d_x c)(x_0, y) \in T_{x_0}^* X.$$

Clearly, the associated mappings:

$$F_Y : (x, y) \in \Omega \rightarrow (y, F_y(x)) \in T^* Y, \quad F_X : (x, y) \in \Omega \rightarrow (x, F_x(y)) \in T^* X,$$

are smooth; we set $D_Y \subset T^* Y$ and $D_X \subset T^* X$ respectively for their image domains. We will require that the cost function c fulfill the following

Generating Condition:

For each $(x_0, y_0) \in X \times Y$, the maps F_{x_0} and F_{y_0} are one-to-one.

Under that condition, we see that the maps F_X and F_Y are one-to-one as well. Setting $F_Y^{-1} : D_Y \rightarrow \Omega$ and $F_X^{-1} : D_X \rightarrow \Omega$ respectively for their set-theoretic inverses³, and

$$E_Y := p_X \circ F_Y^{-1}, \quad E_X := p_Y \circ F_X^{-1},$$

³they are smooth provided F_X (hence also F_Y) has non-vanishing Jacobian determinant, a condition called (A2) in [MTW05, T-W07, K-M09] and *not* required below

for the significant factors of the latter, the following essential property now holds: if $(x, \alpha_x) \in X \times T_x^*X$ lies in D_X (resp. if $(y, \beta_y) \in D_Y$), the point $y = E_X(\alpha_x)$ (resp. $x = E_Y(\beta_y)$) is the unique solution-point of the equation:

$$(4) \quad (-d_x c)(x, y) = \alpha_x \quad (\text{resp. } (-d_y c)(x, y) = \beta_y).$$

One may call the maps E_X, E_Y , c -exponentials⁴. An important example⁵ of cost function satisfying our assumptions occurs when $X = Y$ and $c = \frac{1}{2} d_g^2$ with d_g the distance function defined by a Riemannian metric g on X . In particular, the Generating Condition holds because then c is the generating function of the co-geodesic flow, as well-known [Arn76], hence the map $E_X = E_Y$ is nothing but the usual exponential map (pulled back from TX to T^*X by g). Back to our general setting, we can now state the heuristical result to be proven below by means of a new variational argument:

Proposition 1 *Let the cost function c satisfy the Generating Condition and assume the existence of a diffeomorphism $\varphi : X \rightarrow Y$ minimizing (2) under the constraint (3). There exists a function $f : X \rightarrow \mathbb{R}$ (resp. $f^c : Y \rightarrow \mathbb{R}$) such that df (resp. df^c) ranges in D_X (resp. D_Y) and:*

$$\varphi = E_X \circ df \quad (\text{resp. } \varphi^{-1} = E_Y \circ df^c).$$

Moreover, the function $(x, y) \mapsto f(x) + f^c(y) + c(x, y)$ is constant on the graph of the map φ .

Remark 1 Under the assumption of Proposition 1 and if the manifolds X and Y have no boundary, we see that the domains D_X and D_Y necessarily cross the zero sections of the cotangent bundles T^*X and T^*Y . Specifically, they must contain neighborhoods of the zero sections above *critical points* of the transporting potentials f and f^c respectively.

The outline of the note is the following. We first recall the notion of tangent vector at the identity of a diffeomorphisms group and the definition of the divergence operator associated with a positive measure, inferring the well-known characterization of the Lie algebra of the corresponding group of measure-preserving diffeomorphisms (Section 2). We use this in Section 3 for expressing the variation of a diffeomorphism constrained by (3). Finally (Section 4), we write down the Euler equation of the cost functional (2) restricted to such diffeomorphisms, plugging in the variation expression just obtained and we conclude by applying a lemma of Helmholtz type.

2 Preliminary tools

Let us set $\text{Diff}(X)$ for the group of diffeomorphisms sending the manifold X to itself and $\Gamma_t(TX)$ for the Fréchet space of vector fields on X which, in case $\partial X \neq \emptyset$, are *tangential*. By the latter, we mean that the restriction of each such vector field to ∂X is a section of $T(\partial X)$. In other words, a vector field on X lies in $\Gamma_t(TX)$ if and only if its *flow is complete* on X . Of course, the subscript

⁴Kim and McCann call them cost-exponentials [K-M09]

⁵why we choose the minus sign normalization appearing in (4)

of Γ_t is superfluous in case $\partial X = \emptyset$, but it is convenient to keep it for treating both cases at the same time.

For $t \in \mathbb{R}$ small, let $t \mapsto \varphi_t \in \text{Diff}(X)$ be a path such that $\varphi_0 = I$ (the identity). It is simple but important⁶ to verify that the vector field V defined on X by:

$$x \in X \rightarrow V_x := \left. \frac{d\varphi_t(x)}{dt} \right|_{t=0}$$

must be tangential: $V \in \Gamma_t(TX)$.

Setting $\mathcal{P}(X)$ (resp. $\mathcal{M}_0(X)$) for the set of positive probability (resp. zero average) measures on X , we observe that $\mathcal{P}(X)$ is a domain in an *affine* space modelled on the Fréchet space $\mathcal{M}_0(X)$; in particular, the tangent bundle $T\mathcal{P}(X)$ may be identified with the Cartesian product $\mathcal{P}(X) \times \mathcal{M}_0(X)$.

Identifying the elements of $\mathcal{P}(X)$ with volume-forms, possibly of odd type in case X is not orientable [dRh55], one can readily check that the push-forward $\varphi_{\#}\mu$ of our above measure $\mu \in \mathcal{P}(X)$ by $\varphi \in \text{Diff}(X)$ coincides with the pull-back of the form μ by the inverse diffeomorphism φ^{-1} . We infer that $\varphi_{\#}\mu \in \mathcal{P}(X)$. Setting C_μ^∞ for the Fréchet space of functions $f : X \rightarrow \mathbb{R}$ such that $f\mu \in \mathcal{M}_0(X)$, we are ready to state the following convenient definition:

Definition 1 (Divergence operator) *The divergence operator associated to $\mu \in \mathcal{P}(X)$ is a first order differential operator $\text{div}_\mu : \Gamma_t(TX) \rightarrow C_\mu^\infty$ which can be defined by:*

$$(5) \quad \forall V \in \Gamma_t(TX), \quad \text{div}_\mu(V)\mu := \left. \frac{d}{dt} (\varphi_{t\#}\mu) \right|_{t=0} ,$$

where φ_t stands for the flow of the tangential vector field V .

The range of the operator div_μ is clear because (cf. supra) the right-hand side of (5) lies in $\mathcal{M}_0(X)$. In order to recognize the usual divergence operator, we just fix $V \in \Gamma_t(TX)$ with flow φ_t and differentiate with respect to t at $t = 0$ equation (1) written with $\varphi = \varphi_t$ and h smooth arbitrary. It yields the important identity:

$$(6) \quad \int_X h \text{div}_\mu(V) d\mu \equiv \int_X dh(V) d\mu ,$$

familiar looking, indeed. Finally, let us consider the subgroup:

$$\text{Diff}_\mu(X) := \{\xi \in \text{Diff}(X), \xi_{\#}\mu = \mu\}$$

of μ -preserving diffeomorphisms of X . It is a Fréchet Lie group with Lie algebra $T_I \text{Diff}_\mu(X)$ equal to $\ker \text{div}_\mu$ [E-M70]; the latter equality can readily be figured out by the reader using the above material (exercise).

3 Varying constrained diffeomorphisms

Let $\text{Diff}(Y, X)$ denote the set of diffeomorphisms from Y to X and $\text{Diff}_{\nu, \mu}(Y, X)$ the subset of those which satisfy the constraint:

$$(7) \quad \psi_{\#}\nu = \mu ,$$

⁶we leave it to the reader as an easy exercise

sticking to the measures μ and ν given in the Introduction. We wish to identify the tangent space to $\text{Diff}_{\nu,\mu}(Y, X)$ at a generic constrained diffeomorphism ψ . First of all, we require a Fréchet manifold structure. We define it by declaring that, for some fixed $\psi \in \text{Diff}_{\nu,\mu}(Y, X)$ hence for any, the map:

$$\xi \in \text{Diff}_\mu(X) \rightarrow \xi \circ \psi \in \text{Diff}_{\nu,\mu}(Y, X)$$

is smooth, thus viewing $\text{Diff}_{\nu,\mu}(Y, X)$ as locally modelled on the Fréchet manifold $\text{Diff}_\mu(X)$. Doing so, we can write each path $t \mapsto \psi_t \in \text{Diff}_{\nu,\mu}(Y, X)$, defined (say) for $t \in \mathbb{R}$ small with $\psi_0 = \psi$, uniquely as $\psi_t = \xi_t \circ \psi$ where $t \mapsto \xi_t$ is a path in $\text{Diff}_\mu(X)$ satisfying $\xi_0 = I$. Accordingly, recalling our characterization of $T_I \text{Diff}_\mu(X)$, the tangent vector $\dot{\psi}_0 = \left. \frac{d\psi_t}{dt} \right|_{t=0} \in T_\psi \text{Diff}_{\nu,\mu}(Y, X)$ is equal to $\dot{\xi}_0 \circ \psi$ with $\dot{\xi}_0 \in \ker \text{div}_\mu$. We may conclude that

$$(8) \quad T_\psi \text{Diff}_{\nu,\mu}(Y, X) = \{V \circ \psi, V \in \ker \text{div}_\mu\} .$$

Similarly, switching (X, μ) and (Y, ν) , one would find that

$$(9) \quad T_\varphi \text{Diff}_{\mu,\nu}(X, Y) = \{W \circ \varphi, W \in \ker \text{div}_\nu\} .$$

4 Euler equation, interpretation

Let $\text{Diff}_{\mu,\nu}^\Omega(X, Y)$ (resp. $\text{Diff}_{\nu,\mu}^\Omega(Y, X)$) denote the open subset of $\text{Diff}_{\mu,\nu}(X, Y)$ (resp. $\text{Diff}_{\nu,\mu}(Y, X)$) consisting of diffeomorphisms whose graphs in $X \times Y$ lie in the domain Ω in which the given cost function c is defined. Let $\bar{\mathcal{C}}$ be the restriction of the total cost functional \mathcal{C} given by (2) to the manifold $\text{Diff}_{\mu,\nu}^\Omega(X, Y)$, and $\tilde{\mathcal{C}} : \text{Diff}_{\nu,\mu}^\Omega(Y, X) \rightarrow \mathbb{R}$ be the functional given by:

$$(10) \quad \tilde{\mathcal{C}}(\psi) := \int_Y c(\psi(y), y) d\nu .$$

These functionals are related by:

$$\bar{\mathcal{C}}(\phi) \equiv \tilde{\mathcal{C}}(\phi^{-1})$$

(easy exercise). If $\varphi \in \text{Diff}_{\mu,\nu}^\Omega(X, Y)$ minimize $\bar{\mathcal{C}}$, as assumed in Proposition 1, setting $\psi = \varphi^{-1} \in \text{Diff}_{\nu,\mu}^\Omega(Y, X)$, we infer that ψ minimizes $\tilde{\mathcal{C}}$. We will prove Proposition 1 by writing the Euler equations of $\tilde{\mathcal{C}}$ and $\bar{\mathcal{C}}$ and interpreting them.

Let us do it for $\tilde{\mathcal{C}}$; using (8), we may write:

$$\forall V \in \ker \text{div}_\mu, \quad \int_Y (d_x c)(\psi(y), y)(V \circ \psi(y)) d\nu = 0 .$$

The expression of the preceding integrand prompts us⁷ to consider the 1-form α defined on X by:

$$(11) \quad \forall x \in X, \quad \alpha_x := (-d_x c)(x, \varphi(x)) ,$$

⁷except for the minus sign, inessential here

which thus ranges in D_X . Using the form α , the Euler equation of $\tilde{\mathcal{C}}$ becomes:

$$(12) \quad \forall V \in \ker \operatorname{div}_\mu, \quad \int_X \alpha(V) \, d\mu = 0,$$

as readily checked. The latter can be interpreted as the necessary and sufficient condition under which there exists a function $f : X \rightarrow \mathbb{R}$ such that $\alpha = df$, due to a classical lemma of Helmholtz type (see Appendix A). Combining (11) with (4), we infer the shape of the optimal map φ , namely:

$$\forall x \in X, \quad \varphi(x) = E_X(df_x),$$

as claimed. A similar argument carried out for $\bar{\mathcal{C}}$ in place of $\tilde{\mathcal{C}}$, now using (9), yields the existence of a function $f^c : Y \rightarrow \mathbb{R}$ such that:

$$\forall y \in Y, \quad \psi(y) = E_Y(df_y^c).$$

Finally, from these last two equations⁸ combined with (4), we immediately get

$$df_x + df_y^c + dc_{(x,y)} = 0$$

on the graph of φ . Since the latter is connected, the function

$$(x, y) \mapsto f(x) + f^c(y) + c(x, y)$$

must, indeed, be constant on it. The proof of Proposition 1 is complete.

A Helmholtz lemma

In the course of Section 4, we used the following lemma, which we state sticking to the assumptions made on the manifold X (cf. Introduction):

Lemma 1 *A 1-form α on X satisfies (12) if and only if there exists a function $f : X \rightarrow \mathbb{R}$ such that $\alpha = df$.*

Proof. The 'if' part of the lemma follows from (6). The 'only if' part is fully standard when X has no boundary (see e.g. [Del09, Appendix B]); if $\partial X \neq \emptyset$, it is considered as classical in [E-M70, case (iii) of Theorem 7.1]. For completeness, let us sketch a proof of that case.

We require an auxiliary Riemannian metric g on X ; for simplicity, we take it with volume measure equal to μ . Set $\sharp : T^*X \rightarrow TX$ for its Riesz isomorphism, $\nabla = \sharp d$ (resp. $\Delta = \operatorname{div}_\mu(\nabla)$) for its gradient (resp. Laplace) operator, d_g for the Riemannian distance and N for the inward unit normal vector field on ∂X . We split the proof into three steps.

Step 1: cook up a function $h : X \rightarrow \mathbb{R}$ such that $dh(N) \equiv \alpha(N)$ on ∂X . To proceed bare-handed, we pick the function $\varrho(x) = \text{distance from } x \text{ to } \partial X$. For each point x lying in a small enough neighbourhood \mathcal{N} of ∂X , there exists a unique point $p(x) \in \partial X$ such that $\varrho(x) = d_g(x, p(x))$ with the map p and the function ϱ smooth on \mathcal{N} . Using a cut-off function $t \in [0, \infty) \rightarrow \eta(t) \in [0, 1]$ such that $\eta(t) = 1$ for t small and $\eta \circ \varrho = 0$ outside \mathcal{N} , the function h defined by:

$$\forall x \in X, \quad h(x) = \eta(\varrho(x)) \varrho(x) \alpha_{p(x)}(N_{p(x)})$$

⁸recall $\psi = \varphi^{-1}$ (cf. supra)

is readily seen to fulfill the requirement of Step 1.

Step 2: solve the Neumann problem:

$$\Delta u = v \text{ in } X, \quad u(N) = 0 \text{ on } \partial X,$$

with⁹ $v = \operatorname{div}_\mu(\sharp\alpha - \nabla h)$. Since $v \in C_\mu^\infty$, this linear elliptic problem admits a unique¹⁰ solution $u \in C_\mu^\infty$, as well-known [Hör76, Theorem 10.5.1][Che84, Lemma 1][K-L08, Lemma 2.10].

Step 3: set $f := h + u$ and check that, due to (12), the $L^2(X, g)$ norm of the vector field $\xi := \sharp\alpha - \nabla f$ vanishes. We leave this step as an exercise; hint: use the fact that $\xi \in \ker \operatorname{div}_\mu$ and recall (6). We conclude: $\xi = 0$ or else $\alpha = df$, as claimed.

Acknowledgments.

I thank Craig Evans and Neil Trudinger, who read the first draft of this note, for their encouraging and keen interest. Neil Trudinger mentioned to me the reference [Urb98]; I thank John Urbas for giving me a quick access to it.

References

- [Arn76] V. Arnol'd, *Méthodes Mathématiques de la Mécanique Classique*, Mir, Moscou (1976)
- [Bre91] Y. Brenier, Polar factorization and monotone rearrangement of vector-valued functions, *Comm. Pure Appl. Math.* **44** (1991) 375–417
- [Che84] P. Cherrier, Problèmes de Neumann non linéaires sur les variétés riemanniennes, *J. Funct. Anal.* **57** (1984) 154–206
- [Del09] Ph. Delanoë, Differential geometric heuristics for Riemannian optimal transportation maps, in *Proceedings Abel Symposium 2008*, B. Kruglikov, V. Lychagin & E. Straume Editors, Springer–Verlag (to appear)
- [E-M70] D. Ebin, J. Marsden, Groups of diffeomorphisms and the motion of an incompressible fluid, *Ann. of Math.* **92** (1970) 102–163
- [Eva97] L. C. Evans, Partial differential equations and Monge–Kantorovich mass transfer, in *Current Developments in Mathematics 1997*, International Press, Boston MA (1998), pp.65–126 (latest version downloadable at <http://math.berkeley.edu/~evans/>)
- [Hör76] L. Hörmander, *Linear Partial Differential Operators*, Fourth Printing, Springer–Verlag (1976)
- [K-L08] B. Khesin, P. Lee, A nonholonomic Moser Theorem and optimal mass transport, Preprint (2008) downloadable at: <http://arxiv.org/abs/0802.1551>
- [K-M09] Y.-H. Kim, R. McCann, Continuity, curvature and the general covariance of optimal transportation, *J. Eur. Math. Soc.* (to appear)

⁹note that $\sharp\alpha - \nabla h$ is tangential by Step 1

¹⁰if z is the difference between two solutions, integrating $z\Delta z = 0$ over X yields $dz = 0$

- [MTW05] X.-N. Ma, N. Trudinger, X.-J. Wang, Regularity of potential functions of the optimal transportation problem, *Arch. Rat. Mech. Anal.* **177** (2005) 151–183
- [dRh55] G. de Rham, *Variétés différentiables*, Hermann, Paris (1955)
- [T-W07] N. Trudinger, X.-J. Wang, On the second boundary value problem for Monge–Ampère type equations and optimal transportation, Preprint (2006), revised version (2007) downloadable at: <http://wwwmaths.anu.edu.au/~neilt/RecentPapers.html>
- [Tru06] N. Trudinger, On the regularity of potentials in optimal transportation, talk at the meeting "Optimal Transport and geometric PDE's" (Nice, 15-18 June 2006)
- [Tru07] N. Trudinger, Optimal transportation and nonlinear partial differential equations, (slides), **26th Brazilian Mathematical Colloquium**, August 2007, downloadable at: <http://wwwmaths.anu.edu.au/~neilt/RecentPapers.html>
- [Urb98] J. Urbas, Mass transfer problems, *Lecture Notes* (unpublished) Univ. Bonn (1998)
- [Vil09] C. Villani, *Optimal Transport, Old and New*, Grundlehren der mathematischen Wissenschaften **338**, Springer–Verlag (2009)

Université de Nice–Sophia Antipolis
Laboratoire J.–A. Dieudonné, Parc Valrose
F-06108 Nice CEDEX 2
e-mail: Philippe.DELANOE@unice.fr