Université Paris-Sud

École doctorale de mathématiques
de la région Paris-Sud (ED 142)
Laboratoire de mathématiques d'Orsay

SCUOLA NORMALE SUPERIORE

Scuola Normale Superiore

Classe di Scienze

PhD thesis, mathematics

Thèse de doctorat en mathématiques • Tesi di perfezionamento in matematica

Unidimensional and Evolution Methods for Optimal Transportation

by Nicolas Bonnotte

Defended and approved on December 16, 2013

Committee:	L. Ambrosio,	SNS,	advisor
	J.-D. Benamou,	INRIA,	referee
	J. A. Carrillo,	Imperial College,	referee
	N. Gigli,	UPMC,	examiner
	B. Maury,	Univ. Paris-Sud,	examiner
	F. Santambrogio,	Univ. Paris-Sud,	advisor

Unidimensional and Evolution Methods for Optimal Transportation

In dimension one, optimal transportation is rather straightforward. The easiness with which a solution can be obtained in that setting has recently been used to tackle more general situations, each time thanks to the same method [4, 19, 49]. First, disintegrate your problem to go back to the unidimensional case, and apply the available 1D methods to get a first result; then, improve it gradually using some evolution process.

This dissertation explores that direction more thoroughly. Looking back at two problems only partially solved this way, I show how this viewpoint in fact allows to go even further.

The first of these two problems concerns the computation of Yann Brenier's optimal map. Guillaume Carlier, Alfred Galichon, and Filippo Santambrogio [19] found a new way to obtain it, thanks to an differential equation for which an initial condition is given by the Knothe–Rosenblatt rearrangement. (The latter is precisely defined by a series of unidimensional transformations.) However, they only dealt with discrete target measures; I generalize their approach to a continuous setting [10]. By differentiation, the Monge–Ampère equation readily gives a PDE satisfied by the Kantorovich potential; but to get a proper initial condition, it is necessary to use the Nash–Moser version of the implicit function theorem.

The basics of optimal transport are recalled in the first chapter, and the Nash–Moser theory is exposed in chapter 2. My results are presented in chapter 3, and numerical experiments in chapter 4.

The last chapter deals with the IDT algorithm, devised by François Pitié, Anil C. Kokaram, and Rozenn Dahyot [49]. It builds a transport map that seems close enough to the optimal map for most applications [50]. A complete mathematical understanding of the procedure is, however, still lacking. An interpretation as a gradient flow in the space of probability measures is proposed, with the sliced Wasserstein distance as the functional. I also prove the equivalence between the sliced and usual Wasserstein distances.

Méthodes unidimensionnelles et d'évolution pour le transport optimal

Sur une droite, le transport optimal ne pose pas de difficultés. Récemment, ce constat a été utilisé pour traiter des problèmes plus généraux. En effet, on a remarqué qu'une habile désintégration permet souvent de se ramener à la dimension un, ce qui permet d'utiliser les méthodes afférentes pour obtenir un premier résultat, que l'on fait ensuite évoluer pour gagner en précision [4, 19, 49].

Je montre ici l'efficacité de cette approche, en revenant sur deux problèmes déjà résolus partiellement de cette manière, et en complétant la réponse qui en avait été donnée.

Le premier problème concerne le calcul de l'application de Yann Brenier. En effet, Guillaume Carlier, Alfred Galichon et Filippo Santambrogio [19] ont prouvé que celle-ci peut être obtenue grâce à une équation différentielle, pour laquelle une condition initiale est donnée par le réarrangement de Knothe–Rosenblatt (lui-même défini *via* une succession de transformations unidimensionnelles). Ils n'ont cependant traité que des mesures finales discrètes ; j'étends leur résultat aux cas continus [10]. L'équation de Monge–Ampère, une fois dérivée, donne une EDP pour le potentiel de Kantorovitch ; mais pour obtenir une condition initiale, il faut utiliser le théorème des fonctions implicites de Nash–Moser.

Le chapitre 1 rappelle quelques résultats essentiels de la théorie du transport optimal, et le chapitre 2 est consacré au théorème de Nash–Moser. J'expose ensuite mes propres résultats dans le chapitre 3, et leur implémentation numérique dans le chapitre 4.

Enfin, le dernier chapitre est consacré à l'algorithme IDT, développé par François Pitié, Anil C. Kokaram et Rozenn Dahyot [49]. Celui-ci construit une application de transport suffisamment proche de celle de M. Brenier pour convenir à la plupart des applications [50]. Une interprétation en est proposée en termes de flot de gradients dans l'espace des probabilités, avec pour fonctionnelle la distance de Wasserstein projetée. Je démontre aussi l'équivalence de celle-ci avec la distance usuelle de Wasserstein.

Metodi unidimensionali e di evoluzione per il trasporto ottimale

Sulla retta reale, il trasporto ottimale non presenta nessuna difficoltà. Questo fatto è stato usato di recente per ottenere risultati anche in situazioni più generali. Ogni volta, disintegrando il problema per tornare alla dimensione uno, in modo da utilizzare metodi specifici a questo caso, si ottiene una prima soluzione; e poi, con metodi d'evoluzione, questa viene migliorata [4, 19, 49].

Qui, vorrei mostrare l'efficacia di tale approccio. Rivisito due problemi che avevano ricevuto, in questo modo, solo soluzioni parziali e, continuando nella stessa direzione, li completo.

Il primo problema riguarda la mappa ottimale di Yann Brenier. Guillaume Carlier, Alfred Galichon e Filippo Santambrogio [19] hanno dimostrato che si può calcolarla con un'equazione differenziale ordinaria se il riordinamento di Knothe–Rosenblatt è preso come condizione iniziale. Quest'ultimo viene precisamente definito da una serie di trasformazioni unidimensionali. Tali autori hanno però trattato solo il caso delle misure finali discrete; estendo il loro risultato al caso continuo [10]. Infatti, quando si differenzia l'equazione di Monge–Ampère, si ottiene una PDE per il potenziale di Kantorovič; tuttavia, per avere una condizione iniziale assicurando esistenza e unicità, bisogna usare il teorema di Nash–Moser.

Nel capitolo 1, tratto di qualche risultato essenziale della teoria del trasporto ottimale. Il teorema di Nash e Moser è l'oggetto del capitolo 2. Successsivamente, espongo i miei risultati nel capitolo 3, e la loro implementazione numerica nel capitolo 4.

Infine, nell'ultimo capitolo, studio l'algoritmo IDT, ideato da François Pitié, Anil C. Kokaram, e Rozenn Dahyot [49]. Tale algoritmo produce una mappa così vicina a quella di Brenier, che può essere utilizzata al suo posto in varie situazioni [50]. Un'interpretazione di questo algoritmo è proposta come flusso gradiente nello spazio delle misure di probabilità, rispetto al quadrato della distanza di Super Wasserstein. Mostro anche l'equivalenza tra quest'ultima e la distanza di Wasserstein classica.

Contents

Preface

How I learned of optimal transportation is a bit fortuitous. From time to time, a mathematical education can seem a bit lifeless; at least for me, it felt that way at some point during my scholarship at the École normale supérieure. Yet when I complained to Guillaume Carlier, who was my *tuteur* there, he suggested I should try a new subject: optimal transportation. As it was rooted in a very simple question—roughly, how to move stuff efficiently?—but still involved nice mathematics, he thought it might catch my interest. And it did.

Following his advice, I attended a series of lectures on the subject by François Bolley, Bruno Nazaret, and Filippo Santambrogio—which turned out to be very lively indeed. A year later, in 2010, I was lucky enough to go to the Scuola Normale Superiore in Pisa to write my master thesis under the supervision of Luigi Ambrosio. I was to study one of the most abstract outcome of the theory: gradient flows in the space of probability measures. The months I spent there were intense, and exciting. I was therefore very glad to be able to start a PhD under the joint supervision of Professors Ambrosio and Santambrogio.

Over the three years that followed, I came to learn a lot, and not only about mathematics, but also about perseverance and self-organization, about trust in others' insights as well as in my own intuition—and about a researcher's life and my own aspirations. Of course, going back twice in Pisa for an extended amount of time, I also had the opportunity to learn more about Italy, its language, its culture, and its people.

It was a wonderful experience, for which I am immensely grateful.

— Communay, August 15, 2013

Acknowledgements

In this adventure, I was fortunate enough to count on two advisors to guide me; this thesis owes much to their advice, their patience, and their support. I wish therefore to express all my gratitude to Luigi Ambrosio and Filippo Santambrogio; they always knew how to put me back on track.

I would like to warmly thank Jean-David Benamou and José A. Carillo for accepting to act as referees, and Nicola Gigli and Bertrand Maury for taking part in the defense committee.

Many thanks are also due to Sylvain Faure—if it were not for him, I would not have been able to obtain any numerical results at all.

More generally, I would like to thank all the people of the ANEDP team, including the support staff, and especially my fellow students from the Bureau 256/8, past and present. All along, it was a real pleasure to come to Orsay, and I am not going to forget such a friendly atmosphere any time soon.

Then, I would also like to thank my old tuteur, Guillaume Carlier, for introducing me to optimal transport, and sending me to Pisa in the first place. I owe him much.

I am also much indebted to the Scuola Normale Superiore, the generous hospitality of which I benefitted three times in four years. I have rarely found myself in a environment so favorable to concentration and reflection.

I should also thank the French-Italian University for its financial support, which allowed me to remain in Pisa so long.

At last, I would like to thank my family, and all my friends both in Paris and in Pisa, for their encouragements and comprehension. I owe them my sanity—or at least, what is left thereof.

It was my desire that the introduction be as clear and accessible as possible. I do not know if I have succeeded, but I am grateful to Maxime Gheysens, Arthur Leclaire, Jehanne Richet, and Alexandre Tchao for their useful comments. Any mistake or obscurity that may remain can only be mine.

Introduction

Many illustrations can be found in the literature that try to simply present the problem lying at the heart of optimal transportation. Some talk, for instance, of sand piles to be moved [45, 62], or bread to be sent from bakeries to local cafés [63], or coal to be delivered from mines to steelworks [56]. Let me indulge, however, in giving another example. Readers already familiar with the subject might be excused for skipping the following part; it should get more interesting afterwards.

Imagine you are the head of an industrial complex somewhere in China, maybe producing electronic components for a company called Appell Inc. The labor comes from rural areas all over the country, and needs housing close to the factories; therefore, the complex not only includes many plants, but also dormitories. Your task is to assign to each and every one of your workers a bed. But their commuting costs you money, as you have to pay for buses (or any other transportation system), and you want to minimize your expenses. How would you achieve it?

Assuming there is barely enough accommodation for everyone, we can represent the distributions of workers and beds by two measures μ and ν with the same total mass. Then, given an area A, the values $\mu(A)$ and $\nu(A)$ respectively indicate the numbers of employees and beds in that area. We will denote by $c(x, y)$ the daily cost of transportation between the factory x and the dormitory y, back and forth.

Since there are so many workers—that is, so many variables—, you cannot expect to find a precise solution for everyone, but you need to operate from a "mesoscopic" level. A way to ease the search for a solution is to group your workers by factories, and try to send all the people working at the same place x to sleep in the same dormitory y. In that case, what you are looking for is a *mapping*, $y = T(x)$, telling you, for each factory, where to house its staff—that is, you want to find a map T that minimizes the total cost of transportation,

$$\int c(x, T(x))\, \mathrm{d}\mu(x),$$

and such that $\nu(A) = \mu(T^{-1}(A))$ for any area A, because $T^{-1}(A)$ is where people sleeping in A come from. This version of the problem was historically the first to be studied, by Gaspard Monge [45] in the 18th century—although in term of sand particles rather than workers—, and has therefore come to be known as *Monge's problem*.

However, there might be no such mapping—for instance, if you have no choice but to split the workforce of a given factory between many dormitories. Hence, in the 1940s, Leonid Kantorovich [35, 36] proposed instead to model a solution as a measure γ, such that $\gamma(A \times B)$ represents the number of people working in the area A and sleeping somewhere in B (this implies its marginals should be μ and ν). The total cost of transportation for the plan γ is then given by

$$\int c(x, y)\, \mathrm{d}\gamma(x, y).$$

To find an optimal γ is today called the *Monge–Kantorovich problem*; it really is a generalization of Monge's initial question, for if there is an optimal mapping T, then it corresponds to an optimal measure γ such that

$$\gamma(A \times B) = \mu(A \cap T^{-1}(B)),$$

and the transport costs are the same.

In his papers, Kantorovich also showed you might be able—to keep our story going—to pass on the problem to the workers: just start charging for the accommodation, introduce fares to the transportation system to cover for its operation, and generously hand over a subsidy to compensate for all that. Indeed, values may exist for the subsidies and the bed rates such that the only solution for any employee not to lose money is to find the right spot to sleep. That is, if $S(x)$ is the additional money you grant daily to the people working in the factory x, and $B(y)$ is the price you ask for a bed in the dormitory y, then you could perhaps manage to set S and B in such a way that $S(x) \leq B(y) + c(x, y)$, with the double assurance that: (1) for any given x, there is equality for some y's; (2) if the workers in x comply and go to one of those y's, everyone may have a bed. In the end, you pay the difference between what you hand over and what you get back from the accommodation fares, and if S and B are correctly set, that should be

$$\int S(x)\,\mathrm{d}\mu(x) - \int B(y)\,\mathrm{d}\nu(y) \;=\; \min_{\gamma} \int c(x, y)\,\mathrm{d}\gamma(x, y).$$

The Monge–Kantorovich would then be solved, in some sense—but the difficulty now lies in setting the right values for S and B. Those are called, when optimal, *Kantorovich potentials.*

With Kantorovich's approach, you might have therefore to split a group of coworkers. On the other hand, if the factories are quite small, and not too concentrated, then there are not that many people working at the same place, so it should be easier to assign the same dormitory to them all: the solution might still be a mapping. For a cost equal to the squared distance, this was formally proved by Yann Brenier [13, 14] in the 1980s, who also showed optimal values exist for the bed rates B and the subsidies S that force the employees to find the right spot, which they do by simply following the direction of decreasing subsidies—more precisely, from a factory x, one should go to $y = T(x) = x - \nabla S(x)$. This was to be expected somehow, as the handouts should be fewer where there are more beds nearby.

But then, in practical terms, how to compute the optimal mapping T? When both measures are discrete—that is, when the factories and the dormitories are scattered—, linear programming provides a solution, as does Dimitri P. Bertsekas's algorithm [9]. However, when the distributions are more diffuse, the problem is in general hard to solve—except in dimension one. In that case, there is a formula, which translates into the following method: if the factories and the dormitories are all aligned along the same street, you should do the assignment going from one end to the other, and allocate the first bed you encounter to the first man you meet, and so on. In other terms, if F and G stand for the cumulative distributions of the workers and beds—that is, if $F(t)$ and $G(t)$ are respectively the total numbers of workers and

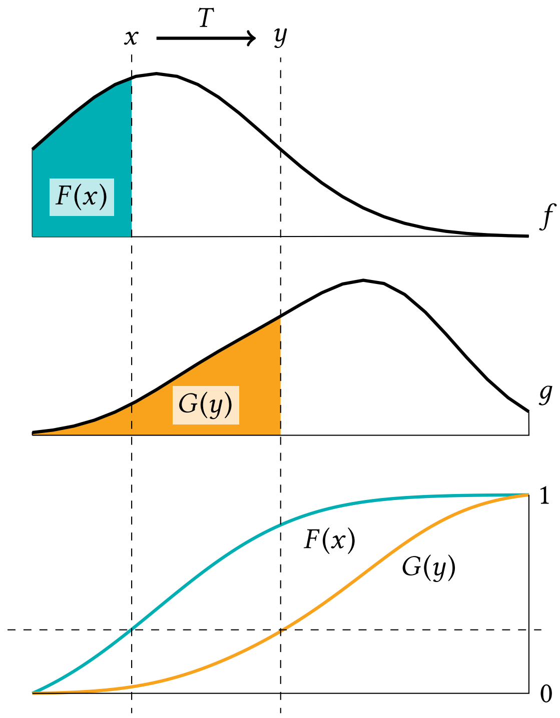

Figure A: Construction of the optimal map T in 1D. The cumulative distributions, F and G, represent the areas below the graphs of the densities of μ and ν, denoted by f and g respectively; the point x is sent onto y, i.e. $y = T(x)$, if and only if $F(x) = G(y)$, which means the filled areas should be equal.

beds located before the point t—, then people working in x should go to sleep to $y = T(x) = G^{-1} \circ F(x)$; see figure A, on this page.

In greater dimensions, even if many numerical methods have been developed [4, 7, 8, 11, 37, 40], the problem remains difficult. It is, for instance, possible to start from a non-optimal mapping, like the Knothe–Rosenblatt rearrangement—which, as we shall see, applies the previous formula on each dimension—, and then alter it through a steepest-descent algorithm so as to make it optimal [4]. Or, using the peculiar form the optimal map should have, $T(x) = x - \nabla S(x)$, one can start from a non-optimal potential S_0, and then apply Newton's method to catch the optimal S [40]. By some aspects, my paper [10] combines these two approaches, since it computes the optimal potential S rather than the map T directly, but it nevertheless manages to start from

Knothe's map. (I will present the results of this paper in chapter 3, with new numerical experiments in chapter 4).

This *Knothe–Rosenblatt rearrangement* was devised independently by Herbert Knothe [38] and Murray Rosenblatt [51] in the 1950s. It is a mapping, assigning to each worker from your industrial complex a bed in a dormitory—although, a priori, not in a very cost-effective way—by solving the problem on each dimension one after the other, thanks to the unidimensional solution to Monge's problem. Let us assume the measures μ and ν have densities, which we denote by f and g; then $f(x)$ is the number of workers in the factory x, and $g(y)$ is the number of beds in the dormitory y. If the complex's roads are divided into avenues (north–south) and streets (west–east), then the position $x = (x_1, x_2)$ of a factory is given by the intersection of an avenue x_1 and a street x_2; the same for a dormitory's position $y = (y_1, y_2)$. To assign the beds, we can start by summing up the workforces on each avenue on the one hand, and the beds on the other hand:

$$\hat{f}(x_1) = \int f(x_1, x_2)\,\mathrm{d}x_2, \qquad \hat{g}(x_1) = \int g(y_1, y_2)\,\mathrm{d}y_2.$$

We denote by $\hat{F}$ and $\hat{G}$ the cumulative distributions of $\hat{f}$ and $\hat{g}$. Then, dealing with each avenue from the west to the east, one after the other, we tell the workers on the avenue x_1 to look for a dormitory on the most western avenue with some spare capacity—and this avenue will be $y_1 = T_\mathrm{K}^1(x_1) = \hat{G}^{-1} \circ \hat{F}(x_1)$. Once everybody has a designated avenue where to find a bed, we proceed likewise to assign a street, and its intersection with the avenue will yield the dormitory's position: starting from the north and moving southward, we tell people working in $x = (x_1, x_2)$ to go to the most northern dormitory they can find on the avenue $y_1 = T_\mathrm{K}^1(x_1)$ with some beds left, which will be at the intersection with the street $y_2 = T_\mathrm{K}^2(x_1, x_2) = \hat{G}_{y_1}^{-1} \circ \hat{F}_{x_1}(x_2)$, with $\hat{F}_{x_1}$ and $\hat{G}_{y_1}$ the (normalized) cumulative distributions of workers and beds on the avenues x_1 and y_1 respectively. The Knothe–Rosenblatt rearrangement is the mapping we thus obtain, $T_\mathrm{K} = (T_\mathrm{K}^1, T_\mathrm{K}^2)$; see figure B, on the next page. Sadly, as this transport map deals with each dimension in a certain order, on which the result strongly depends, it is anisotropic, and thus unsuitable for many applications—e.g., in image processing—because it creates artifacts.

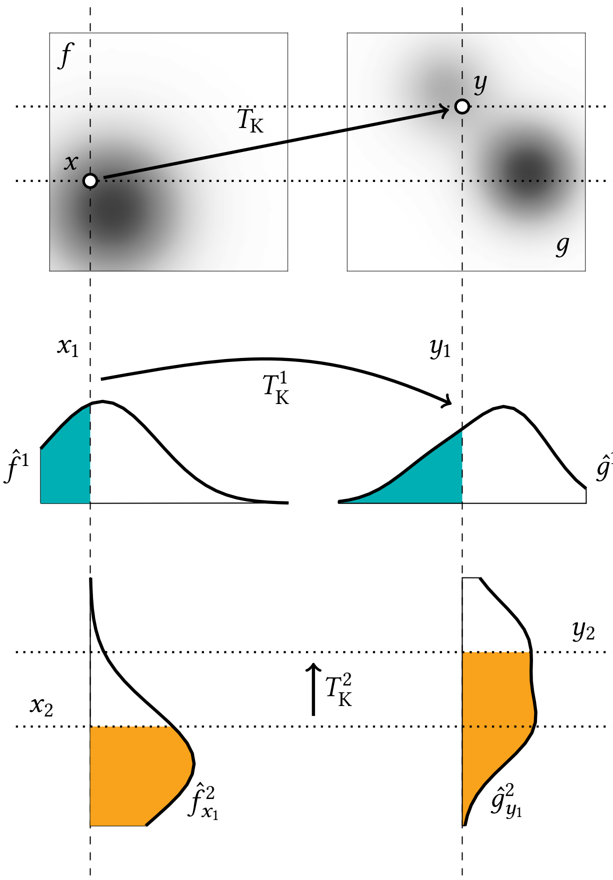

Figure B: Construction of the Knothe–Rosenblatt rearrangement $(y_1, y_2) = T_K(x_1, x_2)$, defined by $y_1 = T_K^1(x_1)$ and $y_2 = T_K^2(x_1, x_2)$. For each dimension, the hashed zones have the same areas, respectively $F^1(x_1) = G^1(y_1)$, and $F^2_{x_1}(x_2) = G^2_{y_1}(y_2)$.

The starting point of the theory I will present in chapter 3 is that this mapping would however be optimal, should the price of a north–south displacement be a lot less expensive than a weast–east one—i.e., the rearrangement would be optimal for a transportation cost $c_\varepsilon(x, y) = |x_1 - y_1|^2 + \varepsilon|x_2 - y_2|^2$, with ε infinitesimally small. But, increasing ε little by little and updating the optimal mapping accordingly, we could get back the optimal map for a regular quadratic cost, at least if we can get to $\varepsilon = 1$. This was achieved by Guillaume Carlier, Alfred Galichon, and Filippo Santambrogio [19], under the assumption the target measure is discrete—that is, when the dormitories are scattered.

Pursuing their work, I was able to deal with more diffuse distributions [10]. They had found a differential equation satisfied by the Kantorovich potential S; I therefore sought to do the same. We have seen that, for a cost equal to the squared distance, $c(x, y) = |x_1 - y_1|^2 + |x_2 - y_2|^2$, the optimal transport map is:

$$T(x) = x - \nabla S(x) = x - \begin{pmatrix} \partial_1 S(x) \\ \partial_2 S(x) \end{pmatrix}.$$

But for a cost $c_\varepsilon(x, y) = |x_1 - y_1|^2 + \varepsilon|x_2 - y_2|^2$, the optimal map can be written as

$$T_\varepsilon(x) = x - \begin{pmatrix} \partial_1 S_\varepsilon(x) \\ \partial_2 S_\varepsilon(x)/\varepsilon \end{pmatrix} = x - A_\varepsilon^{-1}\nabla S_\varepsilon(x) \qquad \text{with} \quad A_\varepsilon = \begin{pmatrix} 1 & 0 \\ 0 & \varepsilon \end{pmatrix}.$$

Since T_ε must still send the measure μ onto the measure ν, that is,

$$\nu(A) = \int_{y \in A} g(y)\,\mathrm{d}y = \int_{T_\varepsilon(x) \in A} f(x)\,\mathrm{d}x = \mu(T_\varepsilon^{-1}(A)) \qquad \text{for any area } A,$$

the following equality, called a *Monge–Ampère equation*, must always hold:

$$f(x) = g(T_\varepsilon(x)) \det(\mathrm{D}T_\varepsilon) = g(x - A_\varepsilon^{-1}\nabla S_\varepsilon(x)) \det(\mathrm{I}_d - A_\varepsilon^{-1}\nabla^2 S_\varepsilon(x)). \tag{a}$$

This equation, along with the further condition $A_\varepsilon - \nabla^2 S_\varepsilon > 0$ (to force uniqueness), completely determines the potential S_ε. The implicit function theorem then allows us to get information on its regularity in the following way: First, for u smooth enough such that $A_\varepsilon - \nabla^2 u > 0$, we set

$$\mathcal{F}(\varepsilon, u) \;:=\; f - g(\mathrm{Id} - A_\varepsilon^{-1}\nabla u)\det(\mathrm{I}_d - A_\varepsilon^{-1}\nabla^2 u),$$

so that $\mathcal{F}(\varepsilon, u) = 0$ if and only if $u = S_\varepsilon$. Then, the differential with respect to u, denoted by $\mathrm{D}_u\mathcal{F}$, is a second-order, strictly elliptic differential operator, which is invertible; hence, $\varepsilon \mapsto S_\varepsilon$ is at least $\mathscr{C}^1$. Differentiating the equation (a) with respect to ε, we therefore get a second-order, elliptic partial differential equation:

$$\operatorname{div}\left(f\left[\mathrm{I}_d - A_\varepsilon^{-1}\nabla^2 S_\varepsilon\right]^{-1} A_\varepsilon^{-1}\left(\nabla\dot{S}_\varepsilon - \dot{A}_\varepsilon A_\varepsilon^{-1}\nabla S_\varepsilon\right)\right) \;=\; 0. \tag{b}$$

The dotted symbols, $\dot{S}_\varepsilon$ and $\dot{A}_\varepsilon$, represent the derivatives with respect to ε; the target density g is here hidden in the determinant of $\mathrm{I}_d - A_\varepsilon^{-1}\nabla^2 S_\varepsilon$.

As long as ε stays away from zero, this last equation can be solved, and $\dot{S}_\varepsilon$ is the unique solution. So, if we know S_{ε_0} for some $\varepsilon_0 > 0$, we can get S_1 back, since we can obtain $\dot{S}_\varepsilon$ by solving the elliptic equation (b), and then compute $S_1 = S_{\varepsilon_0} + \int_{\varepsilon_0}^1 \dot{S}_\varepsilon \,\mathrm{d}\varepsilon$. This is akin to the continuation method, which was used by Philippe Delanoë [23] and John Urbas [60] from a theoretical point of view, and by Grégoire Loeper and Francesca Rapetti [40] for numerical computations.

But what happens when ε is infinitesimally small, and tends to zero? On the one hand, we know T_ε converges to be the Knothe–Rosenblatt rearrangement,

$$y \;=\; T_\mathrm{K}(x) \;=\; (T_\mathrm{K}^1(x_1), T_\mathrm{K}^2(x_1, x_2)).$$

On the other hand, when ε is infinitesimally small but still nonzero,

$$y = T_\varepsilon(x) = x - (\partial_1 S_\varepsilon(x), \partial_2 S_\varepsilon(x)/\varepsilon).$$

To reconcile this with the previous expression, and cancel the $1/\varepsilon$, maybe we can write $S_\varepsilon(x) = S_\varepsilon^1(x_1) + \varepsilon S_\varepsilon^2(x_1, x_2)$. Then,

$$T_\varepsilon(x) = x - \begin{pmatrix} \partial_1 S_\varepsilon^1(x_1) + \varepsilon\partial_1 S_\varepsilon^2(x_1, x_2) \\ \partial_2 S_\varepsilon^2(x_1, x_2) \end{pmatrix} \;\xrightarrow[\varepsilon\to 0]{}\; x - \begin{pmatrix} \partial_1 S_0^1(x_1) \\ \partial_2 S_0^2(x_1, x_2) \end{pmatrix} = T_\mathrm{K}(x),$$

so this viewpoint covers the case $\varepsilon = 0$ as well. This turns out to be the correct approach: in some sense, S^1_ε and S^2_ε are uniquely determined by their initial conditions S^1_0 and S^2_0, which come from the Knothe rearrangement $T_K = \mathrm{Id} - (\partial_1 S^1_0, \partial_2 S^2_0)$.

However, while the implicit function theorem was enough when ε stayed away from zero, results on the behavior of $S_\varepsilon = S^1_\varepsilon + \varepsilon S^2_\varepsilon$ when ε goes to zero prove a lot more difficult to get. The first idea that comes to mind is to try to apply the implicit function theorem once more, but this time to

$$\mathcal{G}(\varepsilon, u^1, u^2) \;:=\; \mathcal{F}(\varepsilon, u_\varepsilon) \;=\; f - g(\mathrm{Id} - A_\varepsilon^{-1}\nabla u_\varepsilon)\det(\mathrm{I}_d - A_\varepsilon^{-1}\nabla^2 u_\varepsilon),$$

defined for $\varepsilon > 0$, with $u_\varepsilon := u^1 + \varepsilon u^2$; when $\varepsilon = 0$, we can set

$$\mathcal{G}(0, u^1, u^2) := f - g(\mathrm{Id} - \partial u)\det(\mathrm{I}_d - \nabla\partial u) \qquad \text{where} \quad \partial u := (\partial_1 u^1, \partial_2 u^2).$$

The problem is, even though it is possible to solve

$$\mathrm{D}_{(u^1,u^2)}\mathcal{G}(0, S^1_0, S^2_0)(v^1, v^2) \;=\; q,$$

for $u^1, u^2 \in \mathscr{C}^{k+2}$ and $q \in \mathscr{C}^k$, the best we can get for the solution (v^1, v^2) is $v^1 \in \mathscr{C}^{k+2}$, which is good, and $\partial_{2,2} v^2 \in \mathscr{C}^k$, which is very bad: we need $v^2 \in \mathscr{C}^{k+2}$. There is, therefore, a loss of regularity, which prevents us from applying the implicit function theorem again. To get around such a difficulty, a solution is to work with $\mathscr{C}^\infty$ maps, so as to have an infinite source of smoothness. But then, we cannot use the implicit function theorem any longer, as $\mathscr{C}^\infty$ is not a Banach space; we need instead to use the stronger *Nash–Moser theorem*, which I will present in chapter 2.

After the theoretical aspects presented in chapter 3, I will show how this method can allow us effectively to compute Brenier's map for the regular quadratic cost, $y = T_1(x)$, in chapter 4. The idea is the go backward, starting from $\varepsilon = 0$ and going up to $\varepsilon = 1$. This numerical material is new, and was not present in my original paper [10]. It is, however, still sketchy: there is yet a considerable amount of work to be done in order to obtain something that can be used practically.

Finally, in the last chapter, a second problem, of a different kind, is introduced; it is however born out of the same overall approach. Since the optimal transport map is so easy to compute in dimension one, and so difficult to get in higher dimensions, the image processing community has devised a way to build another transport map, using only unidimensional mappings [49, 50]—not unlike Knothe's rearrangement therefore, but without its greatest flaw, which is its being anisotropic. Experimentally, it works well enough.

Let us again denote by f and g the densities of our two measures, μ and ν, on $\mathbb{R}^2$. Given any orthonormal basis (e_1, e_2), we can define

$$\hat{f}_{e_1}(x_1) := \int f(x_1 e_1 + x_2 e_2)\, \mathrm{d}x_2 \qquad \text{and} \qquad \hat{g}_{e_1}(y_1) := \int g(y_1 e_1 + y_2 e_2)\, \mathrm{d}x_2.$$

Those are similar to the $\hat{f}$ and $\hat{g}$ defined in the first step of the construction of the Knothe rearrangement; they are, in fact, the same when e is the canonical basis, $e_1 = (1,0)$ and $e_2 = (0,1)$. Then, we have two unidimensional measures, so we know how to send one onto the other—thanks to the map $T_{e_1} = \hat{G}_{e_1}^{-1} \circ \hat{F}_{e_1}$, where $\hat{F}_{e_1}$ and $\hat{G}_{e_1}$ denote again the cumulative distributions. This map should also be a good indicator of how we need to move the original measure μ along the direction e_1 to get ν. Likewise, we can get a map T_{e_2} for the direction e_2, and then combine those two maps into

$$T_e(x) \ := \ T_{e_1}(\langle e_1 | x \rangle)\, e_1 + T_{e_2}(\langle e_2 | x \rangle)\, e_2 \tag{c}$$

It is important to say, however, that this T_e does *not* send μ onto ν. It sends μ onto another measure—let us denote it by μ_1—, which should nevertheless be closer to ν. We can iterate the procedure, with μ_1 instead of μ and using a different basis e, and thus get another map T_e; then we define μ_2 as the measure obtain from μ_1 through the new T_e, and start again. In the end, if all the bases are well chosen, no particular direction should be privileged, and μ_n should converge toward ν. Notice that, at each step, there is a transport map sending μ onto μ_n, which is the composition of all the intermediate T_e.

This algorithm was introduced by François Pitié, Anil C. Kokaram, and Rozenn Dahyot [49], who called it the *Iterative Distribution Transfer algorithm*. To this day, a proper mathematical study is still lacking though. Numerical experiments suggest μ_n converges to ν, but it has not been proved yet—except in a very particular case, when the target measure ν is Gaussian. But even though the transport map between

μ and μ_n does not necessarily converge toward the optimal map between μ and ν, it has nevertheless been successfully used as a replacement [50].

I will present in the last chapter some steps toward a more complete understanding. This algorithm seems to be connected to a gradient flow in the space of probability measures—in the sense of the theory developed by Luigi Ambrosio, Nicola Gigli, and Giuseppe Savaré [3]—with what Marc Bernot called the *sliced Wasserstein distance* as the functional,

$$\mathrm{SW}_p(\mu, \nu) \;=\; \mathrm{SW}_p(f, g) \;:=\; \left(⨍ \mathrm{W}_p(\hat{f}_{e_1}, \hat{g}_{e_1})^p \,\mathrm{d}e \right)^{1/p},$$

the usual Wasserstein distance[1] being the pth root of the minimum value of the Monge–Kantorovich problem for the cost $c(x, y) = |x - y|^p$:

$$\mathrm{W}_p(\mu, \nu) \;:=\; \left(\min_{\gamma} \int |x - y|^p \,\mathrm{d}\gamma(x, y) \right)^{1/p}.$$

Indeed if, instead of defining the transport map T between μ_n and μ_{n+1} by (c) with a random basis e, and hoping for the randomness to homogenize the procedure, we would rather define

$$T(x) \;:=\; ⨍ T_e(\langle e_1 | x \rangle) \,\mathrm{d}e;$$

then, assuming the measures are sums of N Dirac masses—and therefore assimilable to vectors of $\mathbb{R}^{d\times N}$—, we obtain that the measure μ_{n+1} is given by

$$\mu_{n+1} \;:=\; \mu_n - \nabla F(\mu_n) \qquad \text{with} \qquad F(\mu) \;:=\; \frac{1}{2}\mathrm{SW}_2(\mu, \nu)^2.$$

This is nothing but the explicit Euler scheme for the gradient flow equation

$$\dot{\mu}_t \;:=\; -\nabla F(\mu_t).$$

[1] How the name “Wasserstein” came to be associated to this object is a bit strange. According to Ludger Rüschendorf [53], the distance was used in a 1969 paper by Leonid N. Vaserstein [61] and the term “Vasershtein distance” appears a year later, in a paper by Roland Dobrushin [24]. Today, the term “Kantorovich–Rubistein distance” is often used for the case $p = 1$, as the two mathematicians proved the distance could be extended into a norm. The name “Earth Mover’s distance” is also frequent in image processing [52]. See Cédric Villani’s book [63, chapter 6, bibliographical notes].

Following the variational method devised by Richard Jordan, David Kinderlehrer, and Felix Otto [34], and investigated by Luigi Ambrosio, Nicola Gigli, and Giuseppe Savaré [3], we can define an implicit Euler scheme,

$$\mu_{n+1} = \mu_n - h\nabla F(\mu_{n+1}),$$

by taking

$$\mu_{n+1} \in \arg\min_{\mu} \left\{ \frac{1}{2h} \mathrm{W}_2(\mu_n, \mu)^2 + \frac{1}{2} \mathrm{SW}_2(\mu, \nu)^2 \right\}.$$

The Wasserstein distance here replaces the usual Euclidean distance, which is used to define the classical implicit scheme on $\mathbb{R}^d$. Notice this definition works even if the measures are no longer assumed to be discrete. In any case, the sequences $(\mu_n)_{n\in\mathbb{N}}$ converge in some sense to a curve $(\mu_t)_{t\geq 0}$ when the time step tends to 0. This viewpoint could yield a theoretical justification of the algorithm, if we were able to prove the convergence of μ_t toward ν when t tends to infinity; to do so will, however, require more work still.

Chapter 1

Optimal transportation

1.0.1. The aim of this chapter is to recall some well-known facts that shall be needed later on. The presentation has therefore been tailored with a further use in mind, and proofs are only given when they are either very short or of a special interest. Notations are also set here.

For a general introduction to optimal transportation, the reader should rather refer to Cédric Villani's summae [62, 63] or Filippo Santambrogio's forthcoming lecture notes [54]. For a more abstract and more general exposition, see also the monograph by Luigi Ambrosio, Nicola Gigli, and Giuseppe Savaré [3, chapters 5–7].

1.1 The Monge–Kantorovich problem

1.1.1. Monge's problem. Given two probability measures μ and ν on $\mathbb{R}^d$ and a cost function $c : \mathbb{R}^d \times \mathbb{R}^d \to [0, \infty]$, the problem that was first introduced by Gaspard Monge [45] can be stated in modern terms as follows:

$$\text{find} \quad T : \mathbb{R}^d \to \mathbb{R}^d \quad \text{such that} \quad \nu = T_\# \mu \quad \text{and} \quad \int c(x, T(x))\,\mathrm{d}\mu(x) \quad \text{is minimal.} \tag{1.1.1.a}$$

The former condition, $\nu = T_\# \mu$, means that T should transport μ onto ν; that is, ν should be the push-forward of μ by T: for any ξ, $\int \xi(y)\,\mathrm{d}\nu(y) = \int \xi(T(x))\,\mathrm{d}\mu(x)$. The latter asks the total cost of transportation to be minimal.

1.1.2. Monge–Kantorovich problem. Depending on the measures, there might be no transport map sending μ onto ν, for instance if μ is discrete and ν is uniform. Hence, the following generalization was proposed by Leonid Kantorovich [35, 36]: instead of looking for a mapping,

$$\text{find a measure} \quad \gamma \in \Gamma(\mu,\nu) \quad \text{such that} \quad \int c(x,y)\,\mathrm{d}\gamma(x,y) \quad \text{is minimal,} \tag{1.1.2.a}$$

where $\Gamma(\mu,\nu)$ stands for the set of all transport plans between μ and ν, i.e. the probability measures on $\mathbb{R}^d \times \mathbb{R}^d$ with marginals μ and ν. This problem really extends Monge's, for any transport map T sending μ onto ν yields a measure $\gamma \in \Gamma(\mu,\nu)$, which is $\gamma = (\mathrm{Id}, T)_{\#}\mu$, i.e. the only measure γ on $\mathbb{R}^d \times \mathbb{R}^d$ such that

$$\forall\, \xi \in \mathscr{C}_{\mathrm{b}}(\mathbb{R}^d \times \mathbb{R}^d), \qquad \int \xi(x,y)\,\mathrm{d}\gamma(x,y) \;=\; \int \xi(x,T(x))\,\mathrm{d}\mu(x),$$

and the associated costs of transportation are the same. However, unlike in Monge's problem, for which there might be no admissible transport map—not to mention an optimal one—, in Kantorovich's version there is always a transport plan, for instance $\mu \otimes \nu$. Even better, it is not difficult to show there is always a solution:

1.1.3. Proposition. *Let μ,ν be two Borel probability measures on $\mathbb{R}^d$. If the cost function $c : \mathbb{R}^d \times \mathbb{R}^d \to [0,+\infty)$ is lower semicontinuous, then there is a solution to the Monge–Kantorovich problem (1.1.2.a). We denote by $\Gamma_o(\mu,\nu)$ the set of all such solutions.*

Proof. On one hand, as μ and ν are inner regular, the set $\Gamma(\mu,\nu)$ is tight and thus, being obviously closed, compact according to Prokhorov's theorem. On the other hand, as c is lower semicontinuous, the map $\gamma \mapsto \int c(x,y)\,\mathrm{d}\gamma(x,y)$ is also lower semicontinuous; for if

$$c_n(x,y) \;:=\; \inf_{\bar{x},\bar{y}} \left\{ c(\bar{x},\bar{y}) + n\left(|x-\bar{x}|^2 + |y-\bar{y}|^2\right)\right\},$$

then c_n is continuous, $c_n(x,y) \leq c(x,y)$, and c_n converges pointwise to c, and this, as soon as $\gamma_k \to \gamma$, implies

$$\int c\,\mathrm{d}\gamma \;\leq\; \liminf_{n\to\infty} \int c_n \wedge n\,\mathrm{d}\gamma \;\leq\; \liminf_{n\to\infty}\, \liminf_{k\to\infty} \int c_n \wedge n\,\mathrm{d}\gamma_k \;\leq\; \liminf_{k\to\infty} \int c\,\mathrm{d}\gamma_k$$

Thus, any minimizing sequence converges, up to an extraction, to a minimizer. □

1.1.4. Dual formulation. As will be shown in proposition 1.1.6 on the following page, there is a form of duality between the Monge–Kantorovich problem and the following other problem:

$$\text{find} \quad \psi, \varphi \in \mathscr{C}_0(\mathbb{R}^d) \quad \text{such that} \quad \psi(x) + \varphi(y) \le c(x, y) \\ \text{and} \quad \int \psi \, \mathrm{d}\mu + \int \varphi \, \mathrm{d}\nu \quad \text{is maximal.} \tag{1.1.4.a}$$

This is often called the dual or sometimes primal problem, because they are linked (see proposition 1.1.6 on the next page), and the space of signed Radon measures—where the Monge–Kantorovich problem is defined—is the dual of the space of continuous functions vanishing at infinity—where this new problem is defined, even though the condition to vanish at infinity is irrelevant. Whatever the naming, the requirement $\psi, \varphi \in \mathscr{C}_0(\mathbb{R}^d)$ can be relaxed, so that (1.1.4.a) becomes:

$$\text{find} \quad \psi \in \mathrm{L}^1(\mu), \varphi \in \mathrm{L}^1(\nu) \quad \text{such that} \quad \psi(x) + \varphi(y) \le c(x, y) \\ \text{and} \quad \int \psi \, \mathrm{d}\mu + \int \varphi \, \mathrm{d}\nu \quad \text{is maximal.} \tag{1.1.4.b}$$

1.1.5. Kantorovich potential and c-transform. Its seems natural to look for a solution of the new problem (1.1.4.b) among the pairs (ψ, φ) that saturate the condition, and therefore satisfy

$$\varphi(y) = \inf_x \{c(x, y) - \psi(x)\} \qquad \text{and} \qquad \psi(x) = \inf_y \{c(x, y) - \varphi(y)\}.$$

The first equality, when holding, will be written $\varphi = \psi^c$, where ψ^c is called the c-transform of ψ. Similarly, for the second we shall write $\psi = \varphi^c$. If both are verified—that is, if $\psi = \psi^{cc}$—, then ψ is said to be *c-concave*. Then, the problem (1.1.4.b) becomes

$$\text{find} \quad \psi \in \mathrm{L}^1(\mu) \quad \text{such that} \quad \int \psi \, \mathrm{d}\mu + \int \psi^c \, \mathrm{d}\nu \ \text{ is maximal.} \tag{1.1.5.a}$$

Any solution ψ is called a Kantorovich potential between μ and ν.

1.1.6. Proposition. *Let μ, ν be two Borel probability measures on $\mathbb{R}^d$. If the cost function $c : \mathbb{R}^d \times \mathbb{R}^d \to [0, +\infty)$ is lower semicontinuous and*

$$\iint c(x, y)\, \mathrm{d}\mu(x)\, \mathrm{d}\nu(y) \;<\; \infty,$$

then there is a Borel map $\psi : \mathbb{R}^d \to \mathbb{R}$ that is c-concave and optimal for (1.1.5.a). Moreover, the resulting maximum is equal to the minimum of the Monge–Kantorovich problem (1.1.2.a):

$$\min_{\gamma\in\Gamma(\mu,\nu)} \int c(x, y)\, \mathrm{d}\gamma(x, y) \;=\; \max_{\varphi\in \mathrm{L}^1(\mu)} \left\{ \int \varphi(x)\, \mathrm{d}\mu(x) + \int \varphi^c(y)\, \mathrm{d}\nu(y) \right\}.$$

If $\gamma \in \Gamma(\mu, \nu)$ is optimal, then $\psi(x) + \psi^c(y) = c(x, y)$ almost everywhere for γ.

For a proof of this proposition, see the monograph by Luigi Ambrosio, Giuseppe Savaré, and Nicola Gigli [3, Theorem 6.1.5].

1.2 Solution on the real line

1.2.1. In dimension one—that is, when μ and ν are probability measures on the real line—, a solution to the Monge–Kantorovich problem (1.1.2.a) can very often be explicitly computed, and turns out to be a solution of Monge's problem (1.1.1.a) as well. As we will see in chapter 3, my computation of the solution relies on the unidimensional case.

1.2.2. Cumulative distribution and generalized inverse. If μ is a probability measure on $\mathbb{R}$, its cumulative distribution is the map $F : \mathbb{R} \to [0, 1]$ defined by

$$F(x) \;:=\; \mu((-\infty, x]).$$

Its is an nondecreasing and right-continuous function. For such a map, it is possible to define a generalized inverse F^{-1}, also called quantile function, by setting

$$F^{-1}(y) \;:=\; \min\,\{\, x \in [-\infty, \infty] \;|\; y \le F(x) \,\}\,.$$

The values of F^{-1} give the different quantiles: for instance, $F^{-1}(3/4)$ yields the third quartile—hence the alternate name.

1.2.3. Lemma. *If F is a cumulative distribution, then $y \leq F(x)$ if and only if $F^{-1}(y) \leq x$.*

Proof. Since the minimum in the definition of F^{-1} is attained, $y \leq F(F^{-1}(y))$ for any y. Thus, if $F^{-1}(y) \leq x$ for some x, then $y \leq F(F^{-1}(y)) \leq F(x)$, as F is nondecreasing. Conversely, if $y \leq F(x)$, then the definition of F^{-1} implies $F^{-1}(y) \leq x$. □

1.2.4. Proposition. *Let $h \in \mathscr{C}^1(\mathbb{R})$ be a nonnegative, strictly convex function. Let μ and ν be Borel probability measures on $\mathbb{R}$ such that*

$$\iint h(x-y)\,\mathrm{d}\mu(x)\,\mathrm{d}\nu(y) \;<\; \infty. \tag{1.2.4.a}$$

If μ has no atom, and F and G stand for the respective cumulative distribution of μ and ν, then $T := G^{-1} \circ F$ solves Monge's problem for the cost $c(x,y) = h(x-y)$. If γ is the induced transport plan, that is, $\gamma := (\mathrm{Id}, T)_{\#}\mu$, then γ is optimal for the Monge–Kantorovich problem.

Proof. To begin with, notice T is well defined almost everywhere for μ. Indeed, there might be a problem only when $F(x) = 0$, for $G^{-1}(0) = -\infty$. But $F = 0$ only on $(-\infty, a]$ for some $a \in \mathbb{R}$, and, by the very definition of F, we have $\mu((-\infty, a]) = F(a) = 0$.

Notice also that, as F and G are nondecreasing, T must be nondecreasing as well. Then, lemma 1.2.3 on this page applied to the cumulative distribution G yields

$$\begin{aligned} T^{-1}((-\infty, y]) &= \left\{ x \in [-\infty, +\infty] \;\middle|\; G^{-1}(F(x)) \leq y \right\} \\ &= \{ x \in [-\infty, +\infty] \mid F(x) \leq G(y) \}. \end{aligned}$$

First, this set has to be an interval, as T is nondecreasing. Second, since μ has no atom, F is increasing and continuous, so this interval must be closed. Thus, if x is its supremum, we must have $F(x) = G(y)$, and therefore

$$\mu(T^{-1}((-\infty, y])) \;=\; \mu((-\infty, x]) \;=\; F(x) \;=\; G(y) \;=\; \nu((-\infty, y]).$$

This is enough to show $\nu = T_\# \mu$.

Now, let us prove T is optimal. On the one hand, if $u \geq x$, then, as T and h' are nondecreasing, $h'(u - T(u)) \leq h'(u - T(x))$. Integrating between x and some $y \geq x$, we get

$$\begin{aligned}\int_x^y h'(u - T(u))\,\mathrm{d}u \;&\leq\; \int_x^y h'(u - T(x))\,\mathrm{d}u \\ &\leq\; h(y - T(x)) - h(x - T(x)).\end{aligned}$$

On the other hand, if $u \leq x$, then $h'(u - T(u)) \geq h'(u - T(x))$; integrating bewteen x and $y \leq x$, we again get

$$\int_x^y h'(u - T(u))\,\mathrm{d}u \;\leq\; -\int_y^x h'(u - T(x))\,\mathrm{d}u \;\leq\; h(y - T(x)) - h(x - T(x)).$$

Thus, if we set

$$\psi(y) := \int_0^y h'(u - T(u))\,\mathrm{d}u,$$

then, in any case, $\psi(y) - \psi(x) \leq h(y - T(x)) - h(x - T(x))$, which implies

$$\psi^c(T(x)) \;:=\; \inf_y \{h(y - T(x)) - \psi(y)\} \;=\; h(x - T(x)) - \psi(x),$$

and this yields ψ is c-concave. On the other hand, the condition (1.2.4.a) ensures that there are x_0 and y_0 such that

$$\int h(x - y_0)\,\mathrm{d}\mu(x) \;<\; \infty \qquad \text{and} \qquad \int h(x_0 - y)\,\mathrm{d}\nu(y) \;<\; \infty.$$

Since $h(x - y_0) - \psi^c(y_0) \geq \psi(x)$, and $h(x_0 - T(x)) - \psi(x_0) \geq \psi^c(T(x))$, and also $\psi(x) \geq -\psi^c(T(x))$, we have

$$h(x - y_0) - \psi^c(y_0) \;\geq\; \psi(x) \;\geq\; -h(x_0 - T(x)) + \psi(x_0)$$

and as $T_{\#}\mu = \nu$, this implies $\psi \in \mathrm{L}^1(\mu)$. Similarly, $\psi^c \in \mathrm{L}^1(\nu)$. Therefore, integrating the equality $\psi(x) + \psi^c(x) = h(x - T(x))$ with respect to μ gives

$$\int \psi(x)\,\mathrm{d}\mu(x) + \int \psi^c(y)\,\mathrm{d}\nu(y) \;=\; \int c(x, T(x))\,\mathrm{d}\mu(x).$$

Since $\psi(x) + \psi^c(y) \leq c(x, y)$ for all pair (x, y), if γ is any other transport plan, the associated total transport cost is necessarily greater, and thus T is optimal. □

1.3 Yann Brenier's map and its regularity

1.3.1. Gaspard Monge [45] formulated his original problem in the 1780s with the distance as a cost function. But for such a cost, the question is particularly difficult: to give an idea, his characterization of the transport rays was rigorously proved only a century later, by Paul Appell [5, 6]; and in the 1970s, Vladimir Sudakov [58] claimed to have proved the existence of an optimal mapping, but a point in his demonstration was unconvincing—it was corrected by Luigi Ambrosio in 2000 [2], just after another method had been successfully used by Lawrence C. Evans and Wilfrid Gangbo, with stronger assumptions [27].

For a strictly convex cost, however, things are somewhat easier. At the end of the 1980s, Yann Brenier [13, 14] gave a general answer when the cost function is the squared Euclidean distance, and showed the key role convex functions play in that case. Since, his theorem has been extended to arbitrary, strictly convex cost functions, and for measures defined on a variety of domains; those cases will be studied in section 1.4 on page 31.

1.3.2. Subdifferential of a convex function. Let $\varphi : \mathbb{R}^d \to (-\infty, +\infty]$ be a convex, lower semicontinuous function. Then, it follows from the Hahn–Banach theorem applied to the epigraph of φ that, if x belongs to the interior of the domain of φ, there is $p \in \mathbb{R}^d$ such that

$$\forall\, y \in \mathbb{R}^d, \quad \varphi(y) \;\geq\; \varphi(x) + \langle p | y - x \rangle.$$

The set of all those p's is called the subdifferential of φ at x, and is denoted by $\partial\varphi(x)$. It can be shown that φ is locally Lipschitz on the interior of its domain, and therefore is differentiable almost everywhere on it. Should that be the case in x, the subdifferential is then a singleton: $\partial\varphi(x) = \{\nabla\varphi(x)\}$.

1.3.3. Theorem (Brenier). *Let μ and ν be two Borel probability measures on $\mathbb{R}^d$ with finite second-order moments—that is, such that*

$$\int |x|^2 \,\mathrm{d}\mu(x) \;<\; \infty \qquad \textit{and} \qquad \int |y|^2 \,\mathrm{d}\nu(y) \;<\; \infty.$$

Then, if μ is absolutely continuous, there is a unique $T : \mathbb{R}^d \to \mathbb{R}^d$ such that $\nu = T_\# \mu$ and

$$\int |x - T(x)|^2 \,\mathrm{d}\mu(x) \;=\; \min_{\gamma \in \Gamma(\mu,\nu)} \int |x-y|^2 \,\mathrm{d}\gamma(x,y).$$

Moreover, there is only one optimal transport plan γ, which is thus necessarily $(\mathrm{Id}, T)_\#\mu$, and T is the gradient of a convex function φ, which is therefore unique up to an additive constant. There is also a unique (up to an additive constant) Kantorovich potential ψ, which is locally Lipschitz and linked to φ through the relation

$$\varphi(x) \;=\; \frac{1}{2}|x|^2 - \psi(x).$$

Proof. We know from proposition 1.1.6 on page 25 that, for a cost $c(x,y) = \frac{1}{2}|x-y|^2$, there is a c-concave function ψ such that

$$\int \psi(x)\,\mathrm{d}\mu(x) + \int \psi^c(y)\,\mathrm{d}\nu(y) \;=\; \frac{1}{2}\int |x-y|^2 \,\mathrm{d}\gamma(x,y) \tag{1.3.3.a}$$

for some optimal transport plan $\gamma \in \Gamma(\mu,\nu)$. We set

$$\varphi(x) \;:=\; \frac{1}{2}|x|^2 - \psi(x).$$

Then, since $\psi^{cc} = \psi$, this function φ is convex and lower semicontinuous, being a supremum of affine maps:

$$\varphi(x) \;=\; \frac{1}{2}|x|^2 - \psi^{cc}(x)$$

$$= \sup_y \left\{ \frac{1}{2}|x|^2 - \frac{1}{2}|x-y|^2 + \psi^c(y) \right\}$$
$$= \sup_y \left\{ \langle y|x\rangle - \left(\frac{1}{2}|y|^2 - \psi^c(y) \right) \right\}.$$

This computation also yields the Legendre transform of φ, which is

$$\varphi^*(x) = \frac{1}{2}|x|^2 - \psi^c(x).$$

As φ is convex and lower semicontinuous, it is differentiable almost everywhere in the interior of its domain—that is, almost everywhere at least in the interior of the convex hull of the support of μ, since μ is absolutely continuous. All we have to do now is to show that the optimal transport map is

$$T(x) = \nabla\varphi(x) = x - \nabla\psi(x).$$

Notice that equality (1.3.3.a) translates into

$$\int \varphi(x)\,\mathrm{d}\mu(x) + \int \varphi^*(y)\,\mathrm{d}\nu(y) = \int \langle y|x\rangle\,\mathrm{d}\gamma(x,y).$$

As $\varphi(x) + \varphi^*(y) \geq \langle y|x\rangle$, this implies that for γ-a.e. pair (x,y), there is equality. Thus,

$$\forall\, z \in \mathbb{R}^d, \quad \langle y|z\rangle - \varphi(z) \leq \langle y|x\rangle - \varphi(x),$$

which, in turn, means $y \in \partial\varphi(x)$. But φ is differentiable for a.e. x in the support of μ, and in that case the subdifferential is reduced to $\nabla\varphi(x)$. Therefore, $\gamma = (\mathrm{Id}, \nabla\varphi)_{\#}\mu$. This also shows the uniqueness of γ and $T = \nabla\varphi$. This φ is unique up to an additive constant as well, and so is ψ. □

1.3.4. Monge–Ampère equation. Regularity results regarding the convex map φ and the optimal map $T = \nabla\varphi$ have been obtained, most notably by Luis A. Caffarelli [16, 17, 18], using the Monge–Ampère equation: if we denote by f and g the respective densities of μ and ν, then, if it is smooth enough, φ must solve

$$f(x) = g(\nabla\varphi(x))\det(\nabla^2\varphi(x)).$$

1.3.5. Theorem (Caffarelli). *Let U and V be two bounded, open subsets of $\mathbb{R}^d$, and let μ and ν be two probability measures respectively on U and V, with densities f and g. If those densities are bounded and bounded away from 0, and if V is convex, then φ is strictly convex and $\mathscr{C}^{1,\alpha}$ on U. Moreover, if f and g are $\mathscr{C}^k$ with $k \geq 1$, then φ is $\mathscr{C}^{k+2}$.*

If both U and V are strictly convex with smooth boundaries, the regularity of φ holds even on the boundary of U. In that case, $\nabla\varphi$ and $\nabla\varphi^*$ are diffeomorphisms, and inverse of each other.

1.4 Extension to the torus

1.4.1. Existence of an optimal map. Following Yann Brenier's article, an alternate, more general proof was found by Robert J. McCann [42], who then extended it to cover the case of measures defined on a Riemannian manifold[1] [43].

1.4.2. Theorem (McCann). *Let μ and ν be two probability measures on a compact, connected, $\mathscr{C}^3$ manifold without boundary, with μ absolutely continuous. If $\mathrm{d}(x, y)$ stands for the Riemannian distance between x and y, then there is a unique optimal transport plan $\gamma \in \Gamma(\mu, \nu)$ for the cost $c(x, y) = \frac{1}{2}\mathrm{d}(x, y)^2$, which is induced by the transport map $T(x) = \exp_x[-\nabla\psi(x)]$, with ψ Lipschitz and c-concave[2]. The Kantorovich potential ψ is unique up to an additive constant.*

1.4.3. Regularity. The regularity of the Kantorovich potential, for an arbitrary cost, is also very difficult question. During the past decade, a lot of progress has been made: a quite general theorem has been obtained by Xi-Nan Ma, Neil S. Trudinger, and Xu-Jia Wang [41]; a more specific result, on products of spheres, has been recently proved by Alessio Figalli, Young-Heon Kim, and Robert J. McCann [29].

Fortunately, chapter 3 does not require a very abstract theory: all we need is contained in the next theorem (§1.4.5, on page 33), based on Dario Cordero-Erausquin's pioneering work [21]. It gives the existence and regularity of the Kantorovich potential

[1] Dario Cordero-Erausquin [21] had already provided an extension to periodic measures.

[2] On a Riemannian manifold M, for any $v \in \mathrm{T}_x\mathrm{M}$, the point $\exp_x(v)$ is defined as the value at time 1 of the geodesic starting from x with initial velocity v.

for a quadratic cost $c : \mathbb{T}^d \times \mathbb{T}^d \to [0, \infty)$ induced by $\bar{c} : \mathbb{R}^d \times \mathbb{R}^d \to [0, \infty)$ given by

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

where $A \in \mathcal{S}_d^{++}$ is a symmetric, positive-definite matrix, and Az^2 is a shorthand for $\langle Az | z \rangle$. Such a cost arises when one changes the usual metric on $\mathbb{T}^d$ with the one induced by A in the canonical set of coordinates, and then takes half the resulting squared distance as a cost function.

Before stating and proving the theorem, we however need to adapt Yann Brenier's convex point of view to the torus. We have seen in section 1.3 that ψ is a c-concave map if and only if $\varphi(x) := \frac{1}{2}|x|^2 - \psi(x)$ is a lower semicontinuous convex map. Something similar is going on here for a quadratic cost, namely:

1.4.4. Lemma. *A map $\psi : \mathbb{T}^d \to \mathbb{R}$ is c-concave for the cost c induced by $A \in \mathcal{S}_d^{++}$, if and only if $\varphi(x) := \frac{1}{2}Ax^2 - \psi(x)$ is lower semicontinuous and convex on $\mathbb{R}^d$. Then,*

$$\psi^c(y) \;=\; \frac{1}{2}Ay^2 - \varphi^*(y),$$

where φ^* is the Legendre transform of φ for the scalar product induced by A. If ψ is $\mathscr{C}^2$ and such that $A - \nabla^2\psi > 0$, then $x \mapsto x - A^{-1}\nabla\psi(x)$ is a diffeomorphism $\mathbb{T}^d \to \mathbb{T}^d$.

Proof. If ψ is c-concave, then φ is convex and lower semi-continuous, for it can be written as a Legendre transform:

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

This also shows $\varphi^*(y) = \frac{1}{2}Ay^2 - \psi^c(y)$.

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

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

Therefore,

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

i.e. $\psi(x) = \psi^{cc}(x)$.

If ψ is $\mathscr{C}^2$ and such that $A - \nabla^2\psi > 0$, then $A - \nabla^2\psi \geq \varepsilon \, \mathrm{I}_d$ for some $\varepsilon > 0$. Thus, as φ is convex with a super-linear growth, $\nabla\varphi : \mathbb{R}^d \to \mathbb{R}^d$ is a diffeomorphism, and so is the map $T : x \mapsto x - A^{-1}\nabla\psi(x)$. Notice that, if $k \in \mathbb{Z}^d$, then $T(x+k) = T(x) + k$; therefore, T induces a diffeomorphism $\mathbb{T}^d \to \mathbb{T}^d$. □

1.4.5. Proposition. *Let μ and ν be two probability measures on $\mathbb{T}^d$ with smooth, strictly positive densities, and let c be the quadratic cost on $\mathbb{T}^d \times \mathbb{T}^d$ induced by a definite-positive, symmetric matrix A. Then there is a unique c-concave function $\psi : \mathbb{T}^d \to \mathbb{R}$ with $\int \psi \, \mathrm{d}\mu = 0$ such that $T : \mathbb{T}^d \to \mathbb{T}^d$ 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 $\varphi : x \mapsto \frac{1}{2}Ax^2 - \psi(x)$ is a smooth, strictly convex function on $\mathbb{R}^d$. Moreover, the transport map T is optimal for the cost c, and there is no other optimal transport plan but the one it induces.*

Proof. Let us denote by ∇_A the gradient for the metric induced by A. Then according to Robert J. McCann's theorem (§1.4.2, on page 31), there is a Lipschitz function $\psi : \mathbb{T}^d \to \mathbb{R}$ 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 $\int \psi \, \mathrm{d}\mu = 0$ is added, and moreover it is optimal for the Monge–Kantorovich problem. Here on the torus, $\exp_x[-\nabla_A\psi(x)] = x - A^{-1}\nabla\psi(x)$.

For any $x \in \mathbb{R}^d$, 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 $\mathbb{R}^d$. Moreover, according to lemma 1.4.4 on page 32, φ is a convex function. Now, let V be an open, convex subset of $\mathbb{R}^d$, 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 Luis A. Caffarelli's theorem (§1.3.5, on page 31), and thus we get that φ is strictly convex and smooth on U. As U is arbitrary, φ is strictly convex and smooth on $\mathbb{R}^d$. Thus, ψ is also smooth, and T is a diffeomorphism. □

1.5 The Wasserstein space

1.5.1. Wasserstein distance W_p. If μ and ν are two probability measures on a space X, which will be either the Euclidean space or a Riemannian manifold, then the minimal value for the Monge–Kantorovich problem defines a distance, dubbed the Wasserstein distance, when the cost is $c(x, y) = \mathrm{d}(x, y)^p$ with d the distance of X:

$$\mathrm{W}_p(\mu, \nu) := \left(\min_{\gamma \in \Gamma(\mu,\nu)} \int \mathrm{d}(x, y)^p \,\mathrm{d}\gamma(x, y) \right)^{1/p}.$$

1.5.2. Wasserstein space $\mathcal{P}_p(\mathrm{X})$. For the Wasserstein distance between μ and ν to be finite, it is enough for them to have finite pth-order moments. In other words, W_p is a distance on the following subset of the space $\mathcal{P}(\mathrm{X})$ of Borel probability measures on X:

$$\mathcal{P}_p(\mathrm{X}) := \left\{ \mu \in \mathcal{P}(\mathrm{X}) \,\middle|\, \forall x_0 \in \mathrm{X}, \int \mathrm{d}(x, x_0)^p \,\mathrm{d}\mu(x) < \infty \right\}.$$

Thanks to the triangular inequality, the condition "for all x_0" can be replaced by "there is at least one x_0".

1.5.3. Proposition. *For any $\mu, \nu \in \mathcal{P}_1(\mathrm{X})$,*

$$\mathrm{W}_1(\mu, \nu) := \inf_{\psi \in \mathrm{Lip}_1(\mathrm{X})} \int \psi \,\mathrm{d}(\mu - \nu).$$

Proof. This follows from proposition 1.1.6 on page 25, for if ψ is 1-Lipschitz, then $-\psi(y) \leq \mathrm{d}(x, y) - \psi(x)$ for any x, and thus $\psi^c = -\psi$. □

1.5.4. Proposition. *A sequence $(\mu_n)_{n\in\mathbb{N}}$ converges for the Wasserstein distance if and only if it narrowly converges and the pth-order moments converge as well. Therefore, if is X compact, then $\mathcal{P}_p(\mathrm{X}) = \mathcal{P}(\mathrm{X})$ is also compact.*

Proof. See Cédric Villani's first book [62, Theorem 7.12]. □

As will be shown by the next three propositions, optimal transport lies at the heart of the properties of the Wasserstein distance.

1.5.5. Proposition. *Let $\mu_0, \mu_1 \in \mathcal{P}_p(\mathbb{R}^d)$. Then, any $\gamma \in \Gamma(\mu, \nu)$ optimal for the Monge–Kantorovich problem induces a constant-speed geodesic $(\mu_t)_{t\in[0,1]}$, defined by*

$$\mu_t := [(1-t)\mathsf{X} + t\mathsf{Y}]_{\#}\,\gamma,$$

where $\mathsf{X}(x,y) := x$ and $\mathsf{Y}(x,y) := y$; that is,

$$\forall\,\xi \in \mathscr{C}_\mathrm{b}, \qquad \int \xi(z)\,\mathrm{d}\mu_t(z) = \int \xi((1-t)x + ty)\,\mathrm{d}\gamma(x,y).$$

Conversely, any constant-speed geodesic between μ_0 and μ_1 is induced by an optimal transport plan $\gamma \in \Gamma(\mu,\nu)$. Therefore, if μ_0 is absolutely continuous, there is an optimal transport map T between μ_0 and μ_1, and the geodesic is $\mu_t := [(1-t)\,\mathrm{Id} + tT]_{\#}\,\mu_0$.

This shows the Wasserstein space is a length space: the Wasserstein distance coincides with the distance induced by the geodesics.

Proof. Let γ be an optimal transport plan between μ_0 and μ_1. Let also $t \in [0,1]$, and define $\mathsf{Z}_t := (1-t)\mathsf{X} + t\mathsf{Y}$. Then, $\mu_t = [\mathsf{Z}_t]_{\#}\gamma$, and, for any $s \in [0,1]$, $(\mathsf{Z}_s, \mathsf{Z}_t)_{\#}\gamma$ is a transport plan between μ_s and μ_t. Therefore,

$$\begin{aligned} \mathrm{W}_p(\mu_s,\mu_t)^p &\leq \int \big|[(1-s)x + sy] - [(1-t)x + ty]\big|^p \,\mathrm{d}\gamma(x,y) \\ &\leq |t-s|^p\, \mathrm{W}_p(\mu_0,\mu_1)^p, \end{aligned}$$

that is $\mathrm{W}_p(\mu_s,\mu_t) \leq |t-s| \mathrm{W}_p(\mu_0,\mu_1)$. Were that inequality to be strict for a pair (s,t), the triangular inequality would yield $\mathrm{W}_p(\mu_0,\mu_1) < \mathrm{W}_p(\mu_0,\mu_1)$, which is obviously not possible. Thus,

$$\mathrm{W}_p(\mu_s,\mu_t) = |t-s|\mathrm{W}_p(\mu_0,\mu_1).$$

Conversely, if $(\mu_t)_{t\in[0,1]}$ is a constant speed geodesic, then, for any $t \in [0,1]$, it is possible to "glue" together two optimal transport plans to form $\pi \in \Gamma(\mu_0, \mu_t, \mu_1)$ such that $(\mathsf{X},\mathsf{Y})_\#\pi$ and $(\mathsf{Y},\mathsf{Z})_\#\pi$ are optimal plans between, respectively, μ_0 and μ_t on the one hand, and μ_t and μ_1 on the other hand—where

$$\mathsf{X}(x,y,z) = x, \qquad \mathsf{Y}(x,y,z) = y, \qquad \mathsf{Z}(x,y,z) = z.$$

We refer to Cédric Villani's first book on optimal transportation [62, Lemma 7.6] for a proof of this gluing lemma[3]. Then,

$$\begin{aligned} \mathrm{W}_p(\mu_0,\mu_1) \;\leq\; \|\mathsf{X}-\mathsf{Z}\|_{\mathrm{L}^p(\pi)} \;\leq\; \|\mathsf{X}-\mathsf{Y}\|_{\mathrm{L}^p(\pi)} + \|\mathsf{Y}-\mathsf{Z}\|_{\mathrm{L}^p(\pi)} \\ \leq\; \mathrm{W}_p(\mu_0,\mu_t) + \mathrm{W}_p(\mu_t,\mu_1) \;\leq\; \mathrm{W}_p(\mu_0,\mu_1). \end{aligned}$$

Thus, all the inequalities are, in fact, equalities. This implies $(\mathsf{X},\mathsf{Z})_\#\pi$ is optimal and there is $\alpha \in [0,1]$ such that $\mathsf{Y} = (1-\alpha)\mathsf{X} + \alpha\mathsf{Z}$ in $\mathrm{L}^p(\pi)$. Therefore, $\mathrm{W}_p(\mu_0,\mu_t) = t\mathrm{W}_p(\mu_0,\mu_1)$ yields $\alpha = t$. □

1.5.6. Proposition. *Let $\mu, \nu \in \mathcal{P}(K)$, with K a compact subset of $\mathbb{R}^d$ or a compact manifold. Then, for any $\bar\mu \in \mathcal{P}(K)$, there is a Kantorovich potential ψ between μ and ν for the cost $c(x,y) = \mathrm{d}(x,y)^p/p$ such that:*

$$\lim_{\varepsilon\to 0^+} \frac{\mathrm{W}_2((1-\varepsilon)\mu + \varepsilon\bar\mu, \nu)^2 - \mathrm{W}_2(\mu,\nu)^2}{2\varepsilon} \;=\; \int \psi \,\mathrm{d}(\bar\mu - \mu).$$

A priori, the potential ψ may depend on $\bar\mu$. However, it is obviously no longer the case if the Kantorovich potential is uniquely defined—e.g. if μ or ν is absolutely continuous and strictly positive.

Proof. Let ψ_ε be a Kantorovich potential between $(1-\varepsilon)\mu + \varepsilon\bar\mu$ and ν:

$$\int \psi_\varepsilon \,\mathrm{d}[(1-\varepsilon)\mu + \varepsilon\bar\mu] + \int \psi_\varepsilon^c \,\mathrm{d}\nu \;=\; \frac{1}{2}\mathrm{W}_2((1-\varepsilon)\mu + \varepsilon\bar\mu, \nu)^2.$$

[3] The same lemma allows to prove W_p is a distance

Then,

$$\frac{\mathrm{W}_2((1-\varepsilon)\mu+\varepsilon\bar{\mu},\nu)^2-\mathrm{W}_2(\mu,\nu)^2}{2\varepsilon} \;\leq\; \int \psi_\varepsilon\,\mathrm{d}(\bar{\mu}-\mu).$$

Since ψ_ε is c-concave,

$$\psi(x) \;=\; \inf_y\left\{\frac{1}{2}d(x,y)^2-\psi^c(y)\right\},$$

and consequently, as K is bounded, ψ_ε is Lipschitz with a constant that does not depend on ε; so is ψ_ε^c. By the Arzelà–Ascoli theorem, the family $\{(\psi_\varepsilon,\psi_\varepsilon^c)\}$ is therefore relatively compact. Let (ψ,φ) be a limit point such that

$$\limsup_{\varepsilon\to 0^+}\int \psi_\varepsilon\,\mathrm{d}(\bar{\mu}-\mu) \;=\; \int \psi\,\mathrm{d}(\bar{\mu}-\mu).$$

Then, since $\psi(x)+\varphi(y)\leq\frac{1}{2}|x-y|^2$, we have

$$\begin{aligned}
\frac{1}{2}\mathrm{W}_2(\mu,\nu) &\leq \liminf_{\varepsilon\to 0^+}\frac{1}{2}\mathrm{W}_2((1-\varepsilon)\mu+\varepsilon\bar{\mu},\nu)^2\\
&= \liminf_{\varepsilon\to 0^+}\left\{\int \psi_\varepsilon\,\mathrm{d}((1-\varepsilon)\mu+\varepsilon\bar{\mu})+\int \psi_\varepsilon^c\,\mathrm{d}\nu\right\}\\
&\leq \int \psi\,\mathrm{d}\mu+\int \varphi\,\mathrm{d}\nu\\
&\leq \frac{1}{2}\mathrm{W}_2(\mu,\nu)^2.
\end{aligned}$$

Thus, ψ is a Kantorovich potential between μ and ν, and

$$\begin{aligned}
\limsup_{\varepsilon\to 0^+}\frac{\mathrm{W}_2((1-\varepsilon)\mu+\varepsilon\bar{\mu},\nu)^2-\mathrm{W}_2(\mu,\nu)^2}{2\varepsilon} &\leq \limsup_{\varepsilon\to 0^+}\int \psi_\varepsilon\,\mathrm{d}(\bar{\mu}-\mu)\\
&\leq \int \psi\,\mathrm{d}(\bar{\mu}-\mu).
\end{aligned}$$

On the other hand,

$$\begin{aligned}
\frac{1}{2}\mathrm{W}_2((1-\varepsilon)\mu+\varepsilon\bar{\mu},\nu)^2 &\geq \int \psi\,\mathrm{d}((1-\varepsilon)\mu+\varepsilon\bar{\mu})+\int \psi^c\,\mathrm{d}\nu\\
&\geq \frac{1}{2}\mathrm{W}_2(\mu,\nu)^2+\varepsilon\int \psi\,\mathrm{d}(\bar{\mu}-\mu),
\end{aligned}$$

and this yields

$$\liminf_{\varepsilon\to 0^+} \frac{W_2((1-\varepsilon)\mu + \varepsilon\bar{\mu}, \nu)^2 - W_2(\mu,\nu)^2}{2\varepsilon} \geq \int \psi \, \mathrm{d}(\bar{\mu} - \mu). \qquad \square$$

1.5.7. Proposition. *Let $\mu, \nu \in \mathcal{P}(K)$, with K a compact subset of $\mathbb{R}^d$ or $K = \mathbb{T}^d$, and assume μ is absolutely continuous. Let ψ is the (unique up to an additive constant) Kantorovich potential between μ and ν for the cost $c(x, y) = \mathrm{d}(x, y)^p/p$. If ζ is a diffeomorphism of K, then*

$$\lim_{\varepsilon\to 0} \frac{W_2([\mathrm{Id} + \varepsilon\zeta]_\#\mu, \nu)^2 - W_2(\mu,\nu)^2}{2\varepsilon} = \int \langle \nabla\psi | \zeta \rangle \, \mathrm{d}\mu.$$

Proof. As ψ is a Kantorovich potential between μ and ν,

$$\frac{W_2([\mathrm{Id} + \varepsilon\zeta]_\#\mu, \nu)^2 - W_2(\mu,\nu)^2}{2\varepsilon} \geq \int \frac{\psi(x + \varepsilon\zeta(x)) - \psi(x)}{\varepsilon} \, \mathrm{d}\mu(x).$$

Since ψ is Lipschitz (because K is compact), it is differentiable almost everywhere. Thus, Lebesgue's dominated convergence theorem yields

$$\liminf_{\varepsilon\to 0^+} \frac{W_2([\mathrm{Id} + \varepsilon\zeta]_\#\mu, \nu)^2 - W_2(\mu,\nu)^2}{2\varepsilon} \geq \int \langle \nabla\psi(x) \mid \zeta(x) \rangle \, \mathrm{d}\mu(x).$$

Conversely, $\mathrm{Id} - \nabla\psi$ is an optimal map between μ and ν, so $(\mathrm{Id} + \varepsilon\zeta, \mathrm{Id} - \nabla\psi)_\#\mu$ is a transport plan between $[\mathrm{Id} + \varepsilon\zeta]_\#\mu$ and ν, and thus

$$\begin{aligned} W_2([\mathrm{Id} + \varepsilon\zeta]_\#\mu, \nu)^2 &\leq \int \left| [x + \varepsilon\zeta(x)] - [x - \nabla\psi(x)] \right|^2 \mathrm{d}\mu(x) \\ &\leq \int \left\{ \left| x - [x - \nabla\psi(x)] \right|^2 + 2\varepsilon \langle \nabla\psi(x) \mid \zeta(x) \rangle + \varepsilon^2 |\zeta(x)|^2 \right\} \mathrm{d}\mu(x) \\ &\leq W_2(\mu,\nu)^2 + \varepsilon \int \langle \nabla\psi(x) \mid \zeta(x) \rangle \, \mathrm{d}\mu(x) + \varepsilon^2 \int |\zeta(x)|^2 \, \mathrm{d}\mu(x). \end{aligned}$$

Hence

$$\limsup_{\varepsilon\to 0^+} \frac{W_2([\mathrm{Id} + \varepsilon\zeta]_\#\mu, \nu)^2 - W_2(\mu,\nu)^2}{2\varepsilon} \leq \int \langle \nabla\psi(x) \mid \zeta(x) \rangle \, \mathrm{d}\mu(x). \qquad \square$$

1.6 The Benamou–Brenier formula

1.6.1. According to proposition 1.5.5 on page 35, the geodesics in $\mathcal{P}_p(\mathbb{R}^d)$ are all induced by optimal transport plans. Jean-David Benamou and Yann Brenier [7] found another characterization, namely that the geodesics should minimize the average kinetic energy of the particles through the transport. That is, the geodesics should minimze

$$(\mu_t)_{t\in[0,1]} \longmapsto \int_0^1 \int |v_t(x)|^2 \, \mathrm{d}\mu_t(x) \, \mathrm{d}t,$$

among all the absolutely continuous curves $(\mu_t)_{t\in[0,1]}$ with $(v_t)_{t\in[0,1]}$ the associated velocity field given by the continuity equation:

$$\frac{\mathrm{d}v}{\mathrm{d}t} + \operatorname{div}(v\mu) \;=\; 0.$$

From this, they derived a method to numerically solve the Monge–Kantorovich problem.

1.6.2. Metric derivative. If $(\mu_t)_{t\in I}$ is an absolutely continuous curve in $\mathcal{P}_p(X)$, i.e. if there is $g \in \mathrm{L}^1(I)$ such that

$$\forall\, s, t \in I, \quad \mathrm{W}_p(\mu_s, \mu_t) \;\le\; \int_{[s,t]} g(\omega)\, \mathrm{d}\omega,$$

then, for almost every $t \in I$, the limit

$$|\dot{\mu}|_t \;:=\; \limsup_{h\to 0} \frac{\mathrm{W}_p(\mu_t, \mu_{t+h})}{|h|}$$

exists, and is called the metric derivative or μ. Then, $|\dot{\mu}| \le g$ and $\mathrm{W}_p(\mu_s, \mu_t) \le \int_s^t |\dot{\mu}|$.

Proof. Let $\{t_n\}_{n\in\mathbb{N}}$ be a dense subset of I, and let $d_n(t) := \mathrm{W}_p(\mu_{t_n}, \mu_t)$. Then if $s \le t$, we have $|d_n(s) - d_n(t)| \le \mathrm{W}_p(\mu_s, \mu_t) \le \int_s^t g$, so d_n is absolutely continuous and $|d_n'(t)| \le g(t)$. We set $e(t) := \sup |d_n'(t)|$. If all the d_n are differentiable in t—this is the case almost everywhere—, then

$$e(t) = \sup_{n\in\mathbb{N}} |d_n'(t)| = \sup_{n\in\mathbb{N}} \lim_{h\to 0} \frac{|d_n(t) - d_n(t+h)|}{|h|} \le \liminf_{t\to 0} \frac{\mathrm{W}_p(\mu_t, \mu_{t+h})}{|h|}.$$

But $\{t_n\}$ is dense in I, so

$$\mathrm{W}_p(\mu_t, \mu_{t+h}) = \sup_{n\in\mathbb{N}} |d_n(t+h) - d_n(t)| \leq \sup_{n\in\mathbb{N}} \int_{[t,t+h]} |d_n'(\omega)| \,\mathrm{d}\omega \leq \int_{[t,t+h]} e(\omega)\,\mathrm{d}\omega.$$

By the Lebesgue differentiation theorem, this shows that $|\dot{\mu}|$ exists almost everywhere, $|\dot{\mu}| = e$, and since $\mathrm{W}_p(\mu_t, \mu_{t+h}) \leq \int_{[t,t+h]} g$, this also shows $|\dot{\mu}| \leq g$. □

1.6.3. Lemma. *Let $(\mu_t)_{t\in I}$ be an absolutely continuous curve in $\mathcal{P}_p(\mathbb{R}^d)$. Then, there is a vector field $v : I \times \mathbb{R}^d \to \mathbb{R}^d$ such that $|\dot{\mu}|_t = \|v_t\|_{\mathrm{L}^p(\mu_t)}$ for almost all $t \in I$. Moreover, in the distributional sense,*

$$\frac{\mathrm{d}\mu}{\mathrm{d}t} + \mathrm{div}(v\mu) \;=\; 0.$$

Conversely, if there is a vector field $v : I \times \mathbb{R}^d \to \mathbb{R}^d$ such that, in the distributional sense,

$$\frac{\mathrm{d}\mu}{\mathrm{d}t} + \mathrm{div}(v\mu) = 0 \qquad \textit{with} \qquad \int_I \|v_t\|_{\mathrm{L}^p(\mu_t)}\,\mathrm{d}t < \infty,$$

then $(\mu_t)_{t\in I}$ is absolutely continuous, and $|\dot{\mu}|_t \leq \|v_t\|_{\mathrm{L}^p(\mu_t)}$ for almost all $t \in I$.

For the proof of this lemma, we refer to the monograph by Luigi Ambrosio, Nicola Gigli, and Giuseppe Savaré [3, Theorem 8.3.1].

1.6.4. Theorem (Benamou–Brenier). *Let $p \in (1, \infty)$ and $\mu_0, \mu_1 \in \mathcal{P}_p(\mathrm{X})$ with $\mathrm{X} = \mathbb{R}^d$ or $\mathrm{X} = \mathbb{T}^d$. Then*

$$\mathrm{W}_p(\mu_0, \mu_1)^p \;=\; \inf_{v,\rho} \int_{[0,1]} \int_{\mathrm{X}} |v_t(x)|^p \,\mathrm{d}\mu_t(x)\,\mathrm{d}t,$$

where the infimum runs among all pairs (μ, v) such that $(\mu_t)_{t\in[0,1]}$ is a continuous curve between μ_0 and μ_1, and $v : [0,1] \times \mathrm{X} \to \mathbb{R}^d$ is a vector field such that $v_t \in \mathrm{L}^p(\mu_t)$ for almost all $t \in [0,1]$, and, in the distributional sense,

$$\frac{\mathrm{d}\mu}{\mathrm{d}t} + \mathrm{div}(v\mu) \;=\; 0.$$

Proof. The case $\mathrm{X} = \mathbb{R}^d$ directly follows from lemma 1.6.3 on the current page, equality being reached with a constant speed geodesic. Let us nevertheless give a demonstration for $\mathrm{X} = \mathbb{T}^d$, inspired from the original article by Jean-David Benamou and Yann Brenier [7], and the aforementioned book by Luigi Ambrosio, Nicola Gigli, and Giuseppe Savaré [3, Chapter 8]. The reader may also refer to a proof by Kevin Guittet [32].

1. Let (μ_t) be a curve of absolutely continuous, smooth probability measures on $\mathbb{T}^d$, and let v be a vector field smooth in space that, together with (μ_t), solve the continuity equation. Assume

$$\int_0^1 \|v_t\|_{\mathscr{C}^1}\,\mathrm{d}t \;<\; \infty.$$

Then, the solution $t \mapsto X_{s,t}(x)$ of the equation $\mathrm{d}X_{s,t}/\mathrm{d}t = v_t(X_{s,t})$, with $X_{s,s}(x) = x$, is defined for all $t \in [0,1]$ (if we were not working on a compact space without boundary, there would be a difficulty here). For $\xi \in \mathscr{C}^\infty([0,1] \times \mathbb{T}^d)$, we set

$$\varphi_t(x) \;:=\; -\int_t^1 \xi_s(X_{t,s}(x))\,\mathrm{d}s.$$

Since $X_{t,s}(X_{0,t}(x)) = X_{0,s}(x)$,

$$\frac{\mathrm{d}\varphi_t}{\mathrm{d}t}(X_{0,t}) + \langle v_t(X_{0,t}) \,|\, \nabla\varphi_t(X_{0,t})\rangle \;=\; \frac{\mathrm{d}}{\mathrm{d}t}\left[\varphi_t(X_{0,t})\right] \;=\; \xi_t(X_{0,t}),$$

and as $x \mapsto X_{0,t}(x)$ is a diffeomorphism, this implies $\mathrm{d}\varphi/\mathrm{d}t + \langle v \,|\, \nabla\varphi\rangle = \xi$. Thus,

$$\int_0^1\!\!\int \left[\frac{\mathrm{d}\varphi_t}{\mathrm{d}t}(x) + \langle v_t(x) \,|\, \varphi_t(x)\rangle\right] \mathrm{d}\mu_t(x)\,\mathrm{d}t \;=\; \int_0^1\!\!\int \xi_t(x)\,\mathrm{d}\mu_t(x)\,\mathrm{d}t.$$

On the other hand, since (μ, v) solves the continuity equation and $\varphi_1 = 0$,

$$\int_0^1\!\!\int \left[\frac{\mathrm{d}\varphi_t}{\mathrm{d}t}(x) + \langle v_t(x) \,|\, \nabla\varphi_t(x)\rangle\right] \mathrm{d}\mu_t(x)\,\mathrm{d}t \;=\; -\int \varphi_0(x)\,\mathrm{d}\mu_0(x).$$

This implies $\mu_t = [X_{0,t}]_\#\mu_0$. Indeed, let $\bar\mu_t := [X_{0,t}]_\#\mu_0$, and $\sigma := \bar\mu - \mu$. Then, according to the previous computations, which also hold for $\bar\mu$, for any $\xi \in \mathscr{C}^\infty([0,1]\times \mathbb{T}^d)$,

$$\int_0^1\!\!\int \xi_t(x)\,\mathrm{d}\sigma_t(x)\,\mathrm{d}t \;=\; -\int \varphi_0(x)\,\mathrm{d}\sigma_0(x) \;=\; 0.$$

Therefore, since $[X_{0,1}]_\# \mu_0 = \mu_1$,

$$\mathrm{W}_p(\mu_0, \mu_1)^p \;\leq\; \int |X_{0,1}(x) - x|^p \,\mathrm{d}\mu_0(x) \;\leq\; \int \int_0^1 |X'_{0,t}(x)|^p \,\mathrm{d}t \,\mathrm{d}\mu_0(x)$$

and thus, as $X'_{s,t} = v_t(X_{s,t})$ and $\mu_t = [X_{0,t}]_\# \, \mu_0$,

$$\mathrm{W}_p(\mu_0, \mu_1)^p \;\leq\; \int_0^1 \int |v_t(X_{0,t}(x))|^p \,\mathrm{d}\mu_0(x) \,\mathrm{d}t \;\leq\; \int_0^1 \int |v_t(x)|^p \,\mathrm{d}\mu_t(x) \,\mathrm{d}t.$$

2. We no longer assume anything about μ_0 and μ_1, but that they are probability measures on $\mathbb{T}^d$. If (μ_t) is a continuous curve between them, and if v is a vector field solving the continuity equation, with $v_t \in \mathrm{L}^p(\mu_t)$, then, taking a positive mollifier φ_ε, we set $\mu_t^\varepsilon = \varphi_\varepsilon * \mu_t$. As $\varphi_\varepsilon * (v_t \mu_t)$ is absolutely continuous, it has a density v_t^ε with respect to μ_t^ε, which is positive. Thus, $(\mu^\varepsilon, v^\varepsilon)$ also solves the continuity equation, and (μ_t^ε) is still a continuous curve. Moreover, setting $m_\varepsilon = \min \varphi_\varepsilon$,

$$\|v_t^\varepsilon\|_{\mathscr{C}^1} \;\leq\; \frac{\|\varphi_\varepsilon * (v_t \mu_t)\|_{\mathscr{C}^1} (m_\varepsilon + \|\mu_t^\varepsilon\|_{\mathscr{C}^1}}{m_\varepsilon^2}.$$

But, $\|\mu_t^\varepsilon\|_{\mathscr{C}^1} < C_\varepsilon$, and, as

$$\begin{aligned}
\varphi_\varepsilon * (v_t \mu_t)(x) \;&=\; \int \varphi_\varepsilon(x - y) v_t(y) \,\mathrm{d}\mu_t(y) \\
&\leq\; \|\varphi_\varepsilon(x - \cdot\,)\|_{\mathrm{L}^q(\mu_t)} \|v_t\|_{\mathrm{L}^p(\mu_t)} \\
&\leq\; \|\varphi_\varepsilon\|_{\mathscr{C}^0} \|v_t\|_{\mathrm{L}^p(\mu_t)},
\end{aligned}$$

we must also have $\|\varphi_\varepsilon * (v_t \mu_t)\|_{\mathscr{C}^1} \leq C_\varepsilon \|v_t\|_{\mathrm{L}^p(\mu_t)}$. We can therefore assume

$$\int_0^1 \|v_t^\varepsilon\|_{\mathscr{C}^1} \,\mathrm{d}t \;\leq\; C_\varepsilon \int_0^1 \|v_t\|_{\mathrm{L}^p(\mu_t)} \,\mathrm{d}t \;\leq\; C_\varepsilon \left(\int_0^1 |v_t(x)|^p \,\mathrm{d}\mu_t(x) \,\mathrm{d}t \right)^{1/p} < \infty.$$

Then, according to the previous computations,

$$\mathrm{W}_p(\mu_0^\varepsilon, \mu_1^\varepsilon)^p \;\leq\; \int_0^1 \int |v_t^\varepsilon(x)|^p \,\mathrm{d}\mu_t^\varepsilon(x) \,\mathrm{d}t.$$

According to Jensen's inequality, since $(a, b) \mapsto |a|^p / b^{p-1}$ is convex and homogeneous of degree 1 on $\mathbb{R} \times (0, \infty)$, and $\varphi_\varepsilon(x - \cdot)\mu_t$ is a bounded measure,

$$|v_t^\varepsilon(x)|^p \mu_t^\varepsilon(x) \;=\; \frac{\left|\int v_t(y)\varphi_\varepsilon(x-y)\,\mathrm{d}\mu_t(y)\right|^p}{\left(\int \varphi_\varepsilon(x-y)\,\mathrm{d}\mu_t(y)\right)^{p-1}} \;\leq\; \int |v_t(y)|^p \varphi_\varepsilon(x-y)\,\mathrm{d}\mu_t(y).$$

Thus,

$$\mathrm{W}_p(\mu_0^\varepsilon, \mu_1^\varepsilon)^p \;\leq\; \int_0^1 \int |v_t^\varepsilon(x)|^p \,\mathrm{d}\mu_t^\varepsilon(x)\,\mathrm{d}t \;\leq\; \int_0^1 \int |v_t(y)|^p \,\mathrm{d}\mu_t(y)\,\mathrm{d}t.$$

Then, letting $\varepsilon \to 0$, we finally get

$$\mathrm{W}_p(\mu_0, \mu_1)^p \;\leq\; \int_0^1 \int |v_t(y)|^p \,\mathrm{d}\mu_t(y)\,\mathrm{d}t.$$

3. Conversely, if μ_0 and μ_1 are absolutely continuous, with strictly positive, smooth densities, then according to Yann Brenier's theorem (§1.3.3, on page 29) and Luis A. Caffarelli's theorem (§1.3.5, on page 31), there is a diffeomorphism $T : \mathbb{T}^d \to \mathbb{T}^d$ such that $\mu_1 = T_\# \mu_0$. Then, if we set $v_t = (T - \mathrm{Id}) \circ [(1-t)\,\mathrm{Id} + tT]^{-1}$ and let μ_t be the density of $[(1-t)\,\mathrm{Id} + tT]_\# \,\mu$, we get

$$\frac{\mathrm{d}\mu}{\mathrm{d}t} + \mathrm{div}(v\mu) = 0 \qquad \text{and} \qquad \mathrm{W}_p(\mu_0, \mu_1)^p = \int_0^1 \int |v_t(x)|^p \,\mathrm{d}\mu_t(x)\,\mathrm{d}t.$$

4. In the general case, let $\gamma \in \Gamma_\mathrm{o}(\mu_0, \mu_1)$ be an optimal plan, and let $(\mu_t)_{t \in [0,1]}$ be the geodesic induced by γ. Define then a probability measure π on $[0, 1] \times \mathbb{T}^d$ with

$$\begin{aligned} \int \xi(t, z)\,\mathrm{d}\pi(t, z) \;&=\; \int_0^1 \int \xi(t, z)\,\mathrm{d}\mu_t(z)\,\mathrm{d}t \\ &=\; \int_0^1 \int \xi(t, (1-t)x + ty)\,\mathrm{d}\gamma(x, y)\,\mathrm{d}t. \end{aligned}$$

Then, if $\xi \in \mathscr{C}_\mathrm{c}^\infty((0, 1) \times \mathbb{T}^d)$,

$$\int \{\xi_{t+h}(z) - \xi_t(z)\}\,\mathrm{d}\pi(t, z)$$

$$
\begin{aligned}
&= \int_0^1 \left(\int \xi_t(z)\, \mathrm{d}\mu_{t-h}(z) - \int \xi_t(x)\, \mathrm{d}\mu_t(z) \right) \mathrm{d}t \\
&\leq \int_0^1 \int \{\xi_t((1-t+h)x + (t-h)y) - \xi_t((1-t)x + ty)\}\, \mathrm{d}\gamma(x,y)\, \mathrm{d}t \\
&\leq h \int_0^1 \int \int_0^1 \langle \nabla \xi_t((1-t+sh)x + (t-sh)y) \mid x - y \rangle\, \mathrm{d}s\, \mathrm{d}\gamma(x,y)\, \mathrm{d}t \\
&\leq h \mathrm{W}_p(\mu_0, \mu_1) \left(\int_0^1 \iint_0^1 \left| \nabla \xi_t((1-t+sh)x + (t-sh)y) \right|^q \mathrm{d}s\, \mathrm{d}\gamma(x,y) \mathrm{d}t \right)^{1/q}
\end{aligned}
$$

and thus, dividing by h and letting $h \to 0$,

$$
\int \frac{\mathrm{d}}{\mathrm{d}t} \xi_t(z)\, \mathrm{d}\pi(t,z) \;\leq\; \mathrm{W}_p(\mu_0, \mu_1) \|\nabla \xi\|_{\mathrm{L}^q(\pi)}.
$$

For $\xi \in \mathscr{C}^\infty((0,1) \times \mathbb{T}^d)$, we set

$$
L(\nabla \xi) \;:=\; -\int \frac{\mathrm{d}}{\mathrm{d}t} \xi_t(z)\, \mathrm{d}\pi(t,z),
$$

then this L can be extended into a continuous linear form on $\mathrm{L}^q(\pi)$. Thus, there is $v \in \mathrm{L}^p(\pi)$ such that $\|v\|_{\mathrm{L}^p(\pi)} \leq \mathrm{W}_p(\mu_0, \mu_1)$ and

$$
\forall\, \xi \in \mathscr{C}^\infty((0,1) \times \mathbb{T}^d), \qquad \int \left\{ \frac{\mathrm{d}}{\mathrm{d}t} \xi_t(z) + \langle v_t(z) \mid \nabla \xi_t(z) \rangle \right\} \mathrm{d}\pi(t,z) = 0.
$$

This implies that

$$
\frac{\mathrm{d}\mu}{\mathrm{d}t} + \mathrm{div}(v\mu) = 0 \qquad \text{and} \qquad \int_0^1 \int |v_t(z)|^p\, \mathrm{d}\mu_t(z)\, \mathrm{d}t \;\leq\; \mathrm{W}_p(\mu_0, \mu_1)^p. \qquad \square
$$

Chapter 2

The inverse function theorem of Nash and Moser

2.0.1. The Nash–Moser theorem is an extension of the well-known inverse function theorem to maps between Fréchet spaces. The first steps toward such a theorem were made in 1956 by John Nash [48], in his proof that one could embed any Riemannian manifold into some Euclidean space. A decade later, Jürgen Moser [46, 47] exposed a general method, which has ever since known many applications and developments. We will here follow the presentation made by Richard S. Hamilton [33], though keeping only the elements required to come to a minimal working statement, which is enough to satisfy our needs.

2.0.2. Compared with the standard inverse function theorem, two conditions need to be added for a map ζ between two Fréchet spaces to be invertible near 0: first, that $D\zeta$ itself be invertible on a whole neighborhood, since this does not follow any longer from the invertibility of $D\zeta(0)$; second, there should be "reasonable" bounds on $\zeta, D\zeta$, and $[D\zeta]^{-1}$, i.e. $\|\zeta(u)\|_n$ for instance should be bounded at most by $1 + \|u\|_{n+r}$ for some constant $r \geq 0$ independent of n. In mathematical terms, we will say that $\zeta, D\zeta$ and $[D\zeta]^{-1}$ need to satisfy some "tame" estimates.

If $\zeta(0) = 0$, and v is fixed and close to 0, a way to get u such that $v = \zeta(u)$ is to use a continuous version of Newton's method[1] and find a solution of the equation

$$u'(t) = \lambda[\mathrm{D}\zeta(u(t))]^{-1}(v - \zeta(u(t))),$$

starting for instance from $u(0) = 0$, since then $\zeta(u(t)) = (1 - e^{-\lambda t})v \to v$. As such an ODE might not have a solution if ζ is a map between Fréchet spaces, we will use a smooth family of operators $(S_t)_{t\geq 0}$ such that $S_t \to \mathrm{Id}$ when $t \to \infty$ and each S_t takes its values in a finite-dimensional subspace, and then solve

$$u'(t) = \lambda[\mathrm{D}\zeta(S_t u(t))]^{-1} S_t(v - \zeta(u(t))).$$

The existence of an appropriate family of finite-dimensional subspaces will be guaranteed by working on a particular class of Fréchet spaces, the so-called "tame" Fréchet spaces. Fortunately, the Fréchet space we are interested in, namely $\mathscr{C}^\infty(\mathbb{T}^d)$, is tame—as will be shown in the last section.

2.1 Definitions and statements

2.1.1. We will now state the theorem and its implicit-function corollary, but, beforehand, we need to introduce a few definitions.

2.1.2. Graded Fréchet space. This is the name given to a Fréchet space F endowed with a family of increasingly stronger seminorms $(\|\cdot\|_{n\in\mathbb{N}})$, so that

$$\forall\, u \in \mathrm{F},\ \|u\|_0 \leq \|u\|_1 \leq \cdots \leq \|u\|_n \leq \cdots .$$

For any $q \in \mathbb{N}$ and $\rho > 0$, we set

$$B_q(\rho) := \left\{ h \in \mathrm{F} \,\middle|\, \|h\|_q < \rho \right\} \qquad \text{and} \qquad \overline{B}_q(\rho) := \left\{ h \in \mathrm{F} \,\middle|\, \|h\|_q \leq \rho \right\}.$$

[1] Ivar Ekeland recently showed this is not the only way, by proving a more general inverse function theorem using his variational principle instead of Newton's method [26].

2.1.3. Tame linear map, tame isomorphism. A tame linear map $L : \mathrm{F} \to \mathrm{G}$ is a linear map between two graded Fréchet spaces such that, for some $r, b \in \mathbb{N}$, the following *tame estimate* is satisfied:

$$\forall\, n \geq b,\ \exists\, C_n > 0,\ \forall\, y \in \mathrm{F},\ \|Lu\|_n^{\mathrm{G}} \leq C_n \|u\|_{n+r}^{\mathrm{F}}.$$

Such a map L is a tame isomorphism if it is invertible, and both L and L^{-1} are tame linear maps.

2.1.4. Set of exponentially decreasing sequences $\Sigma(\mathrm{E})$. For a Banach space $(\mathrm{E}, \|\cdot\|_{\mathrm{E}})$, the set of exponentially decreasing sequences in E is defined by:

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

Endowed with the seminorms $(\|\cdot\|_n)_{n \in \mathbb{N}}$, it is a graded Fréchet space. Notice that the seminorms could also be defined for $n \leq 0$.

2.1.5. Tame Fréchet space. A graded Fréchet space F is called a tame Fréchet space if there is a Banach space E and two tame linear maps $L : \mathrm{F} \to \Sigma(\mathrm{E})$ and $K : \Sigma(\mathrm{E}) \to \mathrm{F}$ such that $K \circ L$ is the identity of F.

2.1.6. Tame map, tame estimate. Let F and G be two graded Fréchet spaces, and $\Omega \subset \mathrm{F}$ be an open subset. A map $\zeta : \Omega \to \mathrm{G}$ is said to be tame if it is continuous and, for every point $u_0 \in \Omega$, there is a neigkorhood U_0 of u_0 and some $r, b \in \mathbb{N}$ such that

$$\forall\, n \geq b,\ \exists\, C_n > 0,\ \forall\, u \in U_0,\ \|\zeta(u)\|_n^{\mathrm{G}} \leq C_n \left(1 + \|u\|_{n+r}^{\mathrm{F}}\right).$$

2.1.7. Smooth tame map. Let F and G be two graded Fréchet spaces, and $\Omega \subset \mathrm{F}$ be an open subset. A map $\zeta : \Omega \to \mathrm{G}$ is said to be a smooth tame map if it is smooth and all its Gâteau derivatives are tame.

2.1.8. Theorem (Nash–Moser). *Let* F *and* G *be two tame Fréchet spaces, and* $\Omega \subset \mathrm{F}$ *an open subset. Let* $\zeta : \Omega \to \mathrm{G}$ *be a smooth tame map such that, for any* $u \in \Omega$*,*

$D\zeta(u) : F \to G$ *is invertible. If* $[D\zeta]^{-1} : \Omega \times G \to F$ *is a smooth tame map, then* ζ *is locally invertible, and, locally, its inverse is always a smooth tame map.*

2.1.9. Corollary (implicit functions theorem). *Let* F, G *and* H *be tame spaces,* U_0 *an open subset of* F, V_0 *an open subset of* G. *Assume that* $\xi : U_0 \times V_0 \to H$ *is a smooth tame map, and that there are* $u_0 \in U_0$, $v_0 \in V_0$ *such that* $\xi(u_0, v_0) = 0$. *If, for all* $u \in U_0$, $v \in V_0$, $w \in H$, *there is a unique* $h \in G$ *such that* $D_v\zeta(u, v)h = w$, *and* h, *seen as a function of* u, v *and* w, *is a smooth tame map, then there are* $U \subset U_0$ *an open neighborhood of* u_0, $V \subset V_0$ *an open neighborhood of* v_0, *and a smooth tame map* $\nu : U \to V$ *such that*

$$\forall u \in U,\ \forall v \in V, \quad \xi(u, v) = 0 \iff v = \nu(u).$$

Proof of the corollary. We define a smooth tame map $\zeta : U_0 \times V_0 \to F \times H$ by setting

$$\zeta(u, v) = (u, \xi(u, v)).$$

Then, for all $(u, v) \in U_0 \times V_0$,

$$D\zeta(u, v) = \begin{pmatrix} \mathrm{Id} & 0 \\ D_u\zeta(u, v) & D_v\zeta(u, v) \end{pmatrix}$$

is invertible, and $(u, v, q, w) \mapsto [D\zeta(u, v)]^{-1}(q, w)$ is a smooth tame map. Therefore, according to the Nash–Moser theorem, in a neighborhood $U_1 \times V$ of (u_0, v_0), ζ is invertible, and $\zeta^{-1} : \zeta(U_1 \times V) \to U_1 \times V$ is a smooth tame map. Let $U_2 \times W \subset \zeta(U_1 \times V)$ be a neighborhood of $(u_0, 0)$, and $U \times V' \subset \zeta^{-1}(U_2 \times W)$ be a neighborhood of (u_0, v_0). We then take $\nu : U \to V$ such that

$$(u, \nu(u)) = \zeta^{-1}(u, 0).$$

□

2.2 Organization of the proof

2.2.1. In the next paragraphs, let us simplify the proof we need to give by a sequence of reductions to easier situations. The injectivity of ζ will then be proved in section 2.3

(proposition 2.3.1 on page 51). In section 2.4, we will introduce the smoothing operators that will allow us to prove the surjectivity in section 2.5 (proposition 2.5.7 on page 63). At last, we will deal with the smooth-tameness of ζ^{-1} in section 2.6 (proposition 2.6.3 on page 66).

2.2.2. Lemma. *It is possible to assume* $0 \in \Omega$, *with* $\zeta(0) = 0$, *and* $\mathrm{F} = \mathrm{G} = \Sigma(\mathrm{E})$, *for some Banach space* E.

Proof. Since $\mathrm{D}\zeta(0) : \mathrm{F} \to \mathrm{G}$ is linear tame and invertible, with an inverse map $[\mathrm{D}\zeta(0)]^{-1} : \mathrm{G} \to \mathrm{F}$ which is also linear tame, F and G are isomorphic and can be identified. Since F is a tame Fréchet space, we can assume $\mathrm{F} = \mathrm{G} = \Sigma(\mathrm{E})$ for some Banach space E. □

2.2.3. Lemma. *One can assume that there is* $r_0 \in \mathbb{N}$ *such that, for all* $n \geq 0$, *there is* $C_n > 0$ *such that, if* $\|u\|_{r_0} \leq 1$, *for all* $h, k \in \Sigma(\mathrm{E})$,

$$\|\zeta(u)\|_n \leq C_n \|u\|_{n+r_0}, \tag{2.2.3.a}$$

$$\|\mathrm{D}\zeta(u)h\|_n \leq C_n \left(\|h\|_{n+r_0} + \|h\|_0 \|u\|_{n+r_0}\right), \tag{2.2.3.b}$$

$$\|\mathrm{D}^2\zeta(u)h_1h_2\|_n \leq C_n \left(\|h_1\|_0\|h_2\|_{n+r_0} + \|h_1\|_{n+r_0}\|h_2\|_0 + \|h_1\|_0\|h_2\|_0\|u\|_{n+r_0}\right), \tag{2.2.3.c}$$

$$\|[\mathrm{D}\zeta(u)]^{-1}k\|_n \leq C_n \left(\|k\|_n + \|k\|_0\|u\|_{n+r_0}\right). \tag{2.2.3.d}$$

Proof. Since $\zeta, \mathrm{D}\zeta, \mathrm{D}^2\zeta$ and $[\mathrm{D}\zeta]^{-1}$ are all tame, there is a neighborhood U_0 of 0, and $r, b \in \mathbb{N}$ such that, if $u, h, h_1, h_2, k \in U_0$, for any $n \geq b$,

$$\begin{aligned}
\|\zeta(u)\|_n &\leq C_n (1 + \|u\|_{n+r}), \\
\|\mathrm{D}\zeta(u)h\|_n &\leq C_n (1 + \|h\|_{n+r} + \|u\|_{n+r}), \\
\|\mathrm{D}^2\zeta(u)h_1h_2\|_n &\leq C_n (1 + \|h_1\|_{n+r} + \|h_2\|_{n+r} + \|u\|_{n+r}), \\
\|[\mathrm{D}\zeta(u)]^{-1}k\|_n &\leq C_n (1 + \|k\|_{n+r} + \|u\|_{n+r}).
\end{aligned}$$

This neighborhood U_0 necessarily contains a small ball $B_a(2\rho)$, and we can assume $a \geq r$. Then, since for any $h \in \Sigma(\mathrm{E})$, the vector $\rho h/\|h\|_a$ is in U_0, we obtain that for any $u \in \Sigma(\mathrm{E})$ such that $\|u\|_a \leq \rho$ and any $h, h_1, h_2, k \in \Sigma(\mathrm{E})$,

$$\begin{aligned}
\|\zeta(u)\|_n &\leq C_n\left(1+\|u\|_{n+r}\right), \\
\|\mathrm{D}\zeta(u)h\|_n &\leq C_n\left(\|h\|_{n+r}+\|h\|_a\|u\|_{n+r}\right), \\
\|\mathrm{D}^2\zeta(u)h_1h_2\|_n &\leq C_n\left(\|h_1\|_{n+r}\|h_2\|_a+\|h_1\|_a\|h_2\|_{n+r}+\|h_1\|_a\|h_2\|_a\|u\|_{n+r}\right), \\
\|[\mathrm{D}\zeta(u)]^{-1}k\|_n &\leq C_n\left(\|k\|_{n+r}+\|k\|_a\|u\|_{n+r}\right).
\end{aligned}$$

For any $q \in \mathbb{Z}$, we define $\tau : \Sigma(\mathrm{E}) \to \Sigma(\mathrm{E})$ by $\tau_q(u)_k = e^{qk}u_k$. Then,

$$\|\tau_q(u)\|_n = \sum_{k=0}^{\infty} e^{k(n+q)}\|u_k\| = \|u\|_{n+q}$$

Thus, τ_q is a tame linear map. So, for any $p, q \in \mathbb{Z}$, the map $\tau_p \circ \zeta \circ \tau_q$ is still smooth tame, and

$$\begin{aligned}
\mathrm{D}(\tau_p \circ \zeta \circ \tau_q)(u)h &= \tau_p\left([\mathrm{D}\zeta(\tau_q(u))]\tau_q(h)\right), \\
\mathrm{D}^2(\tau_p \circ \zeta \circ \tau_q)(u)h_1h_2 &= \tau_p\left([\mathrm{D}^2\zeta(\tau_q(u))]\tau_q(h_1)\tau_q(h_2)\right), \\
[\mathrm{D}(\tau_p \circ \zeta \circ \tau_q)(u)]^{-1}k &= \tau_{-q}\left([\mathrm{D}\zeta(\tau_q(u))]^{-1}\tau_{-p}(k)\right).
\end{aligned}$$

Therefore, if we replace ζ with $\tau_p \circ \zeta \circ \tau_q$ for some p, q, and compose with a dilatation so as to have an estimate on a ball of radius 1, then, if $n + p \geq b$ and $n - q \geq b$ and $\|u\|_{a+q} \leq 1$,

$$\begin{aligned}
\|\zeta(u)\|_n &\leq C_n\left(1+\|u\|_{n+p+r+q}\right), \\
\|\mathrm{D}\zeta(u)h\|_n &\leq C_n\left(\|h\|_{n+p+r+q}+\|h\|_{a+q}\|u\|_{n+p+r+q}\right), \\
\|\mathrm{D}^2\zeta(u)h_1h_2\|_n &\leq C_n\Big(\|h_1\|_{n+p+r+q}\|h_2\|_{a+q}+\|h_1\|_{a+q}\|h_2\|_{n+p+r+q} \\
&\qquad +\|h_1\|_{a+q}\|h_2\|_{a+q}\|u\|_{n+p+r+q}\Big), \\
\|[\mathrm{D}\zeta(u)]^{-1}k\|_n &\leq C_n\left(\|k\|_{n-q+r-p}+\|k\|_{a-p}\|u\|_{n-q+r+q}\right).
\end{aligned}$$

Increasing a, b, and r if necessary, we will assume $a = b = r$. Then, for $p = 2r$, $q = -r$, we get that, if $n \geq b - p = -r$ and $n \geq q + b = 0$, and if $\|u\|_0 \leq 1$,

$$\begin{aligned}
\|\zeta(u)\|_n &\leq C_n\left(1 + \|u\|_{n+2r}\right), \\
\|D\zeta(u)h\|_n &\leq C_n\left(\|h\|_{n+2r} + \|h\|_0\|u\|_{n+2r}\right), \\
\|D^2\zeta(u)h_1h_2\|_n &\leq C_n\left(\|h_1\|_{n+2r}\|h_2\|_0 + \|h_1\|_0\|h_2\|_{n+2r} + \|h_1\|_0\|h_2\|_0\|u\|_{n+2r}\right), \\
\|[D\zeta(u)]^{-1}k\|_n &\leq C_n\left(\|k\|_n + \|k\|_{-r}\|u\|_{n+r}\right).
\end{aligned}$$

We then set $r_0 = 2r$. As $\zeta(0) = 0$,

$$\zeta(u) = \int_0^1 D\zeta(tu)u \, dt,$$

and thus, if $\|u\|_0 \leq 1$,

$$\|\zeta(u)\|_n \leq C_n\left(\|u\|_{n+r_0} + \|u\|_0\|u\|_{n+r_0}\right) \leq 2C_n\|u\|_{n+r_0}.$$

□

2.3 Injectivity

2.3.1. Proposition. *If ζ satisfies the assumption of theorem 2.1.8 on page 47 and of the previous section, then there exist $\varepsilon > 0$ and some $C > 0$ such that*

$$\forall\, u, v \in B_{r_0}(\varepsilon), \quad \|u - v\|_0 \leq C\|\zeta(u) - \zeta(v)\|_0.$$

Proof. Let $u, v \in B_{r_0}(\varepsilon)$ for some $\varepsilon \in (0,1)$ that will be fixed later on. Then, according to Taylor's formula,

$$\zeta(v) = \zeta(u) + D\zeta(u)(v-u) + \int_0^1 D^2\zeta((1-t)u + tv)(v-u)^2(1-t)\, dt.$$

This implies

$$v - u = [D\zeta(u)]^{-1}\left\{\zeta(v) - \zeta(u) - \int_0^1 D^2\zeta((1-t)u + tv)(v-u)^2(1-t)\, dt\right\}.$$

Using (2.2.3.d), since $\|(1-t)u+tv\|_{r_0} \leq \varepsilon$ we get

$$\|v-u\|_0 \;\leq\; C_0(1+\varepsilon)\left(\|\zeta(u)-\zeta(v)\|_0 + \int_0^1 \|\mathrm{D}^2\zeta((1-t)u+tv)(v-u)^2\|_0 \,\mathrm{d}t\right). \tag{2.3.1.a}$$

According to (2.2.3.c),

$$\|\mathrm{D}^2\zeta((1-t)u+tv)(v-u)^2\|_0 \\ \leq\; C_{r_0}\left(2\|v-u\|_0\|v-u\|_{r_0} + \|v-u\|_0^2\|(1-t)u+tv\|_{r_0}\right),$$

which, since $u, v \in B_{r_0}(\varepsilon)$, implies

$$\|\mathrm{D}^2\zeta((1-t)u+tv)(v-u)^2\|_0 \;\leq\; C\varepsilon\|v-u\|_0. \tag{2.3.1.b}$$

Thus, (2.3.1.a) becomes

$$\|v-u\|_0 \;\leq\; C\left(\|\zeta(u)-\zeta(v)\|_0 + \varepsilon\|v-u\|_0\right),$$

and the result follows as soon as $C\varepsilon < 1/2$. Then, (2.3.1.b) becomes

$$\|\mathrm{D}^2\zeta((1-t)u+tv)(v-u)^2\|_0 \;\leq\; \|v-u\|_0.$$

□

Let us put the last inequality into

2.3.2. Lemma. *If ε is given by proposition 2.3.1 on the preceding page, then*

$$\forall\, u, v \in B_{r_0}(\varepsilon), \quad \|\mathrm{D}^2\zeta((1-t)u+tv)(v-u)^2\|_0 \;\leq\; \|v-u\|_0.$$

2.4 Smoothing operators

2.4.1. In this section, we introduce the operators $(S_t)_{t\geq 0}$ that will enable us to prove ζ is surjective in the next section. In particular, we will study the solutions of the equation

$$x'(t) + \lambda S_t x(t) = y(t),$$

and give estimates on $\|x(t)\|_n$.

2.4.2. Smoothing operator. Let $\sigma : \mathbb{R} \to [0, +\infty)$ be a smooth function such that $\sigma(t) = 0$ when $t \leq 0$ and $\sigma(t) = 1$ when $t \geq 1$, with σ strictly increasing on $(0, 1)$. The smoothing operator $S_t : \Sigma(\mathrm{E}) \to \Sigma(\mathrm{E})$ is defined by:

$$(S_t u)_k \;:=\; \sigma(t-k)u_k \quad \text{for all } k \in \mathbb{N} \text{ and } u \in \Sigma(\mathrm{E}).$$

2.4.3. Lemma. *Let $n, q \in \mathbb{N}$. Then, for any $u \in \Sigma(\mathrm{E})$,*

$$\forall\, t \in \mathbb{R}, \quad \|S_t u\|_{n+q} \leq e^{qt}\|u\|_n \quad \textit{and} \quad \|u - S_t u\|_n \leq C_q e^{-qt}\|u\|_{n+q}.$$

Proof. Since $\sigma \leq 1$ always, and $\sigma(t-k) = 0$ as soon as $t \leq k$,

$$\|S_t u\|_{n+q} \;\leq\; \sum_{k \leq t} e^{(n+q)k}\|u_k\| \;\leq\; e^{qt}\sum_{k \leq t} e^{nk}\|u_k\| \;\leq\; e^{qt}\|u\|_n.$$

On the other hand, since $\sigma(t-k) = 1$ as soon as $t - 1 \geq k$,

$$\|u - S_t u\|_n \leq \sum_{t-1 \leq k} e^{nk}\|u_k\| \leq e^{-q(t-1)} \sum_{t-1 \leq k} e^{(n+q)k}\|u_k\| \leq e^q e^{-qt}\|u\|_{n+q}. \qquad \square$$

2.4.4. Lemma. *Let $T > 0$. Then for $t \leq T$, the smoothing operator S_t takes its values into a finite-dimensional subspace $\Sigma_T(\mathrm{E}) := \mathrm{Span}\{e_i \mid i \leq T\}$, where we have set $e_i := (\delta_{k,i})_{k \in \mathbb{N}}$.*

2.4.5. Lemma (Landau–Kolmogorov inequalities). *Let $p, q \in \mathbb{N}$. Then, for any $\theta \in (0, 1)$, if $(1-\theta)p + \theta q \in \mathbb{N}$,*

$$\forall\, u \in \Sigma(\mathrm{E}), \quad \|u\|_{(1-\theta)p + \theta q} \leq C_{n,p,q}\|u\|_p^{1-\theta}\|u\|_q^{\theta}.$$

The name is usually used for such equalities in $\mathscr{C}^\infty$ with $\|f\|_n := \|f\|_{\mathscr{C}^n}$ or $\|f\|_n := \|f\|_{\mathrm{H}^n}$.

Proof. Let $n = (1-\theta)p + \theta q$, and assume $p \leq q$. According to lemma 2.4.3 on the previous page,

$$\|u\|_n \;\leq\; \|S_t u\|_n + \|u - S_t u\|_n \;\leq\; C\left(e^{t(n-p)}\|u\|_p + e^{-t(q-n)}\|u\|_q\right).$$

Then, if t is such that $e^{t(n-p)}\|u\|_p = e^{-t(q-n)}\|u\|_q$, i.e. $e^{t(q-p)} = \|u\|_q/\|u\|_p$, since $n - p = \theta(q-p)$ and $q - n = (1-\theta)(q-p)$, we get the desired result. □

2.4.6. Lemma. *Let $\lambda > 0$, $x \in \mathscr{C}^1([0,T];\Sigma(\mathrm{E}))$, and $y \in \mathscr{C}^0([0,T];\Sigma(\mathrm{E}))$ be such that*

$$\forall\, t \in [0,T], \quad x'(t) + \lambda S_t x(t) = y(t),$$

where S_t is the operator introduced in definition 2.4.2 on the preceding page. Then, for any $n, q \in \mathbb{N}$ and assuming $q \in (0,\lambda)$,

$$\int_0^T e^{qt}\|x(t)\|_n \,\mathrm{d}t \leq C_{q,\lambda}\left(\|x(0)\|_{n+q} + \int_0^T \left\{\|y(t)\|_{n+q} + e^{qt}\|y(t)\|_n\right\}\mathrm{d}t\right),$$

where the constant $C_{q,\lambda}$ does not depend on T.

Proof. We set $a_k(s,t) := \exp\left(-\lambda\int_s^t \sigma(\omega - k)\,\mathrm{d}\omega\right)$ for $s < t$. Then, since

$$a_k(t,t) = 1 \qquad \text{and} \qquad \frac{\mathrm{d}}{\mathrm{d}s}a(s,t) = \lambda S_s a(s,t),$$

the equation $x'(t) + \lambda S_t x(t) = y(t)$ yields

$$\begin{aligned}
x_k(t) &= a_k(t,t)x_k(t) \\
&= a_k(0,t)x_k(0) + \int_0^t \frac{\mathrm{d}}{\mathrm{d}s}\left[a_k(s,t)x_k(s)\right]\mathrm{d}s \\
&= a_k(0,t)x_k(0) + \int_0^t \lambda S_s a_k(s,t)x_k(s) + a_k(s,t)x_k'(s)\,\mathrm{d}s \\
&= a_k(0,t)x_k(0) + \int_0^t a_k(s,t)y_k(s)\,\mathrm{d}s,
\end{aligned}$$

and therefore

$$\int_0^T e^{qt}\|x(t)\|_n \,\mathrm{d}t \leq \int_0^T e^{qt}\|a(0,t)x(0)\|_n \,\mathrm{d}t + \int_0^T\int_0^t e^{qt}\|a(s,t)y(s)\|_n \,\mathrm{d}s\,\mathrm{d}t. \tag{2.4.6.a}$$

First of all, notice that $a_k(s,t) \leq 1$. But as $\sigma(\omega - k) = 1$ when $\omega \geq k+1$,

- if $s \leq k+1 \leq t$, then $\int_s^t \sigma(\omega - k)\,\mathrm{d}\omega \geq (t-k-1)$ and $a_k(s,t) \leq e^{-\lambda(t-k-1)}$
- if $k+1 \leq s$, then $a_k(s,t) \leq e^{-\lambda(t-s)}$.

Thus, assuming $T > k+1$, and $\lambda > q > 0$,

$$\begin{aligned}
\int_0^T e^{qt} a_k(0,t)\,\mathrm{d}t &\leq \int_0^{k+1} e^{qt}\,\mathrm{d}t + \int_{k+1}^T e^{qt}e^{-\lambda(t-k-1)}\,\mathrm{d}t \\
&\leq C_{q,\lambda}\left\{e^{q(k+1)} + e^{\lambda(k+1)}e^{-(\lambda-q)(k+1)}\right\} \\
&\leq C_{q,\lambda}e^{qk}.
\end{aligned}$$

Therefore,

$$\int_0^T e^{qt}\|a(0,t)x(0)\|_n \,\mathrm{d}t = \int_0^T e^{qt}\sum_{k=0}^{\infty} e^{kn}a_k(0,t)\|x_k(0)\| \leq C_{q,\lambda}\|x(0)\|_{n+q}.$$

All we have to do now is to bound the second part in (2.4.6.a). By Fubini's theorem,

$$\int_0^T\int_0^t e^{qt}\|a(s,t)y(s)\|_n \,\mathrm{d}s\,\mathrm{d}t = \int_0^T\int_s^T e^{qt}\|a(s,t)y(s)\|_n \,\mathrm{d}t\,\mathrm{d}s. \tag{2.4.6.b}$$

Let us fix $s \in [0,T]$ and $k \in \mathbb{N}$. If $s \leq k+1$, then

$$\begin{aligned}
\int_s^T e^{tq}a_k(s,t)\,\mathrm{d}t &\leq \int_s^{k+1} e^{tq}a_k(s,t)\,\mathrm{d}t + \int_{k+1}^{+\infty} e^{tq}a_k(s,t)\,\mathrm{d}t \\
&\leq \int_s^{k+1} e^{tq}\,\mathrm{d}t + \int_{k+1}^{+\infty} e^{tq}e^{-\lambda(t-k-1)}\,\mathrm{d}t \\
&\leq C_q e^{qk} + C_{q,\lambda}e^{qk},
\end{aligned}$$

and if $k+1 \leq s$,

$$\int_s^T e^{tq} a_k(s,t)\,\mathrm{d}t \;\leq\; \int_s^T e^{tq} e^{-\lambda(t-s)}\,\mathrm{d}t \;\leq\; C_{q,\lambda} e^{qs}$$

We can sum up the situation with the following bound,

$$\int_s^T e^{tq} a_k(s,t)\,\mathrm{d}t \;\leq\; C_{q,\lambda}\left\{e^{qk} + e^{qs}\right\}.$$

Thus,

$$\begin{aligned}\int_s^T e^{qt}\|a(s,t)y(s)\|_n\,\mathrm{d}t \;&=\; \int_s^T e^{qt}\sum_{k=0}^{\infty} e^{kn} a_k(s,t)\|y_k(s)\|\,\mathrm{d}t \\ &=\; \sum_{k=0}^{\infty} e^{kn}\|y_k(s)\| \int_s^T e^{qt} a_k(s,t)\,\mathrm{d}t \\ &\leq\; C_{q,\lambda}\left\{\|y(s)\|_{n+q} + e^{qs}\|y(s)\|_n\right\},\end{aligned}$$

and this, injected into (2.4.6.b), completes the proof. □

2.5 Surjectivity

2.5.1. To show that, for every $\bar{v}$ close to 0, there is $\bar{u}$, also close to 0, such that $\zeta(\bar{u}) = \bar{v}$, we will solve the following ODE:

$$u'(t) = \lambda[\mathrm{D}\zeta(S_t u(t))]^{-1} S_t(\bar{v} - \zeta(u(t))),$$

and show that the solution $u(t)$ is defined on $[0, +\infty)$ and converges to some $\bar{u}$ when t tends to infinity, with $\zeta(\bar{u}) = \bar{v}$. The convergence will be proved thanks to a series of estimates involving $u(t)$, $x(t) = \bar{v} - \zeta(u(t))$, and $y(t) = [D\zeta(S_t u(t)) - D\zeta(u(t))]u'(t)$. It will be shown that

$$x'(t) + \lambda S_t x(t) = y(t)$$

and this second ODE will provide useful estimates, thanks to lemma 2.4.6 on page 54.

2.5.2. Lemma. *Let us fix $\lambda \in \mathbb{R}$. Then, possibly decreasing ε (initially given by proposition 2.3.1 on page 51) and increasing r_0 (from lemma 2.2.3 on page 49), for any $\bar{v} \in B_{r_0}(\varepsilon)$, for some $\delta > 0$ there is a unique $u \in \mathscr{C}^1([0,\delta); \overline{B}_{r_0}(\varepsilon))$ such that*

$$u(0) = 0 \quad \text{and} \quad u'(t) = \lambda\,[\mathrm{D}\zeta(S_t u(t))]^{-1}\,\{S_t\,(\bar{v} - \zeta(u(t)))\} \quad \text{for } t \in [0,\delta).$$

Proof. We divide the proof in four steps.

1. If $\Phi_t(u) := (S_t(u), S_t(\bar{v} - \zeta(u))$ and $\Psi(v)h := \lambda[\mathrm{D}\zeta(v)]^{-1}h$, then $u(t)$ is a solution if and only if $u'(t) = \Psi \circ \Phi_t(u(t))$ and $u(0) = 0$.

2. According to lemma 2.4.4 on page 53, for any $t \leq T$ the smoothing operator $S_t : \Sigma(\mathrm{E}) \to \Sigma(\mathrm{E})$ takes its values into a finite-dimensional subspace $\Sigma_T(\mathrm{E})$ where all the seminorms are equivalent norms. Since $\mathrm{D}_u\Phi : [0,T] \times \Omega \times \Sigma(\mathrm{E}) \to \Sigma_T(\mathrm{E}) \times \Sigma_T(\mathrm{E})$ is a smooth tame map, increasing r_0 and decreasing ε if necessary, for any $t \in [0,T]$,

$$\forall\, u \in B_{r_0}(2\varepsilon),\ \forall\, h \in B_{r_0}(2\rho), \quad \|\mathrm{D}_u\Phi_t(u)h\|_0 \leq C\left(1 + \|h\|_{r_0} + \|u\|_{r_0}\right).$$

Then, as for any $h \in \Sigma(\mathrm{E})$, we always have $\rho h/\|h\|_{r_0} \in B_{r_0}(2\rho)$, we can, more generally, say that

$$\forall\, u \in B_{r_0}(2\varepsilon),\ \forall\, h \in \Sigma(\mathrm{E}), \quad \|\mathrm{D}_u\Phi_t(u)h\|_0 \leq C\|h\|_{r_0},$$

and therefore,

$$\forall\, u, v \in \overline{B}_r(\varepsilon), \quad \|\Phi_t(u) - \Phi_t(v)\|_0 \leq C\|u - v\|_{r_0}.$$

Notice that, if $\Phi_t(u) = (v,h)$, then $\|v\|_0 = \|S_t(u)\|_0 \leq \|u\|_0 \leq \varepsilon$, and

$$\|h\|_0 \;\leq\; \|\bar{v} - \zeta(u)\|_0 \;\leq\; \|\bar{v}\|_0 + C\|u\|_{r_0} \;\leq\; C_0\varepsilon.$$

Thus, maybe decreasing ε again, as $\Psi : (\Sigma_T(\mathrm{E}) \cap B_0(2\varepsilon)) \times (\Sigma_T(\mathrm{E}) \cap B_0(2C_0\varepsilon) \to \Sigma(\mathrm{E})$ is also smooth tame and all the seminorms are equivalent on $\Sigma_T(\mathrm{E})$, we could in the same way show

$$\forall\, v, w, \in \Sigma_T(\mathrm{E}) \cap \overline{B}_0(\varepsilon),\ \forall\, h, k \in \Sigma_T(\mathrm{E}) \cap \overline{B}_0(C_0\varepsilon),$$

$$\|\Psi(v)h - \Psi(w)k\|_{r_0} \;\leq\; C\left(\|v-w\|_0 + \|h-k\|_0\right).$$

3. If for some $\delta > 0$ we set $\mathrm{X} = (\Sigma_T(\mathrm{E}) \cap \overline{B}_0(\varepsilon)) \times (\Sigma_T(\mathrm{E}) \cap \overline{B}_0(C_0\varepsilon)$ and

$$\Xi(v,h)(t) := \Phi_t\left(\int_0^t \Psi(v(s))h(s)\,\mathrm{d}s\right) \qquad \text{for } v,h \in \mathscr{C}^0([0,\delta];\mathrm{X}),$$

then Ξ is well-defined and is a contraction, at least for δ small enough, since

$$\|\Xi(v,h) - \Xi(w,k)\|_\infty \leq C\delta(\|v-w\|_\infty + \|h-k\|_\infty).$$

Moreover, if δ is small enough, then $\mathscr{C}^0([0,\delta];\mathrm{X})$ is stable, as $\|v(s)\|_{r_0} \leq C_T\|v(s)\|_0$ and thus

$$\left\|\int_0^t \Psi(v(s))h(s)\,\mathrm{d}s\right\|_0 \;\leq\; C\int_0^\delta \left(\|h(s)\|_0 + \|h(s)\|_0\|v(s)\|_{r_0}\right)\mathrm{d}s \;\leq\; C\delta\varepsilon,$$

and we can take δ such that $C\delta \leq 1$.

4. By the fixed-point theorem, there is a unique $(v,h) \in \mathscr{C}^0([0,\delta];\mathrm{X})$ such that $\Xi(v,h) = (v,h)$. Then the curve

$$u(t) := \int_0^t \Psi(v(s))h(s)\,\mathrm{d}s$$

is such that $(v(t),h(t)) = \Phi_t(u(t))$, so $u(t) = \int_0^t \Psi \circ \Phi_s(u(s))\,\mathrm{d}s$. This proves the existence of a solution, at least on some interval $[0,\delta]$. Moreover, δ has also been chosen so as to ensure $u(t) \in \overline{B}_{r_0}(\varepsilon)$.

□

2.5.3. Lemma. *Let $u \in \mathscr{C}^1([0,T), \overline{B}_{r_0}(\varepsilon))$ be the curve given by lemma 2.5.2 on page 56, defined on some interval $[0,T)$. Then, if $x(t) = \bar{v} - \zeta(u(t))$,*

$$\begin{aligned}&\forall\, n \geq 0,\ \forall\, q \in \mathbb{N},\ \exists\, C_{n,q} > 0,\ \forall\, t \in [0,T),\\ &\qquad \|u'(t)\|_{n+q} \;\leq\; C_{n,q,\lambda}e^{qt}\left(\|x(t)\|_n + \|u(t)\|_{n+r_0}\|x(t)\|_0\right).\end{aligned}$$

Since $\|u(t)\|_{r_0} \leq \varepsilon$*, this implies*

$$\|u'(t)\|_q \leq C_q e^{qt} \|x(t)\|_0.$$

There is no particular condition on q*, and* $C_{n,q,\lambda}$ *does not depend on* T*.*

Proof. Let $x(t) = \bar{v} - \zeta(u(t)$. Then $u'(t) = \lambda \, [\mathrm{D}\zeta(S_t u(t))]^{-1} S_t x(t)$. According to (2.2.3.d), since $\varepsilon < 1$, for any $n \geq 0$, we have

$$\begin{aligned} \|u'(t)\|_q &= \lambda \|[\mathrm{D}\zeta(S_t u(t))]^{-1} S_t x(t)\|_{n+q} \\ &\leq \lambda C_{n+q} \left(\|S_t x(t)\|_{n+q} + \|S_t x(t)\|_0 \|S_t u(t)\|_{n+q+r_0} \right). \end{aligned}$$

Now, the result follows from lemma 2.4.3 on page 53. □

2.5.4. Lemma. *Let* $u \in \mathscr{C}^1([0,T), \overline{B}_{r_0}(\varepsilon))$ *be the curve given by lemma 2.5.2 on page 56, defined on some interval* $[0,T)$*. Then, if*

$$x(t) := \bar{v} - \zeta(u(t)) \qquad \text{and} \qquad y(t) := [\mathrm{D}\zeta(S_t u(t)) - \mathrm{D}\zeta(u(t))] \, u'(t),$$

we have

$$\forall q \geq 0, \ \exists C_q > 0, \ \forall t \in [0,T), \quad \|y(t)\|_q \leq C_q \|u(t)\|_{q+r_0} \|x(t)\|_0.$$

Once again, there is no condition on q*, and the constant* C_q *does not depend on* T*.*

Proof. Notice that:

$$[\mathrm{D}\zeta(v) - \mathrm{D}\zeta(u)] = \int_0^1 \mathrm{D}^2\zeta\left((1-\omega)u + \omega v\right)(v-u) \, \mathrm{d}\omega,$$

therefore

$$y(t) = \int_0^1 \mathrm{D}^2\zeta((1-\omega)u(t) + \omega S_t u(t))(S_t u(t) - u(t)) u'(t) \, \mathrm{d}\omega.$$

Using (2.2.3.c), we get that, for $q \geq 0$,

$$\|y(t)\|_q \leq C_q \left\{ \|S_t u(t) - u(t)\|_0 \|u'(t)\|_{q+r_0} \right.$$

$$+\|S_t u(t)-u(t)\|_{q+r_0}\|u'(t)\|_0 \\ +\|S_t u(t)-u(t)\|_0\|u'(t)\|_0\|u(t)\|_{q+r_0}\Big\}.$$

According to lemma 2.4.3 on page 53, $\|S_t u(t)-u(t)\|_0 \leq e^{-mt}\|u(t)\|_m$, and therefore

$$\|y(t)\|_q \;\leq\; C\Big\{e^{-(q+r_0)t}\|u(t)\|_{q+r_0}\|u'(t)\|_{q+r_0}+\|u(t)\|_{q+r_0}\|u'(t)\|_0 \\ +\|u(t)\|_0\|u'(t)\|_0\|u(t)\|_{q+r_0}\Big\}.$$

Now, lemma 2.5.3 on page 58 yields $\|u'(t)\|_m \leq C_m e^{mt}\|x(t)\|_0$, and thus, as $\|u(t)\|_{r_0} < \varepsilon$, we get:

$$\|y(t)\|_q \;\leq\; C\Big\{2\|u(t)\|_{q+r_0}\|x(t)\|_0+\varepsilon\|x(t)\|_0\|u(t)\|_{q+r_0}\Big\}. \qquad \square$$

2.5.5. Lemma. *Let $u \in \mathscr{C}^1([0,T),\overline{B}_{r_0}(\varepsilon))$ be the curve given by lemma 2.5.2 on page 56. Then, if $\lambda > r_0$ and if $\|\bar{v}\|_{r_0}$ is small enough,*

$$\exists C>0, \qquad \int_0^T e^{r_0 t}\|x(t)\|_0\,\mathrm{d}t \;\leq\; C\|\bar{v}\|_{r_0}$$

The constant does not depend on T.

Proof. According to lemma 2.4.6 on page 54, as long as $\lambda > r_0 > 0$,

$$\int_0^T e^{r_0 t}\|x(t)\|_0\,\mathrm{d}t \;\leq\; C_{r_0,\lambda}\left(\|x(0)\|_{r_0}+\int_0^T\Big\{\|y(t)\|_{r_0}+e^{r_0 t}\|y(t)\|_0\Big\}\,\mathrm{d}t\right)$$

Notice that $x(0)=\bar{v}$. Thanks to lemma 2.5.4 on the preceding page, we get

$$\int_0^T e^{r_0 t}\|x(t)\|_0\,\mathrm{d}t \\ \leq\; C_{r_0,\lambda}\left(\|\bar{v}\|_{r_0}+\int_0^T\Big\{\|u(t)\|_{2r_0}\|x(t)\|_0+e^{r_0 t}\|u(t)\|_{r_0}\|x(t)\|_0\Big\}\,\mathrm{d}t\right).$$

Since according to lemma 2.5.3 on page 58, $\|u'(t)\|_q \leq Ce^{qt}\|x(t)\|_0$, we have

$$\|u(t)\|_{r_0} \;\leq\; Ce^{r_0 t}\int_0^t e^{r_0 s}\|x(s)\|_0\,\mathrm{d}s, \qquad \|u(t)\|_{r_0} \;\leq\; C\int_0^t e^{r_0 s}\|x(s)\|_0\,\mathrm{d}s.$$

Thus,

$$\int_0^T e^{r_0 t}\|x(t)\|_0 \,\mathrm{d}t \;\le\; C\left\{\|\bar{v}\|_{r_0} + 2\left(\int_0^T e^{r_0 t}\|x(t)\|_0 \,\mathrm{d}t\right)^2\right\}.$$

If we set $\kappa := \int_0^T e^{r_0 t}\|x(t)\|_0 \,\mathrm{d}t$, then $\kappa \le C(\|\bar{v}\|_{r_0} + 2\kappa^2)$. Therefore, $\kappa - 2C\kappa^2 \le C\|\bar{v}\|_{r_0}$. But we always have $\kappa - 2C\kappa^2 \le 1/8C$ with equality only for $\kappa = 1/4C$, so if $C\|\bar{v}\|_{r_0} \le 1/8C$, we can ensure $\kappa \in [0, 1/4C]$, as κ depends continuously on T and $\kappa = 0$ for $T = 0$. But then

$$\kappa \;\le\; \frac{C\|\bar{v}\|_{r_0}}{1 - 2C\kappa} \;\le\; 2C\|\bar{v}\|_{r_0}. \qquad \square$$

2.5.6. Lemma. *Let $u \in \mathscr{C}^1([0,T), \overline{B}_{r_0}(\varepsilon))$ be the curve given by lemma 2.5.2 on page 56. We assume $\lambda > r_0 + 1$. Then, if $\|\bar{v}\|_{r_0}$ is small enough,*

$$\forall\, n \ge r_0,\ \forall\, q \in \mathbb{N},\ \exists\, C_{n,q} > 0,\ \forall\, t \in [0,T),\ \int_0^t \|u'(s)\|_{n+q} \,\mathrm{d}s \le C_{n,q} e^{qt}\|\bar{v}\|_n.$$

There is no condition on q, and $C_{n,q}$ does not depend on T.

Proof. We proceed by induction on n, starting from $n = r_0$.

According to lemma 2.5.3 on page 58, $\|u'(t)\|_{r_0+q} \le Ce^{(r_0+q)t}\|x(t)\|_0$; therefore, assuming $\|\bar{v}\|_{r_0}$ small enough and using lemma 2.5.5 on the preceding page,

$$\int_0^t \|u'(s)\|_{r_0+q} \,\mathrm{d}s \;\le\; Ce^{qt}\int_0^t e^{r_0 s}\|x(s)\|_0 \,\mathrm{d}s \;\le\; Ce^{qt}\|\bar{v}\|_{r_0}. \tag{2.5.6.a}$$

The case $n = r_0$ is thus proved.

Let us now proceed with the induction, and assume that, for some $n \ge r_0$, we have $\int_0^t \|u'(s)\|_{n+q} \,\mathrm{d}s \le Ce^{qt}\|\bar{v}\|_n$ for any $q \ge 0$. From lemma 2.5.3 on page 58, as $n - r_0 \ge 0$, we get

$$\int_0^t \|u'(s)\|_{n+1+q} \,\mathrm{d}s \;\le\; C\int_0^t \left\{e^{(r_0+q+1)s}\left(\|x(s)\|_{n-r_0} + \|\bar{v}\|_n\|x(s)\|_0\right)\right\} \mathrm{d}s. \tag{2.5.6.b}$$

Since $\lambda > r_0 + 1$, lemma 2.4.6 on page 54 yields

$$\int_0^t e^{(r_0+1)s}\|x(s)\|_{n-r_0}\,\mathrm{d}s \le C\left(\|\bar{v}\|_{n+1} + \int_0^t \left\{\|y(s)\|_{n+1} + e^{(r_0+1)s}\|y(s)\|_{n-r_0}\right\}\mathrm{d}s\right).$$

But, using lemma 2.5.4 on page 59, as $\|u(s)\|_{n+m} \le Ce^{ms}\|\bar{v}\|_n$, we get

$$\begin{aligned}\|y(s)\|_{n+1} &\le C\|u(s)\|_{n+1+r_0}\|x(s)\|_0 \le Ce^{(r_0+1)s}\|\bar{v}\|_n\|x(s)\|_0,\\ \|y(s)\|_{n-r_0} &\le C\|u(s)\|_n\|x(s)\|_0 \le C\|\bar{v}\|_n\|x(s)\|_0.\end{aligned}$$

Thus,

$$\int_0^t e^{(r_0+1)s}\|x(s)\|_{n-r_0}\,\mathrm{d}s \le C\left(\|\bar{v}\|_{n+1} + 2\|\bar{v}\|_n\int_0^t e^{(r_0+1)s}\|x(s)\|_0\,\mathrm{d}s\right),$$

and (2.5.6.b) becomes

$$\int_0^t \|u'(s)\|_{n+1+q}\,\mathrm{d}s \le Ce^{qt}\left(\|\bar{v}\|_{n+1} + 3\|\bar{v}\|_n\int_0^t e^{(r_0+1)s}\|x(s)\|_0\,\mathrm{d}s\right). \tag{2.5.6.c}$$

Using lemma 2.4.6 on page 54 once again, we obtain

$$\int_0^t e^{(r_0+1)s}\|x(s)\|_0\,\mathrm{d}s \le C_0\left(\|\bar{v}\|_{r_0+1} + \int_0^t \left\{\|y(s)\|_{r_0+1} + e^{(r_0+1)s}\|y(s)\|_0\right\}\mathrm{d}s\right).$$

From lemma 2.5.4 on page 59, as $\|u(s)\|_{r_0+m} \le Ce^{ms}\|\bar{v}\|_{r_0}$, it follows

$$\begin{aligned}\|y(s)\|_{r_0+1} &\le C_0\|u(s)\|_{2r_0+1}\|x(s)\|_0 \le C_0e^{(r_0+1)s}\|\bar{v}\|_{r_0}\|x(s)\|_0,\\ \|y(s)\|_{r_0} &\le C_0\|u(s)\|_{r_0}\|x(s)\|_0 \le C_0\|\bar{v}\|_{r_0}\|x(s)\|_0.\end{aligned}$$

This yields

$$\int_0^t e^{(r_0+1)s}\|x(s)\|_0\,\mathrm{d}s \le C_0\left(\|\bar{v}\|_{r_0+1} + \|\bar{v}\|_{r_0}\int_0^t e^{(r_0+1)s}\|x(s)\|_0\,\mathrm{d}s\right).$$

Thus, if $\|\bar{v}\|_{r_0}$ is small enough,

$$\int_0^t e^{(r_0+1)s}\|x(s)\|_0\,\mathrm{d}s \le C\|\bar{v}\|_{r_0+1},$$

and with this estimate, (2.5.6.c) becomes

$$\int_0^t \|u'(s)\|_{n+1+q}\,\mathrm{d}s \;\leq\; Ce^{qt}\left(\|\bar{v}\|_{n+1} + \|\bar{v}\|_n\|\bar{v}\|_{r_0+1}\right).$$

However, setting $\theta = 1/(n+1-r_0)$, as

$$r_0 + 1 \;=\; (1-\theta)r_0 + \theta(n+1) \qquad \text{and} \qquad n \;=\; \theta r_0 + (1-\theta)(n+1),$$

we infer from lemma 2.4.5 on page 53 that

$$\|\bar{v}\|_{r_0+1}\|\bar{v}\|_n \;\leq\; \|\bar{v}\|_{r_0}\|\bar{v}\|_{n+1} \;\leq\; C\|\bar{v}\|_{n+1}.$$ □

2.5.7. Proposition. *For $\bar{v}$ with $\|\bar{v}\|_{r_0}$ small enough, there is a unique $\bar{u} \in \Sigma(\mathrm{E})$ such that $\zeta(\bar{u}) = \bar{v}$, and*

$$\forall\, n \geq r_0, \quad \|\bar{u}\|_n \leq C_n\|\bar{v}\|_n.$$

Proof. Let $\bar{v} \in B_{r_0}(\delta)$ with δ small enough and $\lambda > r_0 + 1$. If $u \in \mathscr{C}^1([0,T), \overline{B}_{r_0}(\varepsilon))$ is given by lemma 2.5.2 on page 56, and is defined on a maximal interval $[0,T)$, then according to lemma 2.5.6 on page 61 $(u(t))_{t\in[0,T)}$ is Cauchy when $t \to T$, and thus converges to some u_T. From lemma 2.5.6 on page 61, it follows that

$$\forall\, n \geq r_0, \quad \|\bar{u}\|_n \leq C_n\|\bar{v}\|_n.$$

Thus, by taking $\|\bar{v}\|_{r_0}$ small enough, we can ensure $u_T \in \overline{B}_{r_0}(\varepsilon/2)$. But this implies $T = \infty$, since if it were not the case, by starting over from u_T, we could extend u in $\overline{B}_{r_0}(\varepsilon)$ beyond T, contradicting its maximality.

As $u'(t) = \lambda[\mathrm{D}\zeta(S_t u(t))]^{-1}\{S_t(\bar{v} - \zeta(u(t)))\}$, we get $u'(t)$ also converges when t tends to infinity. On the other hand, since the constants from lemma 2.5.6 on page 61 do not depend on T, $\int_0^\infty \|u'(s)\|_n\,\mathrm{d}s < \infty$, and therefore $u'(t) \to 0$ when $t \to \infty$. This, in turn, implies $\zeta(u(t)) \to \bar{v} = \zeta(u_\infty)$.

Uniqueness follows from proposition 2.3.1 on page 51. □

2.6 Smoothness and tame estimates

2.6.1. Lemma. *Let $\varepsilon > 0$ be the radius given by proposition 2.3.1 on page 51. Then, for all $u, v \in B_{r_0}(\varepsilon)$ and $n \in \mathbb{N}$,*

$$\|u - v\|_n \;\leq\; C_n \left[\|\zeta(u) - \zeta(v)\|_n + \left(\|u\|_{n+r_0} + \|v\|_{n+r_0} \right) \|\zeta(u) - \zeta(v)\|_0 \right].$$

Proof. As in the proof of proposition 2.3.1 on page 51, we start from Taylor's formula, which yields

$$v - u \;=\; [\mathrm{D}\zeta(u)]^{-1} \left\{ \zeta(v) - \zeta(u) - \int_0^1 \mathrm{D}^2\zeta((1-t)u + tv)(v-u)^2 (1-t)\,\mathrm{d}t \right\}.$$

and, using (2.2.3.d), we get

$$\begin{aligned} \|v - u\|_n \;\leq\; C_n \Bigg[& \|\zeta(u) - \zeta(v)\|_n + \int_0^1 \left\| \mathrm{D}^2\zeta((1-t)u + tv)(v-u)^2 \right\|_n \mathrm{d}t \\ & + \|u\|_{n+r_0} \left(\|\zeta(u) - \zeta(v)\|_0 + \int_0^1 \left\| \mathrm{D}^2\zeta((1-t)u + tv)(v-u)^2 \right\|_0 \mathrm{d}t \right) \Bigg]. \end{aligned} \tag{2.6.1.a}$$

On the one hand, from (2.2.3.c) it follows

$$\begin{aligned} & \left\| \mathrm{D}^2\zeta((1-t)u + tv)(v-u)^2 \right\|_n \\ & \qquad \leq\; C_n \left(2\|u - v\|_0 \|u - v\|_{n+r_0} + \|u - v\|_0^2 \|(1-t)u + tv\|_{n+r_0} \right) \\ & \qquad\qquad \leq\; C \left(\|u\|_{n+r_0} + \|v\|_{n+r_0} \right) \|u - v\|_0. \end{aligned}$$

Thanks to the bounds from proposition 2.3.1 on page 51, this yields

$$\left\| \mathrm{D}^2\zeta((1-t)u + tv)(v-u)^2 \right\|_n \;\leq\; C_n \left(\|u\|_{n+r_0} + \|v\|_{n+r_0} \right) \|\zeta(u) - \zeta(v)\|_0, \tag{2.6.1.b}$$

On the other hand, from lemma 2.3.2 on page 52 we get

$$\|\mathrm{D}^2\zeta((1-t)u + tv)(v-u)^2\|_0 \;\leq\; \|v - u\|_0 \;\leq\; C\|\zeta(v) - \zeta(u)\|_0.$$

Putting the last two inequalities into (2.6.1.a), we get the result. ◻

2.6.2. Lemma. *Let $\varepsilon > 0$ be the radius given by proposition 2.3.1 on page 51. Then, for $u, v \in B_{r_0}(\varepsilon)$, we have*

$$\forall\, n \geq 0, \quad \|v - u - [D\zeta(u)]^{-1}(\zeta(v) - \zeta(u))\|_n \leq C_n\left(\|u - v\|_{n+r_0}\|u - v\|_0 + \|u\|_{n+r_0}\|u - v\|_{r_0}^2\right).$$

Proof. Once again, we start from Taylor's formula

$$\zeta(v) = \zeta(u) + \mathrm{D}\zeta(u)(v - u) + \int_0^1 \mathrm{D}^2\zeta((1-t)u + tv)(v-u)^2(1-t)\,\mathrm{d}t,$$

which yields

$$v - u - [\mathrm{D}\zeta(u)]^{-1}(\zeta(v) - \zeta(u)) = -[\mathrm{D}\zeta(u)]^{-1}\left(\int_0^1 \mathrm{D}^2\zeta((1-t)u + tv)(v-u)^2(1-t)\,\mathrm{d}t\right),$$

Then, using (2.2.3.d), we get

$$\begin{aligned}\|v - u - [\mathrm{D}\zeta(u)]^{-1}(\zeta(v) - \zeta(u))\|_n &\leq C_n\left(\int_0^1 \|\mathrm{D}^2\zeta((1-t)u + tv)(v-u)^2\|_n(1-t)\,\mathrm{d}t\right.\\ &\quad\left. + \|u\|_{n+r_0}\int_0^1 \|\mathrm{D}^2\zeta((1-t)u + tv)(v-u)^2\|_0(1-t)\,\mathrm{d}t\right).\end{aligned}$$

Thanks to (2.2.3.c),

$$\begin{aligned}\|\mathrm{D}^2\zeta((1-t)u + tv)(v-u)^2\|_n &\leq C_n\left(2\|v-u\|_{n+r_0}\|v-u\|_0 + \|(1-t)u + tv\|_{n+r_0}\|v-u\|_0^2\right)\\ &\leq C_n\left((2 + t\|u-v\|_0)\|v-u\|_{n+r_0}\|v-u\|_0 + \|u\|_{n+r_0}\|v-u\|_0^2\right),\end{aligned}$$

and thus, since $\|u - v\|_{r_0} \leq 2\varepsilon$,

$$\|\mathrm{D}^2\zeta((1-t)u + tv)(v-u)^2\|_n \leq C\left(\|v-u\|_{n+r_0}\|v-u\|_0 + \|u\|_{n+r_0}\|v-u\|_0^2\right).$$

Now, (2.2.3.c) yields

$$\begin{aligned}\|D^2\zeta((1-t)u+tv)(v-u)^2\|_0 \\ \leq C_n\left(2\|v-u\|_0\|v-u\|_{r_0}+\|(1-t)u+tv\|_{r_0}\|v-u\|_0^2\right) \\ \leq C_n(2+\varepsilon)\|v-u\|_{r_0}^2.\end{aligned}$$ □

2.6.3. Proposition. *For any $\bar{v}$ let $\bar{u} = \zeta^{-1}(\bar{v})$ be the unique antecedent given by proposition 2.5.7 on page 63. Then ζ^{-1} is smooth, and all its derivatives satisfy a tame estimate.*

Proof. From proposition 2.5.7 on page 63, we already know there is an estimate

$$\forall n \geq 2r_0, \quad \|\zeta^{-1}(u)\|_n \leq C_n\|u\|_n.$$

Then, lemma 2.6.1 on page 64 also shows that ζ^{-1} is continuous. Furthermore, lemma 2.6.2 on page 64 yields

$$\|\zeta^{-1}(u+h)-\zeta^{-1}(u)-[D\zeta(\zeta^{-1}(u))]^{-1}h\|_n \leq C_n\left(\|h\|_{n+r_0}\|h\|_0+\|u\|_{n+r_0}\|h\|_{r_0}^2\right),$$

which proves that ζ^{-1} is Gâteaux-differentiable, and

$$D\zeta^{-1}(u)h = \left[D\zeta(\zeta^{-1}(u))\right]^{-1}h.$$

Thus ζ^{-1} is smooth, and the tames estimates for its derivatives follow from the tames estimates of $[D\zeta]^{-1}$ and ζ^{-1}. □

2.7 Tameness of some usual spaces

2.7.1. This chapter would not be complete if we did not show that $\mathscr{C}^\infty(\mathbb{T}^d)$ is tame. This is done in three steps: first, we introduce a space F and prove it is tame; then, thanks to the Fourier transform, which sends the space $\mathscr{C}^\infty(B)$ of smooth functions over the closed unit ball B into F, we show $\mathscr{C}^\infty(B)$ is also tame; at last, thanks to Nash's embedding theorem, we conclude that for any compact riemannian manifold M, the space $\mathscr{C}^\infty(M)$ is also tame.

2.7.2. Lemma. *Let* (X, μ) *be a measure space, and let* $w : \mathrm{X} \to \mathbb{R}$ *be a positive weight function. For any map* $f \in \mathrm{L}^1(\mu)$*, we set* $\|f\|_n := \int e^{nw(x)}|f(x)|\,\mathrm{d}\mu(x)$*. Then,*

$$\mathrm{F} := \left\{ f \in \mathrm{L}^1(\mu) \,\middle|\, \forall\, n \in \mathbb{N}, \|f\|_n < \infty \right\} \qquad \textit{is a tame space.}$$

Proof. For any $k \in \mathbb{N}$, let

$$\mathrm{X}_k \;:=\; \{\, x \in \mathrm{X} \mid k \leq w(x) < k+1 \,\},$$

and define

$$L : \begin{cases} \mathrm{F} & \to \quad \Sigma(\mathrm{L}^1(\mu)) \\ f & \mapsto \quad (\mathbb{1}_{\mathrm{X}_k} f)_{k\in\mathbb{N}} \end{cases} \qquad \text{and} \qquad K : \begin{cases} \Sigma(\mathrm{L}^1(\mu)) & \to \quad \mathrm{F} \\ (f_k)_{k\in\mathbb{N}} & \mapsto \quad \sum f_k \mathbb{1}_{\mathrm{X}_k} \end{cases}.$$

Then, for any $n \geq 0$, as $w(x) \geq k$ for $x \in \mathrm{X}_k$,

$$\|\mathbb{1}_{\mathrm{X}_k} f\|_{\mathrm{L}^1} \;\leq\; e^{-kn} \int_{\mathrm{X}\, e^{nw(x)}} \mathbb{1}_{\mathrm{X}_k}(x)|f(x)|\,\mathrm{d}\mu(x).$$

This implies

$$\|Lf\|_n \;=\; \sum_{k=0}^{\infty} e^{kn} \|\mathbb{1}_{\mathrm{X}_k} f\|_{\mathrm{L}^1} \;\leq\; \int_{\mathrm{X}} e^{nw(x)}|f(x)|\,\mathrm{d}\mu(x) \;\leq\; \|f\|_n,$$

so L is a tame linear map. Conversely,

$$\begin{aligned} \left\| K((f_k)_{k\in\mathbb{N}}) \right\|_n \;&=\; \int_{\mathrm{X}} e^{nw(x)} \sum_{k=0}^{\infty} \mathbb{1}_{\mathrm{X}_k}(x)|f_k(x)|\,\mathrm{d}\mu(x) \\ &\leq\; \sum_{k=0}^{\infty} \int_{\mathrm{X}_k} e^{n(k+1)}|f_k(x)|\,\mathrm{d}\mu(x) \\ &\leq\; e^n \sum_{k=0} e^{kn} \|f_k\|_{\mathrm{L}^1} \\ &\leq\; C_n \|(f_k)_{k\in\mathbb{N}}\|_n, \end{aligned}$$

so K is also tame linear. Since $K \circ L = \mathrm{Id}_{\mathrm{F}}$, we conclude F is a tame space. ◻

2.7.3. Lemma. *Let B be the closed unit ball in $\mathbb{R}^d$. Then*

$$\mathscr{C}_0^\infty(B) \;:=\; \{f \in \mathscr{C}_c^\infty \mid \operatorname{supp} f \subset B\}$$

is a tame space, if endowed with

$$\|f\|_n \;:=\; \max_{|\alpha|\le n} \sup_{x\in B} \left|\mathrm{D}^\alpha f(x)\right|,$$

Proof. Let $w(x) = \frac{1}{2}\ln(1+|x|^2)$. Then, if $f \in \mathscr{C}_0^\infty(B)$ and $\hat{f}$ denotes the Fourier transform of f, we have

$$\int e^{nw(x)}|\hat{f}(x)|\,\mathrm{d}x \;=\; \int (1+|x|^2)^{n/2}|\hat{f}(x)|\,\mathrm{d}x.$$

Notice, however, that for any $\alpha \in \mathbb{N}^d$,

$$|x^\alpha||\hat{f}(x)| \;=\; \frac{1}{2\pi}\left|\int e^{-2i\pi\langle\xi|x\rangle}\mathrm{D}^\alpha f(\xi)\,\mathrm{d}\xi\right| \;\le\; \frac{|B|}{2\pi}\|f\|_{|\alpha|},$$

and therefore, for any $m \in \mathbb{N}$,

$$(1+|x|^2)^m|\hat{f}(x)| \;\le\; C_m\|f\|_{2m}.$$

Thus,

$$\begin{aligned}
\int e^{nw(x)}|\hat{f}(x)|\,\mathrm{d}x \;&\le\; \int \frac{(1+|x|^2)^{(d+1+n)/2}}{(1+|x|^2)^{(d+1)/2}}|\hat{f}(x)|\,\mathrm{d}x \\
&\le\; C_{\lceil(d+1+n)/2\rceil}\int \frac{\|f\|_{2\lceil(d+1+n)/2\rceil}}{(1+|x|^2)^{(d+1)/2}}\,\mathrm{d}x \\
&\le\; C_n\|f\|_{n+d+2}.
\end{aligned}$$

This proves that the Fourier transform $\mathcal{F} : \mathscr{C}_0^\infty(B) \to \mathrm{F}$ is tame, if F stands for the tame space introduced in lemma 2.7.2 on the previous page. Conversely, if $u \in \mathrm{F}$ and $\alpha \in \mathbb{N}^d$,

$$\left|\mathrm{D}^\alpha \hat{u}(\xi)\right| \;=\; \left|\int (-2i\pi x)^\alpha u(x) e^{-2i\pi\langle\xi|x\rangle}\,\mathrm{d}x\right|$$

$$
\begin{aligned}
&\leq C_\alpha \int (1+|x|^2)^{|\alpha|/2} |u(x)| \,\mathrm{d}x \\
&\leq C_\alpha \|u\|_{|\alpha|},
\end{aligned}
$$

and this proves $\|\mathcal{F}^{-1}u\|_n \leq C_n \|u\|_n$. As F is a tame space, $\mathscr{C}_0^\infty(B)$ is also tame. □

2.7.4. Proposition. *Let M be a compact Riemannian manifold. Then $\mathscr{C}^\infty(M)$ is a tame space.*

Proof. According to Nash's embedding theorem, M can be isometrically embedded into a bounded subset of $\mathbb{R}^d$, for some $d \in \mathbb{N}$. Thanks to Whitney's extension theorem, $\mathscr{C}^\infty(M)$ can therefore be tamely embedded into $\mathscr{C}_0^\infty(B)$, which is a tame space according to lemma 2.7.3 on page 67. □

Chapter 3

From Knothe's rearrangement to Brenier's optimal map

3.0.1. A few years ago, Guillaume Carlier, Alfred Galichon, and Filippo Santambrogio [19] proved the existence of a connection between the Knothe–Rosenblatt rearrangement (which will be defined in §3.1.3) and Yann Brenier's map, in the form of a differential equation—at least, when one of the two measures is discrete. In this chapter, I extend their result to the case of absolutely continuous measures. Most of what follows is taken from my article [10].

3.1 The Knothe–Rosenblatt rearrangement

3.1.1. As we have seen in section 1.2, if μ and ν are Borel probability measures on $\mathbb{R}$, with μ atomless, and F and G are their respective cumulative distributions, then $G^{-1} \circ F$ sends μ onto ν. In greater dimensions, the Knothe–Rosenblatt rearrangement is a mapping that intends to use this result to send a measure onto an other. To work with unidimensional measures, we first need to disintegrate them both.

3.1.2. Disintegration of a measure. Let $\mathrm{X} = \mathbb{R}$ or $\mathrm{X} = \mathbb{T}$. Any Borel measure μ on X^d can then be disintegrated according to the axes: there exists a family $\{\mu^1, \dots, \mu^d\}$, with $\mu^k : \mathrm{X}^{k-1} \to \mathcal{P}(\mathrm{X})$ Borel, such that, for all $\xi \in \mathscr{C}^0_\mathrm{b}(\mathrm{X}^d)$,

$$\int \xi(x)\,\mathrm{d}\mu(x) = \int \left(\int \cdots \left(\int \xi(x)\,\mathrm{d}\mu^d_{x_1,\dots,x_{d-1}}(x_d) \right) \cdots \mathrm{d}\mu^2_{x_1}(x_2) \right) \mathrm{d}\mu^1(x_1).$$

For the sake of clarity, let us now assume $d = 2$. If μ is absolutely continuous, and f stand for its density, then the disintegrated measures $\mu^1, \mu^2_{x_1}$ also have densities, namely:

$$f^1(x_1) := \int f(x_1, x_2)\,\mathrm{d}x_2 \qquad \text{and} \qquad f^2_{x_1}(x_2) := \frac{f(x_1, x_2)}{f^1(x_1)}.$$

3.1.3. The Knothe–Rosenblatt rearrangement. This transport map was defined in the 1950s, separately by Murray Rosenblatt [51] and by Herbert Knothe [38]. The former had in mind applications to probability theory and statistics; the later used it to study convex bodies and prove an improved isoperimetric inequality[1]—an idea later popularized by Mikhail Gromov [44].

In dimension two, the rearrangement is defined as follows: Let μ and ν be two absolutely continuous measures on X^2, with $\mathrm{X} = \mathbb{R}$ or $\mathrm{X} = \mathbb{T}$. Let $\{\mu^1, \mu^2\}$ and $\{\nu^1, \nu^2\}$ be their disintegrations, and let $F^1, F^2_{x_1}$ and $G^1, G^2_{y_1}$ be the cumulative distributions of $\mu^1, \mu^2_{x_1}$ and $\nu^1, \nu^2_{y_1}$. Then, we set

$$T^1_{\mathrm{K}}(x_1) := \left[G^1\right]^{-1} \circ F^1(x_1), \qquad T^2_{\mathrm{K}}(x_1, x_2) := \left[G^2_{T^1_{\mathrm{K}}(x_1)}\right]^{-1} \circ F^2_{x_1}(x_2),$$

and $T_{\mathrm{K}} := (T^1_{\mathrm{K}}, T^2_{\mathrm{K}})$. The same procedure can be applied in any dimension.

3.1.4. Lemma. *The rearrangement T_{K} thus defined maps μ onto ν.*

Proof. We give a proof only for $d = 2$. Let $\xi \in \mathscr{C}(\mathrm{X}^2)$. Then,

$$\begin{aligned}\int \xi(T_{\mathrm{K}}(x))\,\mathrm{d}\mu(x) &= \int \left(\int \xi\big(T^1_{\mathrm{K}}(x_1), T^2_{\mathrm{K}}(x_1, x_2)\big)\,\mathrm{d}\mu^2_{x_1}(x_2) \right) \mathrm{d}\mu^1(x_1) \\ &= \int \left(\int \xi(T^1_{\mathrm{K}}(x_1), y_2)\,\mathrm{d}\nu^2_{T^1_{\mathrm{K}}(x_1)}(y_2) \right) \mathrm{d}\mu^1(x_1),\end{aligned}$$

for $T^2_{\mathrm{K}}(x_1, x_2)$ sends $\mu^2_{x_1}$ onto $\nu^2_{T^1_{\mathrm{K}}(x_1)}$. Likewise, as T^1_{K} sends μ^1 onto ν^1, we get

$$\int \xi(T_{\mathrm{K}}(x))\,\mathrm{d}\mu(x) = \int \left(\int \xi(y_1, y_2)\,\mathrm{d}\nu^2_{y_1}(y_2) \right) \mathrm{d}\nu^1(y_1). \qquad \square$$

[1]Brenier's map turned out to be more suited to deal with the isoperimetric inequality than Knothe's rearrangement: Alessio Figalli, Francesco Maggi, and Aldo Pratelli [30, 55] were able to obtain a sharp inequality using Optimal Transport.

3.1.5. The starting point of our investigation is the proof, by Guillaume Carlier, Alfred Galichon, and Filippo Santambrogio [19], that this "rearrangement" is the limit of Brenier's map when the quadratic cost degenerates. We have seen in section 1.4 that, if μ and ν are probability measures on $\mathbb{T}^d$ with strictly positive densities, and

$$A_t := \begin{pmatrix} 1 & & & \\ & \lambda_t^1 & & \\ & & \ddots & \\ & & & \lambda_t^1 \cdots \lambda_t^{d-1} \end{pmatrix}$$

with $\lambda^k : \mathbb{R} \to [0, +\infty)$ such that $\lambda_t^k = 0$ only for $t = 0$, then, for any $t > 0$, there is a unique optimal transport map T_t between μ and ν for the quadratic cost c_t induced by A_t, i.e.

$$c_t(x, y) := \inf_{k \in \mathbb{Z}^d} \frac{1}{2} A_t (x - y - k)^2 = \sum_{k=1}^{d} \frac{\lambda_t^1 \cdots \lambda_t^{k-1}}{2} \mathrm{d}(x_k, y_k)^2,$$

with d the usual distance on $\mathbb{T}$.

3.1.6. Theorem (Carlier–Galichon–Santambrogio). *When t tends to zero, the map T_t converges in $\mathrm{L}^2(\mu)$ to the Knothe–Rosenblatt rearrangement.*

Proof. As the proof is much easier for $d = 2$, we give a proof for $d = 3$ to account for the additional difficulty in greater dimensions. We therefore work on the torus $\mathbb{T}^3$, and proceed in 7 steps.

1. Let $\gamma_t := (\mathrm{Id}, T_t)_\# \mu$ be the optimal transport plan for the quadratic cost c_t, and let $\gamma_K := (\mathrm{Id}, T_K)_\# \mu$ be the plan corresponding to the rearrangement. Up to a subsequence, γ_t converges narrowly to some $\gamma \in \Gamma(\mu, \nu)$. On the one hand, γ_t is optimal for c_t, so

$$\int \left(\mathrm{d}(x_1, y_1)^2 + \cdots + \prod_{k<d} \lambda_t^k \mathrm{d}(x_d, y_d)^2 \right) \mathrm{d}\gamma_t(x, y) \\ \leq \int \left(\mathrm{d}(x_1, y_1)^2 + \cdots + \prod_{k<d} \lambda_t^k \mathrm{d}(x_d, y_d)^2 \right) \mathrm{d}\gamma_K(x, y). \quad (3.1.6.a)$$

On the other hand,

$$\int \mathrm{d}(x_1, y_1)^2 \, \mathrm{d}\gamma(x, y) \;=\; \lim_{t\to 0} \int \mathrm{d}(x_1, y_1)^2 \, \mathrm{d}\gamma_t(x, y).$$

Therefore, taking the limit,

$$\int d_1(x_1, y_1)^2 \, \mathrm{d}\gamma(x, y) \;\le\; \int d_1(x_1, y_1)^2 \, \mathrm{d}\gamma_{\mathrm{K}}(x, y).$$

Thus, denoting by X_k and Y_k the projectors, we can say $\gamma^1 := (\mathsf{X}_1, \mathsf{Y}_1)_\# \gamma$ is optimal between the first marginals of μ and ν. Let μ^1 and ν^1 be those first marginals; then γ^1 is equal to $\gamma^1_{\mathrm{K}} := (\mathsf{X}_1, \mathsf{Y}_1)_\# \gamma_{\mathrm{K}} = (\mathrm{Id}, T^1_{\mathrm{K}})_\# \mu^1$.

2. Since inequality (3.1.6.a) and the optimality of $\gamma^1 = (\mathsf{X}_1, \mathsf{Y}_1)_\# \gamma_{\mathrm{K}}$ imply

$$\begin{aligned}\int \mathrm{d}(x_1, y_1)^2 \, \mathrm{d}\gamma_{\mathrm{K}}(x, y) \\ + \lambda_t \int \left(\mathrm{d}(x_2, y_2)^2 + \cdots + \prod_{1<k<d} \lambda_t^k \mathrm{d}(x_d, y_d)^2 \right) \mathrm{d}\gamma_t(x, y) \\ \le \int c_t(x, y) \, \mathrm{d}\gamma_{\mathrm{K}}(x, y),\end{aligned}$$

we also have

$$\int \mathrm{d}(x_2, y_2)^2 \, \mathrm{d}\gamma(x, y) \;\le\; \int \mathrm{d}(x_2, y_2)^2 \, \mathrm{d}\gamma_{\mathrm{K}}(x, y). \tag{3.1.6.b}$$

We now disintegrate $\gamma^{1,2} := ((\mathsf{X}_1, \mathsf{X}_2), (\mathsf{Y}_1, \mathsf{Y}_2))_\# \, \gamma$:

$$\int_{\mathbb{T}^2} \xi \, \mathrm{d}\gamma^{1,2} = \iint \xi(x, y) \, \mathrm{d}\gamma^2_{x_1, y_1}(x_2, y_2) \, \mathrm{d}\gamma^1(x_1, y_1).$$

Let us, for a moment, assume that for γ^1-almost all (x_1, y_1), the marginals of $\gamma^2_{x_1, y_1}$ are $\mu^2_{x_1}$ and $\nu^2_{y_1}$. Then, by the very definition of the rearrangement T_{K}, since $\gamma^1 = \gamma^1_{\mathrm{K}}$, for γ^1-almost every pair (x_1, y_1),

$$\int \mathrm{d}(x_2, y_2)^2 \, \mathrm{d}[\gamma_{\mathrm{K}}]^2_{x_1, y_1}(x_2, y_2) \;\le\; \int \mathrm{d}(x_2, y_2)^2 \, \mathrm{d}\gamma^2_{x_1, y_1}(x_2, y_2). \tag{3.1.6.c}$$

If we then integrate this with respect to $\gamma^1 = \gamma^1_{\mathrm{K}}$, we get

$$\int_{\mathbb{T}^3} \mathrm{d}(x_2, y_2)^2 \,\mathrm{d}\gamma_{\mathrm{K}}(x, y) \;\leq\; \int_{\mathbb{T}^2} \mathrm{d}(x_2, y_2)^2 \,\mathrm{d}\gamma^{1,2}(x, y).$$

But we have just seen the converse inequality, given by equation (3.1.6.b). This is only possible if, γ^1-almost everywhere, there is equality in (3.1.6.c). Therefore for γ^1-almost all (x_1, y_1), the measure $\gamma^2_{x_1,y_1}$ is also optimal. Thus, $\gamma^{1,2} = ((\mathsf{X}_1, \mathsf{X}_2), (\mathsf{Y}_1, \mathsf{Y}_2))_\# \gamma_{\mathrm{K}}$.

3. We must still prove the marginals of $\gamma^2_{x_1,y_1}$ are $\mu^2_{x_1}$ and $\nu^2_{y_1}$, at least almost everywhere for γ^1. Since the measure $\gamma^1 = \gamma^1_{\mathrm{K}}$ is concentrated on the graph $y_1 = T^1_{\mathrm{K}}(x_1)$, and μ^1 is absolutely continuous, all there is to check is that

$$[\mathsf{X}_2]_\# \gamma^2_{x_1, T^1_{\mathrm{K}}(x_1)} = \mu^2_{x_1} \qquad \text{and} \qquad [\mathsf{Y}_2]_\# \gamma^2_{x_1, T^1_{\mathrm{K}}(x_1)} = \nu^2_{T^1_{\mathrm{K}}(x_1)},$$

for almost every x_1. As ν^1 is absolutely continuous, T^1_{K} is a bijection; denoting by S^1 its inverse, the second equality can be replaced with

$$[\mathsf{Y}_2]_\# \gamma^2_{S^1(y_1), y_1} \;=\; \nu^2_{y_1}$$

which should stand for almost every y_1. By symmetry, we thus need to check only one of the two—for instance, that for almost every x_1, for any continuous function $\xi = \xi(x_2)$,

$$\int \xi(x_2) \,\mathrm{d}\gamma^2_{x_1, T^1_{\mathrm{K}}(x_1)}(x_2) \;=\; \int \xi(x_2) \,\mathrm{d}\mu^2_{x_1}(x_2).$$

Equivalently, we need only to show that for all $\eta = \eta(x_1)$ belonging to a proper *countable* subset of continuous functions, for all $\xi = \xi(x_2)$,

$$\iint \eta(x_1)\xi(x_2) \,\mathrm{d}\gamma^2_{x_1, T^1_{\mathrm{K}}(x_1)} \,\mathrm{d}\mu^1(x_1) \;=\; \iint \eta_1(x_1)\xi(x_2) \,\mathrm{d}\mu^2_{x_1} \,\mathrm{d}\mu^1(x_1).$$

It is now clear why the conclusion should holds, since

$$\iint \eta(x_1)\xi(x_2) \,\mathrm{d}\gamma^2_{x_1, T^1_{\mathrm{K}}(x_1)}(x_2) \,\mathrm{d}\mu^1(x_1) = \int \eta(x_1)\xi(x_2) \,\mathrm{d}\gamma(x, y).$$

4. We now proceed with the third component. Let $\gamma_t^{1,2}$ be an optimal transport plan between $(X_1, X_2)_\# \mu$ and $(Y_1, Y_2)_\# \nu$ for the cost

$$c_t^{1,2} = |x_1 - y_1|^2 + \lambda_t^1 |x_2 - y_2|^2.$$

Then, if p_{x_1,x_2,y_1,y_2} denotes an optimal plan between $\mu^3_{x_1,x_2}$ and $\nu^3_{y_1,y_2}$, we define a transport plan $\pi_t \in \Gamma(\mu, \nu)$ by setting

$$\int \xi(x,y)\,\mathrm{d}\pi_t(x,y) \;=\; \iint \xi(x,y)\,\mathrm{d}p_{x_1,x_2,y_1,y_2}(x_3,y_3)\,\mathrm{d}\gamma_t^{1,2}(x_1,x_2,y_1,y_2).$$

Now,

$$\int_{\mathbb{T}^2} c_t^{1,2}\,\mathrm{d}\gamma_t^{1,2} + \lambda_t^1\lambda_t^2 \int_{\mathbb{T}^3} |x_3 - y_3|^2\,\mathrm{d}\gamma_t(x,y) \;\le\; \int_{\mathbb{T}^3} c_t\,\mathrm{d}\gamma_t \;\le\; \int_{\mathbb{T}^3} c_t\,\mathrm{d}\pi_t,$$

and

$$\begin{aligned}\int_{\mathbb{T}^3} c_t\,\mathrm{d}\pi_t \;&=\; \int_{\mathbb{T}^2} c_t^{1,2}\,\mathrm{d}\gamma_t^{1,2} + \lambda_t^1\lambda_t^2 \iint |x_3 - y_3|^2\,\mathrm{d}p(x_3,y_3)\,\mathrm{d}\bar{\gamma}_t \\ &=\; \int_{\mathbb{T}^2} c_t^{1,2}\,\mathrm{d}\gamma_t^{1,2} + 2\lambda_t^1\lambda_t^2 \int_{\mathbb{T}^2} \mathrm{W}_2(\mu^3_{x_1,x_2}, \nu^3_{y_1,y_2})\,\mathrm{d}\bar{\gamma}_t(x_1,x_2,y_1,y_2).\end{aligned}$$

Thus,

$$\int_{\mathbb{T}^3} |x_3 - y_3|^2\,\mathrm{d}\gamma_t(x,y) \;\le\; 2\int_{\mathbb{T}^2} \mathrm{W}_2(\mu^3_{x_1,x_2}, \nu^3_{y_1,y_2})^2\,\mathrm{d}\gamma_t^{1,2}(x_1,x_2,y_1,y_2).$$

Let us, for an instant, assume

$$\int |x_3 - y_3|^2\,\mathrm{d}\gamma(x,y) \;\le\; 2\int \mathrm{W}_2(\mu^3_{x_1,x_2}, \nu^3_{y_1,y_2})^2\,\mathrm{d}\gamma^{1,2}(x,y). \tag{3.1.6.d}$$

We then disintegrate γ with respect to $\gamma^{1,2} = ((X_1, X_2), (Y_1, Y_2))_\#\,\gamma$, so that

$$\int \xi(x,y)\,\mathrm{d}\gamma(x,y) \;=\; \iint \xi(x,y)\,\mathrm{d}\gamma^3_{x_1,x_2,y_1,y_2}(x_3,y_3)\,\mathrm{d}\gamma^{1,2}(x_1,x_2,y_1,y_2).$$

Then, assuming $\gamma^3_{x_1,x_2,y_1,y_2} \in \Gamma(\mu^3_{x_1,x_2}, \nu^3_{y_1,y_2})$, for any x_1, x_2, y_1, y_2,

$$\mathrm{W}_2(\mu^3_{x_1,x_2}, \nu^3_{y_1,y_2})^2 \leq \frac{1}{2}\int |x_3 - y_3|^2 \,\mathrm{d}\gamma^3_{x_1,x_2,y_1,y_2}(x_3, y_3).$$

Thus, (3.1.6.d) implies there must be equality for $\gamma^{1,2}$-almost every x_1, x_2, y_1, y_2. This, in turn, means γ^3 is optimal almost everywhere, and thus $\gamma = \gamma_K$.

5. We therefore need to prove $\gamma^3_{x_1,x_2,y_1,y_2} \in \Gamma(\mu^3_{x_1,x_2}, \nu^3_{y_1,y_2})$. This is done as previously (see the 3rd step).

6. We must still prove (3.1.6.d). Let $\varepsilon > 0$. Since $(x_1, x_2) \mapsto \mu^3_{x_1,x_2}$ and $(y_1, y_2) \mapsto \nu^3_{y_1,y_2}$ are measurable, according to Lusin's theorem there is a compact K of $\mathbb{T}^1$ such that $\mu^1(K) > 1 - \varepsilon$ and $\nu^1(K) > 1 - \varepsilon$, and

$$\begin{cases} K \times K & \to & \mathcal{P}(\mathbb{T}^1) \\ (x_1, x_2) & \mapsto & \mu^3_{x_1,x_2} \end{cases} \quad \text{and} \quad \begin{cases} K \times K & \to & \mathcal{P}(\mathbb{T}^1) \\ (y_1, y_2) & \mapsto & \nu^3_{y_1,y_2} \end{cases} \quad \text{are continuous.}$$

We now extend those two maps into two continuous maps $\tilde{\mu}^3$ and $\tilde{\nu}^3$ on $\mathbb{T}^2$, such that $\tilde{\mu}^3 = \mu^3$ and $\tilde{\nu}^3 = \nu^3$ on $K \times K$. Then,

$$\int_{\mathbb{T}^2} \mathrm{W}_2(\tilde{\mu}^3, \tilde{\nu}^3)^2 \,\mathrm{d}\gamma^{1,2}_t \to \int_{\mathbb{T}^2} \mathrm{W}_2(\tilde{\mu}^3, \tilde{\nu}^3)^2 \,\mathrm{d}\gamma^{1,2}.$$

On the other hand, since W_2 is bounded on $\mathcal{P}(\mathbb{T}^1)$,

$$\left| \int_{\mathbb{T}^2} \mathrm{W}_2(\tilde{\mu}^3, \tilde{\nu}^3)^2 \,\mathrm{d}\gamma^{1,2}_t - \int_{\mathbb{T}^2} \mathrm{W}_2(\mu^3, \nu^3)^2 \,\mathrm{d}\gamma^{1,2}_t \right| \leq C\gamma_t(\mathbb{T}^2 \setminus K \times K)$$

and

$$\begin{aligned} \gamma_t(\mathbb{T}^2 \setminus K \times K) &\leq \gamma_t\left(\complement K \times \mathbb{T}^1\right) + \gamma_t\left(\mathbb{T}^1 \times \complement K\right) \\ &\leq \mu\left(\complement K\right) + \nu\left(\complement K\right) \\ &\leq 2\varepsilon. \end{aligned}$$

For the same reason,

$$\left| \int_{\mathbb{T}^2} \mathrm{W}_2(\tilde{\mu}^3, \tilde{\nu}^3)^2 \,\mathrm{d}\gamma^{1,2} - \int_{\mathbb{T}^2} \mathrm{W}_2(\mu^3, \nu^3)^2 \,\mathrm{d}\gamma^{1,2}_t \right| \leq 2C\varepsilon$$

as well. Thus,

$$\int_{\mathbb{T}^2} \mathrm{W}_2(\mu^3, \nu^3)^2 \, \mathrm{d}\gamma_t^{1,2} \to \int_{\mathbb{T}^2} \mathrm{W}_2(\mu^3, \nu^3)^2 \, \mathrm{d}\gamma^{1,2}.$$

7. At last,

$$\int \mathrm{d}(T_t(x), T_{\mathrm{K}}(x))^2 \, \mathrm{d}\mu(x) = \int \mathrm{d}(y, T_{\mathrm{K}}(x))^2 \, \mathrm{d}\gamma_t(x,y) \longrightarrow \int \mathrm{d}(y, T_{\mathrm{K}}(x))^2 \, \mathrm{d}\gamma_{\mathrm{K}}(x,y) = 0.$$

and this shows T_t converges to T_{K} in L^2. □

3.2 A PDE for positive times

3.2.1. We know from the theorem of Guillaume Carlier, Alfred Galichon, and Filippo Santambrogio (§3.1.6, on page 72) that there is a link between Knothe's rearrangement and Brenier's map for very degenerate costs. Before investigating this relationship any further, we will now examine the dependency of the optimal map on the quadratic cost.

According to proposition 1.4.5 on page 33, given two smooth, positive measures on $\mathbb{T}^d$, for any cost matrix $A \in \mathcal{S}_d^{++}$, there is a smooth Kantorovich potential $\Psi_A : \mathbb{T}^d \to \mathbb{R}$. What can we say of the regularity about $\Psi : A \mapsto \Psi_A$? Since the optimal map $x \mapsto x - A^{-1}\nabla\Psi_A(x)$ sends one measure onto the other, we know that a Monge–Ampère equation is satisfied: denoting by f and g the densities, we have

$$f(x) = g\left(x - A^{-1}\nabla\Psi_A(x)\right) \det\left(\mathrm{I}_2 - A^{-1}\nabla^2\Psi_A\right).$$

Thus, to get any regularity of Ψ_A with respect to A, the implicit function theorem seems a good idea. We therefore set

$$\mathcal{F}(A, u) := f - g\left(\mathrm{Id} - A^{-1}\nabla u\right) \det\left(\mathrm{I}_2 - A^{-1}\nabla^2 u\right),$$

and intend to show $\mathrm{D}_u\mathcal{F}(A, \Psi_A)$ is an isomorphism.

3.2.2. Zero-mean-value functional spaces. Since the potential is uniquely determined up to an additive constant, it seems more appropriate to work only with maps with zero mean values. Likewise, $\mathcal{F}$ obviously takes it values in a space of zero-mean-value maps. To be of zero mean value is thus a property we shall meet very often; there is hence a need for a specific notation. Given any functional space X, we will denote the space formed by the elements of X having a zero mean value with a $\diamond$ subscript—for instance, $\mathscr{C}^2_\diamond$ will be the space of all $u \in \mathscr{C}^2$ such that $\int u = 0$.

3.2.3. Lemma. *For any $A \in \mathcal{S}_d^{++}$, if $u \in \mathscr{C}^2_\diamond(\mathbb{T}^d)$ is such that $A - \nabla^2 u > 0$, then $\mathcal{F}(u, A) = 0$ if and only if u is the Kantorovich potential between μ and ν for the cost induced by A.*

Proof. This follows from proposition 1.4.5 on page 33 and the characterization given by lemma 1.4.4 on page 32. □

3.2.4. Lemma. *The operator $\mathcal{F}$ is smooth. For any $A \in \mathcal{S}_d^{++}$, if $u \in \mathscr{C}^2_\diamond(\mathbb{T}^d)$ is such that $A - \nabla^2 u > 0$, and $v \in \mathscr{C}^2_\diamond(\mathbb{T}^2)$, then*

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

We denote by M^* the transposed matrix of M, and by $\mathrm{Co}\, M$ its cofactor matrix—that is, the matrix formed by the cofactors (first minors).

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

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

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

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

Since $\nabla[\xi \circ T_A u] = [\nabla T_A u]^* \nabla\xi(T_A u)$,

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

and this yields

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

Therefore, as $\xi \circ T_A u$ is arbitrary, we get the first equality. Then, the second expression quickly follows, thanks to the formula $M^{-1} = [\mathrm{Co}\, M]^* / \det(M)$. □

3.2.5. Lemma. *Let $\varepsilon > 0$ and $A \in \mathcal{S}_d^{++}$. If $u \in \mathscr{C}^2_\diamond(\mathbb{T}^d)$ is such that*

$$A - \nabla^2 u \;>\; \varepsilon (\det A)^{1/d-1} \, \mathrm{I}_d,$$

then for any $q \in [\mathrm{H}^1_\diamond(\mathbb{T}^d)]^*$, there is a unique $v \in \mathrm{H}^1_\diamond(\mathbb{T}^d)$ such that

$$\mathrm{D}_u \mathcal{F}(A,u) v \;=\; q. \tag{3.2.5.a}$$

Moreover, $\|v\|_{\mathrm{H}^1} \leq C_\varepsilon \|q\|_{(\mathrm{H}^1_\diamond)^*}$, and the constant C_ε does not depend upon u.

Proof. Since $A - \nabla^2 u > \varepsilon(\det A)^{1/(d-1)}\, \mathrm{I}_d$, the lowest eigenvalue of $\mathrm{Co}\,(A - \nabla^2 u)$ is bounded by $\varepsilon^{d-1} \det A$. Since $g > \delta$ for some $\delta > 0$, for any $\xi \in \mathscr{C}^\infty(\mathbb{T}^d)$,

$$\begin{aligned} \varepsilon^{d-1} \det A \int \left|\nabla\xi\right|^2 &\leq \int \left\langle [\mathrm{Co}\,(A - \nabla^2 u)]^* \nabla\xi \,\middle|\, \nabla\xi \right\rangle \\ &\leq \frac{1}{\delta} \int g\left(\mathrm{Id} - A^{-1}\nabla u\right) \left\langle [\mathrm{Co}\,(A - \nabla^2 u)]^* \nabla\xi \,\middle|\, \nabla\xi \right\rangle, \end{aligned}$$

and thus

$$\int |\nabla\xi|^2 \;\leq\; -\frac{1}{\delta\varepsilon^{d-1}} \int \xi \mathrm{D}_u \mathcal{F}(A,u)\xi. \tag{3.2.5.b}$$

Therefore, thanks to the existence of a Poincaré inequality for $\mathrm{H}^1_\diamond(\mathbb{T}^d)$, the map $(\xi, \eta) \mapsto \int \eta \mathrm{D}_u \mathcal{F}(A,u)\xi$ induces a coercive, continuous bilinear form on $\mathrm{H}^1_\diamond$. We are thus entitled to apply the Lax–Milgram theorem, which yields the existence and the uniqueness, for every $q \in (\mathrm{H}^1_\diamond)^*$, of a $v \in \mathrm{H}^1_\diamond$ satisfying (3.2.5.a). Moreover, (3.2.5.b) immediately gives us $\|v\|_{\mathrm{H}^1} \leq \frac{1}{\delta\varepsilon^{d-1}} \|q\|_{(\mathrm{H}^1_\diamond)^*}$. □

3.2.6. Lemma. *With the same assumptions and notations as in lemma 3.2.5, for any $n \geq 1$, if $u \in \mathscr{C}^{n+2}_\diamond$ and $q \in \mathrm{H}^{n-1}_\diamond$ satisfy $\|u\|_{\mathscr{C}^3} + \|q\|_{(\mathrm{H}^1_\diamond)^*} \leq M$, then $v \in \mathrm{H}^{n+1}_\diamond$, and*

$$\|v\|_{\mathrm{H}^{n+1}} \leq C_{\varepsilon,M,n} \left\{ \|q\|_{\mathrm{H}^{n-1}} + \|u\|_{\mathscr{C}^{n+2}} \right\}. \tag{3.2.6.a}$$

Proof. We proceed by induction. Let $n \geq 1$, $u \in \mathscr{C}^{n+2}_\diamond$, and $q \in \mathrm{H}^n_\diamond$ such that

$$A - \nabla^2 u > \varepsilon (\det A)^{1/(d-1)} \, \mathrm{I}_d \qquad \text{and} \qquad \|u\|_{\mathscr{C}^3} + \|q\|_{(\mathrm{H}^1_\diamond)^*} \leq M.$$

Let us assume we already know the solution v is in $\mathrm{H}^n_\diamond$, and that

$$\|v\|_{\mathrm{H}^n} \leq C_{\varepsilon,M,n-1} \left\{ \|q\|_{\mathrm{H}^{n-2}} + \|u\|_{\mathscr{C}^{n+1}} \right\}. \tag{3.2.6.b}$$

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

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

First, we set $B_A u := (f - \mathcal{F}(A,u))[A - \nabla^2 u]^{-1}$, so that (3.2.5.a) now reads

$$\mathrm{D}_u \mathcal{F}(A,u) v \;=\; \operatorname{div}(B_A u \nabla v) \;=\; q. \tag{3.2.6.c}$$

Next, for $h \in \mathbb{R}^d$ and $\xi \in \mathrm{H}^1$, we define

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

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

Let $\nu \in \mathbb{N}^d$ be a d-index, with $|\nu| := \nu_1 + \cdots + \nu_d = n-1$, and assume $h \in \mathbb{R}^d$ is small enough. We can apply the operator δ_h to (3.2.6.c), and thus obtain

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

Then, by applying $\partial_\nu = \partial^{|\nu|}/\partial x_1^{\nu_1}\cdots\partial x_d^{\nu_d}$, we get

$$\operatorname{div}(B_A u \nabla \delta_h \partial_\nu v) = \delta_h \partial_\nu q - \sum_{0\le\alpha\le\nu} \binom{\nu}{\alpha} \operatorname{div}\left[(\delta_h \partial_{\nu-\alpha} B_A u)\, \nabla \tau_h \partial_\alpha v\right] - \sum_{0\le\alpha<\nu} \binom{\nu}{\alpha} \operatorname{div}\left[(\partial_{\nu-\alpha} B_A u)\, \nabla \delta_h \partial_\alpha v\right]. \quad (3.2.6.d)$$

According to lemma 3.2.5 on page 79, this implies

$$\begin{aligned}\|\delta_h\partial_\nu v\|_{\mathrm{H}^1} \;\le\; & C_\varepsilon \|\delta_h \partial_\nu q\|_{(\mathrm{H}^1_\diamond)^*} \\ & + C_\varepsilon \sum_{0\le\alpha\le\nu}\binom{\nu}{\alpha}\left\|\operatorname{div}\left[(\delta_h\partial_{\nu-\alpha}B_A u)\,\nabla\tau_h\partial_\alpha v\right]\right\|_{(\mathrm{H}^1_\diamond)^*} \\ & + C_\varepsilon \sum_{0\le\alpha<\nu}\binom{\nu}{\alpha}\left\|\operatorname{div}\left[(\partial_{\nu-\alpha}B_A u)\,\nabla\delta_h\partial_\alpha v\right]\right\|_{(\mathrm{H}^1_\diamond)^*}.\end{aligned}$$

Since $\|\delta_h\partial_\nu q\|_{(\mathrm{H}^1_\diamond)^*} \le \|\partial_\nu q\|_{\mathrm{L}^2}$, this bound is uniform in h. Therefore, $v \in \mathrm{H}^{n+1}$ and

$$\|v\|_{\mathrm{H}^{n+1}} \le C\left\{\|q\|_{\mathrm{H}^{n-1}} + \sum_{0\le k\le n-1}(1+\|u\|_{\mathscr{C}^{n-k+2}})\|v\|_{\mathrm{H}^{k+1}}\right\}. \quad (3.2.6.e)$$

When $n > 1$, the following inequalities hold:

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

These are Landau–Kolmogorov inequalities; we have already met them in lemma 2.4.5 on page 53. They can be easily proved by induction from

$$\|\xi\|_{\mathscr{C}^1} \le \sqrt{2\|\xi\|_{\mathscr{C}^0}\|\xi\|_{\mathscr{C}^2}} \qquad \text{and} \qquad \|\xi\|_{\mathrm{H}^1} \le \sqrt{\|\xi\|_{\mathrm{L}^2}\|\xi\|_{\mathrm{H}^2}},$$

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

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

and therefore

$$\|v\|_{\mathrm{H}^{n+1}} \;\leq\; C\left\{\|q\|_{\mathrm{H}^{n-1}} + \left(1 + \|u\|_{\mathscr{C}^3}\right)\|v\|_{\mathrm{H}^n} + \|u\|_{\mathscr{C}^{n+2}}\|v\|_{\mathrm{H}^1}\right\}.$$

This last inequality also holds when $n = 1$, thanks to (3.2.6.e). As $\|v\|_{\mathrm{H}^1} \leq C_\varepsilon\|q\|_{\mathrm{H}^{-1}}$ and $\|u\|_{\mathscr{C}^3} + \|q\|_{\mathrm{H}^{-1}} \leq M$, using our assumption (3.2.6.b) we get

$$\|v\|_{\mathrm{H}^{n+1}} \leq C_{\varepsilon,M,n}\left\{\|q\|_{\mathrm{H}^{n-1}} + \|u\|_{\mathscr{C}^{n+2}}\right\}. \qquad \square$$

3.2.7. Lemma. *Let $\alpha \in (0,1)$. For any $u \in \mathscr{C}^{n+2,\alpha}$ with $A - \nabla^2 u > 0$, and any $q \in \mathscr{C}^{n,\alpha}_\diamond(\mathbb{T}^d)$, there is a unique $v \in \mathscr{C}^{n+2,\alpha}_\diamond(\mathbb{T}^d)$ such that*

$$\mathrm{D}_u\mathcal{F}(A,u)v \;=\; q.$$

Proof. If $q \in \mathscr{C}^{n,\alpha}_\diamond$, then $q \in \mathrm{H}^n_\diamond$, and thus according to lemma 3.2.6 on page 80, there is $v \in \mathrm{H}^{n+2}_\diamond$ such that $\mathrm{D}_u\mathcal{F}(A,u)v = q$ in $[\mathrm{H}^1_\diamond]^*$. But since $\int q = 0$, such an equality in fact holds in H^{-1}. Thus, locally, in a weak sense, $\mathrm{D}_u\mathcal{F}(A,u)v = q$. Then, since $u \in \mathscr{C}^{n+2,\alpha}$, the coefficients of the operator $\mathrm{D}_u\mathcal{F}(A,u)$ are $\mathscr{C}^{0,\alpha}$; this implies $v \in \mathscr{C}^{n+2,\alpha}$ (see for instance the monograph by David Gilbarg and Neil S. Trudinger [31, Theorem 6.13 and 6.17 and 8.22]). $\square$

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

$$\Psi : \begin{cases} \mathcal{S}_d^{++} & \longrightarrow \quad \mathscr{C}^{n+2,\alpha}(\mathbb{T}^d) \\ A & \longmapsto \quad \Psi_A \end{cases} \quad \textit{is} \quad \mathscr{C}^1.$$

Proof. Let us denote by Ω be the set of all $(A,u) \in \mathcal{S}_d^{++} \times \mathscr{C}^{n+2,\alpha}_\diamond(\mathbb{T}^d)$ such that $A - \nabla^2 u > 0$. Then Ω is open, the operator $\mathcal{F} : \Omega \to \mathscr{C}^{n,\alpha}_\diamond(\mathbb{T}^d)$, defined by

$$\mathcal{F}(A,u) \;:=\; f - g\left(\mathrm{Id} - A^{-1}\nabla u\right)\det\left(\mathrm{I}_2 - A^{-1}\nabla^2 u\right),$$

is smooth and, according to lemma 3.2.7 on the preceding page,

$$\mathrm{D}_u\mathcal{F}(A,\psi_A) : \mathscr{C}_\diamond^{n+2,\alpha}(\mathbb{T}^d) \;\to\; \mathscr{C}_\diamond^{n,\alpha}(\mathbb{T}^d)$$

is a bijection. From the Banach–Schauder theorem, we infer it is an isomorphism. Since $\mathcal{F}(A,\Psi_A) = 0$, according to the implicit function theorem, there is a $\mathscr{C}^1$ map Φ such that, for any $(u,B) \in U$, we can have $\mathcal{F}(B,u) = 0$ if and only if $u = \Phi_B$. According to lemma 3.2.3 on page 78, necessarily then $\Phi_B = \Psi_B$. Thus, $\Psi = \Phi$ is a $\mathscr{C}^1$ map $\mathcal{S}_d^{++} \to \mathscr{C}_\diamond^{n+2,\alpha}(\mathbb{T}^d)$. □

3.2.9. We are now going to apply this result to a cost c_t defined by

$$c_t(x,y) \;:=\; \frac{1}{2}\mathrm{d}(x_1,y_1)^2 + \frac{\lambda_t^1}{2}\mathrm{d}(x_2,y_2)^2 + \cdots + \frac{\lambda_t^1\cdots\lambda_t^{d-1}}{2}\mathrm{d}(x_d,y_d)^2,$$

that is, a cost induced by the diagonal matrix $A_t := \mathrm{diag}(1,\lambda_t^1,\lambda_t^1\lambda_t^2,\ldots,\prod\lambda_t^i)$. We assume $\lambda^1,\ldots,\lambda^{d-1} : \mathbb{R} \to [0,+\infty)$ are smooth, with $\lambda_t^k = 0$ if and only if $t = 0$.

For now, we are only interested in positive times. The behavior when $t = 0$ will be studied in the next section.

3.2.10. Proposition. *The map $\psi : t \mapsto \Psi_{A_t}$ is $\mathscr{C}^1$ on $(0,+\infty)$, and satisfies:*

$$\mathrm{div}\left\{f\left[A_t - \nabla^2\psi_t\right]^{-1}\left(\nabla\dot\psi_t - \dot A_t A_t^{-1}\nabla\psi_t\right)\right\} = 0. \tag{3.2.10.a}$$

Moreover, if $u : (0,+\infty) \to \mathscr{C}^{n+2,\alpha}(\mathbb{T}^d)$ is $\mathscr{C}^1$ and satisfies

$$A_t - \nabla^2 u > 0 \quad \textit{and} \quad \mathrm{div}\left\{f\left[A_t - \nabla^2 u_t\right]^{-1}\left(\nabla\dot u_t - \dot A_t A_t^{-1}\nabla u_t\right)\right\} = 0 \tag{3.2.10.b}$$

for all $t \in (0,+\infty)$, and $u_{t_0} = \psi_{t_0}$ for some $t_0 > 0$, then $u_t = \psi_t$ for any $t > 0$.

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

$$\mathrm{D}_u\mathcal{F}(A_t,\psi_t)\dot\psi_t + \mathrm{D}_A\mathcal{F}(A_t,\psi_t)\dot A_t = 0.$$

On the one hand, it follows from lemma 3.2.4 on page 78 that

$$\mathrm{D}_u\mathcal{F}(A_t,\psi_t)\dot\psi_t \;=\; \operatorname{div}\left(f\left[A_t-\nabla^2\psi_t\right]^{-1}\nabla\dot\psi_t\right).$$

On the other hand, a direct computation yields

$$\mathrm{D}_A\mathcal{F}(A_t,\psi_t)\dot A_t \;=\; -\operatorname{div}\left(f\left[A_t-\nabla^2\psi_t\right]^{-1}\dot A_t A^{-1}\nabla\psi_t\right).$$

We thus get (3.2.10.a).

Conversely, if $u:(0,+\infty)\to\mathscr{C}^{n+2,\alpha}(\mathbb{T}^d)$ is $\mathscr{C}^1$ and satisfies (3.2.10.b), 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 3.2.3 on page 78, $u_t=\Psi_{A_t}$ for all times. □

3.3 Initial condition in two dimensions

3.3.1. Due to the very technical nature of the proofs in this section we will only deal with the dimension 2. Then, in section 3.4, we shall explain what changes in higher dimensions.

3.3.2. Let $\lambda:\mathbb{R}\to[0,+\infty)$ 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

$$A_t \;:=\; \begin{pmatrix}1 & 0\\ 0 & \lambda_t\end{pmatrix},$$

which is

$$c_t(x,y) \;:=\; \frac{1}{2}\mathrm{d}(x_1,y_1)^2+\frac{\lambda_t}{2}\mathrm{d}(x_2,y_2)^2.$$

For t nonzero, let ψ_t be the associated Kantorovich potential between the probability measures μ and ν. We assume they have the same properties as before—that is, they are absolutely continuous with strictly positive, smooth densities. Let T_t be the corresponding optimal transport map. Then, according to proposition 3.2.10 on the preceding page, $t\mapsto\psi_t$ and $t\mapsto T_t$ are $\mathscr{C}^1$ on $\mathbb{R}\setminus\{0\}$. Moreover, we know from the theorem of Guillaume Carlier, Alfred Galichon, and Filippo Santambrogio (page 72), that, as t tends to zero, the map T_t converges to the Knothe–Rosenblatt rearrangement R in $\mathrm{L}^2(\mu)$.

3.3.3. Potentials for Knothe's map. By construction, the Knothe–Rosenblatt rearrangement can be written as $T_{\mathrm{K}}(x_1, x_2) = (T^1_{\mathrm{K}}(x_1), T^2_{\mathrm{K}}(x_1, x_2))$, where $x_1 \mapsto T^1_{\mathrm{K}}(x_1)$ is the optimal map between μ^1 and ν^1, and $x_2 \mapsto T^2_{\mathrm{K}}(x_1, x_2)$ is the optimal map between $\mu^2_{x_1}$ and $\nu^2_{T^1_{\mathrm{K}}(x_1)}$. Recall $\{\mu^1, \mu^2\}$ and $\{\nu^1, \nu^2\}$ are the disintegrations of, respectively, μ and ν (definition 3.1.2 on page 70). Thus, there must exist Kantorovich potentials $x_1 \mapsto \phi^1(x_1)$ and $x_2 \mapsto \phi^2(x_1, x_2)$ such that

$$\begin{aligned} T^1_{\mathrm{K}}(x_1) &= x_1 - \partial_1\phi^1(x_1), \\ T^2_{\mathrm{K}}(x_1, x_2) &= x_2 - \partial_2\phi^2(x_1, x_2). \end{aligned}$$

Those potentials are normalized so that $\int \phi^1(x_1)\,\mathrm{d}x_1 = 0$, and $\int \phi^2(x_1, x_2)\,\mathrm{d}x_2 = 0$ for almost all x_1.

3.3.4. As t tends to zero, the optimal map $T_t = \mathrm{Id} - (\partial_1\psi_t, \partial_2\psi_t/\lambda_t)$ converges toward $T_{\mathrm{K}} = \mathrm{Id} - (\partial_1\phi^1, \partial_2\phi^2)$. A first-order expansion might therefore be $\partial_2\psi_t \sim \lambda_t\partial_2\phi^2$. Since ϕ^1 does not depend on x_2, we could simply have $\psi_t \sim \phi^1 + \lambda_t\phi^2$. This leads us to a priori write:

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

with

$$\psi^1_t(x_1) := \int \psi_t(x_1, x_2)\,\mathrm{d}x_2 \qquad \text{and} \qquad \psi^2_t(x) := \frac{1}{\lambda_t}\left(\psi_t(x) - \psi^1_t(x_1)\right).$$

Thus,

$$\int \psi^1_t(x_1)\,\mathrm{d}x_1 = 0 \qquad \text{and} \qquad \int \psi^2_t(x_1, x_2)\,\mathrm{d}x_2 = 0.$$

Such a decomposition allows us to extend our analysis up to $t = 0$.

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

$$\int u^1(x_1)\,\mathrm{d}x_1 = 0 \qquad \text{and} \qquad \int u^2(x_1, x_2)\,\mathrm{d}x_2 = 0,$$

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

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

Next, we define an operator $\mathcal{G} : \Omega \to \mathscr{C}^\infty(\mathbb{T}^2)$. When t is nonzero,

$$\mathcal{G}(t, u^1, u^2) \;:=\; \mathcal{F}(A_t, u^1 + \lambda_t u^2), \tag{3.3.5.a}$$

where $\mathcal{F}$ is the operator introduced in section 3.2:

$$\mathcal{F}(A, u) \;=\; f - g\left(\mathrm{Id} - A^{-1}\nabla u\right) \det\left(\mathrm{I}_2 - A^{-1}\nabla^2 u\right).$$

We then extend $\mathcal{G}$ to include the case $t = 0$; indeed, notice

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

$$\text{and} \qquad A^{-1}\nabla^2(u^1 + \lambda_t u^2) \;=\; \begin{pmatrix} \partial_{1,1}u^1 + \lambda_t \partial_{1,1}u^2 & \lambda_t \partial_{1,2}u^2 \\ \partial_{1,2}u^2 & \partial_{2,2}u^2 \end{pmatrix}.$$

If we use the shorthand $\partial u := (\partial_1 u^1, \partial_2 u^2)$, then $T_{\mathrm{K}} = \mathrm{Id} - \partial\phi$, and

$$\mathcal{G}(0, u^1, u^2) \;=\; f - g\left(\mathrm{Id} - \partial u\right) \det\left(\mathrm{I}_2 - D\partial u\right). \tag{3.3.5.b}$$

Thus, we can just take $\psi_0^1 := \phi^1$ and $\psi_0^2 := \phi^2$.

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

Proof. As u^1, u^2 are uniquely determined by the values of $u_t := u^1 + \lambda_t u^2$, thanks to the formulae

$$u^1(x_1) \;:=\; \int u_t(x_1, x_2)\,\mathrm{d}x_2 \qquad \text{and} \qquad u^2(x) \;:=\; \frac{1}{\lambda_t}\left(u_t(x) - u^1(x_1)\right),$$

the lemma follows directly from lemma 3.2.3 on page 78. □

3.3.7. Alas, the continuity method here seems to fail us: we cannot do the same as in the previous section and apply the implicit function theorem, for if we solve $\mathrm{D}_u\mathcal{G}(0,\psi_0^1,\psi_0^2)(v^1,v^2)=q$, then a priori the solution v^2 is not smooth enough. Indeed, as we will see later, if $q\in\mathrm{H}^n$, then $v^1\in\mathrm{H}^{n+2}$, but we can only get $v^2\in\mathrm{H}^n$. We can, however, bypass this difficulty by considering $\mathscr{C}^\infty$ functions, so as to have an infinite source of smoothness, and use the Nash–Moser implicit function theorem (§2.1.9, on page 48) instead of the usual implicit function theorem.

3.3.8. We need only to use this theorem in a neighborhood of $(0,\psi_0^1,\psi_0^2)\in\Omega$. Let us define this neighborhood, which we denote by Ω_0, in the following way: first, take $\varepsilon>0$ such that $1-\partial_{1,1}\psi_0^1>\varepsilon$ and $1-\partial_{2,2}\psi_0^2>\varepsilon$; then, define Ω_0 as the set of all $(t,u_t^1,u_t^2)\in\Omega$ such that:

$$\text{if}\quad t=0,\quad\text{then}\quad\begin{cases}1-\partial_{1,1}u^1>\varepsilon\\ 1-\partial_{2,2}u^2>\varepsilon,\end{cases}\tag{3.3.8.a}$$

$$\text{if}\quad t\neq0,\quad\text{then}\quad\begin{cases}1-\partial_{1,1}u^1-\lambda_t\partial_{1,1}u^2>\varepsilon\\ A_t-\nabla^2(u^1+\lambda_tu^2)>\varepsilon\lambda_t^{1/2}\,\mathrm{I}_2\,.\end{cases}\tag{3.3.8.b}$$

3.3.9. Zero mean value w.r.t. the 2nd variable. Recall that we denote with a $\diamond$ subscript the sets of maps with zero mean value: $\mathscr{C}^\infty_\diamond$ is thus the set formed by the smooth functions u such that $\int u=0$. When dealing with a space of functions with two variables, we also denote by a "$*,\diamond$" subscript, as in $\mathscr{C}^\infty_{*,\diamond}(\mathbb{T}^2)$ the set formed by the ξ such that $\int\xi(\cdot,x_2)\,\mathrm{d}x_2=0$.

3.3.10. Theorem. *For all $(t,u^1,u^2)\in\Omega_0$, for any $q\in\mathscr{C}^\infty_\diamond(\mathbb{T}^2)$, there is a unique $(v^1,v^2)\in\mathscr{C}^\infty_\diamond(\mathbb{T}^1)\times\mathscr{C}^\infty_{*,\diamond}(\mathbb{T}^2)$ such that*

$$\mathrm{D}_u\mathcal{G}(t,u^1,u^2)(v^1,v^2)=q,\tag{3.3.10.a}$$

Moreover, the inverse operator

$$\mathcal{S}:\begin{cases}\Omega_0\times\mathscr{C}^\infty_\diamond(\mathbb{T}^2) & \to & \mathscr{C}^\infty_\diamond(\mathbb{T}^1)\times\mathscr{C}^\infty_{*,\diamond}(\mathbb{T}^2)\\ \big((t,u^1,u^2),q\big) & \mapsto & (v^1,v^2)\end{cases}\qquad\textit{is smooth tame.}$$

See definition 2.1.7 on page 47 for the precise definition of a smooth tame map.

Proof. We report the proof of the existence of (v^1, v^2) and of the following "tame" estimate

$$\|v^1\|_{\mathrm{H}^{n+2}} + \|\partial_2 v^2\|_{\mathrm{H}^n} \leq C_n \left(\|u^1\|_{\mathscr{C}^{n+3}} + \|u^2\|_{\mathscr{C}^{n+3}} + \|q\|_{\mathrm{H}^n} \right),$$

to the next two subsections. Let us conclude from that point on. Then, all that remains to show is that $\mathcal{S}$ is continuous, and that al the derivatives $\mathrm{D}^k \mathcal{S}$ are tame.

First, if $(t_k, u^1_k, u^2_k, q_k) \in \Omega_0$ converges toward $(t, u^1, u^2, q) \in \Omega_0$, for each k let (v^1_k, v^2_k) be the corresponding inverse. Thanks to the tame estimate (which we have not proved yet), v^1_k and v^2_k 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, as $\mathrm{D}\mathcal{G}$ is continuous, must be the solution of $\mathrm{D}\mathcal{G}(t, u^1, u^2)(v^1, v^2) = q$.

Then, all the derivative $\mathrm{D}^k \mathcal{S}$ are also tame, since they give the solution to the same kind of equation as (3.3.10.a). Indeed, by differentiating (3.3.10.a), we get

$$\mathrm{D}_u \mathcal{G} \, \mathrm{D}\mathcal{S} = \mathrm{D}q - \mathrm{D}(\mathrm{D}_u \mathcal{G})\mathcal{S}.$$ □

3.3.11. Corollary. *The map* $\begin{cases} \mathbb{R} & \to & \mathscr{C}^\infty_\diamond(\mathbb{T}^1) \times \mathscr{C}^\infty_{*,\diamond}(\mathbb{T}^2) \\ t & \mapsto & (\psi^1_t, \psi^2_t) \end{cases}$ *is smooth.*

Proof. On some interval $(-\tau, \tau)$, this is a direct consequence of corollary 2.1.9 on page 48, theorem 3.3.10 on the preceding page, and lemma 3.3.6 on page 86. For larger t, it follows from theorem 3.2.8 on page 82. □

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

$$A_t - \nabla^2 \psi_t > 0 \quad \textit{and} \quad \operatorname{div}\left(f \left[A_t - \nabla^2 \psi_t \right]^{-1} \left(\nabla \dot{\psi}_t - \dot{A}_t A_t^{-1} \nabla \psi_t \right) \right) = 0, \qquad (3.3.12.a)$$

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

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

with ψ^1_0 and ψ^2_0 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 potentials for the Knothe rearrangement, $(0, u_0^1, u_0^2) \in \Omega_0$, so (t, u_t^1, u_t^2) is in Ω_0 at least for t small. For $t \neq 0$, (3.3.12.a) is equivalent to

$$\mathrm{D}_u \mathcal{F}(t, u_t)\dot{u}_t + \mathrm{D}_t \mathcal{F}(t, u_t) = 0,$$

and therefore

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

By assumption, $\mathcal{G}(0, u_0^1, u_0^2) = 0$. Integrating in time, we get $\mathcal{G}(t, u_t^1, u_t^2) = 0$. Therefore, according to lemma 3.3.6 on page 86, we have $u_t^1 = \psi_t^1$ and $u_t^2 = \psi_t^2$, i.e. $u_t = \psi_t$. For larger t's, we apply proposition 3.2.10 on page 83. □

Proof of the invertibility

3.3.13. Let us recall $\mathcal{F}(A, u) = f - g\left(\mathrm{Id} - A^{-1}\nabla u\right) \det\left(\mathrm{I}_2 - A^{-1}\nabla^2 u\right)$, and

$$\mathcal{G}(t, u^1, u^2) := \mathcal{F}(A_t, u^1 + \lambda_t u^2) \qquad \text{with} \qquad A_t := \begin{pmatrix} 1 & 0 \\ 0 & \lambda_t \end{pmatrix}. \tag{3.3.13.a}$$

We want to prove the invertibility of $\mathrm{D}_u \mathcal{G}(t, u^1, u^2)$. The first lemma (§3.3.14, on the current page) will consider the case $t \neq 0$, the second (§3.3.15, on the following page) the case $t = 0$.

3.3.14. Lemma. *For any $(t, u^1, u^2) \in \Omega_0$ with $t \neq 0$, for all $q \in \mathscr{C}_\diamond^\infty(\mathbb{T}^2)$, there is a unique $(v^1, v^2) \in \mathscr{C}_\diamond^\infty(\mathbb{T}^1) \times \mathscr{C}_{*,\diamond}^\infty(\mathbb{T}^2)$ such that*

$$\mathrm{D}_u \mathcal{G}(t, u^1, u^2)(v^1, v^2) = q. \tag{3.3.14.a}$$

Proof. If we set $u_t := u^1 + \lambda_t u^2$, then lemma 3.2.7 on page 82 tells us that there is a unique $v_t \in \mathscr{C}_\diamond^\infty(\mathbb{T}^2)$ such that

$$\operatorname{div}\left(\left(f - \mathcal{G}(t, u^1, u^2)\right)\left[\mathrm{I}_2 - A_t^{-1}\nabla^2 u_t\right]^{-1} A_t^{-1}\nabla v_t\right) = q. \tag{3.3.14.b}$$

Let us define

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

Then, by construction, (v^1, v^2) is the unique pair solving (3.3.14.a). □

3.3.15. Lemma. *For any* $(0, u^1, u^2) \in \Omega_0$, *for all* $q \in \mathscr{C}^\infty_\diamond(\mathbb{T}^2)$, *there is a unique* $(v^1, v^2) \in \mathscr{C}^\infty_\diamond(\mathbb{T}^1) \times \mathscr{C}^\infty_{*,\diamond}(\mathbb{T}^2)$ *such that*

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

Proof. By substitution, for any $\xi \in \mathscr{C}^\infty$, from (3.3.5.b) we get

$$\int \xi\left(x - \partial u(x)\right)\left[f(x) - \mathcal{G}(0, u^1, u^2)(x)\right]\mathrm{d}x \;=\; \int \xi(y)g(y)\,\mathrm{d}y,$$

with $\partial u := (\partial_1 u^1, \partial_2 u^2)$. Therefore, if we differentiate this with respect to u along the direction v, we get

$$-\int \langle \nabla\xi(\mathrm{Id} - \partial u) \,|\, \partial v\rangle \left(f - \mathcal{G}(0, u^1, u^2)\right) - \int \xi(\mathrm{Id} - \partial u)\mathrm{D}_u\mathcal{G}(0, u^1, u^2)(v^1, v^2) \;=\; 0.$$

Since $\nabla[\xi \circ (\mathrm{Id} - \partial u)] = [\mathrm{I}_2 - D\partial u]^* \nabla\xi(\mathrm{Id} - \partial u)$, we have

$$\langle \nabla\xi(\mathrm{Id} - \partial u) \,|\, \partial v\rangle = \left\langle \nabla[\xi \circ (\mathrm{Id} - \partial u)] \,\middle|\, [\mathrm{I}_2 - D\partial u]^{-1}\partial v\right\rangle$$

and this yields

$$\mathrm{D}_u\mathcal{G}(0, u^1, u^2)(v^1, v^2) = \mathrm{div}\left(\left(f - \mathcal{G}(0, u^1, u^2)\right)[\mathrm{I}_2 - D\partial u]^{-1}\,\partial v\right).$$

Notice, then,

$$\left(f - \mathcal{G}(0, u^1, u^2)\right)[\mathrm{I}_2 - D\partial u]^{-1} = g\,(\mathrm{Id} - \partial u)\begin{pmatrix} 1 - \partial_{2,2}u^2 & 0 \\ \partial_{1,2}u^2 & 1 - \partial_{1,1}u^1 \end{pmatrix}.$$

Thus,

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

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

$$
\int \partial_1\left[g\left(x-\partial u(x)\right)\left(1-\partial_{2,2}u^2(x)\right)\partial_1 v^1(x_1)\right]\mathrm{d}x_2 \;=\; \int q(x)\,\mathrm{d}x_2,
$$

which then brings about

$$
\partial_1\left[\left\{\int g\left(x-\partial u(x)\right)\left(1-\partial_{2,2}u^2(x)\right)\mathrm{d}x_2\right\}\partial_1 v^1(x_1)\right] \;=\; \int q(x)\,\mathrm{d}x_2. \tag{3.3.15.a}
$$

As $\int q(x)\,\mathrm{d}x = 0$, there is a smooth $Q:\mathbb{T}^1\to\mathbb{R}$ such that $\partial_1 Q(x_1) = \int q(x_1,x_2)\,\mathrm{d}x_2$, and it is unique if we require $Q(0) = 0$. Thus, taking a primitive of (3.3.15.a), we obtain

$$
\underbrace{\left[\int g\left(x-\partial u(x)\right)\left(1-\partial_{2,2}u^2(x)\right)\mathrm{d}x_2\right]}_{G(x_1)}\partial_1 v^1(x_1) \;=\; Q(x_1)+c,
$$

for some $c\in\mathbb{R}$. 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 with respect to x_1 of the right hand side must be zero. Combined with the condition $\int v^1\,\mathrm{d}x_1 = 0$, we thus have completely characterized v^1.

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

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

and this is exactly the same kind of equation as (3.3.15.a). If we fix $x_1\in\mathbb{T}^1$, the same reasoning can be applied , and in this way we get v^2 as well. □

Proof of the tame estimates

3.3.16. We refer to definition 2.1.6 on page 47 for a precise definition of what a tame estimate is. Basically, our aim here is to show that, locally on $(t, u^1, u^2) \in \Omega_0$ and $q \in \mathscr{C}^\infty_\diamond(\mathbb{T}^2)$, for any $n \in \mathbb{N}$, there is a constant $C_n > 0$ such that, if

$$\mathrm{D}_u\mathcal{G}(t, u^1, u^2)(v^1, v^2) \;=\; q \tag{3.3.16.a}$$

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

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

In fact, we will prove something slightly stronger:

$$\|v^1\|_{\mathrm{H}^{n+2}} + \|\partial_2 v^2\|_{\mathrm{H}^n} \;\leq\; C_n \left(\|u^1\|_{\mathscr{C}^{n+3}} + \|u^2\|_{\mathscr{C}^{n+3}} + \|q\|_{\mathrm{H}^n}\right). \tag{3.3.16.b}$$

Indeed, as $\int v^2(x_1, x_2)\,\mathrm{d}x_2 = 0$, a Poincaré inequality implies

$$\|v^2\|_{\mathrm{H}^n} \leq c_n \|\partial_2 v^2\|_{\mathrm{H}^n}.$$

Notice also that (3.3.16.b) would by itself yields uniqueness in lemma 3.3.14 on page 89.

3.3.17. We start with the case $t \neq 0$. As the bound for $\|v^1\|_{\mathrm{H}^{n+2}}$ simply follows from lemma 3.2.6 on page 80 and an integration with respect to x_2, we just have to find a bound for $\|\partial_2 v^2\|_{\mathrm{H}^n}$. Let us begin with $\|\partial_2 v^2\|_{\mathrm{L}^2}$; we will then proceed by induction.

3.3.18. Lemma. *Let $M, \varepsilon > 0$. There is a constant C, which depends on M and ε, such that, for any $(t, u^1, u^2) \in \Omega_0$ with $t \neq 0$ and for all $q \in \mathscr{C}^\infty_\diamond(\mathbb{T}^2)$ satisfying*

$$\|q\|_{\mathrm{L}^2} + \|u^1\|_{\mathscr{C}^3} + \|u^2\|_{\mathscr{C}^3} \leq M, \tag{3.3.18.a}$$

if $(v^1, v^2) \in \mathscr{C}^\infty_\diamond(\mathbb{T}^1) \times \mathscr{C}^\infty_\diamond(\mathbb{T}^2)$ is a solution of (3.3.16.a), then

$$\|\partial_2 v^2\|_{\mathrm{L}^2} \leq C. \tag{3.3.18.b}$$

Proof. We set $u_t := u_t^1 + \lambda_t u_t^2$ and $v_t := v_t^1 + \lambda_t v_t^2$. Then, $\mathrm{D}_u\mathcal{F}(A_t, u_t)v_t = q$, and (3.3.8.b) in the definition of Ω_0 ensures we an apply lemma 3.2.6 on page 80 and get

$$\|v_t\|_{\mathrm{H}^2} \;\leq\; C_{\varepsilon,M,1}\left\{\|q\|_{\mathrm{L}^2} + \|u_t\|_{\mathscr{C}^3}\right\} \;\leq\; C. \tag{3.3.18.c}$$

We now set

$$\begin{aligned}
B_t &:= \left(f - \mathcal{G}(t, u^1, u^2)\right)\left[\mathrm{I}_2 - A_t^{-1}\nabla^2 u_t\right]^{-1} A_t^{-1} \\
&= \frac{f - \mathcal{G}(t, u^1, u^2)}{\det(A_t - \nabla^2 u_t)}\left[\mathrm{Co}\,(A_t - \nabla^2 u_t)\right]^* \\
&= \frac{g(\mathrm{Id} - A_t^{-1}\nabla u_t)}{\det A_t}\left[\mathrm{Co}\,(A_t - \nabla^2 u_t)\right]^*
\end{aligned}$$

so that, according to (3.3.13.a) and lemma 3.2.4 on page 78, (3.3.16.a) becomes

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

Notice $\det A_t = \lambda_t$ and

$$\mathrm{Co}\,(A_t - \nabla^2 u_t) = \begin{pmatrix} \lambda_t - \lambda_t\partial_{2,2}u^2 & \lambda_t\partial_{1,2}u_t^2 \\ \lambda_t\partial_{1,2}u_t^2 & 1 - \partial_{1,1}u_t^1 \end{pmatrix}.$$

Therefore, we can write

$$B_t \;=\; U_t + V_t/\lambda_t \tag{3.3.18.d}$$

with

$$U_t \;:=\; g(\mathrm{Id} - A_t^{-1}\nabla u_t)\begin{pmatrix} 1 - \partial_{2,2}u^2 & \partial_{1,2}u^2 \\ \partial_{1,2}u^2 & 0 \end{pmatrix}, \tag{3.3.18.e}$$

$$V_t \;:=\; g(\mathrm{Id} - A_t^{-1}\nabla u_t)\begin{pmatrix} 0 & 0 \\ 0 & 1 - \partial_{1,1}u_t \end{pmatrix}. \tag{3.3.18.f}$$

Thus,

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

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

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

that is,

$$\partial_2 \left[g(\operatorname{Id} - A_t^{-1} \nabla u_t)(1 - \partial_{1,1} u_t) \partial_2 v^2 \right] \;=\; q - \operatorname{div}(U_t \nabla v_t). \tag{3.3.18.g}$$

Since $g > \delta$ for some δ, and as (3.3.8.b) in the definition of Ω_0 means $1 - \partial_{1,1} u_t > \varepsilon$, allowing the constant C to change from line to line we get

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

However, $\int v^2(x_1, x_2)\, \mathrm{d}x_2 = 0$ implies $\|v^2\|_{\mathrm{L}^2} \leq C \|\partial_2 v^2\|_{\mathrm{L}^2}$. Therefore,

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

Thus, since $\|U_t\|_{\mathscr{C}^1} \leq C(1 + \|u^1\|_{\mathscr{C}^3} + \|u^2\|_{\mathscr{C}^3}) \leq C$ follows from (3.3.18.e), we obtain

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

Then, using (3.3.18.c), we get the result. □

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

$$\|\partial_2 v^2\|_{\mathrm{H}^n} \;\leq\; C_n \left(\|q\|_{\mathrm{H}^n} + \|u^1\|_{\mathscr{C}^{n+3}} + \|u^2\|_{\mathscr{C}^{n+3}} \right). \tag{3.3.19.a}$$

Proof. Let us assume (3.3.19.a) 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 (3.3.18.g):

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

We already know from lemma 3.2.7 on page 82 that $v_t = v^1 + \lambda_t v^2$ is smooth, therefore, if we apply $\partial_\nu = \partial^{|\nu|}/\partial x_1^{\nu_1} \cdots \partial x_d^{\nu_d}$, we get

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

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

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

Thus,

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

and therefore

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

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

$$\begin{aligned}\|\partial_2\partial_\nu v^2\|_{\mathrm{L}^2} &\le C\sum_{0\le k\le n}\left\|g(\mathrm{Id} - A^{-1}\nabla u_t)(1 - \partial_{1,1}u_t)\right\|_{\mathscr{C}^{n+1-k}}\left\|\partial_2 v^2\right\|_{\mathrm{H}^k} \\ &\qquad + C\left\{\|q\|_{\mathrm{H}^{n+1}} + \|U_t\nabla v_t\|_{\mathrm{H}^{n+2}}\right\}. \quad (3.3.19.b)\end{aligned}$$

On the one hand, we can use the same Landau–Kolmogorov inequalities as in the proof of Lemma 3.2.6, and $a^{1-t}b^t \leq (1-t)a + tb$, to get, for $0 \leq k \leq n$, the following bound:

$$\begin{aligned}\left\|g(\mathrm{Id} - A^{-1}\nabla u_t)(1-\partial_{1,1}u_t)\right\|_{\mathscr{C}^{n+1-k}} \left\|\partial_2 v^2\right\|_{\mathrm{H}^k} \\ \leq c_n \Big(\left\|g(\mathrm{Id} - A^{-1}\nabla u_t)(1-\partial_{1,1}u_t)\right\|_{\mathscr{C}^{n+1}} \left\|\partial_2 v^2\right\|_{\mathrm{L}^2} \\ + \left\|g(\mathrm{Id} - A^{-1}\nabla u_t)(1-\partial_{1,1}u_t)\right\|_{\mathscr{C}^{1}} \left\|\partial_2 v^2\right\|_{\mathrm{H}^n} \Big) .\end{aligned}$$

Recall we have assumed (3.3.19.a) holds true for n; therefore, using (3.3.18.a), we get

$$\begin{aligned}\left\|g(\mathrm{Id} - A^{-1}\nabla u_t)(1-\partial_{1,1}u_t)\right\|_{\mathscr{C}^{n+1-k}} \left\|\partial_2 v^2\right\|_{\mathrm{H}^k} \\ \leq c_n \left(1 + \|q\|_{\mathrm{H}^n} + \|u^1\|_{\mathscr{C}^{n+3}} + \|u^2\|_{\mathscr{C}^{n+3}}\right) .\end{aligned} \tag{3.3.19.c}$$

On the other hand,

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

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

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

Then, using Lemma 3.2.6 we get

$$\|U_t \nabla v_t\|_{\mathrm{H}^{n+2}} \leq c_n \left(\|q\|_{\mathrm{H}^n} + \|u^1\|_{\mathscr{C}^{n+4}} + \|u^2\|_{\mathscr{C}^{n+4}}\right) . \tag{3.3.19.d}$$

Bringing together (3.3.19.b), (3.3.19.c), and (3.3.19.d), we get the estimate we seek. □

3.3.20. Lemma. *The result of lemma 3.3.19 on page 94 still stands when $t = 0$, with the same constants.*

Proof. Let $(0, u^1, u^2) \in \Omega_0$ and $q \in \mathscr{C}^\infty_\diamond(\mathbb{T}^2)$ such that

$$\|q\|_{\mathrm{L}^2} + \|u^1\|_{\mathscr{C}^3} + \|u^2\|_{\mathscr{C}^3} \leq M, \tag{3.3.20.a}$$

Then, since $(s, u^1, u^2) \in \Omega_0$ for s small enough, we can proceed by approximation. Indeed, if (v_s^1, v_s^2) is the solution to $\mathrm{D}_u\mathcal{G}(s, u^1, u^2)(v_s^1, v_s^2) = q$, 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 3.3.19 on page 94. Up to an extraction, there is convergence, which by compact embedding is as strong as we want. But the convergence can only be toward the solution of $\mathrm{D}_u\mathcal{G}(0, u^1, u^2)(v^1, v^2) = q$, hence estimate (3.3.19.a) is still valid for the limit. □

3.4 Higher dimensions

3.4.1. The difficulty in extending those results in higher dimension only comes from the technical nature of section 3.3. We need a decomposition, not only of the potential, but also of the matrix field B, extending (3.3.18.d). The existence of such a decomposition is the only additional difficulty.

Setting and notations

3.4.2. Cost matrix. We consider $d-1$ smooth maps $\lambda^1, \ldots, \lambda^{d-1} : \mathbb{R} \to [0, +\infty)$ such that $\lambda_t^k = 0$ if and only if $t = 0$. We then define

$$A_t := \begin{pmatrix} 1 & & & & \\ & \lambda_t^1 & & & \\ & & \lambda_t^1\lambda_t^2 & & \\ & & & \ddots & \\ & & & & \prod_{i<d} \lambda_t^i \end{pmatrix}.$$

3.4.3. New decomposition of the potential. In that setting, the decomposition of the Kantorovich potential ψ_t becomes

$$\psi_t(x_1, \ldots, x_d) = \psi_t^1(x_1) + \lambda_t^1 \psi_t^2(x_1, x_2) + \ldots + \left(\prod_{i<d} \lambda_t^i\right) \psi_t^d(x_1, \ldots, x_N).$$

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

$$\forall\, x_1, \ldots, x_{k-1}, \qquad \int \psi_t^k(x_1, \ldots, x_{k-1}, y_k)\, \mathrm{d}y_k = 0.$$

For convenience, we set

$$\hat{\psi}^k := \psi^k + \lambda^k \psi^{k+1} + \ldots + \left(\prod_{k \le i < d} \lambda^i \right) \psi^d,$$

so that we may have

$$\hat{\psi}^1 := \psi, \qquad \hat{\psi}^k = \psi^k + \lambda^k \hat{\psi}^{k+1}, \qquad \hat{\psi}^d = \psi^d,$$

and

$$\forall\, x_1, \ldots, x_{k-1}, \qquad \int \cdots \int \hat{\psi}_t^k(x_1, \ldots, x_{k-1}, y_k, \ldots, y_d)\, \mathrm{d}y_k \ldots \mathrm{d}y_d = 0.$$

For instance, if $d = 3$,

$$\psi = \psi^1 + \lambda^1 \psi^2 + \lambda^1 \lambda^2 \psi^3 \qquad \text{and} \qquad \begin{cases} \hat{\psi}^1 = \psi^1 + \lambda^1 \psi^2 + \lambda^1 \lambda^2 \psi^3 \\ \hat{\psi}^2 = \psi^2 + \lambda^2 \psi^3 \\ \hat{\psi}^3 = \psi^3. \end{cases}$$

3.4.4. Domain. We denote by E the set of all $(t, u^1, \ldots, u^d) \in \mathbb{R} \times \prod \mathscr{C}^\infty(\mathbb{T}^k)$ such that

$$\forall\, k \in \{1, \ldots, d\}, \qquad \int u^k \,\mathrm{d}x_k = 0.$$

Then, if $(t, u^1, \ldots, u^d) \in \mathrm{E}$, we set

$$\hat{u}^d := u^d, \qquad \hat{u}^k := u^k + \lambda^k \hat{u}^{k+1}, \qquad u := \hat{u}^1.$$

This is consistent with the previous notation. Notice

$$\nabla u = \begin{pmatrix} \partial_1 \hat{u}^1 \\ \lambda^1 \partial^2 \hat{u}^2 \\ \lambda^1 \lambda^2 \partial_3 \hat{u}^3 \\ \vdots \\ \prod \lambda^k \partial_d \hat{u}^d \end{pmatrix} \qquad \text{and} \qquad A^{-1} \nabla u = \partial \hat{u} = \begin{pmatrix} \partial_1 \hat{u}^1 \\ \partial_2 \hat{u}^2 \\ \partial_3 \hat{u}^3 \\ \vdots \\ \partial_d \hat{u}^d \end{pmatrix},$$

and thus,

$$A^{-1} \nabla^2 u = D \partial \hat{u} = \begin{pmatrix} \partial_{1,2} \hat{u}^1 & 0 & \cdots & 0 \\ \partial_{1,2} \hat{u}^2 & \partial_{2,2} \hat{u}^2 & \ddots & \vdots \\ \vdots & \vdots & \ddots & 0 \\ \partial_{1,d} \hat{u}^d & \partial_{2,d} \hat{u}^d & \cdots & \partial_{d,d} \hat{u}^d \end{pmatrix}. \tag{3.4.4.a}$$

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

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

As previously, we need only to work on a neighborhood Ω_0 of the tuple $(0, u_0^1, u_0^2)$, with u_0^1 and u_0^2 the Kantorovich potentials for the Knothe–Rosenblatt rearrangement. This neighborhood will be defined later on.

Invertibility

3.4.5. We want to solve, for $(0, u) \in \Omega_0$, the equation $\mathrm{D}_u \mathcal{G}(0, u) v = q$. For $t > 0$,

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

Replacing $A^{-1} \nabla^2 u$ and $A^{-1} \nabla v$ with $D \partial \hat{u}$ and $\partial \hat{v}$, we get

$$\mathrm{D}_u \mathcal{G}(t, u) v = \operatorname{div}\left((f - \mathcal{G}(t, u)) \left[\mathrm{I}_d - D \partial \hat{u} \right]^{-1} \partial \hat{v} \right).$$

When $t = 0$, we have $\hat{u}^k = u^k$ and $\partial \hat{u} = \partial u$, so this becomes

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

The trick is to integrate with respect to $x_{k+1}, \dots, x_d$ to get an equation on $v^1, \dots, v^k$. If $v^1, \dots, v^{k-1}$ have already been found, $[\mathrm{I}_d - D\partial u]^{-1}$ being lower triangular thanks to (3.4.4.a), the resulting equation on v^k is of the same kind as the one we have dealt with in lemma 3.3.15 on page 90. The same reasoning can thus be applied.

Tame estimate

3.4.6. As in the two-dimensional case, we need only to find a tame estimate when t is nonzero for the solution $(v^1, \dots, v^d)$ of

$$q = \operatorname{div}(B\nabla v) \qquad \text{with} \qquad B := \frac{g(\mathrm{Id} - A^{-1}\nabla u)}{\det A}\left[\mathrm{Co}\,(A - \nabla^2 u)\right]^* .$$

First, by integrating with respect to x_d, we obtain the same problem as in dimension $d-1$. Therefore, we can proceed by induction on d.

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

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

$$B = B^1 + \frac{1}{\lambda^1}B^2 + \frac{1}{\lambda^1\lambda^2}B^3 + \dots + \frac{1}{\lambda^1\cdots\lambda^{k-2}}B^{k-1} + \frac{1}{\lambda^1\cdots\lambda^{k-1}}\hat{B}^k,$$

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

$$B^i = \begin{pmatrix} & & & \\ & b^i_{i,i} & \cdots & b^i_{i,d} \\ & \vdots & & \\ & b^i_{d,i} & & \end{pmatrix}, \qquad \hat{B}^k = \begin{pmatrix} & & & \\ & \hat{b}^k_{k,k} & \cdots & \hat{b}^k_{k,d} \\ & \vdots & \ddots & \vdots \\ & \hat{b}^k_{d,k} & \cdots & \hat{b}^k_{d,d} \end{pmatrix}.$$

The point is that all the coefficients $b^i_{\alpha,\beta}, \hat{b}^k_{\alpha,\beta}$ can be bounded in $\mathscr{C}^n$ by the norms of the u^i in $\mathscr{C}^{n+2}$ uniformly in t, at least for small t.

3.4.8. Let us assume such a decomposition exists; then,

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

and thus, since

$$v \;=\; v^1 + \lambda^1 v^2 + \ldots + \lambda^1\cdots\lambda_{k-2}v^{k-1} + \lambda^1\cdots\lambda^{k-1}\hat{v}^k,$$

with $\partial_i v^j = 0$ if $i > j$ and as that implies $\partial_i v = \lambda^1\cdots\lambda^{i-1}\partial_i\hat{v}^i$, we have

$$\operatorname{div}(B\nabla v) = \left[\sum_{i<k} \operatorname{div}(B^i\nabla\hat{v}^i)\right] + \operatorname{div}(\hat{B}^k\nabla\hat{v}^k). \tag{3.4.8.a}$$

On the one hand, the matrix $\hat{B}^k$ is symmetric and non-negative, and we can chose the neighborhood Ω_0 so that to ensure

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

On the other hand, since

$$\forall\, x_1,\ldots,x_{k-1}, \quad \int\cdots\int \hat{v}^k(x_1,\ldots,x_d)\,\mathrm{d}x_k\ldots\mathrm{d}x_d = 0,$$

we have a Poincaré inequality:

$$\left\|\hat{v}^k\right\|_{\mathrm{L}^2}^2 \leq C\sum_{i\geq k}\left\|\partial_i\hat{v}^k\right\|_{\mathrm{L}^2}^2.$$

Therefore,

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

and this shows how we can deduce a L^2 estimate for $\hat{v}^k$ from (3.4.8.a) and a series of estimates for $\hat{v}^i$, for $i < k$. Estimates for the norms H^n, $n > 0$, easily follow, by the same reasoning as in lemma 3.3.19 on page 94.

3.4.9. Thus, all we need to do is to prove the existence of the following decomposition:

$$B \;=\; B^1 + \frac{1}{\lambda^1}B^2 + \frac{1}{\lambda^1\lambda^2}B^3 + \ldots + \frac{1}{\lambda^1\cdots\lambda^{d-1}}B^d,$$

with

$$B^i \;=\; \begin{pmatrix} & & & \\ & b^i_{i,i} & \cdots & b^i_{i,d} \\ & \vdots & & \\ & b^i_{d,i} & & \end{pmatrix}.$$

Remember

$$B := \frac{g(\mathrm{Id} - A^{-1}\nabla u)}{\det A}\left[\mathrm{Co}\,(A - \nabla^2 u)\right]^*,$$

and $\det A = \lambda^1\left(\lambda^1\lambda^2\right)\cdots\left(\lambda^1\cdots\lambda^{d-1}\right)$. Therefore, all we have to do is to show how in $\mathrm{Co}\,(A-\nabla^2 u)$ we can gather the λ^k so as to get the decomposition we seek. Since $\partial_{i,j}u = \lambda^1\cdots\lambda^{\max(i,j)-1}\partial_{i,j}\hat{u}^{\max(i,j)}$,

$$\begin{aligned}\left[\mathrm{Co}\,(A-\nabla^2 u)\right]_{i,j} &= \sum_{\substack{\sigma\in\mathfrak{S}_d\\ \sigma(i)=j}} \varepsilon(\sigma)\prod_{\substack{1\le k\le d\\ k\ne i}}(A-\nabla^2 u)_{k,\sigma(k)}\\ &= \sum_{\substack{\sigma\in\mathfrak{S}_d\\ \sigma(i)=j}} \varepsilon(\sigma)\prod_{\substack{1\le k\le d\\ k\ne i}}\lambda^1\cdots\lambda^{\max(k,\sigma(k))-1}\left(\delta_{k,\sigma(k)} - \partial_{k,\sigma(k)}\hat{u}^{\max(k,\sigma(k))}\right).\end{aligned}$$

Thus, for $i \le j$, we set $\omega_{\alpha,\beta} = \lambda^\alpha\cdots\lambda^{\max(\alpha,\beta)-1}\left(\delta_{\alpha,\beta} - \partial_{\alpha,\beta}\hat{u}^{\max(\alpha,\beta)}\right)$. Then,

$$\begin{aligned}\left[\mathrm{Co}\,(A-\nabla^2 u)\right]_{i,j} &= \sum_{\substack{\sigma\in\mathfrak{S}_d\\ \sigma(i)=j}} \varepsilon(\sigma)\prod_{\substack{1\le k\le d\\ k\ne i}}\lambda^1\cdots\lambda^{k-1}\omega_{k,\sigma(k)}\\ &= \sum_{\substack{\sigma\in\mathfrak{S}_d\\ \sigma(i)=j}} \frac{\varepsilon(\sigma)}{\lambda^1\cdots\lambda^{i-1}}\left(\prod_{1\le k\le d}\lambda^1\cdots\lambda^{k-1}\right)\left(\prod_{\substack{1\le k\le d\\ k\ne i}}\omega_{k,\sigma(k)}\right)\\ &= \frac{\det A}{\lambda^1\cdots\lambda^{i-1}}\sum_{\substack{\sigma\in\mathfrak{S}_d\\ \sigma(i)=j}}\varepsilon(\sigma)\prod_{\substack{1\le k\le d\\ k\ne i}}\omega_{k,\sigma(k)}.\end{aligned}$$

Since we have assumed $i \le j$, this is exactly what we wanted.

3.5 An open problem

3.5.1. In this chapter, we have studied the behavior of the optimal transport map when the cost matrix degenerates. But we have done so only for a diagonal cost matrix. It would, however, be possible to consider more general situations; for instance, we could take a cost matrix A that is diagonal in another base $\{u_1, \cdots, u_d\}$.

In dimension two, this other base can be written as

$$u_1 := (\cos\theta, \sin\theta), \qquad u_2 := (-\sin\theta, \cos\theta).$$

Then, the following matrix is diagonal in this basis:

$$A_\lambda := \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & \lambda \end{pmatrix} \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix},$$

that is,

$$A_\lambda = \begin{pmatrix} (\cos\theta)^2 + \lambda(\sin\theta)^2 & (1-\lambda)\cos\theta\sin\theta \\ (1-\lambda)\cos\theta\sin\theta & \lambda(\cos\theta)^2 + (\sin\theta)^2 \end{pmatrix}.$$

Let us assume $\theta \in (-\pi/2, \pi/2)$, and set $\chi = \tan\theta$. Then,

$$A_\lambda = (\cos\theta)^2 \begin{pmatrix} 1 + \lambda\chi^2 & (1-\lambda)\chi \\ (1-\lambda)\chi & \lambda + \chi^2 \end{pmatrix}.$$

We must then pay 1 for each length unit we travel in the direction u_1, and λ in the direction u_2. The associated transport cost is thus

$$c_\lambda(x, y) = \inf_{k\in\mathbb{Z}^2} \frac{1}{2} A_\lambda\,(x - y - k)^2 = \inf_{k\in\mathbb{Z}^2} \frac{1}{2}\langle u_1|x - y - k\rangle^2 + \frac{\lambda}{2}\langle u_2|x - y - k\rangle^2.$$

3.5.2. If χ is rational, e.g. $\chi = p/q$, then the situation remains basically the same. Indeed, considering our two measures as $\mathbb{Z}^2$-periodic measures defined on $\mathbb{R}^2$ and setting $Z = \sqrt{p^2 + q^2}(\mathbb{Z}u_1 + \mathbb{Z}u_2)$, we can see they are also Z-periodic. Then the results of this chapter apply verbatim on $\mathbb{R}^2/Z$.

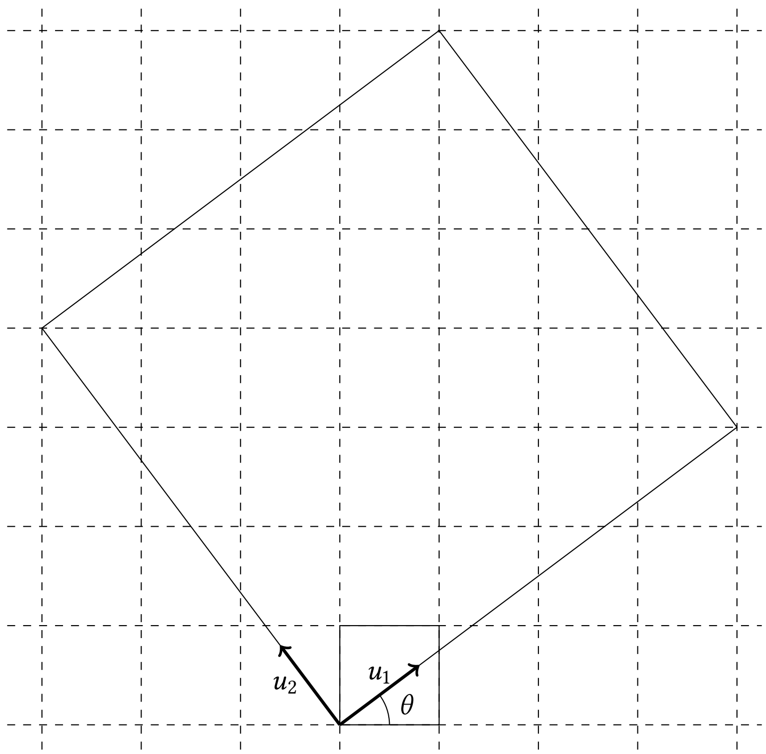

Figure 3.A: When $\chi = \tan\theta$ is rational, we can work on a bigger torus.

3.5.3. But when χ is irrational, then the trajectory $\mathbb{R}u_2$ is dense in $\mathbb{T}^2$. As moving along that direction costs less and less as λ tends to zero, the associated cost tends to zero as well:

3.5.4. Proposition. *If* $\tan\theta \in \mathbb{R} \setminus \mathbb{Q}$*, then* $c_\lambda(x, y) \to 0$ *when* $\lambda \to 0$.

Proof. Without loss of generality, we can assume $y = 0$. Then,

$$\begin{aligned} c_\lambda(x,0) &= \inf_{k\in\mathbb{Z}^2} \frac{1}{2}\left[\langle u_1 | x-k\rangle^2 + \lambda\langle u_2 | x-k\rangle^2\right] \\ &= \inf_{k\in\mathbb{Z}^2} \frac{(\cos\theta)^2}{2}\left[(x_1+\chi x_2-k_1-\chi k_2)^2 + \lambda\,(x_2-\chi x_1-k_2+\chi x_1)^2\right]. \end{aligned}$$

Thus, for any $\varepsilon > 0$, we can find k_1, k_2 such that $|x_1+\chi x_2-k_1-\chi k_2|^2 \leq \varepsilon$, because $\mathbb{Z}+\chi\mathbb{Z}$ is dense in $\mathbb{R}$. Then, taking λ small enough, we get $c_\lambda(x,0) \leq \varepsilon(\cos\theta)^2$. □

3.5.5. On the other hand, the associated optimal transport map is bounded in L^2. So what are its limit points? Is there convergence?

One approach could be to study the Γ-convergence[2] of the functionals

$$F_\lambda : \begin{cases} \Gamma(\mu,\nu) & \to & [0,\infty), \\ \gamma & \mapsto & \int c_\lambda(x,y)\,\mathrm{d}\gamma(x,y). \end{cases}$$

Sadly, even though there are Γ-limit points—there always are—, to identify one of them is not trivial at all. The nature of the irrationality of χ seems to be of some importance, but that makes the problem quite complex.

For more information about Γ-convergence, we refer to Andrea Braides's book on the subject [12].

[2] The Γ in "Γ-convergence" has nothing to do with the set $\Gamma(\mu,\nu)$!

Chapter 4

Numerical computations

4.0.1. When both the source and target measures are discrete, many algorithms already exist for computing the solution to the optimal transport problem, the most famous perhaps being the auction algorithm due to Dimitri Bertsekas [9]. This algorithm was also used by Damien Bosc [11] to deal with continuous measures, by approximation. When only one of the two measures is discrete, Guillaume Carlier, Alfred Galichon, and Filippo Santambrogio [19] showed the optimal transport map could be computed by solving an ODE, starting from the Knothe–Rosenblatt rearrangement. In the general case, Jean-David Benamou and Yann Brenier [7] proposed a method based on their formula (see theorem 1.6.4 on page 40). Sigurd Angenent, Steven Haker, and Allen Tannenbaum [4] developed a steepest-descent algorithm, also starting from the Knothe–Rosenblatt rearrangement. Grégoire Loeper and Francesca Rapetti [40], on the other hand, were able to compute the solution using Newton's method, which is akin to a continuation method.

4.1 A new method

4.1.1. The results exposed in chapter 3 can effectively be applied to compute Brenier's optimal map. This section intends to show how, at least when the underlying space is the torus $\mathbb{T}^2$ and the target measure is uniform. More general cases should be within our reach, even though their implementation is a bit more complex.

4.1.2. As Sigurd Angenent, Steven Haker, and Allen Tannenbaum [4], we start from the Knothe–Rosenblatt rearrangement T_{K}, which is given by two Kantorovich potentials ϕ^1, ϕ^2:

$$T_{\mathrm{K}}(x_1, x_2) \;=\; \begin{pmatrix} x_1 - \partial_1\phi^1(x_1) \\ x_2 - \partial_2\phi^2(x_1, x_2) \end{pmatrix}.$$

Then, as Grégoire Loeper and Francesca Rapetti [40], we use a continuation method: we first set $u_0 = \phi^1$ and $v_0 = \phi^2$; we then increase t little by little, and update u_t and v_t in such a way that $u_t + tv_t$ is always the Kantorovich potential for the cost

$$c_t(x, y) \;:=\; \frac{1}{2}\left|x_1 - y_1\right|^2 + \frac{t}{2}\left|x_2 - y_2\right|^2.$$

Thus, for $t = 1$ we get the Kantorovich potential for the usual quadratic cost, and at that point Brenier's map is just $T_{\mathrm{B}} := \mathrm{Id} - \nabla(u_1 + v_1)$.

In order to update u_t and v_t, we follow the same method as in chapter 3: we use the Monge–Ampère equation. Denoting by f the density of the initial measure, for any t, we should have

$$f \;=\; (1 - \partial^2_{1,1}u_t - t\partial^2_{1,1}v_t)(1 - \partial^2_{2,2}v_t) - t(\partial^2_{1,2}v_t)^2.$$

Therefore, the time derivatives $\dot{u}_t, \dot{v}_t$ are given by the following linearized Monge–Ampère equation:

$$\begin{aligned}(1 - \partial^2_{2,2}v_t)&\partial^2_{1,1}\dot{u}_t \\ &+ t(1 - \partial^2_{2,2}v_t)\partial^2_{1,1}\dot{v}_t + (1 - \partial^2_{1,1}u_t - t\partial^2_{1,1}v_t)\partial^2_{2,2}\dot{v}_t - 2t\partial^2_{1,2}v_t\partial^2_{1,2}\dot{v}_t \\ &\qquad = (\partial^2_{1,2}v_t)^2 - (1 - \partial^2_{2,2}v_t)\partial^2_{1,1}v_t. \end{aligned} \tag{4.1.2.a}$$

This equation, with the aforementioned initial condition, can be broken down as follows:

$$\begin{cases} \partial^2_{1,1}\dot{u}_t(x_1) \;=\; \int p_t(x_1, x_2)\,\mathrm{d}x_2, \\ \operatorname{div}(A_t\nabla\dot{v}_t) \;=\; q_t, \end{cases} \tag{4.1.2.b}$$

with

$$\begin{cases} p_t \;=\; \det(\nabla^2 v_t) + t\operatorname{div}\left(\left[\mathrm{Co}\,\nabla^2 v_t\right]^*\nabla\dot{v}_t\right), \\ q_t \;=\; \det(\nabla^2 v_t) - (1 - \partial^2_{2,2}v_t)\partial^2_{1,1}\dot{u}_t - \partial^2_{1,1}v_t, \end{cases} \tag{4.1.2.c}$$

and

$$A_t = \begin{pmatrix} t(1-\partial^2_{2,2}v) & t\partial^2_{1,2}v \\ t\partial^2_{1,2}v & 1-\partial^2_{1,1}u - t\partial^2_{1,1}v \end{pmatrix},$$

under the conditions

$$\int u_t(x_1)\,\mathrm{d}x_1 = 0 \qquad \text{and} \qquad \int v_t(x_1,x_2)\,\mathrm{d}x_2 = 0. \tag{4.1.2.d}$$

We therefore have four unknown—that is, $u_t, v_t, \dot{u}_t, \dot{v}_t$—and four equations, given by (4.1.2.b) and (4.1.2.d).

4.1.3. Discretization. We proceed with an explicit discretization with respect to time. Given a time step $h > 0$ such that $1/h \in \mathbb{N}$, we compute four sequences of maps: (U_n) and $(\dot{U}_n)$, depending only on the variable x_1, and (V_n) and $(\dot{V}_n)$, depending on x_1 and x_2. The maps U_n and V_n will represent u_{nh} and v_{nh}, and $\dot{U}_n$ and $\dot{V}_n$ will represent $\dot{u}_{nh}$ and $\dot{v}_{nh}$, for $n \in \{0,\dots,1/h\}$. To that end, we first set

$$\begin{cases} U_0 = \phi^1, \\ V_0 = \phi^2, \end{cases} \qquad \text{and} \qquad \begin{cases} \dot{U}_0 = 0, \\ \dot{V}_0 = 0. \end{cases}$$

The values of $\dot{U}_0$ and $\dot{V}_0$ are, in fact, of no consequence. Then, by induction, given the values of $U_n, \dot{U}_n, V_n, \dot{V}_n$, we solve

$$\begin{cases} \partial^2_{1,1}\dot{U}_{n+1} = \int p_n\,\mathrm{d}x_2 \\ \operatorname{div}\Big(A_n\nabla\dot{V}_{n+1}\Big) = q_n \end{cases} \tag{4.1.3.a}$$

with

$$\begin{cases} p_n = \det\Big(\nabla^2 V_n\Big) + nh\operatorname{div}\Big(\big[\operatorname{Co}\nabla^2 V_n\big]^*\nabla\dot{V}_n\Big), \\ q_n = \det\Big(\nabla^2 V_n\Big) - \partial^2_{1,1}V_n - (1-\partial^2_{2,2}V_n)\Big(\partial^2_{1,1}\int p_n\,\mathrm{d}x_2\Big), \end{cases} \tag{4.1.3.b}$$

and

$$A_n = \begin{pmatrix} nh(1-\partial^2_{2,2}V_n) & nh\partial^2_{1,2}V_n \\ nh\partial^2_{1,2}V_n & 1-\partial^2_{1,1}U_n - nh\partial^2_{1,1}V_n \end{pmatrix},$$

under the conditions

$$\int \dot{U}_{n+1}(x_1)\,\mathrm{d}x_1 = 0 \qquad \text{and} \qquad \int \dot{V}_{n+1}(x_1,x_2)\,\mathrm{d}x_2 = 0. \tag{4.1.3.c}$$

The last requirement can be a bit difficult to enforce numerically. However, the following lemma allows us to get $\dot{U}_{n+1}$ and $\dot{V}_{n+1}$ from any pair $(\dot{\underline{U}}_{n+1}, \dot{\underline{V}}_{n+1})$ solving (4.1.3.a):

4.1.4. Lemma. *Let* $(\dot{\underline{U}}_{n+1}, \dot{\underline{V}}_{n+1})$ *solve*

$$\begin{cases} \partial^2_{1,1}\dot{\underline{U}}_{n+1} &= \int p_n\,\mathrm{d}x_2, \\ \operatorname{div}\Big(A_n\nabla\dot{\underline{V}}_{n+1}\Big) &= q_n, \end{cases} \tag{4.1.4.a}$$

without any further condition. Then, (4.1.3.c) *is satisfied by*

$$\begin{aligned} \dot{U}_{n+1} &:= \dot{\underline{U}}_{n+1} - \int \dot{\underline{U}}_{n+1}\,\mathrm{d}x_1 + nh\left(\int \dot{\underline{V}}_{n+1}\,\mathrm{d}x_2 - \iint \dot{\underline{V}}_{n+1}\,\mathrm{d}x_1\,\mathrm{d}x_2\right) \\ \dot{V}_{n+1} &:= \dot{\underline{V}}_{n+1} - \int \dot{\underline{V}}_{n+1}\,\mathrm{d}x_2. \end{aligned}$$

Proof. We set

$$\begin{aligned} \dot{\underline{\underline{U}}}_{n+1} &:= \dot{\underline{U}}_{n+1} - \int \dot{\underline{U}}_{n+1}\,\mathrm{d}x_1 + nh\left(\int \dot{\underline{V}}_{n+1}\,\mathrm{d}x_2 - \iint \dot{\underline{V}}_{n+1}\,\mathrm{d}x_1\,\mathrm{d}x_2\right) \\ \dot{\underline{\underline{V}}}_{n+1} &:= \dot{\underline{V}}_{n+1} - \int \dot{\underline{V}}_{n+1}\,\mathrm{d}x_2. \end{aligned}$$

Then, if $\underline{I}_{n+1} := \int \dot{\underline{V}}_{n+1}\,\mathrm{d}x_2$, since $(\dot{\underline{U}}_{n+1}, \dot{\underline{V}}_{n+1})$ is a solution of (4.1.4.a), we have

$$\begin{aligned} \partial^2_{1,1}\dot{\underline{\underline{U}}}_{n+1} &= \partial^2_{1,1}\dot{\underline{U}}_{n+1} + nh\partial^2_{1,1}\underline{I}_{n+1} = \int p_n\,\mathrm{d}x_2 + nh\partial^2_{1,1}\underline{I}_{n+1}, \\ q_n &= \operatorname{div}\Big(A_n\nabla\dot{\underline{V}}_{n+1}\Big) = \operatorname{div}\Big(A_n\nabla\dot{\underline{\underline{V}}}_{n+1}\Big) + \operatorname{div}(A_n\nabla\underline{I}_{n+1})\,. \end{aligned}$$

Notice that, if w is a function of x_1 only, then

$$\operatorname{div}(A_n w) = nh(1-\partial^2_{2,2}V_n)(\partial^2_{1,1}w).$$

Therefore,

$$\begin{aligned}
\operatorname{div}\left(A_n \nabla \dot{U}_{n+1}\right) &= nh(1-\partial^2_{2,2}V_n)\left(\partial^2_{1,1}\int p_n\right)\\
\operatorname{div}\left(A_n \nabla \dot{\underline{\underline{U}}}_{n+1}\right) &= nh(1-\partial^2_{2,2}V_n)\left(\partial^2_{1,1}\int p_n\,\mathrm{d}x_2 + nh\partial^2_{1,1}\underline{\underline{I}}_{n+1}\right)\\
\operatorname{div}\left(A_n \nabla \dot{\underline{\underline{V}}}_{n+1}\right) &= q_n - nh(1-\partial^2_{2,2}V_n)\left(\partial^2_{1,1}\underline{\underline{I}}_{n+1}\right).
\end{aligned}$$

Then, using (4.1.3.b), we get

$$\operatorname{div}\left(A_n\nabla\left[\dot{\underline{\underline{U}}}_{n+1} + nh\dot{\underline{\underline{V}}}_{n+1}\right]\right) = \operatorname{div}\left(A_n\nabla\left[\dot{U}_{n+1} + nh\dot{V}_{n+1}\right]\right).$$

Since A_n is a positive-definite, symmetric matrix, and both $\dot{\underline{\underline{U}}}_{n+1} + nh\dot{\underline{\underline{V}}}_{n+1}$ and $\dot{U}_{n+1} + nh\dot{V}_{n+1}$ have a zero mean value, they must be equal. Then, it follows from (4.1.3.c) that

$$\dot{U}_{n+1} = \dot{\underline{\underline{U}}}_{n+1} \qquad \text{and} \qquad \dot{V}_{n+1} = \dot{\underline{\underline{V}}}_{n+1}. \qquad \square$$

4.1.5. Thus, we obtain algorithm 4.A, on the next page. How to compute the potentials for the Knothe–Rosenblatt rearrangement is not detailed, as it as been explained elsewhere (see §3.3.3 on page 85).

4.2 Results

4.2.1. FreeFem++. The following results have been obtained with FreeFem++. It is a free software[1] developed at the Jacques-Louis Lions laboratory, in Paris. Its purpose is to solve partial differential equations using the finite-element method. Roughly, this method allows to numerically solve an equation

$$\operatorname{div}(A\nabla u) = q, \qquad u \in \mathrm{H}^1_0 \tag{4.2.1.a}$$

[1]FreeFem++ is *free* in the sense that it can be obtained free of charge, but it also means it is open-source and can be freely shared, studied and modified. It is released under the GNU lesser general public license.

Require: f ▷ Source density
Require: N ▷ Number of time iterations
$t \leftarrow 0$
$U, V \leftarrow$ KnothePotentials$(f, 1)$ ▷ Target is the uniform density
$dU, dV \leftarrow 0, 0$ ▷ This does not really matter
for k **from** 0 **to** $N - 1$ **do**
$p \leftarrow \det(\nabla^2 V) + t\,\mathrm{div}(\left[\mathrm{Co}\,\nabla^2 V\right]^* \nabla dV)$
$dU \leftarrow$ Solve$(\partial^2_{1,1} dU = \int p\,\mathrm{d}x_2)$ ▷ No control over the result
$q \leftarrow \det(\nabla^2 V) - \partial^2_{1,1} V - (1 - \partial^2_{2,2} V)(\partial^2_{1,1} \int p\,\mathrm{d}x_2)$
$A \leftarrow$ Matrix$([[t(1 - \partial^2_{2,2} V), t\partial^2_{1,2} V], [t\partial^2_{1,2} V, 1 - \partial^2_{1,1} U - t\partial^2_{1,1} V]])$
$dV \leftarrow$ Solve$(\mathrm{div}(A\nabla dV) = q)$ ▷ No control over the result
$dU \leftarrow dU - \int dU\,\mathrm{d}x_1 + t\int dV\,\mathrm{d}x_2 - t\iint dV\,\mathrm{d}x_1\,\mathrm{d}x_2$
$dV \leftarrow dV - \int dV\,\mathrm{d}x_2$
$U \leftarrow U + dU/N$
$V \leftarrow V + dV/N$
$t \leftarrow t + 1/N$
end for
return $U + V$

Algorithm 4.a: Computation of the potential for Brenier's map, $T_{\mathrm{B}} = \mathrm{Id} - \nabla(U + V)$

by looking at its variational formulation

$$u \in \operatorname*{arg\,min}_{v \in \mathrm{H}^1_0} \left\{ \frac{1}{2} \int \langle A\nabla v | \nabla v \rangle - \int qv \right\}.$$

(If we work on the torus, there is a priori no boundary condition.) The set H^1_0 is then replaced with a finite-dimensional subspace $V \subset \mathrm{H}^1_0$, for if this subspace V is well chosen, then

$$u_V \in \operatorname*{arg\,min}_{v \in V} \left\{ \frac{1}{2} \int \langle A\nabla v | \nabla v \rangle - \int qv \right\}$$

should be a good approximation of the real solution. Then, if we denote by $e_1, \ldots, e_N$ an orthonormal basis of V, the problem is then equivalent to solve a system

$$\begin{pmatrix} a_{1,1} & \cdots & a_{1,N} \\ \vdots & \ddots & \vdots \\ a_{N,1} & \cdots & a_{N,N} \end{pmatrix} \begin{pmatrix} u_1 \\ \vdots \\ u_N \end{pmatrix} = \begin{pmatrix} q_1 \\ \vdots \\ q_N \end{pmatrix}, \tag{4.2.1.b}$$

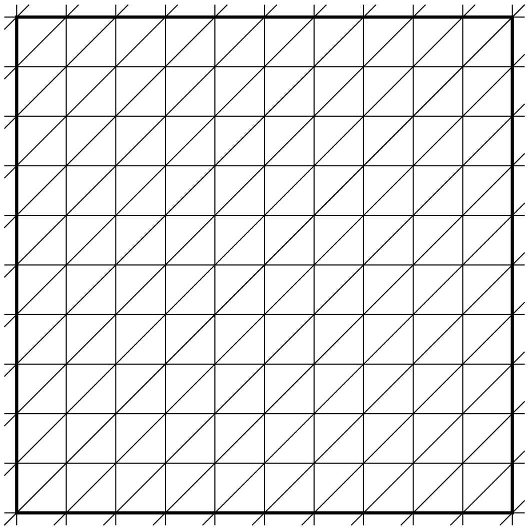

Figure 4.B: A 10×10 mesh on the torus. The space P_1 on this mesh is of dimension $N = 100$.

with

$$a_{i,j} = \int \langle A\nabla e_i | \nabla e_j \rangle, \qquad u_k = \int \langle \nabla e_k | \nabla u \rangle, \qquad q_k = \int q e_k,$$

because

$$u = \sum u_k e_k \qquad \text{and} \qquad q = \sum q_k e_k^*.$$

Solving (4.2.1.a) is therefore reduced to solving the (rather big) linear system (4.2.1.b). The space V used here is P_1, the set of continuous map u over $\mathbb{T}^2$ that are affine on each of the cells of a given mesh—as the one represented on figure 4.B, on this page.

4.2.2. We have tested our algorithm on a 24×24 mesh, with a time step $h = 1/200$, with four initial densities:

1. The first one is a tensor product,

$$f(x, y) = \left(1 + \frac{\sin(2\pi x)}{2}\right)\left(1 + \frac{\sin(2\pi y)}{2}\right).$$

   On figure 4.C, on page 114, it is possible to compare this density f and the density we get with our computation of Brenier's map, that is $\det(\nabla T_B)$. It is not difficult to check that the Knothe–Rosenblatt rearrangement is then theoretically optimal. Indeed, we see on figure 4.D that Brenier's map is very close. The L^2 error that is given is

$$\varepsilon \ := \ \frac{\|f - \det(\nabla T)\|_{L^2}}{\|f\|_{L^2}}.$$

2. The second initial density the algorithm was tested with is:

$$f(x, y) \ = \ 1 - \frac{\cos(2\pi x) + \cos(2\pi y)}{3}.$$

Notice on figure 4.F, on page 115 that the computation of Brenier's map is symmetric, as it should, while Knothe's rearrangement is not. We can see an artifact the left, which is unaccounted for.

3. The third initial density is

$$f(x, y) = 1 + \frac{\sin(2\pi x)\sin(2\pi y)}{2}.$$

This case is interesting, because the projections on the first axis are constant,

$$\int f(x, y)\,\mathrm{d}y \equiv 1,$$

and as a consequence the Knothe–Rosenblatt rearrangement's first component is zero (see figure 4.H, on page 116). As can be seen on figure 4.G, there are more pronounced artifacts on the left and right boundaries, which are hard to explain since all the computations are made on the torus.

4. For the last initial density, we have taken two Gaussian measures that have been made periodic. Artifacts are still present (see figure 4.I, on page 117).

All the FreeFem++ scripts can be found on my website:

http://www.normalesup.org/~bonnotte/thesis/

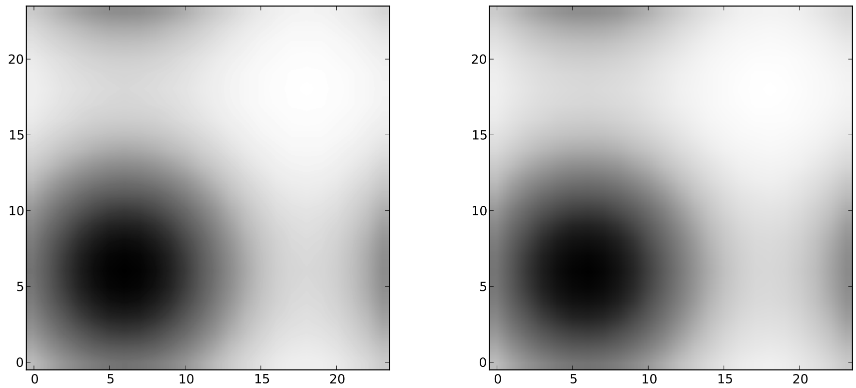

Figure 4.c: The density is $f(x, y) = (1 + \sin(2\pi x)/2)(1 + \sin(2\pi y)/2)$. Left: initial density. Right: density reconstructed using the computed potential. The error in L^2 is $\varepsilon = 1.4\%$.

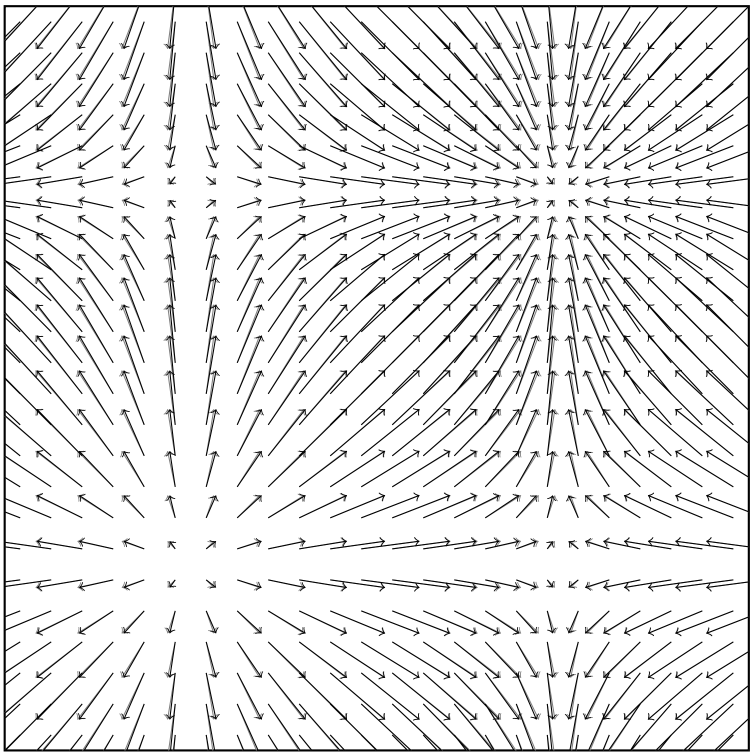

Figure 4.d: As the density is a tensor product, the Knothe rearrangement (gray) is already optimal: nothing changes much, the result of the computations (black) is close.

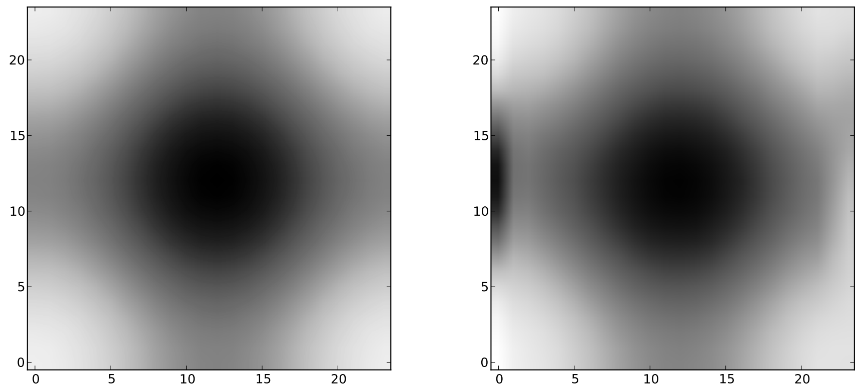

Figure 4.E: The density is $f(x, y) = 1 - (\cos(2\pi x) + \cos(2\pi y))/3$. Left: initial density. Right: density reconstructed using the computed potential. The error in L^2 is $\varepsilon = 7.2\%$.

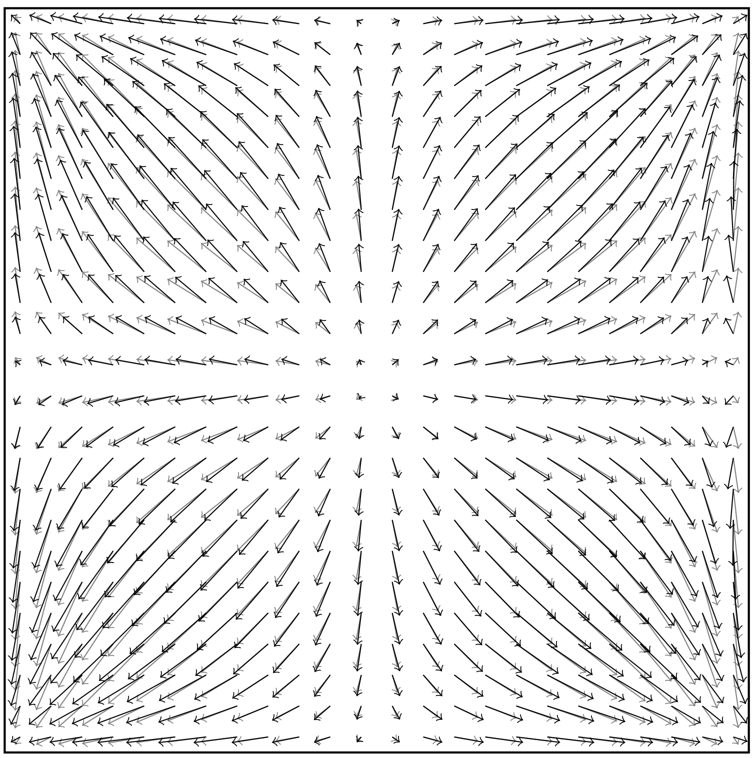

Figure 4.F: The Knothe rearrangement (gray) is close to be optimal, but is not ; compare with the computation of Brenier's map (black), which is symmetric.

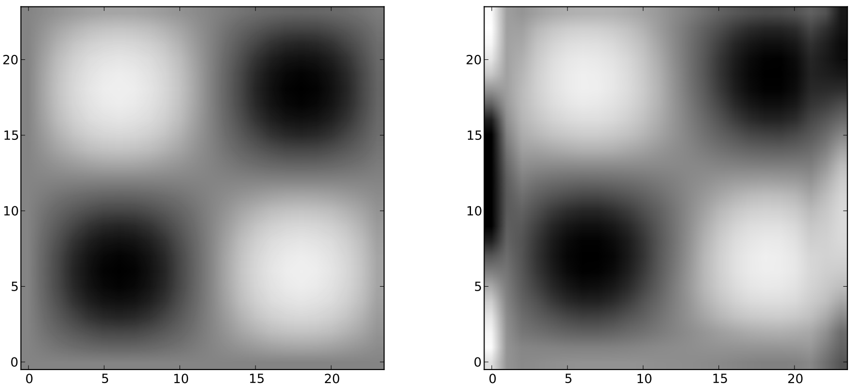

Figure 4.G: The density is $f(x, y) = 1 + \sin(2\pi x)\sin(2\pi y)/2$. Left: initial density. Right: density reconstructed using the computed potential. The error in L^2 is $\varepsilon = 13\%$.

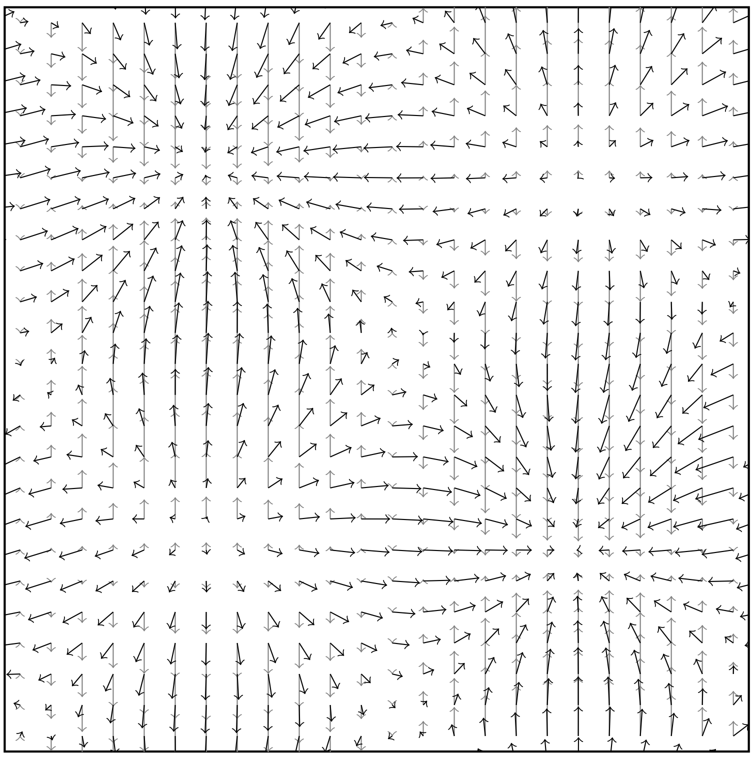

Figure 4.H: Since $\int f(x, y)\,dy = 1$ for any x, the Knothe rearrangement (gray) has a zero first component ; this is not the case with Brenier's map (black).

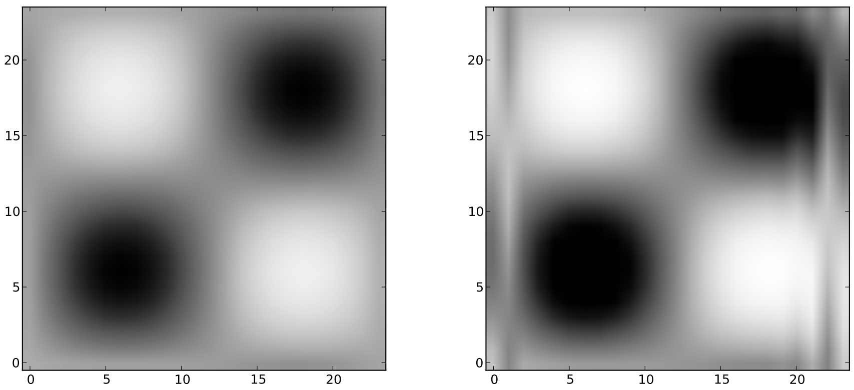

Figure 4.i: Two gaussians, turned periodic. Left: initial density. Right: density reconstructed using the computed potential. The error in L^2 is $\varepsilon = 1.7\%$.

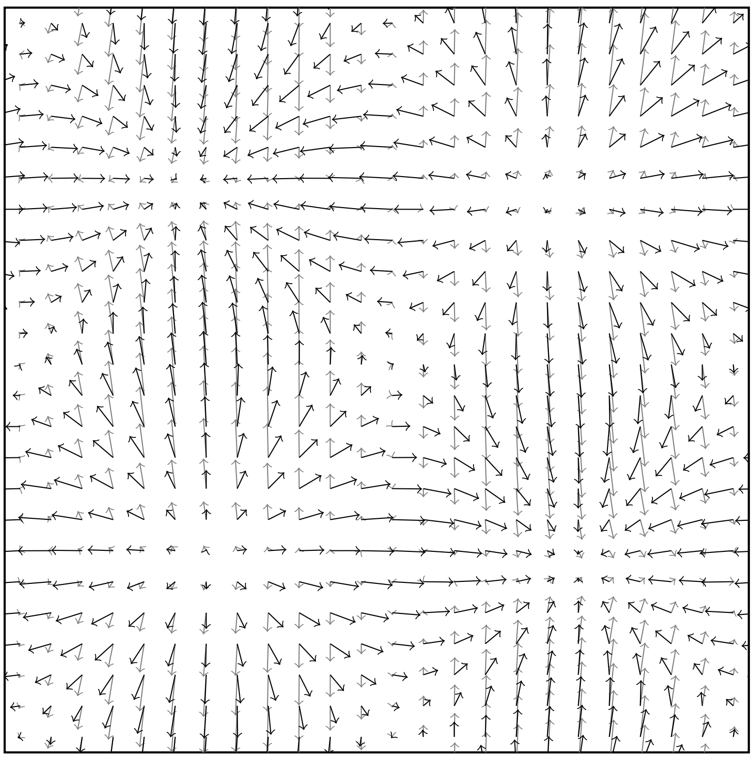

Figure 4.j: The Knothe rearrangement (gray) was not symmetric; Brenier's map (black) is.

4.3 Open questions

In order to have a proper evaluation of this method, four questions need to be addressed:

1. Where do the artifacts come from? Why do they appear on the left and right boundaries, whilst the domain is periodic? An explanation might come from our using an explicit discretization in time; an implicit discretization would probably better, but the computations would become a lot harder.

2. Those numerical experiments presented here were always obtained for a uniform target measure. Is it possible to deal with more general situations? In that case, both the differential equation and the initial condition satisfied by the Kantorovich potential are much more complex, and may need to be carefully handled.

3. Does this method give better results than other algorithms? It would be specially interesting to compare it with the methods of Sigurd Angenent, Steven Haker, and Allen Tannenbaum [4] on the one hand, and Grégoire Loeper and Francesca Rapetti [40] on the other, since both compute the optimal transport map as well. A comparison with the method of Jean-David Benamou and Yann Brenier [7], which computes the geodesic rather than the optimal map, would be less straightforward.

4. At last, numerical convergence and numerical stability are two crucial issues that have been left entirely untouched.

Chapter 5

An isotropic version of the IDT algorithm

5.0.1. In image processing, it is often necessary to transfer the color palette of a reference picture to a target picture—for instance, to homogenize the aspect of a series of shots, e.g. in a film. The two color palettes can be described by measures on the space of all the colors, and any transport map between them yields a possible transfer of coloring. In 2006, François Pitié, Anil C. Kokaram, and Rozenn Dahyot [49] proposed an algorithm to compute such a transfer, which they called "Iterative Distribution Transfer" algorithm. It is based on a succession of unidimensional optimal matching between the projections of the distributions along different axes.

The idea was later taken up by Marc Bernot, who noticed the procedure could be somehow homogenized—indeed, the result of the initial IDT algorithm seems to depend very much on the particular set of axes chosen at each iteration. His remedy was, at each step, to compute matchings for all the axes—instead of selecting a particular subset—, and then average the result. This new version, briefly exposed in a paper he wrote with Julien Rabin, Gabriel Peyré, and Julie Delon [50], can be seen as an explicit Euler scheme for the squared *sliced Wasserstein distance*. Alas, no proof exist for the general convergence of the algorithm toward the target measure, neither for the original nor the homogenized—i.e. isotropic—version.

In this chapter, I would like to present a continuous version of the isotropic IDT algorithm, defined as a gradient flow for the squared sliced Wasserstein distance in the space of probability measures, in the sense of the theory developed by Luigi

Ambrosio, Nicola Gigli and Giuseppe Savaré [3]. I was unable to get convergence toward the target measure, but this point of view might still provide a way to get it.

5.1 The sliced Wasserstein distance

5.1.1. Sliced Wasserstein distance. For any direction $\theta \in \mathbb{S}^{d-1}$, let us denote by θ^* the orthogonal projection on $\mathbb{R}\theta$, that is, $\theta^*(x) := \langle\theta|x\rangle$. Given two probability measures μ and ν, the sliced Wasserstein distance between them is defined as

$$\mathrm{SW}_p(\mu,\nu) \;:=\; \left[\fint_{\mathbb{S}^{d-1}} \mathrm{W}_p(\theta^*_\#\mu, \theta^*_\#\nu)^p \,\mathrm{d}\theta \right]^{1/p}.$$

At first sight, the adjectif "sliced" does not seem to properly describe what the distance represents. It might be more appropriate to talk about a "projected Wasserstein distance" or "Radon–Wasserstein distance", as the projections $\theta^*_\#\mu$ and $\theta^*_\#\nu$ are sometimes called the Radon transforms of μ and ν. However, in Fourier mode, it does result in a slicing, since $\mathcal{F}(\theta^*_\#\mu)(s) = \mathcal{F}\mu(s\theta)$. This is quite convenient, as we can see in the proof of the next statement.

5.1.2. Proposition. *The sliced Wasserstein distance is, indeed, a distance.*

Proof. The triangular inequality is trivial; all there is to show is that $\mathrm{SW}_p(\mu,\nu) = 0$ implies $\mu = \nu$. But if $\mathrm{SW}_p(\mu,\nu) = 0$, then $\theta^*_\#\mu = \theta^*_\#\nu$ for almost every $\theta \in \mathbb{S}^{d-1}$, and this, in turn, yields

$$\mathcal{F}\mu(s\theta) \;=\; \int_{\mathbb{R}^d} e^{-2i\pi s\langle\theta|x\rangle}\,\mathrm{d}\mu(x) \;=\; \mathcal{F}(\theta^*_\#\mu)(s) \;=\; \mathcal{F}(\theta^*_\#\nu)(s) \;=\; \mathcal{F}\nu(s\theta).$$

Since the Fourier transform is injective, we get $\mu = \nu$. □

5.1.3. Proposition. *It $\mu,\nu \in \mathcal{P}_p(\mathbb{R}^d)$, then $\mathrm{SW}_p(\mu,\nu)^p \le c_{d,p}\mathrm{W}_p(\mu,\nu)^p$, with*

$$c_{d,p} \;=\; \frac{1}{d} \fint_{\mathbb{S}^{d-1}} |\theta|_p^p \,\mathrm{d}\theta \;\le\; 1.$$

Notice $c_{d,p} \le 1/d$ as soon as $p \ge 2$.

Proof. Let $\gamma \in \Gamma_o(\mu, \nu)$ be an optimal transport plan. Then $(\theta^* \otimes \theta^*)_\# \gamma$ is a transport plan between $\theta^*_\# \mu$ and $\theta^*_\# \nu$, so

$$\mathrm{W}_p(\theta^*_\# \mu, \theta^*_\# \nu)^p \;\leq\; \int \left|\langle \theta | x\rangle - \langle \theta | y\rangle\right|^p \mathrm{d}\gamma(x, y)$$

Hence, as $\fint \langle \theta | z\rangle^p \,\mathrm{d}\theta = \frac{1}{d} |z|^p \fint |\theta|_p^p \,\mathrm{d}\theta = c_{d,p} |z|^p$,

$$\begin{aligned}
\mathrm{SW}_p(\theta^*_\# \mu, \theta^*_\# \nu)^p \;&\leq\; \int \left(\fint \left|\langle \theta | x\rangle - \langle \theta | y\rangle\right|^p \mathrm{d}\theta \right) \mathrm{d}\gamma(x, y) \\
&\leq\; c_{d,p} \int \left|x - y\right|^p \mathrm{d}\gamma(x, y) \\
&\leq\; c_{d,p} \mathrm{W}_p(\mu, \nu)^p.
\end{aligned}$$

□

5.1.4. Lemma. *There is a constant $C_d > 0$ such that, for all μ, ν supported in $B(0, R)$,*

$$\mathrm{W}_1(\mu, \nu) \;\leq\; C_d R^{d/(d+1)}\, \mathrm{SW}_1(\mu, \nu)^{1/(d+1)}.$$

Proof. First, let us recall proposition 1.5.3 on page 34:

$$\mathrm{W}_1(\mu, \nu) \;=\; \sup \left\{ \int \psi \,\mathrm{d}(\mu - \nu) \,\middle|\, \psi \in \mathrm{Lip}_1(\mathbb{R}^d) \right\}.$$

Then, if we take $\varphi \in \mathscr{C}^\infty_c(\mathbb{R}^d)$ such that φ is radial, $\varphi \geq 0$, $\operatorname{supp} \varphi \subset B(0, 1)$ and $\int \varphi = 1$, and set $\varphi_\lambda(x) := \varphi(x/\lambda)/\lambda^d$, and $\mu_\lambda := \varphi_\lambda * \mu$, and $\nu_\lambda := \varphi_\lambda * \nu$, then, denoting also by $\hat{f}$ the Fourier transform of f,

$$\begin{aligned}
\int \psi \,\mathrm{d}(\mu_\lambda - \nu_\lambda) &= \int \hat{\psi}(\xi)\, [\hat{\mu}(\xi) - \hat{\nu}(\xi)]\, \hat{\varphi}(\lambda \xi) \,\mathrm{d}\xi \\
&= \int_{\mathbb{S}^{d-1}} \int_0^\infty \hat{\psi}(r\theta)\, [\hat{\mu}(r\theta) - \hat{\nu}(r\theta)]\, \hat{\varphi}(\lambda r) r^{d-1} \,\mathrm{d}r \,\mathrm{d}\theta \\
&= \frac{1}{2} \int_{\mathbb{S}^{d-1}} \int_{\mathbb{R}} \hat{\psi}(r\theta)\, [\mathcal{F}(\theta^*_\# \mu)(r) - \mathcal{F}(\theta^*_\# \nu)(r)]\, \hat{\varphi}(\lambda r) r^{d-1} \,\mathrm{d}r \,\mathrm{d}\theta,
\end{aligned}$$

which implies

$$\int \psi \,\mathrm{d}(\mu_\lambda - \nu_\lambda)$$
$$= \frac{1}{2}\int_{\mathbb{S}^{d-1}}\int_{\mathbb{R}}\left[\hat{\psi}(r\theta)e^{2i\pi ru} - \hat{\psi}(r\theta)e^{2i\pi rv}\right]\hat{\varphi}(\lambda r)r^{d-1}\,\mathrm{d}r\,\mathrm{d}\gamma_\theta(u,v)\,\mathrm{d}\theta, \quad (5.1.4.a)$$

where, for each θ, we have taken $\gamma_\theta \in \Gamma_o(\theta^*_\#\mu, \theta^*_\#\nu)$ optimal. However,

$$\int_{\mathbb{R}}\left[\hat{\psi}(r\theta)e^{2i\pi ru} - \hat{\psi}(r\theta)e^{2i\pi rv}\right]\hat{\varphi}(\lambda r)r^{d-1}\,\mathrm{d}r$$
$$= \iint \left[\psi(x)e^{2i\pi r(u-\langle\theta|x\rangle)} - \psi(x)^{2i\pi(v-\langle\theta|x\rangle)}\right]\hat{\varphi}(\lambda r)r^{d-1}\,\mathrm{d}x\,\mathrm{d}r.$$

Dividing the integral in two parts, and replacing x with $x+u\theta$ in the first part, and x with $x+v\theta$ in the second part, we get

$$\int_{\mathbb{R}}\left[\hat{\psi}(r\theta)e^{2i\pi ru} - \hat{\psi}(r\theta)e^{2i\pi rv}\right]\hat{\varphi}(\lambda r)r^{d-1}\,\mathrm{d}r$$
$$= \iint \left[\psi(x+u\theta) - \psi(x+v\theta)\right]e^{-2i\pi r\langle\theta|x\rangle}\hat{\varphi}(\lambda r)r^{d-1}\,\mathrm{d}x\,\mathrm{d}r.$$

Since γ_θ is supported in $[-R,R]^2$, and μ_λ, ν_λ are supported in $B(0, R+\lambda)$, we can assume the map $x \mapsto \psi(x+u\theta) - \psi(x+v\theta)$ is supported in $B(0, 2R+\lambda)$ for almost every u, v, and

$$\left|\int_{\mathbb{R}}\left[\hat{\psi}(r\theta)e^{2i\pi ru} - \hat{\psi}(r\theta)e^{2i\pi rv}\right]e^{-\pi\lambda r^2}r^{d-1}\,\mathrm{d}r\right|$$
$$\begin{aligned}
&\leq (2R+\lambda)^d|\mathbb{S}^{d-1}|\int |u-v||\hat{\varphi}(\lambda r)||r|^{d-1}\,\mathrm{d}r \\
&\leq \frac{(2R+\lambda)^d|\mathbb{S}^{d-1}|}{\lambda^d}\left(\int \hat{\varphi}(r)|r|^{d-1}\,\mathrm{d}r\right)|u-v| \\
&\leq \frac{(2R+\lambda)^d C_d}{\lambda^d}|u-v|.
\end{aligned}$$

Thanks to (5.1.4.a), this yields

$$\mathrm{W}_1(\mu_\lambda, \nu_\lambda) = \sup_\psi \int \psi\,\mathrm{d}(\mu_\lambda - \nu_\lambda) \leq \frac{C_d(2R+\lambda)^d}{\lambda^d}\mathrm{SW}_1(\mu,\nu), \quad (5.1.4.b)$$

although perhaps with a different constant C_d.

Let us now find an upper bound on $\mathrm{W}_1(\mu,\nu) - \mathrm{W}_1(\mu_\lambda,\nu_\lambda)$. Notice

$$\begin{aligned}\int \psi \,\mathrm{d}(\mu-\nu) - \mathrm{W}_1(\mu_\lambda,\nu_\lambda) &\le \int \psi \,\mathrm{d}(\mu-\nu) - \int \psi \,\mathrm{d}(\mu_\lambda-\nu_\lambda) \\ &\le \int (\psi - \varphi_\lambda * \psi)\,\mathrm{d}(\mu-\nu).\end{aligned}$$

But

$$\begin{aligned}\int (\psi - \varphi_\lambda * \psi)\,\mathrm{d}(\mu-\nu) &= \iint [\psi(x) - \psi(x-y)]\,\varphi_\lambda(y)\,\mathrm{d}y\,\mathrm{d}(\mu-\nu)(x) \\ &= \iint [\psi(x) - \psi(x-\lambda y)]\,\varphi(y)\,\mathrm{d}y\,\mathrm{d}(\mu-\nu)(x),\end{aligned}$$

and for any $y \in B(0,1)$,

$$\begin{aligned}\int [\psi(x) - \psi(x-\lambda y)]\,\mathrm{d}(\mu-\nu)(x) &\le \int \left|\psi(x) - \psi(x-\lambda y)\right|\mathrm{d}(\mu+\nu)(x) \\ &\le 2\lambda|y|,\end{aligned}$$

thus

$$\int \psi \,\mathrm{d}(\mu-\nu) - \mathrm{W}_1(\mu_\lambda,\nu_\lambda) \le 2\lambda \int |y|\varphi(y)\,\mathrm{d}y.$$

Taking the supremum over ψ, we get $\mathrm{W}_1(\mu,\nu) - \mathrm{W}_1(\mu_\lambda,\nu_\lambda) \le C_d\lambda$.

Combining this last inequality with (5.1.4.b), we obtain

$$\mathrm{W}_1(\mu,\nu) \le C_d\left(\frac{(2R+\lambda)^d}{\lambda^d}\mathrm{SW}_1(\mu,\nu) + \lambda\right).$$

If we take $\lambda = R^{d/(d+1)}\mathrm{SW}_1(\mu,\nu)^{1/(d+1)}$, we get

$$\mathrm{W}_1(\mu,\nu) \le C_d\left(\frac{(2R+\lambda)^d}{R^d} + 1\right)R^{d/(d+1)}\mathrm{SW}_1(\mu,\nu)^{1/(d+1)}.$$

As $\mathrm{SW}_1(\mu,\nu) \le 2R$, we have $\lambda \le 2^{1/(d+1)}R$, hence the announced inequality, with maybe yet another constant C_d. □

5.1.5. Theorem (Equivalence of SW_p and W_p). *There is a constant $C_{d,p} > 0$ such that, for all $\mu, \nu \in \mathcal{P}(B(0, R))$,*

$$\mathrm{SW}_p(\mu, \nu)^p \;\leq\; c_{d,p} \mathrm{W}_p(\mu, \nu)^p \;\leq\; C_{d,p} R^{p-1/(d+1)}\, \mathrm{SW}_p(\mu, \nu)^{1/(d+1)}.$$

Proof. This follows from the previous lemma, as on the one hand,

$$\mathrm{W}_p(\mu, \nu)^p \leq (2R)^{p-1} \mathrm{W}_1(\mu, \nu),$$

and on the other hand

$$\mathrm{SW}_1(\mu, \nu) \leq \mathrm{SW}_p(\mu, \nu). \qquad \square$$

Notice the exponent $p - 1/(d + 1)$ on R is the only one for which the inequality would be preserved by dilations, given the exponent on SW_p.

5.1.6. Proposition. *Let $\mu, \nu \in \mathcal{P}(K)$, with K a compact subset of $\mathbb{R}^d$, and assume ν absolutely continuous. Then, for each direction $\theta \in \mathbb{S}^{d-1}$, there is a Kantorovich potential ψ_θ between $\theta^*_\# \mu$ and $\theta^*_\# \nu$ for the cost $c_\theta(s, t) = |s - t|^2/2$, and, if $\bar{\mu} \in \mathcal{P}(K)$,*

$$\lim_{\varepsilon \to 0^+} \frac{\mathrm{SW}_2((1-\varepsilon)\mu + \varepsilon\bar{\mu}, \nu)^2 - \mathrm{SW}_2(\mu, \nu)^2}{2\varepsilon} = \fint_{\mathbb{S}^{d-1}} \int_K \psi_\theta(\langle \theta | x \rangle)\, \mathrm{d}(\bar{\mu} - \mu)(x)\, \mathrm{d}\theta.$$

This is to be compared to proposition 1.5.6 on page 36, which dealt with a similar result for the usual Wasserstein distance.

Proof. Since ν is absolutely continuous, for each θ the projected measure $\theta^*_\# \nu$ is also absolutely continuous on $\theta^*(K)$; therefore, there is indeed a Kantorovich potential ψ_θ between $\theta^*_\# \mu$ and $\theta^*_\# \nu$. Since ψ_θ is, a priori, not optimal between $(1 - \varepsilon)\mu + \varepsilon\bar{\mu}$ and ν,

$$\liminf_{\varepsilon \to 0^+} \frac{\mathrm{SW}_2((1-\varepsilon)\mu + \varepsilon\bar{\mu}, \nu)^2 - \mathrm{SW}_2(\mu, \nu)^2}{2\varepsilon} \;\geq\; \fint \int \psi_\theta(\langle \theta | x \rangle)\, \mathrm{d}(\bar{\mu} - \mu)(x)\, \mathrm{d}\theta.$$

Conversely, let ψ^ε_θ be a Kantorovich potential between $\theta^*_\#(1 - \varepsilon)\mu + \varepsilon\bar{\mu}$ and $\theta^*_\# \nu$, with $\int \psi^\varepsilon_\theta \,\mathrm{d}\theta^*_\#[(1 - \varepsilon)\mu + \varepsilon\bar{\mu}] = 0$. Then,

$$\frac{1}{2}\mathrm{SW}_2((1-\varepsilon)\mu + \varepsilon\bar{\mu}, \nu)^2 - \frac{1}{2}\mathrm{SW}_2(\mu, \nu)^2 \;\geq\; \varepsilon \fint \int \psi^\varepsilon_\theta(\langle \theta | x \rangle)\, \mathrm{d}(\bar{\mu} - \mu)(x)\, \mathrm{d}\theta.$$

As in the proof of proposition 1.5.6 on page 36, ψ_θ^ε uniformly converges, when $\varepsilon \to 0$ to a Kantorovich potential for the pair $(\theta^*_\#\mu, \theta^*_\#\nu)$. Then, by Lebesgue's dominated convergence theorem,

$$\limsup_{\varepsilon\to 0^+} \frac{\mathrm{SW}_2((1-\varepsilon)\mu + \varepsilon\bar{\mu}, \nu)^2 - \mathrm{SW}_2(\mu,\nu)^2}{2\varepsilon} \leq \fint\int \psi_\theta(\langle\theta|x\rangle)\, \mathrm{d}(\bar{\mu}-\mu)(x)\, \mathrm{d}\theta. \quad \square$$

5.1.7. Proposition. *Let μ and $\nu \in \mathcal{P}(K)$, with K a compact subset of $\mathbb{R}^d$, and assume μ is absolutely continuous. For any $\theta \in \mathbb{S}^{d-1}$, let ψ_θ is the (unique up to an additive constant) Kantorovich potential between $\theta^*_\#\mu$ and $\theta^*_\#\nu$. If ζ is a diffeomorphism of K, then*

$$\lim_{\varepsilon\to 0} \frac{\mathrm{SW}_2([\mathrm{Id}+\varepsilon\zeta]_\#\mu, \nu)^2 - \mathrm{W}_2(\mu,\nu)^2}{2\varepsilon} = \int\fint \psi'_\theta(\langle\theta|x\rangle)\, \langle\theta|\zeta(x)\rangle\, \mathrm{d}\theta\, \mathrm{d}\mu(x).$$

This is the sliced equivalent of proposition 1.5.7 on page 38.

Proof. As ψ_θ is a Kantorovich potential between $\theta^*_\#\mu$ and $\theta^*_\#\nu$,

$$\frac{\mathrm{SW}_2([\mathrm{Id}+\varepsilon\zeta]_\#\mu, \nu)^2 - \mathrm{W}_2(\mu,\nu)^2}{2\varepsilon} \geq \int\fint \frac{\psi_\theta(\langle\theta|x+\varepsilon\zeta(x)\rangle) - \psi(\langle\theta|x\rangle)}{2\varepsilon}\, \mathrm{d}\theta\, \mathrm{d}\mu(x).$$

Since ψ is differentiable almost everywhere, Lebesgue's dominated convergence theorem ensures

$$\liminf_{\varepsilon\to 0^+} \frac{\mathrm{SW}_2([\mathrm{Id}+\varepsilon\zeta]_\#\mu, \nu)^2 - \mathrm{W}_2(\mu,\nu)^2}{2\varepsilon} \geq \int\fint \psi'_\theta(\langle\theta|x\rangle)\, \langle\theta|\zeta(x)\rangle\, \mathrm{d}\theta\, \mathrm{d}\mu(x).$$

Conversely, let $\gamma_\theta \in \Gamma_o(\theta^*_\#\mu, \theta^*_\#\nu)$ be an optimal plan. Then, we can extend γ_θ into $\pi_\theta \in \Gamma(\mu,\nu)$ such that $(\theta^*\otimes\theta^*)_\#\pi_\theta = \gamma_\theta$; for instance, by disintegrating $\mu\otimes\nu$ with respect to $\theta^*\otimes\theta^*$,

$$\int \xi(x,y)\, \mathrm{d}(\mu\otimes\nu)(x,y) = \int\left(\int \xi(u\theta+\hat{x}, v\theta+\hat{y})\, \mathrm{d}[\mu\otimes\nu]_{u,v}(\hat{x},\hat{y})\right) \mathrm{d}[(\theta^*\otimes\theta^*)_\#(\mu\otimes\nu)](u,v),$$

and then replacing $(\theta^* \otimes \theta^*)_\#(\mu \otimes \nu)$ with γ_θ:

$$\int \xi(x,y)\,\mathrm{d}\pi_\theta(x,y) \;=\; \int \left(\int \xi(u\theta + \hat{x}, v\theta + \hat{y})\,\mathrm{d}[\mu \otimes \nu]_{u,v}(\hat{x},\hat{y}) \right) \mathrm{d}\gamma_\theta(u,v).$$

Now, $[(\theta^* + \varepsilon\theta^*(\zeta)) \otimes \theta^*]_\#\pi_\theta$ is a transport plan between $\theta^*_\#[\mathrm{Id} + \varepsilon\zeta]_\#\mu$ and $\theta^*_\#\nu$; hence,

$$\begin{aligned} &\mathrm{SW}_2([\mathrm{Id} + \varepsilon\zeta]_\#\mu, \nu)^2 - \mathrm{SW}_2(\mu,\nu)^2 \\ &\qquad \le \fint \int \left|\langle\theta|x + \varepsilon\zeta(x) - y\rangle\right|^2 - \left|\langle\theta|x - y\rangle\right|^2 \mathrm{d}\pi_\theta(x,y)\,\mathrm{d}\theta. \end{aligned}$$

But for π_θ-almost every pair (x,y), we have $\langle\theta|y\rangle = \langle\theta|x\rangle - \psi'_\theta(\langle\theta|x\rangle)$, so

$$\begin{aligned} &\mathrm{SW}_2([\mathrm{Id} + \varepsilon\zeta]_\#\mu, \nu)^2 - \mathrm{SW}_2(\mu,\nu)^2 \\ &\qquad \le \fint \int \left|\psi'_\theta(\langle\theta|x\rangle) - \varepsilon\langle\theta|\zeta(x)\rangle\right|^2 - \left|\psi'_\theta(\langle\theta|x\rangle)\right|^2 \mathrm{d}\pi_\theta(x,y)\,\mathrm{d}\theta. \end{aligned}$$

This immediately yields

$$\limsup_{\varepsilon \to 0^+} \frac{\mathrm{SW}_2([\mathrm{Id} + \varepsilon\zeta]_\#\mu, \nu)^2 - \mathrm{W}_2(\mu,\nu)^2}{2\varepsilon} \;\le\; \int \fint \psi'_\theta(\langle\theta|x\rangle)\,\langle\theta|\zeta(x)\rangle\,\mathrm{d}\theta\,\mathrm{d}\mu(x). \quad \square$$

5.2 The Iterative Distribution Transfer algorithm

5.2.1. The algorithm proposed by François Pitié, Anil C. Kokaram, and Rozenn Dahyot [49] starts from a given measure μ, and, for any target measure ν, builds a sequence $(\mu_n)_{n\in\mathbb{N}}$ such that $\mu_0 = \mu$ and μ_n seems to tend to ν when n tends to infinity. Convergence, however, is assured only empirically, as the authors were able to prove it only when ν is a Gaussian measure.

If μ_n has been set, then μ_{n+1} is defined as follows. First, chose an orthogonal basis $B_n = (e_1^n, \ldots, e_d^n)$ in $\mathbb{R}^d$, and take the projections $e_{i\;\#}^{n*}\mu_n$ and $e_{i\;\#}^{n*}\nu$. For each axis i, there is an unidimensional optimal matching between the projections, which we will denote by $t_{e_i^n} : \mathbb{R} \to \mathbb{R}$. Let

$$T_n(x) := \sum_{i=1}^{d} t_{e_i^n}(\langle e_i^n|x\rangle)e_i^n,$$

and set $\mu_{n+1} := T_{n\#}\mu_n$. Then $e^{n*}_{i\,\#}\mu_{n+1} = \langle e^n_i | T_n \rangle_\# \mu_n = [t_{e^n_i} \circ e^{n*}_i]_\# \mu_n = e^{n*}_{i\,\#}\nu$. Thus, μ_{n+1} should be closer to ν than μ_n.

5.2.2. Theorem (Pitié–Kokaram–Dahyot). *We assume μ is absolutely continuous, and ν is a Gaussian measure. Then,*

1. *If B_n are independent, uniform random variables on the set of all orthonormal basis, i.e. on $\mathrm{O}(d)$, then $\mu_n \to \nu$ almost surely.*
2. *Alternatively, if the bases B_n are dense, then $\mu_n \to \nu$.*

It may be said that the original proof by François Pitié, Anil C. Kokaram, and Rozenn Dahyot lacks in precision, on two counts:

- The absolute continuity of the measures μ_n is crucial, but not proved.
- The reader might be misled into believing some kind of uniform continuity for $(\theta, \mu) \mapsto \mathrm{Ent}(\theta_\#\mu | \theta_\#\nu)$ is used, which, of course, is not possible.

The following proof addresses both issues.

Proof. The first thing to check is that the measures μ_n are always absolutely continuous. If we know μ_n is absolutely continuous, then the transport map T_n, which is such that $\mu_{n+1} = T_{n\#}\mu_n$, is $\mathrm{W}^{1,1}$. Moreover, it is easy to check from its definition that T_n is injective on the support of μ_n. Then, there is a μ_n-negligible set N_n and a sequence $(A_k)_{k\in\mathbb{N}}$ of disjoint Borel sets such that

$$\mathbb{R}^d = N_n \cup \bigcup_{k\in\mathbb{N}} A_k \quad \text{and} \quad T_n\big|_{A_k} = a_k$$
$$\text{with} \quad a_k \in \mathscr{C}^1 \quad \text{and} \quad |\det(\mathrm{D}a_k)| \geq \varepsilon_k \quad \mu_n\text{-a.e.},$$

see the book by Lawrence C. Evans & Ronald F. Gariepy [28, Section 6.6.3]. Thus, if N is a negligible set for the Lebesgue measure, and ρ_n stands for the density of μ_n,

$$\mu_{n+1}(N) \;\leq\; \sum_{k\in\mathbb{N}} \frac{1}{\varepsilon_k} \int_{A_k} \mathbb{1}_N(a_k(x))\rho_n(x)|\det(\mathrm{D}a_k)| \,\mathrm{d}x \;=\; 0.$$

Therefore, μ_{n+1} is also absolutely continuous.

Now, the key property of ν is that, being a Gaussian measure, it enjoys a tensorization property: for any basis $B = (e_1, \ldots, e_d)$, we always have $\nu = e^*_{1\,\#}\nu \otimes \cdots \otimes e^*_{d\#}\nu$.

Therefore, if (x_i) are the coordinates of x in the base B and if we denote by f_n, $f_{n,i}$, and g, g_i the respective densities of μ_n, $e^*_{i\,\#}\mu_n$, and ν, $e^*_{i\,\#}\nu$, we get

$$\begin{aligned}\mathrm{Ent}\,(\mu_n \,|\, \nu) &= \int f_n \ln\left[\frac{f_n}{f_{n,1}\cdots f_{n,d}}\right] + \int f_n \ln\left[\frac{f_{n,1}\cdots f_{n,k}}{g_1\cdots g_k}\right] \\ &= \mathrm{Ent}\left(\mu_n \,\middle|\, e^*_{1\,\#}\mu_N \otimes\cdots\otimes e^*_{d\#}\mu_n\right) + \sum_{i=1}^{d}\mathrm{Ent}\left(e^*_{i\,\#}\mu_n \,\middle|\, e^*_{i\,\#}\nu\right).\end{aligned}$$

If B is now the basis chosen to build μ_{n+1}, i.e. $B = B_n = (e^n_1, \ldots, e^n_d)$, we have

$$\mathrm{Ent}\left(e^{n*}_{i\,\#}\mu_{n+1} \,\middle|\, e^{n*}_{i\,\#}\nu\right) = 0.$$

On the other hand,

$$\begin{aligned}\mathrm{Ent}\left(\mu_{n+1} \,\middle|\, e^{n*}_{1\,\#}\mu_{n+1}\otimes\cdots\otimes e^{n*}_{d\#}\mu_{n+1}\right) &= \int \ln\left[\frac{f_{n+1}(y)}{\prod f_{n+1,i}(y_i)}\right]\mathrm{d}\mu_{n+1}(y) \\ &= \int \ln\left[\frac{f_{n+1}(T(x))}{\prod f_{n+1,i}(t_{e^n_i}(x_i))}\right]\mathrm{d}\mu_n(x),\end{aligned}$$

and as $\mathrm{D}T$ is diagonal, with $t'_{e^n_1}, \ldots, t'_{e^n_d}$ on the diagonal, and $f_{n,i} = (f_{n+1,i}\circ t_{e^i_n})t'_{e^i_n}$, we get

$$f_{n+1}(T(x)) = \frac{f_n(x)}{\det \mathrm{D}T(x)} = \frac{f_n(x)}{t'_{e^n_1}(x_1)\ldots t'_{e^n_d}(x_d)} = f_n(x)\prod_{k=1}^{d}\frac{f_{n+1,i}(t_{e^n_i}(x_i))}{f_{n,i}(x_i)}.$$

This implies $\mathrm{Ent}(\mu_{n+1}|e^{n*}_{1\,\#}\mu_{n+1}\otimes\cdots\otimes e^{n*}_{d\#}\mu_{n+1}) = \mathrm{Ent}(\mu_n|e^{n*}_{1\,\#}\mu_n\otimes\cdots\otimes e^{n*}_{d\#}\mu_n)$. Hence,

$$\begin{aligned}\mathrm{Ent}\,(\mu_{n+1} \,|\, \nu) &= \mathrm{Ent}\left(\mu_{n+1} \,\middle|\, e^{n*}_{1\,\#}\mu_{n+1}\otimes\cdots\otimes e^{n*}_{d\#}\mu_{n+1}\right) + \sum_{i=1}^{d}\mathrm{Ent}\left(e^{n*}_{i\,\#}\mu_{n+1} \,\middle|\, e^{n*}_{i\,\#}\nu\right) \\ &= \mathrm{Ent}\left(\mu_n \,\middle|\, e^{n*}_{1\,\#}\mu_n\otimes\cdots\otimes e^{n*}_{d\#}\mu_n\right) \\ &= \mathrm{Ent}\,(\mu_n \,|\, \nu) - \sum_{i=1}^{d}\mathrm{Ent}\left(e^{n*}_{i\,\#}\mu_n \,\middle|\, e^{n*}_{i\,\#}\nu\right).\end{aligned}$$

As the entropy is nonnegative, $\mathrm{Ent}(\mu_n|\nu)$ is nonincreasing, and so converges. But, according to Michel Talagrand's inequality [59],

$$\mathrm{W}_2(e_{i\,\#}^{n*}\mu_n, e_{i\,\#}^{n*}\nu)^2 \;\leq\; C\,\mathrm{Ent}\left(e_{i\,\#}^{n*}\mu_n \,\middle|\, e_{i\,\#}^{n*}\nu\right).$$

Thus,

$$\sum_{i=1}^{d} \mathrm{W}_2(e_{i\,\#}^{n*}\mu_n, e_{i\,\#}^{n*}\nu)^2 \;\leq\; C\left(\mathrm{Ent}\left(\mu_n \,\middle|\, \nu\right) - \mathrm{Ent}\left(\mu_{n+1} \,\middle|\, \nu\right)\right) \;\underset{n\to\infty}{\longrightarrow}\; 0.$$

Then:

1. If $B_n = (e_1^n, \ldots, e_d^n)$ is independent from μ_n, with a uniform law on $\mathrm{O}(d)$, then

$$\mathbb{E}\left[\mathrm{W}_2(e_{k\#}^{n*}\mu_n, e_{k\#}^{n*}\nu)^2\right] \;=\; \mathbb{E}\left[\mathrm{W}_2(e_{k\#}^{n*}\mu_n, e_{k\#}^{n*}\nu)^2 \,\middle|\, \mu_n\right] \;=\; \mathrm{SW}_2(\mu_n, \nu)^2,$$

and thus, by Lebesgue's dominated convergence theorem, $\mathbb{E}[\mathrm{SW}_2(\mu_n, \nu)^2] \to 0$.

2. The sequence $(\mu_n)_{n\geq 1}$ is tight, because for any $n \geq 1$,

$$\int |y|^2 \,\mathrm{d}\mu_n(y) \;=\; \int \sum_{k=1}^{d} \left| t_{e_k^{n-1}}(\langle e_k^{n-1}|x\rangle) \right|^2 \mathrm{d}\mu_{n-1}(x) \;=\; \int |y|^2 \,\mathrm{d}\nu(y).$$

Let μ be a limit point. If all the bases (B_n) form a dense subset of all orthogonal bases of $\mathbb{R}^d$, then for any $\theta \in \mathbb{S}^{d-1}$ we can find an extraction $n_k \to \infty$ such that $e_1^{n_k} \to \theta$, and $\mu_{n_k} \to \mu$ still. Let $\varepsilon > 0$, then for n_k big enough,

$$\mathrm{W}_2\left(e_{1\;\#}^{n_k*}\mu_{n_k}, e_{1\;\#}^{n_k*}\nu\right) \;<\; \varepsilon \qquad \text{and} \qquad \mathrm{W}_2(\mu_{n_k}, \mu) \;<\; \varepsilon,$$

and since e_1^n can be as close to θ as we may desire, we can also impose

$$\mathrm{W}_2\left(e_{1\;\#}^{n_k*}\nu, \theta_\#^*\nu\right) \;<\; \varepsilon \qquad \text{and} \qquad \mathrm{W}_2\left(e_{1\;\#}^{n_k*}\mu, \theta_\#^*\mu\right) \;<\; \varepsilon$$

as well. Because $\mathrm{W}_2\left(e_{1\;\#}^{n_k*}\mu_{n_k}, e_{1\;\#}^{n_k*}\mu\right) < \mathrm{W}_2(\mu_{n_k}, \mu)$, we get

$$\mathrm{W}_2(\theta_\#\mu, \theta_\#\nu) \;<\; 4\varepsilon.$$

Letting $\varepsilon \to 0$, we obtain $\mathrm{SW}_2(\mu, \nu) = 0$. □

5.2.3. If we work only with discrete measures that are sums of N Dirac masses, as

$$\mu = \frac{1}{N} \sum_{k=1}^{N} \delta_{x_k}, \qquad \text{with} \quad x_k \in \mathbb{R}^d,$$

then, setting $\mathbf{x} := (x_1, \dots, x_N)$, we get a vector of $(\mathbb{R}^d)^N$. Letting $\delta_1, \dots, \delta_N$ be the canonical basis of $\mathbb{R}^N$, we can write

$$\mathbf{x} = \sum_{k=1}^{N} x_k \otimes \delta_k.$$

We will write the correspondence between μ and $\mathbf{x}$ as $\mu \sim \mathbf{x}$.

5.2.4. Lemma. *The solution to the Monge–Kantorovich problem between two discrete measures $\mu \sim \mathbf{x}$ and $\nu \sim \mathbf{y}$ is given by a transport map T, such that*

$$T(x_k) = y_{\sigma_T(k)}.$$

for an optimal permutation $\sigma_T \in \mathfrak{S}_N$, such that

$$\mathrm{W}_2(\mu, \nu)^2 = \sum_{k=1}^{N} \left| x_k - y_{\sigma_T(k)} \right|^2 = \min_{\sigma \in \mathfrak{S}_N} \sum_{k=1}^{N} \left| x_k - y_{\sigma(k)} \right|^2.$$

Proof. This follows immediately from Choquet's and Birkhoff's theorems (see Cédric Villani's book [62, p. 5]). □

We will conveniently set $\mathbf{y}_\sigma = (y_{\sigma(1)}, \dots, y_{\sigma(N)})$ for any $\sigma \in \mathfrak{S}_N$, so that, in particular, $\mathrm{W}_2(\mu, \nu)^2 = |\mathbf{x} - \mathbf{y}_{\sigma_T}|^2/2$. Notice $\nu \sim \mathbf{y}_\sigma$ as well.

5.2.5. If μ is the sum of N Dirac masses, then, for any $\theta \in \mathbb{S}^{d-1}$, the projected measure $\theta^*_\# \mu$ is also a sum of Dirac masses:

$$\theta^*_\# \mu = \frac{1}{N} \sum_{k=1}^{N} \delta_{\langle \theta | x_k \rangle},$$

and, with our notation, we can write

$$\theta^*_\# \mu \;\sim\; \sum_{k=1}^{N} \langle \theta | x_k \rangle \delta_k \;=\; \sum_{k=1}^{N} (\theta^* \otimes \mathrm{I}_N)(x_k \otimes \delta_k) \;=\; (\theta^* \otimes \mathrm{I}_N)\mathbf{x}.$$

The IDT algorithm builds a sequence $\mu_n \sim \mathbf{x}^n$, from an initial point $\mu_0 \sim \mathbf{x}^0$ and a reference target measure $\nu \sim \mathbf{y}$, by setting $\mu_{n+1} := T_{n\#}\mu_n$ with

$$T_n \;:=\; \sum_{i=1}^{d} (t_{e_i^n} \circ e_i^{n*}) e_i^n,$$

where $t_{e_i^n}$ is the optimal map between $e_{i\,\#}^{n*}\mu_n$ and $e_{i\,\#}^{n*}\nu$. The basis $e^n = (e_1^n, \ldots, e_d^n)$ changes at each iteration.

5.2.6. Lemma. *Let P_σ denote the permutation matrix associated to a permutation σ, defined by $P_\sigma \delta_k = \delta_{\sigma(k)}$. Then, $\mathbf{y}_\sigma = (\mathrm{I}_d \otimes P_\sigma^{-1})\mathbf{y}$. If σ_θ is the optimal permutation between $\theta^*_\#\mu \sim (\theta^* \otimes \mathrm{I}_N)\mathbf{x}$ and $\theta^*_\#\nu \sim (\theta^* \otimes \mathrm{I}_N)\mathbf{y} \sim (\theta^* \otimes \mathrm{I}_N)\mathbf{y}_{\sigma_\theta}$, we have*

$$\mathrm{W}_2(\theta^*_\#\mu, \theta^*_\#\nu)^2 \;=\; \left|(\theta^* \otimes \mathrm{I}_N)(\mathbf{x} - \mathbf{y}_{\sigma_\theta})\right|^2.$$

Proof. Let $\sigma \in \mathfrak{S}_N$. Then,

$$\mathbf{y}_\sigma = \sum_{k=1}^{N} y_k \otimes \delta_{\sigma^{-1}(k)} = \sum_{k=1}^{N} y_k \otimes P_\sigma^{-1}\delta_k = \sum_{k=1}^{N} (\mathrm{I}_d \otimes P_\sigma^{-1})(y_k \otimes \delta_k) = (\mathrm{I}_d \otimes P_\sigma^{-1})\mathbf{y}.$$

As the optimal map t_θ between $\theta^*_\#\mu$ and $\theta^*_\#\nu$ is given by a permutation $\sigma_\theta \in \mathfrak{S}_N$, such that $\langle \theta | x_k \rangle$ is sent to $t_\theta(\langle \theta | x_k \rangle) = \langle \theta | y_{\sigma_\theta(k)} \rangle$, we have

$$\mathrm{W}_2(\theta^*_\#\mu, \theta^*_\#\nu)^2 \;=\; \sum_{k=1}^{N} \left| \langle \theta | x_k - y_{\sigma_\theta(k)} \rangle \right|^2 \;=\; \left|(\theta^* \otimes \mathrm{I}_N)(\mathbf{x} - \mathbf{y}_{\sigma_\theta})\right|^2. \qquad \square$$

5.2.7. Proposition. *We set, for all $\mathbf{x} \in (\mathbb{R}^d)^N$ and $\sigma \in \mathfrak{S}_N^d$ and any basis e,*

$$\mathcal{F}_e(\mathbf{x}, \sigma) \;=\; \frac{1}{2} \sum_{i=1}^{d} \left|(e_i^* \otimes \mathrm{I}_N)(\mathbf{x} - \mathbf{y}_{\sigma_i})\right|^2.$$

Then, if e^n is the basis which allows us to define μ^{n+1} from μ^n, we have

$$\mathcal{F}_{e^n}(\mathbf{x}^n, \sigma_{e^n}) = \min_{\sigma \in \mathfrak{S}_N^d} \mathcal{F}_{e^n}(\mathbf{x}^n, \sigma) = \frac{1}{2} \sum_{i=1}^{d} \mathrm{W}_2(e_i^{n*}{}_{\#}\mu_n, e_i^{n*}{}_{\#}\nu)^2.$$

If σ_{e^n} represent the sequence of optimal permutations, then

$$\mathbf{x}^{n+1} = \mathbf{x}^n - \nabla_{\mathbf{x}} \mathcal{F}_{e^n}(\mathbf{x}^n, \sigma_{e^n}) \qquad \textit{and} \qquad \frac{1}{2}\left|\mathbf{x}^{n+1} - \mathbf{x}^n\right|^2 = \mathcal{F}_{e^n}(\mathbf{x}^n, \sigma_{e^n}).$$

Thus, the IDT algorithm can be seen as a kind of steepest-descent method. If it were not to depend on an ever-changing basis e^n, it would be (close to) an explicit Euler scheme for a gradient flow.

Proof. On the one hand, $\mathbf{x}^{n+1}$ is defined by

$$x_k^{n+1} := T(x_k^n) = \sum_{i=1}^{d} t_{e_i^n}(\langle e_i^n | x_k^n \rangle)\, e_i^n = \sum_{i=1}^{d} \langle e_i^n | y_{\sigma_{e_i^n}(k)} \rangle e_i^n = \sum_{i=1}^{d} e_i^n e_i^{n*} y_{\sigma_{e_i^n}(k)},$$

where $\sigma_{e_i^n}$ is the optimal permutation associated to $t_{e_i^n}$. Thus,

$$\begin{aligned}
\mathbf{x}^{n+1} &= \sum_{k=1}^{N} \sum_{i=1}^{d} (e_i^n e_i^{n*} y_{\sigma_{e_i^n}(k)}) \otimes \delta_k = \sum_{i=1}^{d} \sum_{k=1}^{N} (e_i^n e_i^{n*} \otimes \mathrm{I}_N)(y_{\sigma_{e_i^n}(k)} \otimes \delta_k) \\
&= \sum_{i=1}^{d} (e_i^n e_i^{n*} \otimes \mathrm{I}_N)(\mathrm{I}_d \otimes P_{\sigma_{e_i^n}}^{-1})\mathbf{y} = \sum_{i=1}^{d} (e_i^n e_i^{n*} \otimes P_{\sigma_{e_i^n}}^{-1})\mathbf{y}.
\end{aligned}$$

On the other hand,

$$\begin{aligned}
\nabla_{\mathbf{x}} \mathcal{F}_e(\mathbf{x}, \sigma) &= \sum_{i=1}^{d} (e_i \otimes \mathrm{I}_N)^* (e_i^* \otimes \mathrm{I}_N)\, (\mathbf{x} - \mathbf{y}_{\sigma_i}) \\
&= \sum_{i=1}^{d} (e_i e_i^* \otimes \mathrm{I}_N)\, (\mathbf{x} - \mathbf{y}_{\sigma_i}) \\
&= \left(\mathbf{x} - \sum_{i=1}^{d} (e_i e_i^* \otimes \mathrm{I}_N)(\mathrm{I}_d \otimes P_{\sigma_i}^{-1})\mathbf{y} \right)
\end{aligned}$$

$$= \left(\mathbf{x} - \sum_{i=1}^{d}(e_i e_i^* \otimes P_{\sigma_i}^{-1})\mathbf{y}\right).$$

Thus, $\nabla_{\mathbf{x}}\mathcal{F}_{e^n}(\mathbf{x}^n, \sigma_{e^n}) = (\mathbf{x}^n - \mathbf{x}^{n+1})$. Moreover,

$$\begin{aligned}\frac{1}{2}\left|\mathbf{x}^n - \mathbf{x}^{n+1}\right|^2 &= \frac{1}{2}\sum_{i=1}^{d}\left|(e_i e_i^* \otimes \mathrm{I}_N)\left(\mathbf{x}^n - \mathbf{y}_{\sigma_{e_i^n}}\right)\right|^2 \\ &= \frac{1}{2}\sum_{i=1}^{d}\left|e_i^* \otimes \mathrm{I}_N\left(\mathbf{x}^n - \mathbf{y}_{\sigma_{e_i^n}}\right)\right|^2 \\ &= \mathcal{F}_{e^n}(\mathbf{x}^n, \sigma_{e^n}).\end{aligned}$$

□

5.3 Marc Bernot's isotropic definition

5.3.1. To remove the dependence vis-à-vis the bases e^n, Marc Bernot suggested to replace $\mathcal{F}_e$ with

$$\mathcal{F}(\mathbf{x}, \boldsymbol{\sigma}) := \frac{1}{2}\fint_{\mathbb{S}^{d-1}}\left|(\theta^* \otimes \mathrm{I}_N)(\mathbf{x} - \mathbf{y}_{\boldsymbol{\sigma}_\theta})\right|^2 \mathrm{d}\theta,$$

defined for $\mathbf{x} \in (\mathbb{R}^d)^N$ and $\boldsymbol{\sigma} : \mathbb{S}^{d-1} \to \mathfrak{S}_N$. In other words,

$$\mathcal{F}(\mathbf{x}, \boldsymbol{\sigma}) = \frac{1}{2d}\fint_{\mathrm{O}(d)}\mathcal{F}_e(\mathbf{x}, (\boldsymbol{\sigma}_{e_1}, \ldots, \boldsymbol{\sigma}_{e_d}))\,\mathrm{d}e.$$

Then, if $\mu \sim \mathbf{x}$ and $\nu \sim \mathbf{y}$,

$$\min_{\boldsymbol{\sigma}}\mathcal{F}(\mathbf{x}, \boldsymbol{\sigma}) = \frac{1}{2}\mathrm{SW}_2(\mu, \nu)^2.$$

We can introduce a parameter $h > 0$, and define a sequence $(\mathbf{x}^n)$ by

$$\mathbf{x}^{n+1} := \mathbf{x}^n - h\nabla_{\mathbf{x}}\mathcal{F}(\mathbf{x}^n, \boldsymbol{\sigma}^n), \tag{5.3.1.a}$$

where $\boldsymbol{\sigma}_\theta^n$ is the optimal permutation between $\theta_\#^*\mu_n$ and $\theta_\#^*\nu$, such that

$$\frac{1}{2}\mathrm{SW}_2(\mu, \nu) = \mathcal{F}(\mathbf{x}^n, \boldsymbol{\sigma}^n).$$

5.3.2. Lemma. *We have, for any* $\mathbf{h} \in (\mathbb{R}^d)^N$,

$$\mathcal{F}(\mathbf{x}^n + \mathbf{h}, \boldsymbol{\sigma}^n) = \mathcal{F}(\mathrm{x}^n, \boldsymbol{\sigma}^n) + \fint_{\mathbb{S}^{d-1}} \left\langle (\theta\theta^* \otimes \mathrm{I}_N)(\mathbf{x}^n - \mathbf{y}_{\sigma_\theta}) \,\middle|\, \mathbf{h} \right\rangle \mathrm{d}\theta + \frac{1}{2d}|\mathbf{h}|^2.$$

Therefore

$$\mathbf{x}^{n+1} = \left(1 - \frac{h}{d}\right)\mathbf{x}^n + h \fint_{\mathbb{S}^{d-1}} (\theta\theta^* \otimes P^{-1}_{\sigma^n_\theta})\mathbf{y},$$

and $\mathrm{SW}_2(\mu_n, \nu)$ *is nonincreasing if* $h < 2d$.

Proof. The expression for $\mathbf{x}^{n+1}$ comes from $\fint \theta\theta^* \,\mathrm{d}\theta = \mathrm{I}_d / d$, and

$$\begin{aligned}
\nabla_{\mathrm{x}} \mathcal{F}(\mathbf{x}^n, \boldsymbol{\sigma}^n) &= \fint (\theta\theta^* \otimes \mathrm{I}_N)\left(\mathbf{x}^n - \mathbf{y}_{\sigma^n_\theta}\right) \mathrm{d}\theta \\
&= \frac{1}{d}\mathbf{x}^n - \fint (\theta\theta^* \otimes \mathrm{I}_N)(\mathrm{I}_d \otimes P^{-1}_{\sigma_\theta})\mathrm{y}\,\mathrm{d}\theta \\
&= \frac{1}{d}\mathbf{x}^n - \fint (\theta\theta^* \otimes P^{-1}_{\sigma_\theta})\mathrm{y}\,\mathrm{d}\theta.
\end{aligned}$$

As for the nonincreasingness of $\mathrm{SW}_2(\mu_n, \nu)$,

$$\begin{aligned}
\frac{1}{2}\mathrm{SW}_2(\mu_{n+1}, \nu)^2 &= \min_{\sigma} \mathcal{F}(\mathbf{x}^{n+1}, \boldsymbol{\sigma}) \\
&\leq \mathcal{F}(\mathbf{x}^n - h\nabla_{\mathrm{x}}\mathcal{F}(\mathbf{x}^n, \boldsymbol{\sigma}^n), \boldsymbol{\sigma}^n) \\
&\leq \mathcal{F}(\mathrm{x}^n, \boldsymbol{\sigma}^n) - h\left|\nabla_{\mathrm{x}}\mathcal{F}(\mathrm{x}^n, \boldsymbol{\sigma}^n)\right|^2 + \frac{h^2}{2d}\left|\nabla_{\mathrm{x}}\mathcal{F}(\mathrm{x}^n, \boldsymbol{\sigma}^n)\right|^2. \qquad \square
\end{aligned}$$

5.4 Implicit version

5.4.1. Equation (5.3.1.a) defines an explicit Euler scheme for the sliced Wasserstein distance. On $\mathbb{R}^d$, given a smooth functional $F : \mathbb{R}^d \to \mathbb{R}$, the explicit Euler scheme yields a sequence $(x_n)_{n\in\mathbb{N}}$, given a starting point x_0 and a time step $h > 0$, by setting

$$x_{n+1} := x_n - h\nabla F(x_n).$$

The *implicit* Euler scheme, on the other hand,

$$x_{n+1} := x_n - h\nabla F(x_{n+1}),$$

can be obtained by, at each step, taking

$$x_{n+1} \in \arg\min_{x} \left\{ \frac{1}{2h} |x - x_n|^2 + F(x) \right\}.$$

In our case, we can define a sequence $\mu_{\mathbf{n}} \sim \mathbf{x}^n$ using such an implicit scheme, by taking

$$\mathbf{x}^{n+1} \in \arg\min_{\mathbf{x}^n} \left\{ \frac{1}{2h} \left| \mathbf{x} - \mathbf{x}^n \right|^2 + \min_{\boldsymbol{\sigma}} \mathcal{F}(\mathbf{x}, \boldsymbol{\sigma}) \right\}.$$

This corresponds to ours setting

$$\mu_{n+1} \in \arg\min_{\mu} \left\{ \frac{1}{2h} \mathrm{W}_2(\mu, \mu_n)^2 + \frac{1}{2} \mathrm{SW}_2(\mu, \nu)^2 \right\}. \tag{5.4.1.a}$$

5.4.2. One of the difficulties of working discrete measures, is that the (sliced) Wasserstein distance is given by an optimal map—or many, for the sliced distance—, but a bijection on a discrete space can only be a permutation. It is hard to find any smoothness of the optimal map with respect to the measures under such circumstances. Things are simpler when the measures are absolutely continuous, as there is some regularity (see the article by Grégoire Loeper [39]). Furthermore, (5.4.1.a) does not lose any meaning if we drop the assumption the measures are all discrete—that is even the starting point of the theory of gradient flows in the space of probability measures, as developed by Luigi Ambrosio, Nicola Gigli, and Giuseppe Savaré [3].

In the rest of this section, we will therefore show that, given two absolutely continuous measures μ_0 and ν, we can define a sequence $(\mu_n)_{n \in \mathbb{N}}$ with

$$\mu_{n+1} \in \arg\min_{\mu} \left\{ \frac{1}{2h} \mathrm{W}_2(\mu, \mu_n)^2 + F(\mu) \right\},$$

where $h > 0$ is a time step and

$$F(\mu) := \frac{1}{2} \mathrm{SW}_2(\mu, \nu)^2.$$

We will work on the closed unit ball $B = \overline{B}(0, 1)$, and assume ν has a strictly positive, smooth density on B. As the algorithm may force μ_n to venture out of B, we will allow it to be defined on rB, with $r > \sqrt{d}$.

5.4.3. Lemma. *Let us fix a time step $h > 0$, and a radius $r > \sqrt{d}$. For a probability measure μ_0 on $rB = \overline{B}(0, r)$ that is absolutely continuous with a strictly positive, smooth density ρ_0, there is a probability measure μ on rB minimizing*

$$\mathcal{G}(\mu) := F(\mu) + \frac{1}{2h}\mathrm{W}_2(\mu, \mu_0)^2 + \delta H(\mu),$$
$$\textit{with} \quad H(\mu) := \begin{cases} \int_{rB} \rho(x)\ln\rho(x)\,\mathrm{d}x & \textit{if } \mathrm{d}\mu(x) = \rho(x)\,\mathrm{d}x, \\ +\infty & \textit{otherwise.} \end{cases}$$

Moreover, this optimal μ has a Lipschitz density ρ, which is strictly positive on rB and such that

$$\|\rho\|_{\mathrm{L}^\infty} \leq \left(1 + h/\sqrt{d}\right)^d \|\rho_0\|_{\mathrm{L}^\infty}.$$

Proof. We follow methods developed by Guillaume Carlier and Filippo Santambrogio [20], and Giuseppe Buttazzo and Filippo Santambrogio [15].

It is well known the entropy H is lower semicontinuous for the Wasserstein distance (see, for instance, the article by Richard Jordan, David Kinderlehrer, and Felix Otto [34, Proposition 41]). Therefore, if $(\mu_n)_{n\in\mathbb{N}}$ is a minimizing sequence in $\mathcal{P}(rB)$, then, up to an extraction, it converges toward a minimizer μ, which must necessarily have a density ρ.

We denote by ψ_θ the Kantorovich potential between $\theta^*_\#\mu$ and $\theta^*_\#\nu$, and φ the Kantorovich potential between μ and μ_0.

Let $\bar{\mu}$ be another probability measure on rB, absolutely continuous with a density $\bar{\rho}$. Then, proposition 1.5.6 on page 36 and proposition 5.1.6 on page 124 together yield

$$\limsup_{\varepsilon\to 0^+} \frac{H(\mu) - H((1-\varepsilon)\mu + \varepsilon\bar{\mu})}{\varepsilon} \leq \frac{1}{\delta}\int_{rB} \Psi(x)(\bar{\rho}(x) - \rho(x))\,\mathrm{d}x$$
$$\text{where} \quad \Psi(x) := \fint_{\mathbb{S}^{d-1}} \psi_\theta(\langle\theta|x\rangle)\,\mathrm{d}\theta + \frac{1}{h}\varphi(x).$$

Since $t \mapsto t\ln t$ is convex, setting $\rho_\varepsilon = (1-\varepsilon)\rho + \varepsilon\bar{\rho}$, we can write

$$\rho\ln\rho - \rho_\varepsilon\ln\rho_\varepsilon \geq \varepsilon(1 + \ln\rho_\varepsilon)(\rho - \bar{\rho}).$$

If $\rho(x) \geq \bar{\rho}(x)$, then $\rho_\varepsilon(x) \geq \bar{\rho}(x)$, and thus

$$\rho \ln \rho - \rho_\varepsilon \ln \rho_\varepsilon \;\geq\; \varepsilon(1 + \ln \bar{\rho})(\rho - \bar{\rho}).$$

Where $\rho(x) \leq \bar{\rho}(x)$, this last inequality still holds, because then $\ln \rho_\varepsilon(x) < \ln \bar{\rho}(x)$, and $\rho(x) - \bar{\rho}(x) < 0$.

In particular, if we take $\bar{\mu}$ uniform on rB, i.e. $\bar{\rho}(x) = 1/(r^d|B|)$, then,

$$\begin{aligned} \rho \ln \rho - \rho_\varepsilon \ln \rho_\varepsilon \;&\geq\; \varepsilon|1 + \ln \bar{\rho}|(\rho + \bar{\rho}) && \text{when } \rho > 0, \\ \rho \ln \rho - \rho_\varepsilon \ln \rho_\varepsilon \;&\geq\; -(1 + \ln(\bar{\varepsilon}\rho))\varepsilon\bar{\rho} && \text{when } \rho = 0. \end{aligned}$$

Integrating, since $\bar{\rho}$ is constant we get

$$\frac{H(\mu) - H((1-\varepsilon)\mu + \varepsilon\bar{\mu})}{\varepsilon} \;\geq\; -2|1 + \ln \bar{\rho}| - (1 + \ln(\varepsilon\bar{\rho}))\frac{|\{\rho = 0\}|}{r^d|B|}.$$

As we have an upper bound when $\varepsilon \to 0$, necessarily $|\{\rho = 0\}| = 0$, i.e. $\rho > 0$ almost everywhere.

Now, let $\bar{\rho} = \eta\rho$ with $\eta \in \mathrm{L}^\infty$. Then,

$$\ln(\rho + \varepsilon(\bar{\rho} - \rho)) = \ln((1 + \varepsilon(\eta - 1))) + \ln(\rho).$$

Therefore, thanks to Lebesgue's dominated convergence theorem,

$$\begin{aligned} \int_{rB} (1 + \ln \rho)(\rho - \bar{\rho}) \;&=\; \lim_{\varepsilon \to 0} \int (1 + \ln \rho_\varepsilon)(\rho - \bar{\rho}) \\ &\leq\; \limsup_{\varepsilon \to 0} \frac{H(\mu) - H((1-\varepsilon)\mu + \varepsilon\bar{\mu})}{\varepsilon} \\ &\leq\; \frac{1}{\delta} \int \Psi(\bar{\rho} - \rho), \end{aligned}$$

and this yields

$$\int [\Psi(x) + \delta \ln \rho(x)]\, \bar{\rho}(x)\, \mathrm{d}x \;\geq\; \int [\Psi(x) + \delta \ln \rho(x)]\, \rho(x)\, \mathrm{d}x.$$

We set $m = \operatorname{ess\,inf}\{\Psi + \delta \ln \rho\}$. For any $m' > m$, by definition, $A := \{m' > \Psi + \delta \ln \rho\}$ has a nonzero measure, so we can take $\eta = \lambda \mathbb{1}_A$ with λ such that $\eta\rho$ is still a probability measure. Then, the previous inequality gives

$$m' \geq \int [\Psi + \delta \ln \rho]\, \eta\rho \geq \int [\Psi + \delta \ln \rho]\, \rho \geq m.$$

Letting m' converge toward m, we get $\Psi + \delta \ln \rho$ is constant, and equal to m almost everywhere. This implies

$$\rho = \exp((m - \Psi)/\delta).$$

As Ψ is Lipschitz, so is ρ. It then follows from theorem 1.3.5 on page 31 that the potential φ between μ and μ_0 is $\mathscr{C}^2$, and $\mathrm{I}_d - \mathrm{D}^2\varphi > 0$ and $1 - \psi_\theta'' > 0$.

Let us denote by f_θ and g_θ the densities of $\theta^*_\# \mu$ and $\theta^*_\# \nu$, and F_θ and G_θ their cumulative distributions. If m_f and m_g stand for the minima of f and g, and M_f and M_g for their maxima, then

$$\begin{aligned} m_f \sqrt{1 - t^2} &\leq f_\theta(t) \leq M_f \sqrt{1 - t^2}, \\ m_g \sqrt{1 - s^2} &\leq g_\theta(s) \leq M_g \sqrt{1 - s^2}. \end{aligned}$$

Let F_θ and G_θ be the cumulative distributions of f_θ and g_θ. If we define

$$\begin{aligned} U_\varepsilon &:= \left\{ x \in rB \;\middle|\; \forall\, \theta \in \mathbb{S}^{d-1}, F_\theta(\langle \theta | x \rangle) \in (\varepsilon, 1 - \varepsilon) \right\}, \\ V_\varepsilon &:= \left\{ y \in B \;\middle|\; \forall\, \theta \in \mathbb{S}^{d-1}, G_\theta(\langle \theta | y \rangle) \in (\varepsilon, 1 - \varepsilon) \right\}, \end{aligned}$$

then f_θ and g_θ are uniformly bounded and bounded above on $\theta^*(U_\varepsilon)$ and $\theta^*(V_\varepsilon)$ respectively. Moreover, it follows from the definition of the optimal map, $t_\theta := G_\theta^{-1} \circ F_\theta$, that $t_\theta(\langle \theta | x \rangle) \in \theta^*(V_\varepsilon)$ for any $x \in U_\varepsilon$. Then, thanks to theorem 1.3.5 on page 31 again, we get that $\psi_\theta \circ \theta^*$ is $\mathscr{C}^2$ on U_ε, and $1 - \psi_\theta'' > 0$. Since $t_\theta = \mathrm{Id} - \psi_\theta' = G_\theta^{-1} \circ F_\theta$, we also have $\psi_\theta \circ \theta^*$ is $\mathscr{C}^{1,\alpha}$ on rB, up to the boundary. By a consequence of Lebesgue's dominated convergence theorem, $\Psi = \varphi/h ⨍ \psi_\theta \circ \theta^* \, \mathrm{d}\theta$ is $\mathscr{C}^2$ on U_ε, and $\mathscr{C}^{1,\alpha}$ up to the boundary. Moreover,

$$\nabla\Psi(x) = \frac{\nabla\varphi}{h} + ⨍ \psi_\theta'(\langle \theta | x \rangle)\theta \, \mathrm{d}\theta,$$

$$\nabla^2\Psi(x) \;=\; \frac{\nabla^2\varphi}{h} + \fint \psi_\theta''(\langle\theta|x\rangle)\theta\otimes\theta\,\mathrm{d}\theta.$$

Therefore, $\rho = \exp((m-\Psi)/\delta)$ is $\mathscr{C}^2$ in the interior, $\mathscr{C}^{1,\alpha}$ up to the boundary, and

$$\nabla\rho \;=\; -\rho\frac{\nabla\Psi}{\delta} \qquad \text{and} \qquad \nabla^2\rho \;=\; \rho\frac{\nabla\Psi\otimes\nabla\Psi}{\delta^2} - \rho\frac{\nabla^2\Psi}{\delta}.$$

If ρ is maximum in x_0 on the boundary, i.e. for $|x_0| = r$, then, as $t\mapsto\rho(tx_0)$ is maximal for $t = 1$, we must have $\langle\nabla\rho(x_0)|x_0\rangle \geq 0$. Thus,

$$\langle\nabla\Psi(x_0)|x_0\rangle \leq 0 \tag{5.4.3.a}$$

But, the transport map $\mathrm{Id}-\nabla\varphi$ between μ and μ_0 takes its values in $\operatorname{supp}\mu_0 = rB$, so

$$\begin{aligned} r^2 \;&\geq\; |x_0 - \nabla\varphi(x_0)|^2 \\ &\geq\; |x_0|^2 + |\nabla\varphi(x_0)|^2 - 2\langle\nabla\varphi(x_0)|x_0\rangle \\ &\geq\; r^2 - 2\langle\nabla\varphi(x_0)|x_0\rangle. \end{aligned}$$

Hence, $\langle\nabla\varphi(x_0)|x_0\rangle \geq 0$. Likewise, for any direction θ, the map $t_\theta = \mathrm{Id}-\psi_\theta'$ takes its values in $\theta^*(B) = [-1,1]$, so

$$\begin{aligned} \left\langle \fint \psi_\theta'(\langle\theta|x_0\rangle)\theta\,\mathrm{d}\theta \,\middle|\, x_0 \right\rangle \;&=\; \fint \psi_\theta'(\langle\theta|x_0\rangle)\langle\theta|x_0\rangle\,\mathrm{d}\theta \\ &\geq\; \frac{1}{2}\fint |\langle\theta|x_0\rangle|^2 - |\langle\theta|x_0\rangle - \psi_\theta'(\langle\theta|x_0\rangle)|^2\,\mathrm{d}\theta \\ &\geq\; \frac{1}{2}\left(\fint |\langle\theta|x_0\rangle|^2\,\mathrm{d}\theta - 1\right) \\ &\geq\; \frac{1}{2}\left(\frac{r^2}{d} - 1\right). \end{aligned}$$

As we have assumed $r > \sqrt{d}$, we finally get $\langle\nabla\Psi(x_0)|x_0\rangle > 0$, and this contradicts (5.4.3.a). Thus, ρ is maximum in a point x_0 in the interior. Since $\nabla\rho(x_0) = 0$ and $\nabla^2\rho(x_0) \leq 0$, we must have $\nabla^2\Psi(x_0) \geq 0$. Hence, as $\psi_\theta'' < 1$,

$$\nabla^2\varphi(x_0) \;\geq\; -h\fint \psi_\theta''(\langle\theta|x_0\rangle)\theta\otimes\theta\,\mathrm{d}\theta$$

$$\geq \; -h \fint \theta \otimes \theta \,\mathrm{d}\theta$$
$$\geq \; -\frac{h}{\sqrt{d}} \mathrm{I}_d \,.$$

This, in turn, yields:

$$\|\rho\|_\infty = \rho(x_0) \;=\; \rho_0(x_0 - \nabla\varphi(x_0)) \det\left(\mathrm{I}_d - \nabla^2\varphi(x)\right) \;\leq\; \left(1 + \frac{h}{\sqrt{d}}\right)^d \|\rho_0\|_\infty, \qquad \square$$

5.4.4. Proposition. *For any time step $h > 0$, and any probability measure $\mu_0 \in \mathcal{P}(rB)$ that is absolutely continuous with a density $\rho_0 \in \mathrm{L}^\infty$, there is $\mu \in \mathcal{P}(rB)$ minimizing*

$$F(\mu) + \frac{1}{2h}\mathrm{W}_2(\mu, \mu_0)^2,$$

which is absolutely continuous, with a density $\rho \in \mathrm{L}^\infty$ such that

$$\|\rho\|_{\mathrm{L}^\infty} \leq \left(1 + h/\sqrt{d}\right)^d \|\rho_0\|_{\mathrm{L}^\infty}.$$

Proof. Let us first assume $\rho_0 \in \mathscr{C}^\infty(rB)$. Then, according to lemma 5.4.3 on page 136, for any $\delta > 0$, there is μ_δ minimizing

$$\mu \;\mapsto\; F(\mu) + \frac{1}{2h}\mathrm{W}_2(\mu, \mu_0)^2 + \delta H(\mu),$$

with a Lipschitz density ρ_δ such that

$$\|\rho_\delta\|_{\mathrm{L}^\infty} \leq \left(1 + h/\sqrt{d}\right)^d \|\rho_0\|_{\mathrm{L}^\infty}.$$

Up to an extraction, we can assume μ_δ converges toward μ in $\mathcal{P}(rB)$ and ρ_δ converges toward ρ for the weak-star topology of L^∞, with ρ the density of μ. Then, $\|\rho\|_{\mathrm{L}^\infty} \leq (1 + h/\sqrt{d})^d \|\rho_0\|_{\mathrm{L}^\infty}$, and this implies

$$H(\mu) \;\leq\; \left[1 + \left(1 + h/\sqrt{d}\right)^d \|\rho_0\|_{\mathrm{L}^\infty}\right] \ln\left[1 + \left(1 + h/\sqrt{d}\right)^d \|\rho_0\|_{\mathrm{L}^\infty}\right] \;<\; \infty,$$

because $t \mapsto t \ln t$ is increasing on $(1/e, \infty)$ and positive on $(1, \infty)$.

Let $\bar{\mu}$ be such that $F(\bar{\mu}) + \mathrm{W}_2(\bar{\mu}, \mu_0)^2/(2h)$ is minimal, and let e_t be the heat kernel, $e_t(x) = \exp(-\pi|x|^2/t)/\sqrt{t}$. We set $\bar{\mu}_t = e_t * \bar{\mu}$. Then, $\bar{\mu}_t \to \bar{\mu}$ in $\mathcal{P}(rB)$, and if $H(\bar{\mu}) = \infty$, then $H(\bar{\mu}_t) \to \infty$ as well. Let t_δ be such that $H(\bar{\mu}_{t_\delta}) < 1/\sqrt{\delta}$; then,

$$\begin{aligned} F(\mu) + \frac{1}{2h}\mathrm{W}_2(\mu, \mu_0)^2 &\leq \liminf_{\delta \to 0} \left\{ F(\mu_\delta) + \frac{1}{2h}\mathrm{W}_2(\mu_\delta, \mu_0)^2 + \delta H(\mu_\delta) \right\} \\ &\leq \liminf_{\delta \to 0} \left\{ F(\bar{\mu}_{t_\delta}) + \frac{1}{2h}\mathrm{W}_2(\bar{\mu}_{t_\delta}, \mu_0)^2 + \delta H(\bar{\mu}_{t_\delta}) \right\} \\ &\leq F(\bar{\mu}) + \frac{1}{2h}\mathrm{W}_2(\bar{\mu}, \mu_0)^2. \end{aligned}$$

Thus, μ is a minimizer as well.

We now drop the assumption $\rho_0 \in \mathscr{C}^\infty$. Then, for any $t > 0$, there is a minimizer $\mu_t \in \mathcal{P}(rB)$ for $\mu \mapsto F(\mu) + \mathrm{W}_2(\mu, \mu_t)^2/(2h)$, which has a density $\rho_t \in \mathrm{L}^\infty$ with $\|\rho_t\|_{\mathrm{L}^\infty} \leq (1 + h/\sqrt{d})^d \|\rho_0\|_{\mathrm{L}^\infty}$. Up to an extraction, μ_t converges toward μ in $\mathcal{P}(rB)$ and ρ_t converges toward ρ for the weak-star topology of L^∞. This implies

$$\|\rho\|_{\mathrm{L}^\infty} \leq \left(1 + h/\sqrt{d}\right)^d \|\rho_0\|_{\mathrm{L}^\infty}.$$

And if $\bar{\mu}$ is a minimizer for $\mu \mapsto F(\mu) + \mathrm{W}_2(\mu, \mu_0)^2/(2h)$, then

$$\begin{aligned} F(\mu) + \frac{1}{2h}\mathrm{W}_2(\mu, \mu_0)^2 &\leq \liminf_{\delta \to 0} \left\{ F(\mu_t) + \frac{1}{2h}\mathrm{W}_2(\mu_t, e_t * \mu_0)^2 + \delta H(\mu_t) \right\} \\ &\leq \liminf_{\delta \to 0} \left\{ F(\bar{\mu}) + \frac{1}{2h}\mathrm{W}_2(\bar{\mu}, e_t * \mu_0)^2 + \delta H(\bar{\mu}) \right\} \\ &\leq F(\bar{\mu}) + \frac{1}{2h}\mathrm{W}_2(\bar{\mu}, \mu_0)^2. \end{aligned}$$

So μ is a minimizer as well. □

5.5 Continous version

5.5.1. Generalized minimizing movements. Given a metric space X, a functional $\mathcal{F} : [0, \infty) \times \mathbb{N} \times \mathrm{X} \times \mathrm{X} \to [-\infty, \infty]$, and an initial point $x_0 \in \mathrm{X}$, a minimizing movement (MM) relative to $\mathcal{F}$ and starting from x_0 is a curve $x : [0, \infty) \to \mathrm{X}$ that is pointwise limit of a family $x_h : [0, \infty) \to \mathrm{X}$ indexed by $h > 0$ such that:

- $x_h(0) = x_0$ for every $h > 0$;

- x_h is constant on each interval $[nh, (n+1)h)$, so $x(t) = x(nh)$ for $n = \lfloor t/h \rfloor$;
- $x_h(t+h)$ minimizes $y \mapsto \mathcal{F}(h, n, y, x(t))$, if $n = \lfloor t/h \rfloor$.

When x is limit of only a sequence x_{h_k}, with $h_k \to 0$, then x is called a generalized minimizing movement (GMM).

5.5.2. The concept of minimizing movements was introduced by Ennio De Giorgi [22], and developed furthermore by Luigi Ambrosio [1]. It is a fundamental tool for the theory of gradient flows in metric spaces, as developed by the latter with Nicola Gigli and Giuseppe Savaré [3, 57].

Indeed, a gradient flow $\dot{x} = -\nabla F(x)$ in $\mathbb{R}^d$ is the limit of the Euler implicit scheme: if $x_h(t+h) = x_h(t) - h\nabla F(x(t))$, with x_h constant on each interval $[nh, (n+1)h)$, then $x_h(t) \to x(t)$; and $x_h(t+h)$ is just obtained from $x_h(t)$ as a minimizer of

$$y \mapsto \mathcal{F}(h, n, y, x_h(t)) \qquad \text{with} \quad \mathcal{F}(h, n, y, x) = \frac{1}{2h}|y - x|^2 + F(y).$$

Thus, a gradient flow in $\mathbb{R}^d$ is a minimizing movement. But, unlike differentiation, minimization can be performed in quite a general framework, as in a metric space. There it is enough to replace the Euclidean distance with the metric distance in the previous expression of $\mathcal{F}$.

5.5.3. Theorem. *Let ν be a probability measure on $B = \overline{B}(0,1)$, with a strictly positive, smooth density. Given an absolutely continuous measure $\mu_0 \in \mathcal{P}(rB)$, with a density $\rho_0 \in \mathrm{L}^p$, there is a Lipschitz generalized minimizing movement $(\mu_t)_{t \geq 0}$ in $\mathcal{P}(rB)$ starting from μ_0 for the functional*

$$\mathcal{F}(h, n, \mu_+, \mu_-) := \frac{1}{2h}\mathrm{W}_2(\mu_+, \mu_-)^2 + F(\mu_+), \qquad \textit{with} \quad F(\mu_+) = \frac{1}{2}\mathrm{SW}_2(\mu_+, \nu)^2.$$

Moreover, for each time $t \geq 0$, the measure μ_t has a density $\rho_t \in \mathrm{L}^p$, and

$$\|\rho_t\|_{\mathrm{L}^p} \leq e^{t\sqrt{d}/q}\|\rho_0\|_{\mathrm{L}^p}.$$

Proof. For any time step $h > 0$, we use proposition 5.4.4 on page 140 to build a curve $(\mu_t^h)_{t \geq 0}$ of absolutely continuous measures by induction, such that:

- $\mu_0^h = \mu_0$;

- μ_t^h is constant on $[nh, (n+1)h)$;
- μ_{t+h}^h minimizes $\mu \mapsto F(\mu) + \mathrm{W}_2(\mu, \mu_t^h)^2/(2h)$;
- denoting by ρ_t^h the density of μ_t^h, and $n = \lfloor t/h \rfloor$,

$$\|\rho_t^h\|_{\mathrm{L}^p} \leq \left(1 + h/\sqrt{d}\right)^{nd/q} \|\rho_0\|_{\mathrm{L}^p}. \tag{5.5.3.a}$$

For an arbitrary $T > 0$, we define a measure μ^h on $[0,T] \times rB$ with

$$\forall\, \xi \in \mathscr{C}^0, \quad \int_{[0,T]\times rB} \xi(t,x)\,\mathrm{d}\mu^h(t,x) \;=\; \int_{[0,T]} \int_{rB} \xi(t,x)\,\mathrm{d}\mu_t^h(x)\,\mathrm{d}t;$$

and it has a density $\rho^h \in \mathrm{L}^p([0,T] \times rB)$, defined by $\rho^h(t,x) = \rho_t^h(x)$. Then, there is $h_n \to 0$ such that $\mu^{h_n} \to \mu$ in $\mathcal{P}(rB)$; moreover, the limit μ has necessarily a density ρ, and ρ^h weakly converges to ρ in L^r for any finite $r \in [1,p]$, with a weak-star convergence in L^∞ when p is infinite.

Let $\rho_t(x) := \rho(t,x)$. We want to show μ_t^h converges $\mathcal{P}(rB)$, and ρ_t^h weakly converges to ρ_t in L^r for all $t \in [0,T]$ and for any finite $r \in [1,p]$ (even though we might have to redefine μ_t and ρ_t on a negligible set of times t). First, there is a measure μ_t whose density is ρ_t, at least for almost any t, since

$$1 \;=\; \frac{1}{2\delta} \int_{t-\delta}^{t+\delta} \int \rho_t^h(x)\,\mathrm{d}x\,\mathrm{d}t \;\underset{h\to 0}{\longrightarrow}\; \int_{t-\delta}^{t+\delta} \int \rho_t(x)\,\mathrm{d}x\,\mathrm{d}t,$$

and this implies ρ_t is indeed a probability density. Next, we show μ_t^h must converge to μ_t. If $\xi \in \mathscr{C}^1(rB)$,

$$\begin{aligned} &\left| \int \xi\,\mathrm{d}\mu_t^{h_n} - \int \xi\,\mathrm{d}\mu_t^{h_m} \right| \\ &\quad \leq \left| \int \xi\,\mathrm{d}\mu_t^{h_n} - \frac{1}{2\delta}\int_{t-\delta}^{t+\delta}\int \xi\,\mathrm{d}\mu_s^{h_n}\,\mathrm{d}s \right| + \left| \int \xi\,\mathrm{d}\mu_t^{h_m} - \frac{1}{2\delta}\int_{t-\delta}^{t+\delta}\int \xi\,\mathrm{d}\mu_s^{h_m}\,\mathrm{d}s \right| \\ &\qquad + \left| \frac{1}{2\delta}\int_{t-\delta}^{t+\delta}\int \xi\,\mathrm{d}\mu_s^{h_n}\,\mathrm{d}s - \frac{1}{2\delta}\int_{t-\delta}^{t+\delta}\int \xi\,\mathrm{d}\mu_s^{h_m}\,\mathrm{d}s \right|. \end{aligned} \tag{5.5.3.b}$$

But, by taking $\gamma \in \Gamma_o(\mu_t^h, \mu_s^h)$, we can first obtain

$$\left|\int \xi \,\mathrm{d}\mu_t^h - \int \xi \,\mathrm{d}\mu_s^h\right| \;\leq\; \int \left|\xi(x) - \xi(y)\right| \mathrm{d}\gamma(x,y) \;\leq\; \|\nabla\xi\|_\infty \mathrm{W}_2(\mu_t^h, \mu_s^h);$$

then, as $\mu_t^h = \mu_{hn_t}^h$ for $n_t = \lfloor t/h \rfloor$, and

$$F(\mu_{h(n+1)}^h) + \frac{1}{2h}\mathrm{W}_2(\mu_{h(n+1)}, \mu_{hn})^2 \;\leq\; F(\mu_{hn}^h) \qquad \text{for every } n,$$

we get

$$\begin{aligned}
\mathrm{W}_2(\mu_t^h, \mu_s^h)^2 \;&=\; \mathrm{W}_2(\mu_{hn_t}^h, \mu_{hn_s}^h)^2 \\
&\leq\; \Big[\sum \mathrm{W}_2\big(\mu_{h(k+1)}^h, \mu_{hk}^h\big)\Big]^2 \\
&\leq\; |n_t - n_s| \sum \mathrm{W}_2\big(\mu_{h(k+1)}^h, \mu_{hk}^h\big)^2 \\
&\leq\; 2h|n_t - n_s|\left|F(\mu_t^h) - F(\mu_s^h)\right| \\
&\leq\; Ch|n_t - n_s|\mathrm{W}_2(\mu_t^h, \mu_s^h) \\
&\leq\; C\,(|t-s| + h)\,\mathrm{W}_2(\mu_t^h, \mu_s^h),
\end{aligned}$$

which implies

$$\mathrm{W}_2(\mu_t^h, \mu_s^h) \;\leq\; C\,(|t-s| + h)\,. \tag{5.5.3.c}$$

Thus,

$$\begin{aligned}
\left|\int \xi \,\mathrm{d}\mu_t^h - \frac{1}{2\delta}\int_{t-\delta}^{t+\delta}\int \xi \,\mathrm{d}\mu_s^h \,\mathrm{d}s\right| \;&\leq\; \frac{1}{2\delta}\int_{t-\delta}^{t+\delta}\left|\int \xi \,\mathrm{d}\mu_t^h - \int \xi \,\mathrm{d}\mu_s^h\right| \mathrm{d}s \\
&\leq\; \frac{C\|\nabla\xi\|_\infty}{2\delta}\int_{t-\delta}^{t+\delta} (|t-s| + h)\,\mathrm{d}s \\
&\leq\; C\|\nabla\xi\|_\infty\,(\delta + h)\,.
\end{aligned}$$

Because μ^{h_n} converge to μ,

$$\frac{1}{2\delta}\int_{t-\delta}^{t+\delta}\int \xi \,\mathrm{d}\mu_s^{h_n}\,\mathrm{d}s \;\longrightarrow\; \frac{1}{2\delta}\int_{t-\delta}^{t+\delta}\int \xi \,\mathrm{d}\mu_s\,\mathrm{d}s;$$

therefore, (5.5.3.b) shows $\int \xi \, \mathrm{d}\mu_t^{h_n}$ is Cauchy. But for almost all $t \in [0, T]$, the limit can only be $\int \xi \, \mathrm{d}\mu_t$. Thus, $\mu_t^{h_n}$ converges to μ_t, and this, with (5.5.3.a), implies the densities $\rho_t^{h_n}$ weakly converge to ρ_t in all L^r for all finite $r \in [1, p]$, with

$$\|\rho_t\|_{\mathrm{L}^p} \;\leq\; e^{t\sqrt{d}/q} \|\rho_0\|_{\mathrm{L}^p}.$$

Moreover, (5.5.3.c) yields

$$\mathrm{W}_2(\mu_s, \mu_t) \;\leq\; \liminf_{n\to\infty} \mathrm{W}_2(\mu_s^{h_n}, \mu_t^{h_n}) \;\leq\; C|t-s|. \qquad \square$$

5.6 Continuity equation

5.6.1. Theorem. *Let $(\mu_t)_{t\geq 0}$ be a generalized minimizing movement given by theorem 5.5.3 on page 142. We denote by ρ_t the density of μ_t. As previously, let $\psi_{t,\theta}$ stand for the Kantorovich potential between $\theta^*_\#\mu_t$ and $\theta^*_\#\nu$ such that $\int \psi_{t,\theta} \, \mathrm{d}\theta^*_\#\mu_t = 0$. Then, in a weak sense,*

$$\frac{\partial \rho_t}{\partial t} + \operatorname{div}(v_t \rho_t) \;=\; 0 \qquad \textit{with} \qquad v_t(x) \;:=\; -\fint_{\mathbb{S}^{d-1}} \psi'_{t,\theta}(\langle\theta|x\rangle)\theta \, \mathrm{d}\theta.$$

More precisely, for any $\xi \in \mathscr{C}_c^\infty([0,\infty) \times B(0,r))$,

$$\int_0^\infty \int_{B(0,r)} \left[\frac{\partial \xi}{\partial t}(t,x) - \fint_{\mathbb{S}^{d-1}} \psi'_{t,\theta}(\langle\theta|x\rangle) \langle\theta|\nabla\xi(t,x)\rangle \, \mathrm{d}\theta \right] \rho_t(x) \, \mathrm{d}x \, \mathrm{d}t$$
$$= \; -\int_{B(0,r)} \xi(0,x)\rho_0(x) \, \mathrm{d}x.$$

The vector field v_t is a tangent vector for the Riemannian structure of $\mathcal{P}(\mathbb{R}^d)$; see the book by Luigi Ambrosio, Nicolas Gigli and Giuseppe Savaré [3, chapter 8] for definitions. Indeed, since $\psi_{t,\theta}$ is Lipschitz, if we set

$$\Psi_t(x) \;:=\; \fint_{\mathbb{S}^{d-1}} \psi_{t,\theta}(\langle\theta|x\rangle) \, \mathrm{d}\theta,$$

then Ψ_t is also Lipschitz, and $v_t = -\nabla\Psi_t$.

Proof. We will proceed in four steps.

1. On the one hand, since $\mu^{h_n} \to \mu$ for some sequence $h_n \to 0$,

$$\int_0^\infty \int_{B(0,r)} \frac{\partial \xi}{\partial t}(t,x)\rho_t^{h_n}(x)\,\mathrm{d}x\,\mathrm{d}t \;\longrightarrow\; \int_0^\infty \int_{B(0,r)} \frac{\partial \xi}{\partial t}(t,x)\rho_t(x)\,\mathrm{d}x\,\mathrm{d}t.$$

On the other hand,

$$\begin{aligned}\int_0^\infty \int_{B(0,r)} &\frac{\partial \xi}{\partial t}(t,x)\rho_t^{h_n}(x)\,\mathrm{d}x\,\mathrm{d}t \\ &= \sum_{k=0}^{\infty} \int_{kh_n}^{(k+1)h_n} \int_{B(0,r)} \frac{\partial \xi}{\partial t}(t,x)\rho_t^{h_n}(x)\,\mathrm{d}x\,\mathrm{d}t \\ &= \sum_{k=0}^{\infty} \int_{B(0,r)} \left[\xi((k+1)h_n,x) - \xi(kh_n,x)\right]\rho_{kh_n}^{h_n}(x)\,\mathrm{d}x\,\mathrm{d}t,\end{aligned}$$

because ρ^h is constant on each interval $[kh,(k+1)h)$. Then,

$$\begin{aligned}\int_0^\infty \int_{B(0,r)} &\frac{\partial \xi}{\partial t}(t,x)\rho_t^{h_n}(x)\,\mathrm{d}x\,\mathrm{d}t \\ &= -\int \xi(0,x)\rho_0^{h_n}(x)\,\mathrm{d}x - \sum_{k=1}^{\infty}\int \xi(kh_n,x)\left[\rho_{kh_n}^{h_n}(x) - \rho_{(k-1)h_n}^{h_n}(x)\right]\mathrm{d}x\,\mathrm{d}t,\end{aligned}$$

and this means, if we set $\xi_k^n(x) := \xi(kh_n,x)$,

$$\begin{aligned}\int \xi(0,x)\rho_0(x)\,\mathrm{d}x &+ \int_0^\infty \int \frac{\partial \xi}{\partial t}(t,x)\rho_t(x)\,\mathrm{d}x\,\mathrm{d}t \\ &\underset{n\to\infty}{\sim} -h_n \sum_{k=1}^{\infty} \int \xi_k^n(x) \frac{\rho_{kh_n}^{h_n}(x) - \rho_{(k-1)h_n}^{h_n}(x)}{h_n}\,\mathrm{d}x\,\mathrm{d}t. \qquad (5.6.1.a)\end{aligned}$$

2. For any $\theta \in \mathbb{S}^{d-1}$, we can find

$$\gamma_{\theta,h,t} \;\in\; \Gamma(\theta^*_\#\mu_t^h, \theta^*_\#\mu_t, \theta^*_\#\nu)$$

such that, if $u, \bar{u}, v$ stand for the variables and $\mathrm{U}, \bar{\mathrm{U}}, \mathrm{V}$ for the corresponding projectors, then $(\mathrm{U},\mathrm{V})_\#\gamma_{\theta,h,t}$ and $(\bar{\mathrm{U}},\mathrm{V})_\#\gamma_{\theta,h,t}$ and $(\mathrm{U},\bar{\mathrm{U}})_\#\gamma_{\theta,h,t}$ are all optimal; indeed, if F, G, H

stand for the cumulative distributions of the three 1D measures, then we can just take

$$\gamma_{\theta,h,t} := (F^{-1}, G^{-1}, H^{-1})_{\#}\mathcal{L}^1.$$

Then, $\gamma_{\theta,h,t}$ can then be extended into a measure $\pi_{\theta,h,t} \in \Gamma(\mu_t^h, \mu_t, \nu)$: first, take a measure $\tilde{\pi} \in \Gamma(\mu_t^h, \mu_t)$ optimal between μ_t^h and μ_t; next, disintegrate $\tilde{\pi} \otimes \nu$ with respect to $(\theta^*, \theta^*, \theta^*)$ into a family $\{[\tilde{\pi} \otimes \nu]_{u,\bar{u},v}\}$, such that, for any η,

$$\begin{aligned}\int \eta(x, \bar{x}, y)\, \mathrm{d}[\tilde{\pi} \otimes \nu](x, \bar{x}, y) \\ = \int_{\mathbb{R}^3} \left(\int_{(\mathbb{R}^{d-1})^3} \eta(u\theta + \hat{x}, \bar{u}\theta + \hat{\bar{x}}, v\theta + \hat{v})\, \mathrm{d}[\tilde{\pi} \otimes \nu]_{u,\bar{u},v}(\hat{x}, \hat{\bar{x}}, \hat{y}) \right) \\ \mathrm{d}[(\theta^*, \theta^*, \theta^*)_{\#}(\tilde{\pi} \otimes \nu)](u, \bar{u}, v);\end{aligned}$$

then define $\pi_{\theta,h,t}$ by replacing $(\theta^*, \theta^*, \theta^*)_{\#}(\tilde{\pi} \otimes \nu)$ with $\gamma_{\theta,h,t}$ in the previous expression:

$$\begin{aligned}\int \eta(x, \bar{x}, y)\, \mathrm{d}\pi_{\theta,h,t}(x, \bar{x}, y) \\ = \int_{\mathbb{R}^3} \left(\int_{(\mathbb{R}^{d-1})^3} \eta(u\theta + \hat{x}, \bar{u}\theta + \hat{\bar{x}}, v\theta + \hat{v})\, \mathrm{d}[\tilde{\pi} \otimes \nu]_{u,\bar{u},v}(\hat{x}, \hat{\bar{x}}, \hat{y}) \right) \\ \mathrm{d}\gamma_{\theta,h,t}(u, \bar{u}, v).\end{aligned}$$

Now, let $\psi_{t,\theta}^h$ be the Kantorovich potential between $\theta_{\#}^*\mu_t^h$ and $\theta_{\#}^*\nu$, and, taking back the same ξ as in the first point, set

$$\begin{aligned}I_{h,t} &:= \int_{B(0,r)} \fint_{\mathbb{S}^{d-1}} \left(\psi_{t,\theta}^h\right)' (\langle\theta|x\rangle)\, \langle\theta|\nabla\xi(t,x)\rangle\, \mathrm{d}\theta\, \mathrm{d}\mu_t^h(x), \\ I_t &:= \int_{B(0,r)} \fint_{\mathbb{S}^{d-1}} (\psi_{t,\theta})' (\langle\theta|\bar{x}\rangle)\, \langle\theta|\nabla\xi(t,\bar{x})\rangle\, \mathrm{d}\theta\, \mathrm{d}\mu_t(\bar{x});\end{aligned}$$

we can write

$$\begin{aligned}I_{h,t} &= \fint\int \langle\theta|x - y\rangle\, \langle\theta|\nabla\xi(t,x)\rangle\, \mathrm{d}\pi_{\theta,h,t}(x)\, \mathrm{d}\theta, \\ I_t &= \fint\int \langle\theta|\bar{x} - y\rangle\, \langle\theta|\nabla\xi(t,\bar{x})\rangle\, \mathrm{d}\pi_{\theta,h_n,t}(x)\, \mathrm{d}\theta.\end{aligned}$$

We conveniently define $\Phi_{\theta,h,t,y}(x) := \langle\theta|x-y\rangle\langle\theta|\nabla\xi(t,x)\rangle$; then,

$$\begin{aligned}\left|I_{h,t}-I_t\right|^2 &\leq \fint\int\left|\Phi_{\theta,h,t,y}(x)-\Phi_{\theta,h,t,y}(\bar{x})\right|^2 \mathrm{d}\pi_{\theta,h,t}(x,\bar{x},y)\,\mathrm{d}\theta\\ &\leq C_\xi\fint\int|x-\bar{x}|^2\,\mathrm{d}\pi_{\theta,h,t}(x,\bar{x},y)\,\mathrm{d}\theta\\ &\leq C_\xi\fint\int 2|\langle\theta|x-\bar{x}\rangle|^2+2|\hat{x}-\hat{\bar{x}}|^2\,\mathrm{d}\pi_{\theta,h,t}(x,\bar{x},y)\,\mathrm{d}\theta,\end{aligned}$$

where $x=\langle\theta|x\rangle\theta+\hat{x}$ and $\bar{x}=\langle\theta|\bar{x}\rangle\theta+\hat{\bar{x}}$. Thus,

$$\begin{aligned}\left|I_{h,t}-I_t\right|^2 &\leq C\fint\left(\int|u-\bar{u}|^2\,\mathrm{d}\gamma_{\theta,h,t}(u,\bar{u},v)+\int|x-\bar{x}|^2\,\mathrm{d}\tilde{\pi}_{\theta,h,t}(x,\bar{x})\right)\mathrm{d}\theta\\ &\leq C\left(\mathrm{SW}_2(\mu_t^h,\mu_t)^2+\mathrm{W}_2(\mu_t^h,\mu_t)^2\right).\end{aligned}$$

As $\operatorname{supp}\xi\subset[0,T]\times B(0,r)$,

$$\int_0^\infty\left|I_{h_n,t}-I_t\right|^2\mathrm{d}t \;\leq\; C\int_0^T \mathrm{W}_2(\mu_t^{h_n},\mu_t)^2\,\mathrm{d}t,$$

and since $\mathrm{W}_2(\mu_t^{h_n},\mu_t)$ tends to zero, by Lebesgue's dominated convergence theorem we get $\int I_t\,\mathrm{d}t\sim\int I_{h_n,t}\,\mathrm{d}t$, which means

$$\int_0^\infty\int\fint(\psi_{t,\theta})'(\langle\theta|\bar{x}\rangle)\,\langle\theta|\nabla\xi(t,\bar{x})\rangle\,\mathrm{d}\theta\,\mathrm{d}\mu_t(\bar{x})\,\mathrm{d}t \underset{n\to\infty}{\sim}\int_0^\infty\int\fint\left(\psi_{t,\theta}^{h_n}\right)'(\langle\theta|x\rangle)\,\langle\theta|\nabla\xi(t,x)\rangle\,\mathrm{d}\theta\,\mathrm{d}\mu_t^{h_n}(x)\,\mathrm{d}t.$$

Recall $\mu_t^{h_n}$ is constant on each interval $[kh_n,(k+1)h_n)$; hence,

$$\int_0^\infty\int\fint(\psi_{t,\theta})'(\langle\theta|\bar{x}\rangle)\,\langle\theta|\nabla\xi(t,\bar{x})\rangle\,\mathrm{d}\theta\,\mathrm{d}\mu_t(\bar{x})\,\mathrm{d}t \underset{n\to\infty}{\sim} h_n\sum_{k=1}^\infty\int\fint\left(\psi_{kh_n,\theta}^{h_n}\right)'(\langle\theta|x\rangle)\,\langle\theta|\nabla\Xi_k^n(x)\rangle\,\mathrm{d}\theta\,\mathrm{d}\mu_{kh_n}^{h_n}(x),$$

where we have set

$$\Xi_k^n(x) \;:=\; \frac{1}{h_n}\int_{kh_n}^{(k+1)h_n}\xi(t,x)\,\mathrm{d}t.$$

However,

$$\begin{aligned}\left|\nabla\xi(kh_n,x) - \nabla\Xi_k^n(x)\right| &\leq \frac{1}{h_n}\int_{kh_n}^{(k+1)h_n}\left|\nabla\xi(kh_n,x) - \nabla\xi(t,x)\,\mathrm{d}x\right|\mathrm{d}t \\ &\leq C_\xi h_n,\end{aligned}$$

so, since $\xi_k^n(x) := \xi(kh_n, x)$,

$$\begin{aligned}&\int_0^\infty \int \fint (\psi_{t,\theta})'\left(\langle\theta|\bar{x}\rangle\right)\langle\theta|\nabla\xi(t,\bar{x})\rangle\,\mathrm{d}\theta\,\mathrm{d}\mu_t(\bar{x})\,\mathrm{d}t \\ &\qquad \underset{n\to\infty}{\sim} h_n\sum_{k=1}^{\infty}\int\fint\left(\psi_{kh_n,\theta}^{h_n}\right)'(\theta^*)\,\langle\theta|\nabla\xi_k^n\rangle\,\mathrm{d}\theta\,\mathrm{d}\mu_{kh_n}^{h_n}. \qquad (5.6.1.b)\end{aligned}$$

3. Using to proposition 1.5.7 on page 38 and proposition 5.1.7 on page 125 and the optimality of μ_{kh}^h, if φ_k^h denotes the Kantorovich potential between μ_{kh}^h and $\mu_{(k-1)h}^h$,

$$\frac{1}{h_n}\int\left\langle\nabla\varphi_k^{h_n}\,\middle|\,\nabla\xi_k^n\right\rangle\mathrm{d}\mu_{kh_n}^{h_n} = -\int\fint\left(\psi_{kh_n,\theta}^{h_n}\right)'(\theta^*)\,\langle\theta|\nabla\xi_k^n\rangle\,\mathrm{d}\theta\,\mathrm{d}\mu_{kh_n}^{h_n}.$$

Let γ be an optimal transport plan between μ_{kh}^h and $\mu_{(k-1)h}^h$; then,

$$\begin{aligned}\int \xi_k^n(x)\frac{\rho_{kh_n}^{h_n}(x) - \rho_{(k-1)h_n}^{h_n}(x)}{h_n}\,\mathrm{d}x &= \frac{1}{h_n}\int\left\{\xi_k^n(y) - \xi_k^n(x)\right\}\mathrm{d}\gamma(x,y), \\ \frac{1}{h_n}\int\left\langle\nabla\varphi_k^{h_n}(x)\,\middle|\,\nabla\xi_k^n(x)\right\rangle\mathrm{d}\mu_{kh_n}^{h_n}(x) &= -\frac{1}{h_n}\int\langle\nabla\xi_k^n(x)|y-x\rangle\,\mathrm{d}\gamma(x,y).\end{aligned}$$

and, since $\left|\xi_k^n(y) - \xi_k^n(x) - \langle\nabla\xi_k^n(x)|y-x\rangle\right| \leq C|x-y|^2$,

$$\int\left|\xi_k^n(y) - \xi_k^n(x) - \langle\nabla\xi_k^n(x)|y-x\rangle\right|\mathrm{d}\gamma(x,y) \leq C\mathrm{W}_2\left(\mu_{(k-1)h_n}^{h_n}, \mu_{kh_n}^{h_n}\right)^2;$$

so, using (5.5.3.c), we get

$$\begin{aligned}\left|\int \xi_k^n(x)\frac{\rho_{kh_n}^{h_n}(x) - \rho_{(k-1)h_n}^{h_n}(x)}{h_n}\,\mathrm{d}x + \int\fint\left(\psi_{kh_n,\theta}^{h_n}\right)'(\theta^*)\langle\theta|\nabla\xi_k^n\rangle\,\mathrm{d}\theta\,\mathrm{d}\mu_{kh_n}^{h_n}\right| \\ \leq Ch_n.\end{aligned}$$

This immediately yields

$$h_n \sum_{k=1}^{\infty} \int \fint \left(\psi^{h_n}_{kh_n,\theta}\right)'(\theta^*) \, \langle\theta|\nabla\xi^n_k\rangle \, \mathrm{d}\theta \, \mathrm{d}\mu^{h_n}_{kh_n}$$

$$\underset{n\to\infty}{\sim} \; -\sum_{k=1}^{\infty} h_n \int \xi^n_k(x) \frac{\rho^{h_n}_{kh_n}(x) - \rho^{h_n}_{(k-1)h_n}(x)}{h_n} \, \mathrm{d}x \qquad (5.6.1.c)$$

4. Combining (5.6.1.a), (5.6.1.b), and (5.6.1.c), we get the result. □

5.7 Open questions

5.7.1. The first and main question that still need to be investigated, is the convergence of μ_t toward the target measure ν when t tends to infinity. When working with discrete measures only, we should not expect any convergence, as symmetry is preserved by the algorithm and a discrete solution might require it to be broken. Nonetheless, convergence might still happen when the measures are absolutely continuous.

The first step toward convergence could be to study the stationary points. We know from theorem 5.6.1 on page 145 that

$$\frac{\partial \mu_t}{\partial t} + \operatorname{div}(v_t\mu_t) = 0 \qquad \text{with} \qquad v_t(x) := -\fint_{\mathbb{S}^{d-1}} \psi'_{t,\theta}(\langle\theta|x\rangle)\theta \, \mathrm{d}\theta.$$

But, does $\fint \psi'_{t,\theta}(\theta^*)\theta \, \mathrm{d}\theta = 0$ implies $\mu_t = \nu$? An answer can easily be given though, if μ_t is absolutely continuous with a strictly positive density:

5.7.2. Lemma. *For any $\mu \in \mathcal{P}(B(0,r))$, if μ is absolutely continuous with a strictly positive density, then $\mu = \nu$ if and only if*

$$\int_{\mathbb{S}^{d-1}} \psi'_\theta(\langle\theta|x\rangle)\theta \, \mathrm{d}\theta = 0 \qquad \text{for } \mu\text{-a.e. } x,$$

with ψ_θ the unidimensional Kantorovich potential between $\theta^*_\#\mu$ and $\theta^*_\#\nu$.

Proof. If $\mu = \nu$, obviously the integral is zero. Conversely, let

$$\Psi(x) := \fint_{\mathbb{S}^{d-1}} \psi_\theta(\langle\theta|x\rangle) \, \mathrm{d}\theta.$$

Then, Ψ is Lipschitz, differentiable almost everywhere, and

$$\nabla\Psi(x) \;=\; \fint_{\mathbb{S}^{d-1}} \psi'_\theta(\langle\theta|x\rangle)\theta\,\mathrm{d}\theta.$$

Thus, if $\nabla\Psi = 0$ almost everywhere, Ψ is constant, and

$$\int \Psi\,\mathrm{d}\mu \;=\; \fint_{\mathbb{S}^{d-1}} \int \psi_\theta(u)\,\mathrm{d}[\theta^*_\#\mu](u)\,\mathrm{d}\theta = 0,$$

yields $\Psi \equiv 0$. On the other hand,

$$\frac{1}{2}\mathrm{W}_2(\theta^*_\#\mu, \theta^*_\#\nu)^2 = \int \psi_\theta D\,[\theta^*_\#\mu] + \int \psi^c_\theta\,\mathrm{d}[\theta^*_\#\nu], \tag{5.7.2.a}$$

and

$$\forall\, u, v \in [-r, r], \quad \psi_\theta(u) + \psi^c_\theta(v) \;\le\; \frac{1}{2}|u - v|^2.$$

Taking $u = v = \langle\theta|y\rangle$ and averaging the last inequality with respect to θ, we get

$$\forall\, y \in B(0, r), \quad \Psi(y) + \fint_{\mathbb{S}^{d-1}} \psi^c_\theta(\theta^*(y))\,\mathrm{d}\theta \;\le\; 0.$$

Then, since $\Psi \equiv 0$, integrating with respect to ν we get

$$\int\fint \psi^c_\theta(\theta^*(y))\,\mathrm{d}\theta\,\mathrm{d}\nu(y) \;\le\; 0.$$

Then, averaging (5.7.2.a) with respect to θ, we also obtain

$$\frac{1}{2}\mathrm{SW}_2(\mu, \nu)^2 = \int\fint \psi^c_\theta(\theta^*(y))\,\mathrm{d}\theta\,\mathrm{d}\nu(y) \;\le\; 0.$$

As the sliced Wasserstein distance is a distance, this implies $\mu = \nu$. □

5.7.3. Another question, although a less important one, regards uniqueness. To obtain the generalized minimizing movement $(\mu_t)_{t\ge 0}$, we have used the compactness of $\mathcal{P}(B(0, r))$ so many times, that there could be a great number of such curve for any given starting point μ_0 and any target measure ν. For gradient flows in the space of probability measures, uniqueness often comes from the convexity of the functional.

However, like the usual Wasserstein distance, it is not difficult to show the sliced Wasserstein distance is 2-*concave* along geodesics: if (μ_t) is a geodesic between μ_0 and μ_1, then

$$\mathrm{SW}_2(\mu_t, \nu)^2 \;\geq\; (1-t)\mathrm{SW}_2(\mu_0, \nu)^2 + t\mathrm{SW}_2(\mu_1, \nu)^2 - t(1-t)\mathrm{W}_2(\mu_0, \mu_1)^2.$$

This does not prevent uniqueness, but if there is only one possible curve (μ_t), we will have to prove it by other means.

Synthèse

Beaucoup d'illustrations ont déjà été proposées pour présenter simplement le problème du transport optimal. On a pu parler de tas de sables à déplacer [45, 62], de cafés parisiens à fournir en pain [63], de charbon à amener depuis les mines jusqu'aux centrales électriques [56], etc. Dans un soucis d'originalité, qu'il soit permis d'ajouter à cette liste l'exemple suivant.

Il existe en Chine de gigantesques complexes industriels, regroupant plusieurs centaines de milliers d'ouvriers sur quelques kilomètres carrés [25]. Ceux-ci sont logés sur place, dans des dortoirs. Dans un soucis d'efficacité, il convient donc d'attribuer à chaque ouvrier un lit qui ne soit pas trop éloigné de son lieu de travail. Admettons qu'assigner à quelqu'un travaillant sur la ligne de montage x un lit dans le dortoir y engendre pour l'entreprise un cout $c(x, y)$, correspondant par exemple aux frais de fonctionnement d'un système de navettes. Quelle est alors la meilleure manière de loger tous les employés ?

La répartition des travailleurs et celle des lits peuvent être représentées par deux mesures μ et ν, de sorte que $\mu(\mathrm{A})$ représente le nombre d'ouvriers travaillant dans la zone A et $\nu(\mathrm{B})$ le nombre de lits disponibles dans la zone B. Supposons qu'il n'y ait pas de logements superflus, et que la capacité des dortoirs corresponde exactement aux besoins d'accueil ; cela se traduit par l'égalité

$$\int \mathrm{d}\mu(x) = \int \mathrm{d}\nu(y).$$

Nous pouvons donc considérer que μ et ν sont des mesures de probabilité. La solution à notre problème est à rechercher sous la forme d'une mesure de probabilité γ sur l'espace produit, telle que $\gamma(\mathrm{A} \times \mathrm{B})$ donne le nombre — ou plutôt, la proportion — d'ouvriers travaillant dans la zone A et logeant dans la zone B ; ceci implique que μ et ν doivent être les marges de γ :

$$\mu(A) = \int_{x\in A} d\gamma(x,y) \qquad \text{et} \qquad \nu(B) = \int_{y\in B} d\gamma(x,y).$$

Notons Γ(μ, ν) l'ensemble de ces mesures γ, appelés « plans de transport ». Le déplacement quotidien des ouvriers entre leur dortoir et leur poste de travail entrainera alors toujours une dépense au moins égale à

$$\inf_{\gamma\in\Gamma(\mu,\nu)} \int c(x,y)\, d\gamma(x,y). \tag{a}$$

Il n'est pas difficile de montrer que cet infimum est toujours atteint : il existe toujours au moins un plan de transport γ correspondant à une allocation optimale des lits. Il est cependant très difficile de caractériser *a priori* les solutions ; l'enjeu est justement de calculer un plan γ optimal.

C'est à Leonid Kantorovitch, mathématicien soviétique et récipiendaire du prix Nobel d'économie en 1975, que l'on doit cette formulation du problème du transport optimal — non pas en termes d'ouvriers, mais de mesures de probabilité. Pour continuer dans la veine industrielle chinoise, Kantorovitch a montré dans les années quarante [35, 36] que le problème pouvait en quelque sorte être transféré sur les employés. Il suffit en effet de leur faire porter le coût du transport, de leur faire payer un loyer pour leur logement, et de compenser cela pour eux par une subvention. Il ne s'agit pas de gagner de l'argent ainsi, mais d'inciter les ouvriers à trouver eux-même le lit le mieux placé ; le prix du transport sera donc égal à son cout de fonctionnement $c(x,y)$. Notons $S(x)$ la subvention accordée aux employés de la chaine de montage x, et $L(y)$ le loyer d'un lit dans le dortoir y ; Kantorovitch a montré que la valeur des subventions S et celle des loyers L peuvent être fixées judicieusement, de telle sorte que :

- les travailleurs reçoivent toujours moins que ce qu'ils ont à dépenser, c'est-à-dire $S(x) \leqslant L(y) + c(x,y)$ quels que soient x et y ;
- chaque ouvrier peut cependant trouver un lit idéal qui ne lui fera pas perdre d'argent, pour lequel $S(x) = L(y) + c(x,y)$;
- il est possible que les employés réussissent tous à trouver un lit idéal.

Kantorovitch a montré que, dans ce cas, le cout total de l'opération pour l'entreprise, qui est donné par la différence entre le montant total des subventions et la somme récupérée par les loyers, est alors

$$\int S(x)\, d\mu(x) - \int L(y)\, d\nu(y) \;=\; \min_{\gamma\in\Gamma(\mu,\nu)} \int c(x,y)\, d\gamma(x,y).$$

Les potentiels S et L sont alors appelés des « potentiels de Kantorovitch ».

Notons que si les chaines de montages sont de petits ateliers qui ne sont pas trop concentrés, il est envisageable que les ouvriers travaillant ensemble en x puissent se retrouver dans le même dortoir $y = T(x)$. Une telle application T fait alors correspondre μ et ν, ce qui se traduit par

$$\nu(B) = \mu(T^{-1}(B)) \qquad \text{quel que soit } B,$$

ou encore

$$\int \xi(y)\, d\nu(y) = \int \xi(T(x))\, d\mu(x)\ ;$$

on dit alors que T envoie μ sur ν, et l'on note $\nu = T_{\#}\mu$. Notons qu'à une telle application est associé un plan de transport γ_T, défini par

$$\gamma_T(A \times B) \;=\; \mu(A \cap T^{-1}(B)).$$

L'intuition que le transport optimal prend la forme d'une telle application si la source est suffisamment diffuse se traduit mathématiquement par un résultat démontré par Yann Brenier [13, 14] à la fin du XX^e siècle : si le cout de transport est égal au carré de la distance, $c(x,y) = |x-y|^2$, et si μ est absolument continue, alors il existe une application T envoyant μ sur ν qui résout le problème du transport optimal, c'est-à-dire que le plan de transport γ_T est optimal, et le coût total de transport est

$$\min_{\gamma\in\Gamma(\mu,\nu)} \int c(x,y)\, d\gamma(x,y) \;=\; \int c(x,y)\, d\gamma_T(x,y) \;=\; \int c(x, T(x))\, d\mu(x).$$

De plus, les potentiels de Kantorovitch S et L sont alors reliés à l'application T par les relations

$$y = \mathrm{T}(x) = x - \nabla \mathrm{S}(x) \qquad \text{et} \qquad x = \mathrm{T}^{-1}(y) = y + \nabla \mathrm{L}(y).$$

Celles-ci traduisent le fait que, pour trouver un logement, il est intéressant d'aller dans la direction des subventions décroissantes, puisque les subventions sont moindres lorsque des dortoirs sont proches. Inversement, des loyers plus élevés signalent une plus grande demande, et donc un plus grand nombre d'ateliers.

Le résultat de M. Brenier résout ainsi un problème ancien, posé d'abord par Gaspard Monge [45] à la fin du xviii^e siècle. Celui-ci avait en effet essayé de résoudre le problème du transport optimal en cherchant la solution, non pas sous la forme d'un plan de transport γ comme Kantorovitch plus tard, mais sous la forme d'une application T. Le coût total de transport est alors au minimum

$$\inf_{\nu = \mathrm{T}_{\#}\mu} \int c(x, \mathrm{T}(x))\, \mathrm{d}\mu(x). \tag{b}$$

Puisque chaque application T qui fait correspondre μ et ν donne un plan de transport γ_{T}, le problème de Kantorovitch (a) est en fait une extension du problème initial de Monge (b). Cependant, à l'inverse du premier, le second peut ne pas avoir de solution ; il peut même ne pas y avoir d'application T telle que $\nu = \mathrm{T}_{\#}\mu$, par exemple si μ est discrète et que ν est uniforme. Mais si μ est absolument continue, et pour un coût quadratique, l'application de M. Brenier est solution du problème de Monge, et résout aussi le problème de Kantorovitch.

Il faut noter qu'en dimension un, il est très facile de résoudre le problème de Monge. S'il l'on note par F et G les fonctions de répartition de μ et ν respectivement,

$$\mathrm{F}(x) \;:=\; \mu(]-\infty, x]) \qquad \text{et} \qquad \mathrm{G}(y) \;:=\; \nu(]-\infty, y]),$$

alors la solution au problème de Monge pour n'importe quel cout strictement convexe, c'est-à-dire n'importe quel cout défini par $c(x, y) = h(y - x)$ avec h positive et strictement convexe, est

$$\mathrm{T}(x) \;:=\; \mathrm{G}^{-1}(\mathrm{F}(x)). \tag{c}$$

En dimension plus grande, il est beaucoup plus difficile de calculer cette solution. Quelques méthodes ont cependant été développées au fil des années :

— lorsque les mesures de départ et d'arrivées sont discrètes, un algorithme célèbre a été mis au point par Dimitri Bertsekas [9] ;
— cet algorithme a ensuite été utilisé par Damien Bosc [11] pour traiter le cas de mesures continues, par approximation ;
— toujours dans le cadre continu, Jean-David Benamou et Yann Brenier [7] ont aussi proposé une méthode, basée sur une interprétation en termes de mécanique des fluides ;
— Sigurd Angenent, Steven Haker et Allen Tannenbaum [4] ont, eux, réussi à utiliser une méthode de descente de gradient ;
— enfin, Grégoire Loeper et Francesca Rapetti [40] ont pu utiliser avec succès la méthode de Newton.

Il y a quelques années, Guillaume Carlier, Alfred Galichon et Filippo Santambrogio [19] ont cependant proposé une nouvelle méthode pour calculer l'application optimale de M. Brenier. Leur approche repose sur l'introduction d'un paramètre $t \in [0,1]$ dans la fonction de cout : par exemple, en dimension deux,

$$c_t(x,y) \;=\; |x_1, y_1|^2 + t|x_2 - y_2|^2.$$

Ceci revient à dire qu'un déplacement suivant l'axe vertical (nord-sud) coute moins qu'un déplacement suivant l'axe horizontal (est-ouest). La solution au problème du transport optimal pour le cout c_t est encore donnée par une application, que l'on notera T_t. Pour $t = 1$, il s'agit bien entendu de l'application de M. Brenier ; pour $t \in]0,1]$, cette application est aussi reliée à un potentiel de Kantorovitch S_t *via* la relation

$$\mathrm{T}_t(x) \;=\; x - \mathrm{A}_t^{-1}\nabla \mathrm{S}_t(x) \qquad \text{avec} \qquad \mathrm{A}_t \;=\; \begin{pmatrix} 1 & 0 \\ 0 & t \end{pmatrix}.$$

MM. Carlier, Galichon et Santambrogio ont montré que, lorsque t tend vers zéro, T_t converge vers le réarrangement de Knothe–Rosenblatt.

Ce « réarrangement », introduit dans les années cinquante séparément par Herbert Knothe [38] et Murrey Rosenblatt [51], envoie encore μ sur ν, et s'obtient par une succession de transformations unidimensionnelles. Par exemple, en dimension deux, si $f(x_1, x_2)$ et $g(y_1, y_2)$ sont les densités de μ et ν, alors

$$x_1 \mapsto \int f(x_1, t)\,\mathrm{d}t \qquad \text{et} \qquad y_1 \mapsto \int g(y_1, t)\,\mathrm{d}t$$

sont deux mesures de probabilité sur la droite réelle, et nous savons comment envoyer la première sur la seconde, grâce à la formule (c) ; notons $x_1 \mapsto \mathrm{T}_\mathrm{K}^1(x_1)$ l'application ainsi obtenue. Alors,

$$x_2 \mapsto \frac{f(x_1, x_2)}{\int f(x_1, t)\,\mathrm{d}t} \qquad \text{et} \qquad y_2 \mapsto \frac{g(\mathrm{T}_\mathrm{K}^1(x_1), y_2)}{\int g(\mathrm{T}_\mathrm{K}^1(x_1), t)\,\mathrm{d}t}$$

sont aussi deux mesures de probabilités ; si l'on note $x_2 \mapsto \mathrm{T}_\mathrm{K}^2(x_1, x_2)$ l'application envoyant l'une sur l'autre, alors le réarrangement de Knothe–Rosenblatt est $\mathrm{T}_\mathrm{K}(x_1, x_2) := (\mathrm{T}_\mathrm{K}^1(x_1), \mathrm{T}_\mathrm{K}^2(x_1, x_2))$.

Contrairement à l'application de M. Brenier, ce réarrangement est donc très facile à calculer explicitement. MM. Carlier, Galichon et Santambrogio ont pourvé que, lorsque l'une des deux mesures est discrète, l'évolution de l'application T_t entre le réarrangement et l'application de M. Brenier est guidée par une équation différentielle.

La première partie de cette thèse a été consacrée à l'extension de leurs résultats aux cas de mesures absolument continues. Le problème gagne alors notablement en complexité.

Tant que t demeure strictement positif, il n'y a pas de grande difficulté. Notons f et g les densités respectives de μ et ν. Puisque $\mathrm{T}_t(x) = x - \nabla \mathrm{S}_t(x)$ envoie μ sur ν, pour n'importe quelle fonction test ξ nous avons

$$\int \xi(y) g(y)\,\mathrm{d}y = \int \xi(y)\,\mathrm{d}\nu(y) = \int \xi(\mathrm{T}_t(x))\mu(x) = \int \xi(\mathrm{T}_t(x)) f(x)\,\mathrm{d}x.$$

Un changement de variable donne alors une équation de Monge–Ampère :

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

Appliquer le théorème des fonctions implicites permet d'obtenir que $t \mapsto \mathrm{S}_t$ est régulière, et dériver l'équation de Monge–Ampère donne alors une équation aux dérivées partielles :

$$\operatorname{div}\left(f\left[\mathrm{I}_d - \mathrm{A}_t^{-1}\nabla^2\mathrm{S}_t\right]^{-1}\mathrm{A}_t^{-1}\left(\nabla\dot{\mathrm{S}}_t - \dot{\mathrm{A}}_t\mathrm{A}_t^{-1}\nabla\mathrm{S}_t\right)\right) = 0. \tag{d}$$

Cette équation devient cependant singulière lorsque t s'annule, puisque A_t^{-1} est une matrice diagonale dont les coefficients font intervenir $1/t$.

Ce problème de singularité peut cependant être contourné en faisant un développement de S_t au premier ordre vis-à-vis du paramètre t. Supposons que nous soyons en dimension deux ; puisque l'application $\mathrm{T}_t = \mathrm{Id} - \mathrm{A}_t^{-1}\nabla\mathrm{S}_t$ converge vers le réarrangement de Knothe-Rosenblatt, et que celui peut s'écrire $\mathrm{T}_\mathrm{K}(x_1, x_2) = (\mathrm{T}_\mathrm{K}^1(x_1), \mathrm{T}_\mathrm{K}^2(x_1, x_2))$, nous écrirons

$$\mathrm{S}_t(x_1, x_2) = u_t(x_1) + tv_t(x_1, x_2).$$

Cette décomposition est unique si l'on impose

$$\int u_t(x_1)\,\mathrm{d}x_1 = \int \mathrm{S}_t(x)\,\mathrm{d}x \qquad \text{et} \qquad \int v_t(x)\,\mathrm{d}x_2 = 0,$$

car dans ce cas

$$u_t(x_1) := \int \mathrm{S}_t(x_1, x_2)\,\mathrm{d}x_2 \qquad \text{et} \qquad s_t(x_1, x_2) := \frac{\mathrm{S}_t(x_1, x_2) - u_t(x_1)}{t}.$$

Alors

$$\mathrm{T}_t(x_1, x_2) = \mathrm{Id} - \begin{pmatrix} \partial_1 u_t(x_1) + t\partial_1 v_t(x_1, x_2) \\ \partial_2 u_t(x_1, x_2) \end{pmatrix},$$

et l'équation (d) perd sa singularité. Elle peut alors être étudiée lorsque t tends vers zéro. Notons que ce raisonnement se peut transposer en n'importe quelle dimension $d > 2$.

Il faut noter que l'aspect le plus délicat de cette étude lorsque t tends vers zéro provient alors d'une perte de régularité vis-à-vis de la seconde variable, qui empêche d'appliquer le théorème des fonctions implicites classique comme précédemment. Cette difficulté peut être contournée en utilisant une version plus forte du théorème, due à John Nash et Jürgen Moser ; celle-ci nécessite néanmoins de ne plus travailler qu'avec des mesures extraordinairement régulières, absolument continues, strictement positives et de classe $\mathscr{C}^\infty$.

Il a cependant été possible d'utiliser cette méthode pour calculer numériquement le transport optimal. L'équation (d), où $S_t(x_1, x_2)$ a été remplacé par $u_t(x_1) + tv_t(x_1, x_2)$, peut être décomposé, ce qui donne le système suivant :

$$\begin{cases} \partial^2_{1,1}\dot{u}_t(x_1) &= \int p_t(x_1, x_2)\,\mathrm{d}x_2, \\ \mathrm{div}(A_t \nabla \dot{v}_t) &= q_t, \end{cases}$$

avec

$$\begin{cases} p_t &= \det(\nabla^2 v_t) + t\,\mathrm{div}\Big(\big[\mathrm{Co}\,\nabla^2 v_t\big]^* \nabla \dot{v}_t\Big), \\ q_t &= \det(\nabla^2 v_t) - (1 - \partial^2_{2,2} v_t)\partial^2_{1,1}\dot{u}_t - \partial^2_{1,1} v_t, \end{cases}$$

Le potentiel $S_t = u_t + tv_t$ n'étant défini qu'à une constante près, les deux équations suivantes peuvent être ajoutées :

$$\int u_t(x_1)\,\mathrm{d}x_1 \;=\; 0 \qquad \text{et} \qquad \int v_t(x_1, x_2)\,\mathrm{d}x_2 \;=\; 0.$$

Une discrétisation explicite en temps permet alors d'obtenir les résultats présentés dans le chapitre 4.

Dans la dernière partie de cette thèse, nous avons étudié l'algorithme IDT (Iterative Distribution Transfer), développé par François Pitié, Anil C. Kokaram et Rozenn Dahyot [49]. Cet algorithme construit une application de transport suffisamment proche de celle de M. Brenier pour convenir à la plupart des applications [50]. Cependant, ses caractéristiques mathématiques sont encore assez mal connues.

Considérons une mesure de référence ν sur $\mathbb{R}^d$ de densité g, et fixons une mesure de départ μ_0 de densité f_0, ainsi qu'une première base orthonormale $e^0 = (e^0_1, \ldots, e^0_d)$ de $\mathbb{R}^d$. Il est possible de projeter chacune des deux mesures sur les axes donnés par e^0 ; nous obtenons ainsi d couples de mesures unidimensionnelles f^0_i et g^0_i, qui sont données par

$$e^0_{i\,\#}\mu_0 = f^0_i(t) = \int_{x_i = t} \mathrm{d}\mu_0(x) \qquad \text{et} \qquad e^0_{i\,\#}\nu = g^0_i(t) \;:=\; \int_{y_i = t} \mathrm{d}\nu(y).$$

Nous savons envoyer l'une sur l'autre, grâce à une application t_i^0 ; posons alors

$$\mathrm{T}_0(x) \;:=\; \sum_{i=1}^{d} t_i^0(x_i) e_i^0.$$

Cette application n'envoie pas μ_0 sur ν, mais la mesure image $\mu_1 := \mathrm{T}_{0\#}\mu_0$ semble néanmoins plus proche de ν que μ_0. Nous pouvons alors choisir une suite de base (e^n) et, en répétant l'opération, construire une suite d'applications (T_n) et une suite de mesures (μ_n) telles que $\mu_{n+1} := \mathrm{T}_{n\#}\mu_n$, où l'application T_n a été construite à l'aide de la base e^n, et les mesures μ_n semblent alors se rapprocher empiriquement de ν si les bases (e^n) forment une suite dense dans l'espace des bases. La convergence mathématique n'a pu cependant être démontrée par MM. Pitié, Kokaram et Dahyot [49] que dans le cas de mesures gaussiennes.

Il se trouve que l'algorithme IDT peut être interprété en termes de flot de gradients pour une fonctionnelle faisant intervenir une certaine distance. En effet, dans le problème du transport optimal, la valeur minimale du cout de transport induit une distance, appelée distance de Wasserstein, entre les deux mesures μ et ν ; notamment, lorsque ce cout est égal au carré de la distance euclidienne, on définit

$$\mathrm{W}_2(\mu,\nu)^2 \;:=\; \min_{\Gamma(\mu,\nu)} \int |x-y|^2 \,\mathrm{d}\gamma(x,y).$$

L'algorithme IDT correspond alors à un schéma d'Euler explicite pour un flot de gradients pour la fonctionnelle

$$\mathcal{F}(\mu) \;=\; d \fint_{\mathbb{S}^{d-1}} \mathrm{W}_2(\theta_\#\mu, \theta_\#\nu)^2 \,\mathrm{d}\theta,$$

où la fonctionnelle $\mathcal{F}$ est aussi discrétisée à chaque étape n et approchée par

$$\mathcal{F}(\mu) \;\approx\; \sum_{i=1}^{d} \mathrm{W}_2({e_i^n}_\#\mu, {e_i^n}_\#\nu)^2.$$

Il se trouve que

$$\mathrm{SW}_2(\mu,\nu)^2 \;:=\; \fint_{\mathbb{S}^{d-1}} \mathrm{W}_2(\theta_\#\mu, \theta_\#\nu)^2 \,\mathrm{d}\theta,$$

définit une nouvelle distance, appelée distance de Wasserstein projetée (ou distance de Super-Wasserstein), qui est équivalente avec la distance de Wasserstein usuelle — comme je l'ai démontré dans le théorème 5.1.5 (page 123).

Les dernières sections sont consacrées à l'étude du flot de gradients pour la fonctionnelle $\mathcal{F}$ dans l'espace des mesures de probabilité, au sens de la théorie développée par Luigi Ambrosio, Nicola Gigli et Giuseppe Savaré [3]. Ce flot, défini comme étant la limite d'un schéma d'Euler implicite pour $\mathcal{F}$, pourrait en effet permettre de mieux comprendre le comportement de l'algorithme IDT.

Bibliography

[1] Luigi Ambrosio. "Movimenti minimizzanti". *Rend. Accad. Naz. Sci. XL Mem. Mat. Appl. (5)*, 1995, vol. 19, pp. 191–246.

[2] ———. "Lecture notes on optimal transport problems". In *Mathematical Aspects of Evolving Interfaces (Funchal, 2000)*, Lecture Notes in Math., vol. 1812, pp. 1–52, doi:10.1007/978-3-540-39189-0_1. Springer, Berlin, 2003.

[3] Luigi Ambrosio, Nicola Gigli, and Giuseppe Savaré. *Gradient Flows, in Metric Spaces and in the Space of Probability Measures*, second ed. Lectures in Mathematics ETH Zürich, doi:10.1007/978-3-7643-8722-8. Birkhäuser Verlag, Basel, 2008.

[4] Sigurd Angenent, Steven Haker, and Allen Tannenbaum. "Minimizing flows for the Monge–Kantorovich problem". *SIAM J. Math. Anal.*, 2003, vol. 35, no. 1, pp. 61–97, doi:10.1137/S0036141002410927.

[5] Paul Appell. "Mémoire sur la théorie des déblais et des remblais". In *Mémoires présentés par divers savants à l'Académie royale des sciences de l'Institut de France*, vol. 29, no. 3, pp. 1–208. Imprimerie nationale, Paris, 1887.

[6] ———. *Le Problème géométrique des déblais et remblais.* Mémorial des sciences mathématiques, vol. 27. Gauthier–Villars, Paris, 1928.

[7] Jean-David Benamou and Yann Brenier. "A computational fluid mechanics solution to the Monge–Kantorovich mass transfer problem". *Numer. Math.*, 2000, vol. 84, no. 3, pp. 375–393, doi:10.1007/s002110050002.

[8] Jean-David Benamou, Adam Oberman, and Froese Britanny. *Numerical Solution of the Second Boundary Value Problem for the Elliptic Monge–Ampère Equation.* Research Report, INRIA, June 2012, http://hal.inria.fr/hal-00703677/PDF/bof.pdf.

[9] Dimitri P. Bertsekas. "Auction algorithms for network flow problems, a tutorial introduction". *Comput. Optim. Appl.*, 1992, vol. 1, no. 1, pp. 7–66, doi:10.1007/BF00247653.

[10] Nicolas Bonnotte. "From Knothe's rearrangement to Brenier's optimal transport map". *SIAM J. Math. Anal.*, 2013, vol. 45, no. 1, pp. 64–87, doi:10.1137/120874850.

[11] Damien Bosc. *Three Essays on Modeling the Dependence Between Financial Assets.* PhD thesis, École polytechnique, June 2012, http://pastel.archives-ouvertes.fr/pastel-00721674.

[12] Andrea Braides. *Γ-Convergence for Beginners.* Oxford Lecture Series in Mathematics and its Applications, vol. 22, doi:10.1093/acprof:oso/9780198507840.001.0001. Oxford University Press, Oxford, 2002.

[13] Yann Brenier. "Décomposition polaire et réarrangement monotone des champs de vecteurs". *C. R. Acad. Sci. Paris Sér. I Math.*, 1987, vol. 305, no. 19, pp. 805–808.

[14] ———. "Polar factorization and monotone rearrangement of vector-valued functions". *Comm. Pure Appl. Math.*, 1991, vol. 44, no. 4, pp. 375–417, doi:10.1002/cpa.3160440402.

[15] Giuseppe Buttazzo and Filippo Santambrogio. "A model for the optimal planning of an urban area". *SIAM J. Math. Anal.*, 2005, vol. 37, no. 2, pp. 514–530, doi:10.1137/S0036141003438313.

[16] Luis A. Caffarelli. "Some regularity properties of solutions of Monge–Ampère equation". *Comm. Pure Appl. Math.*, 1991, vol. 44, no. 8-9, pp. 965–969, doi:10.1002/cpa.3160440809.

[17] ———. "The regularity of mappings with a convex potential". *J. Amer. Math. Soc.*, 1992, vol. 5, no. 1, pp. 99–104, doi:10.2307/2152752.

[18] ———. "Boundary regularity of maps with convex potentials II". *Ann. of Math. (2)*, 1996, vol. 144, no. 3, pp. 453–496, doi:10.2307/2118564.

[19] Guillaume Carlier, Alfred Galichon, and Filippo Santambrogio. "From Knothe's transport to Brenier's map and a continuation method for optimal transport". *SIAM J. Math. Anal.*, 2009/10, vol. 41, no. 6, pp. 2554–2576, doi:10.1137/080740647.

[20] Guillaume Carlier and Filippo Santambrogio. "A variational model for urban planning with traffic congestion". *ESAIM Control Optim. Calc. Var.*, 2005, vol. 11, no. 4, pp. 595–613, doi:10.1051/cocv:2005022.

[21] Dario Cordero-Erausquin. "Sur le transport de mesures périodiques". *C. R. Acad. Sci. Paris Sér. I Math.*, 1999, vol. 329, no. 3, pp. 199–202, doi:10.1016/s0764-4442(00)88593-6.

[22] Ennio De Giorgi. "New problems on minimizing movements". In *Selected papers*, edited by Luigi Ambrosio, Gianni Dal Maso, Marco Forti, Mario Miranda, and Sergio Spagnolo, pp. 699–713. Springer-Verlag, Berlin, 2006. First published in *Boundary Value Problems for Partial Differential Equations and Applications*, RMA Res. Notes Appl. Math., vol. 29, pp. 81–98, Masson, Paris, 1993.

[23] Philippe Delanoë. "Classical solvability in dimension two of the second boundary-value problem associated with the Monge-Ampère operator". *Ann. Inst. H. Poincaré Anal. Non Linéaire*, 1991, vol. 8, no. 5, pp. 443–457, doi:10.1016/j.anihpc.2007.03.001.

[24] R. Dobrushin. "Prescribing a system of random variables by conditional distributions". *Theor. Probab. Appl.*, 1970, vol. 15, no. 3, pp. 458–486, doi:10.1137/1115049.

[25] "When workers dream of a life beyond the factory gates". *The Economist*, December 15, 2012, vol. 405, no. 8815, http://www.economist.com/node/21568384.

[26] Ivar Ekeland. "An inverse function theorem in Fréchet spaces". *Ann. Inst. H. Poincaré Anal. Non Linéaire*, 2011, vol. 28, no. 1, pp. 91–105, doi:10.1016/j.anihpc.2010.11.001.

[27] Lawrence C. Evans and Wilfrid Gangbo. "Differential equations methods for the Monge–Kantorovich mass transfer problem". *Mem. Amer. Math. Soc.*, 1999, vol. 137, no. 653, pp. viii+66, doi:10.1090/memo/0653.

[28] Lawrence C. Evans and Ronald F. Gariepy. *Measure Theory and Fine Properties of Functions*. Studies in Advanced Mathematics. CRC Press, Boca Raton, FL, 1992.

[29] Alessio Figalli, Young-Heon Kim, and Robert J. McCann. "Regularity of optimal transport maps on multiple products of spheres". *J. Eur. Math. Soc. (JEMS)*, 2013, vol. 15, no. 4, pp. 1131–1166, doi:10.4171/jems/388.

[30] Alessio Figalli, Francesco Maggi, and Aldo Pratelli. "A mass transportation approach to quantitative isoperimetric inequalities". *Invent. Math.*, 2010, vol. 182, no. 1, pp. 167–211, doi:10.1007/s00222-010-0261-z.

[31] David Gilbarg and Neil S. Trudinger. *Elliptic Partial Differential Equations of Second Order*. Classics in Mathematics. Springer-Verlag, Berlin, 2001. Reprint of the 1998 edition.

[32] Kevin Guittet. "On the time-continuous mass transport problem and its approximation by augmented Lagrangian techniques". *SIAM J. Numer. Anal.*, 2003, vol. 41, no. 1, pp. 382–399 (electronic), doi:10.1137/S0036142901386069.

[33] Richard S. Hamilton. "The inverse function theorem of Nash and Moser". *Bull. Amer. Math. Soc. (N.S.)*, 1982, vol. 7, no. 1, pp. 65–222, doi:10.1090/s0273-0979-1982-15004-2.

[34] Richard Jordan, David Kinderlehrer, and Felix Otto. "The variational formulation of the Fokker–Planck equation". *SIAM J. Math. Anal.*, 1998, vol. 29, no. 1, pp. 1–17, doi:10.1137/s0036141096303359.

[35] Leonid V. Kantorovich. "On the translocation of masses". *J. Math. Sci.*, 2006, vol. 133, no. 4, pp. 1381–1382, doi:10.1007/s10958-006-0049-2. Translated from *Dokl. Akad. Nauk SSSR*, 1942, vol. 37, no. 7–8, pp. 227–229.

[36] ———. "On a problem of monge". *J. Math. Sci.*, 2006, vol. 133, no. 4, pp. 1383–1383, doi:10.1007/s10958-006-0050-9. Translated from *Uspekhi Mat. Nauk*, 1948, vol. 3, no. 2, 225–226.

[37] Jun Kitagawa. "A parabolic flow toward solutions of the optimal transportation problem on domains with boundary". *J. Reine Angew. Math.*, 2012, vol. 672, pp. 127–160.

[38] Herbert Knothe. "Contributions to the theory of convex bodies". *Michigan Math. J.*, 1957, vol. 4, pp. 39–52.

[39] Grégoire Loeper. "On the regularity of the polar factorization for time dependent maps". *Calc. Var. Partial Differential Equations*, 2005, vol. 22, no. 3, pp. 343–374, doi:10.1007/s00526-004-0280-y.

[40] Grégoire Loeper and Francesca Rapetti. "Numerical solution of the Monge–Ampère equation by a Newton's algorithm". *C. R. Math. Acad. Sci. Paris*, 2005, vol. 340, no. 4, pp. 319–324, doi:10.1016/j.crma.2004.12.018.

[41] Xi-Nan Ma, Neil S. Trudinger, and Xu-Jia Wang. "Regularity of potential functions of the optimal transportation problem". *Arch. Ration. Mech. Anal.*, 2005, vol. 177, no. 2, pp. 151–183, doi:10.1007/s00205-005-0362-9.

[42] Robert J. McCann. "Existence and uniqueness of monotone measure-preserving maps". *Duke Math. J.*, 1995, vol. 80, no. 2, pp. 309–323, doi:10.1215/s0012-7094-95-08013-2.

[43] ———. "Polar factorization of maps on riemannian manifolds". *Geom. Funct. Anal.*, 2001, vol. 11, no. 3, pp. 589–608, doi:10.1007/pl00001679.

[44] Vitali D. Milman and Gideon Schechtman. *Asymptotic Theory of Finite-Dimensional Normed Spaces.* Lecture Notes in Mathematics, vol. 1200. Springer-Verlag, Berlin, 1986. With an appendix by M. Gromov.

[45] Gaspard Monge. "Mémoire sur la théorie des déblais et des remblais". In *Histoire de l'Académie royale des sciences de Paris (année 1781)*, pp. 666–704. Imprimerie royale, 1784.

[46] Jürgen Moser. "A rapidly convergent iteration method and non-linear partial differential equations I". *Ann. Scuola Norm. Sup. Pisa (3)*, 1966, vol. 20, pp. 265–315.

[47] ———. "A rapidly convergent iteration method and non-linear partial differential equations II". *Ann. Scuola Norm. Sup. Pisa (3)*, 1966, vol. 20, pp. 499–535.

[48] John Nash. "The imbedding problem for riemannian manifolds". *Ann. of Math. (2)*, 1956, vol. 63, pp. 20–63.

[49] François Pitié, Anil C. Kokaram, and Rozenn Dahyot. "Automated colour grading using colour distribution transfer". *Computer Vision and Image Understanding*, 2007, vol. 107, no. 1–2, pp. 123–137, doi:10.1016/j.cviu.2006.11.011.

[50] Julien Rabin, Gabriel Peyré, Julie Delon, and Marc Bernot. "Wasserstein barycenter and its application to texture mixing". In *Scale Space and Variational Methods in Computer Vision*, edited by Alfred M. Bruckstein, Bart M. Haar Romeny, Alexander M. Bronstein, and Michael M. Bronstein, Lecture Notes in Computer Science, vol. 6667, pp. 435–446, doi:10.1007/978-3-642-24785-9_37. Springer Berlin Heidelberg, 2012.

[51] Murray Rosenblatt. "Remarks on a multivariate transformation". *Ann. Math. Statistics*, 1952, vol. 23, pp. 470–472.

[52] Yossi Rubner, Carlo Tomasi, and LeonidasJ. Guibas. "The earth mover's distance as a metric for image retrieval". *International Journal of Computer Vision*, 2000, vol. 40, no. 2, pp. 99–121, doi:10.1023/A:1026543900054.

[53] Ludger Rüschendorf. *Wasserstein Metric.* From http://www.stochastik.uni-freiburg.de/~rueschendorf/papers/wasserstein.pdf, 2005, Accessed: Aug. 30, 2013.

[54] Filippo Santambrogio, *Optimal Transport for Applied Mathematicans.* Lecture notes for a series of lessons at Université Paris-Sud, to be published.

[55] ———. "Inégalités isopérimétriques quantitatives via le transport optimal [d'après A. Figalli, F. Maggi et A. Pratelli]". *Astérisque*, 2012, no. 348, pp. 219–231. Séminaire Bourbaki 2010/2011, exposé 1034.

[56] ———. "Trois problèmes de transport à optimiser". *Paris-Sud Magazine*, 2012, vol. 82, p. 20.

[57] ———. *Flots de gradient dans les espaces métriques et leurs applications [d'après Ambrosio-Gigli-Savaré]*, January 2013, http://www.bourbaki.ens.fr/TEXTES/1065.pdf. Séminaire Bourbaki 2012/2013, exposé 1065.

[58] V. N. Sudakov. "Geometric problems in the theory of infinite-dimensional probability distributions". *Proc. Steklov Inst. Math.*, 1979, no. 2, pp. i–v, 1–178. Translated from *Trudy Mat. Inst. Steklov.*, 1976, vol. 141, pp. 1–191.

[59] Michel Talagrand. "Transportation cost for Gaussian and other product measures". *Geom. Funct. Anal.*, 1996, vol. 6, no. 3, pp. 587–600, doi:10.1007/BF02249265.

[60] John Urbas. "On the second boundary value problem for equations of Monge–Ampère type". *J. Reine Angew. Math.*, 1997, vol. 487, pp. 115–124, doi:10.1515/crll.1997.487.115.

[61] L. N. Vasershtein. "Markov processes over denumerable products of spaces describing large system of automata". *Probl. Inform. Transm.*, 1969, vol. 5, no. 3, pp. 47–52. Translated from *Problemy Peredači Informacii*, 1969, vol. 5, no. 3, pp. 64–72.

[62] Cédric Villani. *Topics in Optimal Transportation.* Graduate Studies in Mathematics, vol. 58. American Mathematical Society, Providence, RI, 2003.

[63] ———. *Optimal Transport, Old and New.* Grundlehren der Mathematischen Wissenschaften, vol. 338, doi:10.1007/978-3-540-71050-9. Springer-Verlag, Berlin, 2009.

Additional material, including scripts for numerical simulations, can be found on the author's website: http://www.normalesup.org/~bonnotte/thesis/

Parts of chapters 1 and 3 were published as "From Knothe's Rearrangement to Brenier's Optimal Transport Map", in SIAM *Journal on Mathematical Analysis*, vol. 45, no. 1, pp. 64–87. ©2013, Society for Industrial and Applied Mathematics

The author gratefully acknowledges the generous support of the French-Italian University toward his travel and accomodation expenses (Vinci grant C2-30)

UNIVERSITÉ FRANCO ITALIENNE | UNIVERSITÀ ITALO FRANCESE

AMS subject classification:
28A33, 35G25, 35K55, 35K90, 35J96, 47J07, 49J45, 49Q20, 49K21, 65M60

Key words:
Optimal transport, Knothe–Rosenblatt rearrangement, continuation methods, Nash–Moser inverse function theorem, Iterative Distribution Transfer algorithm, sliced Wasserstein distance, gradient flows