

A NONHOLONOMIC MOSER THEOREM AND OPTIMAL TRANSPORT

BORIS KHESIN AND PAUL LEE

We prove the following nonholonomic version of the classical Moser theorem: given a bracket-generating distribution on a connected compact manifold (possibly with boundary), two volume forms of equal total volume can be isotoped by the flow of a vector field tangent to this distribution. We describe formal solutions of the corresponding nonholonomic mass transport problem and present the Hamiltonian framework for both the Otto calculus and its nonholonomic counterpart as infinite-dimensional Hamiltonian reductions on diffeomorphism groups.

Finally, we define a nonholonomic analog of the Wasserstein (or, Kantorovich) metric on the space of densities and prove that the subriemannian heat equation defines a gradient flow on the nonholonomic Wasserstein space with the potential given by the Boltzmann relative entropy functional.

CONTENTS

1. Introduction	382
2. Around Moser’s theorem	383
2.1. Classical and nonholonomic Moser theorems	383
2.2. The Moser theorem for a fibration	384
2.3. Proofs	385
2.4. The nonholonomic Hodge decomposition and sub-Laplacian	388
2.5. The case with boundary	389
3. Distributions on diffeomorphism groups	391
3.1. A fibration on the group of diffeomorphisms	391
3.2. A nonholonomic distribution on the diffeomorphism group	393

3.3. Accessibility of diffeomorphisms and symplectic structures	394
4. The Riemannian geometry of diffeomorphism groups and mass transport	395
4.1. Optimal mass transport	395
4.2. The Otto calculus	396
5. The Hamiltonian mechanics on diffeomorphism groups	398
5.1. Averaged Hamiltonians	398
5.2. Riemannian submersion and symplectic quotients	399
5.3. Hamiltonian flows on the diffeomorphism groups	401
5.4. Hamiltonian flows on the Wasserstein space	404
6. The subriemannian geometry of diffeomorphism groups	405
6.1. Subriemannian submersion	406
6.2. A subriemannian analog of the Otto calculus.	409
6.3. The nonholonomic heat equation	411
References	413

1. Introduction

The classical Moser theorem establishes that the total volume is the only invariant for a volume form on a compact connected manifold with respect to the diffeomorphism action. In this paper we prove a nonholonomic counterpart of this result and present its applications in the problems of nonholonomic optimal mass transport.

The equivalence for the diffeomorphism action is often formulated in terms of “stability” of the corresponding object: the existence of a diffeomorphism relating the initial object with a deformed one means that the initial object is stable, as it differs from the deformed one merely by a coordinate change. Gray showed in [9] that contact structures on a compact manifold are stable. Moser [16] established stability for volume forms and symplectic structures. A leafwise counterpart of Moser’s argument for foliations was presented by Ghys in [8], while stability of symplectic–contact pairs in transversal foliations was proved in [4]. In this paper we establish stability of volume forms in the presence of any bracket-generating distributions on connected compact manifolds: two volume forms of equal total volume on such a manifold can be isotoped by the flow of a vector field tangent to the distribution. We call this statement a nonholonomic Moser theorem.

Recall that a distribution τ on the manifold M is called *bracket generating*, or *completely nonholonomic*, if local vector fields tangent to τ and their iterated Lie brackets span the entire tangent bundle of the manifold M . Nonholonomic distributions arise in various problems related to rolling or

skating, wherever the “no-slip” condition is present. For instance, a ball rolling over a table defines a trajectory in a configuration space tangent to a nonholonomic distribution of admissible velocities. Note that such a ball can be rolled to any point of the table and stopped at any a priori prescribed position. The latter is a manifestation of the Chow–Rashevsky theorem (see, e.g., [15]): for a bracket-generating distribution τ on a connected manifold M any two points in M can be connected by a horizontal path (i.e., a path everywhere tangent to the distribution τ).¹

Note that for an integrable distribution there is a foliation to which it is tangent and a horizontal path always stays on the same leaf of this foliation. Furthermore, for an integrable distribution, the existence of an isotopy between volume forms requires an infinite number of conditions. On the contrary, the nonholonomic Moser theorem shows that a nonintegrable bracket-generating distribution imposes only one condition on total volume of the forms for the existence of the isotopy between them.

Closely related to the nonholonomic Moser theorem is the existence of a nonholonomic Hodge decomposition, and the corresponding properties of the subriemannian Laplace operator, see Section 2.4. We also formulate the corresponding nonholonomic mass transport problem and describe its formal solutions as projections of horizontal geodesics on the diffeomorphism group for the L^2 -Carnot–Caratheodory metric.

In order to give this description, we first present the Hamiltonian framework for what is now called the Otto calculus — the Riemannian submersion picture for the problems of optimal mass transport. It turns out that the submersion properties can be naturally understood as an infinite-dimensional Hamiltonian reduction on diffeomorphism groups, and this admits a generalization to the nonholonomic setting. We define a nonholonomic analog of the Wasserstein metric on the space of densities. Finally, we extend Otto’s result on the heat equation and prove that the subriemannian heat equation defines a gradient flow on the nonholonomic Wasserstein space with potential given by the Boltzmann relative entropy functional.

2. Around Moser’s theorem

2.1. Classical and nonholonomic Moser theorems. The main goal of this section is to prove the following nonholonomic version of the classical Moser theorem. Consider a distribution τ on a compact manifold M (without boundary unless otherwise stated).

Theorem 2.1. *Let τ be a bracket-generating distribution, and μ_0, μ_1 be two volume forms on M with the same total volume: $\int_M \mu_0 = \int_M \mu_1$. Then there*

¹The motivation for considering volume forms (or, densities) in a space with distribution can be related to problems with many tiny rolling balls. It is more convenient to consider the density of such balls, rather than look at them individually.

exists a diffeomorphism ϕ of M which is the time-one-map of the flow ϕ_t of a nonautonomous vector field V_t tangent to the distribution τ everywhere on M for every $t \in [0, 1]$, such that $\phi^*\mu_1 = \mu_0$.

Note that the existence of the “nonholonomic isotopy” ϕ_t is guaranteed by the only condition on equality of total volumes for μ_0 and μ_1 , just like in the classical case:

Theorem 2.2 [16]. *Let M be a manifold without boundary, and μ_0, μ_1 are two volume forms on M with the same total volume: $\int_M \mu_0 = \int_M \mu_1$. Then there exists a diffeomorphism ϕ of M , isotopic to the identity, such that $\phi^*\mu_1 = \mu_0$.*

Remark 2.3. The classical Moser theorem has numerous variations and generalizations, some of which we would like to mention.

- (a) Similarly one can show that not only the identity, but any diffeomorphism of M is isotopic to a diffeomorphism which pulls back μ_1 to μ_0 .
- (b) The Moser theorem also holds for a manifold M with boundary. In this case a diffeomorphism ϕ is a time-one-map for a (nonautonomous) vector field V on M , tangent to the boundary ∂M .
- (c) Moser also proved in [16] a similar statement for a pair of symplectic forms on a manifold M : if two symplectic structures can be deformed to each other among symplectic structures in the same cohomology class on M , these deformations can be carried out by a flow of diffeomorphisms of M .

Below we describe to which degree these variations extend to the nonholonomic case.

2.2. The Moser theorem for a fibration. Apparently, the most straightforward generalization of the classical Moser theorem is its version “with parameters.” In this case, volume forms on M smoothly depend on parameters and have the same total volume at each value of this parameter: $\int_M \mu_0(s) = \int_M \mu_1(s)$ for all s . The theorem guarantees that the corresponding diffeomorphism exists and depends smoothly on this parameter s .

The following theorem can be regarded as a modification of the parameter version:

Theorem 2.4. *Let $\pi : N \rightarrow B$ be a fibration of an n -dimensional manifold N over a k -dimensional base manifold B . Suppose that μ_0, μ_1 are two smooth volume forms on N . Assume that the pushforwards of these n -forms to B coincide, i.e., they give one and the same k -form on B : $\pi_*\mu_0 = \pi_*\mu_1$. Then, there exists a diffeomorphism ϕ of N which is the time-one-map of a (nonautonomous) vector field V tangent everywhere to the fibers of this fibration and such that $\phi^*\mu_1 = \mu_0$.*

Remark 2.5. Note that in this version the volume forms are given on the *ambient* manifold N , while in the parameter version of the Moser theorem we are given *fiberwise* volume forms. There is also a similar version of this theorem for a foliation, cf., e.g., [8]. In either case, for the corresponding diffeomorphism to exist, the volume forms have to satisfy infinitely many conditions (the equality of the total volumes as functions in the parameter s or as the pushforwards $\pi_*\mu_0$ and $\pi_*\mu_1$). The case of a fibration (or a foliation) corresponds to an integrable distribution τ , and presents the “opposite case” to a bracket-generating distribution. Unlike the case of an integrable distribution, the existence of the corresponding isotopy between volume forms in the bracket-generating case imposes only one condition, the equality of the total volumes of the two forms (regardless, e.g., of the distribution growth vector at different points of the manifold).

2.3. Proofs. First, we recall a proof of the classical Moser theorem. To show how the proof changes in the nonholonomic case, we split it into several steps.

Proof. (1) Connect the volume forms μ_0 and μ_1 by a “segment” $\mu_t = \mu_0 + t(\mu_1 - \mu_0)$, $t \in [0, 1]$. We will be looking for a diffeomorphism g_t sending μ_t to μ_0 : $g_t^*\mu_t = \mu_0$. By taking the t -derivative of this equation, we get the following “homological equation” on the velocity V_t of the flow g_t : $g_t^*(\mathcal{L}_{V_t}\mu_t + \partial_t\mu_t) = 0$, where $\partial_t g_t(x) = V_t(g_t(x))$. This is equivalent to

$$\mathcal{L}_{V_t}\mu_t = \mu_0 - \mu_1,$$

since $\partial_t\mu_t = -(\mu_0 - \mu_1)$.

By rewriting $\mu_0 - \mu_1 = \rho_t\mu_t$ for an appropriate function ρ_t , we reformulate the equation $\mathcal{L}_{V_t}\mu_t = \rho_t\mu_t$ as the problem $\operatorname{div}_{\mu_t} V_t = \rho_t$ of looking for a vector field V_t with a prescribed divergence ρ_t . Note that the total integral of the function ρ_t (relative to the volume μ_t) over M vanishes, which manifests the equality of total volumes for μ_t .

(2) We omit the index t for now and consider a Riemannian metric on M whose volume form is μ . We are looking for a required field V with prescribed divergence among gradient vector fields $V = \nabla u$, which “transport the mass” in the fastest way. This leads us to the elliptic equation $\operatorname{div}_\mu(\nabla u) = \rho$, i.e., $\Delta u = \rho$, where the Laplacian Δ is defined by $\Delta u := \operatorname{div}_\mu \nabla u$ and depends on the Riemannian metric on M .

(3) The key part of the proof is the following

Lemma 2.6. *The Poisson equation $\Delta u = \rho$ on a compact Riemannian manifold M is solvable for any smooth function ρ with zero mean: $\int_M \rho \mu = 0$ (with respect to the Riemannian volume form μ).*

Proof. Describe the space $\text{Coker } \Delta := (\text{Im } \Delta)^{\perp L^2}$, i.e., find the space of all functions h which are L^2 -orthogonal to the image $\text{Im } \Delta$. By applying integration by parts twice, one has

$$0 = \langle h, \Delta u \rangle_{L^2} = -\langle \nabla h, \nabla u \rangle_{L^2} = \langle \Delta h, u \rangle_{L^2}$$

for all smooth functions u on M . Then such functions h must be (weakly) harmonic, and hence they are constant functions on M : $(\text{Im } \Delta)^{\perp L^2} = \{\text{const}\}$. Since the image $\text{Im } \Delta$ is closed, it is the L^2 -orthogonal complement of the space of constant functions $\text{Im } \Delta = \{\text{const}\}^{\perp L^2}$. The condition of orthogonality to constants is exactly the condition of zero mean for ρ : $\langle \text{const}, \rho \rangle_{L^2} = \int_M \rho \mu = 0$. Thus the equation $\Delta u = \rho$ has a weak solution for $\rho \in L^2(M)$ with zero mean, and the ellipticity of Δ implies that the solution is smooth for a smooth function ρ . \square

(4) Now, take $V_t := \nabla u_t$ where u_t is the solution of $\Delta u_t = \rho_t$ and let g_V^t be the corresponding flow on M . Since M is compact and V_t is smooth, the flow exists for all time t . The diffeomorphism $\phi := g_V^1$, the time-one-map of the flow g_V^t , gives the required map which pulls back the volume form μ_1 to μ_0 : $\phi^* \mu_1 = \mu_0$. \square

Proof of Theorem 2.4. The Moser theorem for a fibration: We start by defining the new volume form on the fibers F using the pushforward k -form $\nu_0 := \pi_* \mu_0$ on the base B and the volume n -form μ_0 on N . Namely, consider the pull-back k -form $\pi^* \nu_0$ on N . Then there is a unique $(n-k)$ -form μ_0^F on fibers such that $\mu_0^F \wedge \pi^* \nu_0 = \mu_0$. Similarly we find μ_1^F . Due to the equality of the pushforwards $\pi_* \mu_0$ and $\pi_* \mu_1$, the total volumes of μ_0^F and μ_1^F are fiberwise equal. Hence by the Moser theorem applied to the fibers, there is a smooth vector field tangent to the fibers, smoothly depending on a base point, and whose flow sends one of the $(n-k)$ -forms, μ_1^F , to the other, μ_0^F . This field is defined globally on N , and hence its time-one-map pulls back μ_1 to μ_0 . \square

Now we turn to a nonholonomic distribution on a manifold.

Proof of Theorem 2.1. The nonholonomic version of the Moser theorem:

(1) As before, we connect the forms by a segment $\mu_t, t \in [0, 1]$, and we come to the same homological equation. The latter reduces to $\text{div}_\mu V = \rho$ with $\int \rho \mu = 0$, but the equation now is for a vector field V tangent to the distribution τ .

(2) Consider some Riemannian metric on M . Now we will be looking for the required field V in the form $V := P^\tau \nabla u$, where P^τ is a pointwise orthogonal projection of tangent vectors to the planes of our distribution τ .

We obtain the equation $\text{div}_\mu(P^\tau \nabla u) = \rho$. Rewrite this equation by introducing the sub-Laplacian $\Delta^\tau u := \text{div}_\mu(P^\tau \nabla u)$ associated to the distribution

τ and the Riemannian metric on M . The equation on the potential u becomes $\Delta^\tau u = \rho$.

(3) An analog of Lemma 2.6 is now as follows.

Proposition 2.7. (a) *The sub-Laplacian operator $\Delta^\tau u := \operatorname{div}_\mu(P^\tau \nabla u)$ is a self-adjoint hypoelliptic operator. Its image is closed in L^2 .*

(b) *The equation $\Delta^\tau u = \rho$ on a compact Riemannian manifold M is solvable for any smooth function ρ with zero mean: $\int_M \rho \mu = 0$.*

Proof. (a) The principal symbol δ^τ of the operator Δ^τ is the sum of squares of vector fields forming a basis for the distribution τ : $\delta^\tau = \sum X_i^2$, where X_i form a horizontal orthonormal frame for τ . This is exactly the Hörmander condition of hypoellipticity [10, 11] for the operator Δ^τ . The self-adjointness follows from the properties of projection and integration by parts. The closedness of the image in L^2 follows from the results of [21, 22].

(b) We need to find the condition of weak solvability in L^2 for the equation $\Delta^\tau u = \rho$. Again, we are looking for all those functions h which are L^2 -orthogonal to the image of Δ^τ (or, which is the same, belonging to the kernel of this operator):

$$0 = \langle h, \Delta^\tau u \rangle_{L^2} = \langle h, \operatorname{div}_\mu(P^\tau \nabla u) \rangle_{L^2}$$

for all smooth functions u on M . In particular, this should hold for $u = h$. Integrating by parts we come to

$$0 = \langle h, \operatorname{div}_\mu(P^\tau \nabla h) \rangle_{L^2} = -\langle \nabla h, P^\tau \nabla h \rangle_{L^2} = -\langle P^\tau \nabla h, P^\tau \nabla h \rangle_{L^2},$$

where in the last equality we used the projection property $(P^\tau)^2 = P^\tau = (P^\tau)^*$. Then $P^\tau \nabla h = 0$ on M , and hence the equation $\Delta^\tau u = \rho$ is solvable for any function $\rho \perp_{L^2} \{h \mid P^\tau \nabla h = 0\}$. We claim that all such functions h are constant on M . (More precisely, by setting $u = h$ we implicitly assumed that h is smooth. For any $h \in L^2(M)$ consider a smooth approximation \bar{h} of h with any given precision in the L^2 norm and set $u = \bar{h}$. We obtain that $P^\tau \nabla \bar{h} = 0$ on M . We are going to show that any smooth approximation \bar{h} must be constant, which implies that all such functions h are constant as elements of $L^2(M)$.)

Indeed, the condition $P^\tau \nabla h = 0$ means that $\mathcal{L}_X h = 0$ for any horizontal field X , i.e., a field tangent to the distribution τ . But then h must be constant along any horizontal path, and due to the Chow–Rashevsky theorem it must be constant everywhere on M . Thus the functions ρ must be L^2 -orthogonal to all constants, and hence they have zero mean. This implies that the equation $\operatorname{div}_\mu(P^\tau \nabla u) = \rho$ is solvable for any L^2 function ρ with zero mean. For a smooth ρ the solution is also smooth due to hypoellipticity of the operator. \square

(4) Now consider the horizontal field $V_t := P^\tau \nabla u_t$. As before, the time-one-map of its flow exists for the smooth field V_t on the compact manifold M , and it gives the required diffeomorphism ϕ . \square

2.4. The nonholonomic Hodge decomposition and sub-Laplacian.

According to the classical Helmholtz–Hodge decomposition, any vector field W on a Riemannian manifold M can be uniquely decomposed into the sum $W = \tilde{V} + \tilde{U}$, where $\tilde{V} = \nabla f$ and $\operatorname{div}_\mu \tilde{U} = 0$. Proposition 2.7 suggests the following nonholonomic Hodge decomposition of vector fields on a manifold with a bracket-generating distribution:

Proposition 2.8. (1) *For a bracket-generating distribution τ on a Riemannian manifold M , any vector field W on M can be uniquely decomposed into the sum $W = V + U$, where the field $V = P^\tau \nabla f$ and it is tangent to the distribution τ , while the field U is divergence-free: $\operatorname{div}_\mu U = 0$. Here P^τ is the pointwise orthogonal projection to τ .*

(2) *Moreover, if the vector field \bar{W} is tangent to the distribution τ on M , then $\bar{W} = \bar{V} + \bar{U}$, where $\bar{V} = P^\tau \nabla f \parallel \tau$ as before, while the field \bar{U} is divergence-free, tangent to τ , and L^2 -orthogonal to \bar{V} , see Figure 1.*

Proof. Let $\rho := \operatorname{div}_\mu W$ be the divergence of W with respect to the Riemannian volume μ . First, note that $\int_M \rho \mu = 0$. Indeed, $\int_M (\operatorname{div}_\mu W) \mu = \int_M \mathcal{L}_W \mu = 0$, since the volume of μ is defined in a coordinate-free way, and does not change along the flow of the field W .

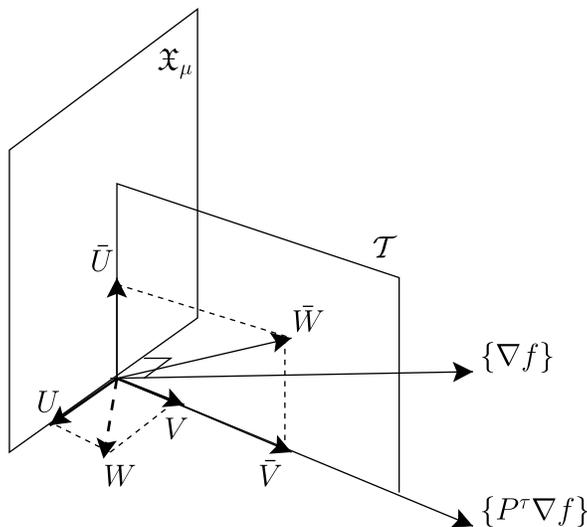


Figure 1. A nonholonomic Hodge decomposition.

Now, apply Proposition 2.7 to find a solution of the equation $\operatorname{div}(P^\tau \nabla f) = \rho$. The field $V := P^\tau \nabla f$ is defined uniquely. Then the field $U := W - V$ is divergence free, which proves (1).

For a field $\bar{W} \parallel \tau$, we define $\bar{V} := P^\tau \nabla f$ in the same way. Note that $\bar{V} \parallel \tau$ as well. Then $\bar{U} := \bar{W} - \bar{V}$ is both tangent to τ and divergence free. Furthermore,

$$\langle \bar{U}, \bar{V} \rangle_{L^2} = \langle \bar{U}, P^\tau \nabla f \rangle_{L^2} = \langle P^\tau \bar{U}, \nabla f \rangle_{L^2} = \langle \bar{U}, \nabla f \rangle_{L^2} = \langle \operatorname{div}_\mu \bar{U}, f \rangle_{L^2} = 0,$$

where we used the properties of \bar{U} established above: $\bar{U} \parallel \tau$ and $\operatorname{div}_\mu \bar{U} = 0$. \square

Above we defined a sub-Laplacian $\Delta^\tau u := \operatorname{div}_\mu(P^\tau \nabla u)$ for a function u on a Riemannian manifold M with a distribution τ .

Proposition 2.9 (cf. [15]). *The sub-Laplacian Δ^τ depends only on a subriemannian metric on the distribution τ and a volume form in the ambient manifold M .*

Proof. Note that the operator $P^\tau \nabla$ on a function u is the horizontal gradient ∇^τ of u , i.e., the vector of the fastest growth of u among the directions in τ . If one chooses a local orthonormal frame X_1, \dots, X_k in τ , then $P^\tau \nabla u = \sum_{i=1}^k (\mathcal{L}_{X_i} u) X_i$. Thus the definition of the horizontal gradient relies on the subriemannian metric only.

The sub-Laplacian $\Delta^\tau \psi = \operatorname{div}_\mu(P^\tau \nabla \psi)$ needs also the volume form μ in the ambient manifold to take the divergence with respect to this form. \square

The corresponding *nonholonomic heat equation* $\partial_t u = \Delta^\tau u$ is also defined by the subriemannian metric and a volume form.

2.5. The case with boundary. For a manifold M with nonempty boundary ∂M and two volume forms μ_0, μ_1 of equal total volume, the classical Moser theorem establishes the existence of diffeomorphism ϕ which is the time-one-map for the flow of a field V_t tangent to ∂M and such that $\phi^* \mu_1 = \mu_0$.

The existence of the required gradient field $V_t = \nabla u$ is guaranteed by the following

Lemma 2.10. *Let μ be a volume form on a Riemannian manifold M with boundary ∂M . The Poisson equation $\Delta u = \rho$ with Neumann boundary condition $\frac{\partial}{\partial n} u = 0$ on the boundary ∂M is solvable for any smooth function ρ with zero mean: $\int_M \rho \mu = 0$.*

Here $\frac{\partial}{\partial n}$ is the differentiation in the direction of outer normal n on the boundary.

Proof of Lemma 2.10. Proceed in the same way as in Lemma 2.6 to find all functions h that are L^2 -orthogonal to the image $\operatorname{Im} \Delta$. The first integration

by parts gives

$$0 = \int_M h(\Delta u)\mu = - \int_M \langle \nabla h, \nabla u \rangle \mu + \int_{\partial M} h \left(\frac{\partial}{\partial n} u \right) \mu = - \int_M \langle \nabla h, \nabla u \rangle \mu,$$

where in the last equality we used the Neumann boundary conditions. The second integration by parts gives

$$0 = \int_M \langle \Delta h, u \rangle \mu - \int_{\partial M} \left(\frac{\partial}{\partial n} h \right) u \mu.$$

This equation holds for all smooth functions u on M , so any such function h must be harmonic in M and satisfy the Neumann boundary condition $\frac{\partial}{\partial n} h = 0$. Hence, these are constant functions on M : $(\text{Im } \Delta)^{\perp L^2} = \{\text{const}\}$ (see the treatment of the Neumann boundary problem in this weak formulation in [24], Chapter 5, Section 7). This gives the same description as in the no-boundary case: the image $(\text{Im } \Delta)$ with the Neumann condition consists of functions ρ with zero mean. \square

Geometrically, the Neumann boundary condition means that there is no flux of density through the boundary ∂M : $0 = \frac{\partial u}{\partial n} = n \cdot \nabla u = n \cdot V$ on ∂M .

For plane distributions on manifolds with boundary, the solution of the Neumann problem becomes a much more subtle issue, as the behavior of the distribution near the boundary affects the flux of horizontal fields across the boundary, and hence the solvability in this problem. However, there is a class of domains in length spaces for which the solvability of the Neumann problem was established.

Let LS be a length space with the distance function $d(x, y)$, defined as infimum of lengths of continuous curves joining $x, y \in \text{LS}$. Consider domains in this space with the property that sufficiently close points in those domains can be joined by a not very long path which does not get too close to the domain boundary. The formal definition is as follows.

Definition 2.11. An open set $\Omega \subseteq \text{LS}$ is called an (ϵ, δ) -domain if there exist $\delta > 0$ and $0 < \epsilon \leq 1$ such that for any pair of points $p, q \in \Omega$ with $d(p, q) \leq \delta$ there is a continuous rectifiable curve $\gamma : [0, T] \rightarrow \Omega$ starting at p and ending at q such that the length $l(\gamma)$ of the curve γ satisfies

$$l(\gamma) \leq \frac{1}{\epsilon} d(p, q)$$

and

$$\min\{d(p, z), d(q, z)\} \leq \frac{1}{\epsilon} d(z, \partial\Omega)$$

for all points z on the curve γ .

A large source of (ϵ, δ) -domains is given by some classes of open sets in Carnot groups, where the Carnot group itself is regarded as a length space with the Carnot–Caratheodory distance, defined via the lengths of

admissible (i.e., horizontal) paths, see, e.g., [17]. There is a natural notion of diameter (or, radius) for domains in length spaces.

Theorem 2.12. *Let τ be a bracket-generating distribution on a subriemannian manifold M with smooth boundary ∂M , and μ_0, μ_1 be two volume forms on M with the same total volume: $\int_M \mu_0 = \int_M \mu_1$. Suppose that the interior of M is an (ϵ, δ) -domain of positive diameter.*

Then there exists a diffeomorphism ϕ of M which is the time-one-map of the flow ϕ_t of a nonautonomous vector field V_t tangent to the distribution τ everywhere on M and to the boundary ∂M for every $t \in [0, 1]$, such that $\phi^ \mu_1 = \mu_0$.*

The proof immediately follows from the result on solvability of the corresponding Neumann problem $\Delta^\tau u = \rho$ with $n \cdot (P^\tau \nabla u)|_{\partial M} = 0$ (or, which is the same, $\frac{\partial u}{\partial(P^\tau n)}|_{\partial M} = 0$) for such domains, established in [17, 18]. Indeed, the same (weak) reduction of the infinitesimal mass transport to the corresponding Neumann problem as in Lemma 2.10 gives

$$0 = \int_M h(\Delta^\tau u) \mu = \int_M \langle \Delta^\tau h, u \rangle \mu - \int_{\partial M} \frac{\partial h}{\partial(P^\tau n)} u \mu.$$

By taking first test functions u vanishing on ∂M and then any test functions satisfying Neumann boundary conditions we obtain $\Delta^\tau h = 0$ and $\frac{\partial h}{\partial(P^\tau n)}|_{\partial M} = 0$. The corresponding solvability and uniqueness result (cf. Theorem 1.5 in [17]) implies that $h = \text{const}$, which in turn gives us the description of the image of the Neumann operator as above.

We note that this solvability in the Neumann problem was shown for $u \in L^{1,2}$ and hence the above theorem is also valid for $V = P^\tau \nabla u$ in the corresponding Sobolev class. Apparently this solvability holds in higher Sobolev classes and in the smooth category, but the proof does not seem to be available in the literature.²

3. Distributions on diffeomorphism groups

3.1. A fibration on the group of diffeomorphisms. Let \mathcal{D} be the group of all (orientation-preserving) diffeomorphisms of a manifold M . Its Lie algebra \mathfrak{X} consists of all smooth vector fields on M . The tangent space to the diffeomorphism group at any point $\phi \in \mathcal{D}$ is given by the right translation of the Lie algebra \mathfrak{X} from the identity $id \in \mathcal{D}$ to ϕ :

$$T_\phi \mathcal{D} = \{X \circ \phi \mid X \in \mathfrak{X}\}.$$

Fix a volume form μ of total volume 1 on the manifold M . Denote by \mathcal{D}_μ the subgroup of volume-preserving diffeomorphisms, i.e., the diffeomorphisms preserving the volume form μ . The corresponding Lie algebra \mathfrak{X}_μ is

²We thank Duy-Minh Nhieu for clarification on this point.

the space of all vector fields on the manifold M which are divergence free with respect to the volume form μ .

Let \mathcal{W} be the set of all smooth normalized volume forms in M , which is called the (*smooth*) *Wasserstein space*. Consider the projection map $\pi^{\mathcal{D}} : \mathcal{D} \rightarrow \mathcal{W}$ defined by the pushforward of the fixed volume form μ by the diffeomorphism ϕ , i.e., $\pi^{\mathcal{D}}(\phi) = \phi_*\mu$. The projection $\pi^{\mathcal{D}} : \mathcal{D} \rightarrow \mathcal{W}$ defines a natural structure of a principal bundle on \mathcal{D} whose structure group is the subgroup \mathcal{D}_μ of volume-preserving diffeomorphisms of M and fibers F are right cosets for this subgroup in \mathcal{D} . Two diffeomorphisms ϕ and $\tilde{\phi}$ lie in the same fiber if they differ by a composition (on the right) with a volume-preserving diffeomorphism: $\tilde{\phi} = \phi \circ s$, $s \in \mathcal{D}_\mu$.

On the group \mathcal{D} we define two vector bundles Ver and Hor whose spaces at a diffeomorphism $\phi \in \mathcal{D}$ consist of right translated divergence-free fields

$$\text{Ver}_\phi = \{X \circ \phi \mid \text{div}_{\phi_*\mu} X = 0\}$$

and gradient fields

$$\text{Hor}_\phi := \{\nabla f \circ \phi \mid f \in C^\infty(M)\},$$

respectively. Note that the bundle Ver is defined by the fixed volume form μ , while Hor requires a Riemannian metric.³

Proposition 3.1. *The bundle Ver of translated divergence-free fields is the bundle of vertical spaces $T_\phi F$ for the fibration $\pi^{\mathcal{D}} : \mathcal{D} \rightarrow \mathcal{W}$. The bundle Hor over \mathcal{D} defines a horizontal distribution for this fibration $\pi^{\mathcal{D}}$.*

Proof. Let ϕ_t be a curve in a fiber of $\pi^{\mathcal{D}} : \mathcal{D} \rightarrow \mathcal{W}$ emanating from the point $\phi_0 = \phi$. Then $\phi_t = \phi_0 \circ s_t$, where $s_0 = \text{id}$ and s_t are volume-preserving diffeomorphisms for each t . Let X_t be a family of divergence-free vector fields, such that $\partial_t s_t = X_t \circ s_t$. Then the vector tangent to the curve $\phi_t = \phi_0 \circ s_t$ is given by $\left. \frac{d}{dt} \right|_{t=0} (\phi_0 \circ s_t) = (\phi_{0*} X_0) \circ \phi_0$. Since X_0 is divergence free with respect to μ , $\phi_{0*} X_0$ is divergence free with respect to $\phi_{0*}\mu$. Hence, any vector tangent to the diffeomorphism group at ϕ is given by $X \circ \phi$, where X is a divergence-free field with respect to the form $\phi_*\mu$.

By the Hodge decomposition of vector fields, we have the direct sum $T\mathcal{D} = \text{Hor} \oplus \text{Ver}$. \square

Remark 3.2. The classical Moser theorem 2.2 can be thought of as the existence of path-lifting property for the principal bundle $\pi^{\mathcal{D}} : \mathcal{D} \rightarrow \mathcal{W}$: any deformation of volume forms can be traced by the corresponding flow, i.e., a path on the diffeomorphism group, projected to the deformation of forms. Its proof shows that this path-lifting property holds and has the uniqueness

³The metric on M does not need to have the volume form μ . In the general case, \mathfrak{X}_μ consists of vector fields divergence free with respect to μ , while the gradients are considered for the chosen metric on M .

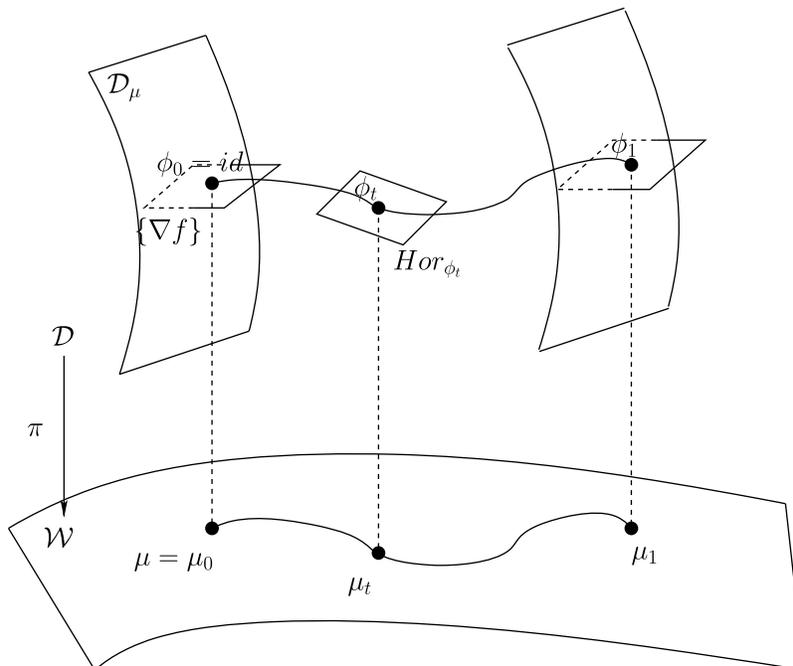


Figure 2. The Moser theorem in both the classical and non-holonomic settings is a path-lifting property in the diffeomorphism group.

property in the presence of the horizontal distribution defined above by using the Hodge decomposition. Namely, given any path μ_t starting at μ_0 in the smooth Wasserstein space \mathcal{W} and a point ϕ_0 in the fiber $(\pi^{\mathcal{D}})^{-1}\mu_0$, there exists a unique path ϕ_t in the diffeomorphism group which is tangent to the horizontal bundle Hor , starts at ϕ_0 , and projects to μ_t , see Figure 2.

3.2. A nonholonomic distribution on the diffeomorphism group.

Let τ be a bracket-generating distribution on the manifold M . Consider the right-invariant distribution \mathcal{T} on the diffeomorphism group \mathcal{D} defined at the identity $id \in \mathcal{D}$ of the group by the subspace in \mathfrak{X} of all those vector fields which are tangent to the distribution τ everywhere on M :

$$\mathcal{T}_\phi = \{V \circ \phi \mid V(x) \in \tau_x \text{ for all } x \in M\}.$$

Proposition 3.3. *The infinite-dimensional distribution \mathcal{T} is a non-integrable distribution in \mathcal{D} . Horizontal paths in this distribution are flows of nonautonomous vector fields tangent to the distribution τ on manifold M .*

Proof. To see that the distribution \mathcal{T} is nonintegrable we consider two horizontal vector fields V and W on M and the corresponding right-invariant

vector fields \widetilde{V} and \widetilde{W} on \mathcal{D} . Then their bracket at the identity of the group is (minus) their commutator as vector fields V and W in M . This commutator does not belong to the plane \mathcal{T}_{id} , since the distribution τ is nonintegrable and hence at least somewhere on M the commutator of horizontal fields V and W is not horizontal.

The second statement immediately follows from the definition of \mathcal{T} . \square

Remark 3.4. Consider now the projection map $\pi^{\mathcal{D}} : \mathcal{D} \rightarrow \mathcal{W}$ in the presence of the distribution \mathcal{T} on \mathcal{D} . The path-lifting property in this case is a restatement of the nonholonomic Moser theorem. Namely, for a curve $\{\mu_t \mid \mu_0 = \mu\}$ in the space \mathcal{W} of smooth densities Theorem 2.1 proves that there is a curve $\{g^t \mid g^0 = \text{id}\}$ in \mathcal{D} , everywhere tangent to the distribution \mathcal{T} and projecting to $\{\mu_t\}$: $\pi^{\mathcal{D}}(g^t) = \mu_t$.

Recall that in the classical case the corresponding path lifting becomes unique once we fix the gradient horizontal bundle $\text{Hor}_{\phi} \subset T_{\phi}\mathcal{D}$ for any diffeomorphism $\phi \in \mathcal{D}$. Similarly, in the nonholonomic case we consider the spaces of gradient projections instead of the gradient spaces: $\text{Hor}_{\text{id}}^{\tau} := \{P^{\tau}\nabla f \mid f \in C^{\infty}(M)\}$, where P^{τ} stands for the orthogonal projection onto the distribution τ in a given Riemannian metric on M . The right-translated gradient projections $\text{Hor}_{\phi}^{\tau} := \{(P^{\tau}\nabla f) \circ \phi \mid f \in C^{\infty}(M)\}$ define a horizontal bundle for the principal bundle $\mathcal{D} \rightarrow \mathcal{W}$ by the nonholonomic Hodge decomposition. (Note also that in both the classical and nonholonomic cases, the obtained horizontal distributions on \mathcal{D} are nonintegrable, cf. [20]. Indeed, the Lie bracket of two gradient fields is not necessarily a gradient field, and similarly for gradient projections. Hence there are no horizontal sections of the bundle $\mathcal{D} \rightarrow \mathcal{W}$, tangent to these horizontal gradient distributions.)

As we will see in Sections 4 and 6, both gradient fields $\{\nabla f\}$ in the classical case and gradient projections $\{P^{\tau}\nabla f\}$ in the nonholonomic case allow one to move the densities in the “fastest way,” and are important in transport problems of finding optimal (“shortest”) paths between densities.

3.3. Accessibility of diffeomorphisms and symplectic structures. Presumably, even a stronger statement holds:

Conjecture 3.5. *Every diffeomorphism in the diffeomorphism group \mathcal{D} can be accessed by a horizontal path tangent to the distribution \mathcal{T} .*

This conjecture can be thought of as an analog of the Chow–Rashevsky theorem in the infinite-dimensional setting of the group of diffeomorphisms, provided that the distribution \mathcal{T} is bracket generating on \mathcal{D} . Note, however, that the Chow–Rashevsky theorem is unknown in the general setting of an infinite-dimensional manifold, while there are only “approximate” analogs of it, e.g., on a Hilbert manifold.

A proof of this conjecture on accessibility of all diffeomorphisms by flows of vector fields tangent to a nonholonomic distribution would imply the

nonholonomic Moser theorem 2.1 on volume forms. Moreover, it would also imply the following nonholonomic version of the Moser theorem on symplectic structures from [16].

Conjecture 3.6. *Suppose that on a manifold M two symplectic structures ω_0 and ω_1 from the same cohomology class can be connected by a path of symplectic structures in the same class. Then for a bracket-generating distribution τ on M there exists a diffeomorphism ϕ of M which is the time-one-map of a nonautonomous vector field V_t tangent to the distribution τ everywhere on M and for every $t \in [0, 1]$, such that $\phi^*\omega_1 = \omega_0$.*

This conjecture follows from the one above since one would consider the diffeomorphism from the classical Moser theorem, and realize it by the horizontal path (tangent to the distribution \mathcal{T}) on the diffeomorphism group, which exists if Conjecture 3.5 holds.

4. The Riemannian geometry of diffeomorphism groups and mass transport

The differential geometry of diffeomorphism groups is closely related to the theory of optimal mass transport, and in particular, to the problem of moving one density to another while minimizing certain cost on a Riemannian manifold. In this section, we review the corresponding metric properties of the diffeomorphism group and the space of volume forms.

4.1. Optimal mass transport. Let M be a compact Riemannian manifold without boundary (or, more generally, a complete metric space) with a distance function d . Let μ and ν be two Borel probability measures on the manifold M which are absolutely continuous with respect to the Lebesgue measure. Consider the following optimal mass transport problem: find a Borel map $\phi : M \rightarrow M$ that pushes the measure μ forward to ν and attains the infimum of the L^2 -cost functional $\int_M d^2(x, \phi(x))\mu$ among all such maps.

The set of all Borel probability measures is called the Wasserstein space. The minimal cost of transport defines a distance \tilde{d} on this space:

$$(4.1) \quad \tilde{d}^2(\mu, \nu) := \inf_{\phi} \left\{ \int_M d^2(x, \phi(x))\mu \mid \phi_*\mu = \nu \right\}.$$

This mass transport problem admits a unique solution ϕ (defined up to measure zero sets), called an optimal map (see [6] for $M = \mathbb{R}^n$ and [14] for any compact connected Riemannian manifold M without boundary). Furthermore, there exists a one-parameter family of Borel maps ϕ_t starting at the identity map $\phi_0 = \text{id}$, ending at the optimal map $\phi_1 = \phi$ and such that ϕ_t is the optimal map pushing μ forward to $\nu_t := \phi_{t*}\mu$ for any $t \in (0, 1)$. The corresponding one-parameter family of measures ν_t describes a geodesic in the Wasserstein space of measures with respect to the distance function

\tilde{d} and is called the displacement interpolation between μ and ν , see [25] for details. (More generally, in mass transport problems one can replace d^2 in the above formula by a cost function $c : M \times M \rightarrow \mathbb{R}$, while we mostly focus on the case $c = d^2/2$ and its subriemannian analog.)

In what follows, we consider a smooth version of the Wasserstein space, cf. Section 3.1. Recall that the smooth Wasserstein space \mathcal{W} consists of smooth volume forms with the total integral equal to 1. One can consider an infinite-dimensional manifold structure on the smooth Wasserstein space, a (weak) Riemannian metric $\langle \cdot, \cdot \rangle^{\mathcal{W}}$, corresponding to the distance function \tilde{d} , and geodesics on this space. Similar to the finite-dimensional case, geodesics on the smooth Wasserstein space \mathcal{W} can be formally defined as projections of trajectories of the Hamiltonian vector field with the “kinetic energy” Hamiltonian in the tangent bundle $T\mathcal{W}$.

4.2. The Otto calculus. For a Riemannian manifold M both spaces \mathcal{D} and \mathcal{W} can be equipped with (weak) Riemannian structures, i.e., can be formally regarded as infinite-dimensional Riemannian manifolds, cf. [7, 13]. (One can consider H^s -diffeomorphisms and H^{s-1} -forms of Sobolev class $s > n/2 + 1$. Both sets can be considered as smooth Hilbert manifolds. However, this is not applicable in the subriemannian case, discussed later, hence we confine to the C^∞ setting applicable in the both cases.)

From now on we fix a Riemannian metric $\langle \cdot, \cdot \rangle^M$ on the manifold M , whose Riemannian volume is the form μ . On the diffeomorphism group we define a Riemannian metric $\langle \cdot, \cdot \rangle^{\mathcal{D}}$ whose value at a point $\phi \in \mathcal{D}$ is given by

$$(4.2) \quad \langle X_1 \circ \phi, X_2 \circ \phi \rangle^{\mathcal{D}} := \int_M \langle X_1 \circ \phi(x), X_2 \circ \phi(x) \rangle_{\phi(x)}^M \mu.$$

The action along a curve (or, “energy” of a curve) $\{\phi_t \mid t \in [0, 1]\} \subset \mathcal{D}$ in this metric is defined in the following straightforward way:

$$E(\{\phi_t\}) = \int_0^1 dt \int_M \langle \partial_t \phi_t, \partial_t \phi_t \rangle^M \mu.$$

If M is flat, \mathcal{D} is locally isometric to the (pre-)Hilbert L^2 -space of (smooth) vector-functions ϕ , see, e.g., [3, 23]. The following proposition is well known.

Proposition 4.1. *Let ϕ_t be a geodesic on the diffeomorphism group \mathcal{D} with respect to the above Riemannian metric $\langle \cdot, \cdot \rangle^{\mathcal{D}}$, and V_t be the (time-dependent) velocity field of the corresponding flow: $\partial_t \phi_t = V_t \circ \phi_t$. Then the velocity V_t satisfies the inviscid Burgers equation on M :*

$$\partial_t V_t + \nabla_{V_t} V_t = 0,$$

where $\nabla_{V_t} V_t$ stands for the covariant derivative of the field V_t on M along itself.

Proof. In the flat case the geodesic equation is $\partial_t^2 \phi_t = 0$: this is the Euler–Lagrange equation for the action functional E . Differentiate $\partial_t \phi_t = V_t \circ \phi_t$ with respect to time t and use this geodesic equation to obtain

$$(4.3) \quad \partial_t V_t \circ \phi_t + \nabla_{V_t} \partial_t \phi_t = 0.$$

After another substitution $\partial_t \phi_t = V_t \circ \phi_t$, the later becomes

$$(\partial_t V_t + \nabla_{V_t} V_t) \circ \phi_t = 0,$$

which is equivalent to the Burgers equation.

The non-flat case involves differentiation in the Levi–Civita connection on M and leads to the same Burgers equation, see details in [7, 12]. \square

Remark 4.2. Smooth solutions of the Burgers equation correspond to non-interacting particles on the manifold M flying along those geodesics on M which are defined by the initial velocities $V_0(x)$. The Burgers flows have the form $\phi_t(x) = \exp^M(tV_0(x))$, where $\exp^M : TM \rightarrow M$ is the Riemannian exponential map on M .

Proposition 4.3 [20]. *The bundle projection $\pi^{\mathcal{D}} : \mathcal{D} \rightarrow \mathcal{W}$ is a Riemannian submersion of the metric $\langle, \rangle^{\mathcal{D}}$ on the diffeomorphism group \mathcal{D} to the Riemannian metric $\langle, \rangle^{\mathcal{W}}$ on the smooth Wasserstein space \mathcal{W} for the L^2 -cost. The horizontal (i.e., normal to fibers) spaces in the bundle $\mathcal{D} \rightarrow \mathcal{W}$ are right-translated gradient fields.*

Recall that for two Riemannian manifolds Q and B , a Riemannian submersion $\pi : Q \rightarrow B$ is a mapping onto B which has maximal rank and preserves lengths of horizontal tangent vectors to Q , see, e.g., [19]. For a bundle $Q \rightarrow B$, this means that there is a distribution of horizontal spaces on Q , orthogonal to the fibers, which is projected isometrically to the tangent spaces to B . One of the main properties of a Riemannian submersion gives the following feature of geodesics:

Corollary 4.4. *Any geodesic, initially tangent to a horizontal space on the full diffeomorphism group \mathcal{D} , always remains horizontal, i.e., tangent to the horizontal distribution. There is a one-to-one correspondence between geodesics on the base \mathcal{W} starting at the measure μ and horizontal geodesics in \mathcal{D} starting at the identity diffeomorphism id .*

Remark 4.5. In the PDE terms, the horizontality of a geodesic means that a solution of the Burgers equation with a potential initial condition remains potential forever. This also follows from the Hamiltonian formalism and the moment map geometry discussed in the next section. Since horizontal geodesics in the group \mathcal{D} correspond to geodesics on the density space \mathcal{W} , potential solutions of the Burgers equation (corresponding to horizontal geodesics) move the densities in the fastest way. The corresponding time-one-maps for Burgers potential solutions provide optimal maps for moving the density μ to any other density ν , see [6, 14].

The Burgers potential solutions have the form $\phi_t(x) = \exp^M(-t\nabla f(x))$ as long as the right-hand side is smooth. The time-one-map ϕ_1 for the flow ϕ_t provides an optimal map between probability measures if the function f is a $(d^2/2)$ -concave function. The notion of c -concavity for a cost function c on M is defined as follows. For a function f its c -transform is $f^c(y) = \inf_{x \in M}(c(x, y) - f(x))$ and the function f is said to be c -concave if $f^{cc} = f$. Here, we consider the case $c = d^2/2$. The family of maps ϕ_t defines the displacement interpolation mentioned in Section 4.1.

Let θ and ν be volume forms with the same total volume and let g and h be functions on the manifold M defined by $\theta = g \text{ vol}$ and $\nu = h \text{ vol}$, where vol be the Riemannian volume form. Then a diffeomorphism ϕ moving one density to the other ($\phi_*\theta = \nu$) satisfies $h(\phi(x)) \det(D\phi(x)) = g(x)$, where $D\phi$ is the Jacobi matrix of the diffeomorphism ϕ . In the flat case the optimal map ϕ is gradient, $\phi = \nabla \tilde{f}$, and the corresponding convex potential \tilde{f} satisfies the Monge–Ampère equation

$$\det(\text{Hess } \tilde{f}(x)) = \frac{g(x)}{h(\nabla \tilde{f}(x))},$$

since $D(\nabla \tilde{f}) = \text{Hess } \tilde{f}$. In the nonflat case, the optimal map is $\phi(x) = \exp^M(-\nabla f(x))$ for a $(d^2/2)$ -concave potential f , and the equation is Monge–Ampère like, see [14, 25] for details. Below we describe the corresponding nonholonomic analogs of these objects.

5. The Hamiltonian mechanics on diffeomorphism groups

In this section we present a Hamiltonian framework for the Otto calculus and, in particular, give a symplectic proof of Proposition 4.3 and Corollary 4.4 on the submersion properties along with their generalizations.

5.1. Averaged Hamiltonians. We fix a Riemannian metric $\langle \cdot, \cdot \rangle^M$ on the manifold M and consider the corresponding Riemannian metric $\langle \cdot, \cdot \rangle^{\mathcal{D}}$ on the diffeomorphism group \mathcal{D} . This defines a map $(X \circ \phi) \mapsto \langle X \circ \phi, \cdot \rangle^{\mathcal{D}}$ from the tangent bundle $T\mathcal{D}$ to the cotangent bundle $T^*\mathcal{D}$. By using this map, one can pull back the canonical symplectic form $\omega^{T^*\mathcal{D}}$ from the cotangent bundle $T^*\mathcal{D}$ to the tangent bundle $T\mathcal{D}$, and regard the latter as a manifold equipped with the symplectic form $\omega^{T\mathcal{D}}$.⁴ Similarly, a symplectic structure ω^{TM} can be defined on the tangent bundle TM by pulling back the canonical symplectic form on the cotangent bundle T^*M via the Riemannian metric $\langle \cdot, \cdot \rangle^M$. The two symplectic forms are related as follows. A tangent vector V in the tangent space $T_{X \circ \phi}T\mathcal{D}$ at the point $X \circ \phi \in T\mathcal{D}$ is a map from M to

⁴The consideration of the tangent bundle $T\mathcal{D}$ (instead of $T^*\mathcal{D}$) as a symplectic manifold allows one to avoid dealing with duals of infinite-dimensional spaces here.

$T(TM) = T^2M$ such that $\pi^{T^2M} \circ V = X \circ \phi$, where $\pi^{T^2M} : T(TM) \rightarrow TM$ is the tangent bundle projection. Let V_1 and V_2 be two tangent vectors in $T_{X \circ \phi}T\mathcal{D}$ at the point $X \circ \phi$, then the symplectic forms are related in the following way:

$$\omega^{T\mathcal{D}}(V_1, V_2) = \int_M \omega^{TM}(V_1(x), V_2(x))\mu(x),$$

where ω^{TM} is understood as the pairing on $T(TM) = T^2M$.

Definition 5.1. Let H^M be a Hamiltonian function on the tangent bundle TM of the manifold M . The *averaged Hamiltonian function* is the function $H^{\mathcal{D}}$ on the tangent bundle $T\mathcal{D}$ of the diffeomorphism group \mathcal{D} obtained by averaging the corresponding Hamiltonian H^M over M in the following way: its value at a point $X \circ \phi \in T_\phi\mathcal{D}$ is

$$(5.1) \quad H^{\mathcal{D}}(X \circ \phi) := \int_M H^M(X \circ \phi(x))\mu(x)$$

for a vector field $X \in \mathfrak{X}$ and a diffeomorphism $\phi \in \mathcal{D}$.

Consider the Hamiltonian flows for these Hamiltonian functions H^M and $H^{\mathcal{D}}$ on the tangent bundles TM and $T\mathcal{D}$, respectively, with respect to the standard symplectic structures on the bundles. The following theorem can be viewed as a generalization of Propositions 4.1 and 4.3.

Theorem 5.2. *Each Hamiltonian trajectory for the averaged Hamiltonian function $H^{\mathcal{D}}$ on $T\mathcal{D}$ describes a flow on the tangent bundle TM , in which every tangent vector to M moves along its own H^M -Hamiltonian trajectory in TM .*

Example 5.3. For the Hamiltonian $K^M(p, q) = \frac{1}{2}\langle p, p \rangle^M$ given by the “kinetic energy” for the metric on M , the above theorem implies that any geodesic on \mathcal{D} is a family of diffeomorphisms of M , in which each particle moves along its own geodesic on M with constant velocity, i.e., its velocity field is a solution to the Burgers equation, cf. Remark 4.2.

Below we discuss this theorem and its geometric meaning in detail. In particular, in the above form, the statement is also applicable to the case of nonholonomic distributions (i.e., subriemannian, or Carnot–Caratheodory spaces) discussed in the next section.

5.2. Riemannian submersion and symplectic quotients. We start with a Hamiltonian proof of Proposition 4.3 on the Riemannian submersion $\mathcal{D} \rightarrow \mathcal{W}$ of diffeomorphisms onto densities. Recall the following general construction in symplectic geometry. Let $\pi : Q \rightarrow B$ be a principal bundle with the structure group G .

Lemma 5.4 (see, e.g., [2]). *The symplectic reduction of the cotangent bundle T^*Q over the G -action gives the cotangent bundle $T^*B = T^*Q//G$.*

Proof. The moment map $J : T^*Q \rightarrow \mathfrak{g}^*$ associated with this action takes T^*Q to the dual of the Lie algebra $\mathfrak{g} = \text{Lie}(G)$. For the G -action on T^*Q the moment map J is the projection of any cotangent space T_a^*Q to cotangent space $T_a^*F \approx \mathfrak{g}^*$ for the fiber F through a point $a \in Q$. The preimage $J^{-1}(0)$ of the zero value is the subbundle of T^*Q consisting of covectors vanishing on fibers. Such covectors are naturally identified with covectors on the base B . Thus factoring out the G -action, which moves the point a over the fiber F , we obtain the bundle T^*B . \square

Suppose also that Q is equipped with a G -invariant Riemannian metric $\langle \cdot, \cdot \rangle^Q$.

Lemma 5.5. *The Riemannian submersion of $(Q, \langle \cdot, \cdot \rangle^Q)$ to the base B with the induced metric $\langle \cdot, \cdot \rangle^B$ is the result of the symplectic reduction.*

Proof. Indeed, the metric $\langle \cdot, \cdot \rangle^Q$ gives a natural identification $T^*Q \approx TQ$ of the tangent and cotangent bundles for Q , and the “projected metric” is equivalent to a similar identification for the base manifold B .

In the presence of metric in Q , the preimage $J^{-1}(0)$ is identified with all vectors in TQ orthogonal to fibers, that is $J^{-1}(0)$ is the horizontal subbundle in TQ . Hence, the symplectic quotient $J^{-1}(0)/G$ can be identified with the tangent bundle TB . \square

Proof of Proposition 4.3. Now we apply this “dictionary” to the diffeomorphism group \mathcal{D} and the Wasserstein space \mathcal{W} . Consider the projection map $\pi^{\mathcal{D}} : \mathcal{D} \rightarrow \mathcal{W}$ as a principal bundle with the structure group \mathcal{D}_μ of volume-preserving diffeomorphisms of M . Recall that the vertical space of this principal bundle at a point $\phi \in \mathcal{D}$ consists of right translations by the diffeomorphism ϕ of vector fields which are divergence free with respect to the volume form $\phi_*\mu$: $\text{Ver}_\phi = \{X \circ \phi \mid \text{div}_{\phi_*\mu} X = 0\}$, and the horizontal space is given by translated gradient fields: $\text{Hor}_\phi = \{\nabla f \circ \phi \mid f \in C^\infty(M)\}$.

For each volume-preserving diffeomorphism $\psi \in \mathcal{D}_\mu$, the \mathcal{D}_μ -action R_ψ of ψ by right translations on the diffeomorphism group is given by

$$R_\psi(\phi) = \phi \circ \psi.$$

The induced action $TR_\psi : T\mathcal{D} \rightarrow T\mathcal{D}$ on the tangent spaces of the diffeomorphism group is given by

$$TR_\psi(X \circ \phi) = (X \circ \phi) \circ \psi.$$

One can see that for volume-preserving diffeomorphisms ψ this action preserves the Riemannian metric (4.2) on the diffeomorphism group \mathcal{D} (it is the

change of variable formula), while for a general diffeomorphism one has an extra factor $\det(D\psi)$, the Jacobian of ψ , in the integral. \square

Remark 5.6. The explicit formula of the moment map $J : TQ \rightarrow \mathfrak{X}_\mu^*$ for the group of volume-preserving diffeomorphisms $G = \mathcal{D}_\mu$ acting on $Q = \mathcal{D}$ is

$$J(X \circ \phi)(Y) = \int_M \langle X, \phi_* Y \rangle^M \phi_* \mu,$$

where $Y \in \mathfrak{X}_\mu$ is any vector field on M divergence-free with respect to the volume form μ , $X \in \mathfrak{X}$, and $\phi \in \mathcal{D}$.

5.3. Hamiltonian flows on the diffeomorphism groups. Let $H^Q : TQ \rightarrow \mathbb{R}$ be a Hamiltonian function invariant under the G -action on the cotangent bundle of the total space Q . The restriction of the function H^Q to the horizontal bundle $J^{-1}(0) \subset TQ$ is also G -invariant, and hence descends to a function $H^B : TB \rightarrow \mathbb{R}$ on the symplectic quotient, the tangent bundle of the base B . Symplectic quotients admit the following reduction of Hamiltonian dynamics:

Proposition 5.7 [2]. *The Hamiltonian flow of the function H^Q preserves the preimage $J^{-1}(0)$, i.e., trajectories with horizontal initial conditions stay horizontal. Furthermore, the Hamiltonian flow of the function H^Q on the tangent bundle TQ of the total space Q descends to the Hamiltonian flow of the function H^B on the tangent bundle TB of the base.*

Now we are going to apply this scheme to the bundle $\mathcal{D} \rightarrow \mathcal{W}$. (Similarly to Section 4.2 we consider this setting either formally or in the corresponding Sobolev spaces, cf. [7].) For a fixed Hamiltonian function H^M on the tangent bundle TM to the manifold M , consider the corresponding averaged Hamiltonian function $H^{\mathcal{D}}$ on $T\mathcal{D}$, given by formula (5.1): $H^{\mathcal{D}}(X \circ \phi) := \int_M H^M(X \circ \phi(x))\mu$. The latter Hamiltonian is \mathcal{D}_μ -invariant (as also follows from the change of variable formula) and it will play the role of the function H^Q . Thus the flow for the averaged Hamiltonian $H^{\mathcal{D}}$ descends to the flow of a certain Hamiltonian $H^{\mathcal{W}}$ on $T\mathcal{W}$.

Describe explicitly the corresponding flow on the tangent bundles of \mathcal{D} and \mathcal{W} . Let $\Psi_t^{H^M} : TM \rightarrow TM$ be the Hamiltonian flow of the Hamiltonian H^M on the tangent bundle of the manifold M and $\Psi_t^{H^{\mathcal{D}}} : T\mathcal{D} \rightarrow T\mathcal{D}$ denotes the flow for the Hamiltonian function $H^{\mathcal{D}}$ on the tangent bundle of the diffeomorphism group.

Theorem 5.8 (=5.2'). *The Hamiltonian flows of the Hamiltonians $H^{\mathcal{D}}$ and H^M are related by*

$$\Psi_t^{H^{\mathcal{D}}}(X \circ \phi)(x) = \Psi_t^{H^M}(X(\phi(x))),$$

where, on the right-hand side, the flow $\Psi_t^{H^M}$ on TM transports the shifted field $X(\phi(x))$, while, on the left-hand side, $X \circ \phi$ is regarded as a tangent vector to \mathcal{D} at the point ϕ .

Proof. Prove this infinitesimally (cf. [7]). Let $X_{H^{\mathcal{D}}}$ and X_{H^M} be the Hamiltonian vector fields corresponding to the Hamiltonians $H^{\mathcal{D}}$ and H^M , respectively. We claim that $X_{H^{\mathcal{D}}}(X \circ \phi) = X_{H^M} \circ X \circ \phi$. Indeed, by the definition of Hamiltonian fields, we have

$$\begin{aligned} \omega^{T\mathcal{D}}(X_{H^{\mathcal{D}}} \circ X \circ \phi, Y) &= \int_M \omega^{TM}(X_{H^M}(X(\phi(x))), Y(x)) \mu \\ &= \int_M dH_{X(\phi(x))}^M(Y(x)) \mu(x) \end{aligned}$$

for any $Y \in T_\phi \mathcal{D}$. By interchanging the integration and exterior differentiation, the latter expression becomes $dH_{X \circ \phi}^{\mathcal{D}}(Y)$ and the result follows. Note that the 2-form $\omega^{T\mathcal{D}}$ is weakly symplectic (see [7]), hence the corresponding Hamiltonian field on $T\mathcal{D}$ is defined uniquely. \square

Remark 5.9. This theorem has a simple geometric meaning for the “kinetic energy” Hamiltonian function $K^M(v) := \frac{1}{2}\langle v, v \rangle^M$ on the tangent bundle TM . One of the possible definitions of geodesics in M is that they are projections to M of trajectories of the Hamiltonian flow on TM , whose Hamiltonian function is the kinetic energy. In other words, the Riemannian exponential map \exp^M on the manifold M is the projection of the Hamiltonian flow $\Psi_t^{K^M}$ on TM . Similarly, the Riemannian exponential $\exp^{\mathcal{D}}$ of the diffeomorphism group \mathcal{D} is the projection of the Hamiltonian flow for the Hamiltonian $K^{\mathcal{D}}(X \circ \phi) := \frac{1}{2} \int_M \langle X \circ \phi, X \circ \phi \rangle^M \mu$ on $T\mathcal{D}$.

Recall that the geodesics on the diffeomorphism group (described by the Burgers equation, see Proposition 4.1) starting at the identity with the initial velocity $V \in T_{\text{id}} \mathcal{D}$ are the flows which move each particle x on the manifold M along the geodesic with the direction $V(x)$. Such a geodesic is well defined on the diffeomorphism group \mathcal{D} as long as the particles do not collide. The corresponding Hamiltonian flow on the tangent bundle $T\mathcal{D}$ of the diffeomorphism group describes how the corresponding velocities of these particles vary (cf. Example 5.3).

For a more general Hamiltonian H^M on the tangent bundle TM , each particle $x \in M$ with an initial velocity $V(x)$ will be moving along the corresponding characteristic, which is the projection to M of the corresponding trajectory $\Psi_t^{H^M}(V(x))$ in the tangent bundle TM .

Now we would like to describe more explicitly horizontal geodesics and characteristics on the diffeomorphism group \mathcal{D} . Recall that $\Psi_t^{H^{\mathcal{D}}}$ denotes the Hamiltonian flow of the averaged Hamiltonian $H^{\mathcal{D}}$ on the tangent bundle $T\mathcal{D}$ of the diffeomorphism group \mathcal{D} . If this Hamiltonian flow is gradient at

the initial moment, it always stays gradient, as implied by Corollary 4.4. Furthermore, the corresponding potential can be described as follows.

Corollary 5.10. *Let f be a function on the manifold M . Then the Hamiltonian flow for H^D with the initial condition $\nabla f \circ \phi \in T_\phi \mathcal{D}$ has the form $\nabla f_t \circ \phi_t$, where $\phi_t \in \mathcal{D}$ is a family of diffeomorphisms and f_t is the family of functions on M starting at $f_0 = f$ and satisfying the Hamilton–Jacobi equation*

$$(5.2) \quad \partial_t f_t + H^M(\nabla f_t(x)) = 0.$$

Proof. This follows from the method of characteristics, which gives the following way of finding f_t , the solution to the Hamilton–Jacobi equation (5.2). Consider the tangent vector $\nabla f(x)$ for each point $x \in M$. Denote by $\Psi_t^{H^M} : TM \rightarrow TM$ the Hamiltonian flow for the Hamiltonian $H^M : TM \rightarrow \mathbb{R}$ and consider its trajectory $t \mapsto \Psi_t^{H^M}(\nabla f(x))$ starting at the tangent vector $\nabla f(x)$. Then project this trajectory to M using the tangent bundle projection $\pi^{TM} : TM \rightarrow M$ to obtain a curve in M . It is given by the formula $t \mapsto \pi^{TM}(\Psi_t^{H^M}(\nabla f(x)))$. As x varies over the manifold M , this defines a flow $\phi_t := \pi^{TM} \circ \Psi_t^{H^M} \circ \nabla f$ on M . (Note that this procedure defines a flow for small time t , while for larger times the map ϕ_t may cease to be a diffeomorphism, i.e., shock waves can appear.) The corresponding time-dependent vector field is gradient and defines the family ∇f_t , the gradient of the solution to the Hamilton–Jacobi equation above, see Figure 3. \square

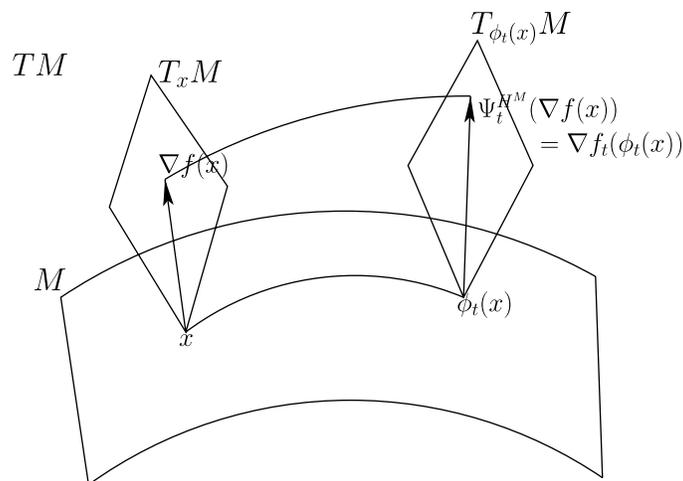


Figure 3. Hamiltonian flow of the Hamiltonian H^M and its projection: the curve $\phi_t(x)$ is the projection of the curve $\Psi_t^{H^M}(\nabla f(x))$ to the manifold M .

Remark 5.11. The above corollary manifests that the Hamilton–Jacobi equation (5.2) can be solved using the method of characteristics due to the built-in symmetry group of all volume-preserving diffeomorphisms.

5.4. Hamiltonian flows on the Wasserstein space. What is the corresponding flow on the tangent bundle $T\mathcal{W}$ of the Wasserstein space, induced by the Hamiltonian flow on $T\mathcal{D}$ for the diffeomorphism group \mathcal{D} after the projection $\pi^{\mathcal{D}} : \mathcal{D} \rightarrow \mathcal{W}$? Fix a Hamiltonian H^M on the tangent bundle TM which defines the averaged Hamiltonian function $H^{\mathcal{D}}$ on the tangent bundle $T\mathcal{D}$, see equation (5.1). Describe explicitly the induced Hamiltonian $H^{\mathcal{W}}$ on the tangent bundle $T\mathcal{W}$.

Let (ν, η) be a tangent vector at a density ν on M , regarded as a point of the Wasserstein space \mathcal{W} . The normalization of densities ($\int \nu = 1$ for all $\nu \in \mathcal{W}$) gives the constraint for tangent vectors: $\int_M \eta = 0$. Let $f : M \rightarrow \mathbb{R}$ be a function that satisfies $(-\operatorname{div}_{\nu} \nabla f) \nu = \eta$. (Given (ν, η) , such a function is defined uniquely up to an additive constant.) Then the induced Hamiltonian on the tangent bundle $T\mathcal{W}$ of the base \mathcal{W} is given by

$$(5.3) \quad H^{\mathcal{W}}(\nu, \eta) = \int_M H^M(\nabla f(x)) \nu,$$

since ∇f is a vector of the horizontal distribution in $T\mathcal{D}$.

Now, the flow $\Psi_t^{H^{\mathcal{W}}}$ of the corresponding Hamiltonian field on $T\mathcal{W}$ can be found explicitly by employing Proposition 5.7. Consider the flow $\phi_t := \pi^{TM} \circ \Psi_t^{H^M} \circ \nabla f$ defined on M for small t in Corollary 5.10.

Theorem 5.12. *The Hamiltonian flow $\Psi_t^{H^{\mathcal{W}}}$ of the Hamiltonian function $H^{\mathcal{W}}$ on the tangent bundle $T\mathcal{W}$ of the Wasserstein space \mathcal{W} is*

$$\Psi_t^{H^{\mathcal{W}}}(\nu, \eta) = (\nu_t, -\mathcal{L}_{\nabla f_t} \nu_t),$$

where \mathcal{L} is the Lie derivative, the family of functions f_t satisfies the Hamilton–Jacobi equation (5.2) for the Hamiltonian function H^M on the tangent bundle TM , and the family $\nu_t = (\phi_t)_* \nu$ is the pushforward of the volume form ν by the map ϕ_t defined above.

Proof. The function $H^{\mathcal{D}}(X \circ \phi) = \int_M H^M(X(\phi(x))) \mu(x)$ on the tangent bundle $T\mathcal{D}$ of the diffeomorphism group induces the Hamiltonian $H^{\mathcal{W}}$ on $T\mathcal{W}$. By virtue of the Hamiltonian reduction, Hamiltonian trajectories of $H^{\mathcal{D}}$ contained in the horizontal bundle $\operatorname{Hor} = \{\nabla f \circ \phi \mid f \in C^\infty(M)\}$ descend to Hamiltonian trajectories of $H^{\mathcal{W}}$. Then the Hamiltonian flow $\Psi^{H^{\mathcal{D}}}$ of the

Hamiltonian $H^{\mathcal{D}}$ is given by $\Psi^{H^{\mathcal{D}}}(X \circ \phi) = \Psi^{H^M} \circ X \circ \phi$, due to Theorem 5.8. By restricting this to the horizontal bundle Hor we have

$$(5.4) \quad \Psi^{H^{\mathcal{D}}}(\nabla f \circ \phi) = \Psi^{H^M} \circ \nabla f \circ \phi.$$

The flow $\Psi^{H^{\mathcal{D}}}$ is described in Corollary 5.10 and has the form $\Psi^{H^{\mathcal{D}}}(\nabla f \circ \phi) = \nabla f_t \circ \phi_t$, where f_t and ϕ_t are defined as required.

On the other hand, recall that the projection $\pi^{\mathcal{D}} : \mathcal{D} \rightarrow \mathcal{W}$ is defined by $\pi^{\mathcal{D}}(\phi) = \phi_*\mu$. The differential $D\pi$ of this map $\pi^{\mathcal{D}}$ is

$$D\pi(X \circ \phi) := (\phi_*\mu, -\mathcal{L}_X(\phi_*\mu)).$$

The application of this relation to (5.4) gives the result. □

Remark 5.13. The time-one-map for the above density flow ν_t in the Wasserstein space \mathcal{W} formally describes optimal transport maps for the Hamiltonian H^M . In particular, it recovers the optimal map recently obtained in [5]. One considers the optimal transport problem for the functional

$$\inf_{\phi} \left\{ \int_M c(x, \phi(x)) \mu \mid \phi_*\mu = \nu \right\}$$

with the cost function c defined by

$$c(x, y) = \inf_{\{\gamma \text{ paths between } x \text{ and } y\}} \int_0^1 L(\gamma, \dot{\gamma}) dt,$$

where the infimum is taken over paths γ joining x and y and the Lagrangian $L : TM \rightarrow \mathbb{R}$ satisfies certain regularity and convexity assumptions, see [5]. The corresponding Hamiltonian H^M in Theorem 5.12 is the Legendre transform of the Lagrangian L . Note that for the “kinetic energy” Lagrangian K^M , the above map becomes the optimal map $\exp^M(-\nabla f)$ mentioned at the beginning of this section, with $\exp^M : TM \rightarrow M$ being the Riemannian exponential of the manifold M .

6. The subriemannian geometry of diffeomorphism groups

In this section we develop the subriemannian setting for the diffeomorphism group. In particular, we derive the geodesic equations for the “nonholonomic Wasserstein metric,” and describe nonholonomic versions of the Monge–Ampère and heat equations.

Let M be a manifold with a fixed distribution τ on it. Recall that a subriemannian metric is a positive definite inner product $\langle \cdot, \cdot \rangle^\tau$ on each plane of the distribution τ smoothly depending on a point in M . Such a metric can be defined by the bundle map $\mathcal{I} : T^*M \rightarrow \tau$, sending a covector $\alpha_x \in T_x^*M$ to the vector V_x in the plane τ_x such that $\alpha_x(U) = \langle V_x, U \rangle^\tau$ on vectors $U \in \tau_x$. The subriemannian Hamiltonian $H^\tau : T^*M \rightarrow \mathbb{R}$ is the corresponding

fiberwise quadratic form:

$$(6.1) \quad H^\tau(\alpha_x) = \frac{1}{2} \langle V_x, V_x \rangle^\tau.$$

Let $\Psi_t^{H^\tau}$ be the Hamiltonian flow for time t of the subriemannian Hamiltonian H^τ on T^*M , while $\pi^{T^*M} : T^*M \rightarrow M$ is the cotangent bundle projection. Then the subriemannian exponential map $\exp^\tau : T^*M \rightarrow M$ is defined as the projection to M of the time-one-map of the above Hamiltonian flow on T^*M :

$$(6.2) \quad \exp^\tau(t\alpha_x) := \pi^{T^*M} \Psi_t^{H^\tau}(\alpha_x).$$

This relation defines a normal subriemannian geodesic on M with the initial covector α_x . Note that the initial velocity of the subriemannian geodesic $\exp^\tau(t\alpha_x)$ is $V_x = \mathcal{I}\alpha_x \in \tau_x$. So, unlike the Riemannian case, there are many subriemannian geodesics having the same initial velocity V_x on M .

Let d_τ be a subriemannian (or, Carnot–Caratheodory) distance on the manifold M , defined as the infimum of the length of all absolutely continuous admissible (i.e., tangent to τ) curves joining given two points. For a bracket-generating distribution τ any two points can be joined by such a curve, so this distance is always finite. Consider the corresponding optimal transport problem by replacing the Riemannian distance d in (4.1) with the subriemannian distance d_τ . Below we study the infinite-dimensional geometry of this subriemannian version of the optimal transport problem. Although, in general, normal subriemannian geodesics might not exhaust all the length minimizing geodesics in subriemannian manifolds (see [15]), we will see that in this problem of subriemannian optimal transport one can confine oneself to only such geodesics! The reason for this is that the *subriemannian* optimal transport induces a *Riemannian* metric structure on the space of densities.

6.1. Subriemannian submersion. Consider the following general setting: Let (Q, \mathcal{T}) be a subriemannian space, i.e., a manifold Q with a distribution \mathcal{T} and a subriemannian metric $\langle \cdot, \cdot \rangle^\tau$ on it. Suppose that $Q \rightarrow B$ is a bundle projection to a Riemannian base manifold B .

Definition 6.1. The projection $\pi : (Q, \mathcal{T}) \rightarrow B$ is a *subriemannian submersion* if the distribution \mathcal{T} contains a *horizontal subdistribution* \mathcal{T}^{hor} , orthogonal (with respect to the subriemannian metric) to the intersections of \mathcal{T} with fibers, and the projection π maps the spaces \mathcal{T}^{hor} isometrically to the tangent spaces of the base B , see Figure 4.

Let a subriemannian submersion $\pi : (Q, \mathcal{T}) \rightarrow B$ be a principal G -bundle $Q \rightarrow B$, where the distribution \mathcal{T} and the subriemannian metric are invariant with respect to the action of the group G . The following theorem is an analog of Corollary 4.4.

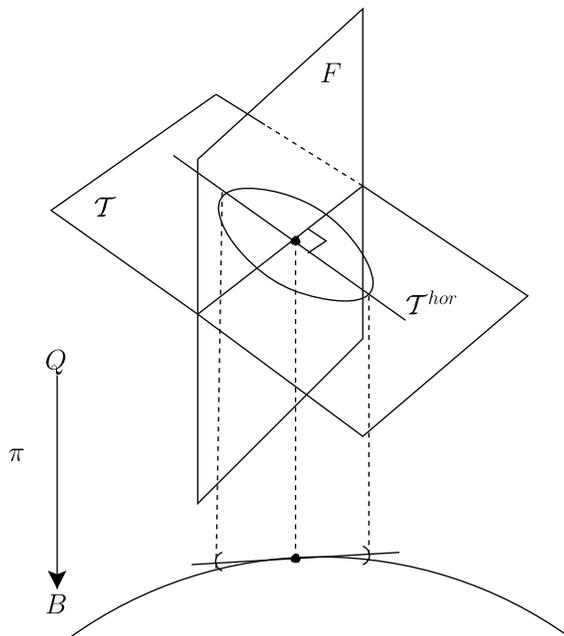


Figure 4. Subriemannian submersion: horizontal subdistribution \mathcal{T}^{hor} is mapped isometrically to the tangent bundle TB of the base.

Theorem 6.2. *For each point b in the base B and a point q in the fiber $\pi^{-1}(b) \subset Q$ over b , every Riemannian geodesic on the base B starting at b admits a unique lift to the subriemannian geodesic on Q starting at q with the velocity vector in \mathcal{T}^{hor} .*

Example 6.3. Consider the standard Hopf bundle $\pi : S^3 \rightarrow S^2$, with the two-dimensional distribution \mathcal{T} transversal to the fibers S^1 . Fix the standard metric on the base S^2 and lift it to a subriemannian metric on S^3 , which defines a subriemannian submersion. If the distribution \mathcal{T} is orthogonal to the fibers, the manifold (S^3, \mathcal{T}) can locally be thought of as the Heisenberg three-dimensional group. Then all subriemannian geodesics on S^3 with a given horizontal velocity project to a one-parameter family of circles on S^2 with a common tangent element. However, only one of these circles, the equator, is a geodesic on the standard sphere S^2 . Thus the equator can be uniquely lifted to a subriemannian geodesic on S^3 with the given initial vector.

Note that the uniqueness of this lifting holds even if the distribution \mathcal{T} is not orthogonal, but only transversal, say at a fixed angle, to the fibers S^1 , see Figure 5.

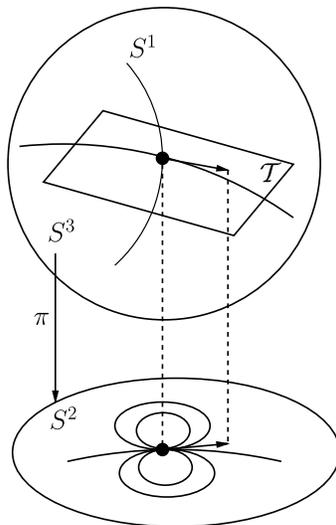


Figure 5. Projections of subriemannian geodesics from (S^3, \mathcal{T}) in the Hopf bundle give circles in S^2 , only one of which, the equator, is a geodesic on the base S^2 .

Proof of Theorem 6.2. To prove this theorem we describe the Hamiltonian setting of the subriemannian submersion.

Let Ver be the vertical subbundle in TQ (i.e., tangent planes to the fibers of the projection $Q \rightarrow B$). Define $\text{Ver}^\perp \subset T^*Q$ to be the corresponding annihilator, i.e., Ver_q^\perp is the set of all covectors $\alpha_q \in T_q^*Q$ at the point $q \in Q$ which annihilate the vertical space Ver_q .

Definition 6.4. The restriction of the subriemannian exponential map $\exp^\tau : T^*Q \rightarrow Q$ to the distribution Ver^\perp is called the *horizontal exponential*

$$\overline{\exp^\tau} : \text{Ver}^\perp \rightarrow Q$$

and the corresponding geodesics are the *horizontal subriemannian geodesics*.

The symplectic reduction identifies the quotient Ver^\perp/G with the cotangent bundle T^*B of the base. Note that the subdistribution \mathcal{T}^{hor} defines a horizontal bundle for the principal bundle $Q \rightarrow B$ in the usual sense. The definition of subriemannian submersion (translated to the cotangent spaces, where we replace \mathcal{T}^{hor} by Ver^\perp) gives that the subriemannian Hamiltonian $H^\mathcal{T}$ defined by (6.1) descends to a Riemannian Hamiltonian $H^{B,\mathcal{T}}$ on T^*B . Moreover, Hamiltonian trajectories of $H^{B,\mathcal{T}}$ starting at the cotangent space T_b^*B are in one-to-one correspondence with the trajectories of $H^\mathcal{T}$ starting at the space Ver_q^\perp . The projection of these Hamiltonian trajectories to the

manifolds B and Q via the cotangent bundle projections π^{T^*B} and π^{T^*Q} , respectively, gives the result. \square

Corollary 6.5. *For a subriemannian submersion, geodesics on the base give rise only to normal geodesics in the total space.*

In order to describe the geodesic geometry on the tangent, rather than cotangent, bundle of the manifold Q , we fix a Riemannian metric on Q whose restriction to the distribution τ is the given subriemannian metric $\langle \cdot, \cdot \rangle^\tau$. This Riemannian metric allows one to identify the cotangent bundle T^*Q with the tangent bundle TQ . Then the exponential map \exp^τ can be viewed as a map $TQ \rightarrow Q$. It is convenient to think of \mathcal{T}^{hor} as the horizontal bundle and identify it with the annihilator Ver^\perp . This way horizontal subriemannian geodesics are geodesics with initial (co)vector in the horizontal bundle \mathcal{T}^{hor} . This identification is particularly convenient for the infinite-dimensional setting, where we work with the tangent bundle of the diffeomorphism group.

6.2. A subriemannian analog of the Otto calculus. Fix a Riemannian metric $\langle \cdot, \cdot \rangle^M$ on the manifold M . Let $P^\tau : TM \rightarrow \tau$ be the orthogonal projection of vectors on M onto the distribution τ with respect to this metric. Let (ν, η_1) and (ν, η_2) be two tangent vectors in the tangent space at the point ν of the smooth Wasserstein space. Recall that for a fixed volume form μ , we define the subriemannian Laplacian as $\Delta^\tau f := \text{div}_\mu(P^\tau \nabla f)$.

Define a *nonholonomic Wasserstein metric* as the (weak) Riemannian metric on the (smooth) Wasserstein space \mathcal{W} given by

$$(6.3) \quad \langle (\nu, \eta_1), (\nu, \eta_2) \rangle^{\mathcal{W}, \mathcal{T}} := \int_M \langle P^\tau \nabla f_1(x), P^\tau \nabla f_2(x) \rangle^M \nu,$$

where functions f_1 and f_2 are solutions of the subriemannian Poisson equation

$$-(\Delta^\tau f_i)\nu = \eta_i$$

for the measure ν .

Theorem 6.6. *The geodesics on the Wasserstein space \mathcal{W} equipped with the nonholonomic Wasserstein metric (6.3) have the form $(\overline{\exp^\tau}(tP^\tau \nabla f))_* \nu$, where $\overline{\exp^\tau} : \mathcal{T}^{\text{hor}} \rightarrow M$ is the horizontal exponential map and ν is any point of \mathcal{W} .*

To prove this theorem we first note that the Riemannian metric $\langle \cdot, \cdot \rangle^{\mathcal{D}}$ defined on the diffeomorphism group restricts to a subriemannian metric $\langle \cdot, \cdot \rangle^{\mathcal{D}, \mathcal{T}}$ on the right invariant bundle \mathcal{T} .

Proposition 6.7. *The map $\pi : (\mathcal{D}, \mathcal{T}) \rightarrow \mathcal{W}$ is a subriemannian submersion of the subriemannian metric $\langle \cdot, \cdot \rangle^{\mathcal{D}, \mathcal{T}}$ on the diffeomorphism group with distribution \mathcal{T} to the nonholonomic Wasserstein metric $\langle \cdot, \cdot \rangle^{\mathcal{W}, \mathcal{T}}$.*

Proof. This statement can be derived from the Hamiltonian reduction, similarly to the Riemannian case.

Here we prove it by an explicit computation. Recall that the map $\pi : \mathcal{D} \rightarrow \mathcal{W}$ is defined by $\pi(\phi) = \phi_*\mu$. Let $X \circ \phi$ be a tangent vector at the point ϕ in the diffeomorphism group \mathcal{D} . Consider the flow ϕ_t of the vector field X , and note that $\pi(\phi_t \circ \phi) = \phi_{t*}\phi_*\mu$. To compute the derivative $D\pi$ we differentiate this equation with respect to time t at $t = 0$:

$$D\pi(X \circ \phi) = \mathcal{L}_{-X}(\phi_*\mu) = -(\operatorname{div}_{\phi_*\mu} X)\phi_*\mu,$$

by the definition of Lie derivative. A vector field X from the horizontal bundle \mathcal{T}^{hor} has the form $(P^\tau \nabla f) \circ \phi$, and for it the equation becomes

$$D\pi((P^\tau \nabla f) \circ \phi) = -(\Delta^\tau f)\phi_*\mu,$$

where the Laplacian Δ^τ is taken with respect to the volume form $\phi_*\mu$.

Therefore, for horizontal tangent vectors $(P^\tau \nabla f_1) \circ \phi$ and $(P^\tau \nabla f_2) \circ \phi$ at the point ϕ their subriemannian inner product is

$$\langle (P^\tau \nabla f_1) \circ \phi, (P^\tau \nabla f_2) \circ \phi \rangle^{\mathcal{D}} = \int_M \langle P^\tau \nabla f_1 \circ \phi, P^\tau \nabla f_2 \circ \phi \rangle^M \mu.$$

After the change of variables this becomes

$$\int_M \langle P^\tau \nabla f_1, P^\tau \nabla f_2 \rangle^M \phi_*\mu = \langle D\pi((P^\tau \nabla f_1) \circ \phi), D\pi((P^\tau \nabla f_2) \circ \phi) \rangle^{\mathcal{W}, \mathcal{T}},$$

which completes the proof. \square

Proof of Theorem 6.6. To describe geodesics in the nonholonomic Wasserstein space we define the Hamiltonian $H^\mathcal{T} : T\mathcal{D} \rightarrow \mathbb{R}$ by

$$(6.4) \quad H^\mathcal{T}(X \circ \phi) := \int_M \langle (P^\tau X) \circ \phi, (P^\tau X) \circ \phi \rangle \mu.$$

The Hamiltonian flow with Hamiltonian $H^\mathcal{T}$ has the form $\exp^\tau((tP^\tau X) \circ \phi)$ according to Theorem 5.8. By taking its restriction to the bundle \mathcal{T}^{hor} and projecting to the base we obtain that the geodesics on the smooth Wasserstein space are

$$(\overline{\exp^\tau}((tP^\tau \nabla f) \circ \phi))_*\nu,$$

where $\nu = \phi_*\mu$ and $P^\tau \nabla f$ is defined by the Hodge decomposition for the field X . This completes the proof of Theorem 6.6. \square

Remark 6.8. For a horizontal subriemannian geodesic $\varphi_t(x) := \overline{\exp^\tau}(tP^\tau \nabla f(x))$ with a smooth function f , the diffeomorphism φ_t satisfies $\frac{d}{dt}\varphi_t = (P^\tau \nabla f_t) \circ \varphi_t$ and f_t is the solution of the Hamilton–Jacobi equation

$$(6.5) \quad \dot{f}_t + H^\mathcal{T}(\nabla f_t) = 0$$

with the initial condition $f_0 = f$, see Corollary 5.10. This equation determines horizontal subriemannian geodesics on the diffeomorphism group \mathcal{D} .

In the Riemannian case, one can see that the vector fields $V_t = \frac{d}{dt}\varphi_t = \nabla f_t \circ \varphi_t$ satisfy the Burgers equation by taking the gradient of the both sides in (6.5), cf. Proposition 4.1. Hence equation (6.5) can be viewed as a subriemannian analog of the potential Burgers equation in \mathcal{D} . However, a subriemannian analog of the Burgers equation for nonhorizontal (i.e., non-potential) normal geodesics on the diffeomorphism group is not so explicit.

Remark 6.9. If the function f is smooth, the time-one-map $\varphi(x) := \overline{\exp^\tau(P^\tau \nabla f(x))}$ along the geodesics described in Theorem 6.6 satisfies the following *nonholonomic* analog of the *Monge–Ampère equation*: $h(\varphi(x)) \det(D\varphi(x)) = g(x)$, where g and h are functions on the manifold M defining two densities $\theta = g \text{ vol}$ and $\nu = h \text{ vol}$.

Furthermore, for the case of the Heisenberg group this formal solution $\varphi(x)$ coincides with the optimal map obtained in [1]. The (minus) potential $-f$ of the corresponding optimal map satisfies the c -concavity condition for $c = d_\tau^2/2$, where d_τ^2 is the subriemannian distance, cf. Remark 4.5.

6.3. The nonholonomic heat equation. Consider the heat equation $\partial_t u = \Delta u$ on a function u on the manifold M , where the operator Δ is given by $\Delta f = \operatorname{div}_\mu \nabla f$. Upon multiplying both sides of the heat equation by the fixed volume form μ , one can regard it as an evolution equation on the smooth Wasserstein space \mathcal{W} . Note that the right-hand side of the heat equation gives a tangent vector $(\Delta u)\mu$ at the point $u\mu$ of the Wasserstein space. The Boltzmann *relative entropy* functional $\operatorname{Ent} : \mathcal{W} \rightarrow \mathbb{R}$ is defined by the integral

$$(6.6) \quad \operatorname{Ent}(\nu) := \int_M \log(\nu/\mu) \nu.$$

The gradient flow of Ent on the Wasserstein space with respect to the metric \tilde{d} gives the heat equation, see [20].

Recall that one can define the subriemannian Laplacian: $\Delta^\tau f := \operatorname{div}_\mu(P^\tau \nabla f)$ for a fixed volume form μ on M . The natural generalization of the heat equation to the nonholonomic setting is as follows.

Definition 6.10. The *nonholonomic* (or, *subriemannian*) *heat equation* is the equation $\partial_t u = \Delta^\tau u$ on a time-dependent function u on M .

Below we show that this equation in the nonholonomic setting also admits a gradient interpretation on the Wasserstein space.

Theorem 6.11. *The nonholonomic heat equation $\partial_t u = \Delta^\tau u$ describes the gradient flow on the Wasserstein space with respect to the relative entropy functional (6.6) and the nonholonomic Wasserstein metric (6.3).*

Namely, for the volume form $\nu_t := g_{t}\mu$ and the gradient $\nabla^{\mathcal{W},\mathcal{T}}$ with respect to the metric $\langle, \rangle^{\mathcal{W},\mathcal{T}}$ on the Wasserstein space one has*

$$\frac{\partial}{\partial t} \nu_t = -\nabla^{\mathcal{W},\mathcal{T}} \operatorname{Ent}(\nu_t) = \Delta^\tau(\nu_t/\mu)\mu.$$

Proof. Denote by (ν, η) a tangent vector to the Wasserstein space \mathcal{W} at a point $\nu \in \mathcal{W}$, where η is a volume form of total integral zero. Let Δ_ν^τ be the subriemannian Laplacian with respect to the volume form ν .

Let h and h_{Ent} be real-valued functions on the manifold M such that $-(\Delta_\nu^\tau h)\nu = \eta$ and $-(\Delta_\nu^\tau h_{\text{Ent}})\nu = \nabla^{\mathcal{W}, \mathcal{T}} \text{Ent}(\nu)$ for the entropy functional Ent . Then, by definition of the metric $\langle \cdot, \cdot \rangle^{\mathcal{W}, \mathcal{T}}$ given by (6.3), we have

$$(6.7) \quad \langle (\nu, \nabla^{\mathcal{W}, \mathcal{T}} \text{Ent}(\nu)), (\nu, \eta) \rangle^{\mathcal{W}, \mathcal{T}} = \int_M \langle P^\tau \nabla h_{\text{Ent}}(x), P^\tau \nabla h(x) \rangle^M \nu.$$

On the other hand, by definitions of Ent and the gradient $\nabla^{\mathcal{W}, \mathcal{T}}$ on the Wasserstein space, one has

$$\begin{aligned} \langle (\nu, \nabla^{\mathcal{W}, \mathcal{T}} \text{Ent}(\nu)), (\nu, \eta) \rangle^{\mathcal{W}, \mathcal{T}} &:= \left. \frac{d}{dt} \right|_{t=0} \text{Ent}(\nu + t\eta) \\ &= \left. \frac{d}{dt} \right|_{t=0} \int_M \left[\log \left(\frac{\nu + t\eta}{\mu} \right) \right] (\nu + t\eta). \end{aligned}$$

After differentiation and simplification the latter expression becomes $\int_M \log(\nu/\mu) \eta$, where we used that $\int_M \eta = 0$. This can be rewritten as

$$\int_M \log(\nu/\mu) \eta = - \int_M \log(\nu/\mu) \mathcal{L}_{P^\tau \nabla h} \nu = \int_M (\mathcal{L}_{P^\tau \nabla h} \log(\nu/\mu)) \nu,$$

by using the Leibnitz property of the Lie derivative \mathcal{L} on the Wasserstein space and the fact that $-(\Delta_\nu^\tau h)\nu = \eta$. Note that the Lie derivative is the inner product with the gradient, and hence

$$\begin{aligned} \int_M (\mathcal{L}_{P^\tau \nabla h} \log(\nu/\mu)) \nu &= \int_M \langle \nabla \log(\nu/\mu), P^\tau \nabla h \rangle^M \nu \\ &= \int_M \langle P^\tau \nabla \log(\nu/\mu), P^\tau \nabla h \rangle^M \nu. \end{aligned}$$

Comparing the latter form with (6.7), we get $P^\tau \nabla h_{\text{Ent}} = P^\tau \nabla \log(\nu/\mu)$, or, after taking the divergence of both parts and using the definition of function h_{Ent} ,

$$\nabla^{\mathcal{W}, \mathcal{T}} \text{Ent}(\nu) = -\Delta_\nu^\tau (\log(\nu/\mu)) \nu.$$

Finally, let us show that the right-hand side of the above equation coincides with $-\Delta_\mu^\tau (\nu/\mu) \mu$. Indeed, the chain rule gives

$$\mathcal{L}_{P^\tau \nabla \log(\nu/\mu)} \nu = \mathcal{L}_{(\mu/\nu) P^\tau \nabla (\nu/\mu)} \nu = (\mu/\nu) \mathcal{L}_{P^\tau \nabla (\nu/\mu)} \nu + d(\mu/\nu) \wedge i_{P^\tau \nabla (\nu/\mu)} \nu.$$

The last term is equal to $(i_{P^\tau \nabla (\nu/\mu)} d(\mu/\nu)) \nu = \mathcal{L}_{P^\tau \nabla (\nu/\mu)} (\mu/\nu) \nu$, which implies that

$$\mathcal{L}_{P^\tau \nabla \log(\nu/\mu)} \nu = \mathcal{L}_{P^\tau \nabla (\nu/\mu)} \mu$$

by the Leibnitz property of Lie derivative. Thus

$$\begin{aligned}\Delta_{\nu}^{\tau}(\log(\nu/\mu))\nu &= \operatorname{div}_{\nu}(P^{\tau}\nabla(\log(\nu/\mu))\nu) \\ &= \mathcal{L}_{P^{\tau}\nabla\log(\nu/\mu)}\nu = \mathcal{L}_{P^{\tau}\nabla(\nu/\mu)}\mu = \Delta_{\mu}^{\tau}(\nu/\mu)\mu.\end{aligned}$$

The above shows that the nonholonomic heat equation is the gradient flow on the Wasserstein space for the same potential as the classical heat equation, but with respect to the nonholonomic Wasserstein metric. \square

Acknowledgments

We are much indebted to R. Beals, Ya. Eliashberg, V. Ivrii, G. Misiolek, D.-M. Nhieu, R. Ponge, and M. Shubin for fruitful discussions and to the anonymous referee for useful remarks. B.K. is grateful to the IHES in Bures-sur-Yvette for its stimulating environment. This research was partially supported by an NSERC research grant.

References

- [1] L. Ambrosio and S. Rigot, *Optimal mass transportation in the Heisenberg group*, J. Func. Anal. **208** (2004), 261–301.
- [2] V.I. Arnold and A.B. Givental, *Symplectic geometry, dynamical systems IV*, Encyclopaedia Math. Sci. **4**, Springer, Berlin, 2001, 1–138.
- [3] V.I. Arnold and B.A. Khesin, *Topological methods in hydrodynamics*, Appl. Math. Sci. **125**, Springer-Verlag, New York, 1998, xvi + 374 pp.
- [4] G. Bande, P. Ghiggini and D. Kotschick, *Stability theorems for symplectic and contact pairs*, Int. Math. Res. Not. **68** (2004), 3673–3688.
- [5] P. Bernard and B. Buffoni, *Optimal mass transportation and Mather theory*, J. Eur. Math. Soc. (JEMS) **9**(1) (2007), 85–121.
- [6] Y. Brenier, *Polar factorization and monotone rearrangements of vector-valued functions*, Comm. Pure Appl. Math. **44**(4) (1991), 375–417.
- [7] D. Ebin and J. Marsden, *Groups of diffeomorphism and the motion of an incompressible fluid*, Ann. Math. **92**(2) (1970), 102–163.
- [8] É. Ghys, *Feuilletages riemanniens sur les variétés simplement connexes*, Ann. Inst. Fourier (Grenoble) **34**(4) (1984), 203–223.
- [9] J.W. Gray, *Some global properties of contact structures*, Ann. Math. **69**(2) (1959), 421–450.
- [10] L. Hörmander, *Hypoelliptic second order differential equations*, Acta Math. **119** (1967), 147–171.
- [11] L. Hörmander, *The analysis of linear partial differential operators III, Pseudo-differential operators*, Springer, Berlin, 2007.
- [12] B. Khesin and G. Misiolek, *Shock waves for the Burgers equation and curvatures of diffeomorphism groups*, Proc. Steklov Inst. Math. **v.250** (2007), 1–9.
- [13] J. Lott, *Optimal transport and Perelman’s reduced volume*, Calc. Var. Partial Differential Equations, **36** (2009), 49–84.

- [14] R. McCann, *Polar factorization of maps in Riemannian manifolds*, *Geom. Funct. Anal.* **11**(3) (2001), 589–608.
- [15] R. Montgomery, *A tour of subriemannian geometries, their geodesics and applications*, *Mathem. Surveys and Monographs*, 91. American Mathematical Society, Providence, RI, 2002.
- [16] J. Moser, *On the volume elements on a manifold*, *Trans. AMS*, **120**(2) (1965), 286–294.
- [17] D.-M. Nhieu, *The Neumann problem for sub-Laplacians on Carnot groups and the extension theorem for Sobolev spaces*, *Ann. Mat. Pura Appl. IV*, **180**(1) (2001), 1–25.
- [18] D.-M. Nhieu and N. Garofalo, *Lipschitz continuity, global smooth approximations and extension theorems for Sobolev functions in Carnot–Carathéodory spaces*, *J. Anal. Math.* **74** (1998), 67–97.
- [19] B. O’Neill, *Submersions and geodesics*, *Duke Math. J.* **34** (1967), 363–373.
- [20] F. Otto, *The geometry of dissipative evolution equations: the porous medium equation*, *Comm. Partial Differential Equations* **26**(1–2) (2001), 101–174.
- [21] L.P. Rothschild and E. Stein, *Hypoelliptic differential operators and nilpotent group*, *Acta Math.* **137** (1976), 247–320.
- [22] L.P. Rothschild and D. Tartakoff, *Parametrices with C^∞ error for cm_b and operators of Hörmander type*, in *Partial differential equations and geometry (Proc. Conf., Park City, Utah, 1977)*, 255–271, *Lecture Notes in Pure and Appl. Math.* **48**, Dekker, New York, 1979.
- [23] A.I. Shnirelman, *The geometry of the group of diffeomorphisms and the dynamics of an ideal incompressible fluid*, *Math. USSR-Sb.* **56** (1987), 79–105.
- [24] M.E. Taylor, *Partial differential equations. I. Basic theory*, *Appl. Math. Sci.* **115**, Springer-Verlag, New York, 1996.
- [25] C. Villani, *Topics in mass transportation*, AMS, Providence, RI, 2003.

DEPARTMENT OF MATHEMATICS,
UNIVERSITY OF TORONTO,
40 ST. GEORGE STREET,
TORONTO, ON M5S 2E4 CANADA
E-mail address: khesin@math.toronto.edu, plee@math.toronto.edu

Received 06/25/2008, accepted 07/07/2009