Mass transportation with LQ cost functions

Ahed Hindawi, Ludovic Rifford, Jean-Baptiste Pomet

▶ **To cite this version:**

Ahed Hindawi, Ludovic Rifford, Jean-Baptiste Pomet. Mass transportation with LQ cost functions. Acta Applicandae Mathematicae, 2011, 113 (2), pp.215-229. 10.1007/s10440-010-9595-1 . hal-00534083v2

HAL Id: hal-00534083
https://hal.science/hal-00534083v2

Submitted on 19 May 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Mass transportation with LQ cost functions

A. Hindawi* L. Rifford† J.-B. Pomet‡

Abstract

We study the optimal transport problem in the Euclidean space where the cost function is given by the value function associated with a Linear Quadratic minimization problem. Under appropriate assumptions, we generalize Brenier's Theorem proving existence and uniqueness of an optimal transport map. In the controllable case, we show that the optimal transport map has to be the gradient of a convex function up to a linear change of coordinates. We give regularity results and also investigate the non-controllable case.

1 Introduction

The optimal transport problem can be stated as follows: given two probability measures μ_0 and μ_1, defined on measurable spaces X and Y respectively and a cost function

$$c : X \times Y \longrightarrow \mathbb{R} \cup \{+\infty\}$$

find a measurable map

$$T : X \longrightarrow Y$$

which pushes forward μ_0 to μ_1, that is

$$T_\sharp \mu_0 = \mu_1 \qquad (\text{i.e. } \mu_1(B) = \mu_0\left(T^{-1}(B)\right) \text{ for all } B \subset Y \text{ measurable}),$$

and which minimizes the transportation cost

$$\text{cost}_c(T) := \int_X c(x, T(x))\, d\mu_0(x).$$

When the transport condition $T_\sharp \mu_0 = \mu_1$ is satisfied, we say that T is a *transport map*, and if T minimizes also the cost, that is if

$$\text{cost}_c(T) = \min_{S_\sharp \mu_0 = \mu_1} \Big\{ \text{cost}_c(S) \Big\}$$

then we call it an *optimal transport map*. Since the seminal famous paper by Gaspard Monge in 1781 [12], there was a revival of interest in mass transportation in the nineties.

*Université de Nice-Sophia Antipolis, Labo. J.-A. Dieudonné, UMR 6621, Parc Valrose, 06108 Nice Cedex 02, France *and* INRIA Sophia Antipolis. (hindawi@unice.fr)

†Université de Nice-Sophia Antipolis, Labo. J.-A. Dieudonné, UMR 6621, Parc Valrose, 06108 Nice Cedex 02, France. On leave to INRIA Sophia Antipolis for one year. (rifford@unice.fr)

‡INRIA, B.P. 93, 06902 Sophia Antipolis cedex, France. (Jean-Baptiste.Pomet@sophia.inria.fr)

In 1987 [3, 4], Brenier proved an existence and uniqueness result of optimal transport maps for the cost $c(x,y) = |x-y|^2$ in $\mathbb{R}^n$ and showed that any such optimal transport map is indeed the gradient of a convex function. Since then, people extended the theory to other costs functions in $\mathbb{R}^n$ or to other types of spaces (see [15]).

The aim of this paper is to study existence, uniqueness, and regularity of optimal transport maps for costs functions coming for LQ minimization problems in $\mathbb{R}^n$. Let us consider a linear control system of the form

$$\dot{x} = Ax + Bu \tag{1.1}$$

where the state x belongs to $\mathbb{R}^n$, the control u belongs to $\mathbb{R}^m$ and A, B are $n \times n$ and $n \times m$ matrices respectively. For every initial state $x \in \mathbb{R}^n$ and every control $u \in L^2\big([0,1];\mathbb{R}^m\big)$, we denote by $x(\cdot\,;x,u) : [0,1] \to \mathbb{R}^n$ the unique solution to the Cauchy problem

$$\begin{cases} \dot{x}(t) = Ax(t) + Bu(t) \quad \text{for a.e. } t \in [0,1], \\ x(0) = x. \end{cases} \tag{1.2}$$

Let us in addition consider a quadratic Lagrangian of the form

$$L(x,u) = \frac{1}{2}\langle x, Wx\rangle + \frac{1}{2}\langle u, Uu\rangle, \tag{1.3}$$

where W is a $n \times n$ symmetric non-negative matrix and U is a $m \times m$ symmetric positive definite matrix. The cost $c : \mathbb{R}^n \times \mathbb{R}^n \to [0, +\infty]$ associated with (1.1) and (1.3) is given by

$$c(x,y) := \inf\left\{ \int_0^1 L(x(t;x,u),u(t))dt \,|\, u \in L^2\big([0,1];\mathbb{R}^m\big) \text{ s.t. } x(1;x,u) = y \right\}, \tag{1.4}$$

where we set $c(x,y) = +\infty$ if there is no $u \in L^2\big([0,1];\mathbb{R}^m\big)$ such that $x(1;x,u) = y$.

We notice that LQ costs as above include as a particular case the Euclidean cost $1/2|x-y|^2$ by taking $B = I_n$, $U = I_n$, $A = 0$, $W = 0$. Costs coming from optimal control have already been studied in [1] (see also [6]). However, this reference does not give regularity results or properties relating the transport map with the gradient of a convex function. Furthermore, the study of a cost that is finite only for points lying in the same leaf of a foliation, like in the non-controllable case is also original.

The structure of the paper is the following. Our results are stated in Section 2. In Section 3, we introduce some preliminaries in optimal transport theory concerned with Kantorovitch duality. In Sections 4 and 5, we provide the proofs of our results. Then, we present some examples in Section 6. Finally in Section 7, we conclude with several remarks about our results.

2 Main results

2.1 Preliminaries on linear systems

Consider system (1.1) and let V be the smallest linear subspace of $\mathbb{R}^n$ that contains the image of the operator $B : \mathbb{R}^m \to \mathbb{R}^n$ and is invariant by A, or, more explicitly,

$$V = \mathrm{Span}_{\mathbb{R}}\Big\{B, AB, A^2B, \ldots, A^{n-1}B\Big\} \subset \mathbb{R}^n\,, \quad d = \dim V. \tag{2.1}$$

The well-known *Kalman criterion* states that system (1.1) is controllable if and only if d; this is stated below with a description of the situation where $d < n$. This decomposition can be found for instance in [13, Lemma 3.3.3].

Proposition 2.1. *Let x, y be in $\mathbb{R}^n$. There exists a control $u \in L^2\big([0,1];\mathbb{R}^m\big)$ such that $x(t;x,u) = y$ if and only if $e^A x - y \in V$ (i.e. y lies in the affine subspace $e^A x + V$). In particular it exists for all x, y if d; the system –or the pair (A,B)– is then called* controllable.

If $d < n$, after a linear change of coordinates in $\mathbb{R}^n$, the control system (1.1) has the following form:

$$\dot{x} = \begin{pmatrix} \dot{x}_1 \\ \dot{x}_2 \end{pmatrix} = \begin{pmatrix} A_1 & A_3 \\ 0 & A_2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{pmatrix} B_1 u \\ 0 \end{pmatrix}, \tag{2.2}$$

where the state x is partitioned into two blocks x_1 and x_2 of dimension d and $n-d$ respectively and A_1, A_2, A_3, B_1 are $d \times d$, $d \times n-d$, $n-d \times n-d$ and $d \times m$ matrices respectively, such that the pair (A_1, B_1) is controllable.

2.2 The controllable case

The following result of existence, uniqueness and regularity follows easily from the classical theory of optimal mass transportation. Throughout the paper, M^* denotes the transpose of the matrix M.

Theorem 2.2. *Assume that the linear control system (1.1) is controllable. Then there are symmetric positive definite $n \times n$ matrices D and F, and an invertible $n \times n$ matrix E such that*

$$c(x,y) = \frac{1}{2}\,\langle x, D\,x\rangle - \langle x, E\,y\rangle + \frac{1}{2}\,\langle y, F\,y\rangle \qquad \forall x, y \in \mathbb{R}^n. \tag{2.3}$$

Let μ_0, μ_1 be two compactly supported probability measures on $\mathbb{R}^n$. Assume that μ_0 is absolutely continuous with respect to the Lebesgue measure $\mathcal{L}^n$. Then, there is existence and uniqueness of an optimal transport map $T : \mathbb{R}^n \to \mathbb{R}^n$. Such a map is characterized by the existence of a convex function $\varphi : \mathbb{R}^n \to \mathbb{R}$ such that

$$T(x) = E^{-1}\,\nabla\varphi(x) \qquad \text{for a.e. } x \in \mathbb{R}^n. \tag{2.4}$$

If in addition, μ_0, μ_1 are associated with probability densities f_0, f_1 on $\mathrm{supp}(\mu_0)$ and $\mathrm{supp}(\mu_1)$ respectively, with f_0 and f_1 bounded from below and above, and if $\mathrm{supp}(\mu_0)$ is connected and $\mathrm{supp}(\mu_1)$ convex, then T is continuous.

We postpone several remarks concerning Theorem 2.2 to Section 7.

2.3 The non-controllable case

Here $d < n$ in (2.1), hence the evolution of the $(n-d)$-dimensional x_2 is totally fixed by (2.2) and $c(x,y)$ will obviously be infinite for almost all pairs x, y: these such that

$y_2 \neq e^{A_2} x_2$. A result similar to the above theorem requires first there exists at least one transport map with finite cost, i.e.

$$\inf_{S_\sharp \mu_0 = \mu_1} \left\{ \mathrm{cost}_c(S) := \int_{\mathbb{R}^n} c(x, S(x))\, d\mu_0(x), \right\} < +\infty. \tag{2.5}$$

Let $\pi_2 : \mathbb{R}^n \to \mathbb{R}^{n-d}$ be the projection on the second block in the coordinates of (2.2): $\pi_2(x) = x_2$ (it is also the projection on the quotient $\mathbb{R}^n/V$).

Theorem 2.3. *Let μ_0 and μ_1 be compactly supported probability measures with* continuous *densities f_0, f_1 ($\mu_0 = f_0 \mathcal{L}^n, \mu_1 = f_1 \mathcal{L}^n$). There exists a transport map with finite cost (2.5) if and only if*

$$(\pi_2)_\sharp \mu_1 = \left(e^{A_2} \circ \pi_2 \right)_\sharp \mu_0 \,. \tag{2.6}$$

If this is satisfied, then there exists a unique optimal transport map $T : \mathbb{R}^n \to \mathbb{R}^n$. Moreover, if both densities f_0, f_1 are bounded from below and above on $\mathrm{supp}(\mu_0)$ *and* $\mathrm{supp}(\mu_1)$ *respectively, and if* $\mathrm{supp}(\mu_0)$ *and* $\mathrm{supp}(\mu_1)$ *are both convex, then T is continuous.*

Again, remarks concerning Theorem 2.3 are postponed to Section 7.

3 Preliminaries in optimal transport theory

Given two probability measures μ_0, μ_1 on $\mathbb{R}^n$ and a cost function $c : \mathbb{R}^n \times \mathbb{R}^n \to [0, +\infty]$, we are looking for a transport map $T : \mathbb{R}^n \to \mathbb{R}^n$ which minimizes the transportation cost $\int_{\mathbb{R}^n} c(x, T(x))\, d\mu_0$. The constraint $T_\# \mu_0 = \mu_1$ being highly non-linear, the optimal transport problem is quite difficult from the viewpoint of calculus of variation. The major advance on this problem was due to Kantorovich, who proposed in [8, 9] a notion of weak solution of the optimal transport problem. He suggested to look for *plans* instead of transport maps, that is probability measures γ in $\mathbb{R}^n \times \mathbb{R}^n$ whose marginals are μ_0 and μ_1, that is

$$(\pi_1)_\sharp \gamma = \mu_0 \qquad \text{and} \qquad (\pi_2)_\sharp \gamma = \mu_1,$$

where $\pi_1 : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$ are the canonical projections on the first and second variable respectively. Denoting by $\Pi(\mu_0, \mu_1)$ the set of plans, the new minimization problem becomes the following:

$$C(\mu_0, \mu_1) = \min_{\gamma \in \Pi(\mu_0, \mu_1)} \left\{ \int_{\mathbb{R}^n \times \mathbb{R}^n} c(x, y)\, d\gamma(x, y) \right\}. \tag{3.1}$$

If γ is a minimizer for the Kantorovich formulation, we say that it is an *optimal plan*. Due to the linearity of the constraint $\gamma \in \Pi(\mu_0, \mu_1)$, it is simple, using weak topologies, to prove existence of solutions to (3.1) as soon as c is lower semi-continuous (see for instance [15]). The connection between the formulation of Kantorovich and that of Monge can be seen by noticing that any transport map T induces the plan defined by $(Id \times T)_\sharp \mu_0$ which is concentrated on the graph of T. Thus, the problem of showing existence of optimal transport maps can be reduced to prove that an optimal transport plan is concentrated on a graph. Moreover, if one can show that *any* optimal plan is concentrated on a graph, since $\frac{\gamma_1 + \gamma_2}{2}$ is optimal if so are γ_1 and γ_2, uniqueness of the transport map easily follows. The following definition is exactly [15, Definition 5.2]:

Definition 3.1. Let $c : \mathbb{R}^n \times \mathbb{R}^n \to [0, \infty]$. A function $\psi : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is said to be *c-convex* if it is not identically $+\infty$ and there exists a function $\zeta : \mathbb{R}^n \to \mathbb{R} \cup \{-\infty, +\infty\}$ such that

$$\psi(x) = \sup_{y \in \mathbb{R}^n} \Big(\zeta(y) - c(x, y) \Big) \qquad \forall x \in \mathbb{R}^n.$$

Then its *c-transform* is the function ψ^c defined by

$$\psi^c(y) = \inf_{x \in \mathbb{R}^n} \Big(\psi(x) + c(x, y) \Big) \qquad \forall y \in \mathbb{R}^n,$$

and its *c-subdifferential* is the set defined by

$$\partial_c \psi := \Big\{ (x, y) \in \mathbb{R}^n \times \mathbb{R}^n \,|\, \psi^c(y) - \psi(x) = c(x, y) \Big\}.$$

Moreover, the *c-subdifferential of* ψ at $x \in \mathbb{R}^n$ is

$$\partial_c \psi(x) := \Big\{ y \in \mathbb{R}^n \,|\, (x, y) \in \partial_c \psi \Big\},$$

or equivalently

$$\psi(x) + c(x, y) \leq \psi(z) + c(z, y) \qquad \forall z \in \mathbb{R}^n.$$

The functions ψ and ψ^c are said to be *c-conjugate*.

Let us denote by $P_c(\mathbb{R}^n)$ the set of compactly supported probability measures. From now on, $\mathrm{supp}(\mu_0)$ and $\mathrm{supp}(\mu_1)$ will denote the supports of μ_0 and μ_1 respectively, i.e. the smallest closed sets on which μ_0 and μ_1 are respectively concentrated. Kantorovitch duality can be stated as follows (see [15, Theorem 5.10]):

Theorem 3.2. *Let $\mu_0, \mu_1 \in P_c(\mathbb{R}^n)$ and $c : \mathbb{R}^n \times \mathbb{R}^n \to [0, \infty)$ be a lower continuous function. Then there exists a c-convex function $\psi : \mathbb{R}^n \to \mathbb{R}$ such that the following holds: a transport plan $\gamma \in \Pi(\mu_0, \mu_1)$ is optimal if and only if $\gamma(\partial_c \psi) = 1$ (that is, γ is concentrated on the c-subdifferential of ψ). Moreover ψ can be chosen such that*

$$\psi(x) = \sup_{y \in \mathrm{supp}(\mu_1)} \{\psi^c(y) - c(x, y)\} \qquad \forall x \in \mathbb{R}^n, \tag{3.2}$$

$$\psi^c(y) = \inf_{x \in \mathrm{supp}(\mu_0)} \{\psi(x) + c(x, y)\} \qquad \forall y \in \mathbb{R}^n. \tag{3.3}$$

By the above theorem we see that, in order to prove existence and uniqueness of optimal transport maps, it suffices to prove that there exist two Borel sets $Z_0, Z_1 \subset \mathbb{R}^n$, with $\mu_0(Z_0) = \mu_1(Z_1) = 1$, such that and $\partial_c \psi$ is a graph inside $Z_0 \times Z_1$ (or equivalently that $\partial_c \psi(x) \cap Z_1$ is a singleton for all $x \in Z_0$).

4 Proof of Theorem 2.2

First, we prove that (2.3) holds. This is not new, but the formulas are not given in textbooks *in extenso*. The case $W = 0$ is more classical and one may find, for instance in [11] the expression of c in that case:

$$c(x, y) = \langle x - e^{-A} y \,,\, \mathcal{G}^{-1}(x - e^{-A} y) \rangle \;\text{ with }\; \mathcal{G} = \int_0^1 e^{\tau A^*} B U^{-1} B^* e^{\tau A} \mathrm{d}\tau \tag{4.1}$$

where the controllability Grammian $\mathcal{G}$ is positive definite because the system is controllable. This is indeed of the form (2.3). Let us derive the general form from the classical linear Pontryagin Maximum principle (see for instance the same reference).

The pseudo Hamiltonian $H_0 : \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ associated with the optimization problem under study is given by

$$H_0(x, p, u) := \langle p, Ax + Bu\rangle - \frac{1}{2}\langle x, Wx\rangle - \frac{1}{2}\langle u, Uu\rangle, \tag{4.2}$$

the control u that maximizes this expression for each (p, x) is given by

$$u = U^{-1}B^*p \tag{4.3}$$

and the Hamiltonian $H : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$, defined as $H(x, p) = \max\{H_0(x, p, u) \,|\, u \in \mathbb{R}^m\}$ is given by

$$H(x, p) \quad = \quad \langle p, Ax\rangle - \frac{1}{2}\langle x, Wx\rangle + \frac{1}{2}\langle p, BU^{-1}B^*p\rangle. \tag{4.4}$$

Therefore the Hamiltonian differential equation $\dot{x} = \partial H/\partial p$, $\dot{p} = -\partial H/\partial x$ associated to our minimization problem is given by

$$\begin{pmatrix} \dot{x} \\ \dot{p} \end{pmatrix} = \begin{pmatrix} A & BU^{-1}B^* \\ W & -A^* \end{pmatrix} \begin{pmatrix} x \\ p \end{pmatrix} \tag{4.5}$$

Denote by $R(\cdot) : [0, 1] \to M_{2n}(\mathbb{R})$ the fundamental solution to the Cauchy problem

$$\dot{R}(t) = \begin{pmatrix} A & BU^{-1}B^* \\ W & -A^* \end{pmatrix} R(t) \quad \forall t \in [0, 1], \qquad R(0) = I_{2n}, \tag{4.6}$$

and write it as

$$R(t) = \begin{pmatrix} R_1(t) & R_2(t) \\ R_3(t) & R_4(t) \end{pmatrix} \qquad \forall t \in [0, 1],$$

where $R_i(\cdot)$ is a $n \times n$ matrix for $i = 1, \cdots 4$. For every $x \in \mathbb{R}^n$ fixed, denote by $\exp_x$ the mapping from $\mathbb{R}^n$ to $\mathbb{R}^n$ which sends via the Hamiltonian vector field the initial adjoint vector $p \in \mathbb{R}^n$ to the final state $x(1)$, that is

$$\exp_x(p) := R_1(1)x + R_2(1)p \qquad \forall p \in \mathbb{R}^n. \tag{4.7}$$

Controllability of the system (1.1) implies that the affine mapping $\exp_x$ is onto, and hence a bijection. Indeed, there is, for any $x, y \in \mathbb{R}^n$, (at least) one control $u \in L^2\big([0, 1]; \mathbb{R}^m\big)$ which minimizes the cost

$$\int_0^1 L(x(t; x, u), u(t))dt$$

among all controls steering x to y in time 1; thanks to the linear maximum principle, there corresponds to each minimizing control a $p \in \mathbb{R}^n$ such that $\exp_x(p) = y$. Hence the matrix $R_2(t)$ is invertible for all t.

Given x and y in $\mathbb{R}^n$, the cost between x and y is given by

$$c(x, y) = \int_0^1 \frac{1}{2}\langle x(t), Wx(t)\rangle + \frac{1}{2}\langle p(t), BU^{-1}B^*p(t)\rangle dt,$$

where $\big(x(\cdot),p(\cdot)\big) : [0,1] \to \mathbb{R}^n \times \mathbb{R}^n$ is the solution to the Hamiltonian system (4.5) satisfying

$$x(0) = x \quad \text{and} \quad \exp_x\big(p(0)\big) = y.$$

Set $p := p(0)$, that is

$$p = R_2(1)^{-1}\Big(y - R_1(1)x\Big). \tag{4.8}$$

We deduce easily that c is a smooth function of the form

$$c(x,y) = \frac{1}{2}\langle x, Q_1 x\rangle + \langle x, Cp\rangle + \frac{1}{2}\langle p, Q_2 p\rangle \qquad \forall x,y \in \mathbb{R}^n, \tag{4.9}$$

where both Q_1, Q_2 are non-negative symmetric $n \times n$ matrices given by

$$Q_1 := \int_0^1 \Big(R_1(t)^* W R_1(t) + R_3(t)^* B U^{-1} B^* R_3(t)\Big) dt,$$

$$Q_2 := \int_0^1 \Big(R_2(t)^* W R_2(t) + R_4(t)^* B U^{-1} B^* R_4(t)\Big) dt,$$

and where C is the $n \times n$ matrix given by

$$C := \int_0^1 \Big(R_1(t)^* W R_2(t) + R_3(t)^* B U^{-1} B^* R_4(t)\Big) dt.$$

Let $x, y \in \mathbb{R}^n$ and $u \in L^2\big([0,1];\mathbb{R}^m\big)$ a control minimizing $c(x,y)$ be fixed. For every control $v \in L^2\big([0,1],\mathbb{R}^m\big)$, denote by $y(\cdot;y,v) : [0,1] \to \mathbb{R}^n$ the unique solution to the Cauchy problem

$$\begin{cases} \dot{y}(t) = Ay(t) + Bv(t) & \text{for a.e. } t \in [0,1], \\ y(1) = y. \end{cases}$$

By definition of the cost c, there holds

$$\int_0^1 L(y(t;y,v),v(t))dt - c\big(y(0;y,v),y\big) \geq 0 \qquad \forall v \in L^2\big([0,1],\mathbb{R}^m\big).$$

Moreover, there is equality in the above inequality whenever $v = u$. Thanks to the linear maximum principle, since the cost c is smooth, this means that

$$y = \exp_x(p) \quad \text{with} \quad p = -\nabla_x c(x,y).$$

Computing $\nabla_x c(x,y)$ by differentiating (4.9)-(4.8) with respect to the variable x yields

$$Q_1 x + Cp - R_1(1)^*\big(R_2(1)^{-1}\big)^*\big(C^* x + Q_2 p\big) = -p,$$

and finally (recall that Q_1 is symmetric) :

$$C = -I_n + R_1(1)^*\big(R_2(1)^{-1}\big)^* Q_2, \quad Q_1 = R_1(1)^*\big(R_2(1)^{-1}\big)^* C^* = C R_2(1)^{-1} R_1(1). \tag{4.10}$$

Plugging this and (4.8) into (4.9) yields the expression (2.3) for $c(x,y)$ with

$$D = R_2(1)^{-1} R_1(1)\,, \quad E = R_2(1)^{-1}\,, \quad F = (R_2(1)^{-1})^* Q_2 R_2(1)^{-1}\,, \tag{4.11}$$

where D is symmetric because (4.10) implies $Q_1 = -E + E^* Q_2 E$ and Q_1, Q_2 are symmetric; Positive definiteness of D and F may be deduced from the fact that there is no solution of (1.1) driving 0 to a nonzero y or a nonzero x to 0 in finite time with a zero cost.

Let us now show the existence and uniqueness of an optimal transport map. Let μ_0, μ_1 be two compactly supported probability measures in $\mathbb{R}^n$ and assume that μ_0 is absolutely continuous with respect to the Lebesgue measure. Let $\psi, \phi := \psi^c : \mathbb{R}^n \to \mathbb{R}$ be the Kantorovitch potentials given by Theorem 3.2 (note that c is non-negative valued). First, since c is smooth and both sets $\mathrm{supp}(\mu_0), \mathrm{supp}(\mu_1)$ are assumed to be compact, (3.2)-(3.3) imply that both potentials ψ, ϕ are locally Lipschitz. Therefore, thanks to Rademacher's Theorem and the fact that μ_0 is absolutely continuous with respect to the Lebesgue measure, ψ is differentiable μ_0-almost everywhere. Let $x \in \mathrm{supp}(\mu_0)$ be a differentiability point of ψ. Let $y \in \mathrm{supp}(\mu_1)$ be such that

$$\phi(y) - \psi(x) = c(x, y),$$

and $u \in L^2\big([0,1], \mathbb{R}^m\big)$ be a control minimizing

$$c(x,y) = \inf \left\{ \int_0^1 L(x(t;x,v), v(t)) dt \,|\, v \in L^2\big([0,1]; \mathbb{R}^m\big) \text{ s.t. } x(1;x,v) = y \right\}.$$

Then we have

$$\int_0^1 L(y(t;y,v), v(t)) dt \geq c\big(y(0;y,v), y\big) \geq \phi(y) - \psi\big(y(0;y,v)\big) \qquad \forall v \in L^2\big([0,1]; \mathbb{R}^m\big),$$

with equality if $v = u$. As above, this implies that

$$y = \exp_x(p) \quad \text{ with } \quad p = \nabla \psi(x).$$

This shows that y is uniquely determined for every differentiability point of ϕ and in turn proves existence and uniqueness of an optimal measurable transport map $x \mapsto y$. Note that the potential ψ can be written as follows:

$$\begin{aligned} \psi(x) &= \sup_{y \in \mathrm{supp}(\mu_1)} \{\phi(y) - c(x,y)\} \\ &= -\frac{1}{2}\langle x, D\, x\rangle + \sup_{y \in \mathrm{supp}(\mu_1)} \left\{ \langle x, E\, y\rangle - \frac{1}{2}\langle y, F\, y\rangle + \phi(y) \right\}. \end{aligned}$$

This shows that the function $\varphi : \mathbb{R}^n \to \mathbb{R}$ defined by

$$\varphi(x) := \psi(x) + \frac{1}{2}\langle x, D\, x\rangle \qquad \forall x \in \mathbb{R}^n,$$

is convex (as the *sup* of a family of convex functions) while, for almost every $x \in \mathrm{supp}(\mu_0)$, deriving the expression of $\exp_x(p)$ from (4.7) and (4.11),

$$T(x) = \exp_x\big(\nabla \psi(x)\big) = E^{-1}\big(Dx + \nabla \psi(x)\big) = E^{-1} \nabla \varphi(x)\ .$$

It remains to show that under additional assumptions, T is continuous. One obviously has $\nabla \varphi(x) = E\, T(x)$ for all x; it is clear that this and $T_\sharp \mu_0 = \mu_1$ imply $(\nabla \varphi)_\sharp \mu_0 = \hat{\mu}_1$ for the measure $\hat{\mu}_1$ defined by

$$\hat{\mu}_1(A) = \mu_1\Big(\big\{E^{-1} x \,|\, x \in A\big\}\Big)$$

for every Borel set $A \subset \mathbb{R}^n$ ($\hat{\mu}_1$ is the pushforward of μ_1 by the map $x \mapsto Ex$). Since φ is convex, $\nabla\varphi$ is then the optimal transport map from μ_0 to $\hat{\mu}_1$ for the cost $\hat{c}(x,y) = \frac{1}{2}|x-y|^2$. This implies the continuity of $\nabla\varphi$ —hence of T— according to [15, Theorem 12.50 (i)] which asserts the following: if $\hat{\mu}_0, \hat{\mu}_1$ are two compactly supported probability measures in $\mathbb{R}^n$ associated with densities $\hat{f}_0, \hat{f}_1$ which are bounded from below and above on $\mathrm{supp}(\hat{f}_0)$ and $\mathrm{supp}(\hat{f}_1)$ respectively, and if $\mathrm{supp}(\hat{f}_0)$ is connected and $\mathrm{supp}(\hat{f}_1)$ is convex, then the optimal transport map $\hat{T} : \mathbb{R}^n \to \mathbb{R}^n$ from $\hat{\mu}_0$ to with respect to the Euclidean quadratic cost $\hat{c}(x,y) = \frac{1}{2}|x-y|^2$ is continuous. Take $\hat{\mu}_1$ defined above and set $\hat{\mu}_0 = \mu_0$; they do satisfy these assumptions for μ_0 and μ_1 and applying an invertible linear map does not change convexity of the support of a density, and we just proved that $\hat{T}$ is $\nabla\varphi$ with the same φ as above.

5 Proof of Theorem 2.3

Let us first treat the case $d = 0$ separately. According to (2.1), either there is no control ($m = 0$) or the matrix B is zero; in both cases the system reads $\dot{x} = Ax$ and

$$c(x,y) = \begin{cases} \frac{1}{2}\int_0^1 \langle e^{tA}\,x, We^{tA}\,x\rangle\,dt & \text{if } y = e^A x\,, \\ +\infty & \text{otherwise.} \end{cases}$$

Then there is only one map S such that $\mathrm{cost}_c(S) < \infty$, given by $S(x) = e^A x$ for all x, the compatibility on measures is $\mu_1 = S_\sharp \mu_0$ with this precise S, and the result is proved in the case $d = 0$.

From now on, we assume that $d \in \{1, \cdots, n-1\}$, i.e. none of the two blocks in (2.2) is void ($d \geq 1$, $n-d \geq 1$). For every $x = \big(x_1, x_2\big) \in \mathbb{R}^d \times \mathbb{R}^{n-d} = \mathbb{R}^n$ and $u \in L^2\big([0,1]; \mathbb{R}^m\big)$, the solution $x(\cdot; x, u) : [0,1] \to \mathbb{R}^n$ to the Cauchy problem (1.2) with the decomposition given by (2.2) satisfies

$$x(t;x,u) = \begin{pmatrix} x_1(t;x,u) \\ x_2\big(t;x_2\big) \end{pmatrix} = \begin{pmatrix} e^{tA_1}x_1 + e^{tA_1}\int_0^t e^{-sA_1}\big[A_3 e^{sA_2}x_2 + B_1 u(s)\big]\,ds \\ e^{tA_2}x_2 \end{pmatrix}, \qquad (5.1)$$

for every $t \in [0,1]$. Denote by $\bar{x}_1(\cdot) = \bar{x}_1\big(\cdot; x_1, u\big) : [0,1] \to \mathbb{R}^d$ the solution to the Cauchy problem

$$\begin{cases} \dot{\bar{x}}_1(t) = A_1\bar{x}_1(t) + B_1 u(t) & \text{for a.e. } t \in [0,1], \\ \bar{x}_1(0) = x_1. \end{cases} \qquad (5.2)$$

Then we have

$$x_1\big(t;x,u\big) = \bar{x}_1\big(t;x_1,u\big) + G(t)x_2 \qquad \forall t \in [0,1],$$

where $G : [0,1] \to M_{d,n-d}(\mathbb{R})$ is defined by

$$G(t) := e^{tA_1}\int_0^t e^{-sA_1}A_3 e^{sA_2}ds \qquad \forall t \in [0,1].$$

Therefore we have for almost every $t \in [0,1]$,

$$\begin{aligned} L\big(x(t;x,u),u(t)\big) &= \frac{1}{2}\langle x(t;x,u), Wx(t;x,u)\rangle + \frac{1}{2}\langle u(t), Uu(t)\rangle \\ &= \frac{1}{2}\big\langle \bar{x}_1\big(t;x_1,u\big), W_1\,\bar{x}_1\big(t;x_1,u\big)\big\rangle + \frac{1}{2}\langle u(t), Uu(t)\rangle \\ &\qquad + \big\langle \bar{x}_1\big(t;x_1,u\big), X\big(t;x_2\big)\big\rangle + l\big(t;x_2\big), \end{aligned}$$

where the symmetric matrix W is decomposed, in the coordinates of (2.2) as

$$W = \begin{pmatrix} W_1 & W_3 \\ {W_3}^* & W_2 \end{pmatrix}$$

with W_1 and W_2 positive definite $d \times d$ and $n-d \times n-d$ matrices respectively, W_3 a $d \times n-d$ matrix, and

$$\begin{aligned} X\big(t; x_2\big) &= W_1\, G(t)x_2 + W_3\, x_2\big(t; x_2\big)\,, \\ l\big(t; x_2\big) &= \frac{1}{2}\left\langle \begin{pmatrix} G(t)x_2 \\ x_2\big(t; x_2\big) \end{pmatrix}, \begin{pmatrix} W_1 & W_3 \\ {W_3}^* & W_2 \end{pmatrix} \begin{pmatrix} G(t)x_2 \\ x_2\big(t; x_2\big) \end{pmatrix} \right\rangle . \end{aligned}$$

In conclusion, we have for every $x = \big(x_1, x_2\big), y = \big(y_1, y_2\big) \in \mathbb{R}^n$,

$$c(x,y) = \begin{cases} \bar{c}_{x_2}\big(x_1, y_1 - G(1)x_2\big) + \int_0^1 l\big(t; x_2\big)dt & \text{if } y_2 = e^{A_2}x_2 \\ +\infty & \text{otherwise,} \end{cases} \tag{5.3}$$

where the cost $\bar{c}_{x_2} : \mathbb{R}^d \times \mathbb{R}^d \to [0, +\infty)$ is defined by

$$\begin{aligned} &\bar{c}_{x_2}\big(x_1, z_1\big) := \\ &\quad \inf\left\{ \int_0^1 \bar{L}_{x_2}\big(t, \bar{x}_1(t; x_1, u), u(t)\big)dt \,|\, u \in L^2\big([0,1]; \mathbb{R}^m\big) \text{ s.t. } \bar{x}_1\big(1; x_1, u\big) = z_1 \right\}, \end{aligned} \tag{5.4}$$

for any $x_1, z_1 \in \mathbb{R}^d$, and where the Lagrangian $\bar{L}_{x_2} : \mathbb{R} \times \mathbb{R}^d \times \mathbb{R}^m \to [0, +\infty)$ is defined by

$$\bar{L}_{x_2}\big(t, z, u\big) := \frac{1}{2}\,\langle z, W_1\, z\rangle + \frac{1}{2}\,\langle u(t), Uu(t)\rangle + \big\langle z, X\big(t; x_2\big)\big\rangle\,. \tag{5.5}$$

We proceed now as in the proof of Theorem 2.2. Let $x_2 \in \mathbb{R}^{n-d}$ be fixed, the pseudo Hamiltonian $H_0 : \mathbb{R} \times \mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^m \to \mathbb{R}$ associated with the cost $\bar{c}_{x_2}$ is given by

$$H_0(t, z, p, u) := \langle p, A_1 z + B_1 u\rangle - \frac{1}{2}\,\langle z, W_1\, z\rangle - \frac{1}{2}\,\langle u, U\, u\rangle - \big\langle z, X\big(t; x_2\big)\big\rangle\,. \tag{5.6}$$

Then $\frac{\partial H_0}{\partial u} = 0$ yields $u = U^{-1}B_1^* p$ and the Hamiltonian $H : \mathbb{R} \times \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ is given by

$$\begin{aligned} H(t, z, p) &= \max\left\{H_0(t, z, p, u) \,|\, u \in \mathbb{R}^m\right\} \\ &= \langle p, A_1\, z\rangle - \frac{1}{2}\,\langle z, W_1\, z\rangle + \frac{1}{2}\,\big\langle p, B_1 U^{-1} B_1^* p\big\rangle - \big\langle z, X\big(t; x_2\big)\big\rangle\,. \end{aligned}$$

Therefore the Hamiltonian system associated to our minimization problem is given by

$$\begin{cases} \dot{z} &= A_1 z + B_1 U^{-1} B_1^* p \\ \dot{p} &= -A_1^* p + W_1 z + X\big(t; x_2\big). \end{cases} \tag{5.7}$$

Denote by $R(\cdot) : [0,1] \to M_{2d}(\mathbb{R})$ the fundamental solution to the Cauchy problem

$$\dot{R}(t) = \begin{pmatrix} A_1 & B_1 U^{-1} B_1^* \\ W_1 & -A_1^* \end{pmatrix} R(t) \quad \forall t \in [0,1], \qquad R(0) = I_{2d}, \tag{5.8}$$

and write it as

$$R(t) = \begin{pmatrix} R_1(t) & R_2(t) \\ R_3(t) & R_4(t) \end{pmatrix} \qquad \forall t \in [0,1],$$

where $R_i(\cdot)$ is a $d \times d$ matrix for $i = 1, \cdots 4$. Any solution $\big(z(\cdot), p(\cdot)\big) : [0,1] \to \mathbb{R}^d \times \mathbb{R}^d$ of (5.7) with $z(0) = z, p(0) = p$ can be written as

$$\begin{cases} z(t) &= R_1(t)z + R_2(t)p + \bar{z}_{x_2}(t) \\ p(t) &= R_3(t)z + R_4(t)p + \bar{p}_{x_2}(t) \end{cases} \qquad \forall t \in [0,1],$$

where $\big(\bar{z}_{x_2}(\cdot), \bar{p}_{x_2}(\cdot)\big) : [0,1] \to \mathbb{R}^d \times \mathbb{R}^d$ of (5.7) satisfying $\bar{z}_{x_2}(0) = 0, \bar{p}_{x_2}(0) = 0$. For every $x_1 \in \mathbb{R}^d$, the mapping $\exp_{x_1} : \mathbb{R}^d \to \mathbb{R}^d$ defined by

$$\exp_{x_1}(p) := R_1(1)x_1 + R_2(1)p + \bar{z}_{x_2}(1) \qquad \forall p \in \mathbb{R}^d,$$

is an affine bijection. For every $x_1, z_1, p \in \mathbb{R}^d$ with $z_1 = \exp_{x_1}(p)$, there holds

$$\bar{c}_{x_2}\big(x_1, z_1\big) = \frac{1}{2}\,\langle x_1, Q_1\, x_1\rangle + \frac{1}{2}\,\langle p, Q_2\, p\rangle + \langle x_1, C\, p\rangle + \langle x_1, v_{x_2}\rangle + \langle p, w_{x_2}\rangle + k_{x_2}, \quad (5.9)$$

where both Q_1, Q_2 are non-negative symmetric $d \times d$ matrices given by

$$Q_1 := \int_0^1 \Big(R_1(t)^* W_1 R_1(t) + R_3(t)^* B_1 U^{-1} B_1^* R_3(t)\Big) dt,$$

$$Q_2 := \int_0^1 \Big(R_2(t)^* W_1 R_2(t) + R_4(t)^* B_1 U^{-1} B_1^* R_4(t)\Big) dt,$$

where C is the $d \times d$ matrix given by

$$C := \int_0^1 \Big(R_1(t)^* W_1 R_2(t) + R_3(t)^* B_1 U^{-1} B_1^* R_4(t)\Big) dt,$$

and the vectors v_{x_2}, w_{x_2} are given by

$$v_{x_2} := \int_0^1 \Big(R_1(t)^* W_1\, \bar{z}_{x_2}(t) + R_3(t)^* B_1 U^{-1} B_1^*\, \bar{p}_{x_2}(t) + R_1(t)^*\, X\big(t; x_2\big)\Big) dt,$$

$$w_{x_2} := \int_0^1 \Big(R_2(t)^* W_1\, \bar{z}_{x_2}(t) + R_4(t)^* B_1 U^{-1} B_1^*\, \bar{p}_{x_2}(t) + R_2(t)^*\, X\big(t; x_2\big)\Big) dt,$$

and

$$k_{x_2} := \int_0^1 \left(\frac{1}{2}\,\langle \bar{z}_{x_2}(t), W_1\, \bar{z}_{x_2}(t)\rangle + \frac{1}{2}\,\big\langle \bar{p}_{x_2}(t), B_1 U^{-1} B_1^*\, \bar{p}_{x_2}(t)\big\rangle + \big\langle \bar{z}_{x_2}(t), X\big(t; x_2\big)\big\rangle\right) dt.$$

We proceed now as in the proof of Theorem 2.2. Since $\bar{c}_{x_2}$ is smooth, using the linear maximum principle, taking the derivative in the x_1 variable in (5.9) and using that $p = R_2(1)^{-1}\big[z_1 - R_1(1)x_1 - \bar{z}_{x_2}(1)\big]$ yields

$$Q_1\, x_1 + C\, p + v_{x_2} - R_1(1)^*\big(R_2(1)^{-1}\big)^*\,(Q_2\, p + C^*\, x_1 + w_{x_2}) = -p$$

and finally $Q_1 - R_1(1)^*(R_2(1)^{-1})^* C^* = C - R_1(1)^*(R_2(1)^{-1})^* Q_2 + I = v_{x_2} - R_1(1)^* w_{x_2} = 0$, whence (recall that Q_2 is symmetric):

$$v_{x_2} = R_1(1)^* \big(R_2(1)^{-1}\big)^* w_{x_2}, \quad \begin{cases} C = -I_n + R_1(1)^* \big(R_2(1)^{-1}\big)^* Q_2 \\ Q_1 = R_1(1)^*(R_2(1)^{-1})^* C^* = C R_2(1)^{-1} R_1(1) \end{cases} \tag{5.10}$$

In conclusion, setting $D := R_2(1)^{-1} R_1(1), E := R_2(1)^{-1}$, and $F := E^* Q_2 E$ yields

$$\begin{aligned} \bar{c}_{x_2}\big(x_1, z_1\big) = \frac{1}{2} \langle x_1, D\, x_1 \rangle - \big\langle x_1, E\, \big(z_1 - \bar{z}_{x_2}(1)\big) \big\rangle \\ + \frac{1}{2} \big\langle \big(z_1 - \bar{z}_{x_2}(1)\big), F\, \big(z_1 - \bar{z}_{x_2}(1)\big) \big\rangle + \big\langle E\big(z_1 - \bar{z}_{x_2}(1)\big), w_{x_2} \big\rangle + k_{x_2}, \end{aligned}$$

which in turn implies (by (5.3))

$$\begin{aligned} c\,\big((x_1, x_2), \big(y_1, e^{A_2} x_2\big)\big) = \\ \frac{1}{2} \langle x_1, D\, x_1 \rangle - \big\langle x_1, E\, \big(y_1 - G(1) x_2 - \bar{z}_{x_2}(1)\big) \big\rangle \\ + \frac{1}{2} \big\langle \big(y_1 - G(1) x_2 - \bar{z}_{x_2}(1)\big), F\, \big(y_1 - G(1) x_2 - \bar{z}_{x_2}(1)\big) \big\rangle \\ + \big\langle E\big(y_1 - G(1) x_2 - \bar{z}_{x_2}(1)\big), w_{x_2} \big\rangle + k_{x_2} + \int_0^1 l\big(t; x_2\big) dt. \end{aligned} \tag{5.11}$$

D and F are symmetric definite positive for the same reason as in the proof of Theorem 2.2.

We are now ready to prove the result. For every $\bar{x}_2 \in \mathbb{R}^{n-d}$, set

$$\begin{aligned} M_{\bar{x}_2} &:= \Big\{ x = (x_1, x_2) \in \mathbb{R}^d \times \mathbb{R}^{n-d} \,|\, x_2 = \bar{x}_2 \Big\}, \\ N_{\bar{x}_2} &:= \Big\{ x = (x_1, x_2) \in \mathbb{R}^d \times \mathbb{R}^{n-d} \,|\, x_2 = e^{A_2} \bar{x}_2 \Big\}, \end{aligned}$$

and let $\mu_0 = f_0 \mathcal{L}^n, \mu_1 = f_1 \mathcal{L}^n$ with densities $f_0, f_1 \in L^1(\mathbb{R}^n)$ be two compactly supported probability measures.

Fact: *A measurable map S satisfies $S_\sharp \mu_0 = \mu_1$ and $cost_c(S) < +\infty$ if and only if there exists, for almost all x_2, a measurable $S_{x_2} : I\!R^d \to I\!R^d$ such that*

$$S(x_1, x_2) = \Big(S_{x_2}(x_1), e^{A_2} x_2 \Big) \qquad \text{for a.e. } x_1 \tag{5.12}$$

and S_{x_2} pushes forward the measure $\mu_0^{x_2}$ on M_{x_2} defined by

$$\mu_0^{x_2} := f_0\big(\cdot, x_2\big) \left|\det\big(e^{-A_2}\big)\right| \, dx_1$$

to the measure $\mu_1^{x_2}$ on N_{x_2} defined by

$$\mu_1^{x_2} := f_1\left(\cdot, e^{A_2} x_2\right) \, dx_1.$$

Indeed, by Fubini's Theorem, the map $x_1 \mapsto S(x_1, x_2)$ is measurable for almost all x_2 and $\mathrm{cost}_c(S) < +\infty$ implies

$$\int_{\mathbb{R}^d} c\big((x_1, x_2), S(x_1, x_2)\big) f_0(x_1, x_2) dx_1 < +\infty \qquad \text{for a.e. } x_2$$

and in turn, from (5.3), this implies, for almost every $x_2 \in \mathbb{R}^{n-d}$, that $S(x_1, x_2) \in N_{x_2}$ for almost all x_1, which means that $S(x_1, x_2)$ has the form (5.12). Now $S_\sharp \mu_0 = \mu_1$ means $\int_{R^n} h(x) d\mu_1(x) = \int_{R^n} h\big(S(x)\big) d\mu_0(x)$ for any $h \in L^1\big(\mathbb{R}^n\big)$. From Fubini's Theorem together with a change of variable $y_2 = e^{A_2} x_2$ in the left-hand side, this yields, for any positive $h \in L^1\big(\mathbb{R}^n\big)$

$$\begin{aligned} & \int_{\mathbb{R}^{n-d}} \left(\int_{\mathbb{R}^d} h(x_1, x_2) f_1(x_1, x_2) dx_1 \right) dx_2 \\ = & \int_{\mathbb{R}^{n-d}} \left(\int_{\mathbb{R}^d} h\left(\left(S_{e^{-A_2} y_2}(x_1), y_2 \right) \right) f_0\left(x_1, e^{-A_2} y_2\right) \left| \det\left(e^{-A_2}\right) \right| dx_1 \right) dy_2 \, . \end{aligned}$$

This is equivalent to $(S_{x_2})_\sharp \mu_0^{x_2} = \mu_1^{x_2}$ and proves the fact.

If S satisfies $S_\sharp \mu_0 = \mu_1$ and $\mathrm{cost}_c(S) < +\infty$, the form (5.12) implies that, for all measurable h that depends on x_2 only, one has

$$\int_{\mathbb{R}^d \times \mathbb{R}^{n-d}} h(x_2) d\mu_1(x_1, x_2) = \int_{\mathbb{R}^d \times \mathbb{R}^{n-d}} h(e^{A_2} x_2) d\mu_0(x_1, x_2),$$

and this implies (2.6). Conversely, assume now that the measures satisfy (2.6); since

$$x_2 \mapsto \int_{M_{x_2}} d\mu_0^{x_2} \quad \text{and} \quad x_2 \mapsto \int_{N_{x_2}} d\mu_1^{x_2}$$

are the densities of $(e^{A_2} \circ \pi_2)_\sharp \mu_0$ and $(\pi_2)_\sharp \mu_1$ respectively, the measures $\mu_0^{x_2}$ and $\mu_1^{x_2}$ have, for almost all x_2, the same total mass, hence can be considered as probability measures after normalisation. Then using (5.11) and arguing as in the proof of Theorem 2.2, we deduce that for almost every $x_2 \in \mathbb{R}^{n-d}$, there is a unique optimal transport map $T_{x_2} : M_{x_2} \to N_{x_2}$ together with a convex function $\varphi_{x_2} : \mathbb{R}^d \to \mathbb{R}$ such that

$$T_{x_2}\big(x_1\big) = \Big(E^{-1} \nabla \varphi_{x_2}(x_1), e^{A_2} x_2 \Big) \qquad \text{for a.e. } x_1 \in \mathbb{R}^d.$$

This shows the existence and uniqueness of an optimal transport map T, defined by $T(x_1, x_2) = (E^{-1} \nabla \varphi_{x_2}(x_1), e^{A_2} x_2)$ from μ_0 to μ_1 (and a fortiori shows (2.5)). Note that joint measurability of T with respect to the two variables may be seen as a consequence of the necessity part of [2, Theorem 3.2] which asserts that any optimal plan is concentrated on a c-monotone Borel subset of $\mathbb{R}^n \times \mathbb{R}^n$.

If we assume in addition that f_0, f_1 are both continuous and bounded from below and above on their support, then each T_{x_2} is continuous. This is a consequence of Theorem 1 applied for each x_2 to the mass transportation problem from $\mu_0^{x_2}$ to $\mu_1^{x_2}$ with respect to the cost given by (5.11). But by [15, Corollary 5.23], there is stability of transport maps. If $\{x_2^k\}$ is a sequence converging to x_2, the measures $\mu_0^{x_2^k}$ and $\mu_1^{x_2^k}$ weakly converge to $\mu_0^{x_2}$ and $\mu_1^{x_2}$ respectively, and if in addition the corresponding costs converge uniformly as well, then the transport $\{T_{x_2^k}\}$ maps converge to T_{x_2} in probability. Since all $T_{x_2}^k$ are indeed uniformly Hölder continuous (see [14, Theorem 50 (i)] or [15, Theorem 12.50]), and references therein), this concludes the proof of Theorem 2.3.

6 Examples

6.1 The case $W = 0$

If $W = 0$ then the second line in (4.5) is an independent first order differential equation in the p variable; hence $R_3(t) = 0$ for any $t \in [0, 1]$ and the form (4.1) of the cost c. If in addition, we assume that $A = 0$ and that the system is controllable, then the matrix B is necessarily invertible. In that case, we leave the reader to show that the $2n \times 2n$ matrix $R(t)$ has the form

$$R(t) = \begin{pmatrix} I_n & BU^{-1}B^* \\ 0_n & I_n \end{pmatrix}.$$

Then there holds

$$c(x, y) = \frac{1}{2} \left\langle \left(BU^{-1}B^*\right)^{-1} \Big(y - x\Big), \Big(y - x\Big) \right\rangle \qquad \forall x, y \in \mathbb{R}^n.$$

And any optimal transport map given by Theorem 2.2 has the form

$$T = \left(BU^{-1}B^*\right) \nabla\varphi,$$

where we used (4.11), with φ is a convex function. This can also be viewed as a consequence of [4] because the above cost is the Euclidean norm corresponding to the positive matrix $(BU^{-1}B^*)^{-1}$ and is $\left(BU^{-1}B^*\right) \nabla\varphi$ is nothing but the gradient of φ for this Euclidean metric.

6.2 The case $A = 0, \quad B = U = I_n$

To have a nice form of $R(t)$ let us assume that W is symmetric definite positive. In this case, we leave the reader to show that the $2n \times 2n$ matrix $R(t)$ has the form :

$$R(t) = \begin{pmatrix} cosh\ \left(tW^{\frac{1}{2}}\right) & sinh\ \left(tW^{\frac{1}{2}}\right)W^{-\frac{1}{2}} \\ W^{\frac{1}{2}}.sinh\ \left(tW^{\frac{1}{2}}\right) & cosh\ \left(tW^{\frac{1}{2}}\right) \end{pmatrix}.$$

Where

$$cosh\ \left(tW^{\frac{1}{2}}\right) = \sum_{n=0}^{\infty} \frac{t^{2n}}{2n!} W^n, \qquad sinh\ \left(tW^{\frac{1}{2}}\right) = \sum_{n=0}^{\infty} \frac{t^{2n+1}}{2n+1\,!} W^{n+\frac{1}{2}}$$

And any optimal transport map, given by Theorem 2.2, has the form

$$T = \left(sinh\ W^{\frac{1}{2}}.W^{\frac{-1}{2}}\right) \nabla\varphi,$$

where we used (4.11) and that φ is a convex function.

7 Final comments

It is worth noticing that the existence and uniqueness part in Theorem 2.2 is not new. It is a consequence of [1, Theorem 4.1] together with the Lipschitz regularity of the cost. The formula (2.3) implies that the Ma-Trudinger-Wang tensor associated with the cost is identically zero. Such a result has been obtained previously by McCann and Lee without

computing explicitly the cost (see [10, Theorem 1.1]). We finally observe that (2.4) means that the optimal transport map T is convex up to a linear change of coordinates. Such a property is related to the vanishing of the Ma-Trudinger-Wang tensor that we mentioned above and [5, Theorem 4.3]. We refer the interested reader to [7, 15] for a more details about the Ma-Trudinger-Wang tensor and its link with regularity of optimal transport maps.

References

[1] A. Agrachev and P. Lee. Optimal transportation under nonholonomic constraints. Preprint, 2007.

[2] L. Ambrosio and A. Pratelli. *Existence and stability results in the L^1 theory of optimal transportation.* In Optimal transportation and applications (Martina Franca, 2001). Lectures Notes in Math., vol 1813, pp. 123-160. Springer, Berlin, 2003.

[3] Y. Brenier. Décomposition polaire et réarrangement monotone des champs de vecteurs. *C.R. Acad. Sci. Paris, Série I*, 44:805–808, 1991.

[4] Y. Brenier. Polar factorization and monotone rearrangement of vector-valued functions. *Comm. Pure Appl. Math.*, 44:375–417, 1991.

[5] A. Figalli, Y. H. Kim and R. J. McCann. Continuity and injectivity of optimal maps for non-negatively cross-curved costs. Preprint, 2009.

[6] A. Figalli and L. Rifford. Mass Transportation on sub-Riemannian Manifolds. *Geom. Funct. Anal.*, 20(1):124–159, 2010.

[7] A. Figalli, L. Rifford and C. Villani. Necessary and sufficient conditions for continuity of optimal transport maps on Riemannian manifolds. Preprint, 2010.

[8] L. V. Kantorovich. On the transfer of masses. *Dokl. Akad. Nauk. SSSR*, 37:227–229, 1942.

[9] L. V. Kantorovich. On a problem of Monge. *Uspekhi Mat. Nauk.*, 3:225–226, 1948.

[10] P.W.Y. Lee and R.J. McCann. The Ma-Trudinger-Wang curvature for natural mechanical actions. Preprint, 2010.

[11] E. B. Lee and L. Markus. *Foundations of optimal control theory.* John Wiley & Sons Inc., New York, 1967.

[12] G. Monge. Mémoire sur la théorie des déblais et des remblais. In *Histoire de l'Académie Royale des Sciences de Paris*, pp. 666–704, 1781.

[13] E.D. Sontag. Mathematical control theory. Deterministic finite-dimensional systems. *Texts in Applied Mathematics*, Vol. 6. Springer-Verlag, New York, 1990.

[14] C. Villani. Topics in Mass Transportation. *Graduate Studies in Mathematics Surveys*, Vol. 58. American Mathematical Society, Providence, RI, 2003.

[15] C. Villani. Optimal transport, old and new. Grundlehren des mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], Vol. 338, Springer-Verlag, Berlin-New York, 2009.