OPTIMAL TRANSPORTATION
UNDER NONHOLONOMIC CONSTRAINTS

ANDREI AGRACHEV AND PAUL LEE

ABSTRACT. We study Monge's optimal transportation problem, where the cost is given by an optimal control cost. We prove the existence and uniqueness of an optimal map under certain regularity conditions on the Lagrangian, absolute continuity of the measures with respect to Lebesgue, and most importantly the absence of sharp abnormal minimizers. In particular, this result is applicable in the case of subriemannian manifolds with a 2-generating distribution and cost given by d^2, where d is the subriemannian distance. Also, we discuss some properties of the optimal plan when abnormal minimizers are present. Finally, we consider some examples of displacement interpolation in the case of the Grushin plane.

1. INTRODUCTION

Let $(\mathcal{X}, \mu)$, $(\mathcal{Y}, \nu)$ be probability spaces and let $c : \mathcal{X} \times \mathcal{Y} \to \mathbb{R} \cup \{+\infty\}$ be a fixed measurable function. The Monge optimal transportation problem is the minimization of the following functional:

$$\int_{\mathcal{X}} c(x, \phi(x)) \, d\mu$$

over all the Borel maps $\phi : \mathcal{X} \to \mathcal{Y}$ which push forward μ to ν: $\phi_* \mu = \nu$. Maps ϕ which achieve the infimum above are called optimal maps. In this paper, we will only consider the case when $\mathcal{X} = \mathcal{Y} = M$ is a manifold.

In 1942, Kantorovich studied a relaxed version of Monge's problem in his famous paper [15]. However, a huge step toward solving the original problem was not achieved until a decade ago by Brenier. In [8], Brenier proved the existence and uniqueness of an optimal map in the case where $M = \mathbb{R}^n$ and the cost function c is given by $c(x, y) = |x - y|^2$. Later, this is generalized, by McCann [18], to the case of a closed Riemannian manifold M with the cost given by the square of the Riemannian distance $c(x, y) = d^2(x, y)$. Recently, Bernard and Buffoni [7] generalized this further to the case where the cost c is the action associated to a Lagrangian function $L : TM \to \mathbb{R}$ on a compact manifold M. More precisely, the cost is given by

$$(1) \qquad c(x, y) = \inf_{x(0)=x, x(1)=y} \int_0^1 L(x(t), \dot{x}(t)) dt,$$

where the infimum is taken over all curves joining the points x and y, and the Lagrangian L is fibrewise strictly convex with superlinear growth.

In this paper, we consider costs similar to (1). However, instead of minimizing among all curves, the infimum is taken over a subcollection of curves, called admissible paths. These paths are given by a control system and the corresponding cost function is called the optimal control cost. More precisely, a control system is a smooth fiber-preserving map F of a locally trivial bundle $P \to M$ over the manifold M into its tangent bundle TM. If the fibres of the bundle $P \to M$ are diffeomorphic to a set U, then the map $F : P \to TM$ can be written locally as $F : (x, u) \mapsto F(x, u)$, where x is in the manifold M and u is in the set U. We assume that U is a closed subset of a Euclidean space. Admissible controls are measurable bounded maps from $[0, 1]$ to U, and admissible paths are Lipschitz curves which satisfy the equation

$$(2) \qquad\qquad \dot{x}(t) = F(x(t), u(t)),$$

where $u(\cdot)$ is an admissible control. Let $L : M \times U \to \mathbb{R}$ be a Lagrangian. Then the corresponding cost c is given by

$$(3) \qquad\qquad \inf_{(x(\cdot), u(\cdot))} \int_0^1 L(x(t), u(t))\, dt,$$

where the infimum is taken over all admissible pairs $(x(\cdot), u(\cdot)) : [0, 1] \to M \times U$ such that $x(0) = x,\ y(0) = y$.

In the interesting cases, the dimension of U is smaller than that of M and, nevertheless, any two points of M can be connected by an optimal admissible path. In other words, the control system works as a nonholonomic constraint. The shortage of admissible velocities does not allow us to recover an optimal path from its initial point and initial velocity and the Euler–Lagrange description of the extremals does not work well. On the other hand, the Hamiltonian approach remains efficient thanks to the Pontryagin maximum principle. Another difficulty is the appearance of so-called abnormal extremals (singularities of the space of admissible paths) which we are obliged to fight with.

In sections 2 and 3, we will recall some basic notions in optimal control theory and the theory of optimal mass transportation which are necessary for this paper.

In section 4, by using the arguments in the theory of optimal mass transportation and the Pontryagin maximum principle in optimal control theory, we show the existence and uniqueness of an optimal map under some regularity assumptions (Theorem 4.1). All these conditions are mild except the Lipschitz continuity of the cost function. However, this is well known in all of the above cases mentioned. So, the theorem generalizes the work in [8, 18, 7].

In section 5, we study the Lipschitz continuity of the cost function. If abnormal minimizers are absent, then the cost is not only Lipschitz but even semi-concave (see [9]). Unfortunately, abnormal minimizers are unavoidable in many interesting problems and, in particular, in all subriemannian problems. It happens, however, that not all abnormal minimizers are dangerous. To keep the Lipschitz property of the cost (though not the semi-concavity), it is sufficient that the, so-called, sharp abnormal minimizers are absent. Sharp paths are essentially singularities of the space of admissible paths whose neighborhoods in the second order approximations are contained in quadrics with a finite Morse index. Geometric control theory provides simple effective conditions of the sharpness (see, for instance, [4, 6]). These

conditions allow us to prove Lipschitz continuity for a large class of optimal control costs, hence proving the existence and uniqueness of an optimal map of the corresponding Monge problem (Theorem 6.3).

In section 6, we apply the above results to some subriemannian manifolds, where the cost function is given by the square of the subriemannian distance (see section 6 for the basic notions in subriemannian geometry). In the case of a subriemannian manifold, all the mild regularity assumptions are satisfied. Using the result in [6] mentioned above (Proposition 5.2), Lipschitz continuity of the cost can be easily proven in the case of a step 2 distribution (Corollary 6.2), hence proving existence and uniqueness of an optimal map (Theorem 6.3). This generalizes the corresponding result by Ambrosio and Rigot [1] on the Heisenberg group.

In sections 7 and 8, we prove certain properties of the optimal plan when abnormal minimizers are present. In section 7, we consider flows whose trajectories are strictly abnormal minimizers. We show that these flows cannot be an optimal plan for all "nice" initial measures if the cost is continuous. On the contrary, in section 8, we show that these flows are indeed optimal for an important class of problems with discontinuous costs.

In section 9, we study two examples on the Grushin plane for which the results in sections 3 and 4 apply.

2. Elementary optimal control theory

In this section, we recall some notions from optimal control theory. See [4, 13, 14] for the details. Let M be a smooth manifold and let U be a closed subset in $\mathbb{R}^m$ which is called the control set. Let $F : M \times U \to TM$ be a Lipschitz continuous function such that $F_u := F(\cdot, u) : M \to TM$ are smooth vector fields for each point u in the control set U. Assume that the function $(x, u) \mapsto \frac{\partial}{\partial x} F(x, u)$ is continuous. Curves $u(\cdot) : [0, 1] \to U$ in the control set U which are locally bounded and measurable (i.e. $u(\cdot) \in L^\infty([0, 1], U)$) are called admissible controls.

A control system is the following ordinary differential equation with parameters varying over all admissible controls:

$$(4) \qquad \dot{x}(t) = F(x(t), u(t)).$$

The solutions $x(t)$ to the above control system are called admissible paths and $(x(t), u(t))$ are called admissible pairs.

By the classical theory of ordinary differential equations, a unique solution to the system (4) exists locally for almost all time t. Moreover, the resulting local flow is smooth in the space variable x and Lipschitz in the time variable t. The control system is complete if the flows of all control vector fields exist globally.

Let x_0 and x_1 be two points on the manifold M. Denote by $\mathfrak{C}_{x_0}$ the set of all admissible pairs $(x(\cdot), u(\cdot))$ for which the corresponding admissible paths $x(\cdot)$ start at the point x_0, and denote by $\mathfrak{C}_{x_0}^{x_1}$ those pairs in $\mathfrak{C}_{x_0}$ whose admissible paths end at x_1. A control system is called controllable if the set $\mathfrak{C}_{x_0}^{x_1}$ is always nonempty for any pair of points x_0 and x_1 on the manifold.

Let $L : M \times U \to \mathbb{R}$ be a smooth function, called a Lagrangian, and define the cost function corresponding to this Lagrangian as follows:

$$(5) \qquad c(x_0, x_1) = \begin{cases} \inf\limits_{(x(\cdot), u(\cdot)) \in \mathfrak{C}_{x_0}^{x_1}} \int_0^1 L(x(t), u(t))\, dt & \text{if } \mathfrak{C}_{x_0}^{x_1} \neq \emptyset, \\ +\infty & \text{otherwise.} \end{cases}$$

The cost function defined above is said to be complete if given any pair of points (x_0, x_1), there exists an admissible pair which achieves the infimum above and the corresponding admissible path starts from x_0 and ends at x_1.

Remark 2.1. The infimum of the problem in (5) can be equivalently characterized by taking the infimum over all admissible controls $u(\cdot)$ such that the corresponding admissible paths start at the point x_1, end at the point x_0 of the manifold and satisfy the following control system:

$$\dot{x}(s) = -F(x(s), u(s)).$$

This point will become important for later discussion.

Consider the following minimization problem, commonly known as the Bolza problem:

Problem 2.2. Find minimizers for

$$\inf_{(x(\cdot), u(\cdot)) \in \mathfrak{C}_{x_0}} \int_0^1 L(x(s), u(s)) \, ds - f(x(1)).$$

Let $\pi : T^*M \to M$ be the cotangent bundle projection. For each point u in the control set U, define the corresponding Hamiltonian function $H_u : T^*M \to \mathbb{R}$ by

$$H_u(p_x) = p_x(F(x, u)) + L(x, u).$$

If $H : T^*M \to \mathbb{R}$ is a function on the cotangent bundle, we denote its Hamiltonian vector field by $\overrightarrow{H}$. Also, recall that α is in the sub-differential $d^- f_x$ of f at x if there is a C^1 function ϕ which satisfies $d\phi_x = \alpha$ and touches f from below at x. By touching f from below at x, we mean that $\phi \leq f$ and $\phi(x) = f(x)$. α is in the super-differential $d^+ f_x$ of f at x if $-\alpha$ is in the sub-differential of $-f$ at x. It is not hard to see that f is differentiable at x if and only if both the super-differential and the sub-differential of f at x are nonempty, and $d^+ f_x = d^- f_x = \{df_x\}$ in this case. See for instance [10, 24] for a detailed discussion on generalized differentials. Next, we present an elementary version of the Pontryagin maximum principle which we prove in the appendix for the convenience of the reader.

Theorem 2.3 (Pontryagin Maximum Principle for Bolza Problem). *Let $(\widetilde{x}(\cdot), \widetilde{u}(\cdot))$ be an admissible pair which achieves the infimum in Problem* 2.2. *Assume that the function f in Problem* 2.2 *is sub-differentiable at the point $\widetilde{x}(1)$. Then, for each α in the sub-differential $d^- f_{\widetilde{x}(1)}$ of f, there exists a Lipschitz path $\widetilde{p} : [0, 1] \to T^*M$ which satisfies the following for almost all time t in the interval $[0, 1]$:*

(6)
$$\begin{cases} \pi(\widetilde{p}(t)) = \widetilde{x}(t), \\ \widetilde{p}(1) = -\alpha, \\ \dot{\widetilde{p}}(t) = \overrightarrow{H}_{\widetilde{u}(t)}(\widetilde{p}(t)), \\ H_{\widetilde{u}(t)}(\widetilde{p}(t)) = \min_{u \in U} H_u(\widetilde{p}(t)). \end{cases}$$

Remark 2.4. Note that if we consider the minimization problem (5) instead of the Bolza problem, it is well known that there are minimizers which do not satisfy the Hamiltonian system (6). They are called abnormal minimizers and their existence is due to the fact that the endpoint mapping is not a submersion. Such a problem does not arise in the Bolza problem since we are minimizing among curves $x(\cdot)$ with only the initial point $x(0)$ fixed. (See below for definitions of endpoint mapping and abnormal minimizers.)

Remark 2.5. Let $\Delta \subset TM$ be a distribution on an n-dimensional manifold M. That is, for each point x in the manifold M, it smoothly assigns a vector subspace Δ_x of the tangent space T_xM. Assume that the distribution Δ is trivializable, i.e. that there exists a system of vector fields $X_1, ..., X_k$ which span Δ at every point: $\Delta_x = \text{span}\{X_1(x), ..., X_k(x)\}$. Consider the following control system:

$$(7) \qquad \dot{x}(t) = \sum_{i=1}^{k} u_i(t) X_i(x(t)),$$

with initial condition $x(0) = x$ and final condition $x(1) = y$. Recall that we denote by $\mathfrak{C}_x^y$ the set of all admissible pairs $(x(\cdot), u(\cdot))$ such that the admissible path $x(\cdot)$ satisfies $x(0) = x$ and $x(1) = y$. Let c be the cost given by

$$(8) \qquad c(x, y) = \inf_{(x(\cdot), u(\cdot)) \in \mathfrak{C}_x^y} \int_0^1 \sum_{i=1}^{k} u_i^2 \, dt.$$

If the number of vector fields k is equal to the dimension n of the manifold M and the vector fields $X_1, ..., X_k$ are everywhere linearly independent, then the distribution Δ is the same as the tangent bundle TM of M and the admissible paths of the control system (7) are all the paths on M. It also defines a Riemannian metric on M by declaring that the vector fields $X_1, ..., X_n$ are orthonormal everywhere. The cost function c is the square of the Riemannian distance d: $c = d^2$, and the minimizers of this system correspond to the length minimizing geodesics on M. However, this does not work for distributions which are not trivializable.

To overcome this difficulty, we can modify the general definition of control systems in the following way. Let P be a locally trivial bundle on M with bundle projection $\pi^P : P \to M$ and let $F : P \to TM$ be a fibre-preserving map, i.e. $F(P_x) \subseteq T_xM$. The control system corresponding to the map F is given by

$$(9) \qquad \dot{x}(t) = F(v(t)).$$

The admissible pairs $v(\cdot) : [0, 1] \to P$ are locally bounded measurable paths in P such that its projection to the manifold M is a Lipschitz path: $x(\cdot) = \pi^P(v(\cdot))$ is Lipschitz. If we let P be the trivial bundle $M \times U$, we get back the system (4). If a Lagrangian $L : P \to \mathbb{R}$ is fixed, then the corresponding cost function c is defined by

$$(10) \qquad c(x, y) = \inf_{v(\cdot) \in \mathfrak{C}_x^y} \int_0^1 L(v(t)) dt,$$

where the infimum is taken over all admissible pairs $v(\cdot) : [0, 1] \to P$ such that the corresponding admissible path $x(\cdot) = \pi^P(v(\cdot))$ satisfies $x(0) = x$ and $x(1) = y$.

Let $\langle, \rangle$ be a Riemannian metric on the manifold M. If P is the tangent bundle TM of M, the map F is the identity map and the Lagrangian $L : P \to \mathbb{R}$ is given by $L(v) = \langle v, v \rangle$, then the cost function c is equal to the square of the Riemannian distance. If $k < n$, then the admissible paths of the control system (7) are paths tangent to the distribution Δ. Similar to the Riemannian case, the control system defines a subriemannian metric $\langle, \rangle^S$. (See section 6 for the basics on subriemannian geometry.) The cost (8) is the square of the subriemannian distance d_S: $c = d_S^2$. For general distributions Δ which are not trivializable, consider the general control system (9) with $P = \Delta$. $F : \Delta \hookrightarrow TM$ is the inclusion map. If the Lagrangian L is

defined by $L(v) = \langle v, v \rangle^S$, then the cost is again the square of the subriemannian distance.

In this paper (except in section 8), we consider the control systems of the form (4) in order to avoid heavy notation. All the results have an easy generalization to the more general intrinsically defined systems just introduced.

3. Optimal mass transport

The theory of optimal mass transportation is about moving one mass to another that minimizes certain costs. More precisely, let M be a manifold and consider a function $c : M \times M \to \mathbb{R} \cup \{+\infty\}$, called the cost function. Let μ and ν be two Borel probability measures on the manifold M. Then the optimal mass transportation is the following problem:

Problem 3.1. Find a Borel map which achieves the following infimum among all Borel maps $\phi : M \to M$ that push the probability measure μ forward to ν:

$$\inf_{\phi_* \mu = \nu} \int_M c(x, \phi(x)) \, d\mu.$$

Here, we recall that the push forward $\phi_* \mu$ of a measure μ by a Borel map ϕ is defined by $\phi_* \mu(B) = \mu(\phi^{-1}(B))$ for all Borel sets B in M. In many cases, such a problem admits a solution which is unique (up to measure zero), assuming absolute continuity of the measure μ with respect to the Lebesgue measure. This unique solution to (3.1) is called the optimal map or the Brenier map.

The first optimal map was found by Brenier in [8] in the case where the manifold was $\mathbb{R}^n$ and the cost was $c(x, y) = |x - y|^2$. Later, it was generalized to arbitrary closed, connected Riemannian manifolds in [18] with the cost given by the square of the Riemannian distance. The case for the Heisenberg group with the cost given by d^2 was done in [1], where d was the subriemannian distance or the gauge distance. In [7], a much more general cost given by the action associated to a Lagrangian function $L : TM \to \mathbb{R}$ on a compact manifold M was considered. More precisely,

$$(11) \qquad c(x, y) = \inf_{x(0)=x, x(1)=y} \int_0^1 L(x(t), \dot{x}(t)) dt,$$

where the infimum is taken over all curves joining the points x and y.

The existence and uniqueness of an optimal map with the cost given by (11) was shown under the following assumptions:

- The Lagrangian L is fibrewise strictly convex; i.e. the map restriction of L to the tangent space $T_x M$ is strictly convex for each fixed x in the manifold M.
- L has superlinear growth; i.e. $L(v)/|v| \to 0$ as $|v| \to \infty$.
- The cost c is complete; i.e. the infimum (11) is always achieved by some C^2 smooth paths.

Recently, the compactness assumption on the manifold or on the measures was eliminated by [12, 11].

In this paper, we consider a connected manifold M without boundary and the cost function c is given by (5). Consider the following relaxed version of Problem 3.1, called Kantorovich reformulation. Let $\pi_1 : M \times M \to M$ and $\pi_2 : M \times M \to M$ be the projections onto the first and the second components, respectively. Let Γ be

the set of all joint measures Π on the product manifold $M \times M$ with marginals μ and ν: $\pi_{1*}\Pi = \mu$ and $\pi_{2*}\Pi = \nu$.

Problem 3.2. Find minimizers for

$$C(\mu, \nu) := \inf_{\Pi \in \Gamma} \int_{M \times M} c(x, y) \, d\Pi(x, y).$$

Remark 3.3. If ϕ is an optimal map in the problem in (3.1), then $(id \times \phi)_* \mu$ is a joint measure in the set Γ. Therefore, Problem 3.2 is a relaxation of the problem in (3.1).

Before we proceed into the existence proof of an optimal map, let us look at the following dual problem of Kantorovich. See [23] for the history and importance of this dual problem for optimal transportation.

Let c be a cost function and let f be a function on the manifold M. The c_1-transform of the function f is the function f^{c_1} given by

$$f^{c_1}(y) := \inf_{x \in M} [c(x, y) - f(x)].$$

Similarly, the c_2-transform of the function f is defined by

$$f^{c_2}(x) := \inf_{y \in M} [c(x, y) - f(y)].$$

The function f is a c-concave function if it satisfies $f^{c_1 c_2} = f$. Let $\mathfrak{F}$ be the set of all pairs of functions (g, h) on the manifold such that $g : M \to \mathbb{R} \cup \{-\infty\}$ and $h : M \to \mathbb{R} \cup \{-\infty\}$ are in $L^1(\mu)$ and $L^1(\nu)$, respectively, and $g(x) + h(y) \le c(x, y)$ for all $(x, y) \in M \times M$. The dual problem of Kantorovich is the following maximization problem:

Problem 3.4. Find maximizers for

$$\sup_{(g,h) \in \mathfrak{F}} \int_M g \, d\mu + \int_M h \, d\nu.$$

The existence of a solution to the above problem is well known. See [23, Theorem 1.3].

Theorem 3.5. *Assume that there exist two functions c_1 and c_2 such that c_1 is μ-measurable, c_2 is ν-measurable and the cost function c satisfies $c(x, y) \le c_1(x) + c_2(y)$ for all (x, y) in $M \times M$. If c is also continuous, bounded below and $C(\mu, \nu) < \infty$, then there exists a c-concave function f such that the function f is in $L^1(\mu)$, its c_1-transform f^{c_1} is in $L^1(\nu)$ and the pair (f, f^{c_1}) achieves the supremum in Problem 3.4.*

The following theorem on the regularity of the dual pair above is also well known. Stronger results can be found in [24, Chapter 12]. We give a simple proof here for the convenience of the reader.

Theorem 3.6. *Assume that the cost $c(x, y)$ is continuous, bounded below and the measures μ and ν are compactly supported. Then the functions f and f^{c_1} are upper semicontinuous. If the function $x \mapsto c(x, y)$ is also locally Lipschitz on a set $\mathcal{U}$ and the Lipschitz constant is independent of y locally, then f can be chosen to be locally Lipschitz on $\mathcal{U}$.*

Proof. Fix $\epsilon > 0$. Since $f(x) = \inf_{x \in M}[c(x,y) - f^{c_1}(y)]$, there exists y such that $f(x) + \epsilon/2 > c(x,y) - f^{c_1}(y)$. Also, we have $f(x') + f^{c_1}(y) \leq c(x',y)$ for any x' in M. So, combining the above equations and the continuity of the cost c, we have

$$f(x') - f(x) < \epsilon$$

for any x' sufficiently close to x. Therefore, f is upper semicontinuous.

Let K be a compact set containing the support of the measures μ and ν. Let

$$g(x) = \begin{cases} f(x), & \text{if } x \in K, \\ -\infty, & \text{if } x \in M \setminus K, \end{cases} \qquad g'(x) = \begin{cases} f^{c_1}(x), & \text{if } x \in K, \\ -\infty, & \text{if } x \in M \setminus K. \end{cases}$$

Then the pair (g, g') achieves the maximum in Problem 3.4. Let $h = (g')^{c_2}$. Then the pair (h, h^{c_1}) also achieves the maximum. By definition of g', we have $h(x) = \inf_{y \in K}[c(x,y) - f^{c_1}(y)]$. By an argument the same as the proof of upper semicontinuity, for any x and x' in the compact subset K' of $\mathcal{U}$, we can find y in K such that

$$h(x') - h(x) < c(x,y) - c(x',y) + \epsilon/2.$$

By the assumption of the cost c, the above inequality becomes

$$h(x') - h(x) \leq kd(x,x') + \epsilon/2$$

for some constant $k > 0$ which is independent of x on the subset K'. By switching the roles of x and x', the result follows. $\square$

The following theorem about minimizers of the Problem 3.2 is well known. See, for instance, [23, Chapter 2], [24, Theorem 5.10].

Theorem 3.7. *In the assumptions as in Theorem* 3.5, *Problem* 3.2 *admits a minimizer. Moreover, the joint measure* Π *in the set* Γ *achieves the infimum in Problem* 3.2 *if and only if* Π *is concentrated on the set*

$$\{(x,y) \in M \times M | f(x) + f^c(y) = c(x,y)\}.$$

4. Existence and uniqueness of an optimal map

In this section, we show that Monge's problem with cost given by an optimal control cost (3) can be solved under certain regularity assumptions. Let $H : T^*M \to \mathbb{R}$ be the function defined by

$$H(p_x) = \max_{u \in U} \left(p_x(F(x,u)) - L(x,u) \right).$$

If H is well defined and C^2, then we denote its Hamiltonian vector field by $\overrightarrow{H}$ and let $e^{t\overrightarrow{H}}$ be its flow. Let f be the function defined in Theorem 3.5 which is Lipschitz for μ-almost all x. Consider the map $\varphi : M \times [0,1] \to M$ defined by $\varphi(x,t) = \pi(e^{t\overrightarrow{H}}(-df_x))$.

Theorem 4.1. *The map* $x \mapsto \varphi_1(x) := \varphi(x,1)$ *is the unique (up to* μ-*measure zero) optimal map to the Problem* (3.1) *with cost* c *given by* (5) *under the following assumptions:*

(1) *The measures* μ *and* ν *are compactly supported and* μ *is absolutely continuous with respect to the Lebesgue measure.*

(2) c *is bounded below and* $c(x,y)$ *is also locally Lipschitz in the* x *variable and the Lipschitz constant is independent of* y *locally.*

(3) *The cost c is complete; i.e. given any pairs of points (x_0, x_1) in the manifold M, there exists an admissible pair $(x(\cdot), u(\cdot))$ such that the pair achieves the infimum in (5), where $u(\cdot)$ is locally bounded measurable, $x(0) = x_0$ and $x(1) = x_1$.*

(4) *The Hamiltonian function H defined in (16) is well defined and C^2.*

(5) *The Hamiltonian vector field $\overrightarrow{H}$ is complete; i.e. global flow exists.*

The rest of this section is devoted to the proof of Theorem 4.1. Let $\widetilde{\mathfrak{C}}_y$ be the set of all admissible pairs such that the corresponding admissible paths $x(\cdot)$ start from the point y $(x(0) = y)$ and satisfies the following control system:

$$\text{(12)} \qquad\qquad\qquad \dot{x}(t) = -F(x(t), u(t)).$$

Let $\widetilde{\mathfrak{C}}_y^x$ be the set of all those pairs in $\widetilde{\mathfrak{C}}_y$ such that the corresponding admissible paths $x(\cdot)$ end at the point x: $x(1) = x$.

First, we have the following observation.

Proposition 4.2. *Let x be a point which achieves the infimum $f^{c_1}(y) = \inf\limits_{x \in M} (c(x, y) - f(x))$ and let $(\widetilde{x}, \widetilde{u})$ be an admissible pair in $\widetilde{\mathfrak{C}}_y^x$ such that the corresponding admissible path $\widetilde{x}$ minimizes the cost given by*

$$c(x, y) = \inf_{(x(\cdot), u(\cdot)) \in \widetilde{\mathfrak{C}}_y^x} \int_0^1 L(x(t), u(t))\, dt.$$

Then $(\widetilde{x}(\cdot), \widetilde{u}(\cdot))$ achieves the following infimum:

$$\text{(13)} \qquad\qquad f^{c_1}(y) = \inf_{(x(\cdot), u(\cdot)) \in \widetilde{\mathfrak{C}}_y} \int_0^1 L(x(s), u(s))\, ds - f(x(1)).$$

If $\hat{x}(t) = \widetilde{x}(1 - t)$, then $\hat{x}$ achieves the following infimum:

$$\text{(14)} \qquad\qquad f^{c_1}(y) = \inf_{(x(\cdot), u(\cdot)) \in \mathfrak{C}^y} \int_0^1 L(x(s), u(s))\, ds - f(x(0)),$$

where $\mathfrak{C}^y$ denotes the set of all admissible pairs $(x(\cdot), u(\cdot))$ satisfying the following control system:

$$\dot{x}(t) = F(x(t), u(t)), \quad x(1) = y.$$

Let $\widetilde{u}(\cdot)$ be as in the above proposition and let $\hat{u}(t) = \widetilde{u}(1-t)$. Let $H_t : T^*M \to \mathbb{R}$ be given by $H_t(p_x) = p_x(F(x, \hat{u}(t))) - L(x, \hat{u}(t))$. The following is a consequence of Theorem 2.3.

Proposition 4.3. *Let $\widetilde{x}$ be a curve that achieves the infimum in (13) and let $\hat{x}(t) = \widetilde{x}(1 - t)$. Assume that α is contained in the sub-differential of the function f at the point $\hat{x}(0)$. Then there exists a Lipschitz curve $\hat{p} : [0, 1] \to T^*M$ in the cotangent bundle such that the following are true for almost all time t in the interval $[0, 1]$:*

$$\text{(15)} \qquad \begin{cases} \pi(\hat{p}(t)) = \hat{x}(t), \\ \dot{\hat{p}}(t) = \overrightarrow{H}_t(\hat{p}(t)), \\ \hat{p}(0) = -\alpha, \\ H_t(\hat{p}(t)) = \max\limits_{u \in U} \left(\hat{p}(t)(F(\hat{x}(t), u)) - L(\hat{x}(t), u) \right). \end{cases}$$

Proof. By Theorem 2.3, there exists a curve $\widetilde{p} : [0,1] \to T^*M$ in the cotangent bundle T^*M such that

$$\begin{cases} \pi(\widetilde{p}(t)) = \widetilde{x}(t), \\ \widetilde{p}(1) = -\alpha, \\ \dot{\widetilde{p}}(t) = \overrightarrow{\widetilde{H}}_{\widetilde{u}(t)}(\widetilde{p}(t)), \\ \widetilde{H}_{\widetilde{u}(t)}(\widetilde{p}(t)) = \min_{u \in U}\left(-\widetilde{p}(t)(F(\widetilde{x}(t), \widetilde{u}(t))) + L(\widetilde{x}(t), \widetilde{u}(t)) \right), \end{cases}$$

where $\widetilde{H}_{\widetilde{u}}(p) = \min_{u \in U}(-\widetilde{p}(F(\widetilde{x}, \widetilde{u}(t))) + L(\widetilde{x}, \widetilde{u}(t)))$.

Let $\hat{p}(t) = \widetilde{p}(1-t)$ and $\hat{u}(t) = \widetilde{u}(1-t)$. Then the equations above become

$$\begin{cases} \pi(\hat{p}(t)) = \hat{x}(t), \\ \hat{p}(0) = -\alpha, \\ \dot{\hat{p}}(t) = \overrightarrow{H}_{\hat{u}(t)}(\hat{p}(t)), \\ H_{\hat{u}(t)}(\hat{p}(t)) = \max_{u \in U}\left(\hat{p}(t)(F(\hat{x}(t), \hat{u}(t))) - L(\hat{x}(t), \hat{u}(t)) \right). \end{cases}$$

$\square$

Assume that the Hamiltonian function $H : T^*M \to \mathbb{R}$ defined by

$$(16) \qquad\qquad H(p_x) = \max_{u \in U}\left(p_x(F(x,u)) - L(x,u) \right)$$

is well defined and C^2. Let $\overrightarrow{H}$ be the Hamiltonian vector field of the function H and let $e^{t\overrightarrow{H}}$ be its flow. The function f defined in Theorem 3.5 is Lipschitz and so it is differentiable almost everywhere by the Rademacher Theorem. Moreover, the map $df : M \to T^*M$ is measurable and locally bounded. So, if we let $\varphi : M \times [0,1] \to M$ be the map defined by $\varphi(x,t) = \pi(e^{t\overrightarrow{H}}(-df_x))$, then the map φ is a Borel map.

Proposition 4.4. *Under the assumptions of Theorem 4.1, the following is true for μ-almost all x: Given a point x in the support of μ, there exists a unique point y such that*

$$f(x) + f^{c_1}(y) = c(x,y).$$

Moreover, the points x and y are related by $y = \varphi(x,1)$.

Proof. We first claim that the infimum $f(x) = \inf_{y \in M}[c(x,y) - f^{c_1}(y)]$ is achieved for μ almost all x. Indeed, by assumption, we have $f(x) + f^{c_1}(y) \leq c(x,y)$ for all (x,y) in $M \times M$. Also, let Π be the measure defined in Theorem 3.7. Then $f(x) + f^{c_1}(y) = c(x,y)$ for Π-almost everywhere. Since the first marginal of the measure Π is μ, the following is true for μ almost all x: Given a point x in the manifold M, there exists y in M such that $f(x) + f^{c_1}(y) = c(x,y)$. This proves the claim.

Fix a point x for which the infimum $\inf_{y \in M}[c(x,y) - f^{c_1}(y)]$ is achieved and let y be the point which achieves the infimum. By the proof of the above claim, x achieves the infimum $f^{c_1}(y) = \inf_{x \in M}[c(x,y) - f(x)]$. Therefore, by completeness of the cost c and Proposition 4.2, there exists an admissible path $\hat{x}$ such that $\hat{x}(0) = x$, $\hat{x}(1) = y$ and $\hat{x}$ achieves the infimum (14).

Since f is Lipschitz on a bounded open set $\mathcal{U}$ containing the support of μ and ν, it is almost everywhere differentiable on $\mathcal{U}$ by the Rademacher Theorem. Since μ is absolutely continuous with respect to the Lebesgue measure, f is also differentiable

μ-almost everywhere. By Theorem 4.3, for μ-almost all x, there exists a curve $\hat{p} : [0, 1] \to T^*M$ in the cotangent bundle T^*M such that

$$\begin{cases} \dot{\hat{p}}(t) = \overrightarrow{H}_t(\hat{p}(t)), \\ \hat{p}(0) = -df_x, \\ \pi(\hat{p}(t)) = \hat{x}(t), \\ H_t(\hat{p}(t)) = \max_{u \in U} \left(\hat{p}(t)(F(\hat{x}(t), u)) - L(\hat{x}(t), u) \right), \end{cases}$$

where H_t is the function on the cotangent bundle T^*M given by $H_t(p_x) = p_x F(x, u(t)) - L(x, u(t))$.

By the definition of H, we have $H(\hat{p}(t)) = H_t(\hat{p}(t))$. But, we also have $H(p) \geq H_t(p)$ for all $p \in T^*M$. Since both H and H_t are C^2, we have $dH(\hat{p}(t)) = dH_t(\hat{p}(t))$. Hence, $\overrightarrow{H}_t(\hat{p}(t)) = \overrightarrow{H}(\hat{p}(t))$ for almost all t. The result follows from the uniqueness of the solution to the ODE. $\qquad \square$

The rest of the arguments for the existence and uniqueness of an optimal map follow from Theorem 3.7.

Proof of Theorem 4.1. As mentioned above, Problem 3.2 is a relaxation of Problem 3.1. We can recover the latter from the former by restricting the minimization to joint measures of the form $(id \times \phi)_*\mu$, where ϕ is any Borel map pushing forward μ to ν. Therefore, the results follow from Theorem 3.7 and Proposition 4.4. $\qquad \square$

5. Regularity of control costs

In Theorem 4.1, we prove the existence and uniqueness of optimal maps under certain regularity conditions on the cost. Most of the conditions in the theorem are easy to verify except conditions (2) and (3). In this section, we will give simple conditions which guarantee this regularity. This includes the completeness and the Lipschitz regularity of the cost. First, we recall some basic notions in the geometry of optimal control problems; see [2, 14] and the references therein for the details.

Fix a point x_0 on the manifold M and assume that the control set U is $\mathbb{R}^k$. In this section, we change our previous convention on admissible controls. From now on, admissible controls are mappings in $L^1([0, 1], U)$ rather than $L^\infty([0, 1], U)$. Denote by $\mathfrak{C}_{x_0}$ the set of all admissible pairs $(x(\cdot), u(\cdot))$ such that the corresponding admissible paths $x(\cdot)$ start at x_0. Moreover, we assume that the control system is of the following form:

$$(17) \qquad \dot{x}(t) = X_0(x(t)) + \sum_{i=1}^{k} u_i(t) X_i(x(t)),$$

where $u(t) = (u_1(t), ..., u_k(t))$ and $X_0, X_1, ..., X_k$ are fixed smooth vector fields on the manifold M. The Cauchy problem for system (17) is well posed for any locally integrable vector function $u(\cdot)$. We assume, throughout this section, that system (17) is complete, i.e. that all solutions of the system are defined on the whole semi-axis $[0, +\infty)$. This completeness assumption is automatically satisfied if one of the following is true: (i) if M is a compact manifold, (ii) if M is a Lie group and the fields X_i are left-invariant, or (iii) if M is a closed submanifold of the Euclidean space and $|X_i(x)| \leq c(1 + |x|)$, $i = 0, 1, \ldots, k$.

Define the endpoint map $End_{x_0} : L^1([0, 1], \mathbb{R}^k) \to M$ by

$$End_{x_0}(u(\cdot)) = x(1),$$

where $(x(\cdot), u(\cdot))$ is the admissible pair corresponding to the control system (17) with initial condition $x(0) = x_0$. It is known that the map End_{x_0} is a smooth mapping. The critical points of the map End_x are called *singular controls*. Admissible paths corresponding to singular controls are called *singular paths*.

We also need the Hessian of the mapping End_{x_0} at the critical point. (See [4] for the details.) Let E be a Banach space which is an everywhere dense subspace of a Hilbert space H. Consider a mapping $\Phi : E \to \mathbb{R}^n$ such that the restriction of this map $\Phi\big|_W$ to any finite-dimensional subspace W of the Banach space E is C^2. Moreover, we assume that the first and second derivatives of all the restrictions $\Phi\big|_W$ are continuous in the Hilbert space topology on the bounded subsets of E. In other words,

$$\Phi(v + w) - \Phi(v) = D_v\Phi(w) + \frac{1}{2}D_v^2\Phi(w) + o(|w|^2),\ w \in W,$$

where $D_v\Phi$ is a linear map and $D_v^2\Phi$ is a quadratic mapping from E to $\mathbb{R}^n$. Moreover, $\Phi(v)$, $D_v\Phi\big|_W$, and $D_v^2\Phi\big|_W$ depend continuously on v in the topology of H while v is contained in a ball of E.

The Hessian $\mathrm{Hess}_v\Phi : \ker D_v\Phi \to \mathrm{coker}\,D_v\Phi$ of the function Φ is the restriction of $D_v^2\Phi$ to the kernel of $D_v\Phi$ with values considered up to the image of $D_v\Phi$. The Hessian is the part of $D_v^2\Phi$ which survives smooth changes of the variables in E and $\mathbb{R}^n$.

Let p be a covector in the dual space $\mathbb{R}^{n*}$ such that $p(D_v\Phi) = 0$. Then $p(\mathrm{Hess}_v\Phi)$ is a well-defined real quadratic form on $\ker D_v\Phi$. We denote the Morse index of this quadratic form by $\mathrm{ind}(p\mathrm{Hess}_v\Phi)$. Recall that the Morse index of a quadratic form is the supremum of the dimensions of the subspaces where the form is negative definite.

Definition 5.1. A critical point v of the map Φ is called *sharp* if there exists a covector $p \neq 0$ such that $p(D_v\Phi) = 0$ and $\mathrm{ind}(p\mathrm{Hess}_v\Phi) < +\infty$.

Needless to say, the spaces E, H and $\mathbb{R}^n$ can be substituted by smooth manifolds (Banach, Hilbert and n-dimensional) in all this terminology.

Going back to the control system (17), let $(x(\cdot), u(\cdot))$ be an admissible pair for this system. We say that the control $u(\cdot)$ and the path $x(\cdot)$ are sharp if $u(\cdot)$ is a sharp critical point of the endpoint map $End_{x(0)}$.

One necessary condition for controls and paths to be sharp is the, so-called, Goh condition.

Proposition 5.2 (The Goh condition). *If $p(Hess_{u(\cdot)}(End_{x(0)})) < +\infty$, then*

$$p(t)(X_i(x(t))) = p(t)([X_i, X_j](x(t))) = 0 \quad i, j = 1, \ldots, k,\ 0 \leq t \leq 1,$$

where $p(t) = P_{t,1}^*p$ and $P_{t,\tau}$ is the local flow of the control system (20) with control equal to $u(\cdot)$ and time parameter τ.

See [4, Proposition 20.3, 20.4], [6] and the references therein for the proof and other effective necessary and sufficient conditions of the sharpness.

Now consider the optimal control problem

$$(18) \qquad c(x, y) = \inf_{(x(\cdot), u(\cdot)) \in \mathfrak{C}_x^y} \int_0^1 L(x(t), u(t))\, dt,$$

where the infimum ranges over all admissible pairs $(x(t), u(t))$ corresponding to the control system (17) with initial condition $x(0) = x$ and final condition $x(1) = y$.

Let $H : T^*M \to \mathbb{R}$ be the Hamiltonian function defined in (16). For simplicity, we assume that the Hamiltonian is C^2. A minimizer $x(\cdot)$ of the above problem is called *normal* if there exists a curve $p : [0, 1] \to T^*M$ in the cotangent bundle T^*M such that $\pi(p(t)) = x(t)$ and $p(\cdot)$ is a trajectory of the Hamiltonian vector field $\overrightarrow{H}$. Singular minimizers are also called *abnormal*. According to this, not so perfect, terminology a minimizer can simultaneously be normal and abnormal. A minimizer which is not normal is called *strictly abnormal*. Under some regularity and growth conditions on the Lagrangian L, all strictly abnormal minimizers are sharp. (See Theorem 5.3.)

The following theorem gives simple sufficient conditions for completeness of the cost function defined in (18). It is a combination of the well-known existence result (see [21]) and necessary optimality conditions (see [4]).

Theorem 5.3 (Completeness of costs). *Let L be a Lagrangian function which satisfies the following:*

    (1) *L is bounded below and there exists a constant $K > 0$ such that the ratio $\frac{|u|}{L(x,u)+K}$ tends to 0 as $|u| \to \infty$ uniformly on any compact subset of M;*

    (2) *for any compact $C \subset M$ there exist constants $a, b > 0$ such that $|\frac{\partial L}{\partial x}(x, u)| \leq a(L(x, u) + |u|) + b, \ \forall x \in C, \ u \in \mathbb{R}^k$;*

    (3) *the function $u \mapsto L(x, u)$ is a strongly convex function for all $x \in M$.*

Then, for each pair of points (x, y) in the manifold M which satisfy $c(x, y) < +\infty$, there exists an admissible pair $(x(\cdot), u(\cdot))$ achieving the infimum in (18). Moreover, the minimizer $x(\cdot)$ is either a normal or a sharp path.

Remark 5.4. Under the assumptions of the theorem, strictly abnormal minimizers are sharp.

Remark 5.5. Theorem 5.3 leads to many examples that satisfy condition (3) in Theorem 4.1. In particular, this applies to the case where the control set is $U = \mathbb{R}^k$ and the Lagrangian is $L(x, u) = \sum_{i=1}^{k} u_i^2$.

Remark 5.6. It was shown that the optimal controls in Theorem 5.3 that are normal are locally bounded. (See [21].) This allows us to restrict the endpoint map to $L^\infty([0, 1], U)$ in Theorem 5.7 below.

Next, we proceed to the main result of this section, which is concerned with the Lipschitz regularity of the cost function. This takes care of condition (2) in Theorem 4.1.

Theorem 5.7 (Lipschitz regularity). *Assume that the system (7) does not admit sharp controls and the Lagrangian L satisfies the conditions of Theorem 5.3. Then the set $D = \{(x, End_x(u(\cdot)))|x \in M, u \in L^\infty([0, 1], \mathbb{R}^k)\}$ is open in the product $M \times M$. Moreover, the function $(x, y) \mapsto c(x, y)$ is locally Lipschitz on the set D, where the cost c is given by (5).*

Remark 5.8. In the case where the endpoint map is a submersion, there is no singular control. Therefore, Theorem 5.7 is applicable. In particular, this theorem, together with Theorems 4.1 and 5.3, can be used to treat the cases considered in [8, 18, 7]. In section 5, we will consider a class of examples where the endpoint map is not necessarily a submersion, but Theorem 5.7 is still applicable.

The rest of the section is devoted to the proof of Theorem 5.7.

Definition 5.9. Given v in the Banach space E, we write $\text{Ind}_v\Phi \geq m$ if

$$\text{ind}(p\text{Hess}_v\Phi) - \text{codim}\,\text{im}D_v\Phi \geq m$$

for any p in $\mathbb{R}^{n*} \setminus \{0\}$ such that $p(D_v\Phi) = 0$.

It is easy to see that $\{v \in E : \text{Ind}_v\Phi \geq m\}$ is an open subset of E for any integer m. Let $\mathbf{B}_v(\varepsilon)$, $B_x(\varepsilon)$ be the balls of radius ε in E and $\mathbb{R}^n$ centered at v and x respectively. The following is a qualitative version of the openness of a mapping Φ and any mapping C^0 close to it.

Definition 5.10. We say that the map $\Phi : E \to \mathbb{R}^n$ is r-solid at the point v of the Banach space E if for some constant $c > 0$ and any sufficiently small $\varepsilon > 0$, the following inclusion holds for any map $\tilde{\Phi} : \mathbf{B}_v(\varepsilon) \to \mathbb{R}^n$ which is C^0 close to the map Φ:

$$B_{\tilde{\Phi}(v)}(c\varepsilon^r) \subset \tilde{\Phi}(\mathbf{B}_v(\varepsilon)).$$

As usual, to be C^0 close to Φ means that there exists $\delta > 0$ such that $\sup\limits_{w\in\mathbf{B}_v(\varepsilon)} |\tilde{\Phi}(w) - \Phi(w)| \leq \delta$.

The Implicit Function Theorem, together with the Brouwer Fixed Point Theorem, implies that Φ is 1-solid at any regular point.

Lemma 5.11. *If* $\text{Ind}_v\Phi \geq 0$, *then* Φ *is 2-solid at* v.

Proof. This lemma is a refinement of Theorem 20.3 from [4]. It can be proved by a slight modification of the proof of the cited theorem. Obviously, we may assume that v is a critical point of Φ. Moreover, by an argument in the proof of the theorem cited above, we may assume that E is a finite-dimensional space, $v = 0$ and $\Phi(0) = 0$.

Let $E = E_1 \oplus E_2$, where $E_2 = \ker D_0\Phi$. For any $w \in E$ we write $v = v_1 + v_2$, where $v_1 \in E_1$, $v_2 \in E_2$. Now consider the mapping

$$Q : v \mapsto D_0\Phi v_1 + \frac{1}{2}D_0^2\Phi(v_2), \ v \in E.$$

It is shown in the proof of [4, Theorem 20.3] that $Q^{-1}(0)$ contains regular points in any neighborhood of 0. Hence, there exists $c > 0$ such that the image of any continuous mapping $\tilde{Q} : \mathbf{B}_0(1) \to \mathbb{R}^n$ sufficiently close (in the C^0-norm) to $Q\big|_{\mathbf{B}_0(1)}$ contains the ball $B_0(c)$. Now, if $\tilde{\Phi}$ is C^0 close to Φ, we set $\Phi_\varepsilon(v) = \frac{1}{\varepsilon^2}\Phi(\varepsilon^2 v_1 + \varepsilon v_2)$ and $\tilde{\Phi}_\varepsilon(v) = \frac{1}{\varepsilon^2}\tilde{\Phi}(\varepsilon^2 v_1 + \varepsilon v_2)$. Then, by differentiating Φ_ε with respect to ε, it is easy to see that $\Phi_\varepsilon(v) = Q(v) + o(1)$ as $\varepsilon \to 0$. This shows that Φ_ε and hence $\tilde{\Phi}_\varepsilon$ are C^0 close to Q for all sufficiently small ε. Therefore, $\tilde{\Phi}_\varepsilon\big|_{\mathbf{B}_0(1)}$ contains the ball $B_0(c)$. This gives

$$B_0(c) \subset \tilde{\Phi}_\varepsilon(B_0(1)) \subset \frac{1}{\varepsilon^2}\tilde{\Phi}(B_0(\varepsilon)),$$

and the result follows. $\qquad\square$

Remark 5.12. The minimization problem (18) can be rephrased into a constrained minimization problem in an infinite-dimensional space. For simplicity, consider the case where $M = \mathbb{R}^n$. Let $(x(\cdot), u(\cdot))$ be an admissible pair of the control system (17) and let $\varphi : \mathbb{R}^n \times L^\infty([0, 1], \mathbb{R}^k) \to \mathbb{R}$ be the function defined by

$$\varphi(x, u(\cdot)) = \int_0^1 L(x(t), u(t))dt.$$

Let $\Phi : \mathbb{R}^n \times L^\infty([0,1], \mathbb{R}^k) \to \mathbb{R}^n \times \mathbb{R}^n$ be the map

$$\Phi(x, u(\cdot)) = (x, End_x(u(\cdot))).$$

Finding the minimum in (18) is now equivalent to minimizing the function φ on the set $\Phi^{-1}(x, y)$.

Due to the above remark, we can consider the following general setting. Consider a function $\varphi : E \to \mathbb{R}$ on the Banach space E such that $\varphi|_W$ is a C^2-mapping for any finite-dimensional subspace W of E. Recall that the Hilbert space H contains E as a dense subset. Assume that the function φ as well as the first and second derivatives of the restrictions $\varphi|_W$ are continuous on the bounded subsets of E in the topology of H. Also, recall that the map $\Phi : E \to \mathbb{R}^n$ is C^2 when restricted to any finite-dimensional subspace of E. Assume that K is a bounded subset of E that is compact in the topology of H and satisfies the following property:

$$\varphi(v) = \min\{\varphi(w) | w \in E, \ \Phi(w) = \Phi(v)\}$$

for any v in the set K.

We define a function μ on $\Phi(K)$ by the formula $\mu(\Phi(v)) = \varphi(v)$ for any v in K. In the case discussed in Remark 5.12, K is the set of all minimizers and the function μ is the cost function.

Lemma 5.13. *If* $\mathrm{Ind}_v \Phi \geq 2$ *for any* $v \in K$, *then* μ *is locally Lipschitz.*

Proof. Given v in the set K, there exists a finite-dimensional subspace W of the Banach space E such that $\mathrm{Ind}_v \left(\Phi|_W\right) \geq 2$. Then $\mathrm{Ind}_v \left(\Phi|_{W \cap \ker D_v \varphi}\right) \geq 0$. Hence $\Phi|_{W \cap \ker D_v \varphi}$ is 2-solid at v and

$$\Phi\left(\mathbf{B}_v(\varepsilon) \cap W \cap \ker D_v \varphi\right) \supset B_{\Phi(v)}(c\varepsilon^2)$$

for some c and any sufficiently small ε.

Let $x = \Phi(v)$ and $|x-y| = c\varepsilon^2$. Then $y = \Phi(w)$ for some $w \in \mathbf{B}_v(\varepsilon) \cap W \cap \ker D_v \varphi$. We have

$$\mu(y) - \mu(x) \leq \varphi(w) - \mu(x) = \varphi(w) - \varphi(v) \leq c'|w - v|^2 \leq c'\varepsilon^2.$$

Here, we use the fact that w is in $\ker D_v \varphi$ for the second to last inequality and that w is in $\mathbf{B_v}(\varepsilon)$ for the last inequality. Moreover, the compactness of K allows us to choose c, c' and the bound for ε for all $v \in K$. In particular, we can exchange x and y in the last inequality. Hence $|\mu(y) - \mu(x)| \leq \frac{c'}{c}|y - x|$. $\square$

Proof of Theorem 5.7. We describe the proof only in the case $M = \mathbb{R}^n$ in order to simplify the language. Generalization to any manifold is straightforward. We set

$$E = \mathbb{R}^n \times L^\infty([0,T], \mathbb{R}^k), \ H = \mathbb{R}^n \times L^2([0,T], \mathbb{R}^k),$$

$$\Phi(x, u(\cdot)) = (x, End_x(u(\cdot))), \ \varphi(x, u(\cdot)) = \int_0^1 L(x(t), u(t)) \, dt,$$

and apply the above results.

First of all, $\mathrm{Ind}_{(x, u(\cdot))} \Phi = \mathrm{Ind}_{u(\cdot)} End_x = +\infty$ for all $(x, u(\cdot))$ since our system does not admit sharp controls. Lemma 5.11 implies that Φ is 2-solid and $\mathcal{D} = \Phi(E)$ is open.

Now let $\mathcal{B}$ be a ball in E equipped with the weak topology of H. The *endpoint mapping* Φ is continuous as a mapping from $\mathcal{B}$ to $\mathbb{R}^{2n}$. The strict convexity of L implies that there is some constant $c > 0$ such that

$$\varphi(x_n, u_n(\cdot)) - \varphi(x, u(\cdot)) \geq c\|u_n(\cdot) - u(\cdot)\|^2_{L^2} + o(1)$$

as $x_n \to x$, $u_n(\cdot) \rightharpoonup u(\cdot)$, and $(x_n, u_n(\cdot)) \in \mathcal{B}$. Therefore, $\lim_{n\to\infty} \varphi(x_n, u_n(\cdot)) \geq \varphi(x, u(\cdot))$ and $\lim_{n\to\infty} \varphi(x_n, u_n(\cdot)) = \varphi(x, u(\cdot))$ if and only if $(x_n, u_n(\cdot))$ converges to $(x, u(\cdot))$ in the strong topology of H.

Assume that $\varphi(x_n, u_n(\cdot)) = \mu(\Phi(x_n, u_n(\cdot)))$ for all n. Inequality $\varphi(x, u(\cdot)) < \lim_{n\to\infty} \varphi(x_n, u_n(\cdot))$ would imply that

$$\mu(\Phi(x, u(\cdot))) < \lim_{n\to\infty} \mu(\Phi(x_n, u_n(\cdot))).$$

On the other hand, the openness of the map Φ implies that the map μ is uppersemicontinuous. Together with the continuity of Φ, we have the following inequality:

$$\mu(\Phi(x, u(\cdot))) \geq \lim_{n\to\infty} \mu(\Phi(x_n, u_n(\cdot))).$$

Hence $\lim_{n\to\infty} \varphi(x_n, u_n(\cdot)) = \varphi(x, u(\cdot))$ and $(x_n, u_n(\cdot))$ converges to $(x, u(\cdot))$ in the strong topology of H.

Let C be a compact subset of $\mathcal{D}$ and let

$$K = \{(x, u(\cdot)) \in E : \Phi(x, u(\cdot)) \in C, \; \varphi(x, u(\cdot)) = \mu(\Phi(x, u(\cdot)))\}.$$

Then K is contained in some ball $\mathcal{B}$. Recall that $\mathcal{B}$, equipped with the weak topology, is compact. Now the calculations of the previous two paragraphs imply compactness of K in the strong topology of H. Finally, we derive the Lipschitz property of $\mu|_C$ from Lemma 5.13. $\qquad\square$

6. Applications: Mass transportation on subriemannian manifolds

In this section, we will apply the results in the previous sections to some subriemannian manifolds. First, let us recall some basic definitions.

Let Δ and Δ' be two (possibly singular) distributions on a manifold M. Define the distribution $[\Delta, \Delta']$ by

$$[\Delta, \Delta']_x = \text{span}\{[v, w](x)|v \text{ is a section of } \Delta, w \text{ is a section of } \Delta'\}.$$

Define inductively the following distributions: $\Delta^2 = \Delta + [\Delta, \Delta]$ and $\Delta^k = \Delta^{k-1} + [\Delta, \Delta^{k-1}]$. A distribution Δ is called k-generating if $\Delta^k = TM$ and the smallest such k is called the degree of nonholonomy. Also, the distribution is called bracket generating if it is k-generating for some k.

If Δ is a bracket generating distribution, then it defines a flag of distributions by

$$\Delta \subset \Delta^2 \subset ... \subset TM.$$

The growth vector of the distribution Δ at the point x is defined by $(\dim \Delta_x, \dim \Delta^2_x, ..., \dim T_x M)$. The distribution Δ is called regular if the growth vector is the same for all x. Let $x(\cdot) : [a, b] \to M$ be an admissible curve, that is, a Lipschitz curve almost everywhere tangent to Δ. The following classical result on bracket generating distributions is the starting point of subriemannian geometry.

Theorem 6.1 (Chow and Rashevskii)**.** *Given any two points x and y on a connected manifold M with a bracket generating distribution, there exists an admissible curve joining the two points.*

Using the Chow-Rashevskii Theorem, we can define the subriemannian distance d. Let $\langle,\rangle$ be a fibre inner product on the distribution Δ, called a subriemannian metric. The length of an admissible curve $x(\cdot)$ is defined in the usual way: length$(x(\cdot)) = \int_a^b \sqrt{\langle \dot{x}(t), \dot{x}(t) \rangle} \, dt$. The subriemannian distance $d(x,y)$ between two points x and y is defined by the infimum of the lengths of all admissible curves joining x and y. There is a quantitative version of the Chow-Rashevskii Theorem, called the Ball-Box Theorem, which gives Hölder continuity of the subriemannian distance. See [20] for the details.

Corollary 6.2. *Let d_S be the metric of a complete subriemannian space with a distribution Δ. The function d_S^2 is locally Lipschitz if and only if the distribution is 2-generating.*

Proof. The systems with 2-generating distributions do not admit sharp paths because these systems are not compatible with the Goh condition. So d_S^2 is locally Lipschitz by Theorem 5.7. Conversely, if the degree of nonholonomy of the distribution is greater than 2, then it follows from the Ball-Box Theorem [20, Theorem 2.10] that the function d_S^2 is Lipschitz. Indeed, let us fix a point x in the manifold M. If d_S^2 is locally Lipschitz, then $d_S^2(x,y) \leq c|x-y|$ for some constant c and for all y in a neighborhood U of x. On the other hand, by the Ball-Box Theorem, there exists a point z in U whose subriemannian distance d_S from the point x is ε and its Euclidean distance from x satisfies $|x-z| < C\varepsilon^k = Cd_S^k(x,z)$ for some constant C and for all sufficiently small ε. Here, $k > 2$ is the degree of nonholonomy of the distribution. This gives a contradiction and so d_S^2 is not Lipschitz. $\square$

Combining Corollary 6.2 with Theorem 4.1, we prove the existence and uniqueness of an optimal map for a subriemannian manifold with a 2-generating distribution.

Theorem 6.3. *Let M be a complete subriemannian manifold defined by a 2-generating distribution. Then there exists a unique (up to μ-measure zero) optimal map to Monge's problem with the cost c given by $c = d_S^2$. Here d_S is the subriemannian distance of M.*

Remark 6.4. The locally Lipschitz property of the distance d off the diagonal is guaranteed for a much larger class of distributions. In particular, it is proved in [3] that a generic distribution of rank > 2 does not admit nonconstant sharp trajectories. In the case of Carnot groups, the following estimates are valid: a generic n-dimensional Carnot group with a rank k distribution does not admit nonconstant sharp trajectories if $n \leq (k-1)k + 1$ and it has nonconstant sharp length minimizing trajectories if $n \geq (k-1)(\frac{k^2}{3} + \frac{5k}{6} + 1)$. Recall that a simply connected Lie group endowed with a left-invariant distribution V_1 is a Carnot group if the Lie algebra $\mathfrak{g}$ is a graded nilpotent Lie algebra such that it is Lie generated by the subspace with lowest grading (i.e. $\mathfrak{g} = V_1 \oplus V_2 \oplus ... \oplus V_k$, $[V_i, V_j] = V_{i+j}$, $V_i = 0$ if $i > k$ and the subspace V_1 Lie-generates $\mathfrak{g}$).

Clearly, if the cost is locally Lipschitz off the diagonal, then the statement of Theorem 4.1 remains valid with the extra assumption that the supports of the initial measure μ and the final measure ν are disjoint: $\text{supp}(\mu) \cap \text{supp}(\nu) = \emptyset$.

7. NORMAL MINIMIZERS AND PROPERTY OF OPTIMAL MAP
WITH CONTINUOUS OPTIMAL CONTROL COST

According to Theorem 5.7, it remains to study the case where sharp controls exist. In this section, we will describe a property of an optimal map when the cost is continuous. Normal minimizers will play a very important role.

We continue to study the optimal control problem (20), (21). As we already mentioned, strictly abnormal minimizers must be sharp. In addition, if $X_0 = 0$ in (17), then the optimal control cost is continuous. According to the discussion at the end of the previous section, we expect strictly abnormal minimizers mainly for generic rank 2 distributions on manifolds of dimension greater than 3 and for generic Carnot groups of large enough corank. In these situations, strictly abnormal minimizers are indeed unavoidable.

The existence of strictly abnormal minimizers for subriemannian manifolds was first done in [19]. In [22] and [16], it was shown that there are many strictly abnormal minimizers in general for subriemannian manifolds. (See, for instance, Theorem 7.1 below.) Finally, a general theory on abnormal minimizers for rank 2 distributions was developed in [5]. See [20] for a detailed account on the history and references on abnormal minimizers.

Here is a sample result in [22] which is of interest to us.

Theorem 7.1 (Liu and Sussman). *Let M be a 4-dimensional manifold with a rank 2 regular bracket-generating distribution Δ and subriemannian metric $\langle,\rangle$. Let X_1 and X_2 be two global sections of Δ such that*

(1) *X_1 and X_2 are everywhere orthonormal,*
(2) *X_1, X_2, $[X_1, X_2]$ and $[X_2, [X_1, X_2]]$ are everywhere linearly dependent,*
(3) *X_2, $[X_1, X_2]$ and $[X_2, [X_1, X_2]]$ are everywhere linearly independent.*

Then any sufficiently short segments of the integral curves of the vector field X_2 are strictly abnormal minimizers.

We call a local flow a strictly abnormal flow if the corresponding trajectories are all strictly abnormal minimizers. An interesting question is whether the time-1 map of an abnormal flow is an optimal map. The following theorem shows that this is not the case for any reasonable initial measure and continuous cost.

Theorem 7.2. *Assume that the cost c in (3.1) is continuous, bounded below and the support of the measure μ is equal to the closure of its interior. If $\varphi : M \to M$ is a continuous map such that $(id \times \varphi)_* \mu$ achieves the infimum in Problem 3.2, then x and $\varphi(x)$ are connected by a normal minimizer on a dense set of x's in the support of μ.*

Proof. By Theorem 3.5, there exists a function $f : M \to \mathbb{R} \cup \{-\infty\}$ such that f and its c_1-transform achieve the supremum in Problem 3.4. Moreover, by Theorem 3.6, the functions f and f^{c_1} are upper semicontinuous. By Theorem 3.7,

$$(19) \qquad\qquad f(x) + f^{c_1}(\varphi(x)) = c(x, \varphi(x))$$

for μ-almost all x. By the upper semicontinuity of f and f^{c_1},

$$f(x) + f^{c_1}(\varphi(x)) \geq c(x, \varphi(x)).$$

But $f(x) + f^c(y) \leq c(x, y)$ for any x, y in the manifold M. So, (19) holds for all x's in the support U of μ. Therefore, x achieves the infimum $f^{c_1}(\phi(x)) =$

$\inf_{z \in M}[c(z, \phi(x)) - f^{c_1}(z)]$ for all x in the support of μ. Moreover, using (19), it is easy to see that the function f is continuous on U. In particular, it is subdifferentiable on a dense set of U. By Proposition 4.2 and Theorem 4.3, x and $\varphi(x)$ are connected by a normal minimizer if f is sub-differentiable at x. This proves the theorem. $\qquad\qquad\square$

8. Optimal maps with abnormal minimizers

In this section, we describe an important class of control systems which admit smooth optimal maps built essentially from abnormal minimizers. Recall that abnormal minimizers are singular trajectories of the control system whose definition does not depend on the Lagrangian.

Let $\rho : M \xrightarrow{G} N$ be a smooth principal bundle where the structural group G is a connected Abelian Lie group. Let $X_1, \ldots, X_k$ be the vertical vector fields which generate the action of G. Consider the following control system:

$$(20) \qquad \dot{x}(t) = X_0(x(t)) + \sum_{i=1}^{k} u_i(t) X_i(x(t)),$$

where X_0 is a smooth vector field on M, and the re-scaled systems

$$(21) \qquad \dot{x}(t) = \varepsilon X_0(x(t)) + \sum_{i=1}^{k} u_i(t) X_i(x(t))$$

for $\varepsilon > 0$.

We define the Hamiltonian $H : T^*N \to \mathbb{R}$ by

$$(22) \qquad H(p_x) = \max\{p_x(d\rho(X_0(y))) | y \in \rho^{-1}(x)\},$$

where p_x is a covector in T^*N. We assume that the maximum above is achieved for any p in T^*N and it is finite.

A typical example is the Hopf bundle $\phi : \mathrm{SU}(2) \xrightarrow{S^1} S^2$ and a left-invariant vector field F_0. Then $H(p) = \alpha|p|$, where α is a constant and $|p|$ is the length of the covector p with respect to the standard (constant curvature) Riemannian structure on the sphere. (See [4, Section 22.2].)

Consider the following control system on N with an admissible pair $y(\cdot)$ contained in the G-bundle $\rho : M \xrightarrow{G} N$ and an admissible trajectory $x(t) = \rho(y(t))$ (see Remark 2.5):

$$(23) \qquad \dot{x}(t) = d\rho(X_0(y(t))).$$

The function H in (22) is the Hamiltonian of the time-optimal problem of the control system (23). (Recall that the time optimal problem is the following minimization problem: Fix two points x_0 and x_1 in N and minimize the time t_1 among all admissible trajectories $x(\cdot)$ of the control system (23) such that $x(t_0) = x_0$ and $x(t_1) = x_1$.)

Remark 8.1. System (23) is the reduced system associated to system (20) according to the reduction procedure described in [4, Chapter 22]. In particular, ρ transforms any admissible trajectory of system (20) to the admissible trajectory of system (23). Also, the smooth extremal trajectories of the time-optimal problem for system (23) are images under the map ρ of singular trajectories of system (20).

For any $\varepsilon > 0$ and any C^2 smooth function $f : N \to \mathbb{R}$, we introduce the map

$$\Phi_f^\varepsilon : N \to N, \quad \Phi_f^\varepsilon(x) = \pi(e^{\varepsilon\vec{H}}(d_x f)), \ x \in N,$$

where $\pi : T^*N \to N$ is the standard projection and $t \mapsto e^{t\vec{H}}$ is the Hamiltonian flow of H. Set

$$\mathcal{D} = \{p \in T^*N : H(p) > 0, \ H \text{ is of class } C^2 \text{ at } p\}.$$

Assume that Φ_f^ε pushes the measure μ' forward to another measure ν' on N. Consider some "lifts" μ and ν of the measures μ' and ν': $\rho_*\mu = \mu', \rho_*\nu = \nu'$. Let $\Psi : M \longrightarrow M$ be an optimal map pushing forward μ to ν. Then the following theorem says that Ψ is a covering of Φ_f^ε: $\rho \circ \Psi = \Phi_f^\varepsilon \circ \rho$.

Theorem 8.2. *Let K be a compact subset of N and $a \in C^2(N)$. Assume that $df|_K \subset \mathcal{D}$. Let μ and ν be Borel probability measures such that $\mathrm{supp}(\rho_*(\mu)) \subset K$. Then, for any sufficiently small $\varepsilon > 0$ and any optimal Borel map $\Psi : M \to M$ of the control system* (21) *with any Lagrangian $L : TM \to \mathbb{R}$, the following is true whenever $\rho_*(\nu) = \Phi_{f_*}^\varepsilon(\rho_*(\mu))$:*

$$\rho \circ \Psi = \Phi_f^\varepsilon \circ \rho.$$

In particular, x and $\Psi(x)$ are connected by singular trajectories.

Proof. We start from the following.

Definition 8.3. We say that a Borel map $Q : K \to N$ is ε-admissible for system (21) if there exists a Borel map $\varphi : K \to L^\infty([0, \varepsilon], G)$ such that

$$Q(x_0) = x\left(\varepsilon; \varphi(x_0)(\cdot)\right), \quad \forall x_0 \in K,$$

where $t \mapsto x\left(t; \varphi(x_0)(\cdot)\right)$ is an admissible trajectory of the reduced control system (23) with initial condition $x\left(0; \varphi(x_0)(\cdot)\right) = x_0$.

We are going to prove that Φ_f^ε is an admissible map, unique up to a $\rho_*\mu$-measure zero set, which transforms $\rho_*\mu$ into $\rho_*\nu$. This fact implies the statement of the theorem.

Inequality $H(d_x f) > 0$ implies that $d\pi(\vec{H}(d_x f))$ is transversal to the level hypersurface of f through x. Hence the map Φ_f^ε is invertible in a neighborhood of K for any sufficiently small ε. Moreover, the curve $t \mapsto \Phi_f^t(y)$, $0 \leq t \leq \varepsilon$, is a unique admissible trajectory of system (23) which starts at the hypersurface $f^{-1}(f(x))$ and arrives at the point $\Phi_f^\varepsilon(x)$ at a time moment not greater than ε. The last fact is proved by a simple adaptation of the standard sufficient optimality condition (see [4, Chapter 17]).

Now we set

$$f_\varepsilon(x) = f\left((\Phi_a^\varepsilon)^{-1}(x)\right) + \varepsilon.$$

Then f_ε is a smooth function defined in a neighborhood of K.

The optimality property of Φ_f^ε implies that

$$f_\varepsilon(Q(x)) \leq f_\varepsilon\left(\Phi_f^\varepsilon(x)\right)$$

for any ε-admissible map Q and any $x \in K$, and the inequality is strict at any point x where $Q(y) \neq \Phi_f^\varepsilon(x)$. In particular, if

$$\rho_*\mu\left(\{x \in K : Q(x) \neq \Phi_f^\varepsilon(x)\}\right) > 0,$$

then

$$\int_N f_\varepsilon \, d(Q_*(\rho_*(\mu))) = \int_N f_\varepsilon \circ Q \, d(\rho_*(\mu))$$

$$< \int_N f_\varepsilon \circ \Phi_f^\varepsilon \, d(\rho_*(\mu)) = \int_N f_\varepsilon \, d(\rho_*(\nu)).$$

Hence $Q_*(\rho_*(\mu)) \neq \rho_*(\nu)$. This proves the first part of the theorem. The fact that x and $\Psi(x)$ are connected by a singular minimizer follows from this and Remark 8.1.   □

9. Example: The Grushin plane

The Grushin plane is the subriemannian space with base space $\mathbb{R}^2$ and a singular distribution defined by the span of the vectors $\{\partial_{x_1}, x_1 \partial_{x_2}\}$ in each tangent space. In other words, the fibre of this distribution is the whole tangent space of $\mathbb{R}^2$ if $x_1 \neq 0$ and it is spanned by ∂_{x_1} otherwise. We define a subriemannian metric by declaring that the two vector fields above are orthonormal. The control system is given by

$$\dot{x}_1 = u_1, \quad \dot{x}_2 = u_2 x_1.$$

The subriemannian distance d is given by $d(x, y) = \inf_{\mathfrak{C}_x^y} \int_0^1 \sqrt{u_1^2 + u_2^2} \, dt$. In this section, we consider the optimal transport problem with cost c given by $c = d^2$.

There is no abnormal minimizer for this problem, so we consider its Hamiltonian H given by

$$H(x_1, x_2, p_1, p_2) = \frac{1}{2}(p_1^2 + x_1^2 p_2^2).$$

The corresponding Hamiltonian equation is

$$\dot{x}_1 = p_1, \quad \dot{x}_2 = x_1^2 p_2, \quad \dot{p}_1 = -x_1 p_2^2, \quad \dot{p}_2 = 0.$$

For simplicity, we consider the case $x_1(0) = 0 = x_2(0)$, and we let $p_1(0) = a$ and $p_2(0) = b$. In this case, the solutions give geodesics emanating from a point $(0, \delta)$ on the y-axis. They are parameterized by (a, b) and are given by

$$(24) \qquad x_1(t) = \frac{a}{b} \sin(bt), \quad x_2(t) = \frac{a^2}{4b^2}(2bt - \sin(2bt)) + \delta$$

if $b \neq 0$ and are given by

$$(25) \qquad\qquad\qquad\qquad x_1(t) = at, \quad x_2(t) = \delta$$

if $b = 0$. A geodesic is length minimizing if and only if $-\pi/b \leq t \leq \pi/b$.

Next, we consider the mass transport problem. Let d be the subriemannian distance of the Grushin plane and consider Problem 3.1 with cost c given by the square of the subriemannian distance d^2. We also specialize to the case where the target measure ν is equal to the delta mass supported at the origin. In this case, the optimal map is clearly given by the constant map $x \mapsto (0, 0)$. We are interested in the displacement interpolation corresponding to this optimal map. Recall that the displacement interpolation is the one-parameter family of maps ϕ_t such that ϕ_t is the optimal map with the cost c_t given by the following:

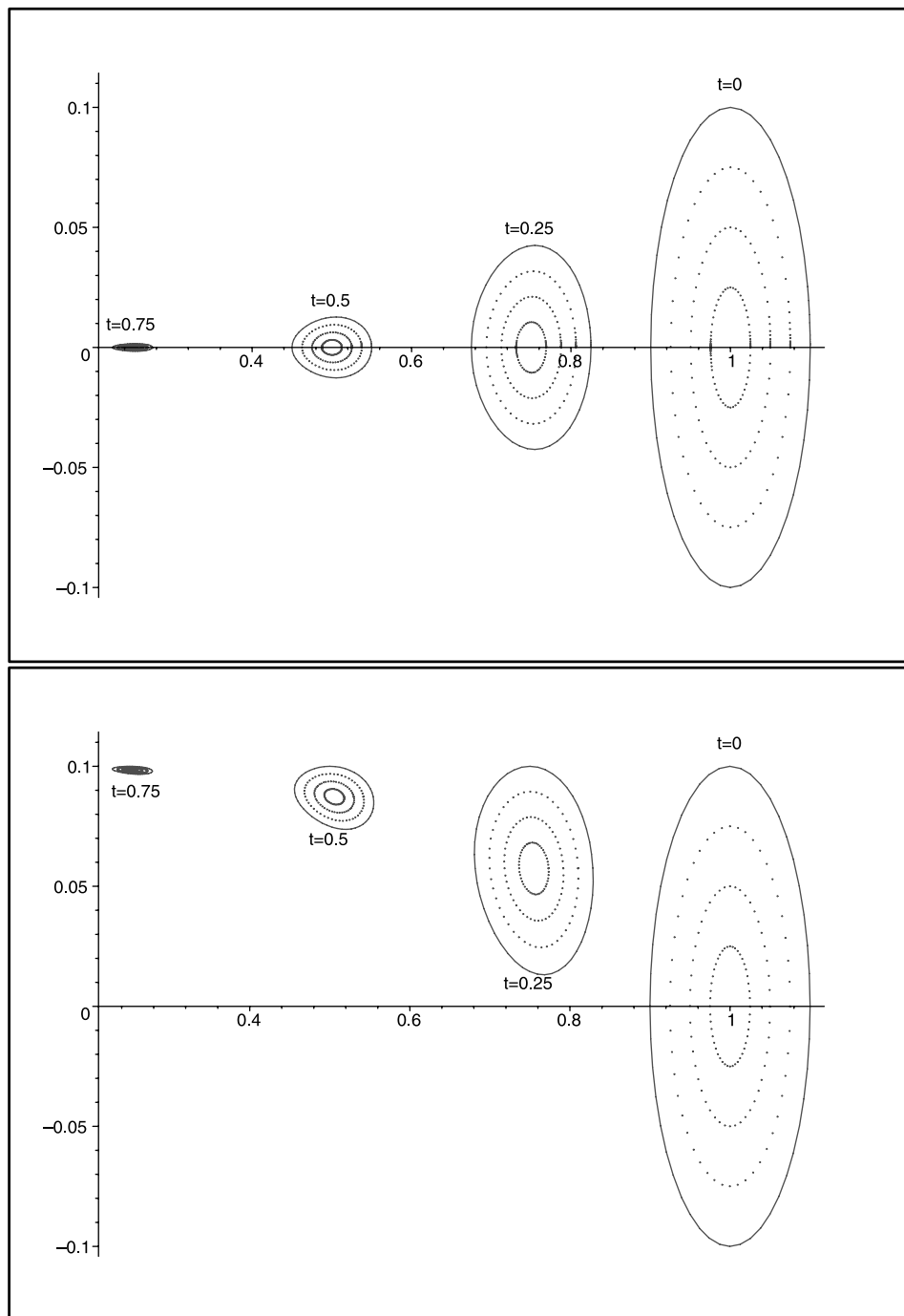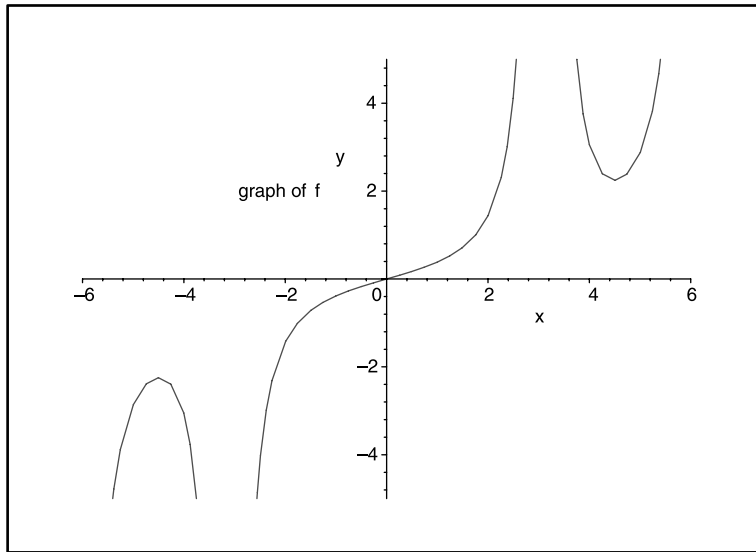$$c_t(x, y) = \inf \int_0^t L(x(s), u(s)) \, ds,$$
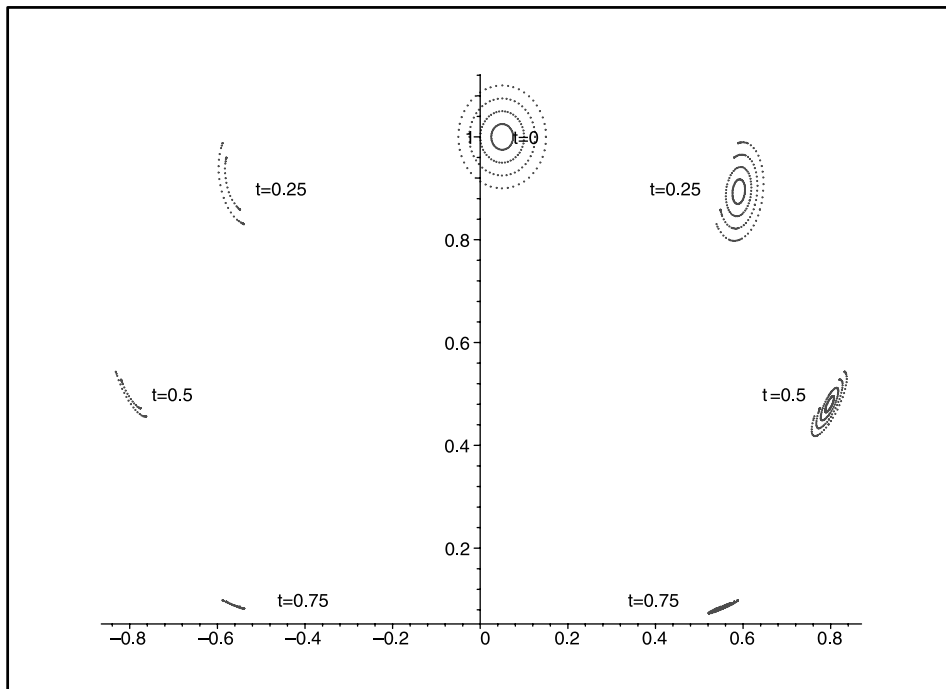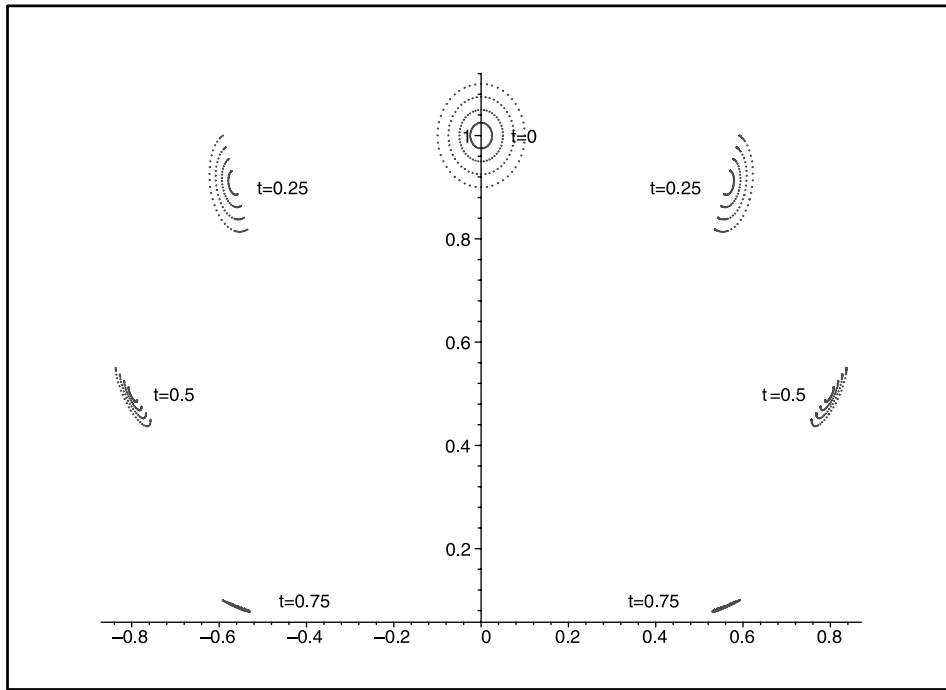
FIGURE 1. Some displacement interpolations

FIGURE 2. Graph of the function f

where the infimum ranges over all admissible pairs $(x(\cdot), u(\cdot))$ of the control system (4) with initial condition $x(0) = x$ and final condition $x(t) = y$. It is easy to see that if $\phi_1 = \pi(e^{\overrightarrow{H}}(-df))$ as in Theorem 4.1, then the displacement interpolation ϕ_t is given by $\pi(e^{t\overrightarrow{H}}(-df))$. (See Figure 1.) Moreover, the displacement interpolation is related to the Hamilton-Jacobi equation via the method of characteristics. See [7] and [10] for details.

To do this, we first evaluate the equations (24) and (25) at $t = 1$. Then we solve a and b in terms of $x_1(1)$ and $x_2(1)$. If $f : (-\pi, \pi) \to \mathbb{R}$ is the function defined by $f(b) = \frac{2b - \sin(2b)}{4\sin^2(b)}$, then f is invertible. (See Figure 2.) A computation shows that

$$a = \frac{f^{-1}\left(\frac{x_2(1) - \delta}{x_1(1)^2}\right) x_1(1)}{\sin\left(f^{-1}\left(\frac{x_2(1) - \delta}{x_1(1)^2}\right)\right)}, \quad b = f^{-1}\left(\frac{x_2(1) - \delta}{x_1(1)^2}\right).$$

Therefore, the displacement interpolation is given by

$$\varphi_t(x_1, x_2) = \left(\frac{a}{b}\sin(b(1-t)), \frac{a^2}{4b^2}(2tb - \sin(2(1-t)b) + \delta\right),$$

where $a = a(x_1, x_2)$ and $b = b(x_1, x_2)$ are given by

$$a(x_1, x_2) = \frac{f^{-1}\left(\frac{x_2 - \delta}{x_1^2}\right) x_1}{\sin\left(f^{-1}\left(\frac{x_2 - \delta}{x_1^2}\right)\right)}, \quad b(x_1, x_2) = f^{-1}\left(\frac{x_2 - \delta}{x_1^2}\right).$$

10. APPENDIX

This appendix is devoted to the proof of Theorem 2.3. The first step is to reduce the problem into a simpler one. Recall that the Bolza problem is the following

minimization problem:

$$\inf_{(x(\cdot), u(\cdot)) \in \mathfrak{C}_{x_0}} \int_0^1 L(x(s), u(s)) \, ds - f(x(1))$$

where the infimum is taken over all admissible pairs $(x(\cdot), u(\cdot))$ satisfying the control system

$$\dot{x}(s) = F(x(s), u(s))$$

and the initial condition $x(0) = x_0$.

Let $\overline{x} = (x, z)$ be a point in the product manifold $M \times \mathbb{R}$ and consider the following extended control system on it:

(26) $$\dot{\overline{x}} = \overline{F}(\overline{x}, u) := (F(x, u), L(x, u)).$$

Note that $\overline{x}(\cdot) = (x(\cdot), z(\cdot))$ satisfies this extended system and the initial condition $\overline{x}(0) = (x_0, 0)$ if and only if $x(\cdot)$ satisfies the original control system in the Bolza problem with the initial condition $x(0) = x_0$ and $z(t) = \int_0^t L(q(s), u(s)) \, ds$. Therefore, Problem 2.2 is equivalent to the following problem.

Problem 10.1. Find minimizers for

(27) $$\inf_{(\overline{x}(\cdot), u(\cdot)) \in \mathfrak{C}_{(x_0, 0)}} (z(1) - f(x(1))),$$

where the infimum is taken over all admissible pairs satisfying the extended control system (26).

Problem 10.1 is an example of the Mayer problem. Letting $g : N \to \mathbb{R}$ be a function on the manifold N, the Mayer problem is the following minimization problem:

Problem 10.2. Find minimizers for

$$\inf_{\mathfrak{C}_{\overline{x}_0}} g(\overline{x}(1)),$$

where the infimum is taken over all admissible pairs $(\overline{x}(\cdot), u(\cdot))$ satisfying the control system

$$\dot{\overline{x}} = \overline{F}(\overline{x}, u)$$

on N and the initial condition $\overline{x}(0) = \overline{x}_0$.

Note that Problem 10.1 is the Mayer problem on the manifold $N = M \times \mathbb{R}$ with function $g : M \times \mathbb{R} \to \mathbb{R}$ given by $g(x, z) = z - f(x)$. Also, if α is in the sub-differential $d^- f_x$ of f at x, then $(-\alpha, 1)$ is in the super-differential $d^+ g_{(x,z)}$ of g at (x, z).

Next, we will prove a version of the Pontryagin maximum principle for the Mayer problem and show how Theorem 2.3 follows from this. For each point u in the control set U, define the corresponding Hamiltonian function $\overline{H}_u : T^* N \to \mathbb{R}$ by

$$\overline{H}_u(p_{\overline{x}}) = p_{\overline{x}}(\overline{F}(\overline{x}, u)).$$

Theorem 10.3 (Pontryagin Maximum Principle for Mayer Problem). *Let $(\widetilde{\overline{x}}(\cdot), \widetilde{u}(\cdot))$ be an admissible pair which achieves the infimum in Problem 10.2. Assume that the function g in Problem 10.2 is super-differentiable at the point $\widetilde{\overline{x}}(1)$ and let $\overline{\alpha}$ be in the super-differential $d^+ g_{\widetilde{\overline{x}}(1)}$ of g. Then there exists a Lipschitz*

path $\widetilde{p}(\cdot) \colon [0,1] \to T^*N$ *which satisfies the following for almost all times* t *in the interval* $[0,1]$*:*

(28)
$$
\begin{cases}
\pi(\widetilde{p}(t)) = \widetilde{\overline{x}}(t), \\
\widetilde{p}(1) = \overline{\alpha}, \\
\dot{\widetilde{p}}(t) = \overrightarrow{\overline{H}}_{\widetilde{u}(t)}(\widetilde{p}(t)), \\
\overline{H}_{\widetilde{u}(t)}(\widetilde{p}(t)) = \min_{u \in U} \overline{H}_u(\widetilde{p}(t)).
\end{cases}
$$

Proof. Fix a point v in the control set and a number τ in the interval $[0,1]$. For each small positive number $\epsilon > 0$, let u_ϵ be the admissible control defined by

$$
u_\epsilon(t) = \begin{cases}
\widetilde{u}(t), & \text{if } t \notin [\tau - \epsilon, \tau]; \\
v, & \text{if } t \in [\tau - \epsilon, \tau].
\end{cases}
$$

Since the optimal control $\widetilde{u}$ is locally bounded, the new control u_ϵ defined above is also locally bounded. Let $P^\epsilon_{t_0,t_1} \colon N \to N$ be the time-dependent local flow of the following ordinary differential equation:

$$
\dot{\overline{x}}(t) = \overline{F}(\overline{x}(t), u_\epsilon(t)).
$$

Here, $P^\epsilon_{0,t}(\overline{x})$ denote the image of the point $\overline{x}$ in the manifold N under the local flow $P^\epsilon_{0,t}$ at time t. It has the property that $P^\epsilon_{t_2,t_3} \circ P^\epsilon_{t_1,t_2} = P^\epsilon_{t_1,t_3}$. Also, recall that $P^\epsilon_{t_0,t_1}$ depends smoothly on the space variable, Lipschitz with respect to the time variables.

Since $\widetilde{\overline{x}}(1) = P^0_{0,1}(\overline{x}_0)$ and the function g is minimizing at $\widetilde{\overline{x}}(1)$, the following is true for all $\epsilon > 0$:

(29)
$$
g(P^\epsilon_{0,1}(\overline{x}_0)) \geq g(P^0_{0,1}(\overline{x}_0)).
$$

Let $\overline{\alpha}$ be a point in the super-differential $d^+ g_{\widetilde{\overline{x}}(1)}$ at the point $\widetilde{\overline{x}}(1)$. Then there exists a C^1 function $\phi \colon N \to \mathbb{R}$ such that $d\phi_{\widetilde{\overline{x}}(1)} = \overline{\alpha}$ and $g - \phi$ has a local maximum at $\widetilde{\overline{x}}(1)$. Combining this with (29), we have

$$
g(P^0_{0,1}(\overline{x}_0)) - \phi(P^\epsilon_{0,1}(\overline{x}_0))
$$
$$
\leq g(P^\epsilon_{0,1}(\overline{x}_0)) - \phi(P^\epsilon_{0,1}(\overline{x}_0)) \leq g(P^0_{0,1}(\overline{x}_0)) - \phi(P^0_{0,1}(\overline{x}_0)).
$$

Simplifying this equation, we get

(30)
$$
\frac{\phi(P^\epsilon_{0,1}(\overline{x}_0)) - \phi(P^0_{0,1}(\overline{x}_0))}{\epsilon} \geq 0.
$$

If R_t denotes the flow of the vector field $\overline{F}_v$, then

(31)
$$
P^\epsilon_{0,1} = P^0_{\tau,1} \circ R_\epsilon \circ P^0_{0,\tau-\epsilon}.
$$

So, if we assume that τ is a point of differentiability of the map $t \mapsto P^0_{0,t}$ which is true for almost all time τ in the interval $[0,1]$, then $P^\epsilon_{0,1}$ is differentiable with respect to ϵ at zero. Therefore, we can let ϵ goes to 0 in (30) and obtain

(32)
$$
\overline{\alpha} \left(\frac{d}{d\epsilon} \Big|_{\epsilon=0} P^\epsilon_{0,1} \right) \geq 0.
$$

If we differentiate equation (31) with respect to ϵ and set it to zero, it becomes

$$
\frac{d}{d\epsilon} \Big|_{\epsilon=0} P^\epsilon_{0,1} = (P^0_{\tau,1})_* (\overline{F}_v - \overline{F}_{\widetilde{u}(\tau)}) \circ P^0_{0,1}.
$$

Substituting this equation back into (32), we get the following:

$$(33) \qquad ((P_{\tau,1}^0)^*\overline{\alpha})(\overline{F}_v(\widetilde{\overline{x}}(\tau)) - \overline{F}_{\widetilde{u}(\tau)}(\widetilde{\overline{x}}(\tau))) \geq 0.$$

Define $\widetilde{p} : [0,1] \to T^*N$ by $\widetilde{p}(t) = (P_{t,1}^0)^*\overline{\alpha}$. Then the first two assertions of the theorem are clearly satisfied.

The following is well known (see [4] or [17]).

Lemma 10.4. *Let $\theta = pdq$ be the tautological 1-form on the cotangent bundle of the manifold N. Then for each diffeomorphism $P : N \to N$, the pull-back map $P^* : T^*N \to T^*N$ on the cotangent bundle of the manifold preserves the 1-form θ.*

Let W_t be the time-dependent vector field on the cotangent bundle of the manifold which satisfies

$$\frac{d}{dt}(P_{t,1}^0)^* = W_t \circ (P_{t,1}^0)^*$$

for almost all time t in $[0,1]$. If $\mathcal{L}_V$ denotes the Lie derivative with respect to a vector field V, then, by Lemma 10.4, the following is true for almost all time t in $[0,1]$:

$$\mathcal{L}_{W_t}\theta = 0.$$

If $\omega = -d\theta$ is the canonical symplectic 2-form on the cotangent bundle, then, by using Cartan's formula, we have

$$i_{W_t}\omega = d(\theta(W_t)).$$

Therefore, the vector field W_t is a Hamiltonian vector field with Hamiltonian given by

$$\overline{H}_{\widetilde{u}(t)}(p) = p(\overline{F}(\overline{x}, \widetilde{u}(t))).$$

The third assertion of the theorem follows from this. The last assertion follows from (33). $\qquad\square$

Going back to Problem 10.1, we can apply the Pontryagin Maximum Principle for the Mayer problem. Let $(\widetilde{x}(\cdot), \widetilde{z}(\cdot))$ be an admissible pair which minimizes Problem 10.1 and let $\overline{H}_t : T^*M \times \mathbb{R} \to \mathbb{R}$ be the function defined by

$$\overline{H}_t(p,l) = p(F(x, \widetilde{u}(t))) + l \cdot L(x, \widetilde{u}(t)).$$

By Theorem 10.3, there exists a curve $(\widetilde{p}(\cdot), \widetilde{l}(\cdot)) : [0,1] \to T_{\widetilde{x}}^*M \times \mathbb{R}$ such that $\widetilde{x}(t) = \pi(\widetilde{p}(t))$ and

$$(34) \qquad \begin{cases} (\dot{\widetilde{p}}, \dot{\widetilde{l}}) = \overrightarrow{\overline{H}}_t(\widetilde{p}, \widetilde{l}), \\ (\widetilde{p}(1), \widetilde{l}(1)) = (-\alpha, 1), \\ \overline{H}_t(\widetilde{p}(t), \widetilde{l}(t)) = \min_{u \in U}\left(\widetilde{p}(t)(F(\widetilde{x}(t), u)) + \widetilde{l}(t) \cdot L(\widetilde{x}(t), u)\right). \end{cases}$$

From the first equation in (34), we get $\dot{\widetilde{l}} = 0$ and $\widetilde{l}(1) = 1$. So, $\widetilde{l}(t) \equiv 1$. Therefore, (34) is simplified to

$$(35) \qquad \begin{cases} \dot{\widetilde{p}} = \overrightarrow{\overline{H}}_{\widetilde{u}}(\widetilde{p}), \\ \widetilde{p}(1) = -\alpha, \\ H_{\widetilde{u}}(\widetilde{p}(t), \widetilde{P}(t)) = \min_{u \in U}\left(\widetilde{p}(t)(F(\widetilde{x}(t), u)) + L(\widetilde{x}(t), u)\right). \end{cases}$$

This finishes the proof of Theorem 2.3.

Acknowledgment

The second author would like to express deep gratitude to his supervisor, Boris Khesin, who suggested to him the problem of optimal mass transportation on sub-riemannian manifolds.

References

[1] L. Ambrosio, S. Rigot: Optimal mass transportation in the Heisenberg group, J. Funct. Anal. 208(2004), 261-301 MR2035027 (2004m:49107)

[2] A. A. Agrachev: Geometry of Optimal Control Problems and Hamiltonian Systems, Lecture Notes, 2004

[3] A. A. Agrachev, J. P. Gauthier, On the subanalyticity of Carnot-Carathéodory distances, Ann. Inst. H. Poincaré Anal. Non Linéaire 18, (2001), 359–382 MR1831660 (2002h:93031)

[4] A. A. Agrachev, Y. L. Sachkov: Control Theory from the Geometric Viewpoint, Springer, 2004 MR2062547 (2005b:93002)

[5] A. A. Agrachev, A. V. Sarychev: Strong minimality of abnormal geodesics for 2-distributions, J. Dynam. Control Systems, 1995, v.1, 139-176 MR1333769 (96d:58022)

[6] A. A. Agrachev, A. V. Sarychev: Abnormal sub-Riemannian geodesics, Morse index and rigidity, Annales de l'Institut Henri Poincaré Analyse Non Linéaire, v.13, 1996, 635-690 MR1420493 (98e:49047)

[7] P. Bernard, B. Buffoni: Optimal mass transportation and Mather theory, J. Eur. Math. Soc., 9 (2007), no. 1, 85-121 MR2283105 (2007j:49050)

[8] Y. Brenier: Polar factorization and monotone rearrangement of vector-valued functions, Comm. Pure Appl. Math. 44, 4(1991), 375-417 MR1100809 (92d:46088)

[9] P. Cannarsa, L. Rifford: Semiconcavity results for optimal control problems admitting no singular minimizing controls, Ann. Inst. H. Poincaré Anal. Non Linéaire 25 (2008), 773-802 MR2436793

[10] P. Cannarsa, C. Sinestrari: Semiconcave Functions, Hamilton-Jacobi Equations, and Optimal Control, Birkhaüser Boston, 2004 MR2041617 (2005e:49001)

[11] A. Fathi, A. Figalli: Optimal transportation on non-compact manifolds, preprint

[12] A. Figalli: Existence, uniqueness and regularity of optimal transport maps, SIAM J. Math. Anal. 39 (2007), 126-137 MR2318378

[13] R. V. Gamkrelidze: Principles of Optimal Control Theory, Plenum Publishing Corporation, New York, 1978 MR0686793 (58:33350c)

[14] V. Jurdjevic: Geometric control theory, Cambridge Studies in Advanced Mathematics, 52. Cambridge University Press, Cambridge, 1997 MR1425878 (98a:93002)

[15] L. Kantorovich: On the translocation of masses, C.R. (Doklady) Acad. Sci. URSS(N.S.), 37, 1942, 199-201 MR0009619 (5:174d)

[16] W. S. Liu, H. J. Sussmann: Shortest paths for sub-Riemannian metrics on rank-2 distributions, Memoirs Amer. Math. Soc., v.118, N. 564, 1995 MR1303093 (96c:53061)

[17] J. Marsden, T. Ratiu: Introduction to Mechanics and Symmetry, Springer, 1999 MR1723696 (2000i:70002)

[18] R. McCann: Polar factorization of maps in Riemannian manifolds, Geometric and Functional Analysis, Vol.11, 2001 MR1844080 (2002g:58017)

[19] R. Montgomery: Abnormal Minimizers , SIAM J. Control and Optimization, vol. 32, no. 6, 1994, 1605-1620. MR1297101 (95g:49006)

[20] R. Montgomery: A tour of subriemannian geometries, their geodesics and applications, Math. Surveys and Monographs, vol. 91, Amer. Math. Soc., 2002 MR1867362 (2002m:53045)

[21] A. Sarychev, D. Torres: Lipschitzian regularity conditions for the minimizing trajectories of optimal control problems, Nonlinear analysis and its applications to differential equations (Lisbon, 1998), 357-368, Progr. Nonlinear Differential Equations Appl., 43, Birkhäuser Boston, Boston, MA, 2001 MR1800636 (2001j:49062)

[22] H. J. Sussmann: A Cornucopia of Abnormal Sub-Riemannian Minimizers. Part I: The Four dimensional Case, IMA technical report no. 1073, December, 1992

[23] C. Villani: Topics in Optimal Transportation, Graduate Studies in Mathematics 58, AMS, Providence, 2003 MR1964483 (2004e:90003)

[24] C. Villani: Optimal transport: Old and new, preprint, `http://www.umpa.ens-lyon.fr/` `~cvillani/Cedrif/B07B.StFlour.pdf`

Scuola Internazionale Superiore di Studi Avanzat, International School for Advanced Studies, Trieste, Italy and Steklov Mathematical Institute, ul. Gubkina 8, Moscow, 119991 Russia

   E-mail address: `agrachev@sissa.it`

Department of Mathematics, University of Toronto, Ontario, Canada M5S 2E4

   E-mail address: `plee@math.toronto.edu`