



HAL
open science

From Knothe's rearrangement to Brenier's optimal transport map

Nicolas Bonnotte

► **To cite this version:**

Nicolas Bonnotte. From Knothe's rearrangement to Brenier's optimal transport map. 2012. hal-00694200v1

HAL Id: hal-00694200

<https://hal.science/hal-00694200v1>

Preprint submitted on 3 May 2012 (v1), last revised 24 Sep 2012 (v2)

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.

From Knothe's Rearrangement to Brenier's Optimal Transport Map

Nicolas BONNOTTE
Université Paris-Sud, Orsay
Scuola Normale Superiore, Pisa

April 24, 2012

Abstract

The Brenier optimal map and Knothe–Rosenblatt rearrangement are two instances of a transport map, that is to say a map sending one measure onto another. The main interest of the former is that it solves the Monge–Kantorovich optimal transport problem, while the latter is very easy to compute, being given by an explicit formula.

A few years ago, Carlier, Galichon, and Santambrogio showed that the Knothe rearrangement could be seen as the limit of the Brenier map when the quadratic cost degenerates. In this paper, we prove that on the torus (to avoid boundary issues), when all the data are smooth, the evolution is also smooth, and is entirely determined by a PDE for the Kantorovich potential (which determines the map), with a subtle initial condition. The proof requires the use of the Nash–Moser inverse function theorem.

This result generalizes the ODE discovered by Carlier, Galichon, and Santambrogio when one measure is uniform and the other is discrete, and could pave the way to new numerical methods for optimal transportation.

Key words. Optimal transport, Knothe–Rosenblatt rearrangement, continuation methods, Nash–Moser inverse function theorem.

AMS classification. 35J96, 47J07, 49K21.

Nicolas BONNOTTE
Département de mathématiques
Université Paris-Sud
91405 Orsay CEDEX
FRANCE

<http://www.normalesup.org/~bonnotte>
nicolas.bonnotte@normalesup.org

1 Introduction

Although optimal transport theory has far-reaching applications, in fields as diverse as continuum mechanics, statistics or image processing, its underlying problem is quite simple: how to send one probability measure onto another, while minimizing some cost of transportation? Let us denote by μ and ν those two measures, defined respectively on X and Y . They could for instance represent the respective distributions of the production of some goods, and the needs for them, and the problem would then be to determine how to organize the supply so that the total cost of transportation is as small as possible.

What we are looking for is a map $T : X \rightarrow Y$ telling us where to send what is in x ; but T will be suitable only if, for any measurable set $A \subset Y$, the goods sent by T in A match the needs of the same region, that is to say if $\mu(T^{-1}(A)) = \nu(A)$. If this condition is satisfied, ν is said to be the push-forward of μ by T , and we write $\nu = T\#\mu$. Let us denote by $c(x, y)$ the cost for going from x to y , then the total cost of transportation we want to minimize is

$$\int_X c(x, T(x)) \, d\mu(x). \quad (1)$$

Notice however that an optimal map may well not exist, and worse, there might even be no map transporting μ onto ν at all, e.g. if μ is discrete and ν is uniform.

The problem of finding a map T minimizing (1), and such that $\nu = T\#\mu$, was first studied by Monge [1] in the 18th century. In the 1940s, Kantorovich [2] introduced the following relaxation of Monge's problem: Instead of sending all that is in x to a unique destination $y = T(x)$, he allowed himself to split the mass. Any strategy for sending μ onto ν can then be represented by a measure γ on $X \times Y$, such that $\gamma(A \times B)$ gives the share of the goods to be moved from A to B . A plan γ is suitable if it matches the production and the needs, i.e. if

$$\gamma(A \times Y) = \mu(A), \quad \gamma(X \times B) = \nu(B).$$

This simply means that μ and ν must be the marginals of γ . Let us denote by $\Gamma(\mu, \nu)$ the set of all such suitable plans. The total cost of transportation with the plan γ is

$$\int_{X \times Y} c(x, y) \, d\gamma(x, y). \quad (2)$$

The Monge–Kantorovich problem consists in finding $\gamma \in \Gamma(\mu, \nu)$ minimizing (2). It is indeed a relaxation of the Monge problem, since if $T : X \rightarrow Y$ sends μ onto ν , then the push-forward $\gamma := (x, T(x))\#\mu$ of μ by $x \mapsto (x, T(x))$ is in $\Gamma(\mu, \nu)$, and the costs (1) and (2) are equal.

At the end of the 1980s, Brenier [3, 4] discovered the optimal transport map for the Monge problem to exist as the gradient of a convex function and to be unique, at least when $X = Y = \mathbb{R}^N$, for the cost $c(x, y) = \frac{1}{2}|x - y|^2$, if μ is absolutely continuous and if μ and ν have finite second order moments. His result was then

extended to measures defined on the torus \mathbb{T}^N by Cordero-Erausquin [5], or more generally on a Riemannian manifold by McCann [6]. While on \mathbb{R}^N the optimal map is $T(x) = \nabla\varphi(x)$ with φ convex, on the torus \mathbb{T}^N the optimal map can be written as $T(x) = x - \nabla\psi(x)$ with $\psi : \mathbb{T}^N \rightarrow \mathbb{R}^N$ such that $\varphi : x \mapsto \frac{1}{2}x^2 - \psi(x)$ defines a convex function on \mathbb{R}^N . More generally, on a Riemannian manifold $T(x) = \exp_x(-\nabla\psi(x))$ for some map ψ , called the Kantorovich potential because it is linked to a dual formulation for the relaxed problem.

Being able to compute the optimal map T , or the underlying potential ψ , is obviously of huge interest. When the measures are discrete, if there is a solution to the Monge problem, it can be obtained for instance with the auction algorithm. In the continuous case, the solution is also easy to compute in dimension 1, for if μ and ν are absolutely continuous, and if F, G stand for their respective cumulative distributions, i.e. $F(x) := \mu((-\infty, x])$ and $G(y) := \nu((-\infty, y])$, then the optimal transport map is $T = G^{-1} \circ F$.

Unfortunately when the dimension is $N > 1$, there is no such easy formula, and it is much more complicated to compute Brenier's map—although not impossible. Among the most notable methods, we could cite the one due to Benamou and Brenier [7], relying on a dynamic formulation of the Monge–Kantorovich problem, in which one tries to minimize the average kinetic energy of the particles during their transportation. On the other hand, Angenent, Haker, and Tannenbaum [8] proposed a steepest descent method, starting from a transport map (for instance, the Knothe–Rosenblatt rearrangement) and letting it evolve so as to reduce the associated transport cost. A couple of years later, Loeper and Rapetti [9] used the characterization of the optimal transport map through the existence of a convex potential, to compute Brenier's map starting from any potential and, with a Newton algorithm, altering it so as to finally get the optimal potential.

Our hope here is that the results presented in this paper might lead to yet another approach for computing Brenier's map. Our starting point is a direct connexion, proved by Carlier, Galichon, and Santambrogio [10] a few years ago but hinted beforehand by Brenier, between the optimal transport map and the Knothe–Rosenblatt rearrangement. This leads us to believe it might be possible to compute Brenier's map starting from the rearrangement (as in the paper by Angenent, Haker, and Tannenbaum [8]), and then proceeding with a continuation method (as in the work by Loeper and Rapetti [9]).

This so-called Knothe–Rosenblatt rearrangement, which is also build so as to send one measure onto another, was first introduced by Rosenblatt [11] and Knothe¹ [13]. It can be defined for absolutely continuous probability measures on \mathbb{R}^2 or on $\mathbb{T}^2 = \mathbb{R}^2/\mathbb{Z}^2$ (in higher dimension, the construction is analogous) as follows: To begin with, let us denote by f, g the densities of μ, ν . Then, take the

¹Interestingly, Knothe used this rearrangement to prove the isoperimetric inequality, for which it is well suited... but in fact not as much as Brenier's map, which Figalli, Maggi, and Pratelli [12] used more recently to prove sharp isoperimetric inequalities.

first marginals, which we denote by μ^1 and ν^1 ; their respective densities are

$$f^1(x_1) = \int f(x_1, x_2) dx_2 \quad \text{and} \quad g^1(y_1) = \int g(y_1, y_2) dy_2.$$

Define R^1 as the optimal transport map between μ^1 and ν^1 . Next, consider the disintegration of μ and ν with respect to μ^1 and ν^1 , that is to say two family of probabilities measures $\{\mu_{x_1}^2\}$ and $\{\nu_{y_1}^2\}$ such that the densities of $\mu_{x_1}^2$ and $\nu_{y_1}^2$ are

$$f_{x_1}^2(x_2) = \frac{f(x_1, x_2)}{f^1(x_1)} \quad \text{and} \quad g_{y_1}^2(y_2) = \frac{g(y_1, y_2)}{g^1(y_1)}.$$

For any x_1 let $R^2(x_1, \cdot)$ be the optimal transport map between $\mu_{x_1}^2$ and $\nu_{R^1(x_1)}^2$. Then the rearrangement is $R(x_1, x_2) := (R^1(x_1), R^2(x_1, x_2))$. It is not difficult to check that it sends μ onto ν .

What Carlier, Galichon and Santambrogio proved is that, if in Monge's problem the cost is, for instance, replaced with

$$c_t(x, y) = \frac{1}{2} \sum_{k=1}^d t^{k-1} |x_k - y_k|^2,$$

then, when the two measures are absolutely continuous, as t goes to 0, the corresponding optimal transport maps T_t converge in L^2 to the rearrangement R . When the initial measure μ is uniform and the final measure ν is discrete, $\nu = \sum a_i \delta_{y_i}$, they could also establish an ODE governing the evolution of the Kantorovich potential ψ_t , at least when the first coordinates of the y_i are distinct.

Thus, the following questions arise: in the continuous case, is it also possible to find a differential equation satisfied by ψ_t ? and if the answer is yes, is there uniqueness, that is to say, given the proper initial condition for $t = 0$, is ψ_t the only solution to this equation? As we are going to see, the answer to both question is positive, at least, to discard boundary issues, on the torus. More precisely, we have the following:

Theorem. *Let A_t be the $(1, t, \dots, t^{N-1})$ diagonal matrix, and μ, ν two probability measures on $\mathbb{T}^N = \mathbb{R}^N / \mathbb{T}^N$ with smooth, strictly positive densities f, g . The optimal transport map for the cost*

$$c_t(x, y) = \frac{1}{2} \sum_{k=1}^N t^{k-1} d(x_k, y_k)^2,$$

d standing for the usual distance on \mathbb{T}^1 , is then $T_t(x) = x - A_t^{-1} \nabla \psi_t(x)$, where the Kantorovich potential ψ_t is chosen so that $\int \psi_t = 0$. The map $t \mapsto \psi_t$ is smooth from $(0, +\infty)$ to $\mathcal{C}^\infty(\mathbb{T}^2)$, with $A_t - D^2 \psi_t > 0$ at all times, and satisfies

$$\operatorname{div} \left(f \left[I - A_t^{-1} D^2 \psi_t \right]^{-1} \left(A_t^{-1} \nabla \psi_t' - A_t^{-1} A_t' A_t^{-1} \nabla \psi_t \right) \right) = 0. \quad (3)$$

Moreover, ψ is the unique solution of (3) such that, if we write for $t \neq 0$,

$$\psi_t(x_1, \dots, x_N) = \psi_t^1(x_1) + t\psi_t^2(x_1, x_2) + \dots + t^{N-1}\psi_t^N(x_1, \dots, x_N), \quad (4)$$

with

$$\forall x_1, \dots, x_{k-1}, \quad \int \psi_t^k(x_1, \dots, x_{k-1}, z) dz = 0,$$

then $t \mapsto (\psi_t^1, \dots, \psi_t^N)$ is \mathcal{C}^2 on $[0, \infty)$, and the Knothe–Rosenblatt rearrangement $R = (R^1, \dots, R^N)$ is given by

$$R^k(x_1, \dots, x_k) = x_k - \partial_k \psi_0^k(x_1, \dots, x_k).$$

The first point is obtained by noticing that, at least when t stays away from 0, ψ_t is the unique solution to a Monge–Ampère equation $\mathcal{F}(A_t, \psi_t) = 0$, where

$$\mathcal{F}(A, u)(x) := f(x) - g\left(x - A^{-1}\nabla u(x)\right) \det\left(I - A^{-1}D^2u(x)\right),$$

is defined on a proper subset of $\mathcal{S}_N^{++} \times \mathcal{C}^2(\mathbb{T}^N)$, and then proving that we can apply the implicit function theorem. As it is well-known, the invertibility of the differential $D_u\mathcal{F}$ in (t, ψ_t) is equivalent to the existence and uniqueness of the solution to a strictly elliptic equation, so the argument is rather straightforward.

For small times, because of the degeneracy of A_t , we need the decomposition (4), which leads us to introduce another operator, namely

$$\mathcal{G}(t, u^1, u^2, \dots, u^N) := \mathcal{F}\left(A_t, \sum t^{k-1}u^k\right),$$

defined on a good subset of $[0, +\infty) \times \mathcal{C}^2(\mathbb{T}^1) \times \mathcal{C}^2(\mathbb{T}^2) \times \dots \times \mathcal{C}^2(\mathbb{T}^N)$, in such a way that $(\psi_t^1, \dots, \psi_t^N)$ is the only (u^1, \dots, u^N) such that $\mathcal{G}(t, u^1, \dots, u^N) = 0$. Unfortunately, a loss of regularity for the solutions (v^1, \dots, v^N) of the equation

$$D_u\mathcal{G}(t, \psi^1, \dots, \psi^N)(v^1, \dots, v^N) = q$$

prevents us from applying the implicit function theorem once more. We circumvent this difficulty by using the smoothness of the Kantorovich potential ψ_t , which allows us to define \mathcal{G} on a subset of $[0, +\infty) \times \mathcal{C}^\infty(\mathbb{T}^1) \times \dots \times \mathcal{C}^\infty(\mathbb{T}^N)$, so that to have an infinite source of regularity, and then by using the Nash–Moser version of the implicit function theorem.

We do not know if there is an equivalent result on \mathbb{R}^N . To be able to construct the Knothe rearrangement, compactness is required, but in \mathbb{R}^N this comes with a boundary. The problem is that the rearrangement is more easily contented with sets whose shapes are somewhat compatible with the axis, e.g. the square, but known regularity results for Brenier’s map fail to apply in that kind of setting.

Acknowledgements. This work is part of a PhD thesis supervised by Luigi Ambrosio (SNS, Pisa) and Filippo Santambrogio (Univ. Paris–Sud), whom the author would like to thank warmly for their advice and support. Financial support for this thesis is provided in part by the Franco–Italian University through the “Vinci” program. Much of this paper is also the result of an extended stay in Pisa in Fall 2011, which was made possible thanks to the ENS–SNS exchange program.

2 General quadratic costs on the torus

Given two probability measures μ, ν on the torus $\mathbb{T}^N = \mathbb{R}^N / \mathbb{Z}^N$, we want to study the evolution with t of the optimal transport map for the cost

$$c_t(x, y) = \frac{1}{2} \sum_{k=1}^N \left(\prod_{i < k} \lambda_i(t) \right) d(x_k, y_k)^2, \quad (5)$$

where $d : \mathbb{T}^1 \times \mathbb{T}^1 \rightarrow [0, +\infty)$ is the usual distance on \mathbb{T}^1 , the $\lambda_i : \mathbb{R} \rightarrow [0, +\infty)$ are smooth and such that $\lambda_i(t) = 0$ if and only if $t = 0$. For $t > 0$, this is a kind of quadratic cost on the torus. Notice that, more generally, we can define a cost given any positive-definite symmetric matrix $A \in \mathcal{S}_N^{++}$ as follow: first, consider $\tilde{c} : \mathbb{R}^N \times \mathbb{R}^N \rightarrow [0, +\infty)$ defined by

$$\tilde{c}(x, y)^2 := \inf_{k \in \mathbb{Z}^N} \frac{1}{2} A(x - y - k)^2,$$

and then take induced map $c : \mathbb{T}^N \times \mathbb{T}^N \rightarrow [0, +\infty)$. This is equivalent to changing the usual metric on \mathbb{T}^N with the one induced by A in the canonical set of coordinates, and then taking half the resulting squared distance as the cost.

An interesting property of such a cost c is that in this case the so-called c -transform of a function $u : \mathbb{T}^N \rightarrow \mathbb{R}$ is strongly connected to the Legendre transform (for the scalar product induced by A) of $x \mapsto \frac{1}{2} Ax^2 - u(x)$, defined on \mathbb{R}^N (we then see u as a periodic function on \mathbb{R}^N). Let us recall that the c -transform of u is the map $u^c : \mathbb{T}^N \rightarrow \mathbb{R}$ defined by

$$u^c(y) = \inf_{x \in \mathbb{T}^N} \{c(x, y) - u(x)\}.$$

This is interesting, because McCann [6] showed that, under suitable assumptions, the optimal transport map T can be written as $T(x) = \exp_x(-\nabla\psi(x))$, for some function ψ such that $\psi^{cc} = \psi$. A map u such that $u^{cc} = u$ is called c -concave.

Lemma 1. *A function $u : \mathbb{T}^N \rightarrow \mathbb{R}$ is c -concave if and only if*

$$v : \begin{cases} \mathbb{R}^N & \rightarrow \mathbb{R} \\ x & \mapsto \frac{1}{2} Ax^2 - u(x) \end{cases}$$

is convex and lower semi-continuous. If u is \mathcal{C}^2 and such that $A - D^2u > 0$, then $x \mapsto x - A^{-1}\nabla u(x)$ induces a diffeomorphism $\mathbb{T}^N \rightarrow \mathbb{T}^N$.

Proof. If u is c -concave, then v is convex and lower semi-continuous, for it can

be written as a Legendre transform:

$$\begin{aligned}
v(x) &= \frac{1}{2}Ax^2 - u^{cc}(x) \\
&= \frac{1}{2}Ax^2 - \inf_{y \in \mathbb{T}^N} \{c(x, y) - u^c(y)\} \\
&= \sup_{y \in \mathbb{R}^N} \left\{ \frac{1}{2}Ax^2 - \tilde{c}(x, y) + u^c(y) \right\} \\
&= \sup_{y \in \mathbb{R}^N} \sup_{k \in \mathbb{Z}^2} \left\{ \frac{1}{2}Ax^2 - \frac{1}{2}A(x - y - k)^2 + u^c(y) \right\} \\
&= \sup_{y \in \mathbb{R}^N} \left\{ \langle Ax, y \rangle - \left[\frac{1}{2}Ay^2 - u^c(y) \right] \right\}.
\end{aligned}$$

Conversely, if v is convex and lower semi-continuous, then it is equal to its double A -Legendre transform:

$$v(x) = \sup_{y \in \mathbb{R}^N} \left\{ \langle Ax, y \rangle - \sup_{z \in \mathbb{R}^N} [\langle Az, y \rangle - v(z)] \right\}.$$

Therefore,

$$\begin{aligned}
u(x) &= \frac{1}{2}Ax^2 - \sup_{y \in \mathbb{R}^N} \left\{ \langle Ax, y \rangle - \sup_{z \in \mathbb{R}^2} [\langle Az, y \rangle - v(z)] \right\} \\
&= \inf_{y \in \mathbb{R}^N} \left\{ \frac{1}{2}A(x - y)^2 - \frac{1}{2}Ay^2 + \sup_{z \in \mathbb{R}^2} [\langle Az, y \rangle - v(z)] \right\} \\
&= \inf_{y \in \mathbb{R}^N} \left\{ \frac{1}{2}A(x - y)^2 - \inf_{z \in \mathbb{R}^2} \left[\frac{1}{2}A(y - z)^2 - u(z) \right] \right\},
\end{aligned}$$

that is to say $u(x) = u^{cc}(x)$.

If u is \mathcal{C}^2 and such that $A - D^2u > 0$, then by compactness $A - D^2u \geq \varepsilon I$ for some $\varepsilon > 0$. Thus, v being convex with a super-linear growth, $\nabla v : \mathbb{R}^N \rightarrow \mathbb{R}^N$ is a diffeomorphism, and so is the map $T : x \mapsto x - A^{-1}\nabla u(x)$. Notice that, if $k \in \mathbb{Z}^N$, then $T(x + k) = T(x) + k$, therefore T induces a diffeomorphism $\mathbb{T}^N \rightarrow \mathbb{T}^N$. \square

In the next proposition, we start from the existence and uniqueness of the Kantorovich potential for such a generalized cost (this comes from McCann [6]), and then apply the results of Caffarelli [14] to get its smoothness, in the exact same way as Cordero-Erausquin [5] did. (More general results regarding the regularity of the Kantorovich potential, and thus, of the optimal transport map, on arbitrary products of spheres have been recently obtained by Figalli, Kim, and McCann [15].)

Proposition 2. *Let μ and ν be two probability measures on \mathbb{T}^N with smooth, strictly positive densities, and let c be the quadratic cost on $\mathbb{T}^N \times \mathbb{T}^N$ induced by a definite-positive symmetric matrix A .*

Then there is a unique c -concave function $\psi : \mathbb{T}^N \rightarrow \mathbb{R}$ with $\int \psi = 0$ such that $T : \mathbb{T}^N \rightarrow \mathbb{T}^2$ defined by $T(x) := x - A^{-1}\nabla\psi(x)$ sends μ onto ν .

The function ψ is a Kantorovich potential, it is smooth, and the application $\varphi : x \mapsto \frac{1}{2}Ax^2 - \psi(x)$ is a smooth, strictly convex function on \mathbb{R}^N .

The transport map T is optimal for the cost c . There is no other optimal transport plan but the one it induces.

Of course in this proposition, instead of $T(x) := x - A^{-1}\nabla\psi(x)$ we should have written $T(x) = x - \pi(A^{-1}\nabla\psi(x))$, where $\pi : \mathbb{R}^N \rightarrow \mathbb{T}^N$ is the usual projection.

Proof. Let us denote by ∇_A the gradient for the metric induced by A . Then according to McCann [6], there is a Lipschitz function $\psi : \mathbb{T}^N \rightarrow \mathbb{R}$ that is c -concave and such that $T : x \mapsto \exp_x[-\nabla_A\psi(x)]$ pushes μ forward to ν . It is uniquely defined if the condition $\int \psi(x) dx = 0$ is added, and moreover it is optimal for the Monge–Kantorovich problem. Notice that here on the torus, $\exp_x[-\nabla_A\psi(x)] = x - A^{-1}\nabla\psi(x)$.

For any $x \in \mathbb{R}^N$, let $\varphi(x) := \frac{1}{2}Ax^2 - \psi(x)$. Then $T(x) = A^{-1}\nabla\varphi(x)$ sends μ onto ν , seen as periodic measures on \mathbb{R}^N . Moreover, according to Lemma 1, φ is a convex function. Now, let V be a open, convex subset of \mathbb{R}^N , and define $U = (\nabla\varphi)^{-1}(V)$; then $\nabla\varphi$ sends $\mu|_U$ onto $A\#\nu|_V$, and both measures are still absolutely continuous with smooth, bounded, strictly positive densities. Therefore we are entitled to apply the results of Caffarelli [14], and thus we get that φ is strictly convex and smooth on U . As U is arbitrary, φ is strictly convex and smooth on \mathbb{R}^2 . Thus, ψ is also smooth, and T is a diffeomorphism. \square

3 PDE satisfied for positive times

Let μ and ν be two probability measures on \mathbb{T}^N with smooth, strictly positive densities f and g . According to Proposition 2, for any $A \in \mathcal{S}_N^{++}$, we have a smooth Kantorovich potential $\Psi_A : \mathbb{T}^N \rightarrow \mathbb{R}$. What can we say of the regularity of $\Psi : A \mapsto \Psi_A$?

As $x \mapsto x - A^{-1}\nabla\Psi_A(x)$ sends μ onto ν , the following Monge–Ampère equation is satisfied:

$$f(x) = g\left(x - A^{-1}\nabla\Psi_A(x)\right) \det\left(I - A^{-1}D^2\Psi_A\right).$$

For $u \in \mathcal{C}^2(\mathbb{T}^N)$ such that $A - D^2u > 0$ and $\int u = 0$, we set

$$\mathcal{F}(A, u) = f - g\left(\text{id} - A^{-1}\nabla u\right) \det\left(I - A^{-1}D^2u\right).$$

Thanks to the characterization of c -concave functions from Lemma 1, and to Proposition 2, we have

Lemma 3. *For any $u \in \mathcal{C}^2(\mathbb{T}^N)$ such that $A - D^2u > 0$ and $\int u = 0$, we can have $\mathcal{F}(A, u) = 0$ if and only if $u = \Psi_A$.*

We are now going to prove that we can apply the implicit function theorem.

In the following, for any function space X we denote with a \diamond subscript the space formed by the elements of X having a zero mean value, e.g. \mathcal{C}_\diamond^2 is the space of all $u \in \mathcal{C}^2$ such that $\int u = 0$.

Lemma 4. *The operator \mathcal{F} is smooth. For any $A \in \mathcal{S}_N^{++}$, if $u \in \mathcal{C}_\diamond^2(\mathbb{T}^N)$ is such that $A - D^2u > 0$, if $v \in \mathcal{C}_\diamond^2(\mathbb{T}^2)$, then*

$$\begin{aligned} D_u \mathcal{F}(A, u)v &= \operatorname{div} \left((f - \mathcal{F}(A, u)) [A - D^2u]^{-1} \nabla v \right) \\ &= \frac{1}{\det A} \operatorname{div} \left(g \left(\operatorname{id} - A^{-1} \nabla u \right) \left[\operatorname{Co} \left(A - D^2u \right) \right]^T \nabla v \right). \end{aligned}$$

We denote by M^T the transposed matrix of M , and by $\operatorname{Co}(M)$ its comatrix, that is to say the matrix formed by the cofactors.

Proof. The smoothness of \mathcal{F} is obvious. By substitution, for any $\xi \in \mathcal{C}^\infty$,

$$\int \xi \left(x - A^{-1} \nabla u(x) \right) [f(x) - \mathcal{F}(A, u)(x)] \, dx = \int \xi(y) g(y) \, dy.$$

Therefore, if we conveniently set $T_A u(x) := x - A^{-1} \nabla u(x)$ and differentiate the previous equation with respect to u along the direction v , we get

$$- \int \left\langle \nabla \xi(T_A u), A^{-1} \nabla v \right\rangle (f - \mathcal{F}(A, u)) - \int \xi(T_A u) D_u \mathcal{F}(A, u)v = 0.$$

Since $\nabla[\xi \circ T_A u] = [DT_A u]^T \nabla \xi(T_A u)$, we have

$$\begin{aligned} \left\langle \nabla \xi(T_A u), A^{-1} \nabla v \right\rangle &= \left\langle \nabla[\xi \circ T_A u], [DT_A u]^{-1} A^{-1} \nabla v \right\rangle \\ &= \left\langle \nabla[\xi \circ T_A u], [I - A^{-1} D^2 u]^{-1} A^{-1} \nabla v \right\rangle, \end{aligned}$$

and this yields

$$\int \xi(T_A u) D_u \mathcal{F}(A, \psi_A)v = \int \xi(T_A u) \operatorname{div} \left([I - A^{-1} D^2 u]^{-1} A^{-1} \nabla v \right),$$

and thus, since $\xi \circ T_A u$ is arbitrary, we get the first equality. Then, we can easily obtain the second expression using the formula $M^{-1} = [\operatorname{Co} M]^T / \det(M)$. \square

Lemma 5. *Let $\varepsilon > 0$ and $A \in \mathcal{S}_N^{++}$. If $u \in \mathcal{C}_\diamond^2(\mathbb{T}^N)$ is such that $A - D^2u > \varepsilon$, then for any $q \in [H_\diamond^1(\mathbb{T}^N)]^*$, there is a unique $v \in H_\diamond^1(\mathbb{T}^N)$ such that*

$$D_u \mathcal{F}(A, u)v = q. \tag{6}$$

Moreover, $\|v\|_{H^1} \leq C_\varepsilon \|q\|_{(H_\diamond^1)^*}$.

Proof. As $A - D^2u > \varepsilon$, the lowest eigenvalue of $\text{Co}(A - D^2u)$ is bounded by ε^{N-1} . Since $g > \delta$ for some $\delta > 0$, for any $\xi \in \mathcal{C}^\infty(\mathbb{T}^N)$,

$$\begin{aligned} \varepsilon^{N-1} \int |\nabla \xi|^2 &\leq \int \langle [\text{Co}(A - D^2u)]^T \nabla \xi, \nabla \xi \rangle \\ &\leq \frac{1}{\delta} \int g (\text{id} - A^{-1} \nabla u) \langle [\text{Co}(A - D^2u)]^T \nabla \xi, \nabla \xi \rangle, \end{aligned}$$

and thus

$$\int |\nabla \xi|^2 \leq -\frac{1}{\delta \varepsilon^{N-1}} \int \xi D_u \mathcal{F}(A, u) \xi. \quad (7)$$

Therefore, thanks to the existence of a Poincaré inequality on $H_\diamond^1(\mathbb{T}^N)$, the map $(\xi, \eta) \mapsto \int \eta D_u \mathcal{F}(A, u) \xi$ induces a coercive, continuous bilinear form on H_\diamond^1 . We are thus entitled to apply the Lax–Milgram theorem, which yields the existence and the uniqueness, for every $q \in (H_\diamond^1)^*$, of a $v \in H_\diamond^1$ satisfying (6). Moreover, (7) immediately gives us $\|v\|_{H^1} \leq \frac{1}{\delta \varepsilon^{N-1}} \|q\|_{(H_\diamond^1)^*}$. \square

The regularity of the solutions to an elliptic equation is well known. However, as in the following we will need some very precise estimates to apply the Nash–Moser theorem, let us give a proof of the following result:

Lemma 6. *Under the same assumptions, and with the same notations, for any $n \geq 1$, if $u \in \mathcal{C}_\diamond^{n+2}$ and $q \in H_\diamond^{n-1}$ satisfy $\|u\|_{\mathcal{C}^3} + \|q\|_{(H_\diamond^1)^*} \leq M$, then $v \in H_\diamond^{n+1}$, and*

$$\|v\|_{H^{n+1}} \leq C_{\varepsilon, M, n} \{ \|q\|_{H^{n-1}} + \|u\|_{\mathcal{C}^{n+2}} \}. \quad (8)$$

Proof. We proceed by induction. Let $n \geq 1$, $u \in \mathcal{C}_\diamond^{n+2}$ and $q \in H_\diamond^n$ such that $A - D^2u > \varepsilon$ and $\|u\|_{\mathcal{C}^3} + \|q\|_{(H_\diamond^1)^*} \leq M$. We assume that we already know that the corresponding solution v is in H_\diamond^n , and that

$$\|v\|_{H^n} \leq C_{\varepsilon, M, n-1} \{ \|q\|_{H^{n-2}} + \|u\|_{\mathcal{C}^{n+1}} \}. \quad (9)$$

Notice that we do have such an inequality for $n = 1$, according to the previous lemma, but with $\|q\|_{(H_\diamond^1)^*}$ instead of $\|q\|_{H^{-1}}$. Let us now show that it implies $v \in H_\diamond^{n+1}$ and

$$\|v\|_{H^{n+1}} \leq C_{\varepsilon, M, n} \{ \|q\|_{H^{n-1}} + \|u\|_{\mathcal{C}^{n+2}} \}.$$

First, we set $B_A u := (f - \mathcal{F}(A, u))[A - D^2u]^{-1}$, so that Equation (6) becomes

$$D_u \mathcal{F}(A, u) v = \text{div}(B_A u \nabla v). \quad (10)$$

Then, for $h \in \mathbb{R}^2$ and $\xi \in H^1$, we also define

$$\tau_h \xi(x) := \xi(x + h) \quad \text{and} \quad \delta_h \xi(x) := \frac{\xi(x + h) - \xi(x)}{h}.$$

Notice then that $\delta_h(\eta \xi) = \eta \delta_h \xi + (\delta_h \eta) \tau_h \xi$, and $\|\delta_h \xi\|_{L^2} \leq \|\xi\|_{H^1}$.

Let $\nu \in \mathbb{N}^2$ be a 2-index, with $|\nu| := \nu_1 + \nu_2 = n - 1$, and let $h \in \mathbb{R}^2$ be small enough. We can apply the operator δ_h to Equation (10), and we then obtain

$$\operatorname{div}(B_A u \nabla \delta_h v) = \delta_h q - \operatorname{div}[(\delta_h B_A u) \nabla \tau_h v]$$

Then, by differentiating this with respect to ν , we get

$$\begin{aligned} \operatorname{div}(B_A u \nabla \delta_h \partial_\nu v) &= \delta_h \partial_\nu q - \sum_{0 \leq \alpha \leq \nu} \binom{\nu}{\alpha} \operatorname{div}[(\delta_h \partial_{\nu-\alpha} B_A u) \nabla \tau_h \partial_\alpha v] \\ &\quad - \sum_{0 \leq \alpha < \nu} \binom{\nu}{\alpha} \operatorname{div}[(\partial_{\nu-\alpha} B_A u) \nabla \delta_h \partial_\alpha v]. \end{aligned} \quad (11)$$

Now, Lemma 5 tells us that this implies

$$\begin{aligned} \|\delta_h \partial_\nu v\|_{H^1} &\leq C_\varepsilon \|\delta_h \partial_\nu q\|_{(H_\delta^1)^*} \\ &\quad + C_\varepsilon \sum_{0 \leq \alpha \leq \nu} \binom{\nu}{\alpha} \|\operatorname{div}[(\delta_h \partial_{\nu-\alpha} B_A u) \nabla \tau_h \partial_\alpha v]\|_{(H_\delta^1)^*} \\ &\quad + C_\varepsilon \sum_{0 \leq \alpha < \nu} \binom{\nu}{\alpha} \|\operatorname{div}[(\partial_{\nu-\alpha} B_A u) \nabla \delta_h \partial_\alpha v]\|_{(H_\delta^1)^*}. \end{aligned}$$

Since this boundary is uniform in h , it is enough to ensure $v \in H^{n+1}$ and

$$\|v\|_{H^{n+1}} \leq C \left\{ \|q\|_{H^{n-1}} + \sum_{0 \leq k \leq n-1} (1 + \|u\|_{\mathcal{C}^{n-k+2}}) \|v\|_{H^{k+1}} \right\}.$$

Notice that, when $n > 1$, the following Landau–Kolmogorov inequalities hold

$$\begin{aligned} \|u\|_{\mathcal{C}^{n-k+2}} &\leq C_{k,n} \|u\|_{\mathcal{C}^3}^{1-\frac{k}{n-1}} \|u\|_{\mathcal{C}^{n+2}}^{\frac{k}{n-1}}, \\ \|v\|_{H^{k+1}} &\leq C_{k,n} \|v\|_{H^1}^{\frac{k}{n-1}} \|v\|_{H^n}^{1-\frac{k}{n-1}}. \end{aligned}$$

They are quite classical and can be easily proved by induction from

$$\|\xi\|_{\mathcal{C}^1} \leq \sqrt{2\|\xi\|_{\mathcal{C}^0}\|\xi\|_{\mathcal{C}^2}} \quad \text{and} \quad \|\xi\|_{H^1} \leq \sqrt{\|\xi\|_{L^2}\|\xi\|_{H^2}},$$

for ξ smooth enough satisfying $\int \xi = 0$. Since $a^{1-t}b^t \leq (1-t)a + tb$, we get

$$\|u\|_{\mathcal{C}^{n-k+2}} \|v\|_{H^{k+1}} \leq \frac{k}{n-1} \|u\|_{\mathcal{C}^3} \|v\|_{H^n} + \left(1 - \frac{k}{n-1}\right) \|u\|_{\mathcal{C}^{n+2}} \|v\|_{H^1},$$

and therefore

$$\|v\|_{H^{n+1}} \leq C \{ \|q\|_{H^{n-1}} + (1 + \|u\|_{\mathcal{C}^3}) \|v\|_{H^n} + \|u\|_{\mathcal{C}^{n+2}} \|v\|_{H^1} \}.$$

This last inequality still holds when $n = 1$. In any case, as $\|v\|_{H^1} \leq C_\varepsilon \|q\|_{H^{-1}}$ and $\|u\|_{\mathcal{C}^3} + \|q\|_{H^{-1}} \leq M$, using our assumption (9), we get

$$\|v\|_{H^{n+1}} \leq C_{\varepsilon, M, n} \{ \|q\|_{H^{n-1}} + \|u\|_{\mathcal{C}^{n+2}} \}.$$

This is exactly what we wanted. \square

Lemma 7. *Under the same assumptions, for any $q \in \mathcal{C}_\diamond^{n,\alpha}(\mathbb{T}^N)$, there is a unique $v \in \mathcal{C}_\diamond^{n+2,\alpha}(\mathbb{T}^N)$ such that*

$$D_u \mathcal{F}(A, u)v = q.$$

Proof. If $q \in \mathcal{C}_\diamond^{n,\alpha}$, then $q \in H_\diamond^n$, and thus according to the previous lemmas, there is $v \in H_\diamond^{n+2}$ such that $D_u \mathcal{F}(A, u)v = q$ in $(H_\diamond^1)^*$. But since $\int q = 0$, given the particular form of $D_u \mathcal{F}(A, u)v$ given by Lemma 4, such an equality in fact holds in H^{-1} . Thus, locally, in a weak sense,

$$D_u \mathcal{F}(A, u)v = q.$$

Then, we can locally use the theory of regularity for the solutions to a strictly elliptic equation in \mathbb{R}^N to get existence and uniqueness of $v \in \mathcal{C}^{n,\alpha}$ (cf. for instance Gilbarg & Trudinger [16], Chapter 6). \square

Theorem 8. *For any $A \in \mathcal{S}_N^{++}$, let Ψ_A be the Kantorovich potential between the probability measure μ and ν , which are still assumed to have smooth, strictly positive densities. Then, for any $n \geq 0$ and $\alpha \in (0, 1)$, the following map*

$$\Psi : \begin{cases} \mathcal{S}_N^{++} & \longrightarrow & \mathcal{C}^{n+2,\alpha}(\mathbb{T}^N) \\ A & \longmapsto & \Psi_A \end{cases} \text{ is } \mathcal{C}^1.$$

Proof. We denote by Ω be the set of all $(A, u) \in \mathcal{S}_N^{++} \times \mathcal{C}_\diamond^{n+2,\alpha}(\mathbb{T}^N)$ such that $A - D^2u > 0$. Then Ω is open, the operator $\mathcal{F} : \Omega \rightarrow \mathcal{C}_\diamond^{n,\alpha}(\mathbb{T}^N)$, defined by

$$\mathcal{F}(A, u) = f - g \left(\text{id} - A^{-1} \nabla u \right) \det \left(I - A^{-1} D^2 u \right),$$

is smooth and, according Lemma 7, $D_u \mathcal{F}(A, \psi_A) : \mathcal{C}_\diamond^{n+2,\alpha}(\mathbb{T}^N) \rightarrow \mathcal{C}_\diamond^{n,\alpha}(\mathbb{T}^N)$ is a bijection. From the Banach–Schauder theorem, we deduce it is an isomorphism. Since $\mathcal{F}(A, \Psi_A) = 0$, according to the implicit function theorem, there is a \mathcal{C}^1 map Φ defined in a neighborhood of A such that $B - D^2\Phi_B > 0$ and, for any $u \in \mathcal{C}_\diamond^{n+2,\alpha}$, $\mathcal{F}(B, u) = 0$ if and only if $u = \Phi_B$. According to Lemma 3, it implies $\Phi_B = \Psi_B$. Thus, globally, $\Psi = \Phi$ is a \mathcal{C}^1 map $\mathcal{S}_N^{++} \rightarrow \mathcal{C}_\diamond^{n,\alpha}(\mathbb{T}^N)$. \square

We are now going to apply this result to the cost c defined by (5), that is to say the cost induced by the matrix

$$A_t := \begin{pmatrix} 1 & & & & \\ & \lambda_1(t) & & & \\ & & \lambda_1(t)\lambda_2(t) & & \\ & & & \ddots & \\ & & & & \prod \lambda_i(t) \end{pmatrix},$$

where $\lambda_1, \dots, \lambda_{N-1} : \mathbb{R} \rightarrow [0, +\infty)$ are assumed to be such that $\lambda_k = 0$ if and only if $t = 0$.

Theorem 9. *If $\lambda_1, \dots, \lambda_{N-1}$ are smooth, the map $\psi : t \mapsto \Psi_{A_t}$ is \mathcal{C}^1 , and satisfies:*

$$\operatorname{div} \left\{ f \left[A_t - D^2 \psi_t \right]^{-1} \left(\nabla \psi'_t - A'_t A_t^{-1} \nabla \psi_t \right) \right\} = 0. \quad (12)$$

Moreover, if $u : (0, +\infty) \rightarrow \mathcal{C}^{n+2, \alpha}(\mathbb{T}^N)$ is \mathcal{C}^1 and satisfies, for all $t \in (0, +\infty)$,

$$A_t - D^2 u > 0 \quad \text{and} \quad \operatorname{div} \left\{ f \left[A_t - D^2 u_t \right]^{-1} \left(\nabla u'_t - A'_t A_t^{-1} \nabla u_t \right) \right\} = 0, \quad (13)$$

and if $u_{t_0} = \psi_{t_0}$ for some $t_0 > 0$, then $u = \psi$.

Proof. If $\psi_t := \Psi_{A_t}$, for all $t > 0$, we have $\mathcal{F}(A_t, \psi_t) = 0$. If we differentiate with respect to t , we get

$$D_u \mathcal{F}(A_t, \psi_t) \psi'_t + D_A \mathcal{F}(A_t, \psi_t) A'_t = 0.$$

We have seen in Lemma 4 that

$$D_u \mathcal{F}(A_t, \psi_t) \psi'_t = \operatorname{div} \left(f \left[A_t - D^2 \psi_t \right]^{-1} \nabla \psi'_t \right).$$

On the other hand,

$$D_A \mathcal{F}(A_t, \psi_t) A'_t = - \operatorname{div} \left(f \left[A_t - D^2 \psi_t \right]^{-1} A'_t A^{-1} \nabla \psi_t \right).$$

We thus get (12).

If $u : (0, +\infty) \rightarrow \mathcal{C}^{n+2, \alpha}(\mathbb{T}^N)$ is \mathcal{C}^1 and satisfies (13), with $u_{t_0} = \psi_{t_0}$ for some $t_0 > 0$, then $\mathcal{F}(A_t, u_t)$ must be constant and equal to $\mathcal{F}(A_{T_0}, u_{t_0}) = 0$. Thus, according to Lemma 3, $u_t = \Psi_{A_t}$. \square

4 Initial condition in dimension 2

Due to the very technical nature of the proofs, the following sections will only deal with the dimension 2, to ease the exposition. In the last section, we shall explain what changes in higher dimensions.

Let $\lambda : \mathbb{R} \rightarrow [0, +\infty)$ be a smooth function such that $\lambda_t = 0$ if and only if $t = 0$. From now on, we will only consider the cost induced by

$$A_t = \begin{pmatrix} 1 & 0 \\ 0 & \lambda_t \end{pmatrix}.$$

For $t \neq 0$, let ψ_t be the associated Kantorovich potential between the probability measures μ, ν , assuming they have the same properties as before (that is, strictly positive and smooth densities), and let T_t be the corresponding optimal transport map. Then, according to Theorem 9, $t \mapsto \psi_t$ and $t \mapsto T_t$ are \mathcal{C}^1 on $\mathbb{R} \setminus \{0\}$. Moreover, Carlier, Galichon, and Santambrogio [10] proved:

Theorem 10 (C.-G.-S.). *As $t \rightarrow 0$, the map T_t converges to the Knothe–Rosenblatt rearrangement R in $L^2(\mathbb{T}^2, \mu; \mathbb{T}^2)$.*

Let us denote by ψ_0^1 and $\psi_0^2(x_1, \cdot)$ the Kantorovich potentials for respectively R^1 and $R^2(x_1, \cdot)$. Indeed, recall that R^1 sends the first part μ^1 of the disintegration of μ onto the first part ν^1 of the disintegration of ν , and that $R^2(x_1, \cdot)$ sends the second part $\mu_{x_2}^1$ onto $\nu_{R^1(x_1)}^2$, in an optimal way for the squared distance on the 1-dimensional torus \mathbb{T}^1 ; hence these transport maps come from some potentials. We have:

$$R(x) = \begin{pmatrix} x_1 - \partial_1 \psi_0^1(x_1) \\ x_2 - \partial_2 \psi_0^2(x_1, x_2) \end{pmatrix} =: x - \partial \psi_0(x).$$

The Carlier–Galichon–Santambrogio theorem suggests some connexion exists between ψ_t and (ψ_0^1, ψ_0^2) . Since $T_t = \text{id} - A_t^{-1} \nabla \psi_t$, let us follow our instinct and set

$$\psi_t(x_1, x_2) = \psi_t^1(x_1) + \lambda_t \psi_t^2(x_1, x_2),$$

and, to ensure uniqueness, require

$$\int \psi_t^1(x_1) dx_1 = 0 \quad \text{and} \quad \int \psi_t^2(x_1, x_2) dx_2 = 0.$$

Notice that ψ_t^1 and ψ_t^2 are then uniquely determined, and are smooth, since

$$\psi_t^1(x_1) = \int \psi_t(x_1, x_2) dx_2 \quad \text{and} \quad \psi_t^2(x) = \frac{1}{\lambda_t} \left(\psi_t(x) - \psi_t^1(x_1) \right).$$

Let us denote by E the set of all $(t, u^1, u^2) \in \mathbb{R} \times \mathcal{C}^\infty(\mathbb{T}^1) \times \mathcal{C}^\infty(\mathbb{T}^2)$ such that

$$\int u^1(x_1) dx_1 = 0 \quad \text{and} \quad \int u^2(x_1, x_2) dx_2 = 0,$$

and by Ω the open subset of E formed by the (t, u^1, u^2) such that:

- either $t \neq 0$, and then $A_t - D^2(u^1 + \lambda_t u^2) > 0$;
- or $t = 0$, and then $1 - \partial_{1,1} u^1 > 0$ and $1 - \partial_{2,2} u^2 > 0$.

Then, thanks to Lemma 1, we can define an operator $\mathcal{G} : \Omega \rightarrow \mathcal{C}^\infty(\mathbb{T}^2)$ by setting, when $t \neq 0$,

$$\mathcal{G}(t, u^1, u^2) := \mathcal{F}(A_t, u^1 + \lambda_t u^2),$$

where \mathcal{F} is the operator introduced in Section 3:

$$\mathcal{F}(A, u) = f - g \left(\text{id} - A^{-1} \nabla u \right) \det \left(I - A^{-1} D^2 u \right).$$

Since, according to Lemma 3, $\mathcal{F}(A_t, u_t) = 0$ if and only if $u_t = \psi_t$, we have

Lemma 11. For any $(t, u^1, u^2) \in \Omega$, $\mathcal{G}(t, u^1, u^2) = 0$ if and only if $u^1 = \psi_t^1$ and $u^2 = \psi_t^2$.

Now, we are going to extend \mathcal{G} for $t = 0$. Notice indeed that

$$A^{-1}\nabla(u^1 + \lambda_t u^2) = \begin{pmatrix} \partial_1 u^1 + \lambda_t \partial_1 u^2 \\ \partial_2 u^2 \end{pmatrix}$$

and

$$A^{-1}D^2(u^1 + \lambda_t u^2) = \begin{pmatrix} \partial_{1,1} u^1 + \lambda_t \partial_{1,1} u^2 & \lambda_t \partial_{1,2} u^2 \\ \partial_{1,2} u^2 & \partial_{2,2} u^2 \end{pmatrix},$$

therefore we can smoothly extend \mathcal{G} . If we conveniently set $\partial u := (\partial_1 u^1, \partial_2 u^2)$, then $R = \text{id} - \partial\psi_0$, and

$$\mathcal{G}(0, u^1, u^2) = f - g(\text{id} - \partial u) \det(I - D\partial u). \quad (14)$$

Alas, we cannot do the same as in the previous section and apply the implicit function theorem, for if we solve $D_u \mathcal{G}(0, \psi_0^1, \psi_0^2)(v^1, v^2) = q$, then *a priori* the solution v^2 is not smooth enough. Indeed, as we will see later, if $q \in H^n$, then $v^1 \in H^{n+2}$, but we can only get $v^2 \in H^n$. However, we can circumvent this difficulty by considering \mathcal{C}^∞ functions, so as to have an infinite source of smoothness, and use the Nash–Moser implicit function theorem instead of the “classical” implicit function theorem.

Before stating our next result, let us recall some definitions from the Nash–Moser theory. For more details, see for instance Hamilton [17].

Let X and Y be two Fréchet spaces, endowed each one with a family of increasingly stronger semi-norms, denoted by $\{\|\cdot\|_n^X\}_{n \geq 0}$ and $\{\|\cdot\|_n^Y\}_{n \geq 0}$. For instance, you can think of $\mathcal{C}^\infty(\mathbb{T}^2)$, with the norms $\|\cdot\|_n = \|\cdot\|_{\mathcal{C}^n}$ or equivalently $\|\cdot\|_n = \|\cdot\|_{H^n}$. A map $\varphi : U \rightarrow Y$ is said to be “tame” if it is defined on an open set $U \subset X$, if it is continuous, and if, in a neighborhood V of every point, one can find $r \geq 0$, $b \geq 0$ and a sequence $(C_n)_{n \geq b}$ of positive constants such that the following “tame estimate” is satisfied:

$$\forall x \in V, \forall n \geq b, \|\varphi(x)\|_n^Y \leq C_n \left(1 + \|x\|_{n+r}^X\right).$$

Notice that r, b, C_n can depend on V , but V cannot change with n . The map φ is “smooth tame” if it is smooth and if all its Gâteaux derivative $D^k \varphi : U \times X \rightarrow Y$ are tame.

The Nash–Moser theorem holds for some Fréchet spaces, the so-called “tame spaces” defined as follows. If E is a Banach space, the space of exponentially decreasing sequences in E is defined as:

$$\Sigma(E) := \left\{ (u_n) \in E^{\mathbb{N}} ; \quad \forall n \in \mathbb{N}, \quad \|u\|_n^{\Sigma(E)} := \sum_{k=0}^{\infty} e^{nk} \|u_k\|^E < \infty \right\}$$

A Fréchet space X is said to be “tame” if there is a Banach space E and two tame linear maps $\Phi : X \rightarrow \Sigma(E)$ and $\Psi : \Sigma(E) \rightarrow X$ such that $\Psi \circ \Phi = \text{id}_X$. For instance, $\mathcal{C}^\infty(\mathbb{T}^2)$ is a tame space. If X and Y are tame, then so is their cartesian product $X \times Y$.

Theorem 12 (Nash–Moser). *Let X and Y be two tame spaces. Let $U \subset X$ be an open subset and $\Phi : U \rightarrow Y$ be a smooth tame map. We assume that, for any $u \in U$ and any $q \in Y$, there is a unique $v \in X$ such that $D\Phi(u)v = q$. If the inverse operator $\mathcal{S} : U \times Y \rightarrow X$ is a smooth tame map, then Φ is locally invertible, and the local inverse maps are smooth tame.*

Corollary 13 (implicit function). *Let X, Y, Z be three tame spaces, and $U \subset X$ and $V \subset Y$ be open subsets. We assume $\Phi : U \times V \rightarrow Z$ is a smooth tame map such that $\Phi(u_0, v_0) = 0$ for some $(u_0, v_0) \in U \times V$. If, for any $u \in U$, $v \in V$ and $q \in Z$, there is a unique $w \in Y$ such that $D_v\Phi(u, v)w = q$, and if the inverse operator $\mathcal{S} : U \times V \times Z \rightarrow Y$ is a smooth tame map, then there is a smooth tame map ψ defined in a neighborhood of u_0 and taking values in a neighborhood of v_0 such that $\Phi(u, \psi(u)) = 0$ if and only if $v = \psi(u)$.*

Recall that we denote with a \diamond subscript the sets of maps with zero mean value: $\mathcal{C}_{\diamond}^\infty$ is thus the set formed by the smooth functions u such that $\int u = 0$. When X is a 2-variable function space, we also denote by a “ $*, \diamond$ ” subscript, as in $\mathcal{C}_{*, \diamond}^\infty(\mathbb{T}^2)$ the set formed by the $\xi \in X$ such that $\int \xi(\cdot, x_2) dx_2 = 0$.

Theorem 14. *For all $(t, u^1, u^2) \in \Omega$, for any $q \in \mathcal{C}_{\diamond}^\infty(\mathbb{T}^2)$, there is a unique $(v^1, v^2) \in \mathcal{C}_{\diamond}^\infty(\mathbb{T}^1) \times \mathcal{C}_{*, \diamond}^\infty(\mathbb{T}^2)$ such that*

$$D_u \mathcal{G}(t, u^1, u^2)(v^1, v^2) = q, \quad (15)$$

Moreover, the inverse operator

$$\mathcal{S} : \begin{cases} \Omega \times \mathcal{C}_{\diamond}^\infty(\mathbb{T}^2) & \rightarrow \mathcal{C}_{\diamond}^\infty(\mathbb{T}^1) \times \mathcal{C}_{*, \diamond}^\infty(\mathbb{T}^2) \\ ((t, u^1, u^2), q) & \mapsto (v^1, v^2) \end{cases}$$

is smooth tame.

Proof. We will show the existence of (v^1, v^2) in Section 5. We also report the proof of the existence of a tame estimate for the inverse operator \mathcal{S} to Section 6.

Let us now show \mathcal{S} is continuous.

If $(t_k, u_k^1, u_k^2, q_k) \in \Omega$ converges towards $(t, u^1, u^2, q) \in \Omega$, for each k let (v_k^1, v_k^2) be the corresponding inverse. Thanks to the tame estimate from Section 6, v_k^1 and v_k^2 are bounded in all the spaces H^n . Hence, compact embeddings provide convergence, up to an extraction, to some v^1, v^2 as strongly as we want, which, since $D\mathcal{G}$ is continuous, must be the solution of $D\mathcal{G}(t, u^1, u^2)(v^1, v^2) = q$.

All the derivative $D^k \mathcal{S}$ are also tame, since they basically give the solution to the same kind of equation as (15). Indeed, by differentiating (15), we get

$$D_u \mathcal{G} D \mathcal{S} = Dq - D(D_u \mathcal{G}),$$

and then we can apply the results of Section 6 once more. \square

Corollary 15. The map $\begin{cases} \mathbb{R} & \rightarrow \mathcal{C}_\diamond^\infty(\mathbb{T}^1) \times \mathcal{C}_{*,\diamond}^\infty(\mathbb{T}^2) \\ t & \mapsto (\psi_t^1, \psi_t^2) \end{cases}$ is smooth.

Proof. This is a consequence of Corollary 13, Theorem 14, and Lemma 11. \square

Theorem 16. The curve in $\mathcal{C}_\diamond^\infty(\mathbb{T}^2)$ formed by the Kantorovich potentials (ψ_t) is the only smooth curve defined on \mathbb{R} such that, for $t \neq 0$,

$$A_t - D^2\psi_t > 0 \text{ and } \operatorname{div} \left(f [A_t - D^2\psi_t]^{-1} (\nabla\dot{\psi}_t - \dot{A}_t A_t^{-1} \nabla\psi_t) \right) = 0, \quad (16)$$

and that can be decomposed into two smooth curves (ψ_t^1) and (ψ_t^2) such that

$$\psi_t(x_1, x_2) = \psi_t^1(x_1) + \lambda_t \psi_t^2(x_1, x_2),$$

with ψ_0^1 and ψ_0^2 being the Kantorovich potentials for the Knothe rearrangement.

Proof. Let $u_t = u_t^1 + \lambda_t u_t^2$ be such a curve, and let us check that $u_t = \psi_t$. Since u_0^1 and u_0^2 are the Kantorovich potentials for the Knothe rearrangement, $(0, u_0^1, u_0^2) \in \Omega$, and as $A_t - D^2\psi_t > 0$ for $t \neq 0$, (t, u_t^1, u_t^2) is always in Ω . For $t \neq 0$, (16) is equivalent to

$$D_u \mathcal{F}(t, u_t) \dot{u}_t + D_t \mathcal{F}(t, u_t) = 0,$$

and therefore

$$D_u \mathcal{G}(t, u_t^1, u_t^2) (\dot{u}_t^1, \dot{u}_t^2) + D_t \mathcal{G}(t, u_t^1, u_t^2) = 0.$$

By assumption, $\mathcal{G}(0, u_0^1, u_0^2) = 0$. Integrating in time, we get $\mathcal{G}(t, u^1, u^2) = 0$. Therefore, according to Lemma 11, $u_t^1 = \psi_t^1$ and $u_t^2 = \psi_t^2$, i.e. $u_t = \psi_t$. \square

5 Proof of the invertibility

We recall that

$$\mathcal{F}(A, u) = f - g \left(\operatorname{id} - A^{-1} \nabla u \right) \det \left(I - A^{-1} D^2 u \right),$$

and

$$\mathcal{G}(t, u^1, u^2) := \mathcal{F}(A_t, u^1 + \lambda_t u^2) \quad \text{with} \quad A_t := \begin{pmatrix} 1 & 0 \\ 0 & \lambda_t \end{pmatrix}. \quad (17)$$

We want to prove the invertibility of $D_u \mathcal{G}(t, u^1, u^2)$. The first lemma will consider the case $t \neq 0$, the second one the case $t = 0$.

Lemma 17. For any $(t, u^1, u^2) \in \Omega$ with $t \neq 0$, for any $q \in \mathcal{C}_\diamond^\infty(\mathbb{T}^2)$, there is a unique $(v^1, v^2) \in \mathcal{C}_\diamond^\infty(\mathbb{T}^1) \times \mathcal{C}_{*,\diamond}^\infty(\mathbb{T}^2)$ such that

$$D_u \mathcal{G}(t, u^1, u^2) (v^1, v^2) = q. \quad (18)$$

Proof. Let $(t, u^1, u^2) \in \Omega$ with $t \neq 0$, and let $q \in \mathcal{C}_\diamond^\infty(\mathbb{T}^2)$. Then, if we set $u_t := u^1 + \lambda_t u^2$, Lemma 6 tells us that there is a unique $v_t \in \mathcal{C}_\diamond^\infty(\mathbb{T}^2)$ such that

$$\operatorname{div} \left((f - \mathcal{G}(t, u^1, u^2)) [I - A_t^{-1} D^2 u_t]^{-1} A_t^{-1} \nabla v_t \right) = q. \quad (19)$$

Let us define

$$v^1(x_1) := \int v_t(x_1, x_2) dx_2 \quad \text{and} \quad v^2(x_1, x_2) := \frac{1}{\lambda_t} (v_t(x_1, x_2) - v^1(x_1)).$$

Then, by construction, (v^1, v^2) is the unique pair solving (18). \square

Lemma 18. *For any $(0, u^1, u^2) \in \Omega$, for any $q \in \mathcal{C}_\diamond^\infty(\mathbb{T}^2)$, there is a unique $(v^1, v^2) \in \mathcal{C}_\diamond^\infty(\mathbb{T}^1) \times \mathcal{C}_{*,\diamond}^\infty(\mathbb{T}^2)$ such that*

$$D_u \mathcal{G}(0, u^1, u^2)(v^1, v^2) = q.$$

Proof. Writing $\partial u = (\partial_1 u^1, \partial_2 u^2)$ and $\partial v = (\partial_1 v^1, \partial_2 v^2)$, we want to solve

$$D_u \mathcal{G}(0, u^1, u^2)(v^1, v^2) = q.$$

By substitution, for any $\xi \in \mathcal{C}^\infty$, Equation (14) yields

$$\int \xi(x - \partial u(x)) [f(x) - \mathcal{G}(0, u^1, u^2)(x)] dx = \int \xi(y) g(y) dy.$$

Therefore, if we differentiate the previous equation with respect to u along the direction v , we get

$$\begin{aligned} - \int \langle \nabla \xi(\operatorname{id} - \partial u), \partial v \rangle (f - \mathcal{G}(0, u^1, u^2)) \\ - \int \xi(\operatorname{id} - \partial u) D_u \mathcal{G}(0, u^1, u^2)(v^1, v^2) = 0. \end{aligned}$$

Since $\nabla[\xi \circ (\operatorname{id} - \partial u)] = [I - D\partial u]^T \nabla \xi(\operatorname{id} - \partial u)$, we have

$$\langle \nabla \xi(\operatorname{id} - \partial u), \partial v \rangle = \langle \nabla[\xi \circ (\operatorname{id} - \partial u)], [I - D\partial u]^{-1} \partial v \rangle$$

and this yields

$$D_u \mathcal{G}(0, u^1, u^2)(v^1, v^2) = \operatorname{div} \left((f - \mathcal{G}(0, u^1, u^2)) [I - D\partial u]^{-1} \partial v \right).$$

Notice then that

$$(f - \mathcal{G}(0, u^1, u^2)) [I - D\partial u]^{-1} = g(\operatorname{id} - \partial u) \begin{pmatrix} 1 - \partial_{2,2} u^2 & 0 \\ \partial_{1,2} u^2 & 1 - \partial_{1,1} u^1 \end{pmatrix},$$

thus,

$$\begin{aligned} D_u \mathcal{G}(0, u^1, u^2)(v^1, v^2) \\ = \partial_1 \left[g(x - \partial u(x)) \left(1 - \partial_{2,2} u^2(x) \right) \partial_1 v^1(x_1) \right] + \partial_2 [\dots]. \end{aligned}$$

Therefore, if $D_u \mathcal{G}(0, u^1, u^2)(v^1, v^2) = q$, integrating with respect to x_2 yields

$$\int \partial_1 \left[g(x - \partial u(x)) \left(1 - \partial_{2,2} u^2(x) \right) \partial_1 v^1(x_1) \right] dx_2 = \int q(x) dx_2,$$

that is to say

$$\partial_1 \left[\left\{ \int g(x - \partial u(x)) \left(1 - \partial_{2,2} u^2(x) \right) dx_2 \right\} \partial_1 v^1(x_1) \right] = \int q(x) dx_2. \quad (20)$$

But there is a smooth map $Q : \mathbb{T}^1 \rightarrow \mathbb{R}$ such that $\partial_1 Q(x_1) = \int q(x_1, x_2) dx_2$, since $\int q(x) dx = 0$, and it is unique if we require $Q(0) = 0$. Thus, taking a primitive of (20), there is a $c \in \mathbb{R}$ such that:

$$\underbrace{\left[\int g(x - \partial u(x)) \left(1 - \partial_{2,2} u^2(x) \right) dx_2 \right]}_{G(x_1)} \partial_1 v^1(x_1) = Q(x_1) + c.$$

Since $G(x_1) > 0$, we get

$$\partial_1 v^1 = \frac{Q + c}{G},$$

and this yields the unique possible value for c since the integral w.r.t. x_1 of the right hand side must be zero. Combined with the condition $\int v^1 dx_1 = 0$, we thus have completely characterized v^1 .

Now, let us do the same for v^2 . We have to solve the equation

$$\begin{aligned} \partial_2 \left[g(\text{id} - \partial u) \left(1 - \partial_{1,1} u^1 \right) \partial_2 v^2 \right] \\ = q - \partial_1 \left[g(\text{id} - \partial u) \left(1 - \partial_{2,2} u^2 \right) \partial_1 v^1 \right] - \partial_2 \left[g(\text{id} - \partial u) \partial_{1,2} u^2 \partial_1 v^1 \right], \end{aligned}$$

and this is exactly the same kind of equation as (20). If we fix $x_1 \in \mathbb{T}^1$, the same reasoning can be applied here, and thus we get v^2 . \square

6 Proof of the tame estimates

Our aim here is to show that, locally on $(t, u^1, u^2) \in \Omega$ and $q \in \mathcal{C}_\diamond^\infty(\mathbb{T}^2)$, for any $n \in \mathbb{N}$, there is a constant $C_n > 0$ such that, if

$$D_u \mathcal{G}(t, u^1, u^2)(v^1, v^2) = q \quad (21)$$

for some $(v^1, v^2) \in \mathcal{C}_\diamond^\infty(\mathbb{T}^1) \times \mathcal{C}_{*,\diamond}^\infty(\mathbb{T}^2)$, then

$$\|v^1\|_{H^{n+2}} + \|v^2\|_{H^n} \leq C_n \left(1 + |t| + \|u^1\|_{H^{n+3}} + \|u^2\|_{H^{n+3}} + \|q\|_{H^n}\right).$$

In fact, we will prove something slightly stronger:

$$\|v^1\|_{H^{n+2}} + \|\partial_2 v^2\|_{H^n} \leq C_n \left(\|u^1\|_{\mathcal{C}^{n+3}} + \|u^2\|_{\mathcal{C}^{n+3}} + \|q\|_{H^n}\right). \quad (22)$$

Indeed then, since $\int v^2(x_1, x_2) dx_2 = 0$, we have a Poincaré inequality, which implies $\|v^2\|_{H^n} \leq c_n \|\partial_2 v^2\|_{H^n}$. Notice also that (22) would by itself prove there is uniqueness.

Lemma 19. *Let $M, \varepsilon > 0$. There are $C = C(M, \varepsilon)$ and $\tau = \tau(M, \varepsilon)$ such that, if $(t, u^1, u^2) \in \Omega$ and $q \in \mathcal{C}_\diamond^\infty(\mathbb{T}^2)$ satisfy*

$$0 < |t| < \tau, \quad 1 - \partial_{1,1} u^1 - \lambda_t \partial_{1,1} u^2 > \varepsilon, \quad \|q\|_{L^2} + \|u^1\|_{\mathcal{C}^3} + \|u^2\|_{\mathcal{C}^3} \leq M, \quad (23)$$

if $(v^1, v^2) \in \mathcal{C}_\diamond^\infty(\mathbb{T}^1) \times \mathcal{C}_\diamond^\infty(\mathbb{T}^2)$ is a solution of (21), then

$$\|\partial_2 v^2\|_{L^2} \leq C. \quad (24)$$

Proof. We set $u_t := u^1 + \lambda_t u^2$ and also $v_t := v^1 + \lambda_t v^2$. Then, by assumption,

$$D_u \mathcal{F}(A_t, u_t) v_t = q.$$

By choosing τ appropriately, we can ensure that (23) implies $A_t - D^2 u_t > \varepsilon$, so that we are entitled to apply Lemma 6 and get

$$\|v_t\|_{H^2} \leq C_{\varepsilon, N, 1} \{\|q\|_{L^2} + \|u_t\|_{\mathcal{C}^3}\} \leq C. \quad (25)$$

We now set

$$\begin{aligned} B_t &:= \left(f - \mathcal{G}(t, u^1, u^2)\right) \left[I - A_t^{-1} D^2 u_t\right]^{-1} A_t^{-1} \\ &= \frac{f - \mathcal{G}(t, u^1, u^2)}{\det(A_t - D^2 u_t)} \left[\text{Co}(A_t - D^2 u_t)\right]^T \\ &= \frac{g(\text{id} - A_t^{-1} \nabla u_t)}{\det A_t} \left[\text{Co}(A_t - D^2 u_t)\right]^T \end{aligned}$$

so that, according to (17) and Lemma 4, Equation (21) becomes

$$\text{div}(B_t \nabla v_t) = q.$$

Notice that

$$\text{Co}(A_t - D^2 u_t) = \begin{pmatrix} \lambda_t - \lambda_t \partial_{2,2} u_t^2 & \lambda_t \partial_{1,2} u_t^2 \\ \lambda_t \partial_{1,2} u_t^2 & 1 - \partial_{1,1} u_t^1 \end{pmatrix},$$

therefore we can write $B_t = U_t + V_t/\lambda_t$ with

$$U_t := g(\text{id} - A_t^{-1}\nabla u_t) \begin{pmatrix} 1 - \partial_{2,2}u^2 & \partial_{1,2}u^2 \\ \partial_{1,2}u^2 & 0 \end{pmatrix}, \quad (26)$$

$$V_t := g(\text{id} - A_t^{-1}\nabla u_t) \begin{pmatrix} 0 & 0 \\ 0 & 1 - \partial_{1,1}u_t \end{pmatrix}. \quad (27)$$

Thus,

$$q = \text{div}(B_t \nabla v_t) = \text{div}(U_t \nabla v_t) + \frac{1}{\lambda_t} \text{div}(V_t \nabla v_t).$$

As $\partial_2 v^1 = 0$, we have $V_t \nabla v^1 = 0$. Since $v_t = v^1 + \lambda_t v^2$, we get

$$\text{div}(U_t \nabla v_t) + \text{div}(V_t \nabla v^2) = q,$$

that is to say

$$\partial_2 \left[g(\text{id} - A_t^{-1}\nabla u_t)(1 - \partial_{1,1}u_t)\partial_2 v^2 \right] = q - \text{div}(U_t \nabla v_t). \quad (28)$$

Since $g > \delta$ for some δ , and as we have assume $1 - \partial_{1,1}u_t > \varepsilon$, allowing the constant C to change from line to line we get

$$\begin{aligned} \|\partial_2 v^2\|_{L^2}^2 &\leq \frac{C}{\delta\varepsilon} \int g(\text{id} - A^{-1}\nabla u_t)(1 - \partial_{1,1}u_t)|\partial_2 v^2|^2 \\ &\leq C \int [q - \text{div}(U_t \nabla v_t)] v^2 \\ &\leq C (\|q\|_{L^2} + \|U_t \nabla v_t\|_{H^1}) \|v^2\|_{L^2} \end{aligned}$$

However, since $\int v^2(x_1, x_2) dx_2 = 0$, we have $\|v^2\|_{L^2} \leq C \|\partial_2 v^2\|_{L^2}$. Therefore,

$$\|\partial_2 v^2\|_{L^2} \|v^2\|_{L^2} \leq C \|\partial_2 v^2\|_{L^2}^2 \leq C (\|q\|_{L^2} + \|U_t \nabla v_t\|_{H^1}) \|v^2\|_{L^2}.$$

Thus, since $\|U_t\|_{\mathcal{C}^1} \leq C(1 + \|u^1\|_{\mathcal{C}^3} + \|u^2\|_{\mathcal{C}^3}) \leq C$ as we can see from (26),

$$\|\partial_2 v^2\|_{L^2} \leq C \{ \|q\|_{L^2} + \|v_t\|_{H^2} \}.$$

Then, using (25), we get the result. \square

We now proceed by induction to get an estimate for any order $n \in \mathbb{N}$.

Lemma 20. *Under the same assumptions than in the previous lemma, for any $n \in \mathbb{N}$, there is a constant $C_n = C_n(M, \varepsilon)$ such that*

$$\|\partial_2 v^2\|_{H^n} \leq C_n \left(\|q\|_{H^n} + \|u^1\|_{\mathcal{C}^{n+3}} + \|u^2\|_{\mathcal{C}^{n+3}} \right). \quad (29)$$

Proof. Let us assume (29) has been proved for some $n \in \mathbb{N}$, and let us show it holds even for $n + 1$. Let $\nu \in \mathbb{N}^2$ be such that $|\nu| := \nu_1 + \nu_2 = n + 1$. Recall (28), that is to say

$$\partial_2 \left[g(\text{id} - A^{-1} \nabla u_t)(1 - \partial_{1,1} u_t) \partial_2 v^2 \right] = q - \text{div}(U_t \nabla v_t).$$

We already know from Lemma 7 that $v_t = v^1 + \lambda_t v^2$ is smooth, therefore, if we differentiate this equation with respect to ν , we get

$$\begin{aligned} & \partial_2 \left[g(\text{id} - A^{-1} \nabla u_t)(1 - \partial_{1,1} u_t) \partial_2 \partial_\nu v^2 \right] \\ &= - \sum_{0 \leq \alpha < \nu} \binom{\nu}{\alpha} \partial_2 \left[\partial_{\nu-\alpha} \left\{ g(\text{id} - A^{-1} \nabla u_t)(1 - \partial_{1,1} u_t) \right\} \partial_2 \partial_\alpha v^2 \right] \\ & \quad + \partial_\nu q - \partial_\nu \text{div}(U_t \nabla v_t). \end{aligned}$$

On the other hand, since $g > \delta$ and $1 - \partial_{1,1} u_t > \varepsilon$, we have

$$\begin{aligned} \|\partial_2 \partial_\nu v^2\|_{L^2}^2 &\leq \frac{1}{\delta \varepsilon} \int g(\text{id} - A^{-1} \nabla u_t)(1 - \partial_{1,1} u_t) |\partial_2 \partial_\nu v^2|^2 \\ &\leq -\frac{1}{\delta \varepsilon} \int \partial_2 \left[g(\text{id} - A^{-1} \nabla u_t)(1 - \partial_{1,1} u_t) \partial_2 \partial_\nu v^2 \right] \partial_\nu v^2. \end{aligned}$$

Thus,

$$\begin{aligned} & \|\partial_2 \partial_\nu v^2\|_{L^2}^2 \\ &\leq \sum_{0 \leq \alpha < \nu} \binom{\nu}{\alpha} \int \left[\partial_{\nu-\alpha} \left\{ g(\text{id} - A^{-1} \nabla u_t)(1 - \partial_{1,1} u_t) \right\} \partial_2 \partial_\alpha v^2 \right] \partial_2 \partial_\nu v^2 \\ & \quad - \frac{1}{\delta \varepsilon} \int \left[\partial_\nu q - \partial_\nu \text{div}(U_t \nabla v_t) \right] \partial_\nu v^2, \end{aligned}$$

and therefore

$$\begin{aligned} & \|\partial_2 \partial_\nu v^2\|_{L^2}^2 \\ &\leq \sum_{0 \leq \alpha < \nu} C \left\| \partial_{\nu-\alpha} \left\{ g(\text{id} - A^{-1} \nabla u_t)(1 - \partial_{1,1} u_t) \right\} \partial_2 \partial_\alpha v^2 \right\|_{L^2} \left\| \partial_2 \partial_\nu v^2 \right\|_{L^2} \\ & \quad + C \left\| \partial_\nu q - \partial_\nu \text{div}(U_t \nabla v_t) \right\|_{L^2} \left\| \partial_\nu v^2 \right\|_{L^2}. \end{aligned}$$

As $\|\partial_\nu v^2\|_{L^2} \leq c \|\partial_2 \partial_\nu v^2\|_{L^2}$, we get

$$\begin{aligned} \|\partial_2 \partial_\nu v^2\|_{L^2} &\leq C \sum_{0 \leq k \leq n} \left\| g(\text{id} - A^{-1} \nabla u_t)(1 - \partial_{1,1} u_t) \right\|_{\mathcal{C}^{n+1-k}} \left\| \partial_2 v^2 \right\|_{H^k} \\ & \quad + C \left\{ \|q\|_{H^{n+1}} + \|U_t \nabla v_t\|_{H^{n+2}} \right\}. \quad (30) \end{aligned}$$

On the one hand, using the same Landau–Kolmogorov inequalities and the same trick than in the proof of Lemma 6, for $0 \leq k \leq n$, we get

$$\begin{aligned} & \left\| g(\text{id} - A^{-1}\nabla u_t)(1 - \partial_{1,1}u_t) \right\|_{\mathcal{C}^{n+1-k}} \left\| \partial_2 v^2 \right\|_{H^k} \\ & \leq c_n \left(\left\| g(\text{id} - A^{-1}\nabla u_t)(1 - \partial_{1,1}u_t) \right\|_{\mathcal{C}^{n+1}} \left\| \partial_2 v^2 \right\|_{L^2} \right. \\ & \quad \left. + \left\| g(\text{id} - A^{-1}\nabla u_t)(1 - \partial_{1,1}u_t) \right\|_{\mathcal{C}^1} \left\| \partial_2 v^2 \right\|_{H^n} \right). \end{aligned}$$

Recall we have assumed (29) holds true for n , therefore, using (23), we get

$$\begin{aligned} & \left\| g(\text{id} - A^{-1}\nabla u_t)(1 - \partial_{1,1}u_t) \right\|_{\mathcal{C}^{n+1-k}} \left\| \partial_2 v^2 \right\|_{H^k} \\ & \leq c_n \left(1 + \|q\|_{H^n} + \|u^1\|_{\mathcal{C}^{n+3}} + \|u^2\|_{\mathcal{C}^{n+3}} \right). \quad (31) \end{aligned}$$

On the other hand,

$$\begin{aligned} \|U_t \nabla v_t\|_{H^{n+2}} &= \|D^{n+1}(U_t \nabla v_t)\|_{H^1} \\ &\leq C \{ \|U_t\|_{\mathcal{C}^{n+2}} \|\nabla v_t\|_{H^1} + \|U_t\|_{\mathcal{C}^1} \|\nabla v_t\|_{H^{n+2}} \}, \end{aligned}$$

which, since $\|u^1\|_{\mathcal{C}^3} + \|u^2\|_{\mathcal{C}^3} \leq M$, implies

$$\|U_t \nabla v_t\|_{H^{n+2}} \leq C \left\{ \left(1 + \|u^1\|_{\mathcal{C}^{n+4}} + \|u^2\|_{\mathcal{C}^{n+4}} \right) \|v_t\|_{H^2} + \|v_t\|_{H^{n+2}} \right\}.$$

Then, using Lemma 6 we get

$$\|U_t \nabla v_t\|_{H^{n+2}} \leq c_n \left(\|q\|_{H^n} + \|u^1\|_{\mathcal{C}^{n+4}} + \|u^2\|_{\mathcal{C}^{n+4}} \right). \quad (32)$$

Bringing together (30), (31), and (32), we get the estimate we sought. \square

Lemma 21. *The result of Lemma 20 still stands for $t = 0$, with the same constants.*

Proof. Let $(0, u^1, u^2) \in \Omega$ and $q \in \mathcal{C}_\diamond^\infty(\mathbb{T}^2)$ such that

$$1 - \partial_{1,1}u^1 > \varepsilon, \quad \|q\|_{L^2} + \|u^1\|_{\mathcal{C}^3} + \|u^2\|_{\mathcal{C}^3} \leq M, \quad (33)$$

Let τ be the constant given by Lemma (20). Then, since $(s, u^1, u^2) \in \Omega$ for s small enough, with $0 < |s| < \tau$, we can proceed by approximation. Indeed, if (v_s^1, v_s^2) is the solution to

$$D_u \mathcal{G}(s, u^1, u^2)(v_s^1, v_s^2) = q,$$

where u^1, u^2, q have been all fixed, then all the H^n norms of v_s^1, v_s^2 are bounded according to Lemma 20. Up to an extraction, there is convergence, which by compact embedding is as strong as we want. But the convergence can only be towards the solution of

$$D_u \mathcal{G}(0, u^1, u^2)(v^1, v^2) = q,$$

hence estimate (29) is still valid for the limit. \square

7 Higher dimensions

The difficulty in extending those results in higher dimensions only come from the technical nature of Sections 5 and 6. We are now going to explain what changes in higher dimension.

Setting and notations

We consider $\lambda_1, \dots, \lambda_{N-1} : \mathbb{R} \rightarrow [0, +\infty)$, assumed to be smooth and such that $\lambda_k = 0$ if and only if $t = 0$. We then define A_t by

$$A_t := \begin{pmatrix} 1 & & & & \\ & \lambda_1(t) & & & \\ & & \lambda_1(t)\lambda_2(t) & & \\ & & & \ddots & \\ & & & & \prod \lambda_i(t) \end{pmatrix}.$$

The decomposition of the Kantorovich potential ψ_t becomes

$$\psi_t(x_1, \dots, x_N) = \psi_t^1(x_1) + \lambda_1 \psi_t^2(x_1, x_2) + \dots + \left(\prod_{i < N} \lambda_i \right) \psi_t^N(x_1, \dots, x_N).$$

where ψ_t^k depends only on the k first variables x_1, \dots, x_k , and is such that

$$\forall x_1, \dots, x_{k-1}, \quad \int \psi_t^k(x_1, \dots, x_{k-1}, y_k) dy_k = 0.$$

For convenience, we set

$$\widehat{\psi}_t^k := \psi_t^k + \lambda_k \psi_t^{k+1} + \dots + \left(\prod_{k \leq i < N} \lambda_i \right) \psi_t^N,$$

so that we have

$$\widehat{\psi}_t^1 := \psi_t, \quad \widehat{\psi}_t^k = \psi_t^k + \lambda_k \widehat{\psi}_t^{k+1}, \quad \widehat{\psi}_t^N = \psi_t^N,$$

and

$$\forall x_1, \dots, x_{k-1}, \quad \int \dots \int \widehat{\psi}_t^k(x_1, \dots, x_{k-1}, y_k, \dots, y_N) dy_k \dots dy_N = 0.$$

For instance, if $d = 3$, we have

$$\psi_t = \psi_t^1 + \lambda_1 \psi_t^2 + \lambda_1 \lambda_2 \psi_t^3 \quad \text{and} \quad \begin{cases} \widehat{\psi}_t^1 = \psi_t^1 + \lambda_1 \psi_t^2 + \lambda_1 \lambda_2 \psi_t^3 \\ \widehat{\psi}_t^2 = \psi_t^2 + \lambda_2 \psi_t^3 \\ \widehat{\psi}_t^3 = \psi_t^3. \end{cases}$$

Let us denote by E the set of all $(t, u^1, \dots, u^N) \in \mathbb{R} \times \prod \mathcal{C}^\infty(\mathbb{T}^k)$ such that

$$\forall k \in \{1, \dots, N\}, \quad \int u^k dx_k = 0.$$

Then, if $(t, u^1, \dots, u^N) \in E$, we set

$$\widehat{u}^N := u^N, \quad \widehat{u}^k := u^k + \lambda_k \widehat{u}^{k+1}, \quad u := \widehat{u}^1,$$

and this is consistent with the previous notation. Notice that

$$\nabla u = \begin{pmatrix} \partial_1 \widehat{u}^1 \\ \lambda_1 \partial_2 \widehat{u}^2 \\ \lambda_1 \lambda_2 \partial_3 \widehat{u}^3 \\ \vdots \\ \prod \lambda_k \partial_N \widehat{u}^N \end{pmatrix} \quad \text{and} \quad A^{-1} \nabla u = \partial \widehat{u} = \begin{pmatrix} \partial_1 \widehat{u}^1 \\ \partial_2 \widehat{u}^2 \\ \partial_3 \widehat{u}^3 \\ \vdots \\ \partial_N \widehat{u}^N \end{pmatrix},$$

and thus,

$$A^{-1} D^2 u = D \partial \widehat{u} = \begin{pmatrix} \partial_{1,2} \widehat{u}^1 & 0 & \cdots & 0 \\ \partial_{1,2} \widehat{u}^2 & \partial_{2,2} \widehat{u}^2 & \ddots & \vdots \\ \vdots & \vdots & \ddots & 0 \\ \partial_{1,N} \widehat{u}^N & \partial_{2,N} \widehat{u}^N & \cdots & \partial_{N,N} \widehat{u}^N \end{pmatrix}. \quad (34)$$

We define Ω as the open subset of E formed by the (t, u) such that:

- either $t \neq 0$, and then $A_t - D^2 u > 0$;
- or $t = 0$, and then $1 - \partial_{k,k} u^k > 0$ for all k .

Invertibility

We want to solve, for $(0, u) \in \Omega$, the equation $D_u \mathcal{G}(0, u)v = q$. Since for $t > 0$,

$$D_u \mathcal{G}(t, u)v = \operatorname{div} \left((f - \mathcal{G}(t, u)) [I - A^{-1} D^2 u]^{-1} A^{-1} \nabla v \right),$$

which, when replacing $A^{-1} D^2 u$ and $A^{-1} \nabla v$ with $D \partial \widehat{u}$ and $\partial \widehat{v}$, becomes

$$D_u \mathcal{G}(t, u)v = \operatorname{div} \left((f - \mathcal{G}(t, u)) [I - D \partial \widehat{u}]^{-1} \partial \widehat{v} \right),$$

and since, when $t = 0$, we have $\widehat{u}^k = u^k$ and $\partial \widehat{u} = \partial u$, what we would like to solve is

$$q = D_u \mathcal{G}(0, u)v = \operatorname{div} \left((f - \mathcal{G}(0, u)) [I - D \partial u]^{-1} \partial v \right).$$

The trick is to integrate with respect to x_{k+1}, \dots, x_N to get an equation on v^1, \dots, v^k . If v^1, \dots, v^{k-1} have already been found, $[I - D \partial u]^{-1}$ being lower triangular thanks to (34), the resulting equation on v^k is of the same kind as the one we have dealt with in Section 5. The same reasoning can thus be applied.

Tame estimate

As in the 2-dimensional case, we need only to find a tame estimate when $t \neq 0$ for the solution (v^1, \dots, v^N) of

$$q = \operatorname{div}(B\nabla v) \quad \text{with} \quad B := \frac{g(\operatorname{id} - A^{-1}\nabla u)}{\det A} \left[\operatorname{Co}(A - D^2 u) \right]^T.$$

First, notice that by integrating with respect to x_N , we are reduced to the $N - 1$ dimensional case. Therefore, we can proceed by induction on N .

So let us assume we already have a tame estimate for v^1, \dots, v^{N-1} . To get an estimate for $v^N = \widehat{v}^N$, we will find one for each \widehat{v}^k , this time by induction on k . Since $\widehat{v}^1 = v$ satisfies a nice strictly elliptic equation, and thus comes with a tame estimate, we need only to show how to get one for \widehat{v}^k if we have one for $\widehat{v}^1, \dots, \widehat{v}^{k-1}$.

The key lies in the following decomposition of the matrix B : for any k ,

$$B = B^1 + \frac{1}{\lambda_1} B^2 + \frac{1}{\lambda_1 \lambda_2} B^3 + \dots + \frac{1}{\lambda_1 \cdots \lambda_{k-2}} B^{k-1} + \frac{1}{\lambda_1 \cdots \lambda_{k-1}} \widehat{B}^k,$$

where the coefficients $(b_{\alpha,\beta}^i)$ of B^i are zero except when $\min(\alpha, \beta) = i$, and where the coefficients $(\widehat{b}_{\alpha,\beta}^k)$ of \widehat{B}^k are zero except for $\min(\alpha, \beta) \geq k$:

$$B^i = \begin{pmatrix} b_{i,i}^i & \cdots & b_{i,N}^i \\ \vdots & & \vdots \\ b_{N,i}^i & & \end{pmatrix}, \quad \widehat{B}^k = \begin{pmatrix} \widehat{b}_{k,k}^k & \cdots & \widehat{b}_{k,N}^k \\ \vdots & & \vdots \\ \widehat{b}_{N,k}^k & \cdots & \widehat{b}_{N,N}^k \end{pmatrix},$$

the point being that all the coefficients $b_{\alpha,\beta}^i, \widehat{b}_{\alpha,\beta}^k$ can be bounded in \mathcal{C}^n by the norms of the u^i in \mathcal{C}^{n+2} uniformly in t , at least for small t . Indeed, if such a decomposition exists, since

$$v = v^1 + \lambda_1 v^2 + \dots + \lambda_1 \cdots \lambda_{i-2} v^{i-1} + \lambda_1 \cdots \lambda_{i-1} \widehat{v}^i,$$

with $\partial_i v^j = 0$ if $i > j$, which implies $\partial_i v = \lambda_1 \cdots \lambda_{i-1} \partial_i \widehat{v}^i$, we have

$$\operatorname{div}(B\nabla v) = \left[\sum_{i < k} \frac{1}{\lambda_1 \cdots \lambda_{i-1}} \operatorname{div}(B^i \nabla v) \right] + \frac{1}{\lambda_1 \cdots \lambda_{k-1}} \operatorname{div}(\widehat{B}^k \nabla v),$$

and thus

$$\operatorname{div}(B\nabla v) = \left[\sum_{i < k} \operatorname{div}(B^i \nabla \widehat{v}^i) \right] + \operatorname{div}(\widehat{B}^k \nabla \widehat{v}^k). \quad (35)$$

On the one hand, the matrix \widehat{B}^k is symmetric and non-negative, and we can work on an open subset of Ω such that

$$\forall \xi \in \mathbb{R}^N, \quad \varepsilon \left(\sum_{i \geq k} |\xi_i|^2 \right) \leq \langle \widehat{B}^k \xi, \xi \rangle.$$

On the other hand, since

$$\forall x_1, \dots, x_{k-1}, \quad \int \cdots \int \widehat{v}^k(x_1, \dots, x_N) dx_k \cdots dx_N = 0,$$

we have a Poincaré inequality:

$$\|\widehat{v}^k\|_{L^2}^2 \leq C \sum_{i \geq k} \|\partial_i \widehat{v}^k\|_{L^2}^2.$$

Therefore,

$$\|\widehat{v}^k\|_{L^2}^2 \leq \frac{C}{\varepsilon} \int \langle \widehat{B}^k \nabla \widehat{v}^k, \nabla \widehat{v}^k \rangle \leq \frac{C}{\varepsilon} \|\operatorname{div}(\widehat{B}^k \nabla \widehat{v}^k)\|_{L^2} \|\widehat{v}^k\|_{L^2},$$

and this shows how we can deduce a L^2 estimate for \widehat{v}^k from (35) and a series of estimates for \widehat{v}^i , $i < k$. Estimates for the norms H^n , $n > 0$, easily follow, by the same reasoning as in Section 6.

Thus, all we need is the decomposition

$$B = B^1 + \frac{1}{\lambda_1} B^2 + \frac{1}{\lambda_1 \lambda_2} B^3 + \cdots + \frac{1}{\lambda_1 \cdots \lambda_{N-1}} B^N,$$

with

$$B^i = \begin{pmatrix} b_{i,i}^i & \cdots & b_{i,N}^i \\ \vdots & & \\ b_{N,i}^i & & \end{pmatrix}.$$

Remember that

$$B := \frac{g(\operatorname{id} - A^{-1} \nabla u)}{\det A} [\operatorname{Co}(A - D^2 u)]^T,$$

and $\det A = \lambda_1 (\lambda_1 \lambda_2) \cdots (\lambda_1 \cdots \lambda_{N-1})$, therefore all we have to do is to show how in $\operatorname{Co}(A - D^2 u)$ we can gather the λ_k so as to get the decomposition we seek. Since $\partial_{i,j} u = \lambda_1 \cdots \lambda_{\max(i,j)-1} \partial_{i,j} \widehat{u}^{\max(i,j)}$,

$$\begin{aligned} [\operatorname{Co}(A - D^2 u)]_{i,j} &= \sum_{\substack{\sigma \in \mathfrak{S}_n \\ \sigma(i)=j}} \prod_{\substack{1 \leq k \leq N \\ k \neq i}} (A - D^2 u)_{k, \sigma(k)} \\ &= \sum_{\substack{\sigma \in \mathfrak{S}_n \\ \sigma(i)=j}} \prod_{\substack{1 \leq k \leq N \\ k \neq i}} \lambda_1 \cdots \lambda_{\max(k, \sigma(k))-1} \left(\delta_{k, \sigma(k)} - \partial_{k, \sigma(k)} \widehat{u}^{\max(k, \sigma(k))} \right). \end{aligned}$$

Thus, if $i \leq j$, we set $\omega_{\alpha, \beta} = \lambda_\alpha \cdots \lambda_{\max(\alpha, \beta)-1} \left(\delta_{\alpha, \beta} - \partial_{\alpha, \beta} \widehat{u}^{\max(\alpha, \beta)} \right)$, and

then we get

$$\begin{aligned} \left[\text{Co}(A - D^2u) \right]_{i,j} &= \sum_{\substack{\sigma \in \mathfrak{S}_n \\ \sigma(i)=j}} \varepsilon(\sigma) \prod_{\substack{1 \leq k \leq N \\ k \neq i}} \lambda_1 \cdots \lambda_{k-1} \omega_{k, \sigma(k)} \\ &= \sum_{\substack{\sigma \in \mathfrak{S}_n \\ \sigma(i)=j}} \frac{\varepsilon(\sigma)}{\lambda_1 \cdots \lambda_{i-1}} \left[\prod_{1 \leq k \leq N} \lambda_1 \cdots \lambda_{k-1} \right] \left[\prod_{\substack{1 \leq k \leq N \\ k \neq i}} \omega_{k, \sigma(k)} \right], \end{aligned}$$

that is to say,

$$\left[\text{Co}(A - D^2u) \right]_{i,j} = \frac{\det A}{\lambda_1 \cdots \lambda_{i-1}} \sum_{\substack{\sigma \in \mathfrak{S}_n \\ \sigma(i)=j}} \varepsilon(\sigma) \prod_{\substack{1 \leq k \leq N \\ k \neq i}} \omega_{k, \sigma(k)}.$$

Since we have assume $i \leq j$, this is exactly what we wanted.

References

- [1] G. MONGE, “Mémoire sur la théorie des déblais et des remblais”, *Histoire de l’Académie royale des sciences de Paris*, pp. 666–704, 1784.
- [2] L. KANTOROVICH, “On the translocation of masses”, *Journal of Mathematical Sciences*, vol. 133, pp. 1381–1382, 2006. The original paper was published in *Dokl. Akad. Nauk SSSR*, vol. 37, no. 7–8, pp. 227–229, 1942.
- [3] Y. BRENIER, “Décomposition polaire et réarrangement monotone des champs de vecteurs”, *C. R. Acad. Sci. Paris Sér. I Math.*, vol. 305, no. 19, pp. 805–808, 1987.
- [4] ———, “Polar factorization and monotone rearrangement of vector-valued functions”, *Comm. Pure Appl. Math.*, vol. 44, no. 4, pp. 375–417, 1991.
- [5] D. CORDERO-ERAUSQUIN, “Sur le transport de mesures périodiques”, *C. R. Acad. Sci. Paris Sér. I Math.*, vol. 329, no. 3, pp. 199–202, 1999.
- [6] R. J. McCANN, “Polar factorization of maps on Riemannian manifolds”, *Geom. Funct. Anal.*, vol. 11, no. 3, pp. 589–608, 2001.
- [7] J.-D. BENAMOU AND Y. BRENIER, “A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem”, *Numer. Math.*, vol. 84, no. 3, pp. 375–393, 2000.
- [8] S. ANGENENT, S. HAKER, AND A. TANNENBAUM, “Minimizing flows for the Monge-Kantorovich problem”, *SIAM J. Math. Anal.*, vol. 35, no. 1, pp. 61–97, 2003.

- [9] G. LOEPER AND F. RAPETTI, “Numerical solution of the Monge-Ampère equation by a Newton’s algorithm”, *C. R. Math. Acad. Sci. Paris*, vol. 340, no. 4, pp. 319–324, 2005.
- [10] G. CARLIER, A. GALICHON, AND F. SANTAMBROGIO, “From Knothe’s transport to Brenier’s map and a continuation method for optimal transport”, *SIAM J. Math. Anal.*, vol. 41, no. 6, pp. 2554–2576, 2009/10.
- [11] M. ROSENBLATT, “Remarks on a multivariate transformation”, *Ann. Math. Statistics*, vol. 23, pp. 470–472, 1952.
- [12] A. FIGALLI, F. MAGGI, AND A. PRATELLI, “A mass transportation approach to quantitative isoperimetric inequalities”, *Invent. Math.*, vol. 182, no. 1, pp. 167–211, 2010.
- [13] H. KNOTHE, “Contributions to the theory of convex bodies”, *Michigan Math. J.*, vol. 4, pp. 39–52, 1957.
- [14] L. A. CAFFARELLI, “The regularity of mappings with a convex potential”, *J. Amer. Math. Soc.*, vol. 5, no. 1, pp. 99–104, 1992.
- [15] A. FIGALLI, Y.-H. KIM, AND R. J. McCANN, *Regularity of optimal transport maps on multiple products of spheres*. To appear in *J. Eur. Math. Soc. (JEMS)*, 2011.
- [16] D. GILBARG AND N. S. TRUDINGER, *Elliptic partial differential equations of second order*, vol. 224 of Grundlehren der Mathematischen Wissenschaften, Springer-Verlag, Berlin, 1977.
- [17] R. S. HAMILTON, “The inverse function theorem of Nash and Moser”, *Bull. Amer. Math. Soc. (N.S.)*, vol. 7, no. 1, pp. 65–222, 1982.