



HAL
open science

Contributions to the optimal transport problem and its regularity

Maxime Prod'Homme

► **To cite this version:**

Maxime Prod'Homme. Contributions to the optimal transport problem and its regularity. Analysis of PDEs [math.AP]. Université Paul Sabatier - Toulouse III, 2021. English. NNT : 2021TOU30122 . tel-03419872v2

HAL Id: tel-03419872

<https://theses.hal.science/tel-03419872v2>

Submitted on 8 Mar 2022

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

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



THÈSE

En vue de l'obtention du
DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par l'Université Toulouse 3 - Paul Sabatier

Présentée et soutenue par
Maxime PROD'HOMME

Le 27 octobre 2021

Contributions au problème du transport optimal et à sa régularité

Ecole doctorale : **EDMITT - Ecole Doctorale Mathématiques, Informatique et Télécommunications de Toulouse**

Spécialité : **Mathématiques et Applications**

Unité de recherche :

IMT : Institut de Mathématiques de Toulouse

Thèse dirigée par

Max FATHI et Felix OTTO

Jury

M. Djalil CHAFAÏ, Rapporteur

M. Filippo SANTAMBROGIO, Rapporteur

Mme Maria COLOMBO, Examinatrice

M. Radu IGNAT, Examineur

M. Max FATHI, Directeur de thèse

M. Felix OTTO, Co-directeur de thèse

Remerciements

Mes premiers remerciements vont à mes directeurs de thèse, Max Fathi et Felix Otto, sans qui cette thèse ne serait pas ce qu'elle est. Je mesure ma chance d'avoir pu travailler à leurs côtés et j'ai apprécié tout particulièrement leur disponibilité, leurs conseils avisés et les nombreuses discussions au cours desquelles j'ai beaucoup appris.

Je tiens également à remercier Djalil Chafaï et Filippo Santambrogio pour leur lecture attentive de cette thèse et leurs remarques bienvenues. Merci à Maria Colombo et à Radu Ignat d'avoir accepté de faire partie de mon jury, j'en suis très honoré.

Je remercie bien sûr les co-auteurs des articles constituant cette thèse. Outre Max et Felix, je remercie ici Nathaël Gozlan. *I would especially like to thank Tobias Ried for our numerous discussions in front of the blackboards of the Max Planck Institute ; it was always a pleasure to work with him.*

Étant peu expansif par nature, je serai bref dans la suite. Ma gratitude va à tous les chercheurs avec qui j'ai échangé au cours de ma thèse et qui m'ont permis d'avancer, aux doctorants et post-docs avec qui j'ai pu discuter à Leipzig ou à Toulouse, à mes amis qui m'ont soutenu et m'ont surtout sorti régulièrement de mes recherches, qualité nécessaire au vu du caractère parfois anxiogène d'une thèse. Merci également à ma famille pour les mêmes raisons. Enfin, je remercie tout spécialement Amélie, pour m'avoir supporté pendant ces 3 ans, dans le sens anglais comme dans le sens français du terme.

Résumé

Cette thèse s'intéresse au problème du transport optimal, en particulier aux propriétés de régularité d'une fonction réalisant un tel transport.

La première partie de ce manuscrit donne une nouvelle démonstration du théorème de contraction de Caffarelli, selon lequel le transport optimal de la gaussienne sur une mesure ayant une densité uniformément log-concave par rapport à la gaussienne est 1-Lipschitz. Au contraire des précédentes preuves de ce résultat, la stratégie employée ici s'appuie directement sur le problème de minimisation, ou plus précisément sur une régularisation entropique du transport optimal et une caractérisation variationnelle de la propriété Lipschitz du transport optimal démontrée par Gozlan et Juillet.

La deuxième partie étend l'approche variationnelle proposée par Goldman et Otto, permettant de prouver la régularité partielle de fonctions de transport optimales, à des fonctions de coût générales. La principale nouveauté est l'introduction, dans le contexte du transport optimal, de la notion de quasi-minimalité déjà bien établie dans le domaine des surfaces minimales. On montre en effet qu'un plan de transport optimal pour un coût proche du coût quadratique est quasi-optimal pour le coût quadratique, permettant alors résoudre le problème posé par l'absence d'équivalence de Benamou-Brenier pour un coût quelconque.

Enfin, on obtient dans la troisième et dernière partie une estimation optimale de la vitesse de convergence, en distances de Wasserstein W_1 et W_2 , de la mesure spectrale empirique de matrices de Ginibre vers la loi circulaire. Ceci s'appuie sur une estimation des fluctuations du nombre de points dans un domaine, grâce à la structure déterminantale du processus de points, et une construction itérative de mesures intermédiaires, de l'échelle microscopique à l'échelle macroscopique, permettant de relier distance W_2 et variance du nombre de points.

Mots-clés : Transport optimal, Régularité partielle, Monge-Ampère, Matrices aléatoires, Ensemble de Ginibre.

Abstract

This thesis deals with the optimal transport problem, in particular with regularity properties shared by optimal transport maps.

The first part of this manuscript provides a new proof of the Caffarelli contraction theorem, stating that the optimal transport map from the gaussian measure to a measure with a uniformly log-concave density with respect to the gaussian measure is 1-Lipschitz. The strategy developed here differs from the previous ones in the fact that it directly exploits the minimization problem, through an entropic regularization of optimal transport and a variational characterization of the Lipschitz property of an optimal transport map proved by Gozlan and Juillet.

The second part extends the variational approach, introduced by Goldman and Otto for the regularity theory of optimal transport maps, to the situation of general cost functions. The main contribution is the introduction, in the context of optimal transport, of the notion of almost-minimality, which is already well established in minimal surfaces. We show that an optimal transport plan for a cost function close to being quadratic is almost-optimal for the quadratic cost, allowing us to solve the issue raised by the non-existence of a Benamou-Brenier equivalence for a general cost function.

In the third part, we obtain the optimal rate of convergence, in W_1 and W_2 Wasserstein distance, of the empirical spectral measure of Ginibre matrices to the circular law. This relies on an estimation of the fluctuations of the number of points in a domain, using the determinantal structure of the point process, and an iterative construction of intermediate measures, from microscopic to macroscopic scales, which allows to link the W_2 distance to the variance of the number of points.

Key words: Optimal transport, Partial regularity, Monge-Ampère, Random matrices, Ginibre ensemble.

Table des matières

Introduction	5
1 A proof of the Caffarelli contraction theorem via entropic regularization	41
2 Variational approach to the regularity of optimal transport maps: general cost functions	57
3 Vitesse de convergence optimale vers la loi circulaire pour l'ensemble de Ginibre	125
Références bibliographiques	161

Introduction

De façon générale, la présente thèse s'intéresse au problème du transport optimal. Ce chapitre introductif s'attachera d'abord à présenter les notions essentielles du transport optimal, en particulier sa structure et ce que l'on peut dire de sa régularité. On s'attardera ensuite sur un sous-problème, celui de l'appariement optimal entre mesures empiriques de points, qu'ils soient i.i.d. ou qu'ils représentent les valeurs propres de matrices aléatoires.

La dernière section consiste en une présentation des résultats obtenus, ceux-ci constituant le reste du manuscrit. Le chapitre 1 reprend l'article [FGP20] et donne une preuve variationnelle du théorème de contraction de Caffarelli. Le chapitre 2 est basé sur l'article [OPR21], où l'on étend l'approche variationnelle de [GO20] à des coûts généraux. Enfin, le chapitre 3 s'intéresse au comportement de la mesure spectrale des matrices de Ginibre, donnant en particulier une estimation optimale de la vitesse de convergence vers la loi circulaire.

Sommaire

I	Le problème du transport optimal	6
I.1	Coût quadratique	11
I.2	Coût général	16
II	Régularité du transport optimal	18
II.1	L'équation de Monge-Ampère	18
II.2	Régularité partielle	23
II.3	Coût général	24
II.4	Régularité quantitative	26
III	Le problème d'appariement optimal	27
III.1	Mesures empiriques de points i.i.d.	28
III.2	Mesures spectrales empiriques	30
IV	Contributions de la thèse	32
IV.1	Une preuve variationnelle du théorème de contraction de Caffarelli	32

IV.2	Généralisation de l'approche variationnelle de [GO20] à des coûts généraux	34
IV.3	Vitesse de convergence optimale vers la loi circulaire pour l'ensemble de Ginibre	37

I Le problème du transport optimal

Posée pour la première fois par Monge en 1781, la question était la suivante : étant donné un tas de sable avec une certaine forme, comment déplacer chaque grain de sable pour former un autre tas ayant une configuration prescrite, de façon à minimiser la distance moyenne de transport ? En termes mathématiques, étant données deux distributions de masse dans \mathbb{R}^d (dans le problème de Monge, la dimension d de l'espace ambiant est 2 ou 3 mais on autorise ici une dimension quelconque), modélisées par deux densités $f : \mathbb{R}^d \supset X \rightarrow \mathbb{R}_+$ et $g : \mathbb{R}^d \supset Y \rightarrow \mathbb{R}_+$ ayant la même masse totale, on cherche une fonction $T : X \rightarrow Y$ transportant f sur g et telle que la distance totale $\int_X |T(x) - x|f(x) dx$ soit minimale.¹

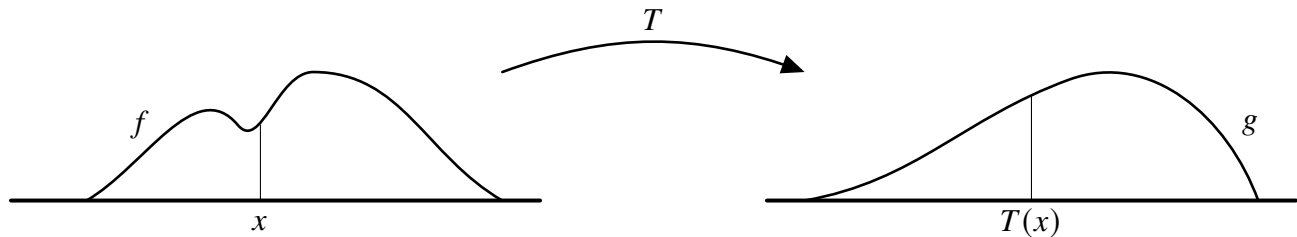


FIGURE 1: Transport entre deux densités.

On peut généraliser cette question à des distributions données par des mesures positives de même masse sur \mathbb{R}^d , ainsi qu'à d'autres façons de calculer le coût du transport de x à y . Ainsi, si μ , respectivement ν , est une mesure sur X , respectivement Y , avec $\mu(X) = \nu(Y)$, et si $c : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}_+$ est une fonction telle que $c(x, y)$ représente le coût de déplacement de x à y , existe-t-il une fonction de transport T telle que $T_{\#}\mu = \nu$ et vérifiant

$$\int_X c(x, T(x)) \mu(dx) = \inf_{S_{\#}\mu = \nu} \int_X c(x, S(x)) \mu(dx) \quad ? \tag{I.1}$$

La mesure $T_{\#}\mu$, appelée *push-forward* de μ par T , est une mesure sur Y définie par

$$(T_{\#}\mu)(A) = \mu(T^{-1}(A)),$$

ou de façon équivalente par la propriété que, pour toute fonction ζ mesurable,

$$\int_Y \zeta d(T_{\#}\mu) = \int_X \zeta \circ T d\mu.$$

¹Dans l'ensemble du manuscrit, $|a|$ désigne la norme euclidienne du vecteur a et si A est un ensemble mesurable, $|A|$ désigne la mesure de Lebesgue de A .

On retrouve au passage dans cette définition le fait que μ et $T_{\#}\mu$ ont la même masse totale. Dans la suite, on normalise cette masse commune à 1 : μ et ν sont des mesures de probabilité, dont on note l'ensemble $\mathcal{P}(\mathbb{R}^d)$.

Le premier obstacle à l'existence d'un tel *transport optimal* est l'absence possible de fonction de transport. En effet, prenons l'exemple d'une mesure de Dirac δ_a comme distribution de départ. Dans ce cas, toute fonction S transporte δ_a sur la mesure de Dirac $\delta_{S(a)}$, de sorte qu'il n'existe de fonction de transport que si la mesure d'arrivée est elle-même un Dirac. Le deuxième obstacle est bien sûr la possibilité que l'infimum dans (I.1) ne soit pas atteint, ce qui peut arriver lorsque la mesure de départ donne une mesure non nulle à des ensembles de dimension strictement inférieure à d . Voir [San15, Sec. 1.4] où l'exemple du transport entre les mesures $\mu := \mathcal{H}^1 \llcorner_{\{0\} \times [0;1]}$ et $\nu := \frac{1}{2} \mathcal{H}^1 \llcorner_{\{-1\} \times [0;1]} + \frac{1}{2} \mathcal{H}^1 \llcorner_{\{1\} \times [0;1]}$ dans \mathbb{R}^2 est détaillé.

Suivant les idées de Kantorovitch, remarquons qu'une *fonction de transport* T est un cas particulier de *plan de transport*, c'est-à-dire une mesure π sur $X \times Y$ qui forme un couplage entre μ et ν dans le sens où la première marginale (d -dimensionnelle) de π est μ et sa seconde marginale est ν , ou encore : pour toute fonction $\phi : \mathbb{R}^d \rightarrow \mathbb{R}$,

$$\int_{X \times Y} \phi(x) \pi(\mathrm{d}x \mathrm{d}y) = \int_X \phi \mathrm{d}\mu \quad \text{et} \quad \int_{X \times Y} \phi(y) \pi(\mathrm{d}x \mathrm{d}y) = \int_Y \phi \mathrm{d}\nu.$$

En effet, en notant $\pi_T = (\mathrm{Id} \times T)_{\#}\mu$ le couplage induit par T , on a

$$\int_X c(x, T(x)) \mu(\mathrm{d}x) = \int_{X \times Y} c(x, y) \pi_T(\mathrm{d}x \mathrm{d}y).$$

L'idée est alors de relaxer le problème de minimisation (I.1) en autorisant tous les couplages entre μ et ν , pas seulement ceux induits par une fonction de transport. On arrive au problème suivant : étant données deux mesures μ et ν et une fonction c , on cherche un couplage π tel que

$$\int_{X \times Y} c \mathrm{d}\pi = \inf_{\gamma \in \Pi(\mu, \nu)} \int_{X \times Y} c \mathrm{d}\gamma, \tag{I.2}$$

où $\Pi(\mu, \nu)$ désigne l'ensemble des couplages entre μ et ν .

D'un point de vue physique, le problème (I.2), à la différence du problème (I.1), permet de séparer la masse. En effet, si on a une fonction de transport T alors toutes les particules présentes en position x vont au même endroit, donné par $T(x)$. Le transport est dit *déterministe*. Avec un couplage $\pi \in \Pi(\mu, \nu)$, il est possible d'envoyer des particules partant de la même position x en plusieurs endroits différents, dont la distribution est donnée par la désintégration de π par rapport à μ : le théorème de désintégration permet d'écrire $\pi(\mathrm{d}x \mathrm{d}y) = \pi_x(\mathrm{d}y) \mu(\mathrm{d}x)$, ce qui indique que la masse provenant de x est redistribuée selon π_x (remarquons que la désintégration de π_T donne $(\pi_T)_x = \delta_{T(x)}$). Dit autrement, $\pi(A \times B)$ est la quantité de matière passant de l'ensemble A à l'ensemble B .

Mathématiquement, le problème (I.2) est beaucoup plus simple à résoudre que (I.1). D'une part, la fonctionnelle à minimiser est linéaire par rapport au couplage et d'autre part, l'ensemble $\Pi(\mu, \nu)$ est convexe ; on est donc face à un problème d'optimisation convexe. Remarquons également que cette fois, l'ensemble sur lequel on optimise n'est jamais vide et qu'il est compact pour la topologie de la convergence étroite, par application du théorème de Prokhorov. Une application de la méthode directe du calcul des variations nous donne alors notre premier résultat d'existence (voir par exemple [San15, Th. 1.7]).

Théorème I.1. Soient $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$ et $c : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}_+$ semi-continue inférieurement. Alors (I.2) admet une solution.

Distance sur l'espace des mesures de probabilité

Un intérêt théorique important du transport optimal est qu'il fournit une famille de distances "naturelles" sur l'espace $\mathcal{P}(\mathbb{R}^d)$. Plus précisément, si $p \geq 1$, le coût optimal de transport entre deux mesures avec fonction de coût $c(x, y) := |y - x|^p$ vérifie les axiomes d'une distance.

Proposition I.1 (Distances de Wasserstein W_p). Soit $1 \leq p < \infty$, soient μ et ν deux mesures ayant chacun un moment d'ordre p fini, ce que l'on notera $\mu, \nu \in \mathcal{P}_p(\mathbb{R}^d)$. Définissons

$$W_p(\mu, \nu) := \inf \left\{ \int |y - x|^p d\pi \mid \pi \in \Pi(\mu, \nu) \right\}^{\frac{1}{p}}.$$

Alors W_p est une distance sur $\mathcal{P}_p(\mathbb{R}^d)$.

Exemple I.2. Un cas très simple où la distance de Wasserstein est calculable est celle de deux mesures de Dirac :

$$W_p(\delta_x, \delta_y) = |x - y|.$$

La topologie induite par W_p est celle de la convergence faible des mesures, si on se place sur $\mathcal{P}_p(\Omega)$ avec Ω un compact. Sur \mathbb{R}^d , on a l'équivalence

$$W_p(\mu_n, \mu) \rightarrow 0 \iff \left(\mu_n \rightharpoonup \mu \quad \text{et} \quad \int |x|^p d\mu_n \rightarrow \int |x|^p d\mu \right).$$

De plus, $(\mathcal{P}_p(\mathbb{R}^d), W_p)$ est complet (voir par exemple [AGS08, Prop. 7.1.5]).

La distance W_1 admet une formulation duale qui prend la forme de la distance de Kantorovich-Rubinstein ([Vil03, Th. 1.14]),

$$W_1(\mu, \nu) = \sup_{|\psi|_{\text{Lip}} \leq 1} \int \psi d(\mu - \nu),$$

où $|\psi|_{\text{Lip}}$ est définie par

$$|\psi|_{\text{Lip}} = \sup_{x \neq y} \frac{|f(x) - f(y)|}{|x - y|}.$$

On verra au cours de cette introduction que l'espace $(\mathcal{P}_2(\mathbb{R}^d), W_2)$ peut être muni d'une structure géométrique particulièrement intéressante.

Transport déterministe dans un cadre discret

Après le résultat positif donné par le Théorème I.1, le but est de montrer que, sous certaines conditions sur μ et ν ainsi que sur c , le plan de transport optimal est en fait induit par une fonction de transport. Examinons cela dans un cadre simplifié.

Exemple I.3. Etudions ici un problème d'appariement optimal (cf. section III), c'est-à-dire le cas où les deux mesures μ et ν sont discrètes et ont le même nombre N d'atomes, chacun de ces atomes ayant le même poids $\frac{1}{N}$:

$$\mu = \frac{1}{N} \sum_{i=1}^N \delta_{x_i} \quad \text{et} \quad \nu = \frac{1}{N} \sum_{j=1}^N \delta_{y_j}.$$

Un couplage π entre μ et ν est alors représenté par une matrice $(\pi_{ij})_{1 \leq i, j \leq N}$ où $0 \leq \pi_{ij} \leq 1$ donne la portion de masse allant de x_i à y_j . Toute la masse de μ doit être relocalisée et doit former toute la masse de ν , ce qui implique que π est bistochastique :

$$\sum_{k=1}^N \pi_{kj} = \sum_{l=1}^N \pi_{il} = 1 \quad \forall (i, j).$$

Notre problème de transport optimal devient donc

$$\min \left\{ \sum_{i,j=1}^N c(x_i, y_j) \pi_{ij} \mid \pi \text{ bistochastique} \right\}.$$

Comme on l'a vu plus haut, c'est un problème d'optimisation linéaire avec contraintes convexes donc l'infimum est atteint en un point extrémal de l'ensemble des matrices bistochastiques. D'après le théorème de Birkhoff-von Neumann, un tel point extrémal est une matrice de permutation. On obtient donc un minimiseur π induit par une fonction de transport T . Il n'y a donc pas d'intérêt à séparer la masse de μ , il suffit de relier les points deux à deux de façon optimale. Remarquons également qu'il n'y a pas forcément unicité du transport optimal dans ce cadre discret. Par exemple si μ et ν sont constituées de deux points chacune de telle façon que les deux atomes de μ sont placés sur la médiatrice du segment reliant les deux atomes de ν , voir la Figure 2, alors les deux fonctions possibles sont optimales dès que le coût est par exemple fonction de la distance.



FIGURE 2: Non-unicité du transport optimal.

Dans ce cadre simple, peut-on exhiber une condition nécessaire et suffisante pour que T soit optimale ? Remarquons que toute autre fonction de transport \tilde{T} de μ à ν peut être déduite de T en permutant les indices : $\tilde{T}(x_i) = T(x_{\sigma(i)})$ pour une certaine permutation $\sigma \in \mathfrak{S}_N$. Le transport T est donc optimal si et seulement si

$$\sum_{i=1}^N c(x_i, T(x_i)) \leq \sum_{i=1}^N c(x_i, T(x_{\sigma(i)})) \quad \forall \sigma \in \mathfrak{S}_N,$$

ou, présenté autrement : pour tout $k \geq 1$, pour tous $(x'_1, \dots, x'_k) \in \{x_1, \dots, x_N\}^k$, en notant $y'_i := T(x'_i) \in \{y_1, \dots, y_N\}$, on a

$$\sum_{i=1}^k c(x'_i, y'_i) \leq \sum_{i=1}^k c(x'_i, y'_{\sigma(i)}) \quad \forall \sigma \in \mathfrak{S}_k.$$

Cette condition nécessaire et suffisante d'optimalité découle ici de notre connaissance préalable de la structure déterministe du plan de transport optimal. Pourtant, on va voir dans la suite que cette condition, que l'on appelle *c-monotonie cyclique* est encore nécessaire et suffisante pour qu'un plan de transport π entre deux mesures de probabilité quelconques soit optimal. C'est grâce à cette propriété que l'on va prouver dans certains cas que π est déterministe.

Aspects numériques

D'un point de vue numérique, le problème d'appariement optimal présenté dans l'exemple I.3 peut être résolu par l'algorithme hongrois, ou l'algorithme des enchères. Plus généralement, le transport optimal entre deux mesures discrètes est un problème adapté à des méthodes d'optimisation linéaire de type simplexe ou points intérieurs, voir [PC19, Section 3] ou [San15, Section 6.4] pour plus de détails. Cependant, ces techniques ont toutes une complexité au mieux cubique par rapport au nombre de points. Une autre approche a été proposée dans [Cut13], permettant une amélioration substantielle de la rapidité de calcul. L'idée est de minimiser la fonctionnelle de (I.2) pénalisée par un terme entropique que l'on peut choisir petit. Ceci rend le problème strictement convexe et permet l'application de l'algorithme de Sinkhorn, ou IPFP, convergeant très rapidement. On va également utiliser cet algorithme à des fins théoriques dans le chapitre 1 de cette thèse.

Applications du transport optimal

L'économie est un domaine d'application historique du transport optimal puisque c'est dans ce cadre que Kantorovitch (et d'autres mathématiciens) a entre autres développé les techniques de résolution du problème (I.2). Beaucoup de problèmes en économie ou logistique sont formulés en terme de couplage : allocation de ressources, planification de production, appariement entre salariés et emplois de façon à ce que les compétences de chacun correspondent au mieux à l'emploi et que les salaires soient optimisés, le problème principal-agent... On renvoie à [Gal16] pour une présentation des applications modernes du transport optimal en économie. Un deuxième domaine d'application majeur, bien que plus récent, se trouve en science des données, que ce soit le traitement d'images, de données, l'interpolation de formes, la classification supervisée et l'adaptation de domaine en machine learning, voir [KPTSR17, PC19] pour ces notions et beaucoup d'autres applications. On verra également dans la suite de cette introduction des liens avec la mécanique des fluides. Enfin, mentionnons une application théorique importante à la géométrie riemannienne, à savoir la théorie développée indépendamment par Lott et Villani dans [LV09] et Sturm dans [Stu06a, Stu06b] pour donner une notion de courbure de Ricci minorée à des espaces métriques généraux via la convexité de certaines fonctionnelles le long des géodésiques définies par le transport optimal.

I.1 Coût quadratique

On va maintenant focaliser notre présentation sur le cas d'un coût fonction quadratique de la distance euclidienne, c'est-à-dire $c(x, y) = \frac{|x-y|^2}{2}$, ou de façon équivalente $c(x, y) = -x \cdot y$, puisque ce coût donne exactement les mêmes minimiseurs. Ce coût quadratique est le plus utilisé dans la littérature, du fait de la simplicité des résultats associés et de la riche structure que confère la distance W_2 à l'espace des mesures de probabilité. Voyons d'abord ce que la notion de c -monotonie cyclique devient dans ce cadre : un ensemble $\Gamma \subset \mathbb{R}^d \times \mathbb{R}^d$ est dit *cycliquement monotone* si pour tout $k > 0$, tous k -uplets (x_1, \dots, x_k) et (y_1, \dots, y_k) tels que $(x_i, y_i) \in \Gamma$ pour tout i , on a

$$\sum_{i=1}^k |x_i - y_i|^2 \leq \sum_{i=1}^k |x_i - y_{\sigma(i)}|^2 \quad \forall \sigma \in \mathfrak{S}_k,$$

ce qui est équivalent à

$$\sum_{i=1}^k x_i \cdot (y_i - y_{\sigma(i)}) \geq 0 \quad \forall \sigma \in \mathfrak{S}_k.$$

On comprend avec ce cas particulier l'emploi du terme "monotonie". En particulier, cette définition implique que pour deux couples $(x_1, y_1), (x_2, y_2) \in \Gamma$, le vecteur reliant x_1 à y_1 ne peut pas croiser celui reliant x_2 à y_2 . Si ces vecteurs représentent un transport de masse de x_i à y_i , on s'attend à ce qu'un transport optimal vérifie cette propriété. La Figure 3 présente deux transports entre deux ensembles de 4 points, celui de droite est cycliquement monotone et optimal, celui de gauche n'est pas monotone. En fait, la propriété de monotonie est plus

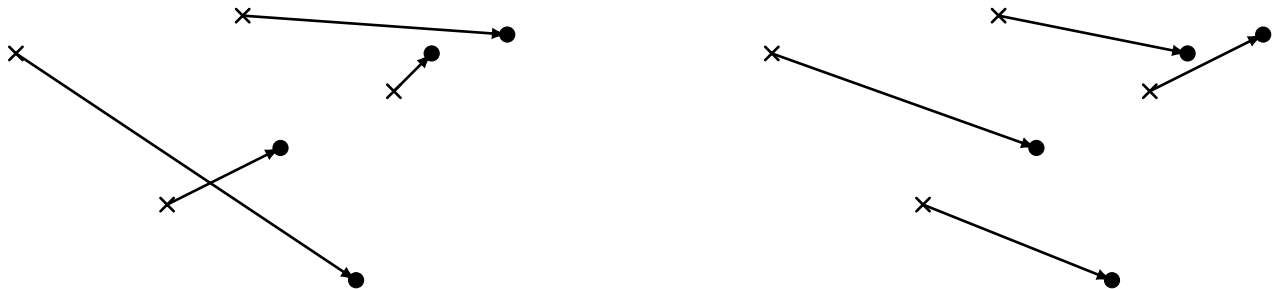


FIGURE 3: Appariement entre deux ensembles de 4 points.

forte que celle de non-croisement, comme le montre le coin supérieur droit de la Figure 3. Dans ce coin, on transporte deux points sur deux autres points, les deux transports possibles donnent des trajectoires qui ne se croisent pas, pourtant le transport de droite est monotone et optimal alors que celui de gauche n'est pas monotone donc pas optimal.

Proposition I.4. *Si π est un plan de transport optimal pour le coût quadratique alors le support de π ($\text{Spt } \pi$) est cycliquement monotone.*

Dans le cas où μ et ν ont des moments d'ordre 2 finis, la réciproque est vraie : si $\pi \in \Pi(\mu, \nu)$ a un support cycliquement monotone alors π est optimal. Combiné à un théorème de Rockafellar [Roc66] caractérisant les ensembles cycliquement monotones comme étant ceux

inclus dans le graphe de la sous-différentielle d'une fonction convexe propre et semi-continue inférieurement u , la Proposition I.4 induit que si μ est absolument continue par rapport à la mesure de Lebesgue, de façon à ce que l'ensemble des points où u n'est pas différentiable soit de μ -mesure nulle, alors $\text{Spt } \pi$ est inclus dans le graphe du gradient d'une fonction convexe. On obtient donc en partie le résultat suivant.

Théorème I.2 (Brenier [Bre87], McCann [McC95]). *Soient $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$ avec $\mu \ll \mathcal{L}^d$.*

Si μ et ν ont des moments d'ordre 2 finis, il existe une unique fonction T telle que $T\#\mu = \nu$ et vérifiant (I.1), avec $c(x, y) = \frac{|x-y|^2}{2}$. De plus, T est le gradient d'une fonction convexe : $T = \nabla u$.

Si le moment d'ordre 2 de μ ou ν est infini, il existe un unique gradient ∇u d'une fonction convexe telle que $\nabla u\#\mu = \nu$.

Enfin, si on ajoute l'hypothèse $\nu \ll \mathcal{L}^d$ alors ∇u^ est l'unique transport optimal, en coût quadratique, de ν à μ , où u^* est la transformée de Legendre de u . En particulier, $\nabla u \circ \nabla u^* = \text{Id}$ presque partout.*

On dira que T dérive d'un *potentiel* convexe u . La réciproque est vraie, dans le sens où le gradient d'une fonction convexe qui transporte μ sur ν est optimal pour le coût quadratique.

Exemple I.5. En dimension 1, la propriété $T = u'$ avec u convexe revient à dire que T est une application croissante. En fait, on peut montrer directement que si $\mu \in \mathcal{P}(\mathbb{R})$ n'a pas d'atome alors $T := \mathcal{F}_\nu^{-1} \circ \mathcal{F}_\mu$ est l'unique fonction croissante transportant μ sur ν , où \mathcal{F}_λ est la fonction de répartition de la mesure λ et \mathcal{F}_λ^{-1} sa pseudo-inverse. Cette fonction est donc le transport optimal en coût quadratique de μ à ν . Elle est également optimale pour d'autres coûts, en particulier tous ceux de la forme $c(x, y) = h(|x - y|)$ avec h strictement convexe.



FIGURE 4: Transport croissant en dimension 1.

Dualité de Kantorovitch

Le potentiel convexe u est lié à une formulation duale du problème de transport optimal. Plus précisément, la contrainte $\pi \in \Pi(\mu, \nu)$ se réécrit

$$\sup \left\{ \int \phi \, d\mu + \int \psi \, d\nu - \int (\phi(x) + \psi(y)) \, d\pi \mid \phi, \psi \text{ continues bornées} \right\} = 0,$$

car cette quantité est infinie si $\pi \notin \Pi(\mu, \nu)$. En injectant cette contrainte dans notre problème de minimisation et en intervertissant inf et sup, on obtient un problème dual équivalent à (I.1) :

$$\sup \left\{ \int \phi \, d\mu + \int \psi \, d\nu \mid \phi \oplus \psi \leq \frac{|x-y|^2}{2}, \quad \phi \in L^1(\mu), \quad \psi \in L^1(\nu) \right\}, \quad (\text{I.3})$$

où $\phi \oplus \psi(x, y) = \phi(x) + \psi(y)$. Le couple (ϕ, ψ) optimisant (I.3) est formé d'une fonction c -concave ϕ et de sa c -conjuguée ψ , dans le sens où $\psi(y) = \inf_{x \in \mathbb{R}^d} \left\{ \frac{|y-x|^2}{2} - \phi(x) \right\}$ (c'est-à-dire que ψ est l'inf-convolution de $-\phi$ et $\frac{|\cdot|^2}{2}$). La fonction ϕ est alors le *potentiel de Kantorovitch* associé au problème. En particulier, on a $u(x) := \frac{|x|^2}{2} - \phi(x)$ convexe et semi-continue inférieurement et le transport optimal s'écrit alors $T(x) = x - \nabla \phi(x)$.

Formulation eulérienne du transport optimal et formule de Benamou-Brenier

Prenons l'exemple d'un couplage déterministe, donné par une fonction de transport T . Une fois que l'on connaît T , c'est-à-dire que l'on sait comment la masse est redistribuée, on peut représenter ce couplage de façon dynamique par des segments reliant chacun une position initiale x à sa destination $T(x)$. On définit ainsi une famille de fonctions

$$T_t(x) := (1-t)x + tT(x) \quad \text{pour } t \in [0, 1].$$

On a ainsi une vision *lagrangienne* du transport : on suit au cours du temps chaque particule x le long de sa trajectoire $(T_t(x))_{t \in [0,1]}$. Comme en mécanique des fluides, on peut passer à une description *eulérienne* du même phénomène, introduite dans le cadre du transport optimal par Benamou et Brenier dans [BB99]. En effet, définissons la famille de mesures

$$\rho_t := T_t \# \mu \quad \text{pour } t \in [0, 1].$$

ρ_t est la distribution de masse à l'instant t , on a en particulier $\rho_0 = \mu$ et $\rho_1 = \nu$. Pour que la description soit complète, il faut donner le champ de vitesse à l'instant t . Or le trajet de x à $T(x)$, de $t = 0$ à $t = 1$, se fait à vitesse constante $T(x) - x$, ce qui nous donne

$$v_t(T_t(x)) = T(x) - x, \quad \forall t \in [0, 1].$$

On peut facilement vérifier que la paire (ρ, ν) vérifie, au sens des distributions, l'équation de continuité

$$\partial_t \rho + \nabla \cdot (v\rho) = 0,$$

où $(v\rho)_t$ est la mesure $j_t(dx) = v_t(x)\rho_t(dx)$, qui correspond au *flux*.

Plus généralement, si $\pi \in \Pi(\mu, \nu)$, on introduit la densité dynamique

$$\rho_t := X_t \# \pi \quad \text{où } X_t(x, y) := (1-t)x + ty, \tag{I.4}$$

ce qui se réécrit

$$\int \zeta(x) \rho_t(dx) = \int \zeta((1-t)x + ty) \pi(dx dy) \quad \forall \zeta \in C_c^0(\mathbb{R}^d).$$

Le flux est alors la mesure j_t définie par

$$\int \xi(x) \cdot j_t(dx) = \int \xi((1-t)x + ty) \cdot (y-x) \pi(dx dy) \quad \forall \xi \in C_c^0(\mathbb{R}^d)^d. \tag{I.5}$$

La paire densité-flux (ρ, j) est solution, au sens des distributions, de l'équation de continuité

$$\partial_t \rho + \nabla \cdot j = 0, \tag{I.6}$$

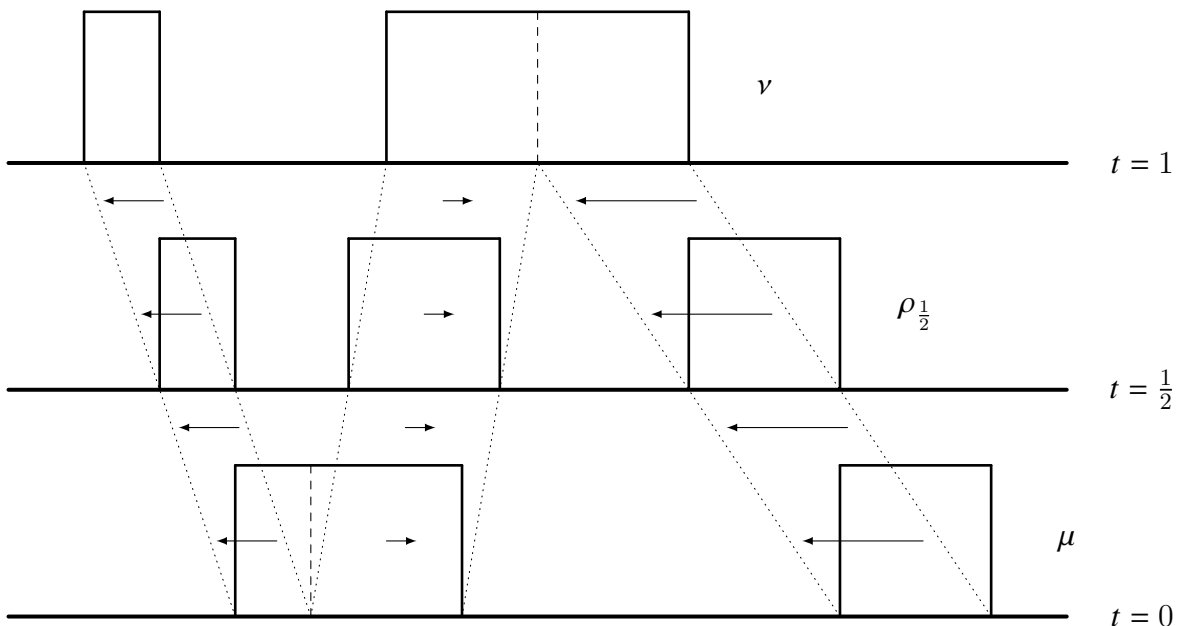


FIGURE 5: Description eulérienne du transport.

c'est-à-dire que pour tout $\zeta \in C_c^1(\mathbb{R}^d \times [0, 1])$,

$$\int_0^1 \left(\int \partial_t \zeta \, d\rho_t + \nabla \zeta \cdot dj_t \right) dt = \int \zeta_1 \, d\nu - \int \zeta_0 \, d\mu. \quad (\text{I.7})$$

Introduisons l'énergie cinétique associée à l'équation de continuité (I.6) ([AFP00, Eq. 2.26]) :

$$\int \frac{1}{\rho} |j|^2 := \begin{cases} \int_0^1 \int \left| \frac{dj_t}{d\rho_t}(x) \right|^2 \rho_t(dx) \, dt & \text{si } j \ll \rho, \\ +\infty & \text{sinon.} \end{cases}$$

Dans [BB99] (voir aussi [Vil03, Th. 8.1], [AGS08, Chap. 8]), Benamou et Brenier ont prouvé que la formulation lagrangienne du transport optimal avec le coût quadratique, consistant à minimiser la longueur quadratique moyenne des trajectoires, est bien équivalente à sa version eulérienne, qui consiste à minimiser l'énergie cinétique globale de la paire (ρ, j) :

$$\inf_{\pi \in \Pi(\mu, \nu)} \int |y - x|^2 \pi(dx dy) = \inf \left\{ \int \frac{1}{\rho} |j|^2 \mid \partial_t \rho + \nabla \cdot j = 0, \quad \rho_0 = \mu, \quad \rho_1 = \nu \right\}. \quad (\text{I.8})$$

En particulier, la paire (ρ, j) minimisant le membre de droite est bien la paire densité-flux associée au minimiseur du membre de gauche. Cette formulation équivalente conduit à une méthode numérique de calcul du transport optimal, voir [BB99] et [San15, Chap. 6].

Structure géométrique de l'espace $(\mathcal{P}_2(\mathbb{R}^d), W_2)$

Dans la version dynamique du transport, on étudie des courbes dans l'espace de mesures $\mathcal{P}(\mathbb{R}^d)$. Si μ et ν sont dans $\mathcal{P}_p(\mathbb{R}^d)$, la courbe $(\rho_t)_{t \in [0,1]}$ associée au plan de transport optimal

π par l'équation (I.4) définit en fait une géodésique à vitesse constante dans l'espace métrique $(\mathcal{P}_p(\mathbb{R}^d), W_p)$, dans le sens où, pour tous $t, s \in [0, 1]$,

$$W_p(\rho_t, \rho_s) = |t - s|W_p(\mu, \nu).$$

L'espace $(\mathcal{P}_p(\mathbb{R}^d), W_p)$ est alors un espace géodésique : pour tous $\mu, \nu \in \mathcal{P}_p(\mathbb{R}^d)$, il existe une géodésique à vitesse constante qui les relie. De plus, une courbe dans $\mathcal{P}_p(\mathbb{R}^d)$ est une géodésique à vitesse constante de $(\mathcal{P}_p(\mathbb{R}^d), W_p)$ si et seulement si elle est de la forme (I.4) pour une mesure $\pi \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$ optimale pour le coût $|x - y|^p$, en tant que couplage entre ses d -marginales.

Même si l'équivalence (I.8) peut être généralisée aux coûts $|x - y|^p$ pour tout $p > 1$, c'est dans le cas euclidien $p = 2$ que la formule est la plus riche de conséquences. En effet, réécrivons cette équivalence de la façon suivante :

$$W_2^2(\mu, \nu) = \inf \left\{ \int_0^1 \|v_t\|_{L^2(\rho_t)}^2 dt \mid \partial_t \rho_t + \nabla \cdot (v_t \rho_t) = 0, \quad \rho_0 = \mu, \quad \rho_1 = \nu \right\}.$$

Une telle formule ressemble fortement à la définition d'une métrique riemannienne, c'est-à-dire

$$d^2(x, y) = \inf \left\{ \int_0^1 \|\gamma'(t)\|_{\gamma(t)}^2 dt \mid \gamma(0) = x, \quad \gamma(1) = y \right\},$$

où $\|\cdot\|_m$ est la norme associée au produit scalaire sur l'espace tangent au point m . Il apparaît alors que l'on peut munir $(\mathcal{P}_2(\mathbb{R}^d), W_2)$ d'une structure riemannienne formelle : en définissant la norme

$$\|\partial_t \rho_t\|_{\rho_t} = \inf \left\{ \|v_t\|_{L^2(\rho_t)} \mid \partial_t \rho_t + \nabla \cdot (v_t \rho_t) = 0 \right\}, \quad (\text{I.9})$$

on a

$$W_2^2(\mu, \nu) = \inf \left\{ \int_0^1 \|\partial_t \rho_t\|_{\rho_t}^2 dt \mid \rho_0 = \mu, \quad \rho_1 = \nu \right\}.$$

L'infimum dans (I.9) est atteint en un unique $v_t \in L^2(\rho_t)$ caractérisé par

$$\langle v_t, w \rangle_{L^2(\rho_t)} = 0, \quad \forall w \in L^2(\rho_t), \quad \nabla \cdot (w \rho_t) = 0.$$

On peut donc représenter chaque vecteur tangent $\partial_t \rho_t$ en ρ_t par un unique $v_t \in L^2(\rho_t)$ de norme minimale parmi les champs de vecteurs compatibles avec l'équation de continuité. L'espace tangent formel en un point $\rho \in \mathcal{P}_2(\mathbb{R}^d)$ est donc

$$\text{Tan}_\rho \mathcal{P}_2(\mathbb{R}^d) := \{v \in L^2(\rho) \mid \langle v, w \rangle_{L^2(\rho)} = 0, \quad \forall w \in L^2(\rho), \quad \nabla \cdot (w \rho) = 0\},$$

muni du produit scalaire usuel sur $L^2(\rho)$. Cette structure riemannienne formelle a été identifiée pour la première fois par Otto dans [Ott01], où il développe à partir de là des règles de différentiation formelle de certaines fonctionnelles sur l'espace de mesures $\mathcal{P}_2(\mathbb{R}^d)$, voir également [Lot08] et [Gig12], ainsi que [Vil09, Chap. 15].

I.2 Coût général

Une partie de cette thèse s'inscrit dans un contexte plus général que le coût quadratique. Voyons ici comment s'adaptent les résultats précédents si $c : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}_+$ est maintenant une fonction de coût quelconque.

Le premier résultat que l'on peut adapter directement est celui de c -monotonie cyclique du support d'un plan de transport optimal ([Rüs96]).

Proposition I.6. *Si c est continue et π est optimal pour le coût c alors $\text{Spt } \pi$ est c -cycliquement monotone.*

La réciproque est vraie si le coût est continu en prenant éventuellement la valeur $+\infty$ ([Pra08]) et si le coût est semi-continu inférieurement à valeurs finies ([ST09]), et même dans des cas plus généraux ([BGMS09, BC10]). Cependant, elle n'est pas vraie en toute généralité ([AP03]). Comme dans le cadre quadratique, un analogue du théorème de Rockafellar caractérise de tels ensembles. Γ est c -cycliquement monotone si et seulement s'il existe une fonction u c -convexe telle que $\Gamma \subset \text{graph } \partial_c u$. Une fonction u est dite c -convexe si elle s'écrit

$$u(x) = \sup_{y \in Y} \{-c(x, y) + \lambda_y\},$$

pour une famille de constantes λ_y . La c -sous-différentielle de u en x est alors

$$\partial_c u(x) = \{y \in Y \mid u(z) + c(z, y) \geq u(x) + c(x, y), \quad \forall z \in X\}.$$

Si $c(x, y) = -x \cdot y$, on retrouve bien la définition de la sous-différentielle d'une fonction convexe. Les ensembles X et Y étant bornés et si c est régulière, il est possible de montrer qu'une fonction c -convexe est presque partout deux fois différentiable. Soit $y \in \partial_c u(x)$ et supposons que u est différentiable en x alors la fonction

$$z \mapsto u(z) + c(z, y)$$

atteint un minimum en x , ce qui se traduit par

$$\nabla u(x) + \nabla_x c(x, y) = 0. \tag{I.10}$$

En résumé, si $(x, y) \in \text{Spt } \pi$, où π est un plan de transport c -optimal, et si u est différentiable en x , alors (I.10) est vérifié. Ainsi, si la fonction $y \mapsto -\nabla_x c(x, y)$ est injective, on déduit de cette égalité que π est déterministe. Ceci motive les hypothèses suivantes sur le coût :

(C1) $c \in \mathcal{C}^2(X \times Y)$.

(C2) Pour tout $x \in X$, la fonction $Y \ni y \mapsto -\nabla_x c(x, y) \in \mathbb{R}^d$ est injective.

On obtient alors le résultat suivant :

Théorème I.3. *Soit $c : X \times Y \rightarrow \mathbb{R}_+$ vérifiant (C1)–(C2), soient $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$ avec μ absolument continue par rapport à la mesure de Lebesgue. Alors il existe une unique fonction $T : X \rightarrow Y$ telle que $T_{\#}\mu = \nu$ et T est c -optimale.*

De plus, il existe une fonction c -convexe u telle que

$$T(x) = c\text{-exp}_x(\nabla u(x)),$$

où la c -exponentielle est définie par

$$y = c\text{-exp}_x(p) \iff p = -\nabla_x c(x, y).$$

Si on ajoute l'hypothèse $\nu \ll \mathcal{L}^d$ ainsi que

(C3) Pour tout $y \in Y$, la fonction $X \ni x \mapsto -\nabla_y c(x, y) \in \mathbb{R}^d$ est injective,

alors l'unique transport optimal de ν à μ s'écrit

$$T^*(y) := c^*\text{-exp}_y(\nabla u^c(y)),$$

avec $c^*(x, y) := c(y, x)$ et $u^c(y) := \sup_{x \in X} \{-c(x, y) - u(x)\}$.

Dans le cas particulier où $c(x, y) = h(x - y)$ avec h une fonction strictement convexe, ce résultat a été prouvé de manière indépendante dans [GM95] et [Caf96]. Il semble que le cas général ait été présenté pour la première fois, en utilisant la même stratégie, dans [MTW05].

Le problème original de Monge

Le Théorème I.3 ne s'applique pas lorsque $c(x, y) = |y - x|$, puisqu'en particulier l'hypothèse **(C2)** n'est pas vérifiée. Voyons tout d'abord que, à cause de la convexité non stricte du coût, il n'y a en général pas unicité du transport optimal.

Exemple I.7. Soient $\mu := \sum_{k=0}^{n-1} \delta_k$ et $\nu := \sum_{k=1}^n \delta_k$. Alors on peut montrer (et facilement se convaincre) que le transport $T(k) = k + 1$, qui correspond à un décalage de chacun des points d'un cran vers la droite, est optimal et correspond à un coût global de déplacement égal à n . Cependant, le transport

$$\tilde{T}(k) := \begin{cases} n & \text{si } k = 0, \\ k & \text{sinon,} \end{cases}$$

qui revient à laisser tous les points au même endroit, sauf le premier qui est envoyé en dernière position, engendre également un coût global égal à n et est donc lui aussi optimal. Remarquons que pour le coût quadratique, seul T est optimal.

L'existence d'une fonction de transport optimale est assurée dès que l'hypothèse d'absolue continuité de μ est satisfaite (voir [San15, Chapitre 2] pour plus de détails).

Théorème I.4 (Sudakov [Sud79], Trudinger - Wang [TW01], Caffarelli - Feldman - McCann [CFM02], Ambrosio [Amb03]). *Supposons que $\mu \ll \mathcal{L}^d$ alors il existe une fonction de transport T optimale pour le coût $|y - x|$.*

En dimension 1, si μ n'a pas d'atome, le transport monotone construit dans l'Exemple I.5 est optimal pour le coût $|y - x|$. Comme on l'a vu, ce n'est a priori pas le seul.

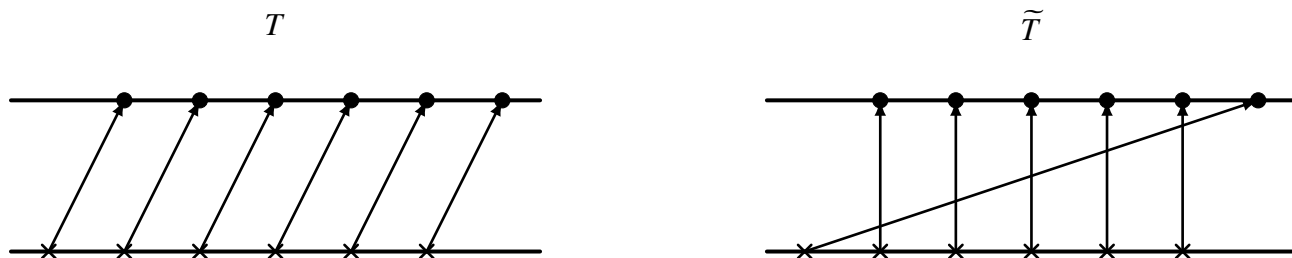


FIGURE 6: Les transports de l'Exemple I.7.

II Régularité du transport optimal

Après avoir résumé les principales notions liées au transport optimal, on s'intéresse maintenant à la régularité du minimiseur de (I.1). Pour cela, on suppose dans cette section que les mesures μ et ν sont absolument continues par rapport à la mesure de Lebesgue, de façon à ce qu'il existe un transport optimal déterministe dans les deux directions. Les deux premières sous-sections traiteront exclusivement du coût quadratique, d'abord en présentant les liens entre le transport optimal et une équation de Monge-Ampère, la majorité des résultats de régularité étant obtenus par cette équation. On verra que dans la plupart des cas, espérer une régularité globale, même aussi simple que la continuité, de la fonction de transport est irréaliste, ce qui nous amènera à des résultats de régularité partielle.

II.1 L'équation de Monge-Ampère

La situation est la suivante : $f, g : \mathbb{R}^d \rightarrow \mathbb{R}$ sont deux densités de probabilité à supports compacts $\text{Spt } f \subset X$ et $\text{Spt } g \subset Y$, X et Y étant deux ouverts bornés de \mathbb{R}^d , et on suppose qu'elles sont bornées et minorées par une constante strictement positive sur leurs supports. Soit maintenant $T : \mathbb{R}^d \rightarrow \mathbb{R}^d$ une fonction transportant f sur g , c'est à dire $T_{\#}f = g$.² Cette propriété se réécrit

$$\int \zeta(T(x))f(x) dx = \int \zeta(y)g(y) dy \quad \forall \zeta,$$

ce qui, après un changement de variables formel $y = T(x)$ sur l'intégrale de droite, donne

$$\int \zeta(T(x))f(x) dx = \int \zeta(T(x))g(T(x))|\det DT(x)| dx \quad \forall \zeta.$$

On obtient donc formellement, en utilisant la minoration strictement positive de g ,

$$|\det DT(x)| = \frac{f(x)}{g(T(x))} \quad \text{pour } x \in X. \quad (\text{II.1})$$

Cette EDP découle uniquement d'un changement de variables et de la condition $T_{\#}f = g$, en particulier, le caractère optimal ou non de T pour un certain coût n'est pas utilisé. Cependant, il est impossible d'obtenir une quelconque régularité de T , même en supposant f et g aussi régulières que possible, de l'équation (II.1). D'abord, le calcul que l'on a fait est formel

²C'est un abus de notation : $T_{\#}f = g$ signifie en fait $T_{\#}(f\mathcal{L}^d) = (g\mathcal{L}^d)$.

et le changement de variables doit être justifié, ce qui nécessite une régularité préalable de T . Ensuite, il existe en général une infinité de solutions à (II.1), la plupart n'ayant aucune régularité. Prenons un exemple extrêmement simple : $f = g = 1_D$ où D est le disque unité de \mathbb{R}^2 . Le transport optimal est bien sûr l'identité mais n'importe quelle fonction qui échange des parties du disque est un transport admissible. Il est donc clair que l'on ne peut s'attendre à aucune régularité pour un transport quelconque.

Si maintenant T est en plus optimal pour le coût quadratique, le Théorème I.2 donne une information supplémentaire sur la structure de T : c'est le gradient d'une fonction convexe u . Un théorème d'Alexandrov ([Vil09, Th. 14.25]) assure alors que u est presque partout deux fois différentiable, ce qui permet de définir une hessienne $\nabla^2 u$ positive, et de prouver que $T = \nabla u$ vérifie bien l'équation (II.1) presque partout, celle-ci prenant alors la forme d'une équation de type Monge-Ampère en u .

Théorème II.1 (McCann [McC97]). *Avec les hypothèses sur f et g , si u est une fonction convexe telle que $\nabla u_{\#} f = g$ alors pour presque tout $x \in \mathbb{R}^d$,*

$$\det \nabla^2 u(x) = \frac{f(x)}{g(\nabla u(x))}. \quad (\text{II.2})$$

Puisque u est bien une solution dans un sens classique de l'équation, grâce au Théorème II.1, on dira qu'une fonction convexe u telle que $\nabla u_{\#} f = g$ est une *solution au sens de Brenier* de l'équation de Monge-Ampère (II.2).

L'équation de Monge-Ampère

$$\det \nabla^2 u = F, \quad (\text{II.3})$$

où $F : \mathbb{R}^d \rightarrow \mathbb{R}_+$, bénéficie d'une théorie de régularité bien établie. En effet, Alexandrov ([Ale58]) a introduit la notion de solution généralisée suivante : si ν est une mesure sur un ouvert Ω , une fonction convexe $u : \Omega \rightarrow \mathbb{R}$ est *solution au sens d'Alexandrov* de $\det \nabla^2 u = \nu$, si la mesure d'Alexandrov μ_u coïncide avec ν , μ_u étant définie par

$$\mu_u(A) := |\cup_{x \in A} \partial u(x)|,$$

avec $\partial u(x)$ la sous-différentielle de u en x . Ceci est motivé par le fait que si u est convexe C^2 alors

$$\int_A \det \nabla^2 u(x) \, dx = |\nabla u(A)|.$$

Alexandrov a en particulier prouvé l'existence et l'unicité d'une solution généralisée au problème de Dirichlet associé. Il a également démontré un principe du maximum pour cette équation, résultat extrêmement utile pour l'étude d'une EDP. Suite aux contributions de nombreux auteurs, il apparaît que si le domaine Ω est uniformément convexe et à bord régulier, si F est régulière et minorée par une constante strictement positive sur $\overline{\Omega}$ alors la solution u au sens d'Alexandrov de (II.3) est régulière sur $\overline{\Omega}$. En particulier, comme on s'y attend, u gagne deux ordres de régularité par rapport à F : si $F \in C^{k,\alpha}(\overline{\Omega})$, avec $k \geq 2$, alors $u \in C^{k+2,\alpha}(\overline{\Omega})$ (voir [Fig17] pour une présentation bien plus détaillée de ces résultats).

Cependant, pour l'équation de Monge-Ampère (II.2) dérivant du problème de transport optimal, on a besoin d'une théorie de régularité pour un second membre beaucoup moins régulier, une hypothèse de régularité sur $\frac{f}{g \circ \nabla u}$ revenant à une hypothèse de régularité sur u .

Au début des années 90, Caffarelli a développé une théorie de régularité pour (II.3) avec un second membre seulement supposé majoré et minoré par des constantes strictement positives, ce qui correspond exactement à notre hypothèse sur les densités f et g .

Théorème II.2 (Caffarelli [Caf91]). *Soit $u : \Omega \rightarrow \mathbb{R}$ une fonction strictement convexe vérifiant, pour un certain $K > 0$,*

$$\frac{1}{K} \leq \det \nabla^2 u \leq K \quad \text{sur } \Omega, \quad (\text{II.4})$$

au sens d'Alexandrov alors $u \in C_{loc}^{1,\gamma}(\Omega)$ pour un certain γ universel.

Il reste alors à montrer qu'une solution au sens de Brenier de (II.2) est une solution strictement convexe au sens d'Alexandrov de (II.4). Pour cela, une hypothèse cruciale à ajouter est la convexité du support de g .

Exemple II.1. Le contre-exemple classique pour se convaincre que la convexité du domaine d'arrivée est nécessaire part de la situation suivante en dimension 2 : prenons $f = \mathbf{1}_B$ et $g = \mathbf{1}_{(B \cap \{x_1 < 0\}) - e_1} + \mathbf{1}_{(B \cap \{x_1 > 0\}) + e_1}$, où B est le disque unité de \mathbb{R}^2 . Il est clair que le transport optimal $T(x) = x + \text{sgn}(x) e_1$ est discontinu sur $\{x_1 = 0\}$. Il est alors possible de montrer que cette discontinuité est conservée dans la situation voisine où le support de g est modifié en une forme d'haltère avec un pont entre les deux demi-disques, en faisant un domaine connexe mais non convexe. Par stabilité du transport optimal, si le pont est suffisamment mince, le

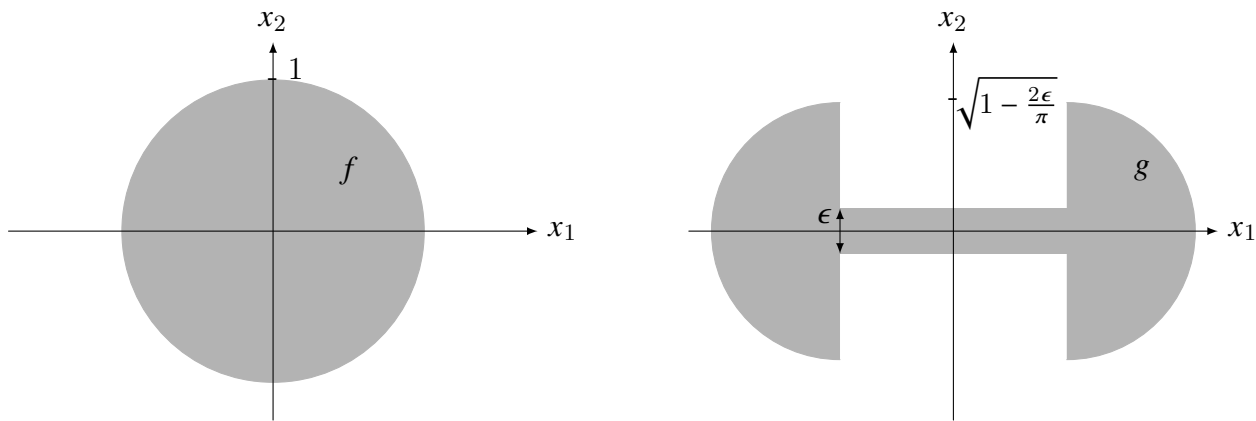


FIGURE 7: Les densités modifiées de l'Exemple II.1.

potentiel u associé est une solution non régulière de (II.2) au sens de Brenier et n'est donc pas une solution au sens d'Alexandrov de (II.4) puisque la mesure μ_u a une partie singulière. Pourtant, les densités sont aussi régulières que possibles.

Exemple II.2. Dans [Wol97] et [Urb07], les auteurs donnent un critère sur les courbures des bords de deux domaines qui assure que le transport optimal entre ces deux domaines n'est pas un C^1 -difféomorphisme. Plus récemment, l'article [CJLPR15] fournit un critère du même type sur la courbure du bord du domaine d'arrivée suffisant pour que le transport optimal soit discontinu. Enfin, on peut se demander si un "petit" défaut de convexité peut induire une

discontinuité du transport, le support de la mesure-cible dans le contre-exemple II.1 étant très fortement non-convexe. C'est effectivement le cas : dans [Jha19], il est démontré que pour tout $\varepsilon > 0$, l' ε -perturbation Lipschitz \tilde{Q} d'un carré Q présentée dans la Figure 8, induit un transport optimal discontinu de $\mathbf{1}_Q$ sur $\mathbf{1}_{\tilde{Q}}$, ce qui illustre bien l'importance de la convexité du domaine d'arrivée.

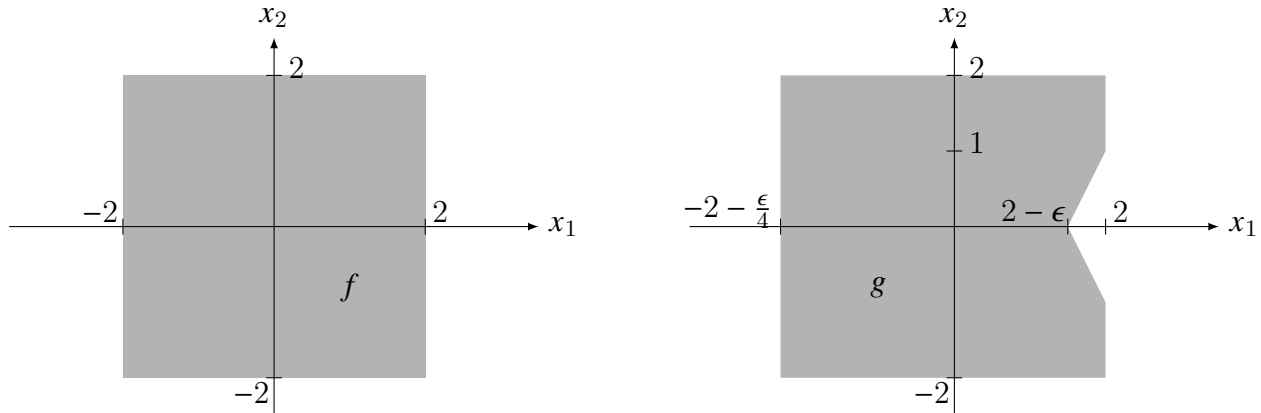


FIGURE 8: Le contre-exemple de [Jha19].

En ajoutant cette hypothèse de convexité, Caffarelli a prouvé qu'une solution u convexe au sens de Brenier de (II.2) est bien une solution au sens d'Alexandrov de cette même équation. Partant de là, il a obtenu le résultat suivant.

Théorème II.3 (Caffarelli [Caf92a], [Caf92b]). *En plus des hypothèses sur f et g au début de cette sous-section, supposons que $\text{Spt } g$ est convexe. Alors le transport optimal T de f à g est $C^{0,\beta}$ à l'intérieur de $\text{Spt } f$, pour un certain $\beta \in (0, 1)$.*

Si de plus, $\text{Spt } f$ est convexe alors $T \in C^{0,\beta}(\text{Spt } f)$, pour un certain $\beta \in (0, 1)$.

Dans [Caf90], Caffarelli montre également que si u est une solution convexe au sens d'Alexandrov de $\det \nabla^2 u = F$ sur B_1 et si F est suffisamment proche de 1 alors u est $W^{2,p}$ sur $B_{\frac{1}{2}}$. On en déduit (toujours dans le cas où $\text{Spt } g$ est convexe) que pour tout $p > 1$, il existe $\varepsilon(p) > 0$ tel que si $|f - 1| < \varepsilon$ et $|g - 1| < \varepsilon$, alors le potentiel convexe u tel que $\nabla u \# f = g$ est $W_{\text{loc}}^{2,p}$ dans l'intérieur de $\text{Spt } f$.

Linéarisation de l'équation de Monge-Ampère

Pour terminer cette sous-section sur l'équation de Monge-Ampère associée au transport optimal, on présente ici une linéarisation de cette équation, dont une version quantitative nous sera utile dans le chapitre 2. La fonction $u(x) := \frac{|x|^2}{2}$ est solution triviale de l'équation de Monge-Ampère $\det \nabla^2 u = 1$, ce qui correspondrait à $f(x) = g(\nabla u(x)) = g(x)$. Perturbons cette situation en écrivant $u(x) = \frac{|x|^2}{2} + \varepsilon v(x) + O(\varepsilon^2)$ et $g(x) = (1 + \varepsilon h(x) + O(\varepsilon^2))f(x)$, on obtient au premier ordre en ε l'équation linéarisée

$$\left(-\Delta - \frac{\nabla f}{f} \cdot \nabla\right)v = h. \quad (\text{II.5})$$

Supposons maintenant que f et g sont tous les deux proches de 1 : $f(x) = 1 + \varepsilon\gamma(x) + O(\varepsilon^2)$ et $g(x) = 1 + \varepsilon\delta(x) + O(\varepsilon^2)$. Le même calcul que précédemment donne l'équation linéarisée

$$-\Delta v = \delta - \gamma.$$

Ainsi, si f et g sont proches de 1, la solution u de l'équation de Monge-Ampère (II.2) associée est proche de $\frac{|x|^2}{2} + \varphi(x)$ avec φ solution de l'équation de Poisson

$$-\Delta \varphi = g - f. \tag{II.6}$$

En d'autres termes, le transport optimal de f à g est proche de l'identité perturbée par le gradient d'une solution de (II.6). Dans [GHO21], une version quantitative de cette linéarisation est obtenue et est un des ingrédients de base de la stratégie développée dans le Chapitre 2 de cette thèse.

Lien entre W_2 et une norme de Sobolev négative

La linéarisation précédente permet également de linéariser formellement la distance W_2 entre deux mesures μ et ν proches l'une de l'autre. A nouveau, le potentiel convexe est $u(x) = \frac{|x|^2}{2} + \varepsilon v(x) + O(\varepsilon^2)$ et les densités f de μ et g de ν sont telles que $g(x) = (1 + \varepsilon h(x) + O(\varepsilon^2))f(x)$. On a alors

$$W_2^2(\mu, \nu) = \int |\nabla u(x) - x|^2 \mu(dx) = \varepsilon^2 \int |\nabla v|^2 d\mu + O(\varepsilon^4).$$

On a vu que v est solution de l'équation linéarisée (II.5), c'est-à-dire en notant $L := -\Delta - \frac{\nabla f}{f} \cdot \nabla$,

$$Lv = h.$$

L'opérateur L vérifiant une formule d'intégration par parties contre μ , on obtient

$$\int \nabla v \cdot \nabla v d\mu = \int (Lv)v d\mu = \int hv d\mu.$$

En définissant la norme à poids $\dot{H}^{-1}(\mu)$ par

$$\|h\|_{\dot{H}^{-1}(\mu)} := \sup \left\{ \int hk d\mu \mid \|\nabla k\|_{L^2(\mu)} \leq 1 \right\},$$

et en remarquant que pour tout k ,

$$\int hk d\mu = \int (Lv)k d\mu = \int \nabla v \cdot \nabla k d\mu \leq \|\nabla v\|_{L^2(\mu)} \|\nabla k\|_{L^2(\mu)},$$

avec égalité si $k = \frac{v}{\|\nabla v\|_{L^2(\mu)}}$, on en déduit que

$$\|\nabla v\|_{L^2(\mu)} = \|h\|_{\dot{H}^{-1}(\mu)}.$$

On voit formellement qu'au premier ordre (voir [OV00, § 7]),

$$W_2(\mu, (1 + h)\mu) = \|h\|_{\dot{H}^{-1}(\mu)} + o(h).$$

Il est possible d'intégrer cette estimation infinitésimale et de passer d'une norme à poids à la norme \dot{H}^{-1} associée à la mesure de Lebesgue. Ainsi, si $0 < a \leq f, g \leq b < \infty$, on a (voir [Pey16], [San15, Th. 5.34])

$$\frac{1}{\sqrt{b}} \|f - g\|_{\dot{H}^{-1}} \leq W_2(f, g) \leq \frac{1}{\sqrt{a}} \|f - g\|_{\dot{H}^{-1}}. \tag{II.7}$$

II.2 Régularité partielle

On a vu dans l'Exemple II.1 que si le domaine d'arrivée n'est pas convexe, on ne peut pas espérer une régularité globale pour la fonction de transport optimale. Voyons dans cette sous-section ce que l'on peut tout de même dire dans ce cas. Les hypothèses sur les densités f et g sont les mêmes que dans la sous-section précédente, en particulier f , g , $\frac{1}{f}$ et $\frac{1}{g}$ sont bornées par K sur leurs supports. Rappelons également que le coût est quadratique.

Le premier résultat dans ce sens est dû à Figalli ([Fig10]) en dimension 2, et est rapidement étendu aux dimensions supérieures par Figalli et Kim ([FK10]).

Théorème II.4 (Figalli - Kim [FK10]). *Soit T le transport optimal de f à g . Il existe deux ouverts $X' \subset X$ et $Y' \subset Y$ avec $|X \setminus X'| = |Y \setminus Y'| = 0$, tels que $T : X' \rightarrow Y'$ soit un $C^{0,\alpha}$ -difféomorphisme.*

Dans [Fig10], une description précise de l'ensemble des points singuliers du potentiel convexe u est donnée, ce qui n'a pas pu être étendu en dimension supérieure. La preuve de [FK10] suit la stratégie de [Caf92a], en montrant en particulier qu'en un point x pour lequel $\partial u(x) \cap Y \neq \emptyset$, le potentiel convexe u est strictement convexe et différentiable. Il reste alors à montrer que cet ensemble est un ouvert de mesure pleine.

Preuve variationnelle de régularité partielle

Le chapitre 2 de cette thèse s'appuie sur une preuve variationnelle du Théorème II.4 proposée par Goldman et Otto dans [GO20], dont on va maintenant donner un aperçu.

Le résultat principal est un résultat d' ε -régularité qui, couplé au théorème d'Alexandrov, permet rapidement de montrer la régularité partielle de T mentionnée dans le Théorème II.4. Pour énoncer ce résultat d' ε -régularité, on introduit l'énergie quadratique de T à l'échelle R

$$\mathcal{E}_R(T) := \frac{1}{R^{d+2}} \left(\int_{B_R} |T(x) - x|^2 dx + \int_{B_R} |T^{-1}(y) - y|^2 dy \right),$$

qui mesure l'erreur en norme L^2 entre T et l'identité dans B_R et entre T^{-1} et l'identité dans B_R . L'idée de l' ε -régularité est de montrer que si T est proche de l'identité à une certaine échelle, c'est-à-dire ici si l'énergie $\mathcal{E}_R(T)$ est petite, et donc si f et g sont proches de 1 à cette échelle, alors T est hölderienne à une plus petite échelle.

Théorème II.5. *Supposons que $f(0) = g(0) = 1$ et que pour un certain $R > 0$,*

$$\mathcal{E}_R(T) + R^{2\alpha} \left([f]_{\alpha,R}^2 + [g]_{\alpha,R}^2 \right) \ll 1, \quad (\text{II.8})$$

alors $T \in C^{1,\alpha}(B_{\frac{R}{2}})$ et

$$R^{2\alpha} [\nabla T]_{\alpha, \frac{R}{2}}^2 \lesssim \mathcal{E}_R(T) + R^{2\alpha} \left([f]_{\alpha,R}^2 + [g]_{\alpha,R}^2 \right).$$

La stratégie de preuve du Théorème II.5 est proche de celle utilisée par De Giorgi pour obtenir un résultat d' ε -régularité pour des surfaces minimales, voir en particulier [SS82] pour une variante très proche de cette stratégie.

- L'ingrédient principal est une version quantitative de la linéarisation de l'équation de Monge-Ampère en équation de Poisson, présentée dans la sous-section II.1. En effet, la situation est proche de celle décrite dans cette sous-section : le transport $T := \nabla u$ est proche de l'identité et les densités f et g sont proches de 1. Ainsi, $T - \text{Id}$ devrait être proche du gradient d'une fonction ϕ solution du problème de Poisson associé dans B_R avec condition au bord liée au flux j . La version quantitative [GO20, Prop. 4.6] est la suivante : pour tout $\tau > 0$, sous les mêmes hypothèses que le Théorème II.5, en particulier (II.8) qui dépend maintenant de τ , il existe un gradient harmonique $\nabla\phi$ dans $B_{\frac{R}{2}}$ tel que

$$\frac{1}{R^{d+2}} \int_{B_{\frac{R}{2}}} |T(x) - x - \nabla\phi(x)|^2 dx \lesssim \tau \mathcal{E}_R(T) + R^{2\alpha} \left([f]_{\alpha,R}^2 + [g]_{\alpha,R}^2 \right), \quad (\text{II.9})$$

et $\frac{1}{R^{d+2}} \int_{B_{\frac{R}{2}}} |\nabla\phi|^2 \lesssim \mathcal{E}_R(T) + R^{2\alpha} \left([f]_{\alpha,R}^2 + [g]_{\alpha,R}^2 \right).$

Cette approximation harmonique est d'abord prouvée dans le cadre eulérien, en construisant un compétiteur à la paire densité-flux associée à T , puis transférée dans le cadre lagrangien grâce à l'équivalence (I.8).

- Après développement de Taylor de ϕ , il est clair que T doit être très proche d'une fonction affine, celle-ci étant très proche de l'identité par estimations elliptiques. En particulier, le fait que l'on puisse choisir τ aussi petit que l'on veut dans (II.9) fait que T est plus proche d'une fonction affine à l'échelle $\frac{R}{2}$ qu'elle ne l'était grâce à (II.8) à l'échelle R . Ceci suggère d'effectuer un changement de variables affine pour obtenir un transport \widehat{T} , toujours optimal entre deux densités \widehat{f} et \widehat{g} , et telle qu'à une plus petite échelle θR et pour un certain $\beta > 0$,

$$\mathcal{E}_{\theta R}(\widehat{T}) \leq \theta^{2\beta} \mathcal{E}_R(T) + CR^{2\alpha} \left([f]_{\alpha,R}^2 + [g]_{\alpha,R}^2 \right).$$

- La dernière étape est une itération du résultat précédent pour obtenir la formulation de Campanato de la régularité $C^{1,\alpha}$ de T ([Cam64]).

Ce schéma de preuve a été généralisé à des mesures μ et ν quelconques dans [GHO21] et adapté dans [MO21] pour obtenir un résultat de régularité au bord, améliorant ainsi un résultat de [CF15]. Le chapitre 2 de cette thèse est une généralisation de cette stratégie au cas de coûts généraux et utilise le résultat d'approximation harmonique obtenu dans [GHO21]. Le cas où les densités sont seulement continues est également traité dans [Gol20].

II.3 Coût général

La régularité du transport optimal pour un coût général $c(x, y)$ est beaucoup plus complexe. Voyons d'abord ce que devient l'équation de transport dans ce cas. Supposons que le coût satisfait aux hypothèses (C1)–(C2). On a toujours

$$|\det DT(x)| = \frac{f(x)}{g(T(x))},$$

avec $T(x) = c\text{-exp}_x(\nabla u(x))$ d'après le Théorème I.3, où u est une fonction c -convexe. Par définition de la c -exponentielle,

$$\nabla u(x) + \nabla_x c(x, T(x)) = 0,$$

ce qui peut être différencié (le calcul est à nouveau formel) en

$$\nabla^2 u(x) + \nabla_{xx} c(x, T(x)) + DT(x) \nabla_{xy} c(x, T(x)) = 0,$$

où $\nabla_{xx} c = \left(\frac{\partial^2 u}{\partial x_i \partial x_j} \right)_{i,j}$ et $\nabla_{xy} c = \left(\frac{\partial^2 u}{\partial x_i \partial y_j} \right)_{i,j}$. Ce calcul suggère l'introduction de l'hypothèse suivante sur le coût :

(C4) $\det \nabla_{xy} c(x, y) \neq 0$ pour tout $(x, y) \in X \times Y$.

En supposant **(C4)** et en remarquant que la c -convexité de u implique que si u est 2 fois différentiable en x alors

$$\nabla^2 u(x) + \nabla_{xx} c(x, T(x)) \geq 0,$$

on obtient une équation du type Monge-Ampère vérifiée par u :

$$\det (\nabla^2 u(x) + \nabla_{xx} c(x, T(x))) = |\det \nabla_{xy} c(x, T(x))| \frac{f(x)}{g(T(x))}, \quad (\text{II.10})$$

en rappelant que $T(x) = c\text{-exp}_x(\nabla u(x))$.

Dans [MTW05], Ma, Trudinger et Wang ont montré qu'en plus des hypothèses **(C1)**–**(C4)** sur le coût et une hypothèse de c -convexité sur les supports de f et g , une certaine hypothèse, appelée depuis la condition MTW, sur les dérivées de c assure la régularité globale du transport optimal sur le support de f : si f et g sont \mathcal{C}^2 alors T est \mathcal{C}^3 . La condition MTW est une identité compliquée faisant intervenir les dérivées de c jusqu'à l'ordre 4. Dans [Loe09], Loeper a ensuite montré que la condition MTW est nécessaire : s'il existe un point où elle n'est pas vérifiée, il est possible de construire des densités f et g régulières, dont les supports vérifient l'hypothèse de c -convexité, telles que le transport c -optimal ne soit pas continu.

La condition MTW est très restrictive, beaucoup de coûts intéressants ne la vérifient pas. Il est donc d'un intérêt certain d'obtenir un résultat de régularité partielle dans l'esprit du Théorème II.4, cette fois pour un coût général. Un tel résultat a été prouvé par De Philippis et Figalli :

Théorème II.6 (De Philippis - Figalli [DF15]). *Supposons que $c : X \times Y \rightarrow \mathbb{R}$ est $\mathcal{C}^{2,\alpha}$ et vérifie **(C2)**–**(C4)** et que les densités f et g vérifient les mêmes hypothèses que précédemment. Si f et g sont continues alors il existe deux ouverts $X' \subset X$ et $Y' \subset Y$ avec $|X \setminus X'| = |Y \setminus Y'| = 0$, tels que $T : X' \rightarrow Y'$ soit un $\mathcal{C}^{0,\beta}$ -difféomorphisme, pour tout $\beta < 1$. Si de plus $c \in \mathcal{C}^{k+2,\alpha}$ et $f, g \in \mathcal{C}^{k,\alpha}$ alors T est un $\mathcal{C}^{k+1,\alpha}$ -difféomorphisme.*

Le chapitre 2 s'attachera à donner une preuve variationnelle de ce résultat, généralisant ainsi la stratégie de [GO20] (cf. sous-section IV.2 pour une présentation générale de cette preuve). Le résultat crucial de [DF15] est un théorème d' ε -régularité, voir [DF15, Th. 4.3 & 5.3], qui prouve la régularité dans un contexte local et normalisé. L'argument utilisé dans [GO20], et donc dans le chapitre 2, est assez similaire à celui de [DF15]. En effet, les deux sont

basés sur une itération de Campanato, sur des boules dont le rayon décroît géométriquement, qui reproduit à chaque étape une amélioration de l'erreur après changement de variables affine. Ce changement de variables provient dans les deux cas de la proximité entre la solution du problème de transport optimal (pour le coût quadratique dans [GO20], pour un coût général dans [DF15]) et la solution d'un problème plus simple admettant une théorie de régularité bien établie. La principale différence réside dans ce problème simplifié : si dans [DF15], la comparaison se fait avec une équation de Monge-Ampère avec source constante, dans [GO20], c'est un simple problème de Poisson qui est utilisé. De plus, cette comparaison se base sur le principe du maximum dans [DF15] alors qu'elle s'appuie dans [GO20] sur le fait que la paire densité-flux optimale est un minimiseur local (ou un quasi-minimiseur dans le chapitre 2) avec ses propres conditions au bord.

La stratégie mise en place dans [DF15, Th. 4.4 & 5.3] a été développée ensuite dans [CF15] pour prouver un résultat de régularité au bord, dans [CF16] pour obtenir la stabilité de la régularité hölderienne du transport et dans [CF17] où est établie une régularité $W^{2,p}$ partielle du transport optimal.

II.4 Régularité quantitative

On présente ici un résultat de régularité plus quantitatif dû à nouveau à Caffarelli [Caf00, Caf02]. Celui-ci énonce que la fonction de transport optimale, en coût quadratique, de la gaussienne standard $\gamma_d(dx) = (2\pi)^{-\frac{d}{2}} e^{-\frac{|x|^2}{2}} dx$ sur \mathbb{R}^d vers une mesure ayant une densité uniformément log-concave par rapport à γ_d est 1-Lipschitz.

Théorème II.7 (Caffarelli [Caf00, Caf02]). *Soit ν une mesure de probabilité de la forme $\nu(dx) = e^{-V} \gamma_d(dx)$ avec V une fonction convexe. Alors il existe une fonction $u : \mathbb{R}^d \rightarrow \mathbb{R}$ convexe de classe C^1 avec $\nabla u \# \gamma_d = \nu$ et $\nabla^2 u \leq I_d$, où $\nabla^2 u$ est la Hessienne de u au sens d'Alexandrov. Ainsi, la fonction de transport $T = \nabla u$ de γ_d à ν est 1-Lipschitz.*

Notons qu'on peut facilement en déduire une version un peu plus quantitative. En effet, on peut reformuler le Théorème II.7 de la façon suivante : si u est convexe et $\nabla u \# \gamma_d = \nu$ avec $\nu(dx) = e^{-W(x)} dx$ et $\nabla^2 W \geq I_d$, alors $\nabla^2 u \leq I_d$, où I_d est la matrice identité de taille d . Prenons maintenant une mesure de probabilité de la forme $\tilde{\nu}(dx) := e^{-\tilde{W}(x)} dx$ avec $\nabla^2 \tilde{W} \geq c I_d$ avec $c > 0$ et définissons $W(x) = \tilde{W}(\frac{x}{\sqrt{c}})$ et \tilde{u} le transport optimal de γ_d à $\tilde{\nu}$. On a donc $\nabla^2 \tilde{W} \geq I_d$, ce qui implique que le transport optimal ∇u de γ_d à $\nu(dx) := c^{-\frac{d}{2}} e^{-W(x)} dx$ vérifie $\nabla^2 u \leq I_d$. Or on vérifie facilement que

$$u(x) = \sqrt{c} \tilde{u}(x),$$

de sorte qu'on obtient

$$\nabla^2 \tilde{u} \leq \frac{1}{\sqrt{c}} I_d.$$

En particulier, si ν est strictement plus log-concave que la gaussienne, c'est-à-dire si $c > 1$, alors le transport optimal est contractant.

Le théorème de contraction de Caffarelli joue un rôle important dans le domaine des inégalités fonctionnelles, où il permet de transférer une inégalité vraie pour la gaussienne γ_d à des mesures uniformément log-concaves. Voyons un exemple très simple.

Exemple II.3. Il est bien connu ([Gro75]) que la gaussienne γ_d vérifie une inégalité log-Sobolev avec constante optimale 2 :

$$\text{Ent}_{\gamma_d}(f) \leq 2 \int_{\mathbb{R}^d} |\nabla f|^2 d\gamma_d, \quad \text{pour tout } f \in H^1(\gamma_d),$$

où l'entropie de f relative à une mesure μ est définie par

$$\text{Ent}_{\mu}(f) := \int_{\mathbb{R}^d} f^2 \log f^2 d\mu - \int_{\mathbb{R}^d} f^2 d\mu \log \left(\int_{\mathbb{R}^d} f^2 d\mu \right).$$

Prenons maintenant μ une mesure à densité uniformément log-concave par rapport à la mesure de Lebesgue, c'est-à-dire $\mu(dx) = e^{-V(x)} dx$ avec $\nabla^2 V \geq cI_d$ et $c > 0$. Soit ∇u le transport optimal de γ_d à μ , on sait d'après le théorème de contraction que $\nabla^2 u \leq \frac{1}{\sqrt{c}}I_d$. On a alors

$$\begin{aligned} \text{Ent}_{\mu}(f) &= \text{Ent}_{\gamma_d}(f \circ \nabla u) \leq 2 \int_{\mathbb{R}^d} |\nabla(f \circ \nabla u)|^2 d\gamma_d \\ &\leq 2 \int_{\mathbb{R}^d} |\nabla^2 u|^2 |\nabla f \circ \nabla u|^2 d\gamma_d \\ &\leq \frac{2}{c} \int_{\mathbb{R}^d} |\nabla f \circ \nabla u|^2 d\gamma_d \\ &= \frac{2}{c} \int_{\mathbb{R}^d} |\nabla f|^2 d\mu. \end{aligned}$$

On a ainsi montré que μ vérifie une inégalité log-Sobolev avec constante $\frac{2}{c}$. Cette constante est optimale, puisque la gaussienne $\left(\frac{c}{2\pi}\right)^{\frac{d}{2}} e^{-\frac{c|x|^2}{2}}$ admet $\frac{2}{c}$ pour constante optimale. On retrouve alors une conséquence du critère de courbure-dimension de Bakry-Émery [BÉ85]. On renvoie à [Led99] pour une présentation des inégalités de Sobolev logarithmiques. Voir également [OT20] pour un résultat de rigidité pour l'inégalité log-Sobolev.

Des applications à d'autres types d'inégalités géométriques, comme l'inégalité isopérimétrique gaussienne, peuvent être trouvées dans [Cor02, Har99, Har01, Mil18]. D'autre part, le Théorème II.7 a été utilisé pour des résultats quantitatifs de rigidité et de stabilité d'inégalités fonctionnelles [DF17, CFP18, CF20]. Pour ce genre d'applications, le fait que la borne sur la constante de Lipschitz du transport optimal ne dépend pas de la dimension est crucial. Mentionnons quelques extensions récentes, telles que des estimations de type Sobolev ne dépendant pas de la dimension [Kol10, Kol13] et des variantes pour des perturbations à support compact de la gaussienne standard [CFJ17].

III Le problème d'appariement optimal

On présente dans cette section un cas particulier de transport optimal déjà esquissé dans l'Exemple I.3, à savoir le transport entre deux mesures discrètes ayant le même nombre d'atomes, ceux-ci ayant la même masse. Ici, ces mesures discrètes seront empiriques dans le sens où elles proviennent de processus de points et l'objet de l'étude est le comportement

moyen du coût de transport optimal ou du transport optimal lui-même. Après avoir rappelé les principaux résultats pour l'appariement optimal classique entre une mesure empirique de points indépendants et identiquement distribués et la mesure de référence, on s'intéressera à une famille de processus de points ni indépendants, ni identiquement distribués, prenant la forme de mesures spectrales empiriques de matrices aléatoires.

III.1 Mesures empiriques de points i.i.d.

Le problème d'*appariement optimal* (*optimal matching* en anglais) consiste principalement en l'étude du problème d'optimisation suivant

$$\min_{\sigma \in \mathfrak{S}_N} \frac{1}{N} \sum_{k=1}^N |X_k - Y_{\sigma(k)}|^p, \quad (\text{III.1})$$

où $p \geq 1$ et où les variables aléatoires X_k et Y_k sont indépendantes et identiquement distribuées (i.i.d.) selon une mesure de référence μ . Remarquons que ce problème est exactement un problème de transport optimal entre les mesures empiriques discrètes $\frac{1}{N} \sum_{k=1}^N \delta_{X_k}$ et $\frac{1}{N} \sum_{k=1}^N \delta_{Y_k}$. En effet,

$$W_p^p \left(\frac{1}{N} \sum_{k=1}^N \delta_{X_k}, \frac{1}{N} \sum_{k=1}^N \delta_{Y_k} \right) = \min_{\sigma \in \mathfrak{S}_N} \frac{1}{N} \sum_{k=1}^N |X_k - Y_{\sigma(k)}|^p.$$

Par inégalité triangulaire, il est clair que la valeur du minimum dans (III.1) est estimée par la valeur de

$$W_p^p \left(\frac{1}{N} \sum_{k=1}^N \delta_{X_k}, \mu \right), \quad (\text{III.2})$$

ce que l'on va donc étudier dans la suite de cette sous-section. Le but est maintenant d'estimer la valeur de (III.2), d'établir des estimations de concentration de la mesure empirique $\mu_N := \frac{1}{N} \sum_{k=1}^N \delta_{X_k}$ autour de sa moyenne μ et si possible d'étudier le comportement de la fonction réalisant le transport optimal de μ_N à μ en coût $|x - y|^p$.

Cette thèse ne s'intéressera pas à des mesures empiriques où les atomes sont indépendants et identiquement distribués, mais il nous semble tout de même intéressant de rappeler quelques résultats principaux dans ce domaine.

- **Dimension 1** ([BL16]). Pour une large classe de mesures $\mu \in \mathcal{P}(\mathbb{R})$, par exemple dès que μ possède un moment d'ordre $q > 2$ fini, on a³

$$\mathbb{E}W_1(\mu_N, \mu) \lesssim \frac{1}{\sqrt{N}}.$$

Si $p > 1$, la situation est plus complexe. Par exemple, si $\mu = \mathcal{L}|_{[0,1]}$, alors

$$\mathbb{E}W_p^p(\mu_N, \mu) \approx N^{-\frac{p}{2}},$$

³Dans cette thèse, on utilisera la notation \lesssim pour signifier qu'une quantité est plus petite qu'une autre quantité multipliée par une constante. Si cette constante dépend d'un paramètre α , on notera \lesssim_α . La notation $A \approx B$ signifie alors : $A \lesssim B$ et $B \lesssim A$.

mais si $\mu = \mathcal{N}(0, 1)$,

$$\mathbb{E}W_p^p(\mu_N, \mu) \approx \begin{cases} N^{-\frac{p}{2}} & \text{si } 1 \leq p < 2, \\ \frac{\log \log N}{N_1} & \text{si } p = 2, \\ \frac{1}{N(\log N)^{\frac{p}{2}}} & \text{si } p > 2. \end{cases}$$

- **Dimension $d \geq 2$ avec μ uniforme.** Soit $Q^d := (0, 1)^d$ et $\mu := \mathcal{L}^d \llcorner_{Q^d}$. Les N points X_1, \dots, X_N sont dans le cube Q^d donc la distance moyenne entre les points doit être d'ordre $N^{-\frac{1}{d}}$, ce qui suggère que $W_p^p(\mu_N, \mu)$ doit être d'ordre $N^{-\frac{p}{d}}$. Si cette intuition est vraie dès que $d \geq 3$, Ajtai, Komlós et Tusnády ont montré qu'en dimension 2, un facteur logarithmique apparaît :

$$\mathbb{E}W_p^p(\mu_N, \mu) \approx \begin{cases} \left(\frac{\log N}{N}\right)^{\frac{p}{2}} & \text{si } d = 2 \quad ([\text{AKT84}]), \\ N^{-\frac{p}{d}} & \text{si } d \geq 3 \quad ([\text{Tal92, DY95, Led19}]). \end{cases}$$

Cette particularité de la dimension 2 a fait l'objet de beaucoup d'articles depuis, citons en particulier [AST19] où la limite

$$\lim_{N \rightarrow \infty} \frac{\log N}{N} \mathbb{E}W_2^2(\mu_N, \mu) = \frac{1}{4\pi}, \quad \text{si } d = 2,$$

est établie, prouvant ainsi un ansatz provenant de physiciens [CLPS14] (cet article contient également d'autres conjectures sur le comportement de l'appariement optimal). Avec la même stratégie, il est montré dans [AGT19] que la fonction de transport optimale peut être approchée en norme L^2 par la somme de l'identité et du gradient de la solution à un problème de Poisson. Mentionnons que dans [BB13], les auteurs prouvent qu'en dimension $d \geq 3$, la limite de $N^{-\frac{p}{d}} \mathbb{E}W_p^p(\mu_N, \tilde{\mu}_N)$, avec $\tilde{\mu}_N$ la mesure empirique de N points i.i.d. suivant μ , existe pour $1 \leq p < \frac{d}{2}$. Ce résultat a été récemment étendu pour tout $p \geq 1$ dans [GT21].

- **Dimension $d \geq 2$ pour d'autres mesures.** Si μ est la gaussienne standard dans \mathbb{R}^2 alors

$$\mathbb{E}W_p^p(\mu_N, \mu) \approx \begin{cases} \left(\frac{\log N}{N}\right)^{\frac{p}{2}} & \text{si } 1 \leq p < 2 \quad ([\text{Led19}]), \\ \frac{(\log N)^2}{N} & \text{si } p = 2 \quad ([\text{Led19, Tal18}]). \end{cases}$$

Le cas $p > 2$ est toujours ouvert. Plus généralement, si μ est une mesure ayant un moment fini d'ordre suffisamment grand, on a

$$\mathbb{E}W_1(\mu_N, \mu) \lesssim \sqrt{\frac{\log N}{N}} \quad ([\text{Yuk92}]).$$

Si maintenant μ est la gaussienne standard sur \mathbb{R}^d avec $d \geq 3$, alors

$$\mathbb{E}W_p^p(\mu_N, \mu) \approx N^{-\frac{p}{d}} \quad \text{si } 1 \leq p < d \quad ([\text{LZ19}]).$$

Enfin, de façon très générale, il est montré dans [FG15] que si $p > 0$ et si μ a un moment fini d'ordre suffisamment grand par rapport à p , alors

$$\mathbb{E}W_p^p(\mu_N, \mu) \lesssim \begin{cases} N^{-\frac{1}{2}} & \text{si } p > \frac{d}{2}, \\ N^{-\frac{1}{2}} \log(1+N) & \text{si } p = \frac{d}{2}, \\ N^{-\frac{p}{d}} & \text{si } p < \frac{d}{2}. \end{cases}$$

III.2 Mesures spectrales empiriques

Le chapitre 3 s'intéressera au comportement de la mesure spectrale empirique de matrices aléatoires à entrées gaussiennes indépendantes, en particulier la vitesse de convergence vers une mesure limite en distance W_2 . La mesure spectrale d'une matrice $M \in \mathbb{C}^{N \times N}$ est

$$\mu(M) := \frac{1}{N} \sum_{k=1}^N \delta_{\lambda_k},$$

où $\lambda_1, \dots, \lambda_N$ sont les valeurs propres de M . On donne ici les résultats principaux de convergence quantitative, en commençant par les matrices hermitiennes.

Matrices hermitiennes

On sait depuis [Wig55] et [Wig58] que la mesure spectrale $\mu_N := \mu\left(\frac{M_N}{\sqrt{N}}\right)$ d'une matrice de Wigner $M_N \in \mathbb{C}^{N \times N}$ (matrice hermitienne avec entrées indépendantes sur et au-dessus de la diagonale et identiquement distribuées avec moyenne et variance égales à 1 et moments de tous ordres) renormalisée par \sqrt{N} converge faiblement en probabilité vers la mesure du demi-cercle

$$\nu_{\text{sc}}(dx) := \frac{1}{2\pi} \sqrt{4-x^2} \mathbf{1}_{|x| \leq 2} dx.$$

Cette *loi du demi-cercle* constitue un premier exemple d'*universalité* du comportement des matrices aléatoires, dans le sens où la mesure déterministe limite ne dépend pas de la distribution des entrées de la matrice. Des versions quantitatives de ce résultat ont d'abord été démontrées pour les ensembles GOE et GUE, cas particuliers où les entrées de la matrice sont gaussiennes, ce qui permet d'obtenir une formule pour la distribution des valeurs propres, voir [Meh04, Chap. 3] ou [AGZ10, Chap. 2]. La distance de Kolmogorov entre $\mathbb{E}\mu_N$ et ν_{sc} est de l'ordre de $\frac{1}{N}$ pour le GUE ([GT05]) et pour le GOE ([KTT07]). Il a été montré plus récemment ([GT16]) qu'elle est également de cet ordre pour des matrices de Wigner générales. Enfin, la distance de Kolmogorov entre μ_N et ν_{sc} est bornée par $\frac{\log^5 N}{N}$ avec haute probabilité, voir [GT14].

La distance de Wasserstein entre μ_N et ν_{sc} a été étudiée dans [MM13], où la borne

$$\mathbb{E}W_1(\mu_N, \mathbb{E}\mu_N) \lesssim N^{-\frac{2}{3}}$$

est obtenue, en supposant que la distribution des entrées satisfait une inégalité log-Sobolev. Elle est ensuite améliorée en norme W_2 dans [Dal12] en

$$\mathbb{E}W_2^2(\mu_N, \nu_{\text{sc}}) \lesssim \frac{\log N}{N^2},$$

estimation vraie pour une large classe de matrices de Wigner.

Matrices non nécessairement hermitiennes

Les matrices aléatoires dont on va étudier le spectre dans le chapitre 3 constituent l'ensemble de Ginibre. Ce sont les matrices aléatoires complexes $M_N \in \mathbb{C}^{N \times N}$ à entrées indépendantes suivant la loi gaussienne $\mathcal{N}(0, \frac{1}{2}I_2)$. Ainsi la distribution suivie par M_N est

$$\pi^{-N^2} \exp\left(-\sum_{i,j=1}^N |z_{ij}|^2\right) \prod_{i,j=1}^N dz_{ij}. \quad (\text{III.3})$$

Soient $\lambda_1, \dots, \lambda_N$ les valeurs propres de M_N rangées par module croissant. Ginibre a montré dans [Gin65] que le N -uplet $(\lambda_1, \dots, \lambda_N)$ a pour densité

$$\varphi_N(z_1, \dots, z_N) := \frac{1}{Z_N} \exp\left(-\sum_{i=1}^N |z_i|^2\right) \prod_{1 \leq i < j \leq N} |z_i - z_j|^2$$

par rapport à la mesure de Lebesgue sur \mathbb{C}^N , où

$$Z_N := \pi^N \prod_{k=1}^N k!.$$

Grâce à cette formule, Ginibre ([Gin65], voir aussi [Meh67]) a pu montrer que la mesure spectrale empirique moyenne de $\frac{M_N}{\sqrt{N}}$ converge faiblement vers la mesure uniforme sur le disque unité $\nu_c(dz) := \frac{1}{\pi} \mathbf{1}_{|z| \leq 1} dz$. Un argument de Silverstein, publié dans [Hwa84], améliore ce résultat en prouvant la convergence presque sûre de la mesure spectrale empirique renormalisée.

Théorème III.1 (Loi circulaire). *Soit*

$$\mu'_N := \frac{1}{N} \sum_{k=1}^N \delta_{\frac{\lambda_k}{\sqrt{N}}}$$

la mesure spectrale de $\frac{M_N}{\sqrt{N}}$. Alors $\mu'_N \rightarrow \nu_c$ presque sûrement.

Ce théorème a été depuis étendu à des matrices aléatoires non hermitiennes très générales, voir [TV10] pour le résultat avec les hypothèses les plus faibles, donnant ainsi un nouvel exemple d'universalité. Voir également [BC12] pour une présentation détaillée de ces résultats.

Voyons maintenant les résultats quantitatifs autour de la loi circulaire. La vitesse de convergence de $\mathbb{E}\mu'_N$ vers ν_c en variation totale et en norme de Kolmogorov est en $\frac{1}{\sqrt{N}}$, ce qui correspond bien au taux attendu (voir [GJa20]). Mentionnons qu'une borne quasi-optimale pour des matrices générales (dont la distribution des entrées n'est pas nécessairement gaussienne) est obtenue dans [GJa20]. En ce qui concerne les distances de Wasserstein, [MM14] obtient les estimations

$$\mathbb{E}W_1(\mu'_N, \nu_c) \lesssim \frac{1}{N^{\frac{1}{4}}} \quad \text{et} \quad \mathbb{E}W_2^2(\mu'_N, \nu_c) \lesssim \sqrt{\frac{\log N}{N}},$$

ainsi que des bornes pour W_p . L'estimation L^1 est améliorée dans [CHM18] en

$$\mathbb{E}W_1(\mu'_N, \nu_c) \lesssim \sqrt{\frac{\log N}{N}}, \quad (\text{III.4})$$

mais cela ne donne pas de meilleure borne pour W_2^2 . On va montrer dans le chapitre 3 la borne optimale

$$\mathbb{E}W_2^2(\mu'_N, \nu_c) \lesssim \frac{1}{N}, \quad (\text{III.5})$$

qui permet également de remplacer $\log N$ par 1 dans (III.4). Montrons ici que (III.5) est optimale en N et plus généralement que la distance de Wasserstein W_p entre μ'_N et ν_c est au moins de l'ordre $\frac{1}{\sqrt{N}}$. Prenons $\varepsilon > 0$ et définissons $B_k := B_\varepsilon\left(\frac{\lambda_k}{\sqrt{N}}\right)$. En notant $\mathbb{D} := D(0, 1)$, on a

$$|\mathbb{D} \setminus \cup_{k=1}^N B_k| \geq (1 - N\varepsilon^2)\pi.$$

Or toute la masse de ν_c dans le domaine $\mathbb{D} \setminus \cup_{k=1}^N B_k$ doit se déplacer d'une distance supérieure à ε , ce qui donne

$$W_p^p(\mu'_N, \nu_c) \geq \frac{1}{\pi} \int_{\mathbb{D} \setminus \cup_{k=1}^N B_k} |T(z) - z|^p dz \geq \varepsilon^p (1 - N\varepsilon^2).$$

Après optimisation sur ε , on obtient

$$W_p^p(\mu'_N, \nu_c) \geq \frac{2}{p+2} \left(\frac{p}{p+2}\right)^{\frac{p}{2}} N^{-\frac{p}{2}}. \quad (\text{III.6})$$

IV Contributions de la thèse

IV.1 Une preuve variationnelle du théorème de contraction de Caffarelli

Dans le chapitre 1, on donne une preuve variationnelle du théorème de contraction II.7. La preuve originale par Caffarelli s'appuie sur l'équation de Monge-Ampère vérifiée par les fonctions de transport optimales et sur des arguments de type principe du maximum. En particulier, elle n'exploite pas directement le problème d'optimisation (I.1), ce qui est également le cas d'autres preuves comme [Kol13, KM12].

Stratégie de preuve

Pour montrer le Théorème II.7, on utilise la régularisation entropique et l'algorithme de Sinkhorn. Pour tout $\varepsilon > 0$, on définit un nouveau problème d'optimisation par perturbation de (I.2) par un terme entropique,

$$\inf_{\pi \in \Pi(\gamma_d, \nu)} \left\{ \frac{1}{2} \int |y - x|^2 d\pi + \varepsilon H(\pi \mid \gamma_d \otimes \nu) \right\}, \quad (\text{IV.1})$$

où l'entropie relative à une mesure de référence R est définie par

$$H(\pi | R) := \begin{cases} \int_{\mathbb{R}^d} \log \left(\frac{d\pi}{dR} \right) d\pi & \text{si } \pi \ll R, \\ +\infty & \text{sinon.} \end{cases}$$

On remarque que le problème d'optimisation (IV.1) est équivalent à

$$\inf_{\pi \in \Pi(\gamma_d, \nu)} \varepsilon H(\pi | \mathcal{K}^\varepsilon), \quad (\text{IV.2})$$

où la mesure de référence est

$$\mathcal{K}^\varepsilon(dx dy) := \frac{1}{(2\pi\varepsilon)^{\frac{d}{2}}} e^{-\frac{|y-x|^2}{2\varepsilon}} dy \gamma_d(dx).$$

Heuristiquement, grâce à (IV.1), il paraît assez clair que lorsque ε se rapproche de 0, le problème (IV.2) se rapproche du problème de transport optimal. C'est effectivement le cas [Mik04, Léo12, Léo14]. Il est de plus possible de montrer que la fonctionnelle $\varepsilon H(\pi | \mathcal{K}^\varepsilon)$ Γ -converge vers la fonctionnelle $\pi \mapsto \int |y-x|^2 d\pi$, voir [CDPS17, Th. 2.7], ce qui prouve que le minimiseur π^ε de (IV.2) converge vers le plan de transport optimal.

Il est de plus bien connu ([Csi75]) qu'il existe un unique minimiseur $\pi^\varepsilon \in \Pi(\gamma_d, \nu)$ de (IV.2) et que celui-ci est de la forme

$$\pi^\varepsilon(dx dy) = f^\varepsilon(x) g^\varepsilon(y) \mathcal{K}^\varepsilon(dx dy) \quad \text{avec } \log f^\varepsilon \in L^1(\gamma_d) \text{ et } \log g^\varepsilon \in L^1(\nu). \quad (\text{IV.3})$$

On a donc en particulier

$$\lim_{\varepsilon \rightarrow 0} \varepsilon H(\pi^\varepsilon | \mathcal{K}^\varepsilon) = \frac{1}{2} W_2^2(\gamma_d, \nu). \quad (\text{IV.4})$$

Les potentiels $\log f^\varepsilon$ et $\log g^\varepsilon$ sont caractérisés par la contrainte sur les marginales de π^ε , en effet pour tous x et y ,

$$\begin{aligned} \frac{1}{(2\pi\varepsilon)^{\frac{d}{2}}} f^\varepsilon(x) \int_{\mathbb{R}^d} g^\varepsilon(y) e^{-\frac{|y-x|^2}{2\varepsilon}} dy &= 1, \\ \frac{1}{(2\pi\varepsilon)^{\frac{d}{2}}} g^\varepsilon(y) \int_{\mathbb{R}^d} f^\varepsilon(x) e^{-\frac{|y-x|^2}{2\varepsilon}} e^{-\frac{|x|^2}{2}} dx &= e^{-\left(V(y) + \frac{|y|^2}{2}\right)}. \end{aligned}$$

On reconnaît dans ces formules deux semigroupes de type chaleur. En fait, pour simplifier un peu ces formules, on va plutôt choisir le semigroupe d'Ornstein-Uhlenbeck qui, étant auto-adjoint dans $L^2(\gamma_d)$, va donner deux fois le même semigroupe dans ces formules. Cela signifie que l'on va remplacer dans (IV.2) la mesure de référence \mathcal{K}^ε par

$$\mathcal{R}^\varepsilon(dx dy) := \frac{1}{(2\pi(1-e^{-\varepsilon}))^{\frac{d}{2}}} \exp\left(-\frac{|y - e^{-\frac{\varepsilon}{2}}x|^2}{2(1-e^{-\varepsilon})}\right) dy \gamma_d(dx).$$

Les propriétés restent inchangées, en particulier (IV.3) et (IV.4) sont vraies, et les équations des marginales deviennent

$$\begin{aligned} f^\varepsilon(x) P^\varepsilon g^\varepsilon(x) &= 1, \\ g^\varepsilon(y) P^\varepsilon f^\varepsilon(y) &= e^{-V(y)}, \end{aligned}$$

où P^ε est le semigroupe d’Ornstein-Uhlenbeck. Un argument utilisant un schéma itératif de type Sinkhorn permet alors de prouver que le potentiel $\log f^\varepsilon$, respectivement $\log g^\varepsilon$, est convexe, respectivement concave. Ceci repose de façon cruciale sur une propriété partagée par P^ε et le semigroupe de la chaleur, à savoir la conservation de la log-concavité. Remarquons que ce sont essentiellement les seuls semigroupes de diffusion à vérifier cela ([Kol01]).

A partir de cette propriété des potentiels régularisés, il est possible de conclure de deux façons : celle du chapitre 1 fait intervenir une caractérisation variationnelle de la régularité Lipschitz du transport optimal obtenue dans [GJu20]. Une deuxième possibilité plus directe est rendue possible par le récent article [NW21] dans lequel il est montré que les potentiels renormalisés $\varepsilon \log f^\varepsilon$ et $\varepsilon \log g^\varepsilon$ convergent bien vers les potentiels de Kantorovitch associés au problème du transport optimal de γ_d à ν , voir (I.3). La convexité étant conservée par passage à la limite simple, on obtient que le potentiel de Kantorovitch ϕ est convexe. Puisque le transport optimal s’écrit $\nabla u(x) = x - \nabla \phi(x)$, on déduit bien que $\nabla^2 u \leq I_d$. Cependant, cette deuxième stratégie ne semble pas permettre de prouver la régularité C^1 du transport optimal, au contraire de la caractérisation variationnelle de [GJu20].

IV.2 Généralisation de l’approche variationnelle de [GO20] à des coûts généraux

Le chapitre 2 est consacré à une preuve entièrement variationnelle du résultat d’ ε -régularité pour une fonction de transport optimale entre deux densités hölderiennes, avec une fonction de coût générale c , comme établi par De Philippis et Figalli dans [DF15]. Ce travail s’inscrit dans la continuité des articles [GO20], [GHO21] et [MO21]. Comme précisé dans la sous-section II.2, l’approche variationnelle a été introduite dans [GO20], établissant un résultat d’ ε -régularité dans le cas du coût quadratique $c(x, y) = \frac{1}{2}|x - y|^2$. Dans [GHO21], cette approche a été étendue au transport optimal entre des mesures générales, ce qui demandait en particulier un substitut au principe de *displacement convexity* de McCann. Même si le résultat principal de ce chapitre traite d’un transport optimal entre deux densités, comme dans [GO20], cette généralisation va se révéler cruciale ici. En effet, la preuve de [GO20, Lemme 4.2], qui montre que l’interpolation (I.4) est contrôlée en norme L^∞ , repose sur l’équation de Monge-Ampère et donc sur la structure du transport pour le coût quadratique. Ceci permet ensuite de montrer que le flux (I.5) est contrôlé en norme L^2 , ce qui est à la base du résultat d’approximation harmonique [GO20, Prop. 4.3].

Soient $X, Y \subset \mathbb{R}^d$ deux compacts et $f, g : \mathbb{R}^d \rightarrow \mathbb{R}$ deux densités de probabilité, avec $\text{Spt } f \subset X$ and $\text{Spt } g \subset Y$. Supposons que la fonction de coût $c : X \times Y \rightarrow \mathbb{R}$ satisfait les hypothèses (C1)–(C4). Considérons de plus la situation normalisée suivante : $0 \in \text{Spt } f \cap \text{Spt } g$, $f(0) = g(0) = 1$ et $\nabla_{xy} c(0, 0) = -I$. Le résultat principal du chapitre est le résultat d’ ε -régularité suivant.

Théorème IV.1. *Soit $\pi \in \Pi(f, g)$ c -optimal. Il existe $R_0 = R_0(c) > 0$ et $\varepsilon = \varepsilon(c) > 0$ tels que pour tout $R \leq R_0$ satisfaisant*

$$\frac{1}{R^{d+2}} \int_{B_{4R} \times \mathbb{R}^d} |x - y|^2 d\pi + R^{2\alpha} \left([f]_{\alpha, 4R}^2 + [g]_{\alpha, 4R}^2 + [\nabla_{xy} c]_{\alpha, 4R}^2 \right) \leq \varepsilon, \quad (\text{IV.5})$$

il existe une fonction $T \in C^{1,\alpha}(B_R)$ telle que $(B_R \times \mathbb{R}^d) \cap \text{Spt } \pi \subset \text{graph } T$.

Les valeurs de R_0 et ε ne dépendent du coût c qu'à travers les hypothèses qualitatives (C1)–(C4) sur c . Remarquons que l'on n'utilise pas la structure déterministe du plan optimal π , comme décrite dans la sous-section I.2, mais que cette structure est en fait une conséquence du Théorème IV.1. Cependant, dans la preuve du résultat préliminaire crucial établissant une borne L^∞ a priori sur le déplacement $|y - x|$ dans le support de π , on s'appuie sur la c -monotonie de $\text{Spt } \pi$, notion proche du caractère déterministe de π .

Un corollaire du Théorème IV.1 est le Théorème II.6, établissant que le transport optimal est un $C^{1,\alpha}$ -difféomorphisme entre deux ouverts de mesure pleine. En effet, l'hypothèse (C1) et le fait que Y est borné assurent que le potentiel c -convexe u (voir la sous-section I.2 pour une présentation de la structure du transport optimal avec un coût général) est semi-convexe, c'est-à-dire qu'il existe une constante $C < \infty$ telle que $u + C|x|^2$ soit convexe. Ainsi, le théorème d'Alexandrov s'applique ici : u est presque partout deux fois différentiable. Le Théorème II.6 s'obtient alors en combinant la preuve de [GO20, Th. 1.1] et les changements de variable utilisés dans [DF15].

Une motivation importante pour considérer des coûts généraux est l'étude du transport optimal sur des variétés riemanniennes avec un coût $c(x, y) := \frac{1}{2}d(x, y)^2$, où d est la distance riemannienne. Dans ce contexte, un résultat d' ε -régularité est particulièrement intéressant puisque, même si c est une perturbation compacte du cas euclidien, d'autres mécanismes peuvent créer des singularités, comme la courbure. En effet, la condition MTW (voir sous-section II.3) est très restrictive, même pour le coût $\frac{d^2}{2}$. Il est donc nécessaire de disposer d'une théorie de régularité qui s'affranchisse de conditions supplémentaires sur le coût et de critères géométriques sur les supports. Le Théorème II.6 est utilisé dans [DF15] pour montrer que si f et g sont deux densités $C^{0,\alpha}$ sur la variété avec $f, g, \frac{1}{f}$ et $\frac{1}{g}$ bornées, le transport optimal entre f et g avec coût $\frac{d^2}{2}$ est un $C^{1,\alpha}$ -difféomorphisme entre deux ouverts de mesure pleine sur la variété.

La condition (C4) n'est pas satisfaite par la famille de coûts $c_p(x, y) := |x - y|^p$ si $p \neq 2$, notre résultat ne s'applique donc pas. Une borne L^∞/L^p a priori sur le déplacement est établie dans le très récent article [GM21], ce qui constitue une première étape cruciale pour généraliser le Théorème IV.1 aux coûts c_p . Une telle borne L^∞/L^p , mais seulement globale, a auparavant été obtenue dans [BJR07].

Stratégie de preuve

La preuve du Théorème IV.1 suit de près celle de [GO20], esquissée dans la sous-section II.2. La principale nouveauté est l'introduction, dans le contexte du transport optimal, de la notion de *quasi-minimiseur*, déjà bien établie dans le contexte des surfaces minimales ([Alm76, Bom82]). L'idée est de montrer qu'à une échelle donnée, un plan minimiseur du transport optimal pour le coût c minimise de façon approchée (c'est-à-dire à une erreur près, cette erreur s'améliorant lorsque l'échelle diminue) le transport en coût quadratique. Grâce à cette propriété, on peut appliquer directement le résultat d'approximation harmonique obtenu dans [GHO21].

Il y a essentiellement deux problèmes importants lorsque l'on veut transposer la stratégie de [GO20] dans ce cadre général :

- Le résultat préliminaire crucial est une borne L^∞ a priori sur le déplacement dans le

support de π ([GO20, Lem. 3.1], [GHO21, Lem. 2.9]). Cette estimation en fonction de l'énergie repose sur la monotonie du support du couplage optimal. Ici, on établit un analogue de cette borne L^∞ en s'appuyant sur la c -monotonie du support du couplage c -optimal, voir I.6. Cependant, il est nécessaire d'établir d'abord une borne qualitative sur le déplacement, obtenue pour toute échelle inférieure à un seuil R_0 dépendant du coût à travers les hypothèses (C1)–(C4). On vérifie ensuite que cette borne L^∞ qualitative se reproduit à chaque étape de l'itération.

- L'approximation harmonique [GHO21, Th. 1.5] s'effectue dans le cadre eulérien du transport optimal. Le passage du cadre lagrangien au cadre eulérien, et vice-versa, est facilité par l'équivalence (I.8) entre énergie lagrangienne $\int |y - x|^2 d\pi$ et énergie eulérienne $\int \frac{1}{\rho} |j|^2$ si π est un couplage optimal pour le coût quadratique et (ρ, j) sa paire densité-flux associée (définitions (I.4) et (I.5)). Il se trouve que cette équivalence n'est plus vraie si π est un couplage général. L'énergie eulérienne est toujours dominée par l'énergie lagrangienne, mais cette domination est en général stricte. On a pu montrer néanmoins que si π est un couplage quasi-minimiseur, l'erreur entre les deux énergies peut être estimée par l'erreur dans la propriété de quasi-minimiseur de π . Ceci nous permet alors de passer d'un cadre à l'autre de la même façon que dans [GHO21].

Comparaison avec [DF15]

Notons tout d'abord que les hypothèses (C1)–(C4) sur la fonction de coût sont les mêmes que dans [DF15]. Ensuite, comme on l'a vu dans la sous-section II.3, la stratégie que l'on suit est proche de celle de [DF15]. Mentionnons tout de même que comme dans [GO20], on obtient dans le Théorème IV.1 une estimation de la semi-norme $C^{0,\alpha}$ sous la forme

$$R^{2\alpha} [\nabla T]_{\alpha,R}^2 \lesssim \frac{1}{R^{d+2}} \int_{B_{4R} \times \mathbb{R}^d} |x - y|^2 d\pi + R^{2\alpha} \left([f]_{\alpha,4R}^2 + [g]_{\alpha,4R}^2 + [\nabla_{xy} c]_{\alpha,4R}^2 \right),$$

qui est linéaire en les données. De plus, l'itération de Campanato que l'on effectue au terme de l'argument se fait directement au niveau $C^{1,\alpha}$ de T , à la différence de [DF15] où trois itérations sont nécessaires, d'abord pour obtenir $C^{0,1^-}$, puis $C^{0,1}$ et finalement $C^{1,\alpha}$.

On remarque également que l'hypothèse (IV.5) que l'on fait sur les données n'est pas plus restrictive que celle faite dans [DF15, Th. 4.3 & 5.3]. Les hypothèses sur les densités et le coût sont essentiellement les mêmes, mais alors que l'on mesure l'écart entre le transport et l'identité en norme L^2 , dans [DF15] les auteurs se placent au niveau du potentiel c -convexe u , qu'ils supposent proche en norme L^∞ de la fonction $\frac{1}{2}|x|^2$. On montre dans le Corollaire 2.1.3 que dans la situation où $\nabla_{xx} c(0,0) = 0$ et $R^\alpha [\nabla_{xx} c]_{\alpha,R} \ll 1$ (ce qui veut dire que le coût est considéré comme proche de $-x \cdot y$, ce qui n'est pas forcément le cas dans le Théorème II.5), notre hypothèse découle de celle faite dans [DF15].

Densités continues

On traite également dans la sous-section 2.4.3 le cas où les densités sont seulement continues, pour lequel on prouve comme dans [DF15] que le transport optimal est $C^{0,\beta}$ pour tout $\beta > 0$ lorsque que la situation est normalisée et que les données habituelles sont suffisamment petites.

La preuve de ce résultat d' ε -régularité est très proche de la précédente et les modifications suivent [Gol20]. Il apparaît clairement que le coût doit encore être $C^{2,\alpha}$ pour obtenir une régularité $C^{0,\beta}$ de T , puisque l'itération ne peut pas se faire indéfiniment si l'on suppose uniquement $c \in C^2$ avec l'hypothèse que la distance $\|\nabla_{xy}c + \mathbb{I}\|_\infty$ est petite localement.

Quasi-minimiseurs

Enfin, on montre dans la sous-section 2.4.4 comment il est possible d'obtenir le Théorème IV.1 en relaxant l'hypothèse de minimalité : le résultat est encore vrai si l'on suppose seulement que π est un couplage déterministe quasi-minimiseur à toutes échelles. L'obstacle majeur ici est la borne L^∞/L^2 a priori sur le déplacement. En effet, la preuve de ce résultat a toujours reposé sur la monotonie du support ([GO20, Lem. 3.1], [GHO21, Lem. 2.9] et [MO21, Prop. 2.2]) ou sur sa c -monotonie dans le Lemme 2.1.5. Si le transport est quasi-minimiseur, il n'y a a priori pas de raison que son graphe vérifie une quelconque propriété de monotonie. On a néanmoins pu montrer qu'un tel transport vérifie une borne L^p/L^2 sur le déplacement pour tout $p < \infty$, ce qui est suffisant pour notre preuve. En effet, il nous est possible de montrer que l'énergie associée aux trajectoires de particules trop excentriques, c'est-à-dire celles qui dépassent la borne L^∞ que l'on attend, est très petite par rapport à l'énergie globale du transport. Ainsi, la part significative de l'énergie concerne les trajectoires satisfaisant la borne L^∞ et on peut donc appliquer notre stratégie au couplage amputé des mauvaises trajectoires. Il reste ensuite à combiner les deux pour obtenir le même résultat d'approximation harmonique.

IV.3 Vitesse de convergence optimale vers la loi circulaire pour l'ensemble de Ginibre

On s'intéresse dans le chapitre 3 au transport optimal de la mesure ν_c uniforme sur le disque unité \mathbb{D} à la mesure spectrale renormalisée d'une matrice choisie selon la loi de Ginibre (III.3). D'après le Théorème III.1, une estimation de la distance de Wasserstein entre les deux fournit la vitesse de convergence de cette mesure empirique vers sa limite ν_c . Le résultat principal obtenu ici est une estimation optimale (par rapport à la taille N de la matrice) de l'espérance de la distance de Wasserstein W_2 entre ces deux mesures.

Plus précisément, si M_N est une matrice de l'ensemble de Ginibre, c'est-à-dire distribuée selon (III.3), on définit sa mesure spectrale renormalisée

$$\mu'_N := \frac{1}{N} \sum_{k=1}^N \delta_{\frac{\lambda_k}{\sqrt{N}}}, \quad (\text{IV.6})$$

dont les atomes sont les valeurs propres de $\frac{M_N}{\sqrt{N}}$. D'après le Théorème III.1, la mesure limite est alors la mesure $\nu_c(dz) := \frac{1}{\pi} \mathbf{1}_{\mathbb{D}}(z) dz$, uniforme sur le disque unité \mathbb{D} . On montre le résultat suivant.

Théorème IV.2. *Il existe une constante $C < \infty$ telle que pour tout $N \geq 2$,*

$$\mathbb{E}W_2^2(\mu'_N, \nu_c) \leq \frac{C}{N}.$$

Au vu de (III.6), ce résultat est optimal par rapport à N et améliore, pour la distance W_2 , l'estimation obtenue dans [MM14]. Par l'inégalité de Hölder et puisque les mesures considérées sont de masse 1, on obtient immédiatement une estimation optimale de la distance W_p pour tout $1 \leq p \leq 2$:

Corollaire IV.1. *Il existe $C < \infty$ telle que pour tout $1 \leq p \leq 2$, pour tout $N \geq 2$,*

$$\mathbb{E}W_p(\mu'_N, \nu_c) \leq \frac{C}{\sqrt{N}}.$$

Stratégie de preuve

Pour prouver le Théorème III.1, on va en fait se placer au niveau de la mesure spectrale

$$\mu_N := \sum_{k=1}^N \delta_{\lambda_k}$$

et prouver que la distance $W_2(\mu_N, \nu)$ entre μ_N et la mesure $\nu := \frac{1}{\pi} \mathbf{1}_{\mathbb{D}_N}(z) dz$ uniforme sur le disque $\mathbb{D}_N := D(0, \sqrt{N})$ est en moyenne de l'ordre de N . L'estimation de cette distance passe par une mesure intermédiaire ν_N définie par

$$\nu_N := \frac{\mu_N(\mathbb{D}')}{|\mathbb{D}'|} dz|_{\mathbb{D}'} + \frac{\mu_N(\mathbb{C} \setminus \mathbb{D}')}{|\mathbb{D}_N \setminus \mathbb{D}'|} dz|_{\mathbb{D}_N \setminus \mathbb{D}'},$$

où $\mathbb{D}' := D(0, \sqrt{N}(1 - \varepsilon_N))$ avec $\varepsilon_N := \sqrt{\frac{\ln N}{N}}$. Cette distance ε_N est choisie pour deux raisons. La première est que les calculs que l'on utilise pour estimer la variance du nombre de points dans un carré nécessitent que ce carré soit à distance au moins $\sqrt{N}\varepsilon_N$ du bord de \mathbb{D}_N et la seconde est que la distance typique entre le rayon spectral $\rho(M_N)$ et le bord du disque \mathbb{D}_N est de l'ordre de $\sqrt{N}\varepsilon_N$, voir [Rid03].

On montre alors que la distance moyenne $\mathbb{E}W_2(\nu_N, \nu)$ est négligeable devant \sqrt{N} . Il reste donc à estimer la distance $\mathbb{E}W_2(\mu_N, \nu_N)$. Pour cela, on établit que les deux quantités

$$\mathbb{E}W_2(\mu_N|_{\mathbb{D}'}, \nu_N|_{\mathbb{D}'}) \quad \text{et} \quad \mathbb{E}W_2(\mu_N|_{\mathbb{C} \setminus \mathbb{D}'}, \nu_N|_{\mathbb{D}_N \setminus \mathbb{D}'})$$

sont majorées par N , en découpant les deux domaines \mathbb{D}' et $\mathbb{D}_N \setminus \mathbb{D}'$ en sous-domaines de tailles dyadiques, jusqu'à l'échelle microscopique. On utilise alors à chaque échelle les estimations sur la variance du nombre de points obtenues plus tôt. Cette stratégie de construction itérative m'a été suggérée, dans le cas du carré, par F. Otto.

Concentration et comportement du transport optimal

Le chapitre 3 contient également la preuve d'un résultat de concentration de la distance $W_2(\mu_N, \nu)$ lorsque l'on considère les mesures restreintes à un carré.

Théorème IV.3. *Il existe deux constantes $C < \infty$ et $c > 0$ telles que pour tout carré $Q \subset D(0, \sqrt{N} - \sqrt{\ln N})$ de côté $L \geq 1$, on ait pour tout $t \gg L\sqrt{\ln L}$,*

$$\mathbb{P}_N \left(W_Q \left(\mu_N, \frac{\mu_N(Q)}{|Q|} \right) \geq t \right) \leq C \exp \left(-c \frac{t^2}{L^2} \right). \quad (\text{IV.7})$$

La même stratégie que pour un carré permettrait de prouver une inégalité de concentration pour un disque D centré en 0 de rayon $R_N \leq \sqrt{N} - \sqrt{\ln N}$ de la forme

$$\mathbb{P}_N \left(W_D \left(\mu_N, \frac{\mu_N(D)}{|D|} \right) \geq t \right) \lesssim \exp \left(-c \frac{t^2}{R_N^2} \right).$$

En revanche, pour un anneau A au bord du disque de rayon \sqrt{N} , la contrainte sur le rayon spectral limite ce que l'on peut obtenir. En effet, écrivons

$$\begin{aligned} & \mathbb{P}_N \left(W_2 \left(\mu_N \llcorner_{\mathbb{C} \setminus \mathcal{D}'}, \frac{\mu_N(\mathbb{C} \setminus \mathcal{D}')}{|A|} dz \llcorner_A \right) \geq t \right) \\ & \leq \mathbb{P}_N \left(\left\{ W_2 \left(\mu_N \llcorner_{\mathbb{C} \setminus \mathcal{D}'}, \frac{\mu_N(\mathbb{C} \setminus \mathcal{D}')}{|A|} dz \llcorner_A \right) \geq t \right\} \cap \left\{ \rho(M_N) \leq r\sqrt{N} \right\} \right) \end{aligned} \quad (\text{IV.8})$$

$$+ \mathbb{P}_N \left(\rho(M_N) > r\sqrt{N} \right). \quad (\text{IV.9})$$

La probabilité (IV.9) peut être estimée par $N^{-\frac{M}{2}}$ si $r = 1 + M\varepsilon_N$ ([Rid03]) ou par e^{-Nr^2} si $r \gg 1$ ([CHM18]). Dans les deux cas, le choix de r pour que (IV.9) soit suffisamment petit implique que l'échelle $\theta_0 R_N$ soit de l'ordre de R_N , ce qui ne permet pas d'obtenir une estimation intéressante pour (IV.8). Pour cette raison, on ne donne pas d'inégalité de concentration pour la loi circulaire.

On peut cependant utiliser le Théorème IV.3 pour obtenir des informations sur le comportement de la fonction de transport optimale. En effet, l'inégalité de concentration (IV.7) est la même que celle obtenue dans [GHO18, Prop. 2.7] pour un processus de points de Poisson. Cette inégalité permet de montrer l'existence d'une variable aléatoire r_* satisfaisant

$$\mathbb{E} \left[\exp \left(\frac{cr_*^2}{\ln(2r_*)} \right) \right] < \infty,$$

telle que pour tout l dyadique avec $2r_* \leq l \leq \sqrt{N}$, on ait

$$\frac{1}{l^4} W_{Q_l}^2 \left(\mu_N, \frac{\mu_N(Q_l)}{|Q_l|} \right) \lesssim \frac{\ln \left(\frac{l}{r_*} \right)}{\left(\frac{l}{r_*} \right)^2}.$$

Ce résultat, combiné à la théorie développée dans [GHO21] et étendue dans le chapitre 2 de cette thèse, permet de prouver que pour $L \sim \sqrt{N}$, il existe une variable aléatoire $r_{*,L}$ satisfaisant

$$\sup_N \mathbb{E} \left[\exp \left(\frac{cr_{*,L}^2}{\ln(2r_{*,L})} \right) \right] < \infty \quad (\text{IV.10})$$

et telle que si T réalise le transport optimal en coût quadratique de la mesure de Lebesgue sur Q_L à $\mu_N \llcorner_{Q_L}$, on ait pour tout $2r_{*,L} \leq l \leq L$,

$$\frac{1}{l^4} \int_{B_l(x_L)} |T - (x - x_L)|^2 \lesssim \frac{\ln^3 \left(\frac{L}{r_{*,L}} \right)}{\left(\frac{L}{r_{*,L}} \right)^2}.$$

Ainsi, entre une échelle $2r_{*,L}$, proche de l'échelle microscopique par (IV.10), et l'échelle macroscopique, le transport optimal est quantitativement proche d'une fonction affine $x - x_L$ en norme L^2 . De plus, la dérive x_L est estimée par

$$|x_L|^2 \lesssim r_{*,L}^2 \ln^3 \left(\frac{L}{r_{*,L}} \right).$$

Ce comportement correspond (à l'exposant du logarithme près) à ce à quoi on s'attend dans le cas du processus de Poisson car on sait que la distance moyenne de transport diverge logarithmiquement à grande échelle. Dans notre cas, il serait intéressant de savoir si l'on peut obtenir une borne non divergente sur la dérive.

CHAPITRE 1

A proof of the Caffarelli contraction theorem via entropic regularization

Contents

1.1	Introduction	41
1.2	Entropic transport costs and main results	43
1.2.1	Entropic costs and their zero-noise limit	43
1.2.2	Entropic cost in the framework of Caffarelli theorem	44
1.3	Proofs	47
1.4	Variational proof of Theorem 1.1.2	52

1.1 Introduction

The aim of this chapter is to give a new proof of the celebrated Caffarelli contraction theorem [Caf00, Caf02], which states that the Brenier optimal transport map sending the standard Gaussian measure on \mathbb{R}^d , denoted by γ_d in all the chapter, onto a probability measure ν having a log-concave density with respect to γ_d is a contraction. More precisely, let us recall the generalized version of Caffarelli's theorem:

Theorem 1.1.1 (Caffarelli [Caf00, Caf02]). *For any probability measures μ and ν respectively of the form $\mu(dx) = e^{V(x)}\gamma_d(dx)$ and $\nu(dx) = e^{-W(x)}\gamma_d(dx)$ with V and W convex functions, and further assuming μ has a finite second moment and ν is compactly supported, there exists a continuously differentiable and convex function $\phi : \mathbb{R}^d \rightarrow \mathbb{R}$ such that $\nabla\phi$ is 1-Lipschitz and $\nu = \nabla\phi\#\mu$.*

Caffarelli's original result was only stated for the important particular case where μ is the Gaussian measure γ_d (i.e. $V = 0$), but his proof readily extends to this more general setting [Kol13]. Note that the assumption that ν is compactly supported can be removed via

approximation. See [Vil09, Corollary 5.21] for details. Note that in all the chapter we allow convex function to take the value $+\infty$.

Here, we develop an approach based on a variational characterization of Lipschitz regularity of optimal transport maps obtained by Gozlan and Juillet in [GJu20]. To recall this result, we need to introduce the following definition: if $\eta_1, \eta_2 \in \mathcal{P}_1(\mathbb{R}^d)$, one says that η_1 is dominated by η_2 for the convex order if $\int f d\eta_1 \leq \int f d\eta_2$ for all convex function $f : \mathbb{R}^d \rightarrow \mathbb{R}$. In this case, we write $\eta_1 <_{\text{conv}} \eta_2$. With this notation in hand, the variational characterization of [GJu20] reads as follows:

Theorem 1.1.2. *Let $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$; the following assertions are equivalent:*

- (i) *There exists a continuously differentiable and convex function $\phi : \mathbb{R}^d \rightarrow \mathbb{R}$ such that $\nabla\phi$ is 1-Lipschitz and $\nu = \nabla\phi\#\mu$,*
- (ii) *For all $\eta \in \mathcal{P}_2(\mathbb{R}^d)$ such that $\eta <_{\text{conv}} \nu$,*

$$W_2(\nu, \mu) \leq W_2(\eta, \mu).$$

In other words the Brenier map between μ and ν is a contraction if and only if ν is the closest point to μ among all probability measures dominated by ν in the convex order. We will give an alternative proof of this Theorem in Section 1.4 in the particular case where the support of ν is convex (which is enough for our purpose here) using Kantorovich duality and variational arguments.

Our strategy to recover Theorem 1.1.1 is thus to show the following monotonicity property of the W_2 distance: if μ and ν satisfy the assumptions of Theorem 1.1.1, it holds

$$W_2(\nu, \mu) \leq W_2(\eta, \mu), \quad \forall \eta <_{\text{conv}} \nu. \tag{1.1.1}$$

For that purpose, we will establish a similar inequality at an information theoretic level replacing W_2 by the so called entropic transport cost $\mathcal{T}_H^\varepsilon$ (presented in details in the next section) that is defined in terms of the minimization of the relative entropy between π and a reference measure R^ε involving some small noise parameter $\varepsilon > 0$. We will prove the following monotonicity property of the entropic cost:

$$\mathcal{T}_H^\varepsilon(\nu, \mu) \leq \mathcal{T}_H^\varepsilon(\eta, \mu) \tag{1.1.2}$$

for all $\eta <_{\text{conv}} \nu$ with a finite Shannon information. As observed by Mikami [Mik04] and extensively developed by Léonard [Léo12, Léo14] the zero noise limit of $\varepsilon\mathcal{T}_H^\varepsilon$ is $\frac{1}{2}W_2^2$. Thus letting $\varepsilon \rightarrow 0$ in (1.1.2) will give (1.1.1).

This first chapter is organized as follows. Section 1.2 introduces entropic transport costs and the main results. Section 1.3 gives proofs of these results. Section 1.4 presents the alternative proof of Theorem 1.1.2.

1.2 Entropic transport costs and main results

1.2.1 Entropic costs and their zero-noise limit

Consider the classical Ornstein-Uhlenbeck process $(Z_t)_{t \geq 0}$ on \mathbb{R}^d , defined by the following stochastic differential equation:

$$dZ_t = -\frac{1}{2}Z_t dt + dW_t, \quad t \geq 0,$$

where $(W_t)_{t \geq 0}$ is a standard d dimensional Brownian motion and $Z_0 \sim \gamma_d$. As it is well known, the process Z admits the following explicit representation

$$Z_t = Z_0 e^{-t/2} + e^{-t/2} \int_0^t e^{s/2} dW_s, \quad t \geq 0.$$

The joint law of (Z_0, Z_ε) will be denoted by R^ε . It is therefore given by

$$R^\varepsilon = \text{Law} \left(X, X e^{-\varepsilon/2} + \sqrt{1 - e^{-\varepsilon}} Y \right),$$

with X, Y two independent standard Gaussian random vectors on \mathbb{R}^d . In other words,

$$R^\varepsilon(dx dy) = \gamma_d(dx) r_x^\varepsilon(dy),$$

where $x \mapsto r_x^\varepsilon$ is the probability kernel defined by $r_x^\varepsilon = \mathcal{N}(x e^{-\varepsilon/2}, (1 - e^{-\varepsilon}) I_d)$.

Recall that the relative entropy of a probability measure α with respect to another probability measure β on some measurable space $(\mathcal{X}, \mathcal{A})$ is defined by

$$H(\alpha|\beta) = \int \log \left(\frac{d\alpha}{d\beta} \right) d\alpha,$$

if α is absolutely continuous with respect to β . If this is not the case, one sets $H(\alpha|\beta) = +\infty$. The relative entropy is a non-negative quantity that vanishes only when the two probability measures are equal, this is why it is often called Kullback-Leibler distance (even though it is not a true distance).

Definition 1.2.1 (Entropic transport cost). For all probability measures μ, ν on \mathbb{R}^d , the entropic transport cost associated to R^ε is defined by

$$\mathcal{T}_H^\varepsilon(\mu, \nu) = \inf_{\pi \in \Pi(\mu, \nu)} H(\pi|R^\varepsilon).$$

As shown by Mikami, Léonard and others [Mik04, Léo12] the zero noise limit of $\varepsilon \mathcal{T}_H^\varepsilon$ is $\frac{1}{2} W_2^2$. At a heuristic level, this phenomenon can be easily understood from the following identities:

$$\begin{aligned} \varepsilon H(\pi|R^\varepsilon) &= \varepsilon \int \log \left(\frac{d\pi}{dx} \right) d\pi - \varepsilon \int \log \left(\frac{dR^\varepsilon}{dx} \right) d\pi \\ &= \varepsilon \int \log \left(\frac{d\pi}{dx} \right) d\pi + \frac{\varepsilon}{2(1 - e^{-\varepsilon})} \int |y - e^{-\varepsilon/2} x|^2 \pi(dx dy) + \frac{\varepsilon}{2} \int |x|^2 \mu(dx) + c(\varepsilon), \end{aligned}$$

where $c(\varepsilon) \rightarrow 0$ (and is independent of μ, ν, π). So for small ε , minimizing $\pi \mapsto H(\pi|R^\varepsilon)$ amounts to minimizing $\pi \mapsto \frac{1}{2} \int |x - y|^2 \pi(\mathrm{d}x\mathrm{d}y)$.

In the sequel we will use the following result, which can be easily deduced from a general convergence theorem due to Carlier, Duval, Peyré and Schmitzer [CDPS17, Theorem 2.7]. We will say that a probability measure η is of finite (Shannon) entropy if it is absolutely continuous with respect to the Lebesgue measure and if $\int \log\left(\frac{\mathrm{d}\eta}{\mathrm{d}x}\right) \mathrm{d}\eta$ is finite. Note that, if $\eta \in \mathcal{P}_2(\mathbb{R}^d)$, then it is of finite entropy if and only if $H(\eta|\gamma_d) < +\infty$.

Theorem 1.2.2 (Carlier et al. [CDPS17]). *Suppose that $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$ are of finite entropy. Then, it holds*

$$\varepsilon \mathcal{T}_H^\varepsilon(\mu, \nu) \rightarrow \frac{1}{2} W_2^2(\mu, \nu) \quad \text{as } \varepsilon \rightarrow 0.$$

We state now a technical lemma that will be needed to apply Theorem 1.2.2 in our framework:

Lemma 1.2.3. *If μ and ν satisfy the assumptions of Theorem 1.1.1, then they are of finite entropy.*

The proof is postponed to Section 1.3.

1.2.2 Entropic cost in the framework of Caffarelli theorem

As explained above, the key step in our approach consists in showing that on the set of probability measures dominated by ν in the convex order, the closest point to μ for the entropic cost distance is ν itself (when ν satisfies the assumptions of Theorem 1.1.1).

Theorem 1.2.4. *Let μ and ν satisfy the assumptions of Theorem 1.1.1. Additionally assume that V is bounded from below. If η is such that $\eta \prec_{\text{conv}} \nu$, then for all $\varepsilon > 0$*

$$\mathcal{T}_H^\varepsilon(\mu, \eta) \geq \mathcal{T}_H^\varepsilon(\mu, \nu).$$

Let us admit Theorem 1.2.4 for a moment and complete the proof of Theorem 1.1.1.

Proof of Theorem 1.1.1. Let us temporarily assume that V is bounded from below. According to Lemma 1.2.3, μ and ν have finite entropy. So using Theorem 1.2.2, one concludes by letting $\varepsilon \rightarrow 0$ that for all compactly supported probability measures ν of the form $\nu(\mathrm{d}x) = e^{-W(x)} \gamma_d(\mathrm{d}x)$, with $W : \mathbb{R}^d \rightarrow \mathbb{R} \cup \{+\infty\}$ convex, it holds

$$W_2(\mu, \nu) \leq W_2(\mu, \eta)$$

for all η of finite entropy and such that $\eta \prec_{\text{conv}} \nu$. Now, fix some compactly supported ν_0 of the form $\nu_0(\mathrm{d}x) = e^{-W_0(x)} \gamma_d(\mathrm{d}x)$, with $W_0 : \mathbb{R}^d \rightarrow \mathbb{R} \cup \{+\infty\}$ convex and let us show that the inequality (1.1.1) holds for any $\eta \prec_{\text{conv}} \nu$. Take $\eta \prec_{\text{conv}} \nu_0$ and define, for all $\theta \in (0, \pi/2)$,

$$\nu_\theta = \text{Law}(\cos \theta X + \sin \theta Z) \quad \text{and} \quad \eta_\theta = \text{Law}(\cos \theta Y + \sin \theta Z),$$

where $X \sim \nu_0$, $Y \sim \eta$ and Z is independent of X and Y and has density $\frac{1}{C} \mathbf{1}_B(x) e^{-\frac{|x|^2}{2}}$, where B is the Euclidean unit ball. According to Lemma 1.3.6, ν_θ is compactly supported and of

the form $e^{-W_\theta} \gamma_d$, with W_θ convex, η_θ is of finite entropy and $\eta_\theta \prec_{\text{conv}} \nu_\theta$. Therefore, it holds $W_2(\mu, \nu_\theta) \leq W_2(\mu, \eta_\theta)$. Letting $\theta \rightarrow 0$ gives $W_2(\mu, \nu_0) \leq W_2(\mu, \eta)$ for all $\eta \prec_{\text{conv}} \nu_0$, which, according to Theorem 1.1.2, completes the proof when μ has finite entropy.

Finally, let us remove the assumption that V is bounded from below. Since V is convex, it is bounded from below by some affine function. Thus there exists some $a \in \mathbb{R}^d$ such that $x \mapsto V(x) + a \cdot x$ is bounded from below. Consider the probability measure $\tilde{\mu}$ defined as the push forward of μ under the translation $x \mapsto x + a$. An easy calculation shows that the density of $\tilde{\mu}$ with respect to γ_d is $C e^{V(x-a) + a \cdot (x-a)}$, with C a normalizing constant, and so $\tilde{\mu}$ satisfies our assumptions. Therefore, there exists a continuously differentiable convex function $\tilde{\phi} : \mathbb{R}^d \rightarrow \mathbb{R}$ such that $\nabla \tilde{\phi}$ is 1-Lipschitz and $\nu = \nabla \tilde{\phi} \# \tilde{\mu}$. Setting $\phi(x) = \tilde{\phi}(x + a)$, $x \in \mathbb{R}^d$, one gets $\nu = \nabla \phi \# \mu$ which completes the proof. ■

Before proving Theorem 1.2.4, let us informally explain why one can guess the statement is easier to prove at the level of entropic cost than directly for the Wasserstein distance. If we consider the plain relative entropy, we have the variational formula

$$H(\rho|\mu) = \sup \int f \, d\rho - \log \int e^f \, d\mu,$$

where the supremum runs over the set of functions f such that $\int e^f \, d\mu < +\infty$. Hence, taking $f = -(V + W)$, gives

$$H(\rho|\mu) \geq \int -(V + W) \, d\rho \geq \int -(V + W) \, d\nu = H(\nu|\mu)$$

as soon as $\rho \prec_{\text{conv}} \nu$. So this trivial bound hints at the fact that comparison is easier for entropies when we have a log concavity condition on the relative density.

To prove Theorem 1.2.4, we need to know more about the optimal coupling π for $\mathcal{T}_H^\varepsilon(\mu, \nu)$. The following is classical in entropic transport literature and goes back to the study of the so called Schrödinger bridges [Sch32].

Proposition 1.2.5. *Let $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$ be such that $H(\mu|\gamma_d) < +\infty$ and $H(\nu|\gamma_d) < +\infty$*

1. *There exists a unique coupling $\pi^\varepsilon \in \Pi(\mu, \nu)$ such that*

$$\mathcal{T}_H^\varepsilon(\mu, \nu) = H(\pi^\varepsilon|R^\varepsilon) < +\infty$$

2. *There exist two measurable functions $f^\varepsilon, g^\varepsilon : \mathbb{R}^d \rightarrow \mathbb{R}^+$ such that $\log f^\varepsilon \in L^1(\mu)$, $\log g^\varepsilon \in L^1(\nu)$ and*

$$\pi^\varepsilon(dx dy) = f^\varepsilon(x) g^\varepsilon(y) R^\varepsilon(dx dy).$$

Sketch of proof. (1) We equip the set $\mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$ with the usual topology of narrow convergence. For this topology, the function $\pi \mapsto H(\pi|R^\varepsilon)$ is lower-semicontinuous and the set $\Pi(\mu, \nu)$ is compact. Therefore, the function $H(\cdot |R^\varepsilon)$ attains its minimum at some point π^ε of $\Pi(\mu, \nu)$. It is easily checked that the coupling $\pi_0 = \mu \otimes \nu$ is such that $H(\pi_0|R^\varepsilon) < +\infty$, so $H(\pi^\varepsilon|R^\varepsilon) < +\infty$. Uniqueness comes from the strict convexity of $H(\cdot |R^\varepsilon)$. For the proof of (2) we refer to [Csi75, Corollary 3.2]. In the special case where μ and ν satisfy our log-convexity/concavity assumptions we will give a self-contained proof in Section 1.3. ■

In the setting of Theorem 1.1.1, it turns out that much more can be said about the functions f and g . This is explained in the following result, which seems of independent interest.

Theorem 1.2.6. *With the same notation as in Proposition 1.2.5, let μ be a probability measure of the form $\mu(dx) = e^{V(x)}\gamma_d(dx)$ with a finite second moment and ν be a compactly supported probability measure on \mathbb{R}^d of the form $\nu(dx) = e^{-W(x)}\gamma_d(dx)$, with V, W convex and V bounded from below. There exist a log-convex function $f^\varepsilon : \mathbb{R}^d \rightarrow [1, +\infty)$ and a log-concave function $g^\varepsilon : \mathbb{R}^d \rightarrow [0, +\infty)$ such that the unique optimal coupling $\pi^\varepsilon \in \Pi(\mu, \nu)$ is of the form $\pi^\varepsilon(dxdy) = f^\varepsilon(x)g^\varepsilon(y)R^\varepsilon(dxdy)$. Moreover, the function $\log f^\varepsilon$ is integrable with respect to μ and the function $\log g^\varepsilon$ is integrable with respect to ν and it holds*

$$\mathcal{T}_H^\varepsilon(\mu, \nu) = H(\pi^\varepsilon|R^\varepsilon) = \int \log f^\varepsilon d\mu + \int \log g^\varepsilon d\nu.$$

We now give a brief heuristic explanation as to why one can expect this statement to imply the Caffarelli contraction theorem. Informally, from the convergence of the entropic cost to the Wasserstein distance, we expect from the dual formulation that $\varepsilon \log f$ converges to $|x|^2/2 - \varphi$ (up to some additive constant), where φ is a potential giving rise to the optimal transport map $T = \nabla\varphi$. However convexity is preserved by pointwise convergence, so we expect $|x|^2/2 - \varphi$ to also be convex. But this is equivalent to $\nabla\varphi$ being 1-Lipschitz, since the eigenvalues of the Hessian of φ must then be bounded by 1. Theorem 1.1.2 will allow us to avoid having to prove convergence of $\varepsilon \log f$ to a Kantorovich potential.

Section 1.3 is essentially devoted to the proof of Theorem 1.2.6. With Theorem 1.2.6 in hand, the proof of Theorem 1.2.4 becomes almost straightforward:

Proof of Theorem 1.2.4. Recall the following duality inequality for the relative entropy : if α, β are two probability measures on a measurable space (X, \mathcal{A}) such that $H(\alpha|\beta) < +\infty$, then for any measurable function $h : X \rightarrow \mathbb{R}$ such that $\int e^h d\beta < +\infty$, it holds $\int [h]_+ d\alpha < +\infty$ and

$$H(\alpha|\beta) \geq \int h d\alpha - \log \left(\int e^h d\beta \right) \quad (1.2.1)$$

Let $\pi \in \Pi(\mu, \eta)$ be a coupling between μ and some probability $\eta \prec_{\text{conv}} \nu$ such that $H(\pi|R^\varepsilon) < +\infty$; applying the inequality above to $\alpha = \pi$, $\beta = R^\varepsilon$ and $h(x, y) = \log (f^\varepsilon(x)g^\varepsilon(y))$, $x, y \in \mathbb{R}^d$ gives

$$\begin{aligned} H(\pi|R^\varepsilon) &\geq \int \log f^\varepsilon(x) + \log g^\varepsilon(y) \pi(dxdy) \\ &= \int \log f^\varepsilon(x) \mu(dx) + \int \log g^\varepsilon(y) \eta(dy) \\ &\geq \int \log f^\varepsilon(x) \mu(dx) + \int \log g^\varepsilon(y) \nu(dy) \\ &= H(\pi^\varepsilon|R^\varepsilon) = \mathcal{T}_H^\varepsilon(\mu, \nu), \end{aligned}$$

where the second inequality comes from the fact that $\log g^\varepsilon$ is a *concave* function and $\eta \prec_{\text{conv}} \nu$. Optimizing over π , gives the inequality $\mathcal{T}_H^\varepsilon(\mu, \eta) \geq \mathcal{T}_H^\varepsilon(\mu, \nu)$ and completes the proof. ■

To conclude this section, we mention some perspectives. The most natural question is whether this scheme of proof can be adapted to establish a version of Caffarelli's theorem in other settings than \mathbb{R}^d , such as on manifolds or in free probability [GS14], [BJ21]. Another question is about integrated or non-local quantitative regularity estimates, such as those in [Kol10, Kol13]. The role of 1-Lipschitz bounds in Theorem 1.1.2 is very specific, we do not know if there is an analogue of that equivalence adapted to other types of regularity bounds. However, it could be possible to prove stable a priori bounds on $\varepsilon \log f^\varepsilon$ and pass to the limit. Of particular interest is whether we can establish integrated gradient bounds for non-uniformly convex potentials, since such estimates can still be used to establish Poincaré inequalities [Mil09, Kla13]. Finally, [DF17] proves a rigidity/stability result for the Caffarelli contraction theorem, and it would be interesting to find a way to improve the quantitative bounds.

1.3 Proofs

This section contains the material needed to prove Theorem 1.2.4. The ideas developed here are adapted from a paper by Fortet [For40]. We warmly thank Christian Léonard for mentioning us this paper and explaining to us the ingredients of Fortet's proof. Fortet's work was recently revisited in [EP19, Léo19].

We will denote by P^ε the Ornstein-Uhlenbeck semi-group at time ε defined for all non-negative measurable function ψ by

$$P^\varepsilon \psi(x) = \mathbb{E}[\psi(Z_\varepsilon) | Z_0 = x] = \frac{1}{(2\pi)^{d/2}} \frac{1}{(1 - e^{-\varepsilon})^{d/2}} \int_{\mathbb{R}^d} \psi(y + e^{-\varepsilon/2}x) e^{-\frac{|y|^2}{2(1-e^{-\varepsilon})}} dy, \quad x \in \mathbb{R}^d.$$

Suppose that $f^\varepsilon, g^\varepsilon$ are measurable non-negative functions such that $\pi^\varepsilon(dx dy) = f^\varepsilon(x)g^\varepsilon(y)R^\varepsilon(dx dy)$ belongs to $\Pi(\mu, \nu)$. Then, writing the marginals condition, one sees that f^ε and g^ε are related to each other by the identities: for all $x, y \in \mathbb{R}^d$

$$f^\varepsilon(x)P^\varepsilon g^\varepsilon(x) = e^{V(x)} \quad \text{and} \quad g^\varepsilon(y)P^\varepsilon f^\varepsilon(y) = e^{-W(y)}. \quad (1.3.1)$$

These relations suggest to introduce the functional Φ^ε defined as follows: for all measurable function $h : \mathbb{R}^d \rightarrow \mathbb{R} \cup \{+\infty\}$,

$$\Phi^\varepsilon(h) = V - \log \left(P^\varepsilon \left(e^{-W} \frac{1}{P^\varepsilon(e^h)} \right) \right).$$

With this notation, a couple $(f^\varepsilon, g^\varepsilon)$ satisfies (1.3.1) if and only if $g^\varepsilon = e^{-W} \frac{1}{P^\varepsilon(f^\varepsilon)}$ and $f^\varepsilon = e^{h^\varepsilon}$ with h^ε such that

$$h^\varepsilon = \Phi^\varepsilon(h^\varepsilon).$$

This fixed point equation suggests that the unknown function h^ε could be obtained as the limit when $n \rightarrow +\infty$ of a sequence $(h_n)_{n \geq 0}$ satisfying the recursive scheme

$$h_{n+1} = \Phi^\varepsilon(h_n), \quad n \geq 0 \quad (1.3.2)$$

and initialized with some function h_0 . This fixed point scheme is actually at the core of the use of Sinkhorn's algorithm to numerically approximate optimal transport via entropic regularization [Cut13, BCCNP15].

The convexity of h^ε can then be established if we can initiate this fixed point scheme (1.3.2) with some convex initial data h_0 , thanks to the following key result:

Lemma 1.3.1. *If $h : \mathbb{R}^d \rightarrow \mathbb{R} \cup \{+\infty\}$ is convex, then $\Phi^\varepsilon(h)$ is also convex.*

Proof. This property is inherited from the following classical properties of P^ε :

- If f is log-convex, then $P^\varepsilon(f)$ is log-convex. This simply follows from Hölder inequality.
- If g is log-concave, then $P^\varepsilon(g)$ is log-concave. This follows from the fact that the set of log-concave functions is stable under convolution which is a well known consequence of Prekopa Theorem [Pré73].

■

The line of reasoning sketched above is essentially the one adopted in the proof of Theorem 1.2.6, except that the recurrence scheme (1.3.2) needs to be properly modified in order to force its convergence (this modification is the same as the one proposed by Fortet in [For40]).

Remark 1.3.2. In the compact setting, the map Φ^ε is actually a contraction with respect to a well-chosen metric, see for example [Gen19, Lemma 1] or [CGP16] (following the earlier ideas of [FL89] in the discrete setting). This would ensure that the fixed point must belong to any stable subspace. Here, we work in a noncompact setting (μ has non-compact support) and it seems the map is globally only 1-Lipschitz at that level of generality. One could however expect that it remains a contraction on a suitable stable subspace of convex functions.

Remark 1.3.3. A natural question is whether our scheme of proof can be used directly at the level of the Kantorovich dual formulation of optimal transport, rather than on the regularized version. The answer seems to be no, as in the limit while the minimizers in the dual formulation of entropic transport, suitably rescaled, converge to the Kantorovich potentials, the fixed point problem becomes degenerate in the limit, and only selects so-called c -convex functions (with the cost here being the quadratic distance), so we lose uniqueness. Indeed, there is no known fixed point scheme similar to (1.3.2) for Kantorovich potentials, which is why Sinkhorn's algorithm is only used to numerically approximate the regularized problem [Cut13].

Before moving on to the proof, let us present two other essential properties of Φ^ε .

Lemma 1.3.4.

1. *The map Φ^ε is monotone: $h \leq k \Rightarrow \Phi^\varepsilon(h) \leq \Phi^\varepsilon(k)$.*
2. *For any measurable $h : \mathbb{R}^d \rightarrow \mathbb{R}$, it holds*

$$\int \exp(h(x) - \Phi^\varepsilon(h)(x)) \, d\mu \leq 1,$$

with equality if h is bounded from above.

Proof. The first point is straightforward. Let us prove the second point. Since the operator P^ε is symmetric in $L^2(\gamma_d)$, for any function $h : \mathbb{R}^d \rightarrow \mathbb{R}$ it holds

$$\int e^{h \wedge a - \Phi^\varepsilon(h)} d\mu = \int e^{h \wedge a} P^\varepsilon \left(e^{-W} \frac{1}{P^\varepsilon(e^h)} \right) d\gamma_d = \int P^\varepsilon \left(e^{h \wedge a} \right) e^{-W} \frac{1}{P^\varepsilon(e^h)} d\gamma_d.$$

Letting $a \rightarrow +\infty$, one gets by monotone convergence $\int e^{h - \Phi^\varepsilon(h)} d\mu = \int_{\{P^\varepsilon(e^h) < +\infty\}} e^{-W} d\gamma_d$ which gives the claim. \blacksquare

The existence of a coupling of the desired form can be established under more general conditions on μ and ν :

Theorem 1.3.5. *Let μ be a probability measure of the form $\mu(dx) = e^{V(x)} \gamma_d(dx)$ with $V : \mathbb{R}^d \rightarrow \mathbb{R}$ convex and bounded from below, and let ν be a probability measure on \mathbb{R}^d of the form $\nu(dx) = e^{-W(x)} \gamma_d(dx)$, with $W : \mathbb{R}^d \rightarrow \mathbb{R} \cup \{+\infty\}$ a convex function such that $\{W < -m\}$ is bounded for $m = \inf_{\mathbb{R}^d} V \leq 0$. There exist a log-convex function $f^\varepsilon : \mathbb{R}^d \rightarrow [1, +\infty)$ and a log-concave function $g^\varepsilon : \mathbb{R}^d \rightarrow [0, +\infty)$ such that the measure π^ε defined by $\pi^\varepsilon(dx dy) = f^\varepsilon(x) g^\varepsilon(y) R^\varepsilon(dx dy)$ belongs to $\Pi(\mu, \nu)$.*

Proof of Theorem 1.3.5. Let us show that there exists a convex function $\bar{h} : \mathbb{R}^d \rightarrow \mathbb{R}^+$ such that $\Phi^\varepsilon(\bar{h}) = \bar{h}$. Then, defining $f^\varepsilon = e^{\bar{h}}$ and $g^\varepsilon = e^{-V}/P^\varepsilon(f^\varepsilon)$, we see that f^ε is log-convex, g^ε is log-concave (we use again the fact that P^ε preserves log-convexity) and satisfy (1.3.1).

Let us define by induction the sequence $(h_n)_{n \geq 0}$ as follows: $h_0 = 0$ and for all $n \geq 0$

$$h_{n+1} = [\Phi^\varepsilon(h_n)]_+ \wedge n. \tag{1.3.3}$$

By construction, note that h_n takes values in $[0, n - 1]$. Let us show by induction that the sequence $(h_n)_{n \geq 0}$ is non-decreasing. First observe that $h_1 = 0 = h_0$ and so in particular $h_0 \leq h_1$. According to Item (1) of Lemma 1.3.4, the operator Φ^ε is non-decreasing. Therefore, if $h_{n+1} \geq h_n$ for some $n \geq 0$, then

$$h_{n+2} = [\Phi^\varepsilon(h_{n+1})]_+ \wedge (n+1) \geq [\Phi^\varepsilon(h_n)]_+ \wedge (n+1) \geq [\Phi^\varepsilon(h_n)]_+ \wedge n = h_{n+1}.$$

Let us denote by h_∞ the pointwise limit of h_n as $n \rightarrow \infty$. The function h_∞ takes values in $\mathbb{R}^+ \cup \{+\infty\}$. Let us show that h_∞ solves the following fixed point equation

$$h_\infty = [\Phi^\varepsilon(h_\infty)]_+. \tag{1.3.4}$$

Indeed, by monotone convergence, $P^\varepsilon(e^{h_n}) \rightarrow P^\varepsilon(e^{h_\infty})$. Then, by dominated convergence,

$$P^\varepsilon \left(e^{-W} \frac{1}{P^\varepsilon(e^{h_n})} \right) \rightarrow P^\varepsilon \left(e^{-W} \frac{1}{P^\varepsilon(e^{h_\infty})} \right)$$

which implies that $h_n \rightarrow [\Phi^\varepsilon(h_\infty)]_+$ and gives (1.3.4).

Now let us show that h_∞ is in fact a fixed point of Φ^ε . Let us admit for now that $\Phi^\varepsilon(h_\infty)(x) < +\infty$ for all $x \in \mathbb{R}^d$. Then, thanks to (1.3.4), $h_\infty(x) < +\infty$ for all $x \in \mathbb{R}^d$. According to Item (2) of Lemma 1.3.4, it holds $\int e^{h_\infty - \Phi^\varepsilon(h_\infty)} d\mu \leq 1$. Since $h_\infty \geq \Phi^\varepsilon(h_\infty)$, the function $e^{h_\infty - \Phi^\varepsilon(h_\infty)}$ is bounded from below by 1. Therefore, $h_\infty = \Phi^\varepsilon(h_\infty)$ almost everywhere.

The function $\Phi^\varepsilon(h_\infty)$ is easily seen to be continuous and since h_∞ satisfies (1.3.4) it is also continuous. The functions h_∞ and $\Phi^\varepsilon(h_\infty)$ being continuous, the equality $h_\infty = \Phi^\varepsilon(h_\infty)$ holds in fact everywhere. To complete the proof that h_∞ is a fixed point of Φ^ε , it remains to prove that $\Phi^\varepsilon(h_\infty)(x) < +\infty$ for all $x \in \mathbb{R}^d$. Let us assume, by contradiction, that there exists some $x_o \in \mathbb{R}^d$ such that $\Phi^\varepsilon(h_\infty)(x_o) = +\infty$. This easily implies that $P^\varepsilon(e^{h_\infty}) = +\infty$ almost everywhere, which in turn implies that $\Phi^\varepsilon(h_\infty) \equiv \infty$. Since $h_\infty \geq \Phi^\varepsilon(h_\infty)$, one concludes also that $h_\infty \equiv +\infty$. Now let us show that there exists n_0 such that for all $n \geq n_0$

$$\inf_{x \in \mathbb{R}^d} \Phi^\varepsilon(h_n)(x) \geq 0. \quad (1.3.5)$$

For any $x \in \mathbb{R}^d$, it holds (denoting by $C = (2\pi)^{d/2}(1 - e^{-\varepsilon})^{d/2}$ and by $m = \inf_{\mathbb{R}^d} V$)

$$\begin{aligned} P^\varepsilon \left(e^{-W} \frac{1}{P^\varepsilon(e^{h_n})} \right) (x) &= \frac{1}{C} \int_{\mathbb{R}^d} e^{-W(y)} \frac{1}{P^\varepsilon(e^{h_n})}(y) e^{-\frac{|y-e^{-\varepsilon}/2x|^2}{2(1-e^{-\varepsilon})}} dy \\ &= \frac{1}{C} \int_{\{W \leq -m\}} e^{-W(y)} \frac{1}{P^\varepsilon(e^{h_n})}(y) e^{-\frac{|y-e^{-\varepsilon}/2x|^2}{2(1-e^{-\varepsilon})}} dy \\ &\quad + \frac{1}{C} \int_{\{W > -m\}} e^{-W(y)} \frac{1}{P^\varepsilon(e^{h_n})}(y) e^{-\frac{|y-e^{-\varepsilon}/2x|^2}{2(1-e^{-\varepsilon})}} dy \\ &= \frac{1}{C} \int_{\{W \leq -m\}} e^{-\frac{|y-e^{-\varepsilon}/2x|^2}{2(1-e^{-\varepsilon})}} dy \max_{z \in \{W \leq -m\}} e^{-W(z)} \frac{1}{P^\varepsilon(e^{h_n})}(z). \\ &\quad + \frac{e^m}{C} \int_{\{W > -m\}} e^{-\frac{|y-e^{-\varepsilon}/2x|^2}{2(1-e^{-\varepsilon})}} dy, \end{aligned}$$

where we used the fact that $P^\varepsilon(e^{h_n}) \geq 1$, since $h_n \geq 0$. The sequence of functions $\frac{1}{P^\varepsilon(e^{h_n})}$ is a non-increasing sequence of continuous functions converging to 0. Therefore, according to Dini's Theorem, the convergence is uniform on the compact set $K = \{W < 0\}$. Since W is convex, it is bounded from below on K . Therefore, there exists n_0 such that $\max_{z \in K} e^{-W(z)} \frac{1}{P^\varepsilon(e^{h_n})}(z) \leq e^m$ for all $n \geq n_0$. Plugging this inequality into the inequality above, one easily gets (1.3.5). Now, according to (1.3.5), there exists some n_o such that $\Phi^\varepsilon(h_{n_o}) \geq 0$. Therefore $h_{n_o+1} = \Phi^\varepsilon(h_{n_o}) \leq \Phi^\varepsilon(h_{n_o+1})$. Since h_{n_o+1} is bounded, Item (2) of Lemma 1.3.4 yields $\int e^{h_{n_o+1} - \Phi^\varepsilon(h_{n_o+1})} d\mu = 1$, which implies that $h_{n_o+1} = \Phi^\varepsilon(h_{n_o+1})$. Therefore, $h_\infty = h_{n_o+1}$, which contradicts the fact that $h_\infty \equiv +\infty$.

Finally, let $k_0 = h_\infty^{**}$ be the convex regularization of h_∞ (which is well defined since h_∞ is bounded from below). By definition $k_0 \leq h_\infty$ and since $h_\infty \geq 0$, it holds $k_0 \geq 0$. Define by induction $(k_n)_{n \geq 1}$ by $k_{n+1} = \max(\Phi^\varepsilon(k_n); k_0)$. Since according to Lemma 1.3.1 Φ^ε preserves convexity and k_0 is convex, k_n is convex for all n . The sequence k_n is non-decreasing and satisfies $k_n \leq h_\infty$ for all n . Therefore, k_n converges pointwise to some k_∞ , which is also convex and finite valued. Reasoning as above one sees that $k_\infty = \max(\Phi^\varepsilon(k_\infty); k_0)$ and so in particular $k_\infty \geq \Phi^\varepsilon(k_\infty)$. Using again the fact that $\int e^{k_\infty - \Phi^\varepsilon(k_\infty)} d\mu \leq 1$, one concludes that k_∞ is a fixed point of Φ^ε . Setting $\bar{h} = k_\infty$ completes the proof. \blacksquare

Proof of Theorem 1.2.6. First, let us note that $H(\mu \otimes \nu | R^\varepsilon) < +\infty$. Since μ and ν have finite second moment, this is easily seen to be equivalent to $H(\mu | \gamma_d) < +\infty$ and $H(\nu | \gamma_d) < +\infty$, which is true according to Lemma 1.2.3. According to Theorem 1.3.5, there exists a coupling

$\pi^\varepsilon(dx dy) = f^\varepsilon(x)g^\varepsilon(y)R^\varepsilon(dx dy) \in \Pi(\mu, \nu)$ such that f^ε is log-convex and g^ε is log-concave. It remains to show that this coupling is optimal for $\mathcal{T}_H^\varepsilon(\mu, \nu)$. Since $P^\varepsilon f^\varepsilon(y)g^\varepsilon(y) = e^{-W(y)}$, $y \in \mathbb{R}^d$, one sees that $\log g^\varepsilon(y) = -W(y) - \log P^\varepsilon f^\varepsilon(y)$. The function $\log P^\varepsilon f^\varepsilon$ is bounded continuous on the support of ν and W is integrable with respect to ν . Therefore, $\log g^\varepsilon$ is integrable with respect to ν . On the other hand, since $\log f^\varepsilon \geq 0$, the integral $\int \log f^\varepsilon d\mu$ makes sense in $[0, +\infty]$. Let $\pi \in \Pi(\mu, \nu)$ be a coupling such that $H(\pi|R^\varepsilon) < +\infty$ (this set is non empty, since it contains $\mu \otimes \nu$). Applying Inequality (1.2.1) with $\alpha = \pi$, $\beta = R^\varepsilon$ and $h(x, y) = \log f^\varepsilon(x) + \log g^\varepsilon(y)$, $x, y \in \mathbb{R}^d$, gives

$$H(\pi|R^\varepsilon) \geq \int \log f^\varepsilon(x) + \log g^\varepsilon(y) \pi(dx dy) = \int \log f^\varepsilon(x) \mu(dx) + \int \log g^\varepsilon(y) \nu(dy),$$

which shows that $\log f^\varepsilon$ is integrable with respect to μ . A simple calculation shows that

$$H(\pi^\varepsilon|R^\varepsilon) = \int \log f^\varepsilon(x) \mu(dx) + \int \log g^\varepsilon(y) \nu(dy),$$

which shows its optimality. ■

Finally let us prove the technical lemmas used in the proof of Theorem 1.1.1.

Proof of Lemma 1.2.3. Let us first show that the probability measure μ has finite entropy. Since μ has a finite second moment, it is enough to show that $H(\nu|\gamma_d) < +\infty$, which amounts to show that V is μ integrable. Since V is bounded from below by some affine function, it is clear that $[V]_-$ is μ integrable. Moreover, since the convex function V is such that $\int e^{V(x)} \gamma_d(dx) = 1$, this implies according to [GMRS17, Lemma 2.1] that $[V]_+(x) \leq \frac{|x|^2}{2}$, for all $x \in \mathbb{R}^d$, and so $[V]_+$ is also μ integrable. Similarly, to see that $H(\nu|\gamma_d) < +\infty$, it is enough to show that W is ν integrable. On the one hand, $\int [W]_+ d\nu = \int [\log(e^{-W})e^{-W}]_- d\gamma_d \leq \frac{1}{e}$. On the other hand, $[W]_-$ is ν integrable since W is bounded from below by some affine function. ■

The following Lemma was used in the proof of Theorem 1.1.1:

Lemma 1.3.6. *Let $\nu(dx) = e^{-W(x)} \gamma_d(dx)$ with $W : \mathbb{R}^d \rightarrow \mathbb{R} \cup \{+\infty\}$ convex and $\eta \prec_{conv} \nu$. Assume furthermore that ν has compact support. Define, for all $\theta \in (0, \pi/2)$,*

$$\nu_\theta = \text{Law}(\cos \theta X + \sin \theta Z) \quad \text{and} \quad \eta_\theta = \text{Law}(\cos \theta Y + \sin \theta Z),$$

where $X \sim \nu$, $Y \sim \eta$ and Z independent of X, Y and such that the law α of Z is given by $\alpha(dz) = \frac{1}{C} \mathbf{1}_B \gamma_d(dz)$ where B is the Euclidean unit ball and C a normalizing constant. Then, for all $\theta \in (0, \pi/2)$,

1. the probability ν_θ has a density of the form e^{-W_θ} with respect to γ_d , with $W_\theta : \mathbb{R}^d \rightarrow \mathbb{R} \cup \{+\infty\}$ convex,
2. the probability measures ν_θ and η_θ are compactly supported,
3. it holds $\eta_\theta \leq \nu_\theta$,
4. the probability η_θ has finite entropy.

Proof. The density f_θ of ν_θ is given by

$$f_\theta(x) = \frac{1}{C'} \int_B e^{-W\left(\frac{x-\sin\theta y}{\cos\theta}\right)} e^{-\frac{|x-\sin\theta y|^2}{2\cos^2\theta}} e^{-\frac{|y|^2}{2}} dy,$$

where C' is a normalizing constant. A simple calculation shows that

$$e^{\frac{|x|^2}{2}} f_\theta(x) = \frac{1}{C'} \int_B e^{-W\left(\frac{x-\sin\theta y}{\cos\theta}\right)} e^{-\frac{|\sin\theta x-y|^2}{2\cos^2\theta}} dy$$

and, according to Prekopa Theorem [Pr 73], the right hand side is log-concave, which completes the proof of Item (1). The proofs of Items (2) and (3) are straightforward and left to the reader. The density of η_θ is $g_\theta(x) = \frac{1}{C} \int e^{-\frac{|\cos\theta y-x|^2}{2\sin^2\theta}} \mathbf{1}_B\left(\frac{\cos\theta y-x}{\sin\theta}\right) \eta(dy)$ and so $g_\theta \leq \frac{1}{C}$. On the other hand, $g_\theta \log g_\theta \geq -1/e$. Since the support of η_θ is compact, one sees that $g_\theta \log g_\theta$ is integrable and so η_θ has finite entropy. ■

1.4 Variational proof of Theorem 1.1.2

The goal of this section is to give an alternative proof of Theorem 1.1.2. The original proof in [GJu20] uses a weak version of optimal transport as an intermediary, but we give here a new proof relying only on the variational problem solved by the Brenier map. However, we need to restrict the proof to the case where ν is absolutely continuous with respect to Lebesgue (and μ too but this is hardly a restriction when we assume that a Brenier map exists), with its support being convex. Note that for the purpose of proving the Caffarelli contraction theorem, these assumptions are not a restriction.

Let us recall some classical facts about quadratic transport that will be used in the proof (see [Vil03, Vil09] for proofs and more general statements). If $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$, the quadratic transport cost $\frac{1}{2}W_2^2$ admits the following dual formulation due to Kantorovich:

$$\frac{1}{2}W_2^2(\mu, \nu) = \sup_{\varphi, \psi} \left\{ \int \frac{|x|^2}{2} - \varphi(x) \mu(dx) + \int \frac{|y|^2}{2} - \psi(y) \nu(dy) \right\}, \quad (1.4.1)$$

where the supremum runs over couples of *convex conjugate* functions (φ, ψ) , that is to say that $\varphi, \psi : \mathbb{R}^d \rightarrow \mathbb{R} \cup \{+\infty\}$ are convex, lower semi-continuous and such that $\psi = \varphi^*$ and $\varphi = \psi^*$, where we recall that the Legendre transform h^* of a function $h : \mathbb{R}^d \rightarrow \mathbb{R} \cup \{+\infty\}$ is defined by

$$h^*(y) = \sup_{x \in \mathbb{R}^d} \{x \cdot y - h(x)\}, \quad \forall y \in \mathbb{R}^d.$$

A classical result in optimal transport tells moreover that the supremum in (1.4.1) is always attained. If (φ, ψ) is such a dual optimizer, we will say that φ (resp. ψ) is a transport potential from μ to ν (resp. ν to μ). This terminology is justified by the fact that if μ is absolutely continuous with respect to Lebesgue measure, then according to Brenier theorem, if φ is a transport potential from μ to ν , the map $\nabla\varphi$ (which is well defined μ almost surely) is (the μ almost surely unique) optimal transport map between μ and ν , *i.e.* $\nu = \nabla\varphi\#\mu$ and

$W_2^2(\mu, \nu) = \int |x - \nabla\varphi(x)|^2 \mu(dx)$. Finally, if ν is also absolutely continuous with respect to Lebesgue measure, then $\nabla\psi$ is the optimal transport map between ν and μ and it holds

$$\nabla\psi \circ \nabla\varphi(x) = x \quad \text{and} \quad \nabla\psi \circ \nabla\varphi(y) = y$$

for μ almost every x and ν almost every y .

For reader's convenience, let us reformulate Theorem 1.1.2 in a slightly different form and with the extra assumptions mentioned above :

Theorem 1.4.1. *Let $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$ be absolutely continuous with respect to Lebesgue measure and suppose that ν has a convex support. The following are equivalent:*

- (i) *There exists a transport potential $\varphi : \mathbb{R}^d \rightarrow \mathbb{R}$ from μ to ν which is continuously differentiable on \mathbb{R}^d and such that $\nabla\varphi$ is 1-Lipschitz on \mathbb{R}^d ,*
- (ii) *There exists a transport potential $\psi : \mathbb{R}^d \rightarrow \mathbb{R} \cup \{+\infty\}$ from ν to μ such that the function $\mathbb{R}^d \rightarrow \mathbb{R} \cup \{+\infty\} : x \mapsto \psi(x) - \frac{|x|^2}{2}$ is convex.*
- (iii) *For all $\eta \in \mathcal{P}_2(\mathbb{R}^d)$ such that $\eta \prec_{\text{conv}} \nu$,*

$$W_2(\nu, \mu) \leq W_2(\eta, \mu).$$

The equivalence between (i) and (ii) uses the following classical fact of convex analysis:

$$\varphi \in C^1 \text{ and } \nabla\varphi \text{ 1-Lipschitz} \iff \varphi^* - \frac{|\cdot|^2}{2} \text{ is convex,}$$

see [HL01, Theorem E 4.2.1] or [GJu20, Lemma 2.1]. It turns out that condition (ii) will be easier to handle than condition (i).

Proof of Theorem 1.4.1, (ii) \Rightarrow (iii). Assume ψ is a transport potential from ν to μ such that $\psi - \frac{|\cdot|^2}{2}$ is convex and let $\varphi = \psi^*$. For any $\eta \prec_{\text{conv}} \nu$, it holds

$$\begin{aligned} \frac{1}{2}W_2^2(\mu, \nu) &= \int \frac{|x|^2}{2} - \varphi(x) \mu(dx) + \int \frac{|y|^2}{2} - \psi(y) \nu(dy) \\ &\leq \int \frac{|x|^2}{2} - \varphi(x) \mu(dx) + \int \frac{|y|^2}{2} - \psi(y) \eta(dy) \\ &\leq \frac{1}{2}W_2^2(\mu, \eta), \end{aligned}$$

where the first equality comes from the optimality of (φ, ψ) , the second inequality from the fact that $\eta \prec_{\text{conv}} \nu$ and the concavity of $\frac{|\cdot|^2}{2} - \psi$ and the third inequality from Kantorovich duality. \blacksquare

For any probability measure $\rho \in \mathcal{P}_2(\mathbb{R}^d)$, we will denote from now on as

$$C_\rho := \{\eta \in \mathcal{P}_2(\mathbb{R}^d) : \eta \prec_{\text{conv}} \rho\}$$

the set of all probability measures which are dominated by ρ in the convex order. Note that this set is geodesically convex in $(\mathcal{P}_2(\mathbb{R}^d), W_2)$.

In order to prove the converse implication (iii) \Rightarrow (ii), we will proceed by contradiction and show that if (ii) is not true then one can construct a competitor $\eta \in C_\nu$ with a smaller Wasserstein distance to μ . For that purpose, we will use the following simple localization lemma.

Lemma 1.4.2. *Let $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$ and suppose that $T : \mathbb{R}^d \rightarrow \mathbb{R}^d$ is an optimal transport map from μ to ν . Let $A \subset \mathbb{R}^d$ be a Borel set such that $\nu(A) > 0$, define ν_A as the renormalized restriction of ν to A : $\nu_A := \frac{\nu|_A}{\nu(A)}$ and μ_A as the renormalized restriction of μ to $T^{-1}(A)$: $\mu_A := \frac{\mu|_{T^{-1}(A)}}{\mu(T^{-1}(A))}$. Suppose that there exists $\eta_A \in C_{\nu_A}$ such that $W_2(\mu_A, \eta_A) < W_2(\mu_A, \nu_A)$. Then the probability measure η defined by $\eta = \nu|_{A^c} + \nu(A)\eta_A$ is such that $\eta \in C_\nu$ and $W_2(\mu, \eta) < W_2(\mu, \nu)$.*

Proof. It is clear that $\eta <_{\text{conv}} \nu$. Let us show that η is closer to μ than ν . For the sake of simplicity, we will assume that there exists an optimal transport map S between μ_A and η_A . According to [Vil09, Theorem 4.6], the map T is still the optimal transport map from μ_A to ν_A . The map R defined by $R(x) = S(x)$ if $x \in T^{-1}(A)$ and $R(x) = T(x)$ if $x \in T^{-1}(A)^c$ is a transport map between μ and η (not necessarily optimal) and it holds:

$$\begin{aligned} W_2^2(\mu, \eta) &\leq \int |x - R(x)|^2 \mu(dx) \\ &= \mu(T^{-1}(A)) \int |x - S(x)|^2 \mu_A(dx) + \int_{T^{-1}(A)^c} |x - T(x)|^2 \mu(dx) \\ &< \mu(T^{-1}(A)) W_2^2(\mu_A, \nu_A) + \int_{T^{-1}(A)^c} |x - T(x)|^2 \mu(dx) \\ &= \mu(T^{-1}(A)) \int |x - T(x)|^2 \mu_A(dx) + \int_{T^{-1}(A)^c} |x - T(x)|^2 \mu(dx) \\ &= W_2^2(\mu, \nu). \end{aligned}$$

■

Before completing the proof of Theorem 1.4.1, let us state a lemma about strongly convex functions. Given a convex function f , we will denote $\text{dom}(f) = \{x \in \mathbb{R}^d : f(x) < +\infty\}$ the domain of f .

Lemma 1.4.3. *Let $f : \mathbb{R}^d \rightarrow \mathbb{R} \cup \{+\infty\}$ be a lower semi-continuous convex function such that $\text{dom}(f)$ has a non-empty interior. The function f is such that $f - \frac{|\cdot|^2}{2}$ is convex if and only if for all $x, y \in \text{int dom}(f)$ where f is differentiable it holds*

$$(\nabla f(x) - \nabla f(y)) \cdot (x - y) \geq |x - y|^2. \quad (1.4.2)$$

Note that if f is continuously differentiable, the conclusion of the lemma is straightforward. The proof of Lemma 1.4.3 is postponed at the end of the section.

Proof of Theorem 1.4.1, (iii) \Rightarrow (ii). Assume that (ii) does not hold, that is to say that whenever ψ is a transport potential from ν to μ then the function $\psi - \frac{|\cdot|^2}{2}$ is not convex. First we want to make sure that the convexity problem occurs on the support of ν , denoted

by $\text{Spt}(\nu)$ in what follows. Take $\tilde{\psi}$ an arbitrary transport potential from ν to μ . Then, since $\text{Spt}(\nu)$ is closed and convex, the function ψ defined by $\psi(y) = \tilde{\psi}(y)$ if $y \in \text{Spt}(\nu)$ and $\psi(y) = +\infty$ otherwise is still convex and lower semi-continuous. Defining $\varphi := \psi^*$, one easily sees that $\varphi \leq \tilde{\psi}^*$ and so (φ, ψ) is a dual optimizer. In all what follows we will deal with this special potential ψ .

Since (ii) does not hold, Lemma 1.4.3 applied with $f = \psi$ shows that there exist two points x_0, y_0 in the interior of $\text{Spt}(\nu)$ where f is differentiable and such that

$$(x_0 - y_0) \cdot (\nabla f(x_0) - \nabla f(y_0)) < |x_0 - y_0|^2. \quad (1.4.3)$$

By continuity of $r \mapsto \nu(B_r(z))$, for $z \in \text{Spt}(\nu)$, we can find two functions $r_\varepsilon \rightarrow 0$ and $s_\varepsilon \rightarrow 0$ as $\varepsilon \rightarrow 0$ such that for all $\varepsilon > 0$, $\nu(B_{r_\varepsilon}(x_0)) = \nu(B_{s_\varepsilon}(y_0))$. We then define A_ε to be the union of these two (disjoint) balls:

$$A_\varepsilon := B_{r_\varepsilon}(x_0) \cup B_{s_\varepsilon}(y_0).$$

Then ν_{A_ε} converges weakly to $\frac{1}{2}\delta_{x_0} + \frac{1}{2}\delta_{y_0}$, as $\varepsilon \rightarrow 0$. Let $b_\varepsilon = \int y \nu_{A_\varepsilon}(dy)$ be the barycenter of A_ε with respect to ν_{A_ε} . Then we have that

$$\lim_{\varepsilon \rightarrow 0} b_\varepsilon = b := \frac{x_0 + y_0}{2}.$$

In order to construct a competitor, let us collapse the mass of ν_{A_ε} towards b_ε using the displacement interpolant between ν_{A_ε} and δ_{b_ε}

$$\rho_t^\varepsilon := [(1-t)\text{Id} + tb_\varepsilon] \# \nu_{A_\varepsilon}.$$

Then it is easily seen that for all $t \in [0, 1]$, $\rho_t^\varepsilon <_{\text{conv}} \nu_{A_\varepsilon}$. Note that we must go towards b_ε , instead of b directly, in order to stay in $C_{\nu_{A_\varepsilon}}$. As $\varepsilon \rightarrow 0$, this will not make a difference.

Let us now compute $W_2(\mu_{A_\varepsilon}, \rho_t^\varepsilon)$, where μ_{A_ε} is defined as in Lemma 1.4.2, and show that for t and ε small enough it is strictly less than $W_2(\mu_{A_\varepsilon}, \nu_{A_\varepsilon})$. Note that ρ_t^ε is the image of μ_{A_ε} under the map $(1-t)\nabla\varphi + tb_\varepsilon$ which is clearly the gradient of the convex function $x \mapsto (1-t)\varphi(x) + tb_\varepsilon \cdot x$ and is thus optimal. Therefore,

$$\begin{aligned} W_2^2(\mu_{A_\varepsilon}, \rho_t^\varepsilon) &= \int |x - (1-t)\nabla\varphi(x) - tb_\varepsilon|^2 \mu_{A_\varepsilon}(dx) \\ &= W_2^2(\mu_{A_\varepsilon}, \nu_{A_\varepsilon}) + 2t \int (x - \nabla\varphi(x)) \cdot (\nabla\varphi(x) - b_\varepsilon) \mu_{A_\varepsilon}(dx) + t^2 \int |\nabla\varphi(x) - b_\varepsilon|^2 \mu_{A_\varepsilon}(dx). \end{aligned}$$

So, for any fixed $\varepsilon > 0$, the derivative at $t = 0$ is thus given by

$$\frac{d}{dt} \Big|_{t=0} W_2^2(\mu_{A_\varepsilon}, \rho_t^\varepsilon) = 2 \int (x - \nabla\varphi(x)) \cdot (\nabla\varphi(x) - b_\varepsilon) \mu_{A_\varepsilon}(dx)$$

Note that this formula is a particular case of [Vil09, Theorem 23.9] which gives the time derivative of the Wasserstein distance along general curves of probability measures. Now, our goal is to show that the quantity calculated above is negative for all ε small enough. Since $\mu_{A_\varepsilon} = \nabla\psi \# \nu_{A_\varepsilon}$ and $\nabla\varphi \circ \nabla\psi(y) = y$ for ν_{A_ε} -almost every y , one gets

$$\frac{d}{dt} \Big|_{t=0} W_2^2(\mu_{A_\varepsilon}, \rho_t^\varepsilon) = 2 \int (\nabla\psi(y) - y) \cdot (y - b_\varepsilon) \nu_{A_\varepsilon}(dy).$$

To conclude, we use the following continuity property of the subgradient : if $\psi(x) < +\infty$, then for any $\delta > 0$, there exists $r > 0$ such that if $z \in B_r(x)$ then $\partial\psi(z) \subset \partial\psi(x) + B_\delta(0)$ (see [HL93, Theorem 6.2.4]). Since $\partial\psi(x_0) = \{\nabla\psi(x_0)\}$ and $\partial\psi(y_0) = \{\nabla\psi(y_0)\}$, it follows easily that

$$\begin{aligned} \lim_{\varepsilon \rightarrow 0} \frac{d}{dt} \Big|_{t=0} W_2^2(\mu_{A_\varepsilon}, \rho_t^\varepsilon) &= (\nabla\psi(x_0) - x_0) \cdot (x_0 - b) + (\nabla\psi(y_0) - y_0) \cdot (y_0 - b) \\ &= \frac{1}{2}(\nabla\psi(x_0) - x_0) \cdot (x_0 - y_0) + \frac{1}{2}(\nabla\psi(y_0) - y_0) \cdot (y_0 - x_0) \\ &= \frac{1}{2}(\nabla\psi(x_0) - \nabla\psi(y_0)) \cdot (x_0 - y_0) - \frac{1}{2}|x_0 - y_0|^2 \stackrel{(1.4.3)}{<} 0. \end{aligned}$$

Therefore, for ε and t small enough $W_2(\mu_{A_\varepsilon}, \rho_t^\varepsilon) < W_2(\mu_{A_\varepsilon}, \nu_{A_\varepsilon})$, which according to Lemma 1.4.2 shows that there exists $\eta \in C_\nu$ such that $W_2(\mu, \eta) < W_2(\mu, \nu)$ and completes the proof. ■

Proof of Lemma 1.4.3. Let us denote by $D = \text{dom}(f)$. According to [HL93, Theorem 6.1.2], $f - \frac{|\cdot|^2}{2}$ is convex if and only if for all $x, y \in D$, it holds

$$(x - y) \cdot (u - v) \geq |x - y|^2, \quad \forall u \in \partial f(x), \forall v \in \partial f(y), \quad (1.4.4)$$

where, we recall that for any $x \in D$, we denote by $\partial f(x)$ the sub-gradient of the convex function f at point x which is defined as the set of all vectors $u \in \mathbb{R}^d$ such that $f(z) \geq f(x) + u \cdot (z - x)$, for all $z \in \mathbb{R}^d$. We recall also that when f is differentiable at x (which is true for Lebesgue almost every x in the interior of D) then $\partial f(x) = \{\nabla f(x)\}$. Therefore, if f satisfies (1.4.4) it satisfies (1.4.2).

Let us show the converse. According to [Roc70, Theorem 25.6], for any $a \in D$, it holds

$$\partial f(a) = \text{Conv}(S(a)) + R(a),$$

where $R(a)$ is the normal cone to D at a , i.e

$$R(a) = \{h \in \mathbb{R}^d : h \cdot (z - a) \leq 0, \forall z \in D\}$$

and $S(a)$ is the set of vectors u such that there exists a sequence of points $a_k \in \text{int } D$ where f is differentiable such that $a_k \rightarrow a$ and $\nabla f(a_k) \rightarrow u$ as $k \rightarrow \infty$. Let $x, y \in D$ and $u \in \partial f(x)$ and $v \in \partial f(y)$, with decomposition $u = u' + h$, $v = v' + k$ with $u' \in \text{Conv}(S(x))$, $v' \in \text{Conv}(S(y))$, $h \in R(x)$ and $k \in R(y)$. Since $(x - y) \cdot (h - k) \geq 0$, it is enough to show that $(x - y) \cdot (u' - v') \geq |x - y|^2$. By convexity, it is enough to prove that $(x - y) \cdot (u' - v') \geq |x - y|^2$ for all $u' \in S(x)$ and $v' \in S(y)$. If a_k and b_k are sequences converging to x and y respectively in such a way $\nabla f(a_k) \rightarrow u'$ and $\nabla f(b_k) \rightarrow v'$, then according to (1.4.2), it holds

$$(\nabla f(a_k) - \nabla f(b_k)) \cdot (a_k - b_k) \geq |a_k - b_k|^2$$

and letting $k \rightarrow \infty$ gives the desired inequality. ■

CHAPITRE 2

Variational approach to the regularity of optimal transport maps: general cost functions

Contents

2.1	Introduction	58
2.1.1	Main results	58
2.1.2	Strategy of the proofs	61
2.2	An L^∞ bound on the displacement	71
2.3	Almost-minimality and harmonic approximation	76
2.3.1	Almost-minimality with respect to Euclidean cost	76
2.3.2	Eulerian point of view	78
2.3.3	Harmonic Approximation	81
2.4	The ε-regularity result	83
2.4.1	One-step improvement	83
2.4.2	Proof of ε -regularity	90
2.4.3	ε -regularity for continuous densities	100
2.4.4	ε -regularity for almost-minimizers	107
2.5	Partial regularity: Proof of Corollary 3.1.4	110
2.A	Some technical lemmas	113
2.A.1	Properties of the support of couplings	113
2.A.2	Bound on \mathcal{D}_R	115
2.A.3	A property of semi-convex functions	116
2.B	The change of coordinates Lemma 3.1.5	117
2.C	Some aspects of Campanato's theory	117

2.D An L^p bound on the displacement for almost-minimizing transport maps 119

2.1 Introduction

2.1.1 Main results

Let $X, Y \subset \mathbb{R}^d$ be compact. We assume that the cost function $c : X \times Y \rightarrow \mathbb{R}$ satisfies

(C1) $c \in C^2(X \times Y)$.

(C2) For any $x \in X$, the map $Y \ni y \mapsto -\nabla_x c(x, y) \in \mathbb{R}^d$ is one-to-one.

(C3) For any $y \in Y$, the map $X \ni x \mapsto -\nabla_y c(x, y) \in \mathbb{R}^d$ is one-to-one.

(C4) $\det \nabla_{xy} c(x, y) \neq 0$ for all $(x, y) \in X \times Y$.

Let $\rho_0, \rho_1 : \mathbb{R}^d \rightarrow \mathbb{R}$ be two probability densities, with $\text{Spt } \rho_0 \subseteq X$ and $\text{Spt } \rho_1 \subseteq Y$. It is well-known that under (an even milder regularity assumption than) condition (C1), the optimal transportation problem

$$\inf_{\pi \in \Pi(\rho_0, \rho_1)} \int_{\mathbb{R}^d \times \mathbb{R}^d} c(x, y) \, d\pi, \quad (2.1.1)$$

where the infimum is taken over all couplings π between the measures $\rho_0 \, dx$ and $\rho_1 \, dy$, admits a solution π , which we call a c -optimal coupling.

For $R > 0$ we define the set

$$\#_R := (B_R \times \mathbb{R}^d) \cup (\mathbb{R}^d \times B_R), \quad (2.1.2)$$

which is quite natural in the context of optimal transportation, because it allows for a symmetric treatment of the transport problem: it is suitable to describe all the mass that gets transported out of B_R , and all the mass that is transported into B_R . For $\alpha \in (0, 1)$ we write

$$[\nabla_{xy} c]_{\alpha, R} := \sup_{(x, y) \neq (x', y') \in \#_R} \frac{|\nabla_{xy} c(x, y) - \nabla_{xy} c(x', y')|}{|(x, y) - (x', y')|^\alpha} \quad (2.1.3)$$

for the $C^{0, \alpha}$ -semi-norm of the mixed derivative $\nabla_{xy} c$ of the cost function in the cross $\#_R$, and denote by

$$[\rho]_{\alpha, R} := \sup_{x \neq x' \in B_R} \frac{|\rho(x) - \rho(x')|}{|x - x'|^\alpha}$$

the $C^{0, \alpha}$ -semi-norm of ρ in B_R .

Fixing $\rho_0(0) = \rho_1(0) = 1$, we think of the densities as non-dimensional objects. This means that $\pi(\#_R) = \int_{B_R} (\rho_0 + \rho_1) - \pi(B_R \times B_R)$ has units of $(\text{length})^d$, so that the *Euclidean*

transport energy $\int_{\#_R} \frac{1}{2} |x - y|^2 d\pi$ has dimensionality $(\text{length})^{d+2}$, and explains the normalization by $R^{-(d+2)}$ in assumption (2.1.4) and in the definition (2.1.8) of \mathbb{E}_R below, making it a non-dimensional quantity¹. Similarly, the normalization $\nabla_{xy}c(0, 0) = -\mathbb{I}$ makes the second derivatives of the cost function non-dimensional.

The main result of this chapter is the following ε -regularity result:

Theorem 2.1.1. *Assume that (C1)–(C4) hold and that $\rho_0(0) = \rho_1(0) = 1$, as well as $\nabla_{xy}c(0, 0) = -\mathbb{I}$. Assume further that 0 is in the interior of $X \times Y$.*

Let π be a c -optimal coupling from ρ_0 to ρ_1 . There exists $R_0 = R_0(c) > 0$ such that for all $R \leq R_0$ with²

$$\frac{1}{R^{d+2}} \int_{B_{4R} \times \mathbb{R}^d} |x - y|^2 d\pi + R^{2\alpha} \left([\rho_0]_{\alpha, 4R}^2 + [\rho_1]_{\alpha, 4R}^2 + [\nabla_{xy}c]_{\alpha, 4R}^2 \right) \ll_c 1, \quad (2.1.4)$$

there exists a function $T \in C^{1,\alpha}(B_R)$ such that $(B_R \times \mathbb{R}^d) \cap \text{Spt } \pi \subseteq \text{graph } T$, and the estimate

$$[\nabla T]_{\alpha, R}^2 \lesssim \frac{1}{R^{d+2+2\alpha}} \int_{B_{4R} \times \mathbb{R}^d} |x - y|^2 d\pi + [\rho_0]_{\alpha, 4R}^2 + [\rho_1]_{\alpha, 4R}^2 + [\nabla_{xy}c]_{\alpha, 4R}^2 \quad (2.1.5)$$

holds.

We stress that the implicit constant in (2.1.5) is independent of the cost c . The scale R_0 below which our ε -regularity result holds has to be such that $B_{2R_0} \subseteq X \cap Y$ and such that the qualitative L^∞/L^2 bound (Lemma 2.2.1) holds. We note that the dependence of R_0 on c and the implicit dependence on c in the smallness assumption (2.1.4) are only through the qualitative information (C1)–(C4), see Remark 2.1.7 and Lemma 2.2.1 for details. Note also that, without appealing to the well-known result that the solution of (2.1.1) is a deterministic coupling $\pi = (\text{Id} \times T)_{\#} \rho_0$, this structural property of the optimal coupling is an outcome of our iteration.

Remark 2.1.2. Under the same assumptions as in Theorem 2.1.1, in particular only asking for the one-sided energy $\frac{1}{R^{d+2}} \int_{B_{4R} \times \mathbb{R}^d} |x - y|^2 d\pi$ to be small in (2.1.4), we can also prove the existence of a function $T^* \in C^{1,\alpha}(B_R)$ such that $(\mathbb{R}^d \times B_R) \cap \text{Spt } \pi \subseteq \{(T^*(y), y) : y \in B_R\}$, with the same estimate on the semi-norm of ∇T^* . This follows from the symmetric nature of the assumptions (C1)–(C4), of the normalization conditions on the densities and the cost, and of the smallness assumption (2.1.4). We refer the reader to STEP 1 of the proof of Theorem 2.1.1 to see how (2.1.4) entails smallness of a symmetric version of the Euclidean transport energy, as defined in (2.1.8), at a smaller scale.

As in [DF15], Theorem 2.1.1 leads to a partial regularity result for a c -optimal transport map T , that is, a map such that the c -optimal coupling between ρ_0 and ρ_1 is of the form

¹Notice that we are using a different convention here than in [GHO21], since it is more natural to work with the non-dimensional energy in our context.

²An assumption of the form $f \ll 1$ means that there exists $\varepsilon > 0$, typically only depending on the dimension d and Hölder exponents, such that if $f \leq \varepsilon$, then the conclusion holds. We write \ll_Λ to indicate that ε also depends on the parameter Λ . The symbols \sim , \gtrsim and \lesssim indicate estimates that hold up to a global constant C , which typically only depends on the dimension d and Hölder exponents. For instance, $f \lesssim g$ means that there exists such a constant with $f \leq Cg$. $f \sim g$ means that $f \lesssim g$ and $g \lesssim f$.

$\pi_T := (\text{Id} \times T)_{\#} \rho_0$. We refer to the subsection I.2 for details on the structure of T , in particular its relation to a c -convex potential u . Note that by assumption **(C1)** and the boundedness of Y , the function u is semi-convex, i.e., there exists a constant C such that $u + C|x|^2$ is convex. Hence, by Alexandrov's Theorem (see, for instance, [EG15, Theorem 6.9], or [Vil09, Theorem 14.25]), u is twice differentiable at a.e. $x \in X$.

Before stating the partial regularity result, let us mention that our L^2 -based assumption on the smallness of the Euclidean energy of the forward transport is not more restrictive than the L^∞ -based assumption of closeness of the Kantorovich potential u to $\frac{1}{2}|\cdot|^2$ in [DF15, Theorems 4.3 & 5.3]. However, the assumption on u is not invariant under transformations of c and u that preserve optimality, whereas the optimal transport map T_u , and hence our assumption on the energy $R^{-d-2} \int_{B_{4R}} |x - T_u(x)|^2 \rho_0(x) dx$, are unaffected. For that reason we additionally have to fix $\nabla_{xx}c(0, 0) = 0$, and $\nabla_x c(0, 0) = 0$ in the following corollary³, and ask for $[\nabla_{xy}c]_{\alpha, 4R}$ to be small. Hence, in this result we think of the cost c as being close to $-x \cdot y$, which is not necessarily the case in Theorem 2.1.1. Indeed, in Theorem 2.1.1, we only assume that $\nabla_{xy}c$ is close to $-\mathbb{I}$, which is the case if c is close in C^2 norm to $-x \cdot y$. Now, if $\nabla_{xy}c$ is close to $-\mathbb{I}$, one may add to the cost c terms that only depend on x or y : it does not affect the mixed second derivatives and neither does it affect the optimization problem (2.1.1), as a coupling π only sees mixed terms. The added terms thus change the value of the minimum but not the optimizer.

Corollary 2.1.3. *Assume that **(C1)**–**(C4)** hold and that $\rho_0(0) = \rho_1(0) = 1$, as well as $\nabla_{xy}c(0, 0) = -\mathbb{I}$, $\nabla_{xx}c(0, 0) = 0$, and $\nabla_x c(0, 0) = 0$. Assume further that 0 is in the interior of $X \times Y$.*

Let T_u be the c -optimal transport map from ρ_0 to ρ_1 . There exists $R_0 = R_0(c) > 0$ such that for all $R \leq R_0$ with

$$\frac{1}{R^2} \|u - \frac{1}{2}|\cdot|^2\|_{C^0(B_{8R})} + R^\alpha \left([\rho_0]_{\alpha, 4R} + [\rho_1]_{\alpha, 4R} + [\nabla_{xy}c]_{\alpha, 4R} + [\nabla_{xx}c]_{\alpha, 4R} \right) \ll_c 1, \quad (2.1.6)$$

$T_u \in C^{1, \alpha}(B_R)$ with

$$[\nabla T_u]_{\alpha, R} \lesssim \frac{1}{R^{2+\alpha}} \|u - \frac{1}{2}|\cdot|^2\|_{C^0(B_{8R})} + [\rho_0]_{\alpha, 4R} + [\rho_1]_{\alpha, 4R} + [\nabla_{xy}c]_{\alpha, 4R} + [\nabla_{xy}c]_{\alpha, 4R}. \quad (2.1.7)$$

The partial regularity statement is then as follows:

Corollary 2.1.4. *Let $\rho_0, \rho_1 : \mathbb{R}^d \rightarrow \mathbb{R}$ be two probability densities with the properties that $X = \text{Spt } \rho_0$ and $Y = \text{Spt } \rho_1$ are bounded with⁴ $|\partial X| = |\partial Y| = 0$, ρ_0, ρ_1 are positive on their supports, and $\rho_0 \in C^{0, \alpha}(X)$, $\rho_1 \in C^{0, \alpha}(Y)$. Assume that $c \in C^{2, \alpha}(X \times Y)$ and that **(C2)**–**(C4)** hold. Then there exist open sets $X' \subseteq X$ and $Y' \subseteq Y$ with $|X \setminus X'| = |Y \setminus Y'| = 0$ such that the c -optimal transport map T between ρ_0 and ρ_1 is a $C^{1, \alpha}$ -diffeomorphism between X' and Y' .*

As recently pointed out in [Gol20] for the quadratic case, the variational approach is flexible enough to also obtain ε -regularity for optimal transport maps between merely continuous densities. The modifications presented in [Gol20] can be combined with our results to prove

³Assuming that $T_u(0) = 0$, the assumption $\nabla_x c(0, 0) = 0$ fixes $\nabla u(0) = -\nabla_x c(0, T_u(0)) = 0$.

⁴ $|\cdot|$ denotes the Lebesgue measure on \mathbb{R}^d .

an ε -regularity result for the class of general cost functions considered above. This will be the context of section 2.4.3.

Finally, Theorem 2.1.1 can also be applied to optimal transportation on a Riemannian manifold \mathcal{M} with cost given by the square of the Riemannian distance function d : if $\rho_0, \rho_1 \in \mathcal{C}^{0,\alpha}(\mathcal{M})$ are two probability densities, locally bounded away from zero and infinity on \mathcal{M} , then the optimal transport map $T : \mathcal{M} \rightarrow \mathcal{M}$ sending ρ_0 to ρ_1 for the cost $c = \frac{d^2}{2}$ is a $\mathcal{C}^{1,\alpha}$ -diffeomorphism outside two closed sets $\Sigma_X, \Sigma_Y \subset \mathcal{M}$ of measure zero. See [DF15, Theorem 1.4] for details.

2.1.2 Strategy of the proofs

In this section we sketch the proof of the ε -regularity Theorem 2.1.1. As in [GO20, GH021] one of the key steps is a harmonic approximation result, which can be obtained by an explicit construction and (approximate) orthogonality on an *Eulerian* level.

L^∞ bound on the displacement

A crucial ingredient to the variational approach is a local L^∞/L^2 -estimate on the level of the displacement. More precisely, given a scale R , it gives a pointwise estimate on the non-dimensionalized displacement $\frac{y-x}{R}$ in terms of the (non-dimensionalized) Euclidean transport energy

$$\mathcal{E}_R(\pi) := \frac{1}{R^{d+2}} \int_{\#_R} \frac{1}{2} |x - y|^2 d\pi, \quad (2.1.8)$$

which amounts to a squared L^2 -average of the displacement. While this looks like an inner regularity estimate in the spirit of the main result, Theorem 2.1.1, it is not. In fact, it is rather an interpolation estimate with the c -monotonicity of $\text{Spt } \pi$ providing an invisible second control next to the energy. This becomes most apparent in the simple context of [GO20, Lemma 3.1] where monotonicity morally amounts to a (one-sided) L^∞ -control of the gradient of the displacement. The interpolation character of the estimate still shines through in the fractional exponent $\frac{2}{d+2} \in (0, 1)$ on the L^2 -norm.

Following [GH021], we here allow for general measures μ and ν ; the natural local control of these data on the energy scale is given by

$$\mathcal{D}_R(\mu, \nu) := \frac{1}{R^{d+2}} W_{B_R}^2(\mu, \kappa_\mu) + \frac{(\kappa_\mu - 1)^2}{\kappa_\mu} + \frac{1}{R^{d+2}} W_{B_R}^2(\nu, \kappa_\nu) + \frac{(\kappa_\nu - 1)^2}{\kappa_\nu}, \quad (2.1.9)$$

which measures locally at scale $R > 0$ the distance from given measures μ and ν to the Lebesgue measure, where

$$\kappa_\mu = \frac{\mu(B_R)}{|B_R|} \quad \text{and} \quad W_{B_R}^2(\mu, \kappa_\mu) = W^2(\mu|_{B_R}, \kappa_\mu dx|_{B_R}) \quad (2.1.10)$$

is the quadratic Wasserstein distance between $\mu|_{B_R}$ and $\kappa_\mu dx|_{B_R}$.⁵ Notice that if $\mu = \rho_0 dx$ and $\nu = \rho_1 dy$ with α -Hölder continuous probability densities such that $\frac{1}{2} \leq \rho_j \leq 2$ on B_R ,

⁵We use the convention that $W^2(\mu, \nu) = \inf_{\pi \in \Pi(\mu, \nu)} \int \frac{1}{2} |x - y|^2 d\pi$ in this chapter.

$j = 0, 1$, then⁶

$$\mathcal{D}_R \lesssim R^{2\alpha} \left([\rho_0]_{\alpha,R}^2 + [\rho_1]_{\alpha,R}^2 \right), \quad (2.1.11)$$

see Lemma 2.A.4 in the appendix.

The new aspect compared to [GHO21, Lemma 2.9] is the general cost function c . Not surprisingly, it turns out that the result still holds provided c is close to Euclidean and that the closeness is measured in the non-dimensional \mathcal{C}^2 -norm. We stress the fact that this closeness is not required on the entire “cross” $\#_{5R}$, cf. (2.1.2), but only to the “finite cross”

$$\mathbb{B}_{5R,\Lambda R} := (B_{5R} \times B_{\Lambda R}) \cup (B_{\Lambda R} \times B_{5R}). \quad (2.1.12)$$

This is crucial, since only this smallness is guaranteed by the finiteness of the $\mathcal{C}^{2,\alpha}$ -norm, cf. (2.1.17) below. This sharpening is a consequence of the qualitative hypotheses (C1)–(C4).

Proposition 2.1.5. *Assume that the cost function c satisfies (C1)–(C4), and let $\pi \in \Pi(\mu, \nu)$ be a coupling with c -monotone support.*

For all $\Lambda < \infty$ and for all $R > 0$ such that

$$\#_{5R} \cap \text{Spt } \pi \subseteq \mathbb{B}_{5R,\Lambda R} \quad (2.1.13)$$

and for which

$$\mathcal{E}_{6R} + \mathcal{D}_{6R} \ll 1, \quad (2.1.14)$$

$$\text{and } \|\nabla_{xy} c + \mathbb{I}\|_{\mathcal{C}^0(\mathbb{B}_{5R,\Lambda R})} \ll 1, \quad (2.1.15)$$

we have that

$$(x, y) \in \#_{4R} \cap \text{Spt } \pi \quad \Rightarrow \quad |x - y| \lesssim R (\mathcal{E}_{6R} + \mathcal{D}_{6R})^{\frac{1}{d+2}}. \quad (2.1.16)$$

Remark 2.1.6. A close look at the proof of Proposition 2.1.5 actually tells us that if \mathcal{E}_{6R} is replaced in (2.1.14) by the one-sided energy, that is, if we assume

$$\frac{1}{R^{d+2}} \int_{B_{6R} \times \mathbb{R}^d} |y - x|^2 d\pi \ll 1,$$

and if (2.1.13) is replaced by the one-sided inclusion

$$(B_{5R} \times \mathbb{R}^d) \cap \text{Spt } \pi \subseteq \mathbb{B}_{5R,\Lambda R},$$

then we still get a one-sided L^∞ bound in the form of

$$(x, y) \in (B_{4R} \times \mathbb{R}^d) \cap \text{Spt } \pi \quad \Rightarrow \quad |x - y| \lesssim R \left(\frac{1}{R^{d+2}} \int_{B_{6R} \times \mathbb{R}^d} |y - x|^2 d\pi + \mathcal{D}_{6R} \right)^{\frac{1}{d+2}}.$$

This observation will be useful in the proof of Theorem 2.1.1 to relate the one-sided energy in (2.1.4) to the full energy in Proposition 2.1.16.

⁶Whenever there is no room for confusion, we will drop the dependence of \mathcal{D} on the measures μ and ν .

Note that due to assumption (2.1.13) Proposition 2.1.5 might appear rather useless: indeed, one basically has to assume a (qualitative) L^∞ bound in the sense that there is a constant $\Lambda < \infty$ such that if $x \in B_{5R}$, then $y \in B_{\Lambda R}$, in order to obtain the L^∞ bound (2.1.16). However, as we show in Lemma 2.2.1, due to the global assumptions (C1)–(C4) alone, there exists a scale $R_0 > 0$ and a constant $\Lambda_0 < \infty$ such that (2.1.13) holds. Moreover, in the Campanato iteration used to prove Theorem 2.1.1, which is based on suitable affine changes of coordinates, the qualitative L^∞ bound (2.1.13) is reproduced in each step of the iteration (with a constant Λ that after the first step can be fixed throughout the iteration, e.g. $\Lambda = 27$ works).

Remark 2.1.7. There is an apparent mismatch with respect to the domains involved in the closeness assumptions on c in Theorem 2.1.1 and Proposition 2.1.5: we assume⁷ $R^{2\alpha} [\nabla_{xy} c]_{\alpha, 6R}^2 \ll 1$ in Theorem 2.1.1 and $\|\nabla_{xy} c + \mathbb{I}\|_{C^0(\mathbb{B}_{5R, \Lambda R})} \ll 1$ in Proposition 2.1.5. We are able to relate the two assumptions due to the qualitative L^∞ bound (2.1.13): If $\nabla_{xy} c(0, 0) = -\mathbb{I}$, we have for all $\Lambda < \infty$, using the inclusion $\mathbb{B}_{5R, \Lambda R} \subseteq \#_{6R}$,

$$\|\nabla_{xy} c + \mathbb{I}\|_{C^0(\mathbb{B}_{5R, \Lambda R})} = \sup_{(x, y) \in \mathbb{B}_{5R, \Lambda R}} |\nabla_{xy} c(x, y) - \nabla_{xy} c(0, 0)| \lesssim_\Lambda R^\alpha [\nabla_{xy} c]_{\alpha, 6R}. \quad (2.1.17)$$

Thus, if $R^{2\alpha} [\nabla_{xy} c]_{\alpha, 6R}^2$ is chosen small enough, then the assumption (2.1.15) in Proposition 2.1.5 is fulfilled.

Almost-minimality with respect to Euclidean cost

One of the main new contributions of this work is showing that the concept of almost-minimality, which is well-established in the theory of minimal surfaces, can lead to important insights also in optimal transportation. The key observation is that if c is quantitatively close to Euclidean cost, then a minimizer of (2.1.1) is almost-minimizing for the quadratic cost.

One difficulty in applying the concept of almost-minimality is that we are dealing with local quantities, for which local minimality (being minimizing with respect to its own boundary condition) would be the right framework to adopt.

Lemma 2.1.8. *Let $\pi \in \Pi(\mu, \nu)$ be a c -optimal coupling between the measures μ and ν . Then for any Borel set $\Omega \subseteq \mathbb{R}^d \times \mathbb{R}^d$ the coupling $\pi_\Omega := \pi|_\Omega$ is c -optimal given its own marginals, i.e. c -optimal between the measures μ_Ω and ν_Ω defined via*

$$\mu_\Omega(A) = \pi((A \times \mathbb{R}^d) \cap \Omega), \quad \nu_\Omega(A) = \pi((\mathbb{R}^d \times A) \cap \Omega), \quad (2.1.18)$$

for any Borel measurable $A \subseteq \mathbb{R}^d$.

This lemma allows us to restrict any c -optimal coupling π to a “good” set, where particle trajectories are well-behaved in the sense that they satisfy an L^∞ bound. In particular, we have the following corollary:

Corollary 2.1.9. *Let $\pi \in \Pi(\mu, \nu)$ be a c -optimal coupling between the measures μ and ν with the property that for all $(x, y) \in \#_R \cap \text{Spt } \pi$*

$$|x - y| \leq R.$$

⁷The assumption of Theorem 2.1.1 is scaled to $6R$ here to match the scale on which smallness of \mathcal{E} and \mathcal{D} is assumed in both statements.

Then the coupling $\pi_R := \pi|_{\#_R}$ is c -optimal between the measures μ_R and ν_R as defined in (2.1.18) and we have that $\text{Spt } \mu_R, \text{Spt } \nu_R \subseteq B_{2R}$ (in particular $\text{Spt } \pi_R \subseteq B_{2R} \times B_{2R}$), $\mu_R = \mu$ and $\nu_R = \nu$ on B_R , and $\mu_R \leq \mu$, $\nu_R \leq \nu$.

One of the main observations now is that c -optimal couplings of the type considered in Corollary 2.1.9 are almost-minimizers of the Euclidean transport cost. The following assumptions (2.1.19) and (2.1.20) should be read as properties satisfied by the marginal measures μ and ν of the restriction of a c -optimal coupling to a finite cross on which the L^∞ bound (2.1.16) holds. Moreover, one of the marginals should be close to the Lebesgue measure in the sense that $\mu(B_R) \lesssim R^d$.

Proposition 2.1.10. *Let μ and ν be two measures such that*

$$\text{Spt } \mu \subseteq B_R, \text{Spt } \nu \subseteq B_R, \quad (2.1.19)$$

$$\mu(B_R) \leq 2|B_R|, \quad (2.1.20)$$

for some $R > 0$. Let $\pi \in \Pi(\mu, \nu)$ be a c -optimal coupling between the measures μ and ν . Then π is almost-minimizing for the Euclidean cost, in the sense that for any $\tilde{\pi} \in \Pi(\mu, \nu)$ we have that

$$\int \frac{1}{2}|x - y|^2 d\pi \leq \int \frac{1}{2}|x - y|^2 d\tilde{\pi} + R^{d+2}\Delta_R, \quad (2.1.21)$$

where

$$\Delta_R := C\|\nabla_{xy}c + \mathbb{I}\|_{C^0(B_R \times B_R)} \mathcal{E}_R(\pi)^{\frac{1}{2}} \quad (2.1.22)$$

for some constant C depending only on d .

The above statement is most naturally formulated in terms of couplings, that is, in the Kantorovich framework. However, the way (almost-)minimality enters in the proof of the harmonic approximation result (see Theorem 2.1.13 below), it is needed in the Eulerian picture, where the construction of a competitor is done.

The Eulerian side of optimal transportation

As already presented in (I.4) and onwards, given a coupling $\pi \in \Pi(\mu, \nu)$ between measures μ and ν , we can define its Eulerian description, i.e. the density-flux pair (ρ_t, j_t) associated with the coupling π by

$$\int \zeta d\rho_t := \int \zeta((1-t)x + ty) d\pi, \quad \int \xi \cdot dj_t := \int \xi((1-t)x + ty) \cdot (y - x) d\pi \quad (2.1.23)$$

for $t \in [0, 1]$ and for all test functions $\zeta \in C_c^0(\mathbb{R}^d \times [0, 1])$ and fields $\xi \in C_c^0(\mathbb{R}^d \times [0, 1])^d$. It is easy to check that (ρ_t, j_t) is a distributional solution of the continuity equation

$$\partial_t \rho_t + \nabla \cdot j_t = 0, \quad \rho_0 = \mu, \quad \rho_1 = \nu, \quad (2.1.24)$$

that is, for any $\zeta \in C_c^1(\mathbb{R}^d \times [0, 1])$ there holds

$$\int_0^1 \left(\int \partial_t \zeta \, d\rho_t + \nabla \zeta \cdot dj_t \right) dt = \int \zeta_1 \, d\nu - \int \zeta_0 \, d\mu. \quad (2.1.25)$$

For brevity, we will often write $(\rho, j) := (\rho_t dt, j_t dt)$. Being divergence-free in (t, x) , the density-flux pair (ρ, j) admits internal (and external) traces on $\partial(B_R \times (0, 1))$ for any $R > 0$, see [CF03] for details, i.e., there exists a measure f_R on $\partial B_R \times (0, 1)$ such that

$$\int_{B_R \times [0, 1]} (\partial_t \zeta \, d\rho + \nabla \zeta \cdot dj) = \int_{B_R} \zeta_1 \, d\nu - \int_{B_R} \zeta_0 \, d\mu + \int_{\partial B_R \times [0, 1]} \zeta \, df_R, \quad (2.1.26)$$

We also introduce the time-averaged measure \bar{f}_R on ∂B_R defined via

$$\int_{\partial B_R} \zeta \, d\bar{f}_R := \int_{\partial B_R \times [0, 1]} \zeta \, df_R. \quad (2.1.27)$$

Similarly, defining the measure $\bar{j} := \int_0^1 dj(\cdot, t)$, it is easy to see that

$$\nabla \cdot \bar{j} = \mu - \nu$$

and that therefore \bar{j} admits internal and external traces, which agree for all $R > 0$ with $|\bar{j}|(\partial B_R) = \mu(\partial B_R) = \nu(\partial B_R) = 0$, and the internal trace agrees with \bar{f}_R .

Note that we have the duality [San15, Proposition 5.18]

$$\frac{1}{2} \int \frac{1}{\rho} |j|^2 = \sup_{\xi \in C_c^0(\mathbb{R}^d \times [0, 1])^d} \int \xi \cdot dj - \frac{|\xi|^2}{2} \, d\rho, \quad (2.1.28)$$

which immediately implies the subadditivity of $(\rho, j) \mapsto \int \frac{1}{\rho} |j|^2$. A localized version of (2.1.28), in the form of

$$\frac{1}{2} \int_{B \times [0, 1]} \frac{1}{\rho} |j|^2 = \sup_{\xi \in C_c^0(B \times [0, 1])^d} \int \xi \cdot dj - \frac{|\xi|^2}{2} \, d\rho, \quad (2.1.29)$$

also holds for any open set $B \subseteq \mathbb{R}^d$. From the inequality $\xi \cdot (y - x) - \frac{1}{2} |\xi|^2 \leq \frac{1}{2} |x - y|^2$, which is true for any $\xi, x, y \in \mathbb{R}^d$, the duality formula (2.1.28) immediately implies that the Eulerian cost of the density-flux pair (ρ, j) corresponding to a coupling π via (2.1.23) is always dominated by the Lagrangian cost of π , i.e.

$$\frac{1}{2} \int \frac{1}{\rho} |j|^2 \leq \frac{1}{2} \int |x - y|^2 \, d\pi. \quad (2.1.30)$$

We stress that this inequality is in general strict.

Contrary to the case of quadratic cost $c(x, y) = \frac{1}{2} |x - y|^2$, or, equivalently, $\tilde{c}(x, y) = -x \cdot y$, given an optimal coupling π for the cost c , the density-flux pair (ρ, j) associated with π in the sense of (2.1.23) is not optimal for the Benamou–Brenier formulation [BB99] of optimal transportation, i.e.,

$$W^2(\mu, \nu) = \inf \left\{ \frac{1}{2} \int \frac{1}{\rho} |j|^2 : \partial_t \rho + \nabla \cdot j = 0, \rho_0 = \mu, \rho_1 = \nu \right\}, \quad (2.1.31)$$

where the continuity equation and boundary conditions are understood in the weak sense (2.1.25), see [Vil03, Chapter 8] for details.

Remark 2.1.11. For minimizers π of the Euclidean transport cost, i.e. $W^2(\mu, \nu) = \frac{1}{2} \int |y - x|^2 d\pi$, we have *equality* in (2.1.30),

$$\frac{1}{2} \int \frac{1}{\rho} |j|^2 = \frac{1}{2} \int |x - y|^2 d\pi. \quad (2.1.32)$$

Indeed, if (ρ, j) is the Eulerian description of π , then by (2.1.31)

$$\frac{1}{2} \int |x - y|^2 d\pi = W^2(\mu, \nu) \leq \frac{1}{2} \int \frac{1}{\rho} |j|^2,$$

which together with (2.1.30) implies (2.1.32).

As another consequence, while displacement convexity guarantees in the Euclidean case that the Eulerian density $\rho \leq 1$ (up to a small error), c.f. [GO20, Lemma 4.2], in our case ρ is in general merely a measure. This complication is already present in [GHO21] and led to important new insights in dealing with marginals that are not absolutely continuous with respect to Lebesgue measure in the Euclidean case, upon which we are also building in this work.

The Eulerian version of the almost-minimality Proposition 2.1.10 can then be obtained via the following lemma:

Lemma 2.1.12. *Let $\pi \in \Pi(\mu, \nu)$ be a coupling between the measures μ and ν with the property that there exists a constant $\Delta < \infty$ such that*

$$\int \frac{1}{2} |x - y|^2 d\pi \leq \int \frac{1}{2} |x - y|^2 d\tilde{\pi} + \Delta \quad (2.1.33)$$

for any $\tilde{\pi} \in \Pi(\mu, \nu)$, and let (ρ, j) be its Eulerian description defined in (2.1.23). Then

$$\int \frac{1}{2} |x - y|^2 d\pi \leq \frac{1}{2} \int \frac{1}{\rho} |j|^2 + \Delta,$$

and

$$\frac{1}{2} \int \frac{1}{\rho} |j|^2 \leq \frac{1}{2} \int \frac{1}{\tilde{\rho}} |\tilde{j}|^2 + \Delta$$

for any pair of measures $(\tilde{\rho}, \tilde{j})$ satisfying

$$\int \left(\partial_t \zeta d\tilde{\rho} + \nabla \zeta \cdot d\tilde{j} \right) = \int \zeta_1 d\nu - \int \zeta_0 d\mu \quad (2.1.34)$$

for all $\zeta \in C_c^\infty(\mathbb{R}^d \times [0, 1])$.

The harmonic approximation result

The main ingredient in the proof of Theorem 2.1.1 is the harmonic approximation result, which states that if a coupling between two measures supported on a ball (say of radius $7R$ for some $R > 0$) satisfies the L^∞ bound Proposition 2.1.5 globally on its support and is almost-minimizing with respect to the Euclidean cost, then the displacement $y - x$ is quantitatively close to a harmonic gradient field $\nabla \Phi$ in $\# \#_R$. This is actually a combination of a harmonic approximation result in the Eulerian picture (Theorem 2.1.13) and Lemma 2.1.14, which allows us to transfer the Eulerian information back to the Lagrangian framework.

Theorem 2.1.13 (Harmonic approximation). *Let $R > 0$ and μ, ν be two measures with the property that*

$$\text{Spt } \mu \subseteq B_{7R}, \quad \text{Spt } \nu \subseteq B_{7R}, \quad \text{and} \quad \mu(B_{7R}) \leq 2|B_{7R}|. \quad (2.1.35)$$

Let further $\pi \in \Pi(\mu, \nu)$ be a coupling between the measures μ and ν , such that:

1. π satisfies a global L^∞ -bound, that is, there exists a constant $M \leq 1$ such that

$$|x - y| \leq MR \quad \text{for any } (x, y) \in \text{Spt } \pi. \quad (2.1.36)$$

2. If (ρ, j) is the Eulerian description of π as defined in (2.1.23), then there exists a constant $\Delta_R < \infty$ such that

$$\int \frac{1}{\rho} |j|^2 \leq \int \frac{1}{\bar{\rho}} |\bar{j}|^2 + R^{d+2} \Delta_R \quad (2.1.37)$$

for any Eulerian competitor, i.e. any pair of measures $(\bar{\rho}, \bar{j})$ satisfying

$$\int \partial_t \zeta \, d\bar{\rho} + \nabla \zeta \cdot d\bar{j} = \int_{B_R} \zeta_1 \, d\nu - \int_{B_R} \zeta_0 \, d\mu + \int_{\partial B_R \times [0,1]} \zeta \, df_R.$$

Then for every $0 < \tau \ll 1$, there exist $\varepsilon_\tau > 0$ and $C, C_\tau < \infty$ such that, provided⁸

$$\mathcal{E}_{6R}(\pi) + \mathcal{D}_{6R}(\mu, \nu) \leq \varepsilon_\tau, \quad (2.1.38)$$

the following holds: There exists a radius $R_ \in (3R, 4R)$ such that if Φ is the solution, unique up to an additive constant, of⁹*

$$\Delta \Phi = \kappa_\mu - \kappa_\nu \quad \text{in } B_{R_*} \quad \text{and} \quad \nu \cdot \nabla \Phi = \nu \cdot \bar{j} = \bar{f}_{R_*} \quad \text{on } \partial B_{R_*}, \quad (2.1.39)$$

then¹⁰

$$\frac{1}{R^{d+2}} \int_{B_{2R} \times [0,1]} \frac{1}{\rho} |j - \rho \nabla \Phi|^2 \leq \left(\tau + \frac{CM}{\tau} \right) \mathcal{E}_{6R} + C_\tau \mathcal{D}_{6R} + \Delta_R, \quad (2.1.40)$$

and

$$\frac{1}{R^2} \sup_{B_{2R}} |\nabla \Phi|^2 + \sup_{B_{2R}} |\nabla^2 \Phi|^2 + R^2 \sup_{B_{2R}} |\nabla^3 \Phi|^2 \lesssim \mathcal{E}_{6R} + \mathcal{D}_{6R}. \quad (2.1.41)$$

⁸Note that by assumption (2.1.35) the coupling π is supported on $B_{7R} \times B_{7R}$, so that by means of the estimate $\mathcal{E}_{6R}(\pi) \lesssim \mathcal{E}_{7R}(\pi) = \int |x - y|^2 \, d\pi$ the smallness assumption of the Euclidean energy of π on scale $6R$ could be replaced by a smallness assumption on the global Euclidean energy of π . However, since \mathcal{D} behaves nicely under restriction only on average, the applicability of the harmonic approximation result derived in [GHO21] becomes more apparent in the form of assumption (2.1.38).

⁹We recall that since the boundary flux \bar{f}_{R_*} (as defined in (2.1.27)) is a measure, equation (2.1.39) has to be understood in the distributional sense, and that $\kappa_\mu = \frac{\mu(B_{R_*})}{|B_{R_*}|}$ and $\kappa_\nu = \frac{\nu(B_{R_*})}{|B_{R_*}|}$.

¹⁰We refer to (2.3.16) for how to understand the left-hand side of (2.1.40).

From the Eulerian version of the harmonic approximation Theorem 2.1.13 we can also obtain a Lagrangian version via almost-minimality:

Lemma 2.1.14. *Let $R > 0$ and let $\pi \in \Pi(\mu, \nu)$ be a coupling between the measures μ and ν , such that*

1. π satisfies a global L^∞ -bound, that is, there exists a constant $M \leq 1$ such that

$$|x - y| \leq MR \quad \text{for any } (x, y) \in \text{Spt } \pi; \quad (2.1.42)$$

2. if (ρ, j) is the Eulerian description of π as defined in (2.1.23), then there exists a constant $\Delta_R < \infty$ such that

$$\int |x - y|^2 d\pi \leq \int \frac{1}{\rho} |j|^2 + R^{d+2} \Delta_R. \quad (2.1.43)$$

Then for any smooth function Φ there holds

$$\frac{1}{R^{d+2}} \int_{\#_R} \int_0^1 |x - y + \nabla \Phi(tx + (1-t)y)|^2 dt d\pi \leq \frac{1}{R^{d+2}} \int_{B_{2R} \times [0,1]} \frac{1}{\rho} |j - \rho \nabla \Phi|^2 + \Delta_R. \quad (2.1.44)$$

One-step improvement and Campanato iteration

With the harmonic approximation result at hand, we can derive a one-step improvement result, which roughly says that if the coupling π is quantitatively close to $(\text{Id} \times \text{Id})_{\#} \rho_0$ on some scale R , expressed in terms of the estimate

$$\mathcal{E}_R(\pi) + R^{2\alpha} \left([\rho_0]_{\alpha, R}^2 + [\rho_1]_{\alpha, R}^2 + [\nabla_{xy} c]_{\alpha, R}^2 \right) \ll 1,$$

and the fact that the (qualitative) L^∞ bound on the displacement (2.1.13) holds, then on a smaller scale θR , after an affine change of coordinates, it is even closer to $(\text{Id} \times \text{Id})_{\#} \rho_0$. This is the basis of a Campanato iteration to obtain the existence of the optimal transport map T and its $C^{1,\alpha}$ regularity.

We start with the affine change of coordinates and its properties:

Lemma 2.1.15. *Let $\pi \in \Pi(\mu, \nu)$ be an optimal transport plan with respect to the cost function c between the measures $\mu(dx) = \rho_0(x) dx$ and $\nu(dy) = \rho_1(y) dy$.*

Given a non-singular matrix $B \in \mathbb{R}^{d \times d}$ and a vector $b \in \mathbb{R}^d$, we perform the affine change of coordinates¹¹

$$\begin{pmatrix} \widehat{x} \\ \widehat{y} \end{pmatrix} = \begin{pmatrix} Bx \\ \gamma B^{-*} D^*(y - b) \end{pmatrix} =: Q(x, y), \quad (2.1.45)$$

¹¹We use the notation $B^{-*} = (B^*)^{-1}$, where B^* is the transpose of B , and $\nabla^2 c(x, y) = \begin{pmatrix} \nabla_{xx} c(x, y) & \nabla_{yx} c(x, y) \\ \nabla_{xy} c(x, y) & \nabla_{yy} c(x, y) \end{pmatrix}$.

where $D = -\nabla_{xy}c(0, b)$ and $\gamma = \left(\frac{\rho_1(b)}{\rho_0(0)} \frac{|\det B|^2}{|\det D|}\right)^{\frac{1}{d}}$.¹² If we let

$$\widehat{\rho}_0(\widehat{x}) = \frac{\rho_0(x)}{\rho_0(0)}, \quad \widehat{\rho}_1(\widehat{y}) = \frac{\rho_1(y)}{\rho_1(b)}, \quad \widehat{c}(\widehat{x}, \widehat{y}) = \gamma c(x, y),$$

so that in particular $\widehat{\rho}_0(0) = \widehat{\rho}_1(0) = 1$ and

$$\nabla_{\widehat{x}\widehat{y}}\widehat{c}(\widehat{x}, \widehat{y}) = BD^{-1}\nabla_{xy}c(x, y)B^{-1}, \quad (2.1.46)$$

from which it follows that $\nabla_{\widehat{x}\widehat{y}}\widehat{c}(0, 0) = -\mathbb{I}$, then the coupling

$$\widehat{\pi} := \frac{|\det B|}{\rho_0(0)} Q_{\#}\pi \quad (2.1.47)$$

is an optimal coupling between the measures $\widehat{\mu}(d\widehat{x}) = \widehat{\rho}_0(\widehat{x}) d\widehat{x}$ and $\widehat{\nu}(d\widehat{y}) = \widehat{\rho}_1(\widehat{y}) d\widehat{y}$ with respect to the cost function \widehat{c} .

In the change of variables we perform, the role of D is to ensure that we get a normalized cost, i.e. $\nabla_{\widehat{x}\widehat{y}}\widehat{c}(0, 0) = -\mathbb{I}$, while γ and $\det B$ in (2.1.47) are needed for $\widehat{\pi}$ to define a transportation plan between the new densities. We refer the reader to Appendix 2.B for a proof of this lemma.

Proposition 2.1.16. *Assume that $\rho_0(0) = \rho_1(0) = 1$ and $\nabla_{xy}c(0, 0) = -\mathbb{I}$, and let $\pi \in \Pi(\rho_0, \rho_1)$ be c -optimal.*

Then for all $\beta \in (0, 1)$, there exist $\theta \in (0, 1)$ and $C_\beta < \infty$ such that for all $\Lambda < \infty$ and $R > 0$ for which

$$\#_{8R} \cap \text{Spt } \pi \subseteq \mathbb{B}_{8R, \Lambda R}, \quad (2.1.48)$$

$$\mathcal{E}_{9R}(\pi) + R^{2\alpha} \left([\rho_0]_{\alpha, 9R}^2 + [\rho_1]_{\alpha, 9R}^2 + [\nabla_{xy}c]_{\alpha, 9R}^2 \right) \ll_{\Lambda} 1, \quad (2.1.49)$$

there exist a symmetric matrix $B \in \mathbb{R}^{d \times d}$ and a vector $b \in \mathbb{R}^d$ with

$$|B - \mathbb{I}|^2 + \frac{1}{R^2}|b|^2 \lesssim \mathcal{E}_{9R}(\pi) + R^{2\alpha} \left([\rho_0]_{\alpha, 9R}^2 + [\rho_1]_{\alpha, 9R}^2 \right), \quad (2.1.50)$$

such that, performing the change of variables in Lemma 2.1.15, the coupling $\widehat{\pi}$ is \widehat{c} -optimal between the measures with densities $\widehat{\rho}_0$ and $\widehat{\rho}_1$ and there holds

$$\mathcal{E}_{\theta R}(\widehat{\pi}) \leq \theta^{2\beta} \mathcal{E}_{9R}(\pi) + C_\beta R^{2\alpha} \left([\rho_0]_{\alpha, 9R}^2 + [\rho_1]_{\alpha, 9R}^2 + [\nabla_{xy}c]_{\alpha, 9R}^2 \right). \quad (2.1.51)$$

Moreover, we have the inclusion

$$\#_{\frac{8}{\theta}R} \cap \text{Spt } \widehat{\pi} \subseteq \mathbb{B}_{\frac{8}{\theta}R, 3\theta R}. \quad (2.1.52)$$

¹²Note that D is non-singular by assumption (C4).

Let us give a rough sketch of how the one-step improvement result can now be iterated: In a first step, the qualitative bound on the displacement is obtained from the global assumptions (C1)–(C4) on the cost function, see Lemma 2.2.1. This yields an initial scale $R_0 > 0$ below which the cost function is close enough to the Euclidean cost function for (2.1.13) to hold. We may therefore apply Proposition 2.1.16, so that after an affine change of coordinates the energy inequality (2.1.51) holds, the transformed densities and cost function are again normalized at the origin, optimality is preserved, and the qualitative L^∞ bound (2.1.52) holds for the new coupling. We can therefore apply the one-step improvement Proposition 2.1.16 again, going to smaller and smaller scales. Together with Campanato’s characterization of Hölder regularity, this yields the claimed existence and $C^{1,\alpha}$ regularity of T .

The details of the above parts of the proof of our main Theorem 2.1.1 are explained in the sections below, with a full proof of Theorem 2.1.1 in Section 2.4.2. The proof of Corollary 2.1.4 is essentially a combination of the ideas in [GO20] and [DF15], and is given for the convenience of the reader in Section 2.5.

We conclude the introduction with a comment on the extension of the results presented above to general almost-minimizers with respect to Euclidean cost in the following sense:

Definition 2.1.17 (Almost-minimality w.r.t. Euclidean cost (on all scales)). A coupling $\pi \in \Pi(\mu, \nu)$ is almost-minimal with respect to Euclidean cost if there exists $R_0 > 0$ and $\Delta_r : (0, R_0] \rightarrow [0, \infty)$ non-decreasing such that for all $r \leq R_0$ and (x_0, y_0) in the interior of $X \times Y$ there holds

$$\int |y - x|^2 d\pi \leq \int |y - x|^2 d\tilde{\pi} + r^{d+2}\Delta_r \quad (2.1.53)$$

for all $\tilde{\pi} \in \Pi(\mu, \nu)$ such that $\text{Spt}(\pi - \tilde{\pi}) \subseteq (B_r(x_0) \times \mathbb{R}^d) \cup (\mathbb{R}^d \times B_r(y_0))$.

We will restrict our attention to almost-minimizers in the class of deterministic transport plans coming from a Monge map T , i.e. $\pi = \pi_T = (\text{Id}, T)_\# \mu$, and call a transport map almost-minimizing with respect to Euclidean cost if for all $r \leq R_0$ and $x_0 \in \text{int}X$ there holds

$$\int |T(x) - x|^2 \mu(dx) \leq \int |\tilde{T}(x) - x|^2 \mu(dx) + r^{d+2}\Delta_r \quad (2.1.54)$$

for all \tilde{T} such that $\tilde{T}_\# \mu = \nu$ and $\text{graph}\tilde{T} = \text{graph}T$ outside $(B_r(x_0) \times \mathbb{R}^d) \cup (\mathbb{R}^d \times B_r(T(x_0)))$.

In this situation, we get the following generalization¹³ of Theorem 2.1.1, whose proof will be sketched in Section 2.4.4.

Theorem 2.1.18. *Assume that $\rho_0(0) = \rho_1(0) = 1$, and that 0 is in the interior of $X \times Y$. Let $T : X \rightarrow Y$ be an almost-minimizing transport map from μ to ν with rate function $\Delta_r = Cr^{2\alpha}$ for some $C < \infty$. Assume further that T is invertible. There exists $R_1 > 0$ such that for any $R \leq R_1$ with*

$$\frac{1}{R^{d+2}} \left(\int_{B_{4R}} |T(x) - x|^2 \mu(dx) + \int_{B_{4R}} |T^{-1}(y) - y|^2 \nu(dy) \right) + R^{2\alpha} \left([\rho_0]_{\alpha, 4R}^2 + [\rho_1]_{\alpha, 4R}^2 \right) \ll 1,$$

there holds $T \in C^{1,\alpha}(B_R)$.

¹³A more quantitative version of this result is work in progress.

2.2 An L^∞ bound on the displacement

In this section we establish an L^∞ bound on the displacement for transference plans $\pi \in \Pi(\mu, \nu)$ with c -monotone support, that is,

$$c(x, y) + c(x', y') \leq c(x, y') + c(x', y) \quad \text{for all } (x, y), (x', y') \in \text{Spt } \pi, \quad (2.2.1)$$

provided that the transport cost is small, the marginals μ, ν are close to the Lebesgue measure, and the cost function c is close to the Euclidean cost function. In Lemma 2.2.1 we use the c -monotonicity (2.2.1) combined with the qualitative hypotheses (C1)–(C4) in conjunction with compactness to obtain a more qualitative version of the L^∞/L^2 -bound, which just expresses finite expansion. In Proposition 2.1.5 this qualitative L^∞/L^2 bound in form of (2.1.13) is upgraded to the desired quantitative version under the scale-invariant smallness assumption (2.1.15). The latter is a consequence of the quantitative smallness hypothesis $R^{2\alpha} [\nabla_{xy} c]_{\alpha, R}^2 \ll 1$, as we pointed out in Remark 2.1.7. In both steps, we need to ensure that there are sufficiently many points in $\text{Spt } \pi$ close to the diagonal. This is formulated in Lemma 2.A.1, which does not rely on monotonicity.

Proof of Proposition 2.1.5. Let $\Lambda < \infty$ and $R > 0$ be such that (2.1.13), (2.1.14) and (2.1.15) hold. We only prove the bound (2.1.16) for a couple $(x, y) \in (B_{4R} \times \mathbb{R}^d) \cap \text{Spt } \pi$, as the other case $(x, y) \in (\mathbb{R}^d \times B_{4R}) \cap \text{Spt } \pi$ follows by symmetry.

STEP 1 (Rescaling). Let $(\tilde{x}, \tilde{y}) = (S_R(x), S_R(y)) := (R^{-1}x, R^{-1}y)$ and set

$$\tilde{\mu} := S_{R\#}\mu, \quad \tilde{\nu} := S_{R\#}\nu, \quad \tilde{\pi} := R^{-d}(S_R \times S_R)_{\#}\pi \quad \text{and} \quad \tilde{c}(\tilde{x}, \tilde{y}) := R^{-2}c(R\tilde{x}, R\tilde{y}),$$

so that \tilde{c} still satisfies properties (C1)–(C4), and we have $\tilde{\pi} \in \Pi(\tilde{\mu}, \tilde{\nu})$, and $\text{Spt } \tilde{\pi}$ is \tilde{c} -monotone. We also have

$$\mathcal{E}_6(\tilde{\pi}) = \mathcal{E}_{6R}(\pi), \quad \mathcal{D}_6(\tilde{\mu}, \tilde{\nu}) = \mathcal{D}_{6R}(\mu, \nu) \quad \text{and} \quad \|\nabla_{\tilde{x}\tilde{y}} \tilde{c} + \mathbb{I}\|_{C^0(\mathbb{B}_{5, \Lambda})} = \|\nabla_{xy} c + \mathbb{I}\|_{C^0(\mathbb{B}_{5R, \Lambda R})}.$$

This allows us to only consider the case $R = 1$ in the following. We will abbreviate

$$\mathcal{E} := \mathcal{E}_6 \quad \text{and} \quad \mathcal{D} := \mathcal{D}_6.$$

STEP 2 (Use of c -monotonicity of $\text{Spt } \pi$). Let $(x, y) \in (B_4 \times \mathbb{R}^d) \cap \text{Spt } \pi$. We first show that for all $(x', y') \in (B_5 \times \mathbb{R}^d) \cap \text{Spt } \pi$ we have

$$(x - y) \cdot (x - x') \leq 3|x - x'|^2 + |x' - y'|^2 + \delta|x - x'| |x - y|, \quad (2.2.2)$$

where, recalling (2.1.12),

$$\delta := \|\nabla_{xy} c + \mathbb{I}\|_{C^0(\mathbb{B}_{5, \Lambda})}. \quad (2.2.3)$$

Indeed, setting $x_t := tx + (1-t)x'$ and $y_s := sy + (1-s)y'$ for $t, s \in [0, 1]$, c -monotonicity (2.2.1) of $\text{Spt } \pi$ implies that

$$\begin{aligned} 0 &\geq (c(x, y) - c(x', y)) - (c(x, y') - c(x', y')) = \int_0^1 \int_0^1 (x - x') \cdot \nabla_{xy} c(x_t, y_s) (y - y') \, ds dt \\ &= -(x - x') \cdot (y - y') + \int_0^1 \int_0^1 (x - x') \cdot (\nabla_{xy} c(x_t, y_s) + \mathbb{I}) (y - y') \, ds dt. \end{aligned} \quad (2.2.4)$$

By assumption (2.1.13), we have $(x, y), (x', y') \in B_5 \times B_\Lambda$ and thus $\{(x_t, y_t)\}_{t \in [0,1]} \subseteq B_5 \times B_\Lambda \subset \mathbb{B}_{5,\Lambda}$. Hence, we obtain from (2.2.3)

$$\left| \int_0^1 \int_0^1 (x - x') \cdot (\nabla_{xy} c(x_t, y_s) + \mathbb{I}) (y - y') \, ds dt \right| \leq \delta |x - x'| |y - y'|,$$

so that (2.2.4) turns into

$$0 \geq -(x - x') \cdot (y - y') - \delta |x - x'| |y - y'|. \quad (2.2.5)$$

Upon writing $y - y' = (y - x) + (x - x') + (x' - y')$, it follows that

$$\begin{aligned} 0 \geq & -(x - x') \cdot (y - x) - |x - x'|^2 - (x - x') \cdot (x' - y') - \delta |x - x'| |y - x| \\ & - \delta |x - x'|^2 - \delta |x - x'| |x' - y'|, \end{aligned}$$

from which we obtain the estimate

$$\begin{aligned} (x - y) \cdot (x - x') & \leq \delta |x - x'| |x - y| + (1 + \delta) |x - x'|^2 + (1 + \delta) |x - x'| |x' - y'| \\ & \leq \delta |x - x'| |x - y| + \frac{3}{2} (1 + \delta) |x - x'|^2 + \frac{1}{2} (1 + \delta) |x' - y'|^2. \end{aligned}$$

Note that $\delta \ll 1$ by assumption (2.1.15), hence (2.2.2) follows.

STEP 3 (Proof of estimate (2.1.16)). Let $r \ll 1$ and $e \in \mathcal{S}^{d-1}$ be arbitrary, to be fixed later. Let η be supported in $B_{\frac{r}{2}}(x - re)$ and satisfy the bounds

$$\sup |\eta| + r \sup |\nabla \eta| + r^2 \sup |\nabla^2 \eta| \lesssim 1. \quad (2.2.6)$$

We make the additional assumption that η is normalized in such a way that

$$\int (x - x') \eta(x') \, dx' = r^{d+1} e. \quad (2.2.7)$$

Note that since $x \in B_4$ and $r \ll 1$, we have

$$\text{Spt } \eta \subseteq B_{\frac{r}{2}}(x - re) \subset B_5.$$

Integrating inequality (2.2.2) against the measure $\eta(x') \pi(dx' dy')$, it follows that

$$\begin{aligned} (x - y) \cdot \int (x - x') \eta(x') \pi(dx' dy') & \leq 3 \int |x - x'|^2 \eta(x') \pi(dx' dy') \\ & \quad + \int |x' - y'|^2 \eta(x') \pi(dx' dy') \\ & \quad + \delta |x - y| \int |x - x'| \eta(x') \pi(dx' dy'). \end{aligned} \quad (2.2.8)$$

Note that by (2.2.7) the integral on the left-hand-side of inequality (2.2.8) can be expressed as

$$\begin{aligned} \int (x - x')\eta(x') \pi(dx'dy') &= \int (x - x')\eta(x')\mu(dx') \\ &= \kappa_\mu r^{d+1} e + \int (x - x')\eta(x') (\mu(dx') - \kappa_\mu dx'). \end{aligned}$$

To estimate the latter integral, we recall the following result from [GHO21, Lemma 2.8]: for any $\zeta \in C^\infty(B_R)$,

$$\left| \int_{B_R} \zeta (d\mu - \kappa_\mu dx) \right| \leq \left(\kappa_\mu \int_{B_R} |\nabla \zeta|^2 dx W_{B_R}^2(\mu, \kappa_\mu) \right)^{\frac{1}{2}} + \frac{1}{2} \sup_{B_R} |\nabla^2 \zeta| W_{B_R}^2(\mu, \kappa_\mu). \quad (2.2.9)$$

By this estimate with $\zeta = (x - \cdot)\eta$ and using that $\kappa_\mu \sim 1$ by assumption (2.1.14), we obtain with (2.2.6) that

$$\left| \int (x - x')\eta(x') (\mu(dx') - \kappa_\mu dx') \right| \lesssim (r^d \mathcal{D})^{\frac{1}{2}} + \frac{1}{r} \mathcal{D} \lesssim \varepsilon r^{d+1} + \frac{1}{\varepsilon r} \mathcal{D}$$

for some $0 < \varepsilon \ll 1$ to be fixed later. Hence,

$$\begin{aligned} (x - y) \cdot \int (x - x')\eta(x') \pi(dx'dy') \\ \geq \kappa_\mu r^{d+1} (x - y) \cdot e - C \left(\varepsilon r^{d+1} + \frac{1}{\varepsilon r} \mathcal{D} \right) |x - y|. \end{aligned} \quad (2.2.10)$$

We now estimate each term on the right-hand-side of inequality (2.2.8) separately:

1) For the first term we estimate

$$\begin{aligned} \int |x - x'|^2 \eta(x') \pi(dx'dy') &= \int |x - x'|^2 \eta(x') \mu(dx') \\ &\leq \kappa_\mu \int |x - x'|^2 \eta(x') dx' + \left| \int |x - x'|^2 \eta(x') (\mu(dx') - \kappa_\mu dx') \right|. \end{aligned}$$

Using again (2.2.6) and $\kappa_\mu \sim 1$ for the first term on the right-hand side, estimate (2.2.9) with $\zeta = |x - \cdot|^2 \eta$, and Young's inequality for the second term we obtain

$$\int |x - x'|^2 \eta(x') \pi(dx'dy') \lesssim r^{d+2} + \mathcal{D}. \quad (2.2.11)$$

2) For the second term on the right-hand-side of (2.2.8) we use that $\text{Spt } \eta \subseteq B_5$ and (2.2.6), recalling also the definition (2.1.8) of \mathcal{E} , to estimate

$$\int |x' - y'|^2 \eta(x') \pi(dx'dy') \lesssim \int_{B_5 \times \mathbb{R}^d} |x' - y'|^2 \pi(dx'dy') \lesssim \mathcal{E}. \quad (2.2.12)$$

3) We may bound the integral in the third term on the right-hand-side of (2.2.8) as for (2.2.11) by using (2.2.6), $\kappa_\mu \sim 1$ and estimate (2.2.9) with¹⁴ $\zeta = |x - \cdot|\eta$ to get

$$\int |x - x'|\eta(x') \pi(dx'dy') \lesssim r^{d+1} + \frac{1}{r}\mathcal{D}. \quad (2.2.13)$$

Inserting the estimates (2.2.10), (2.2.11), (2.2.12), and (2.2.13) into inequality (2.2.8) yields

$$\kappa_\mu r^{d+1}(x-y) \cdot e \lesssim \left(\varepsilon r^{d+1} + \frac{1}{\varepsilon r} \mathcal{D} \right) |x-y| + r^{d+2} + \mathcal{D} + \varepsilon + \delta \left(r^{d+1} + \frac{1}{r} \mathcal{D} \right) |x-y|.$$

Since e is arbitrary and $\kappa_\mu \sim 1$, this turns into

$$|x-y| \lesssim (\varepsilon + \delta)|x-y| + \left(\delta + \frac{1}{\varepsilon} \right) \left(\frac{1}{r} \right)^{d+2} \mathcal{D} |x-y| + r + \left(\frac{1}{r} \right)^{d+1} (\varepsilon + \mathcal{D}). \quad (2.2.14)$$

We first choose ε and the implicit constant in (2.1.15), which in view of (2.2.3) governs δ , so small that we may absorb the first term on the right-hand-side into the left-hand-side. We then choose r to be a large multiple of $(\varepsilon + \mathcal{D})^{\frac{1}{d+2}}$, so that also the second right-hand-side term in (2.2.14) can be absorbed. This choice of r is admissible in the sense of $r \ll 1$ provided the implicit constant in (2.1.14) is small enough. This yields (2.1.16). \blacksquare

The next lemma shows that due to the global qualitative information on the cost function c , that is, **(C1)–(C4)**, there is a scale below which we can derive a qualitative bound on the displacement. It roughly says that there is a small enough scale after which the cost essentially behaves like Euclidean cost, with an error that is uniformly small due to compactness of the set $X \times Y$.

Lemma 2.2.1. *Assume that the cost function c satisfies **(C1)–(C4)** and let $\pi \in \Pi(\mu, \nu)$ be a coupling with c -monotone support.*

There exist $\Lambda_0 < \infty$ and $R_0 > 0$ such that for all $R \leq R_0$ for which

$$\mathcal{E}_{6R} + \mathcal{D}_{6R} \ll 1, \quad (2.2.15)$$

we have the inclusion

$$\#_{5R} \cap \text{Spt } \pi \subseteq \mathbb{B}_{5R, \Lambda_0 R}.$$

Proof. We only prove the inclusion $(\mathbb{B}_{5R} \times \mathbb{R}^d) \cap \text{Spt } \pi \subseteq \mathbb{B}_{5R} \times \mathbb{B}_{\Lambda_0 R}$, the other inclusion $(\mathbb{R}^d \times \mathbb{B}_{5R}) \cap \text{Spt } \pi \subseteq \mathbb{B}_{\Lambda_0 R} \times \mathbb{B}_{5R}$ follows analogously since the assumptions are symmetric in x and y .

STEP 1 (Use of c -monotonicity of $\text{Spt } \pi$). Let $R > 0$ be such that (2.2.15) holds, in the sense that we may use Lemma 2.A.1(ii), and set

$$\tilde{c}(x, y) := c(x, y) - c(x, 0) - c(0, y) + c(0, 0). \quad (2.2.16)$$

¹⁴Notice that by the assumption that η is supported on $B_{\frac{\varepsilon}{2}}(x - re)$, the function $x' \mapsto |x - x'|\eta(x')$ has no singularity at $x' = x$.

We claim that there exists a constant $\lambda < \infty$, depending only on $\|c\|_{C^2(X \times Y)}$, such that

$$\nabla_x \tilde{c}(x, y) \in B_{\lambda R} \quad \text{for all } (x, y) \in (B_{5R} \times \mathbb{R}^d) \cap \text{Spt } \pi.$$

To show this, we use the c -monotonicity (2.2.1) of $\text{Spt } \pi$. Notice that c -monotonicity of $\text{Spt } \pi$ implies its \tilde{c} -monotonicity:

$$\tilde{c}(x, y) - \tilde{c}(x', y) \leq \tilde{c}(x, y') - \tilde{c}(x', y') \quad \text{for all } (x, y), (x', y') \in \text{Spt } \pi. \quad (2.2.17)$$

With $x_t := tx + (1-t)x'$ we can write

$$\begin{aligned} \tilde{c}(x, y) - \tilde{c}(x', y) &= \int_0^1 \nabla_x \tilde{c}(x_t, y) dt \cdot (x - x') \\ &= \nabla_x \tilde{c}(x, y) \cdot (x - x') + \int_0^1 (\nabla_x \tilde{c}(x_t, y) - \nabla_x \tilde{c}(x, y)) dt \cdot (x - x'), \end{aligned}$$

and, using that $\nabla_x \tilde{c}(0, 0) = 0$,

$$\begin{aligned} \tilde{c}(x, y') - \tilde{c}(x', y') &= (\nabla_x \tilde{c}(0, y') - \nabla_x \tilde{c}(0, 0)) \cdot (x - x') \\ &\quad + \int_0^1 (\nabla_x \tilde{c}(x_t, y') - \nabla_x \tilde{c}(0, y')) dt \cdot (x - x'). \end{aligned}$$

Inserting these two identities into inequality (2.2.17) gives

$$\begin{aligned} \nabla_x \tilde{c}(x, y) \cdot (x - x') &\leq \int_0^1 |\nabla_x \tilde{c}(x_t, y) - \nabla_x \tilde{c}(x, y)| dt |x - x'| + |\nabla_x \tilde{c}(0, y') - \nabla_x \tilde{c}(0, 0)| |x - x'| \\ &\quad + \int_0^1 |\nabla_x \tilde{c}(x_t, y') - \nabla_x \tilde{c}(0, y')| dt |x - x'|. \end{aligned}$$

Using the boundedness of $\|c\|_{C^2(X \times Y)}$ we estimate this expression further by

$$\begin{aligned} \nabla_x \tilde{c}(x, y) \cdot (x - x') &\leq \|\nabla_{xx} \tilde{c}\|_{C^0(X \times Y)} \left(\int_0^1 |x_t - x| dt + \int_0^1 |x_t| dt \right) |x - x'| \\ &\quad + \|\nabla_{xy} \tilde{c}\|_{C^0(X \times Y)} |y'| |x - x'| \\ &\lesssim \|c\|_{C^2(X \times Y)} (|x'| + |x| + |y'|) |x - x'|, \end{aligned} \quad (2.2.18)$$

where in the last step we estimated the integrals and used that

$$\|\nabla_{xx} \tilde{c}\|_{C^0(X \times Y)} \leq 2\|\nabla_{xx} c\|_{C^0(X \times Y)}, \quad \|\nabla_{xy} \tilde{c}\|_{C^0(X \times Y)} = \|\nabla_{xy} c\|_{C^0(X \times Y)}.$$

Now by Lemma 2.A.1(ii), given $x \in B_{5R}$, we have $(S_R(x, e) \times B_{7R}) \cap \text{Spt } \pi \neq \emptyset$ for any direction $e \in S^{d-1}$. Hence, letting $e = \frac{\nabla_x \tilde{c}(x, y)}{|\nabla_x \tilde{c}(x, y)|}$, we can find a point $(x', y') \in (S_R(x, e) \times B_{7R}) \cap \text{Spt } \pi$. Since the opening angle of $S_R(x, e)$ is $\frac{\pi}{2}$, we have

$$\nabla_x \tilde{c}(x, y) \cdot (x - x') = |\nabla_x \tilde{c}(x, y)| |x - x'| e \cdot \frac{x - x'}{|x - x'|} \gtrsim |\nabla_x \tilde{c}(x, y)| |x - x'|.$$

It follows with (2.2.18) that there exists $\lambda < \infty$ such that

$$|\nabla_x \tilde{c}(x, y)| \lesssim \|c\|_{C^2(X \times Y)} (|x'| + |x| + |y'|) \leq \lambda R.$$

STEP 2 (Use of global information on c). We claim that there exist $R_0 > 0$ and $\Lambda_0 < \infty$ such that for all $R \leq R_0$ and $x \in X$, we have that

$$B_{\lambda R} \subseteq -\nabla_x \tilde{c}(x, B_{\Lambda_0 R}).$$

Indeed, by assumption (C2), for any $x \in X$, the map $-\nabla_x \tilde{c}(x, \cdot) = -\nabla_x c(x, \cdot) + \nabla_x c(x, 0)$ is one-to-one on Y . Since $\nabla_{xy} \tilde{c}(x, y) = \nabla_{xy} c(x, y)$, it follows by (C4) that $\det \nabla_{xy} \tilde{c}(x, y) \neq 0$ for all $(x, y) \in X \times Y$. Hence, the map

$$F_x : -\nabla_x \tilde{c}(x, Y) \rightarrow Y, \quad v \mapsto [-\nabla_x \tilde{c}(x, \cdot)]^{-1}(v)$$

is well-defined and a C^1 -diffeomorphism, so that in particular

$$F_x(v) = F_x(0) + DF_x(0)v + o_x(|v|).$$

Using that $-\nabla_x \tilde{c}(x, 0) = 0$, which translates into $F_x(0) = 0$, this takes the form

$$F_x(v) = DF_x(0)v + o_x(|v|).$$

By compactness of X , there thus exist a radius $R_0 > 0$ and a constant $\Lambda_0 < \infty$ such that

$$\lambda |F_x(v)| \leq \Lambda_0 |v| \quad \text{for all } x \in X \quad \text{and} \quad |v| \leq \lambda R_0,$$

which we may reformulate as

$$F_x(B_{\lambda R}) \subseteq B_{\Lambda_0 R}, \quad \text{i.e.} \quad B_{\lambda R} \subseteq F_x^{-1}(B_{\Lambda_0 R}) = -\nabla_x \tilde{c}(x, B_{\Lambda_0 R})$$

for all $R \leq R_0$ and $x \in X$.

STEP 3 (Conclusion). If $(x, y) \in (B_{5R} \times \mathbb{R}^d) \cap \text{Spt } \pi$, then we claim that $|y| \leq \Lambda_0 R$ for $R \leq R_0$.

Indeed, by STEP 1 we have $\nabla_x \tilde{c}(x, y) \in B_{\lambda R}$. Since $B_{\lambda R} \subseteq -\nabla_x \tilde{c}(x, B_{\Lambda_0 R})$ by STEP 2, injectivity of $y \mapsto \nabla_x \tilde{c}(x, y)$ implies that we must have $y \in B_{\Lambda_0 R}$. \blacksquare

Remark 2.2.2. Note that if (2.2.15) is replaced by smallness of the one-sided energy, i.e.,

$$\frac{1}{R^{d+2}} \int_{B_{6R} \times \mathbb{R}^d} |y - x|^2 d\pi + \mathcal{D}_{6R} \ll 1,$$

then Lemma 2.A.1 still applies and we obtain the one-sided qualitative bound

$$(B_{5R} \times \mathbb{R}^d) \cap \text{Spt } \pi \subseteq \mathbb{B}_{5R, \Lambda_0 R}.$$

2.3 Almost-minimality and harmonic approximation

2.3.1 Almost-minimality with respect to Euclidean cost

In this section we show that a minimizer of the optimal transport problem with cost function c is an approximate minimizer for the problem with Euclidean cost function. However, in order to make full use of the Euclidean harmonic approximation result from [GHO21, Proposition 1.6] on the Eulerian side, we have to be careful in relating Lagrangian and Eulerian energies. This is where the concept of almost-minimality shows its strength, since it provides us with the missing bound of Lagrangian energy in terms of its Eulerian counterpart.

Proof of Proposition 2.1.10. First, let us observe that we may assume in the following that

$$\frac{1}{2} \int |x - y|^2 d(\pi - \tilde{\pi}) \geq 0, \quad (2.3.1)$$

since otherwise there is nothing to show. By the support assumption (2.1.19) on μ and ν , the couplings π and $\tilde{\pi}$ satisfy

$$\text{Spt } \pi \subseteq B_R \times B_R, \quad \text{Spt } \tilde{\pi} \subseteq B_R \times B_R. \quad (2.3.2)$$

Together with (2.3.1) this implies that

$$\mathcal{E}_R(\tilde{\pi}) \leq \mathcal{E}_R(\pi). \quad (2.3.3)$$

By the admissibility of $\tilde{\pi}$, i.e. that π and $\tilde{\pi}$ have the same marginals, we may write

$$\begin{aligned} \frac{1}{2} \int |x - y|^2 d(\pi - \tilde{\pi}) &= - \int (c(x, y) + x \cdot y) d(\pi - \tilde{\pi}) + \int c(x, y) d(\pi - \tilde{\pi}) \\ &= - \int (\tilde{c}(x, y) + x \cdot y) d(\pi - \tilde{\pi}) + \int c(x, y) d(\pi - \tilde{\pi}), \end{aligned} \quad (2.3.4)$$

where \tilde{c} is defined as in (2.2.16). Abbreviating

$$\zeta(x, y) := -(\tilde{c}(x, y) + x \cdot y),$$

optimality of π with respect to the cost function c implies that

$$\frac{1}{2} \int |x - y|^2 d(\pi - \tilde{\pi}) \leq \int \zeta(x, y) d(\pi - \tilde{\pi}). \quad (2.3.5)$$

Using again the admissibility of $\tilde{\pi}$, we may write

$$\int \zeta(x, y) d(\pi - \tilde{\pi}) = \int (\zeta(x, y) - \zeta(x, x)) d(\pi - \tilde{\pi}).$$

Note that by the definition (2.2.16) of \tilde{c} , the function ζ satisfies

$$\nabla_x \zeta(x, 0) = 0 \quad \text{for all } x \quad \text{and} \quad \nabla_y \zeta(0, y) = 0 \quad \text{for all } y, \quad (2.3.6)$$

$$\nabla_{xy} \zeta(x, y) = -(\nabla_{xy} \tilde{c}(x, y) + \mathbb{I}) = -(\nabla_{xy} c(x, y) + \mathbb{I}). \quad (2.3.7)$$

Now, by (2.3.6),

$$\begin{aligned} \zeta(x, y) - \zeta(x, x) &= \int_0^1 (\nabla_y \zeta(x, sy + (1-s)x) - \nabla_y \zeta(0, sy + (1-s)x)) \cdot (y - x) ds \\ &= \int_0^1 \int_0^1 x \cdot \nabla_{xy} \zeta(tx, sy + (1-s)x)(y - x) dt ds, \end{aligned}$$

so that, using (2.3.7) and (2.3.2), it follows that

$$\begin{aligned} \left| \int \zeta(x, y) d(\pi - \tilde{\pi}) \right| &\lesssim \|\nabla_{xy} c + \mathbb{I}\|_{L^\infty(B_R \times B_R)} \int_{B_R \times B_R} |x| |y - x| d(\pi + \tilde{\pi}) \\ &\lesssim R \|\nabla_{xy} c + \mathbb{I}\|_{L^\infty(B_R \times B_R)} \int |y - x| d(\pi + \tilde{\pi}). \end{aligned} \quad (2.3.8)$$

By the estimate

$$(\pi + \tilde{\pi})(\mathbb{R}^d \times \mathbb{R}^d) = 2\mu(\mathbb{R}^d) = 2\mu(B_R) \stackrel{(2.1.20)}{\lesssim} R^d,$$

and Hölder's inequality, we get

$$\int |x - y| d(\pi + \tilde{\pi}) \lesssim R^{\frac{d}{2}} \left(\int |x - y|^2 d(\pi + \tilde{\pi}) \right)^{\frac{1}{2}} \stackrel{(2.3.3)}{\lesssim} R^{d+1} \mathcal{E}(\pi)^{\frac{1}{2}}.$$

■

2.3.2 Eulerian point of view

The purpose of this section is to translate almost-minimality from the Lagrangian setting, as encoded by Proposition 2.1.10, to the Eulerian setting so that it may be plugged into the proof of the harmonic approximation result [GHO21, Proposition 1.7]. This is the purpose of Lemma 2.1.12 below, which relates a (Lagrangian) coupling π , which we think of being an almost-minimizer, to its Eulerian description (ρ, j) (introduced in (2.1.23)).

The proof of Lemma 2.1.12 relies on the fact that the Eulerian cost is always dominated by the Lagrangian one, while the other inequality in general only holds for minimizers of the Euclidean transport cost. However, in the proof of Lemma 2.1.12 we can use almost-minimality of π , together with the equality of Eulerian and Lagrangian energy for minimizers of the Euclidean transport cost (Remark 2.1.11) to overcome this nuisance.

Proof of Lemma 2.1.12. Let $\pi^* \in \Pi(\mu, \nu)$ be a minimizer of the Euclidean transport cost between μ and ν and let (ρ^*, j^*) be its Eulerian description. Then by (2.1.30) and (2.1.33) it follows that

$$\int \frac{1}{\rho} |j|^2 \leq \int |x - y|^2 d\pi \leq \int |x - y|^2 d\pi^* + \Delta.$$

Since $W^2(\mu, \nu) = \frac{1}{2} \int |y - x|^2 d\pi^*$, we have that

$$\int |x - y|^2 d\pi^* = \int \frac{1}{\rho^*} |j^*|^2,$$

see Remark 2.1.11, so that by minimality of π^* , which implies minimality of (ρ^*, j^*) in the Benamou–Brenier formulation (2.1.31) of $W^2(\mu, \nu)$, we obtain

$$\int \frac{1}{\rho} |j|^2 \leq \int |x - y|^2 d\pi \leq \int \frac{1}{\rho^*} |j^*|^2 + \Delta \leq \int \frac{1}{\tilde{\rho}} |\tilde{j}|^2 + \Delta$$

for any $(\tilde{\rho}, \tilde{j})$ satisfying (2.1.34), in particular also for $(\tilde{\rho}, \tilde{j}) = (\rho, j)$. ■

The Eulerian version of almost-minimality also implies the following localized version, which will be needed for the harmonic approximation:

Corollary 2.3.1. *Let $\pi \in \Pi(\mu, \nu)$ be a coupling between the measures μ and ν with the property that there exists a constant $\Delta < \infty$ such that*

$$\int \frac{1}{2}|x - y|^2 d\pi \leq \int \frac{1}{2}|x - y|^2 d\tilde{\pi} + \Delta \quad (2.3.9)$$

for any $\tilde{\pi} \in \Pi(\mu, \nu)$, and let (ρ, j) be its Eulerian description defined in (2.1.23). For any $r > 0$ small enough¹⁵, let f_r be the inner trace of j on $\partial B_r \times (0, 1)$ in the sense of (2.1.26), i.e.

$$\int_{B_r \times [0,1]} \partial_t \zeta d\rho + \nabla \zeta \cdot dj = \int_{B_r} \zeta_1 d\nu - \int_{B_r} \zeta_0 d\mu + \int_{\partial B_r \times [0,1]} \zeta df_r. \quad (2.3.10)$$

Then for any density-flux pair $(\tilde{\rho}, \tilde{j})$ satisfying

$$\int \partial_t \zeta d\tilde{\rho} + \nabla \zeta \cdot d\tilde{j} = \int_{B_r} \zeta_1 d\nu - \int_{B_r} \zeta_0 d\mu + \int_{\partial B_r \times [0,1]} \zeta df_r \quad (2.3.11)$$

there holds

$$\frac{1}{2} \int_{B_r \times [0,1]} \frac{1}{\tilde{\rho}} |j|^2 \leq \frac{1}{2} \int \frac{1}{\tilde{\rho}} |\tilde{j}|^2 + \Delta. \quad (2.3.12)$$

Proof. Let $(\tilde{\rho}, \tilde{j})$ satisfy (2.3.11). Then the density-flux pair $(\widehat{\rho}, \widehat{j}) := (\tilde{\rho}, \tilde{j}) + (\rho, j) \llcorner_{B_r^c \times [0,1]}$ obeys (2.1.34). Indeed,

$$\begin{aligned} \int \partial_t \zeta d\widehat{\rho} + \nabla \zeta \cdot d\widehat{j} &= \int \partial_t \zeta d\tilde{\rho} + \nabla \zeta \cdot d\tilde{j} + \int_{B_r^c \times [0,1]} \partial_t \zeta d\rho + \nabla \zeta \cdot dj \\ &= \int \partial_t \zeta d\tilde{\rho} + \nabla \zeta \cdot d\tilde{j} + \int \partial_t \zeta d\rho + \nabla \zeta \cdot dj - \int_{B_r \times [0,1]} \partial_t \zeta d\rho + \nabla \zeta \cdot dj \\ &\stackrel{(2.3.11) \& (2.1.25) \& (2.3.10)}{=} \int \zeta_1 d\nu - \int \zeta_0 d\mu. \end{aligned}$$

Hence, by Lemma 2.1.12 and subadditivity of the Eulerian cost it follows that

$$\frac{1}{2} \int \frac{1}{\tilde{\rho}} |j|^2 \leq \frac{1}{2} \int \frac{1}{\widehat{\rho}} |\widehat{j}|^2 + \Delta \leq \frac{1}{2} \int \frac{1}{\tilde{\rho}} |\tilde{j}|^2 + \frac{1}{2} \int_{B_r^c \times [0,1]} \frac{1}{\rho} |j|^2 + \Delta,$$

which implies (2.3.12). ■

Lemma 2.1.12, in the form of the bound (2.1.43), allows us to relate Eulerian and Lagrangian side of the harmonic approximation result, which will be central in the application to the one-step-improvement Proposition 2.1.16. The proof of this Lagrangian version is very similar to [GHO21, Proof of Theorem 1.5], however, we stress again that since we are not dealing with minimizers of the Euclidean transport cost, one has to be careful when passing from Eulerian to Lagrangian energies.

¹⁵ r must be small enough so that $B_r \subset \text{Spt } \mu \cap \text{Spt } \nu$.

Proof of Lemma 2.1.14. By scaling, it is enough to show that

$$\int_{\#_1} \int_0^1 |x - y + \nabla\Phi(tx + (1-t)y)|^2 dt d\pi \leq \int_{B_2 \times [0,1]} \frac{1}{\rho} |j - \rho \nabla\Phi|^2 + \Delta_1. \quad (2.3.13)$$

By the L^∞ -bound, we know that if $(x, y) \in \#_1 \cap \text{Spt } \pi$, then for any $t \in [0, 1]$ there holds $tx + (1-t)y \in B_2$. Hence, we may estimate

$$\begin{aligned} & \int_{\#_1} \int_0^1 |x - y + \nabla\Phi(tx + (1-t)y)|^2 dt d\pi \\ & \leq \iint_0^1 |x - y + \nabla\Phi(tx + (1-t)y)|^2 1_{\{tx+(1-t)y \in B_2\}} dt d\pi. \end{aligned} \quad (2.3.14)$$

Multiplying out the square and using the definition of (ρ, j) from (2.1.23), we may write¹⁶

$$\begin{aligned} & \iint_0^1 |x - y + \nabla\Phi(tx + (1-t)y)|^2 1_{\{tx+(1-t)y \in B_2\}} dt d\pi \\ & = \iint_0^1 |x - y|^2 1_{\{tx+(1-t)y \in B_2\}} dt d\pi + \iint_0^1 |\nabla\Phi(tx + (1-t)y)|^2 1_{\{tx+(1-t)y \in B_2\}} dt d\pi \\ & \quad + 2 \iint_0^1 (x - y) \cdot \nabla\Phi(tx + (1-t)y) 1_{\{tx+(1-t)y \in B_2\}} dt d\pi \\ & = \int |x - y|^2 d\pi - \iint_0^1 |x - y|^2 1_{\{tx+(1-t)y \in B_2^c\}} dt d\pi + \int_{B_2 \times [0,1]} |\nabla\Phi|^2 d\rho - 2 \int_{B_2 \times [0,1]} \nabla\Phi \cdot dj. \end{aligned}$$

Now note that (2.1.29) implies a local counterpart of (2.1.30): for any open set $B \subseteq \mathbb{R}^d$,

$$\int_{B \times [0,1]} \frac{1}{\rho} |j|^2 \leq \iint_0^1 |x - y|^2 1_{\{tx+(1-t)y \in B\}} dt d\pi. \quad (2.3.15)$$

Arguing with an open ε -neighborhood of B and continuity of the right-hand side with respect to B , one may show that (2.3.15) holds for any closed set B , so that in particular

$$\int_{B_2^c \times [0,1]} \frac{1}{\rho} |j|^2 \leq \iint_0^1 |x - y|^2 1_{\{tx+(1-t)y \in B_2^c\}} dt d\pi,$$

which, together with (2.1.43), gives

$$\begin{aligned} \int |x - y|^2 d\pi - \iint_0^1 |x - y|^2 1_{\{tx+(1-t)y \in B_2^c\}} dt d\pi & \leq \int \frac{1}{\rho} |j|^2 + \Delta_1 - \int_{B_2^c \times [0,1]} \frac{1}{\rho} |j|^2 \\ & = \int_{B_2 \times [0,1]} \frac{1}{\rho} |j|^2 + \Delta_1, \end{aligned}$$

¹⁶Note that by an approximation argument, we may use $\zeta := 1_{B_2} |\nabla\Phi|^2$ and $\xi := 1_{B_2} \nabla\Phi$ as test functions in (2.1.23).

so that (2.3.13) follows from the identity

$$\begin{aligned}
 & \int_{B_2 \times [0,1]} \frac{1}{\rho} |j|^2 + \int_{B_2 \times [0,1]} |\nabla \Phi|^2 \, d\rho - 2 \int_{B_2 \times [0,1]} \nabla \Phi \cdot dj \\
 &= \int_{B_2 \times [0,1]} \left(\left| \frac{dj}{d\rho} \right|^2 + |\nabla \Phi|^2 - 2 \nabla \Phi \cdot \frac{dj}{d\rho} \right) d\rho = \int_{B_2 \times [0,1]} \left| \frac{dj}{d\rho} - \nabla \Phi \right|^2 d\rho \\
 &= \int_{B_2 \times [0,1]} \frac{1}{\rho} |j - \rho \nabla \Phi|^2. \tag{2.3.16}
 \end{aligned}$$

■

2.3.3 Harmonic Approximation

In this section we sketch the proof of the (Eulerian) harmonic approximation Theorem 2.1.13. As already noted in the introduction, the proof of Theorem 2.1.13 is done at the Eulerian level (as in [GO20, GHO21]) by constructing a suitable competitor.

Proof of Theorem 2.1.13. By scaling we may without loss of generality assume that $R = 1$. Let (ρ, j) be the Eulerian description of the coupling $\pi \in \Pi(\mu, \nu)$. The proof of the Eulerian version of the harmonic approximation consists of the following four steps, at the heart of which is the construction of a suitable competitor (STEP 3). Note that since we want to make the dependence on the parameter M in the L^∞ -bound (2.1.36) precise, one actually has to look a bit closer into the proofs of the corresponding statements in [GHO21], since in their presentation of the results the estimate $M \lesssim (\mathcal{E}_6 + \mathcal{D}_6)^{\frac{1}{d+2}}$ is used.¹⁷

STEP 1 (Passage to a regularized problem). Choose a good radius $R_* \in (3, 4)$ for which the flux \bar{f} is well-behaved on ∂B_{R_*} . Actually, since we are working with L^2 -based quantities, to be able to get L^2 bounds on $\nabla \Phi$, we would have to be able to estimate \bar{f} in L^2 or at least in the Sobolev trace space $L^{\frac{2(d-1)}{d}}$. However, since the boundary fluxes j are just measures and since for the approximate orthogonality (see STEP 2) a regularity theory is required up to the boundary, one first has to go to a solution φ (with $\int_{B_{R_*}} \varphi \, dx = 0$) of a regularized problem

$$\Delta \varphi = \text{const} \quad \text{in } B_{R_*} \quad \text{and} \quad \nu_{R_*} \cdot \nabla \varphi = \widehat{g} \quad \text{on } \partial B_{R_*}, \tag{2.3.17}$$

where const is the generic constant for which the equation has a solution, and \widehat{g} is a regularization through rearrangement of \bar{f} with good L^2 bounds (see [GHO21, Section 3.1.2] for details).

Using properties of the regularized flux \widehat{g} and elliptic regularity, the error made by this approximation can be quantified as

$$\sup_{B_2} |\nabla(\Phi - \varphi)|^2 + \sup_{B_2} |\nabla^2(\Phi - \varphi)|^2 + \sup_{B_2} |\nabla^3(\Phi - \varphi)|^2 \lesssim M^\alpha (\mathcal{E}_6 + \mathcal{D}_6)^{1+\frac{\alpha}{2}}, \tag{2.3.18}$$

¹⁷We do not assume that this bound holds in assumption (2.1.36). However, in the one-step improvement Proposition 2.1.16 and the consecutive iteration to obtain the ε -regularity Theorem 2.1.1, this is of course an important ingredient, which holds in view of Proposition 2.1.5.

for any $\alpha \in (0, 1)$, see [GHO21, Proof of Proposition 1.7]. Note that by the definition of ρ in (2.1.23) and assumption (2.1.35) we have that

$$\int_{B_2 \times [0,1]} d\rho \leq \int d\rho = \int d\pi = \mu(\mathbb{R}^d) = \mu(B_7) \lesssim 1,$$

so that together with (2.3.18) we may estimate, using $M \leq 1$,

$$\int_{B_2 \times [0,1]} \frac{1}{\rho} |j - \rho \nabla \Phi|^2 \lesssim \int_{B_2 \times [0,1]} \frac{1}{\rho} |j - \rho \nabla \varphi|^2 + (\mathcal{E}_6 + \mathcal{D}_6)^{1+\frac{\alpha}{2}}$$

for any $\alpha \in (0, 1)$. Note that the error term is superlinear in $\mathcal{E}_6 + \mathcal{D}_6$.

STEP 2 (Approximate orthogonality [GHO21, Proof of Lemma 1.8]). For every $0 < \tau \ll 1$ there exist $\varepsilon_\tau > 0$, $C_\tau < \infty$ such that if $\mathcal{E}_6 + \mathcal{D}_6 \leq \varepsilon_\tau$, then

$$\int_{B_2 \times [0,1]} \frac{1}{\rho} |j - \rho \nabla \varphi|^2 - \left(\int_{B_{R_*} \times [0,1]} \frac{1}{\rho} |j|^2 - \int_{B_{R_*}} |\nabla \varphi|^2 \right) \leq \tau \mathcal{E}_6 + C_\tau \mathcal{D}_6. \quad (2.3.19)$$

The proof of (2.3.19) essentially relies on the representation formula

$$\begin{aligned} \int_{B_{R_*} \times [0,1]} \frac{1}{\rho} |j - \rho \nabla \varphi|^2 &= \int_{B_{R_*} \times [0,1]} \frac{1}{\rho} |j|^2 - \int_{B_{R_*}} |\nabla \varphi|^2 \\ &\quad + 2 \int_{B_{R_*}} \varphi d(\mu - \nu) + 2 \int_{\partial B_{R_*}} \varphi d(\widehat{g} - \overline{f}) + \int_{B_{R_*} \times [0,1]} |\nabla \varphi|^2 (d\rho - dx). \end{aligned}$$

The three error terms in the second line of this equality are then bounded as follows. The first term uses that μ and ν are close in Wasserstein distance. An estimate on the second term relies on the fact that \widehat{g} and \overline{f} are close¹⁸. This bound relies on the choice of a good radius in STEP 1 and L^2 estimates up to the boundary on $\nabla \varphi$. The bound on the third error term uses elliptic regularity theory and a restriction result for the Wasserstein distance, which implies that $W_{B_{R_*}}^2(\int_0^1 \rho, \kappa) \lesssim \mathcal{E}_6 + \mathcal{D}_6$.¹⁹ This estimate actually requires a further regularization of \widehat{g} , and by relying on interior regularity estimates explains why one has to go from B_{R_*} to the slightly smaller ball B_2 in the estimate (2.3.19). A close inspection of [GHO21, Proof of Lemma 1.8] shows that the term involving M in these error estimates comes in product with a superlinear power of $\mathcal{E}_6 + \mathcal{D}_6$ as in STEP 1, so that we may bound $M \leq 1$ and still be able to obtain an arbitrarily small prefactor in (2.3.19) by choosing $\mathcal{E}_6 + \mathcal{D}_6$ small enough.

STEP 3 (Construction of a competitor [GHO21, Proof of Lemma 1.9]). For every $0 < \tau \ll 1$ there exist $\varepsilon_\tau > 0$ and $C, C_\tau < \infty$ such that if $\mathcal{E}_6 + \mathcal{D}_6 \leq \varepsilon_\tau$, then there exists a density-flux pair $(\widetilde{\rho}, \widetilde{j})$ satisfying (2.3.11) for $r = R_*$, and such that

$$\int \frac{1}{\widetilde{\rho}} |\widetilde{j}|^2 - \int_{B_{R_*}} |\nabla \varphi|^2 \leq \left(\tau + \frac{CM}{\tau} \right) \mathcal{E}_6 + C_\tau \mathcal{D}_6. \quad (2.3.20)$$

¹⁸Closeness here means closeness of \widehat{g}_\pm and \overline{f}_\pm (the positive and negative parts of the measures) with respect to the geodesic Wasserstein distance on ∂B_{R_*} .

¹⁹We note that for the case of quadratic cost and Hölder continuous densities treated in [GO20] a bound on this term is easy due to McCann's displacement convexity, which implies that $\rho \leq 1$ up to a small error.

STEP 4 (Almost-minimality on the Eulerian level). Since the density-flux pair $(\tilde{\rho}, \tilde{j})$ satisfies (2.3.11) for $r = R_*$, Corollary 2.3.1 implies that

$$\int_{B_{R_*} \times [0,1]} \frac{1}{\rho} |j|^2 - \int_{B_{R_*}} |\nabla \varphi|^2 \leq \int_{B_{R_*}} \frac{1}{\tilde{\rho}} |\tilde{j}|^2 - \int_{B_{R_*}} |\nabla \varphi|^2 + \Delta. \quad (2.3.21)$$

Combining the above steps, we have proved that for any $0 < \tau \ll 1$ there exist $\varepsilon_\tau > 0$ and $C_\tau < \infty$ such that if Φ is the solution of (2.1.39), then

$$\int_{B_2 \times [0,1]} \frac{1}{\rho} |j - \rho \nabla \Phi|^2 \leq \left(\tau + \frac{CM}{\tau} \right) \mathcal{E}_6 + C_\tau \mathcal{D}_6 + \Delta. \quad (2.3.22)$$

■

2.4 The ε -regularity result

2.4.1 One-step improvement

The following proposition is a one-step improvement result, which will be the basis of a Campanato iteration in Theorem 2.1.1. Note that the iteration is more complicated than in [GO20], because at each step we have to restrict the c -optimal coupling to the smaller cross where the L^∞ -bound holds to be able to apply the harmonic approximation result. As a consequence, we have to make sure that the qualitative bound (2.2.15) on the displacement (which is an important ingredient in obtaining quantitative version of the L^∞ -bound (2.1.16)) is propagated in each step of the iteration.²⁰

We start with the short proof of Lemma 2.1.8, which is the starting point of each iteration step.

Proof of Lemma 2.1.8. Let $\pi \in \Pi(\mu, \nu)$ be a c -optimal coupling, and let $\pi_\Omega := \pi|_\Omega$ be its restriction to a subset $\Omega \subseteq \mathbb{R}^d \times \mathbb{R}^d$. Then $\pi_\Omega \in \Pi(\mu_\Omega, \nu_\Omega)$ where the marginal measures μ_Ω, ν_Ω are defined in (2.1.18).

Given any coupling $\tilde{\pi} \in \Pi(\mu_\Omega, \nu_\Omega)$, we can define $\hat{\pi} := \pi - \pi_\Omega + \tilde{\pi}$. It is easy to see that $\hat{\pi}$ is an admissible coupling between the measures μ and ν , hence by c -optimality of π , we obtain by the additivity of the cost functional with respect to the transference plan

$$\int c(x, y) d\pi \leq \int c(x, y) d\hat{\pi} = \int c(x, y) d\pi - \int c(x, y) d\pi_\Omega + \int c(x, y) d\tilde{\pi},$$

hence

$$\int c(x, y) d\pi_\Omega \leq \int c(x, y) d\tilde{\pi} \quad \text{for any } \tilde{\pi} \in \Pi(\mu_\Omega, \nu_\Omega),$$

that is, π_Ω is a c -optimal coupling between μ_Ω and ν_Ω . ■

²⁰Alternatively, one could devise an argument based on the fact that the qualitative L^∞ bound only depends on the cost through its global properties (C1)–(C4) and that the set of cost functions considered in the iteration is relatively compact.

As a direct application and as a further preparation for the one-step-improvement we present the short proof of Corollary 2.1.9.

Proof of Corollary 2.1.9. Let $\pi \in \Pi(\mu, \nu)$ be c -optimal. Then by Lemma 2.1.8 the coupling $\pi_R := \pi \llcorner_{\#_R}$ is a c -optimal coupling between the measures μ_R and ν_R defined via

$$\mu_R(A) = \pi((A \times \mathbb{R}^d) \cap \#_R), \quad \nu_R(A) = \pi((\mathbb{R}^d \times A) \cap \#_R),$$

for any $A \subseteq \mathbb{R}^d$ Borel. In particular, we have that

$$\mu_R(A) \leq \pi(A \times \mathbb{R}^d) = \mu(A).$$

If $A \subseteq B_{2R}^c$, then

$$\mu_R(A) = \pi(((A \cap B_R) \times \mathbb{R}^d) \cup (A \times B_R)) = \pi(A \times B_R) = 0,$$

since for any $(x, y) \in \text{Spt } \pi_R$, by assumption we have that $|x - y| \leq MR \leq R$, so $A \times B_R \cap \text{Spt } \pi = \emptyset$. Hence, $\text{Spt } \mu_R \subseteq B_{2R}$. Similarly, if $A \subseteq B_R$, then

$$\mu_R(A) = \pi((A \times \mathbb{R}^d) \cup (A \times B_R)) = \pi(A \times \mathbb{R}^d) = \mu(A),$$

which implies that $\mu_R = \mu$ on B_R .

By symmetry, the same properties hold for ν_R . ■

We now give the proof of the one-step improvement Proposition 2.1.16, which is the working horse of the Campanato iteration.

Proof of Proposition 2.1.16. STEP 1 (Rescaling). Without loss of generality we can assume that $R = 1$. Hence, to simplify notation we will also use the shorthand

$$\begin{aligned} \mathcal{E} &:= \mathcal{E}_g(\pi), \quad \mathcal{D} := \mathcal{D}_g(\rho_0, \rho_1), \\ \text{and } [\rho_0]_\alpha &:= [\rho_0]_{\alpha,9}, \quad [\rho_1]_\alpha := [\rho_1]_{\alpha,9}, \quad [\nabla_{xy}c]_\alpha := [\nabla_{xy}c]_{\alpha,9}. \end{aligned}$$

Note that by the normalization $\rho_0(0) = \rho_1(0) = 1$ we have that $\frac{1}{2} \leq \rho_j \leq 2$ for $j = 1, 2$, provided that $[\rho_0]_\alpha + [\rho_1]_\alpha$ is small enough. In particular, Lemma 2.A.4 then implies that

$$\mathcal{D} \lesssim [\rho_0]_\alpha^2 + [\rho_1]_\alpha^2.$$

Indeed, let $(\tilde{x}, \tilde{y}) = S_R(x, y) := (R^{-1}x, R^{-1}y)$ and set

$$\tilde{\rho}_0(\tilde{x}) = \rho_0(R\tilde{x}), \quad \tilde{\rho}_1(\tilde{y}) = \rho_1(R\tilde{y}), \quad \tilde{c}(\tilde{x}, \tilde{y}) = R^{-2}c(R\tilde{x}, R\tilde{y}),$$

and $\tilde{b} = R^{-1}b$. We still have $\tilde{\rho}_0(0) = \tilde{\rho}_1(0) = 1$ and $\nabla_{\tilde{x}\tilde{y}}\tilde{c}(0, 0) = -\mathbb{I}$, and $\tilde{\pi} := R^{-d}(S_R)_\# \pi$ is the \tilde{c} -optimal coupling between $\tilde{\rho}_0$ and $\tilde{\rho}_1$.

STEP 2 (L^∞ -bound for π). We claim that the coupling π satisfies the L^∞ -bound

$$|x - y| \leq M \quad \text{for all } (x, y) \in \text{Spt } \pi \llcorner_{\#_6} \quad (2.4.1)$$

with $M = C_\infty(\mathcal{E} + \mathcal{D})^{\frac{1}{d+2}}$ for some constant $C_\infty < \infty$ depending only on d .

In view of the harmonic approximation Theorem 2.1.13 it is therefore natural to consider the coupling $\pi \llcorner_{\#_6}$, which by Lemma 2.1.8 and Corollary 2.1.9 is a c -optimal coupling between its own marginals, denoted by μ and ν .

Indeed, since π is c -optimal, its support is c -monotone. By assumptions (2.1.48), (2.1.49), and Remark 2.1.7 we may therefore estimate

$$\|\nabla_{xy}c + \mathbb{I}\|_{C^0(\mathbb{B}_{8,\Lambda})} \lesssim_\Lambda [\nabla_{xy}c]_\alpha \ll_\Lambda 1. \quad (2.4.2)$$

Hence, appealing to Proposition 2.1.5 (suitably rescaled), we obtain the claimed L^∞ -bound (2.4.1). Notice that the dependence of the smallness assumption (2.1.49) on Λ only enters through the Λ -dependence of (2.4.2).

STEP 3 (Properties of the marginals μ and ν of $\pi \llcorner_{\#_6}$). We claim that the marginals μ and ν of $\pi \llcorner_{\#_6}$ satisfy

$$\text{Spt } \mu, \text{Spt } \nu \subseteq B_7, \quad \text{and} \quad \mu(B_7) \leq 2|B_7|.$$

Indeed, due to the L^∞ -bound (2.4.1) the marginal measures are supported on B_7 if $\mathcal{E} + \mathcal{D} \ll 1$ (such that $M \leq 1$). Furthermore, from Corollary 2.1.9 we have that $\mu = \rho_0 dx$ and $\nu = \rho_1 dy$ on B_6 , as well as $\mu \leq \rho_0 dx$, which implies that

$$\mu(B_7) \leq \int_{B_7} \rho_0 dx \leq (1 + [\rho_0]_\alpha)|B_7| \leq 2|B_7|$$

since by assumption (2.1.49) we may assume that $[\rho_0]_\alpha \leq 1$.

STEP 4 (Almost-minimality of $\pi \llcorner_{\#_6}$ and applicability of the harmonic approximation Theorem 2.1.13). We show next that the coupling $\pi \llcorner_{\#_6}$ is an almost-minimizer of the Euclidean transport problem in the sense that for any $\tilde{\pi} \in \Pi(\mu, \nu)$ there holds

$$\frac{1}{2} \int |x - y|^2 d\pi \llcorner_{\#_6} \leq \frac{1}{2} \int |x - y|^2 d\tilde{\pi} + \Delta, \quad (2.4.3)$$

with $\Delta = C[\nabla_{xy}c]_\alpha \mathcal{E}^{\frac{1}{2}}$, and that $\pi \llcorner_{\#_6}$ satisfies the assumptions of Theorem 2.1.13.

Indeed, by STEP 3 and the c -optimality of $\pi \llcorner_{\#_6}$, we may apply Proposition 2.1.10 to obtain inequality (2.4.3) with

$$\Delta = C\|\nabla_{xy}c + \mathbb{I}\|_{C^0(B_7 \times B_7)} \mathcal{E}_7^{\frac{1}{2}} \lesssim [\nabla_{xy}c]_\alpha \mathcal{E}^{\frac{1}{2}}, \quad (2.4.4)$$

where we also used that $\mathcal{E}_7 \lesssim \mathcal{E}$.

Note that in view of Lemma 2.1.12 the Eulerian description of $\pi \llcorner_{\#_6}$ satisfies (2.1.37). Together with STEP 2 and STEP 3 this implies that the assumptions of Theorem 2.1.13 are fulfilled.

STEP 5 (Definition of b and B and proof of estimate (2.1.50)). For given $\tau > 0$, to be chosen later, let ε_τ and C_τ be as in the harmonic approximation result (Theorem 2.1.13). We claim that, with the harmonic gradient field $\nabla\Phi$ defined in (2.1.39),

$$b := \nabla\Phi(0), \quad A := \nabla^2\Phi(0), \quad \text{and} \quad B := \mathcal{E}^{\frac{A}{2}}, \quad (2.4.5)$$

satisfy the bounds (2.1.50).

Since $\rho_0(0) = \rho_1(0) = 1$, we have $\frac{1}{2} \leq \rho_j \leq 2$, $j = 0, 1$, provided $[\rho_0]_\alpha^2 + [\rho_1]_\alpha^2$ is small enough, so that (2.1.11) holds. In particular, since μ agrees with $\rho_0 dx$ on B_6 and ν agrees with $\rho_1 dy$ on B_6 , we may estimate

$$\mathcal{D}_6(\mu, \nu) = \mathcal{D}_6(\rho_0, \rho_1) \stackrel{(2.1.11)}{\lesssim} [\rho_0]_{\alpha,6}^2 + [\rho_1]_{\alpha,6}^2 \lesssim [\rho_0]_\alpha^2 + [\rho_1]_\alpha^2.$$

Assume now that

$$\mathcal{E}_6(\pi) + \mathcal{D}_6(\mu, \nu) \lesssim \mathcal{E} + [\rho_0]_\alpha^2 + [\rho_1]_\alpha^2 \leq \varepsilon_\tau, \quad (2.4.6)$$

so that Theorem 2.1.13 allows us to define the vector b and the matrices A, B as in (2.4.5). Note that since A is symmetric, so is the matrix B . By (2.1.41) and (2.1.11), we then have

$$|b|^2 \lesssim \mathcal{E}_6(\pi) + \mathcal{D}_6(\mu, \nu) \lesssim \mathcal{E} + [\rho_0]_\alpha^2 + [\rho_1]_\alpha^2,$$

and the same estimate holds for $|A|^2$, so that

$$|B - \mathbb{I}|^2 \lesssim \mathcal{E} + [\rho_0]_\alpha^2 + [\rho_1]_\alpha^2.$$

This proves the estimate (2.1.50). Furthermore, recalling the definition of Φ in (2.1.39), we bound

$$\begin{aligned} |1 - \det B^2|^2 &= |1 - \mathcal{E}^{\text{tr}A}|^2 = \left|1 - \mathcal{E}^{\Delta\Phi(0)}\right|^2 = |1 - \mathcal{E}^{\kappa_\mu - \kappa_\nu}|^2 \\ &\lesssim |\kappa_\nu - \kappa_\mu|^2 \stackrel{(2.1.9)}{\lesssim} \mathcal{D}_{R^*}(\mu, \nu) \stackrel{(2.1.11)}{\lesssim} R_*^{2\alpha} \left([\rho_0]_{\alpha, R^*}^2 + [\rho_1]_{\alpha, R^*}^2\right) \lesssim [\rho_0]_\alpha^2 + [\rho_1]_\alpha^2. \end{aligned} \quad (2.4.7)$$

In view of (2.1.50) we may assume that $b \in B_9$, so that²¹

$$|\rho_1(b)^{\frac{1}{d}} - \rho_1(0)^{\frac{1}{d}}|^2 \lesssim |\rho_1(b) - \rho_1(0)|^2 \leq |b|^{2\alpha} [\rho_1]_\alpha^2 \lesssim [\rho_1]_\alpha^2, \quad (2.4.8)$$

and similarly for $D = -\nabla_{xy}c(0, b)$,

$$|D - \mathbb{I}|^2 \lesssim [\nabla_{xy}c]_\alpha^2, \quad (2.4.9)$$

which implies that

$$|1 - \det D|^2 \lesssim |D - \mathbb{I}|^2 \lesssim [\nabla_{xy}c]_\alpha^2. \quad (2.4.10)$$

²¹The first inequality follows from $\rho_1 \geq \frac{1}{2}$ and the Lipschitz continuity of the function $t \mapsto t^{\frac{1}{d}}$ away from zero.

Now, noticing that $\gamma = \rho_1(b)^{\frac{1}{d}} \left(\frac{|\det B|^2}{|\det D|} \right)^{\frac{1}{d}}$ because $\rho_0(0) = 1$, (2.4.7), (2.4.8) together with $\rho_1(0) = 1$, and (2.4.10) imply

$$|1 - \gamma|^2 \lesssim [\rho_0]_\alpha^2 + [\rho_1]_\alpha^2 + [\nabla_{xy}c]_\alpha^2. \quad (2.4.11)$$

STEP 6 (Mapping properties of Q). We show that if $\mathcal{E} + \mathcal{D} + [\nabla_{xy}c]_\alpha^2 \ll \theta^2$,²² then

$$Q^{-1}(\#_\theta) \subseteq \#_{2\theta}. \quad (2.4.12)$$

Indeed, we have

$$\begin{aligned} Q^{-1}(\#_\theta) &= Q^{-1}(B_\theta \times \mathbb{R}^d) \cup Q^{-1}(\mathbb{R}^d \times B_\theta) \\ &= \left(B^{-1}B_\theta \times \mathbb{R}^d \right) \cup \left(\mathbb{R}^d \times (\gamma^{-1}D^{-*}BB_\theta + b) \right), \end{aligned}$$

and from (2.1.50), (2.4.9), and (2.4.11) it follows that $B^{-1}B_\theta \subseteq B_{2\theta}$ and $\gamma^{-1}D^{-*}BB_\theta + b \subseteq B_{2\theta}$ if $\mathcal{E} + \mathcal{D} + [\nabla_{xy}c]_\alpha^2 \ll \theta^2$, which we assume to be true from now on.

STEP 7 (Transformation of the displacement) We next show that for all $(\widehat{x}, \widehat{y}) = Q(x, y) \in \mathbb{R}^d \times \mathbb{R}^d$ and $t \in [0, 1]$ there holds

$$B(\widehat{y} - \widehat{x}) = y - x - \nabla\Phi(tx + (1-t)y) + \delta, \quad (2.4.13)$$

where the error δ is controlled by

$$\begin{aligned} |\delta| &\lesssim (|\gamma - 1| + |D - \mathbb{I}|) (|y| + |b|) + |\nabla^2\Phi(0)|^2|x| + |\nabla^2\Phi(0)||x - y| \\ &\quad + \sup_{s \in [0,1]} |\nabla^3\Phi(s(tx + (1-t)y))| |tx + (1-t)y|^2. \end{aligned} \quad (2.4.14)$$

Indeed, by (2.1.45) we have

$$B(\widehat{y} - \widehat{x}) = \gamma D^*(y - b) - B^2x = y - b - B^2x + (\gamma D^* - \mathbb{I})(y - b),$$

and the second term, which will turn out to be an error term, can be bounded by

$$|(\gamma D^* - \mathbb{I})(y - b)| \leq (|\gamma - 1||D^*| + |D^* - \mathbb{I}|) |y - b| \stackrel{(2.4.9)}{\lesssim} (|\gamma - 1| + |D - \mathbb{I}|) (|y| + |b|). \quad (2.4.15)$$

We show next that $B^2x + b \approx x + \nabla\Phi(tx + (1-t)y)$, with an error that can be controlled. This relies on the fact that,

$$B^2 = \mathcal{E}^A = \mathbb{I} + A + (\mathcal{E}^A - \mathbb{I} - A), \quad (2.4.16)$$

²²Note that we have not yet fixed θ .

where

$$|\mathcal{E}^A - \mathbb{I} - A| \lesssim |A|^2, \quad (2.4.17)$$

and Taylor approximation. Indeed,

$$\begin{aligned} & |\nabla\Phi(tx + (1-t)y) - (\nabla\Phi(0) + \nabla^2\Phi(0)(tx + (1-t)y))| \\ & \lesssim \sup_{s \in [0,1]} |\nabla^3\Phi(s(tx + (1-t)y))| |tx + (1-t)y|^2 \end{aligned} \quad (2.4.18)$$

so that with (2.4.16), (2.4.17), and (2.4.5),

$$\begin{aligned} & |B^2x + b - (x + \nabla\Phi(tx + (1-t)y))| \lesssim |\nabla^2\Phi(0)x + \nabla\Phi(0) - \nabla\Phi(tx + (1-t)y)| + |\nabla^2\Phi(0)|^2|x| \\ & \leq |\nabla^2\Phi(0)(tx + (1-t)y) + \nabla\Phi(0) - \nabla\Phi(tx + (1-t)y)| + |1-t||\nabla^2\Phi(0)||x-y| + |\nabla^2\Phi(0)|^2|x| \\ & \stackrel{(2.4.18)}{\lesssim} \sup_{s \in [0,1]} |\nabla^3\Phi(s(tx + (1-t)y))| |tx + (1-t)y|^2 + |\nabla^2\Phi(0)||x-y| + |\nabla^2\Phi(0)|^2|x|. \end{aligned} \quad (2.4.19)$$

Collecting all the error terms from (2.4.15) and (2.4.19) then yields (2.4.13).

STEP 8 (Proof of the energy estimate (2.1.51)). In this final step we prove the energy estimate (2.1.51).

To this end, let us compute

$$\begin{aligned} \theta^{d+2}\mathcal{E}_\theta(\widehat{\pi}) &= \int_{\#_\theta} |\widehat{x} - \widehat{y}|^2 d\widehat{\pi} \leq |B^{-1}|^2 \int_{\#_\theta} |B(\widehat{x} - \widehat{y})|^2 d\widehat{\pi} = |B^{-1}|^2 \int_0^1 \int_{\#_\theta} |B(\widehat{x} - \widehat{y})|^2 d\widehat{\pi} dt \\ & \stackrel{(2.1.47)\&(2.4.13)}{\lesssim} |\det B||B^{-1}|^2 \int_{Q^{-1}(\#_\theta)} \int_0^1 (|y - x - \nabla\Phi(tx + (1-t)y)|^2 + |\delta|^2) dt d\pi \\ & \stackrel{(2.4.12)\&(2.1.50)}{\lesssim} \int_{\#_{2\theta}} \int_0^1 |y - x - \nabla\Phi(tx + (1-t)y)|^2 dt d\pi + \int_{\#_{2\theta}} \int_0^1 |\delta|^2 dt d\pi. \end{aligned} \quad (2.4.20)$$

For the first term on the right-hand side of (2.4.20), we use the harmonic approximation theorem in its Lagrangian version (Lemma 2.1.14), which is applicable by STEP 4. In particular, we may estimate, assuming that $\theta \leq \frac{1}{2}$,

$$\begin{aligned} \int_{\#_{2\theta}} \int_0^1 |y - x - \nabla\Phi(tx + (1-t)y)|^2 dt d\pi &\leq \int_{\#_1} \int_0^1 |y - x - \nabla\Phi(tx + (1-t)y)|^2 dt d\pi \\ &\leq \int_{B_2 \times [0,1]} \frac{1}{\rho} |j - \rho \nabla\Phi|^2 + \Delta \\ &\leq \left(\tau + \frac{CM}{\tau} \right) \mathcal{E} + C_\tau \mathcal{D} + \Delta, \end{aligned}$$

so that by (2.4.4), (2.1.11), and Young's inequality,

$$\begin{aligned} & \int_{\#_{2\theta}} \int_0^1 |y - x - \nabla\Phi(tx + (1-t)y)|^2 dt d\pi \\ & \lesssim \left(\tau + \frac{M}{\tau} \right) \mathcal{E} + C_\tau ([\rho_0]_\alpha^2 + [\rho_1]_\alpha^2 + [\nabla_{xy}c]_\alpha^2). \end{aligned} \quad (2.4.21)$$

To estimate the error term $\int_{\#_{2\theta}} \int_0^1 |\delta|^2 dt d\pi$ we use the following consequence of the L^∞ -bound (2.4.1): if $\mathcal{E} + [\rho_0]_\alpha^2 + [\rho_1]_\alpha^2 \ll \theta^{d+2}$, which holds in view of (2.1.49), then

$$\#_{2\theta} \cap \text{Spt } \pi \subseteq B_{3\theta} \times B_{3\theta}. \quad (2.4.22)$$

With this in mind, we will bound each term in (2.4.14) separately. Note that

$$\begin{aligned} \int_{\#_{2\theta}} (|x|^2 + |y|^2) d\pi & \stackrel{(2.4.22)}{\leq} \int_{B_{3\theta} \times \mathbb{R}^d} |x|^2 d\pi + \int_{\mathbb{R}^d \times B_{3\theta}} |y|^2 d\pi \\ & = \int_{B_{3\theta}} |x|^2 \rho_0(x) dx + \int_{B_{3\theta}} |y|^2 \rho_1(y) dy \lesssim \theta^{d+2}, \end{aligned}$$

and similarly $\int_{\#_{2\theta}} d\pi \lesssim \theta^d$. Moreover, by (2.4.22), we may estimate for any $(x, y) \in \#_{2\theta} \cap \text{Spt } \pi$

$$\sup_{s \in [0,1]} |\nabla^3 \Phi(s(tx + (1-t)y))| |tx + (1-t)y|^2 \lesssim \theta^2 \sup_{B_{3\theta}} |\nabla^3 \Phi| \lesssim \theta^2 \sup_{B_2} |\nabla^3 \Phi|.$$

Together with the bound (2.1.50) and since we assumed that $\mathcal{E} + [\rho_0]_\alpha^2 + [\rho_1]_\alpha^2 \ll \theta^2$, which implies that $|b|^2 \lesssim \theta^2$, it follows that

$$\begin{aligned} \int_{\#_{2\theta}} \int_0^1 |\delta|^2 dt d\pi & \lesssim \theta^{d+2} (|1 - \gamma|^2 + |D - \mathbb{I}|^2) + \theta^{d+2} |\nabla^2 \Phi(0)|^4 \\ & \quad + |\nabla^2 \Phi(0)|^2 \int_{\#_{2\theta}} |x - y|^2 d\pi + \theta^{d+4} \sup_{B_2} |\nabla^3 \Phi|^2. \end{aligned}$$

Hence with the bounds (2.4.11), (2.4.9), and (2.1.41), together with $\mathcal{E} + [\rho_0]_\alpha^2 + [\rho_1]_\alpha^2 \ll \theta^2$ and $\theta \leq 1$, we obtain

$$\begin{aligned} \int_{\#_{2\theta}} \int_0^1 |\delta|^2 dt d\pi & \lesssim \theta^{d+2} ([\rho_0]_\alpha^2 + [\rho_1]_\alpha^2 + [\nabla_{xy}c]_\alpha^2) + \theta^{d+2} (\mathcal{E} + [\rho_0]_\alpha^2 + [\rho_1]_\alpha^2)^2 \\ & \quad + (\mathcal{E} + [\rho_0]_\alpha^2 + [\rho_1]_\alpha^2) \theta^{d+2} \mathcal{E} + \theta^{d+4} (\mathcal{E} + [\rho_0]_\alpha^2 + [\rho_1]_\alpha^2) \\ & \lesssim \theta^{d+2} ([\rho_0]_\alpha^2 + [\rho_1]_\alpha^2 + [\nabla_{xy}c]_\alpha^2) + \theta^{d+4} \mathcal{E}. \end{aligned} \quad (2.4.23)$$

Inserting the estimates (2.4.21) and (2.4.23) into (2.4.20) yields the existence of a constant $C = C(d, \alpha)$ such that

$$\mathcal{E}_\theta(\widehat{\pi}) \leq C \left(\tau \theta^{-(d+2)} + \frac{M}{\tau \theta^{d+2}} + \theta^2 \right) \mathcal{E} + C_{\tau, \theta} ([\rho_0]_\alpha^2 + [\rho_1]_\alpha^2 + [\nabla_{xy}c]_\alpha^2).$$

We now fix $\theta = \theta(d, \alpha, \beta)$ such that $C\theta^2 \leq \frac{1}{3}\theta^{2\beta}$, which is possible since $\beta < 1$, and then τ small enough so that $C\tau\theta^{-(d+2)} \leq \frac{1}{3}\theta^{2\beta}$. This fixes ε_τ given by Theorem 2.1.13. Finally, we may choose $\mathcal{E} + [\rho_0]_\alpha^2 + [\rho_1]_\alpha^2$ small enough so that

$$C \frac{M}{\tau\theta^{d+2}} \stackrel{(2.4.1)}{=} C \frac{C_\infty(\mathcal{E} + \mathcal{D})^{\frac{1}{d+2}}}{\tau\theta^{d+2}} \leq \frac{1}{3}\theta^{2\beta},$$

yielding (2.1.51).

STEP 9 (Proof of (2.1.52)). Let $(\widehat{x}, \widehat{y}) \in \#_{\frac{8}{9}\theta} \cap \text{Spt } \widehat{\pi}$ and set $(x, y) := \mathcal{Q}^{-1}(\widehat{x}, \widehat{y})$, see Lemma 2.1.15 for the definition of \mathcal{Q} . Then by 2.A.3, there holds $(x, y) \in \text{Spt } \pi$. Similarly to STEP 6, we have that $\mathcal{Q}^{-1}(\#_{\frac{8}{9}\theta}) \subseteq \#_\theta$, provided that $\mathcal{E} + \mathcal{D} + [\nabla_{xy}c]_\alpha^2 \ll \theta^2$, so that

$$(x, y) \in \#_\theta \cap \text{Spt } \pi.$$

Using the L^∞ bound (2.1.16) (notice that $\theta \leq \frac{2}{3}$) and the fact that we assumed $\mathcal{E} + \mathcal{D} \ll \theta^{d+2}$, we obtain

$$|x - y| \leq M = C_\infty(\mathcal{E} + \mathcal{D})^{\frac{1}{d+2}} \leq \theta,$$

so that $(x, y) \in \mathbb{B}_{\theta, 2\theta}$. It thus follows that $(\widehat{x}, \widehat{y}) \in \mathcal{Q}(\mathbb{B}_{\theta, 2\theta})$. Following the proof of STEP 6, we have

$$\mathcal{Q}(\mathbb{B}_{\theta, 2\theta}) \subseteq \mathbb{B}_{2\theta, 3\theta},$$

and conclude that $(\widehat{x}, \widehat{y}) \in \mathbb{B}_{2\theta, 3\theta} \cap \#_{\frac{8}{9}\theta} = \mathbb{B}_{\frac{8}{9}\theta, 3\theta}$. ■

2.4.2 Proof of ε -regularity

To lighten the notation in this subsection, let us set

$$\mathcal{H}_R(\rho_0, \rho_1, c) := R^{2\alpha} \left([\rho_0]_{\alpha, R}^2 + [\rho_1]_{\alpha, R}^2 + [\nabla_{xy}c]_{\alpha, R}^2 \right).$$

We are now in the position to give the proof of our main ε -regularity theorem, which we restate for the reader's convenience:

Theorem 2.1.1. Assume that (C1)–(C4) hold and that $\rho_0(0) = \rho_1(0) = 1$, as well as $\nabla_{xy}c(0, 0) = -\mathbb{I}$. Assume further that 0 is in the interior of $X \times Y$.

Let π be a c -optimal coupling from ρ_0 to ρ_1 . There exists $R_0 = R_0(c) > 0$ such that for all $R \leq R_0$ with

$$\frac{1}{R^{d+2}} \int_{B_{4R} \times \mathbb{R}^d} |y - x|^2 d\pi + \mathcal{H}_{4R}(\rho_0, \rho_1, c) \ll_c 1, \quad (2.4.24)$$

there exists a function $T \in C^{1, \alpha}(B_R)$ such that $\text{Spt } \pi \cap (B_R \times \mathbb{R}^d) \subseteq \text{graph } T$, and the estimate

$$[\nabla T]_{\alpha, R}^2 \lesssim \frac{1}{R^{2\alpha}} \left(\frac{1}{R^{d+2}} \int_{B_{4R} \times \mathbb{R}^d} |y - x|^2 d\pi + \mathcal{H}_{4R}(\rho_0, \rho_1, c) \right) \quad (2.4.25)$$

holds.

Proof of Theorem 2.1.1. To simplify notation, we write \mathcal{E}_R for $\mathcal{E}_R(\pi)$ and \mathcal{H}_R for $\mathcal{H}_R(\rho_0, \rho_1, c)$. Note that since $\rho_0(0) = \rho_1(0) = 1$ and $R^\alpha ([\rho_0]_{\alpha, 4R} + [\rho_1]_{\alpha, 4R})$ is small by (2.4.24), we may assume throughout the proof that $\frac{1}{2} \leq \rho_0, \rho_1 \leq 2$ on B_{4R} .

STEP 1 (Control of the full energy at scale $2R$). We show that under assumption (2.4.24) we can bound

$$\mathcal{E}_{2R} \lesssim \frac{1}{R^{d+2}} \int_{B_{4R} \times \mathbb{R}^d} |y - x|^2 d\pi. \quad (2.4.26)$$

Indeed, from Remarks 2.1.6 and 2.2.2, we know that (2.4.24) implies the L^∞ bound

$$\begin{aligned} (x, y) &\in (B_{3R} \times \mathbb{R}^d) \cap \text{Spt } \pi \\ \Rightarrow |x - y| &\lesssim R \left(\frac{1}{R^{d+2}} \int_{B_{4R} \times \mathbb{R}^d} |y - x|^2 d\pi + \mathcal{D}_{4R}(\rho_0, \rho_1) \right)^{\frac{1}{d+2}}. \end{aligned} \quad (2.4.27)$$

Let us now prove that

$$(\mathbb{R}^d \times B_{2R}) \cap \text{Spt } \pi \subseteq B_{4R} \times B_{2R}, \quad (2.4.28)$$

from which we get

$$\int_{\mathbb{R}^d \times B_{2R}} |y - x|^2 d\pi \leq \int_{B_{4R} \times \mathbb{R}^d} |y - x|^2 d\pi,$$

thus yielding (2.4.26). To this end, assume there exists $(x, y) \in (B_{4R}^c \times B_{2R}) \cap \text{Spt } \pi$. Let also $x' \in [x, y] \cap \partial B_{\frac{5}{2}R}$ and $y' \in \mathbb{R}^d$ such that $(x', y') \in \text{Spt } \pi$, see Figure 2.1. Then by (2.4.27), (2.4.24) and (2.1.11), we have

$$|x' - y'| \ll R. \quad (2.4.29)$$

From (2.2.5), recalling the definition (2.2.3) of δ and the fact that $\delta \ll 1$, we get, upon writing $y - y' = y - x' + x' - y'$,

$$\begin{aligned} 0 &\leq (x - x') \cdot (y - x') + (x - x') \cdot (x' - y') + \delta |x - x'| |y - x'| + \delta |x - x'| |x' - y'| \\ &\leq (\delta - 1) |x - x'| |y - x'| + (1 + \delta) |x - x'| |x' - y'| \\ &\leq |x - x'| \left(-\frac{1}{2} \frac{R}{2} + (1 + \delta) |x' - y'| \right), \end{aligned}$$

which together with (2.4.29) yields a contradiction, proving (2.4.28).

In the following, STEP 2 – STEP 6 are devoted to prove that under the assumption

$$\mathcal{E}_{2R} + \mathcal{H}_{2R} \ll 1, \quad (2.4.30)$$

the following Campanato estimate holds:

$$\sup_{0 < r < \frac{R}{2}} \sup_{x_0 \in B_R} \frac{1}{r^{d+2+2\alpha}} \inf_{A, b} \int_{(B_r(x_0) \cap B_R) \times \mathbb{R}^d} |y - (Ax + b)|^2 d\pi \lesssim \frac{1}{R^{2\alpha}} (\mathcal{E}_{2R} + \mathcal{H}_{2R}). \quad (2.4.31)$$

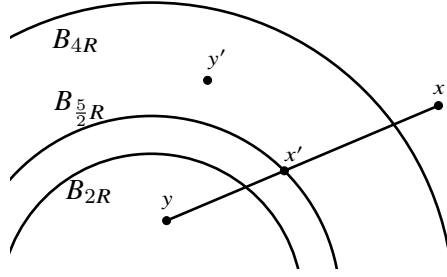


FIGURE 2.1: The definition of x' and y' in the proof of (2.4.28).

STEP 2 (Getting to a normalized setting). We claim that it is enough to prove that if

$$\mathcal{E}_R + \mathcal{H}_R \ll 1 \tag{2.4.32}$$

then for all $r \leq \frac{R}{2}$,

$$\frac{1}{r^{d+2}} \inf_{A,b} \int_{B_r \times \mathbb{R}^d} |y - (Ax + b)|^2 d\pi \lesssim \frac{r^{2\alpha}}{R^{2\alpha}} (\mathcal{E}_R + \mathcal{H}_R). \tag{2.4.33}$$

Let us first notice that for any $x_0 \in B_R$, $(B_r(x_0) \cap B_R) \times \mathbb{R}^d \subseteq \#_{2R}$ so that

$$\begin{aligned} \mathcal{E}_{x_0,R} &:= \frac{1}{R^{d+2}} \int_{(B_r(x_0) \cap B_R) \times \mathbb{R}^d} |y - x|^2 d\pi \lesssim \mathcal{E}_{2R}, \\ \text{and } \mathcal{H}_{x_0,R} &:= R^{2\alpha} \left([\rho_0]_{\alpha, B_R(x_0)}^2 + [\rho_1]_{\alpha, R}^2 + [\nabla_{xy} c]_{\alpha, (B_r(x_0) \cap B_R) \times \mathbb{R}^d}^2 \right) \lesssim \mathcal{H}_{2R}. \end{aligned}$$

Therefore, in view of (2.4.30), it is sufficient to show for all $x_0 \in B_R$ that, if

$$\mathcal{E}_{x_0,R} + \mathcal{H}_{x_0,R} \ll 1, \tag{2.4.34}$$

then for all $r \leq \frac{R}{2}$,

$$\frac{1}{r^{d+2}} \inf_{A,b} \int_{(B_r(x_0) \cap B_R) \times \mathbb{R}^d} |y - (Ax + b)|^2 d\pi \lesssim \frac{r^{2\alpha}}{R^{2\alpha}} (\mathcal{E}_{x_0,R} + \mathcal{H}_{x_0,R}). \tag{2.4.35}$$

Performing a similar change of coordinates as Lemma 2.1.15, namely letting $M := -\nabla_{xy} c(x_0, 0)$, $\gamma := (\rho_0(x_0) \det M)^{\frac{1}{d}}$ and $S(y) := \gamma M^{-1} y$ and defining

$$\tilde{\pi} := \frac{1}{\rho_0(x_0)} (\text{Id} \times S^{-1})_{\#} \pi, \quad \tilde{\rho}_0 := \frac{\rho_0}{\rho_0(x_0)}, \quad \tilde{\rho}_1 := \rho_1 \circ S, \quad \tilde{c}(x, y) := \gamma^{-1} c(x, S(y)),$$

we have that $\tilde{\rho}_0(x_0) = \tilde{\rho}_1(0) = 1$ and $\nabla_{xy} \tilde{c}(x_0, 0) = -\text{I}$. Furthermore, $\tilde{\pi}$ is a \tilde{c} -optimal coupling between $\tilde{\rho}_0$ and $\tilde{\rho}_1$ and $\tilde{\pi}, \tilde{\rho}_0, \tilde{\rho}_1$ and \tilde{c} still satisfy (2.4.34). Without loss of generality, we may thus assume $x_0 = 0$ and then (2.4.34) turns into (2.4.32) and (2.4.35) turns into (2.4.33).

STEP 3 (First step of the iteration). Recall that (2.1.11) holds, i.e., $\mathcal{D}_R \lesssim R^{2\alpha} \left([\rho_0]_{\alpha,R}^2 + [\rho_1]_{\alpha,R}^2 \right)$. By (a rescaling of) Lemma 2.2.1, there exist $\Lambda_0 < \infty$ and $R_0 > 0$ such that for all $R \leq R_0$ for which (2.4.32) holds, we have

$$\#_{\frac{8}{9}R} \cap \text{Spt } \pi \subseteq \mathbb{B}_{\frac{8}{9}R, \frac{\Lambda_0}{9}R}. \quad (2.4.36)$$

Let $\beta > 0$. In view of (2.4.32) and (2.4.36), we may apply Proposition 2.1.16 to get the existence of a symmetric matrix B_1 and a vector b_1 such that, defining $D_1 := -\nabla_{xy}c_1(0, b_1)$, $\gamma_1 := (\rho_1(b_1) \det B_1^2 D_1^{-1})^{\frac{1}{d}}$ and $\rho_{0,1}$, $\rho_{1,1}$, c_1 and π_1 as $\widehat{\rho}_0$, $\widehat{\rho}_1$, \widehat{c} and $\widehat{\pi}$ in Lemma 2.1.15, we get

$$\mathcal{E}^1 := \mathcal{E}_{\theta R}(\pi_1) \leq \theta^{2\beta} \mathcal{E}_R + C_\beta \mathcal{H}_R, \quad (2.4.37)$$

and that π_1 is a c_1 -optimal coupling between $\rho_{0,1}$ and $\rho_{1,1}$. From (2.1.52) we also have the inclusion

$$\#_{\frac{8}{9}\theta R} \cap \text{Spt } \pi_1 \subseteq \mathbb{B}_{\frac{8}{9}\theta R, 3\theta R}, \quad (2.4.38)$$

so that we may fix $\Lambda = 27$ from now on (assuming w.l.o.g. that $\Lambda_0 \geq 27$).

STEP 4 (Iterating Proposition 2.1.16). We now show that we can iterate Proposition 2.1.16 a finite number of times.

Notice that from the estimates (2.1.50), (2.4.9) and (2.4.11) we have the inclusion

$$\gamma_1^{-1} D_1^{-*} B_1 B_{\theta R} + b_1 \subseteq B_R. \quad (2.4.39)$$

Hence, we can bound

$$\begin{aligned} [\rho_{1,1}]_{\alpha, \theta R} &= \rho_1(b_1)^{-1} \sup_{y \neq y' \in B_{\theta R}} \frac{|\rho_1(\gamma_1^{-1} D_1^{-*} B_1 y + b_1) - \rho_1(\gamma_1^{-1} D_1^{-*} B_1 y' + b_1)|}{|y - y'|^\alpha} \\ &\stackrel{(2.4.8)}{\lesssim} |\gamma_1^{-1} D_1^{-*} B_1|^\alpha \sup_{y \neq y' \in B_{\theta R}} \frac{|\rho_1(\gamma_1^{-1} D_1^{-*} B_1 y + b_1) - \rho_1(\gamma_1^{-1} D_1^{-*} B_1 y' + b_1)|}{|(\gamma_1^{-1} D_1^{-*} B_1 y + b_1) - (\gamma_1^{-1} D_1^{-*} B_1 y' + b_1)|^\alpha} \\ &\stackrel{(2.4.39)}{\leq} \gamma_1^{-\alpha} |D_1^{-1}|^\alpha |B_1|^\alpha \sup_{\tilde{y} \neq \tilde{y}' \in B_R} \frac{|\rho_1(\tilde{y}) - \rho_1(\tilde{y}')|}{|\tilde{y} - \tilde{y}'|^\alpha}. \end{aligned}$$

Estimates (2.1.50), (2.4.8), (2.4.9), and (2.4.11) thus yield for $i \in \{0, 1\}$ ($\rho_{0,1}$ is treated similarly)

$$[\rho_{i,1}]_{\alpha, \theta R}^2 \leq \left(1 + C(\mathcal{E}_R^{\frac{1}{2}} + \mathcal{H}_R^{\frac{1}{2}}) \right) [\rho_i]_{\alpha, R}^2,$$

where C is a constant depending only on d and α . Using (2.1.46), the same kind of computation gives

$$[\nabla_{xy}c_1]_{\alpha, \theta R}^2 \leq \left(1 + C(\mathcal{E}_R^{\frac{1}{2}} + \mathcal{H}_R^{\frac{1}{2}}) \right) [\nabla_{xy}c]_{\alpha, R}^2, \quad (2.4.40)$$

so that

$$\mathcal{H}^1 := \mathcal{H}_{\theta R}(\rho_{0,1}, \rho_{1,1}, c_1) \leq \theta^{2\alpha} \left(1 + C(\mathcal{E}_R^{\frac{1}{2}} + \mathcal{H}_R^{\frac{1}{2}})\right) \mathcal{H}_R. \quad (2.4.41)$$

Therefore, (2.4.32), (2.4.37), (2.4.38), and (2.4.41) imply that we may iterate Proposition 2.1.16 K times, $K \geq 2$, to find sequences of matrices B_k and D_k , a sequence of vectors b_k , and a sequence of real numbers γ_k such that, setting for $1 \leq k \leq K$

$$\rho_{0,k}(x) := \rho_{0,k-1}(B_k^{-1}x), \quad \rho_{1,k}(y) := \frac{\rho_{1,k-1}(\gamma_k^{-1}D_k^{-*}B_k y + b_k)}{\rho_{1,k-1}(b_k)},$$

$$Q_k(x, y) := (B_k x, \gamma_k B_k^{-1} D_k^* (y - b_k)), \quad \pi_k := \det B_k Q_k \# \pi_{k-1}, \quad (2.4.42)$$

$$\text{and } c_k(x, y) := \gamma_k c_{k-1}(B_k^{-1}x, \gamma_k^{-1}D_k^{-*}B_k y + b_k), \quad (2.4.43)$$

we recover $\rho_{0,k}(0) = \rho_{1,k}(0) = 1$ and $\nabla_{xy}c_k(0, 0) = -\mathbb{I}$, π_k is a c_k -optimal coupling between $\rho_{0,k}$ and $\rho_{1,k}$, and from (2.1.52) we have

$$\#_{\frac{8}{9}\theta^k R} \cap \text{Spt } \pi_k \subseteq \mathbb{B}_{\frac{8}{9}\theta^k R, 3\theta^k R}. \quad (2.4.44)$$

Moreover, defining

$$\mathcal{E}^k := \mathcal{E}_{\theta^k R}(\pi_k) \quad \text{and} \quad \mathcal{H}^k := \mathcal{H}_{\theta^k R}(\rho_{0,k}, \rho_{1,k}, c_k), \quad (2.4.45)$$

we have

$$\mathcal{E}^k \stackrel{(2.1.51)}{\leq} \theta^{2\beta} \mathcal{E}^{k-1} + C_\beta \mathcal{H}^{k-1}, \quad (2.4.46)$$

$$|B_k - \mathbb{I}|^2 + \frac{1}{(\theta^{k-1}R)^2} |b_k|^2 \stackrel{(2.1.50)}{\lesssim} \mathcal{E}^{k-1} + \theta^{2(k-1)\alpha} R^{2\alpha} \left([\rho_{0,k-1}]_{\alpha, \theta^{k-1}R}^2 + [\rho_{1,k-1}]_{\alpha, \theta^{k-1}R}^2 \right), \quad (2.4.47)$$

$$|D_k - \mathbb{I}|^2 \stackrel{(2.4.9)}{\lesssim} \theta^{2(k-1)\alpha} R^{2\alpha} [\nabla_{xy}c_{k-1}]_{\alpha, \theta^{k-1}R}^2, \quad (2.4.48)$$

$$|\gamma_k - 1|^2 \stackrel{(2.4.11)}{\lesssim} \mathcal{H}^{k-1}. \quad (2.4.49)$$

Arguing as for (2.4.41), there exists a constant $C_1 = C_1(d, \alpha) < \infty$ such that

$$\mathcal{H}^k \leq \theta^{2\alpha} \left(1 + C_1 \left((\mathcal{E}^{k-1})^{\frac{1}{2}} + (\mathcal{H}^{k-1})^{\frac{1}{2}} \right)\right) \mathcal{H}^{k-1}. \quad (2.4.50)$$

STEP 5 (Smallness at any step of the iteration). From now on, we fix $\beta > \alpha$. Let us show by induction that there exists a constant $C_2 = C_2(d, \alpha, \beta) < \infty$ such that for all $K \in \mathbb{N}$

$$\mathcal{E}^K \leq C_2 \theta^{2K\alpha} (\mathcal{E}_R + \mathcal{H}_R), \quad (2.4.51)$$

$$\mathcal{H}^K \leq \theta^{2\alpha} (1 + \theta^{K\alpha}) \mathcal{H}^{K-1}. \quad (2.4.52)$$

This will show, together with (2.4.44), that we can keep on iterating Proposition 2.1.16. As an outcome of this induction, the estimate

$$\mathcal{H}^K \lesssim C_3 \theta^{2K\alpha} \mathcal{H}_R \quad (2.4.53)$$

holds for all $K \in \mathbb{N}$ with $C_3 := \prod_{k=1}^{+\infty} (1 + \theta^{k\alpha}) < \infty$.

We set

$$C_2 := \max \left\{ \theta^{2(\beta-\alpha)}, \frac{C_3 C_\beta \theta^{-2\alpha}}{1 - \theta^{2(\beta-\alpha)}} \right\}. \quad (2.4.54)$$

By (2.4.37) and (2.4.41), the estimates (2.4.51) and (2.4.52) hold for $K = 1$ provided $\mathcal{E}_R + \mathcal{H}_R$ is small enough, since $C_2 \geq \max\{\theta^{2(\beta-\alpha)}, C_\beta \theta^{-2\alpha}\}$. Assume now that (2.4.51) and (2.4.52) hold for all $1 \leq k \leq K - 1$. By induction hypothesis, we have

$$\mathcal{E}^{K-1} \leq C_2 \theta^{2(K-1)\alpha} (\mathcal{E}_R + \mathcal{H}_R) \quad \text{and} \quad \mathcal{H}^{K-1} \leq C_3 \theta^{2(K-1)\alpha} \mathcal{H}_R.$$

Combining these two estimates with (2.4.50) for $k = K$ and assuming that

$$C_1 \left(C_2^{\frac{1}{2}} (\mathcal{E}_R^{\frac{1}{2}} + \mathcal{H}_R^{\frac{1}{2}}) + C_3^{\frac{1}{2}} \mathcal{H}_R^{\frac{1}{2}} \right) \leq \theta^\alpha,$$

which is possible provided $\mathcal{E}_R + \mathcal{H}_R$ is small enough, we get (2.4.52). Similarly, by (2.4.46) with $k = K$ we obtain, using that $C_2 \geq \frac{C_3 C_\beta \theta^{-2\alpha}}{1 - \theta^{2(\beta-\alpha)}}$ by (2.4.54), the bound

$$\theta^{-2K\alpha} \mathcal{E}^K \leq \left(\theta^{2(\beta-\alpha)} C_2 + C_\beta C_3 \theta^{-2\alpha} \right) (\mathcal{E}_R + \mathcal{H}_R) \leq C_2 (\mathcal{E}_R + \mathcal{H}_R).$$

This concludes the induction.

STEP 6 (Campanato estimate). We can now prove the main estimate, that is, assuming (2.4.32), we show that (2.4.33) holds.

Let

$$\bar{\gamma}_k := \gamma_k \dots \gamma_1, \quad \bar{B}_k := B_k \dots B_1, \quad \bar{D}_k := B_k^{-1} D_k^* \dots B_1^{-1} D_1^*, \quad \bar{b}_k := \sum_{i=1}^k (\gamma_k B_k^{-1} D_k^*) \dots (\gamma_i B_i^{-1} D_i^*) b_i$$

and

$$\bar{Q}_k(x, y) := (\bar{B}_k x, \bar{\gamma}_k \bar{D}_k y - \bar{b}_k). \quad (2.4.55)$$

We see that, recalling (2.4.42) and noticing that $\bar{Q}_k = Q_k \circ \dots \circ Q_1$,

$$\pi_k = \det \bar{B}_k \bar{Q}_k \# \pi \quad \text{and} \quad \rho_{0,k}(x) = \rho_0(\bar{B}_k^{-1} x). \quad (2.4.56)$$

Moreover, from (2.4.47), (2.4.51), and (2.4.52), we have the estimate

$$|\bar{B}_k - \mathbb{I}|^2 \lesssim \mathcal{E}_R + \mathcal{H}_R \ll 1, \quad (2.4.57)$$

so that

$$B_{\frac{1}{2}\theta^k R} \times \mathbb{R}^d \subseteq \bar{B}_k^{-1} (B_{\theta^k R}) \times \mathbb{R}^d = \bar{Q}_k^{-1} (B_{\theta^k R} \times \mathbb{R}^d). \quad (2.4.58)$$

Similarly, from (2.4.47), (2.4.48) and (2.4.49),

$$|\bar{\gamma}_k - 1|^2 + |\bar{D}_k - \mathbb{I}|^2 \lesssim \mathcal{E}_R + \mathcal{H}_R \ll 1. \quad (2.4.59)$$

Let us now compute

$$\begin{aligned} & \frac{1}{(\frac{1}{2}\theta^k R)^{d+2}} \inf_{A,b} \int_{B_{\frac{1}{2}\theta^k R} \times \mathbb{R}^d} |y - (Ax + b)|^2 d\pi \\ & \stackrel{(2.4.58)}{\lesssim} \frac{1}{(\theta^k R)^{d+2}} \int_{\bar{Q}_k^{-1}(B_{\theta^k R} \times \mathbb{R}^d)} |y - \bar{\gamma}_k^{-1} \bar{D}_k^{-1} \bar{B}_k x - \bar{\gamma}_k^{-1} \bar{D}_k^{-1} \bar{b}_k|^2 d\pi \\ & \stackrel{(2.4.56)}{=} \frac{(\det \bar{B}_k)^{-1}}{(\theta^k R)^{d+2}} \int_{B_{\theta^k R} \times \mathbb{R}^d} |\bar{\gamma}_k^{-1} \bar{D}_k^{-1} (y - x)|^2 d\pi_k \\ & \stackrel{(2.4.57) \& (2.4.59)}{\lesssim} \frac{1}{(\theta^k R)^{d+2}} \int_{B_{\theta^k R} \times \mathbb{R}^d} |y - x|^2 d\pi_k \stackrel{(2.4.45)}{=} \mathcal{E}^k. \end{aligned}$$

By (2.4.51), we obtain

$$\frac{1}{(\frac{1}{2}\theta^k R)^{d+2}} \inf_{A,b} \int_{B_{\frac{1}{2}\theta^k R} \times \mathbb{R}^d} |y - (Ax + b)|^2 d\pi \lesssim \theta^{2k\alpha} (\mathcal{E}_R + \mathcal{H}_R),$$

from which (2.4.33) follows, concluding the proof of (2.4.31)

STEP 7 (Spt π is contained in the graph of a function T within $B_R \times \mathbb{R}^d$). We claim that (2.4.31) implies the existence of a function $T : B_R \rightarrow Y$ such that

$$(B_R \times \mathbb{R}^d) \cap \text{Spt } \pi \subseteq \text{graph } T. \quad (2.4.60)$$

In the following, we abbreviate

$$[[\pi]]_\alpha^2 := \sup_{0 < r < \frac{R}{2}} \sup_{x_0 \in B_R} \frac{1}{r^{d+2+2\alpha}} \inf_{A,b} \int_{(B_r(x_0) \cap B_R) \times \mathbb{R}^d} |y - (Ax + b)|^2 d\pi.$$

To prove the claim, fix $x_0 \in B_R$ and notice that (2.4.31) implies that for any $r > 0$ small enough, there holds

$$\frac{1}{r^{d+2}} \inf_{A,b} \int_{B_r(x_0) \times \mathbb{R}^d} |y - (Ax + b)|^2 d\pi \lesssim \frac{r^{2\alpha}}{R^{2\alpha}} \left(\frac{1}{R^{d+2}} \int_{B_{4R} \times \mathbb{R}^d} |x - y|^2 d\pi + \mathcal{H}_{4R} \right). \quad (2.4.61)$$

STEP 7.A. It is easy to see that the infimum in (2.4.61) is attained at some $A_r = A_r(x_0)$ and $b_r = b_r(x_0)$. Analogous to [Cam64, Lemma 3.IV] one can show that there exist a matrix $A_0 = A_0(x_0)$ and a vector $b_0 = b_0(x_0)$ such that $A_r \rightarrow A_0$ and $b_r \rightarrow b_0$ as $r \rightarrow 0$ (uniformly in x_0) with rates

$$|A_r - A_0| \lesssim [[\pi]]_\alpha r^\alpha \quad \text{and} \quad |b_r - b_0| \lesssim [[\pi]]_\alpha r^{\alpha+1}. \quad (2.4.62)$$

We refer the reader to Appendix 2.C for a proof of the convergences and (2.4.62).

STEP 7.B. We claim that

$$\frac{1}{r^d} \int_{B_r(x_0) \times \mathbb{R}^d} |y - (A_0 x_0 + b_0)|^2 d\pi \rightarrow 0 \quad \text{as } r \searrow 0. \quad (2.4.63)$$

Indeed, we can split

$$\begin{aligned} \frac{1}{r^d} \int_{B_r(x_0) \times \mathbb{R}^d} |y - (A_0 x_0 + b_0)|^2 d\pi &\lesssim \frac{1}{r^d} \int_{B_r(x_0) \times \mathbb{R}^d} |y - (A_r x + b_r)|^2 d\pi \\ &\quad + \frac{1}{r^d} \int_{B_r(x_0) \times \mathbb{R}^d} |(A_r - A_0)x + (b_r - b_0)|^2 d\pi \\ &\quad + \frac{1}{r^d} \int_{B_r(x_0) \times \mathbb{R}^d} |A_0(x - x_0)|^2 d\pi. \end{aligned} \quad (2.4.64)$$

By definition of A_r, b_r , we have

$$\frac{1}{r^d} \int_{B_r(x_0) \times \mathbb{R}^d} |y - (A_r x + b_r)|^2 d\pi = \frac{1}{r^d} \inf_{A,b} \int_{B_r(x_0) \times \mathbb{R}^d} |y - (Ax + b)|^2 d\pi \lesssim [[\pi]]_\alpha^2 r^{2\alpha+2}.$$

Using (2.4.62), $\rho_0 \leq 2$, and $x_0 \in B_R$, it follows that

$$\begin{aligned} \frac{1}{r^d} \int_{B_r(x_0) \times \mathbb{R}^d} |(A_r - A_0)x + (b_r - b_0)|^2 d\pi &= \frac{1}{r^d} \int_{B_r(x_0)} |(A_r - A_0)x + (b_r - b_0)|^2 \rho_0(x) dx \\ &\lesssim |A_r - A_0|^2 R^2 + |b_r - b_0|^2 \\ &\lesssim [[\pi]]_\alpha^2 r^{2\alpha} (R^2 + r^2). \end{aligned}$$

Finally, the last term in (2.4.64) is estimated by

$$\frac{1}{r^d} \int_{B_r(x_0) \times \mathbb{R}^d} |A_0(x - x_0)|^2 d\pi = \frac{1}{r^d} \int_{B_r(x_0)} |A_0(x - x_0)|^2 \rho_0(x) dx \lesssim r^2.$$

Letting $r \rightarrow 0$ in the above estimates proves the claim (2.4.63).

STEP 7.C. By disintegration, there exists a family of measures $\{\pi_x\}_{x \in X}$ on Y such that

$$\frac{1}{r^d} \int_{B_r(x_0) \times \mathbb{R}^d} |y - (A_0 x_0 + b_0)|^2 d\pi = \frac{1}{r^d} \int_{B_r(x_0)} \int |y - (A_0 x_0 + b_0)|^2 \pi_x(dy) \rho_0(x) dx. \quad (2.4.65)$$

Since the left-hand side of (2.4.65) tends to zero as $r \rightarrow 0$ by Step 7.B, it follows that if x_0 is a Lebesgue point, we must have

$$\int |y - (A_0 x_0 + b_0)|^2 \pi_{x_0}(dy) = 0,$$

hence $\pi_{x_0} = \delta_{A_0 x_0 + b_0}$.

STEP 7.D. For any Lebesgue point $x_0 \in B_R$, define $T(x_0) := A_0(x_0)x_0 + b_0(x_0)$. Then the previous Step 7.C shows that

$$\pi|_{B_R \times \mathbb{R}^d} = (\text{Id} \times T)_\# \rho_0,$$

that is (2.4.60).

STEP 8 By boundedness of ρ_0 , (2.4.31) implies the bound

$$\begin{aligned} & \sup_{0 < r < \frac{R}{2}} \sup_{x_0 \in B_R} \frac{1}{r^{d+2+2\alpha}} \inf_{A,b} \int_{B_r(x_0) \cap B_R} |T(x) - (Ax + b)|^2 dx \\ & \lesssim \frac{1}{R^{2\alpha}} \left(\frac{1}{R^{d+2}} \int_{B_{4R} \times \mathbb{R}^d} |x - y|^2 d\pi + \mathcal{H}_{4R} \right), \end{aligned}$$

which by means of Campanato's theory [Cam64] proves that $T \in C^{1,\alpha}(B_R)$ and that the Hölder seminorm of ∇T satisfies (2.4.25). \blacksquare

Remark 2.4.1. The deterministic structure of the c -optimal coupling, that is, the existence of T such that $\pi = (\text{Id} \times T)_\# \rho_0$, is a classical result in optimal transportation. If we had used this result, the proof would have become shorter, as STEP 7 would not have been needed.

Before we give the proof of Corollary 2.1.3, let us remark that one can show the following variant of our qualitative L^∞ bound on the displacement (Lemma 2.2.1):

Lemma 2.4.2. *Assume that the cost function c satisfies (C1)–(C4) and that $\nabla_x c(0, 0) = 0$. Let u be a c -convex function. There exist $\Lambda_0 < \infty$ and $R'_0 > 0$ such that for all $R \leq R'_0$ for which $\frac{1}{R^2} \| |u - \frac{1}{2}| \cdot |\cdot|^2 \|_{C^0(B_{8R})} \leq 1$ we have*

$$\text{esssup}_{x \in B_{4R}} |c\text{-exp}_x(\nabla u(x))| \leq \Lambda_0 R. \quad (2.4.66)$$

Proof. Since u is c -convex, it is differentiable a.e. For any $x \in B_{4R}$ such that $\nabla u(x)$ exists, let $y = c\text{-exp}_x(\nabla u(x))$, that is,

$$\nabla u(x) + \nabla_x c(x, y) = 0. \quad (2.4.67)$$

Let \tilde{c} be defined as in (2.2.16). Then, using $\nabla_x c(0, 0) = 0$, we have

$$\begin{aligned} -\nabla_x \tilde{c}(x, y) &= -\nabla_x c(x, y) + \nabla_x c(x, 0) \stackrel{(2.4.67)}{=} \nabla u(x) + \nabla_x c(x, 0) - \nabla_x c(0, 0) \\ &= \nabla u(x) - x + x + \int_0^1 \nabla_{xx} c(tx, 0) dt x, \end{aligned}$$

so that

$$|\nabla_x \tilde{c}(x, y)| \leq |\nabla u(x) - x| + |x| + \|\nabla_{xx} c\|_{C^0(X \times Y)} |x|.$$

Being c -convex, the function u , and therefore also the function $x \mapsto u(x) - \frac{1}{2}|x|^2$, is semi-convex, which implies that

$$\operatorname{ess\,sup}_{x \in B_{4R}} |\nabla u(x) - x| \lesssim \frac{1}{R} \sup_{x \in B_{8R}} |u - \frac{1}{2}|x|^2|, \quad (2.4.68)$$

see Lemma 2.A.6 in the appendix. By the closeness assumption on u and **(C1)** we may therefore bound

$$|\nabla_x \tilde{c}(x, y)| \leq \lambda R.$$

Steps 2 and 3 of the proof of Lemma 2.2.1 then imply that there exist $\Lambda_0 < \infty$ and $R'_0 > 0$ (depending on c only through assumptions **(C1)**–**(C4)**) such that for all $R \leq R'_0$ we have

$$|\nabla_x \tilde{c}(x, y)| \leq \lambda R \quad \Rightarrow \quad |y| \leq \Lambda_0 R,$$

that is, (2.4.66) holds. ■

Proof of Corollary 2.1.3. By Lemma 2.4.2 there exist $\Lambda_0 < \infty$ and $R'_0 > 0$, depending only on the qualitative assumptions **(C1)**–**(C4)** on c such that for all $R \leq R'_0$ for which (2.1.6) holds, we have

$$\|T_u\|_{L^\infty(B_{4R})} \leq \Lambda_0 R. \quad (2.4.69)$$

We claim that

$$\frac{1}{R} \|x - T_u\|_{L^\infty(B_{4R})} \lesssim \frac{1}{R^2} \|u - \frac{1}{2}|\cdot|^2\|_{C^0(B_{8R})} + R^\alpha ([\nabla_{xy}c]_{\alpha, 4R} + [\nabla_{xx}c]_{\alpha, 4R}), \quad (2.4.70)$$

which immediately implies that

$$\begin{aligned} \frac{1}{R^{d+2}} \int_{B_{4R} \times \mathbb{R}^d} |x - y|^2 d\pi &= \frac{1}{R^{d+2}} \int_{B_{4R}} |x - T_u(x)|^2 \rho_0(x) dx \\ &\lesssim \frac{1}{R^4} \|u - \frac{1}{2}|\cdot|^2\|_{C^0(B_{8R})}^2 + R^{2\alpha} \left([\nabla_{xy}c]_{\alpha, 4R}^2 + [\nabla_{xx}c]_{\alpha, 4R}^2 \right) \ll 1. \end{aligned} \quad (2.4.71)$$

In particular, it follows by Theorem 2.1.1, that there exists a potentially smaller scale $R_0 \leq R'_0$ such that for all $R \leq R_0$ for which (2.1.6) holds, we have that $T_u \in C^{1,\alpha}(B_R)$ and ∇T_u satisfies the bound (2.1.5). Applying (2.4.71) once more, we see that (2.1.7) holds.

To prove the claim (2.4.70), we appeal to semi-convexity of the c -convex function u (which implies semi-convexity of the function $x \mapsto u(x) - \frac{1}{2}|x|^2$), in particular Lemma 2.A.6, to bound

$$\begin{aligned} \|x - T_u\|_{L^\infty(B_{4R})} &\leq \|x - \nabla u\|_{L^\infty(B_{4R})} + \|\nabla u - T_u\|_{L^\infty(B_{4R})} \\ &\lesssim \frac{1}{R} \|u - \frac{1}{2}|\cdot|^2\|_{C^0(B_{8R})} + \|\nabla u - T_u\|_{L^\infty(B_{4R})}. \end{aligned}$$

It remains to estimate the latter term. To this end, notice that for a.e. $x \in B_{4R}$ we have $\nabla u(x) = -\nabla_x c(x, T_u(x))$, so that with the normalization assumption $\nabla_x c(0, 0) = 0$ we may bound

$$\begin{aligned} |\nabla u(x) - T_u(x)| &= |\nabla_x c(x, T_u(x)) + T_u(x)| \\ &\leq |\nabla_x c(x, T_u(x)) - \nabla_x c(x, 0) + T_u(x)| + |\nabla_x c(x, 0) - \nabla_x c(0, 0)| \\ &\leq \int_0^1 |(\nabla_{xy} c(x, sT_u(x)) + \mathbb{I})T_u(x)| ds + \int_0^1 |\nabla_{xx} c(tx, 0)x| dt. \end{aligned}$$

It now follows with (2.4.69), definition (2.1.3), and $\nabla_{xx} c(0, 0) = 0$, $\nabla_{xy} c(0, 0) = -\mathbb{I}$, that

$$\begin{aligned} |\nabla u(x) - T_u(x)| &\leq [\nabla_{xy} c]_{\alpha, 4R} (|x|^\alpha + |T_u(x)|^\alpha) |T_u(x)| + [\nabla_{xx} c]_{\alpha, 4R} |x|^{\alpha+1} \\ &\stackrel{(2.4.69)}{\leq} C_{\Lambda_0} R^\alpha ([\nabla_{xy} c]_{\alpha, 4R} + [\nabla_{xx} c]_{\alpha, 4R}) R. \end{aligned} \quad (2.4.72)$$

In view of (2.1.6), we may assume

$$C_{\Lambda_0} R^\alpha ([\nabla_{xy} c]_{\alpha, 4R} + [\nabla_{xx} c]_{\alpha, 4R}) \leq 1,$$

so that

$$|\nabla u(x) - T_u(x)| \leq R. \quad (2.4.73)$$

Using that by Lemma 2.A.6 and the smallness assumption (2.1.6), we have

$$|\nabla u(x) - x| \lesssim \frac{1}{R} \|u - \frac{1}{2}|\cdot|^2\|_{C^0(B_{8R})} \lesssim R,$$

and writing $T_u(x) = (T_u(x) - \nabla u(x)) + (\nabla u(x) - x) + x$, the estimate (2.4.72) turns into

$$\begin{aligned} |\nabla u(x) - T_u(x)| &\lesssim [\nabla_{xy} c]_{\alpha, 4R} (R^\alpha + |\nabla u(x) - T_u(x)|^\alpha) (|\nabla u(x) - T_u(x)| + R) + R^{1+\alpha} [\nabla_{xx} c]_{\alpha, 4R} \\ &\stackrel{(2.4.73)}{\lesssim} R^{1+\alpha} ([\nabla_{xy} c]_{\alpha, 4R} + [\nabla_{xx} c]_{\alpha, 4R}). \end{aligned}$$

This proves the claimed inequality (2.4.70). ■

2.4.3 ε -regularity for continuous densities

In this subsection only, we modify the assumptions on the densities: ρ_0 and ρ_1 are only supposed to be continuous on X and Y respectively. As in [Gol20] we still derive Hölder regularity (but only $C^{0,\beta}$) for the optimal map in a small ball, assuming that the lagrangian energy, the L^∞ distance between 1 and the densities and the semi-norm of $\nabla_{xy} c$ are small enough at a larger scale.

Theorem 2.4.3. *Assume that (C1)–(C4) hold, and that $\nabla_{xy} c(0, 0) = -\mathbb{I}$.*

Let π be the c -optimal coupling between ρ_0 and ρ_1 . There exists $R_0 > 0$ such that for all $R \leq R_0$ with

$$\mathcal{E}_{2R}(\pi) + \|\rho_0 - 1\|_{L^\infty(B_R)}^2 + \|\rho_1 - 1\|_{L^\infty(B_R)}^2 + [\nabla_{xy} c]_{\alpha, 2R}^2 \ll 1, \quad (2.4.74)$$

then there exists a map T such that $\text{Spt } \pi \cap (B_R \times \mathbb{R}^d) \subset \text{graph } T$ and for any $\beta \in (0, 1)$, we have $T \in C^{0,\beta}(B_R)$.

Notice that we do not give an estimate on the semi-norm of ∇T , in fact we can only say that it is bounded. Indeed, the estimate (2.1.5) comes from the data getting geometrically smaller at each step of the iteration. Here the data concerning the densities do not improve as we zoom in, we only have

$$\|\widehat{\rho}_i - 1\|_{L^\infty(B_{\theta R})}^2 \leq \|\rho_i - 1\|_{L^\infty(B_R)}^2.$$

A second point to notice is that we still need to assume Hölder regularity of the second derivative of the cost, as c being C^2 would not be enough. This was already the case in [DF15]. In our iteration, we see clearly why it is needed: after the change of variables that we give in Proposition 2.4.4 below, the new cost is such that

$$\nabla_{\widehat{x}\widehat{y}}\widehat{c}(\widehat{x}, \widehat{y}) = B\nabla_{xy}c(B\widehat{x}, \gamma^{-1}D^{-1}B^{-1}\widehat{y} + b)D^{-1}B^{-1}.$$

From there, the best L^∞ estimate we can hope for is

$$\|\nabla_{\widehat{x}\widehat{y}}\widehat{c} + \mathbb{I}\|_{L^\infty(\#_{\theta R})} \leq |B||B^{-1}||D^{-1}| \|\nabla_{xy}c + \mathbb{I}\|_{L^\infty(\#_R)}.$$

From estimates (2.4.78) and (2.4.82), we get

$$\|\nabla_{\widehat{x}\widehat{y}}\widehat{c} + \mathbb{I}\|_{L^\infty(\#_{\theta R})} \leq \left(1 + C\varepsilon^{\frac{1}{2}}\right) \|\nabla_{xy}c + \mathbb{I}\|_{L^\infty(\#_R)},$$

where ε is the left-hand side of (2.4.74). Now, the issue is that the norms of the densities will not improve as we iterate this result, so that ε will stay at best of the same order, which causes the L^∞ norm of the iterated cost to grow geometrically.

The proof is very close to the one in the previous subsection and to [Gol20]. The idea is to iterate a one-step improvement result that we state but do not prove, as it is the same as Proposition 2.1.16 up to Lemma 2.A.4 replaced by the estimate

$$\mathcal{D}_R(\rho_0, \rho_1) \lesssim \|\rho_0 - 1\|_{L^\infty(B_R)}^2 + \|\rho_1 - 1\|_{L^\infty(B_R)}^2 \quad \text{if } \frac{1}{2} \leq \rho_{0,1} \leq 2 \text{ in } B_R, \quad (2.4.75)$$

that is proved exactly the same way. Notice also that this change of variables does not need to normalize the densities to be 1 at 0, because we keep track of closeness to 1 with the L^∞ norms.

Proposition 2.4.4. *Assume that $\pi \in \Pi(\rho_0, \rho_1)$ is c -optimal.*

For all $\beta \in (0, 1)$, there exist $\theta \in (0, 1)$ and $C_\beta > 0$ such that for all $\Lambda < \infty$ and $R > 0$ for which

$$\#_{5R} \cap \text{Spt } \pi \subseteq \mathbb{B}_{5R, \Lambda R} \quad (2.4.76)$$

and

$$\mathcal{E}_{6R}(\pi) + \|\rho_0 - 1\|_{L^\infty(B_{6R})}^2 + \|\rho_1 - 1\|_{L^\infty(B_{6R})}^2 + [\nabla_{xy}c]_{\alpha, 6R}^2 \ll_\Lambda 1, \quad (2.4.77)$$

there exist a symmetric matrix $B \in \mathbb{R}^{d \times d}$ with $\det B = 1$ and a vector $b \in \mathbb{R}^d$ with

$$|B - \mathbb{I}|^2 + \frac{1}{R^2}|b|^2 \lesssim \mathcal{E}_{6R}(\pi) + \|\rho_0 - 1\|_{L^\infty(B_{6R})}^2 + \|\rho_1 - 1\|_{L^\infty(B_{6R})}^2, \quad (2.4.78)$$

such that the following holds: Let

$$\begin{pmatrix} \widehat{x} \\ \widehat{y} \end{pmatrix} = \begin{pmatrix} B^{-1}x \\ \gamma BD(y - b) \end{pmatrix} =: Q(x, y),$$

where $D := -\nabla_{xy}c(0, b)$ and set

$$\begin{aligned} \widehat{\rho}_0(\widehat{x}) &= \rho_0(x), \quad \widehat{\rho}_1(\widehat{y}) = \rho_1(y), \\ \widehat{c}(\widehat{x}, \widehat{y}) &= \gamma (c(x, y) - c(x, b) - c(0, y) + c(0, b)), \end{aligned}$$

where $\gamma = (\det D)^{-\frac{1}{d}}$.²³ Then the coupling

$$\widehat{\pi} := Q_{\#}\pi \tag{2.4.79}$$

is an optimal coupling between the measures $\widehat{\mu}(d\widehat{x}) = \widehat{\rho}_0(\widehat{x}) d\widehat{x}$ and $\widehat{\nu}(d\widehat{y}) = \widehat{\rho}_1(\widehat{y}) d\widehat{y}$ with respect to the cost function \widehat{c} , and the energy estimate

$$\mathcal{E}_{\theta R}(\widehat{\pi}) \leq \theta^{2\beta} \mathcal{E}_{6R}(\pi) + C_{\beta} \left(\|\rho_0 - 1\|_{L^{\infty}(B_{6R})}^2 + \|\rho_1 - 1\|_{L^{\infty}(B_{6R})}^2 + [\nabla_{xy}c]_{\alpha, 6R}^2 \right) \tag{2.4.80}$$

holds. Moreover, we have the inclusion

$$\#_{5\theta R} \cap \text{Spt } \widehat{\pi} \subseteq \mathbb{B}_{5\theta R, 8\theta R}. \tag{2.4.81}$$

From this construction, we get as before the estimate

$$|D - \mathbb{I}| = |\nabla_{xy}c(0, b) - \nabla_{xy}c(0, 0)| \lesssim [\nabla_{xy}c]_{\alpha}, \tag{2.4.82}$$

so that in particular $|1 - \det D|^2 \lesssim [\nabla_{xy}c]_{\alpha}$, hence

$$|1 - \gamma|^2 \lesssim [\nabla_{xy}c]_{\alpha}. \tag{2.4.83}$$

We are now ready to give the proof of Theorem 2.4.3. To lighten the notations, let us set

$$\mathcal{E}_R := \mathcal{E}_R(\pi_T), \quad \mathcal{L}_R := \|\rho_0 - 1\|_{L^{\infty}(B_R)}^2 + \|\rho_1 - 1\|_{L^{\infty}(B_R)}^2, \quad \mathcal{H}_R := R^{2\alpha} [\nabla_{xy}c]_{\alpha, R}^2.$$

Proof of Theorem 2.4.3. The first goal of the proof is to show that under the assumption (2.4.74), the following Campanato-type estimate holds for any $\beta \in (0, 1)$:

$$\sup_{0 < r < 2R} \sup_{x_0 \in B_R} \frac{1}{r^{d+2\beta}} \inf_b \int_{(B_R \cap B_r(x_0)) \times \mathbb{R}^d} |y - b|^2 d\pi \lesssim R^{2-2\beta}. \tag{2.4.84}$$

This will be done in STEP 1–STEP 5.

STEP 1 (Getting to a normalized setting). Let $\beta \in (0, 1)$. We claim that it is enough to prove that if

$$\mathcal{E}_R + \mathcal{L}_R + \mathcal{H}_R \ll 1, \tag{2.4.85}$$

²³Note that M is non-singular by assumption (C4).

then for all $r \leq \frac{R}{2}$,

$$\frac{1}{r^d} \inf_b \int_{B_r \times \mathbb{R}^d} |y - b|^2 d\pi \lesssim r^{2\beta} R^{2-2\beta}. \quad (2.4.86)$$

Indeed, notice that for any $x_0 \in B_R$,

$$\mathcal{E}_{x_0, R} := \frac{1}{R^{d+2}} \int_{\#_{x_0, R}} |y - x|^2 d\pi \lesssim \mathcal{E}_{2R},$$

$$\mathcal{L}_{x_0, R} := \|\rho_0 - 1\|_{L^\infty(B_R(x_0))}^2 + \|\rho_1 - 1\|_{L^\infty(B_R)}^2 \lesssim \mathcal{L}_{2R} \quad \text{and} \quad \mathcal{H}_{x_0, R} := R^{2\alpha} [\nabla_{xy} c]_{\alpha, \#_{x_0, R}}^2 \lesssim \mathcal{H}_{2R},$$

where $\#_{x_0, R} := (B_R(x_0) \times \mathbb{R}^d) \cup (\mathbb{R}^d \times B_R)$. Therefore, it is sufficient to show for all $x_0 \in B_R$ that, if

$$\mathcal{E}_{x_0, R} + \mathcal{L}_{x_0, R} + \mathcal{H}_{x_0, R} \ll 1, \quad (2.4.87)$$

then for all $r \leq \frac{R}{2}$,

$$\frac{1}{r^d} \inf_b \int_{B_r(x_0)} |y - b|^2 dx \lesssim r^{2\beta} R^{2-2\beta}. \quad (2.4.88)$$

Letting $D := -\nabla_{xy} c(x_0, 0)$, $\gamma := (\det D)^{1/d}$ and $S(y) := \gamma D^{-1}y$ and defining

$$\tilde{\rho}_0 := \rho_0, \quad \tilde{\rho}_1 := \rho_1 \circ S, \quad \tilde{c}(x, y) := \gamma^{-1} c(x, S(y)), \quad \tilde{\pi} := (\text{Id} \times S^{-1})_{\#} \pi,$$

we have that $\nabla_{xy} \tilde{c}(x_0, 0) = -\mathbb{I}$, $\tilde{\pi}$ is a \tilde{c} -optimal coupling between $\tilde{\rho}_0$ and $\tilde{\rho}_1$ and $\tilde{\pi}$, $\tilde{\rho}_0$, $\tilde{\rho}_1$ and \tilde{c} still satisfy (2.4.87).

Without loss of generality, we may thus assume $x_0 = 0$, so that (2.4.87) becomes (2.4.85) and (2.4.88) becomes (2.4.86).

STEP 2 (First step of the iteration). By Lemma 2.2.1, there exist $\Lambda_0 < \infty$ and $R_0 > 0$ such that for all $R \leq R_0$ for which (2.4.85) holds, recall also (2.4.75), we have

$$\#_{\frac{5}{6}R} \cap \text{Spt } \pi \subseteq \mathbb{B}_{\frac{5}{6}R, \frac{\Lambda_0}{6}R}. \quad (2.4.89)$$

In view of (2.4.85) and (2.4.89), we may apply Proposition 2.4.4 to get the existence of a symmetric matrix B_1 and a vector b_1 such that, defining $D_1 := -\nabla_{xy} c_1(0, b_1)$, $\gamma_1 := (\det D_1)^{-1/d}$ and $\rho_0^1, \rho_1^1, \pi_1$ and c_1 as $\hat{\rho}_0, \hat{\rho}_1, \hat{\pi}$ and \hat{c} in Proposition 2.4.4, we get

$$\mathcal{E}^1 := \mathcal{E}_{\theta R}(\pi_1) \leq \theta^{2\beta} \mathcal{E}_R + C_\beta (\mathcal{L}_R + \mathcal{H}_R), \quad (2.4.90)$$

and π_1 is a c_1 -optimal coupling between ρ_0^1 and ρ_1^1 . We also get

$$\#_{\frac{5}{6}\theta R} \cap \text{Spt } \pi_1 \subseteq \mathbb{B}_{\frac{5}{6}\theta R, \frac{4}{3}\theta R}. \quad (2.4.91)$$

STEP 3 (Iterating Proposition 2.4.4). We are going to show that we can iterate Proposition 2.4.4 a finite number of times.

Let us compute, using in the last inequality the inclusion $\gamma_1^{-1}D_1^{-*}B_1B_{\theta R} + b_1 \subset B_R$ which comes from the estimates (2.4.78), (2.4.82) and (2.4.83):

$$\|\rho_1^1 - 1\|_{L^\infty(B_{\theta R})} = \sup_{\widehat{y} \in B_{\theta R}} |\rho_1(\gamma_1^{-1}D_1^{-*}B_1\widehat{y} + b_1) - 1| = \sup_{y \in \gamma_1^{-1}D_1^{-*}B_1B_{\theta R} + b_1} |\rho_1(y) - 1| \leq \|\rho_1 - 1\|_{L^\infty(B_R)}.$$

Similarly, using $B_1^{-1}B_{\theta R} \subset B_R$,

$$\|\rho_0^1 - 1\|_{L^\infty(B_{\theta R})} \leq \|\rho_0 - 1\|_{L^\infty(B_R)},$$

so that

$$\mathcal{L}^1 := \mathcal{L}_{\theta R}(\rho_0^1, \rho_1^1) \leq \mathcal{L}_R. \quad (2.4.92)$$

Finally, using the same inclusions and setting $Q_1(x, y) := (B_1x, \gamma_1 B_1^{-1}D_1^*(y - b_1))$ so that $Q_1^{-1}(\#_{\theta R}) \subset \#_R$, we get

$$\begin{aligned} [\nabla_{xy}c_1]_{\alpha, \theta R} &:= \sup_{(\widehat{x}, \widehat{y}) \neq (\widehat{x}', \widehat{y}') \in \#_{\theta R}} \frac{|\nabla_{xy}c_1(\widehat{x}, \widehat{y}) - \nabla_{xy}c_1(\widehat{x}', \widehat{y}')|}{|\widehat{x} - \widehat{x}'|^\alpha + |\widehat{y} - \widehat{y}'|^\alpha} \\ &= \sup_{(\widehat{x}, \widehat{y}) \neq (\widehat{x}', \widehat{y}') \in \#_{\theta R}} \frac{|B_1^{-1}(\nabla_{xy}c(Q_1^{-1}(\widehat{x}, \widehat{y})) - \nabla_{xy}c(Q_1^{-1}(\widehat{x}', \widehat{y}'))))D_1^{-1}B_1|}{|B_1(B_1^{-1}\widehat{x} - B_1^{-1}\widehat{x}')|^\alpha + |\gamma_1 B_1^{-1}D_1^*(\gamma_1^{-1}D_1^{-*}B_1\widehat{y} - \gamma_1^{-1}D_1^{-*}B_1\widehat{y}')|^\alpha} \\ &\leq |B_1| \|B_1^{-1}\| \|D_1^{-1}\| \max\{|B_1^{-1}|^\alpha, (\gamma_1^{-1}|B_1| \|D_1^{-1}\|)^\alpha\} \\ &\quad \times \sup_{(x, y) \neq (x', y') \in Q_1^{-1}(\#_{\theta R})} \frac{|\nabla_{xy}c(x, y) - \nabla_{xy}c(x', y')|}{|x - x'|^\alpha + |y - y'|^\alpha}. \end{aligned}$$

Estimates (2.4.78), (2.4.82) and (2.4.83) thus yield

$$[\nabla_{xy}c_1]_{\alpha, \theta R} \leq \left(1 + C(\mathcal{E}_R^{1/2} + \mathcal{L}_R^{1/2} + \mathcal{H}_R^{1/2})\right) [\nabla_{xy}c]_{\alpha, R},$$

so that

$$\mathcal{H}^1 := \mathcal{H}_{\theta R}(c_1) \leq \theta^{2\alpha} \left(1 + C(\mathcal{E}_R^{1/2} + \mathcal{L}_R^{1/2} + \mathcal{H}_R^{1/2})\right) \mathcal{H}_R. \quad (2.4.93)$$

Therefore, (2.4.85), (2.4.90), (2.4.91), (2.4.92), and (2.4.93) imply that we may iterate Proposition 2.4.4 $K > 1$ times to find sequences of matrices B_k and D_k with $\det B_k = 1$, a sequence of vectors b_k , and sequences of real numbers γ_k such that, setting for $1 \leq k \leq K$

$$\begin{aligned} \rho_0^k(x) &:= \rho_0^{k-1}(B_kx), & \rho_1^k(y) &:= \rho_1^{k-1}(\gamma_k^{-1}D_k^{-1}B_k^{-1}y + b_k), \\ Q_k(x, y) &:= (B_k^{-1}x, \gamma_k B_k D_k(y - b_k)), & \pi_k &:= Q_{k\#}\pi_{k-1}, \\ \text{and } c_k(x, y) &:= \gamma_k c_{k-1}(B_kx, \gamma_k^{-1}D_k^{-1}B_k^{-1}y + b_k) \end{aligned} \quad (2.4.94)$$

we have $\nabla_{xy}c_k(0,0) = -\mathbb{I}$, π_k is a c_k -optimal coupling between ρ_0^k and ρ_1^k , and the inclusion

$$\#_{\frac{5}{6}\theta^k R} \cap \text{Spt } \pi_k \subseteq \mathbb{B}_{\frac{5}{6}\theta^k R, \frac{4}{3}\theta^k R}. \quad (2.4.95)$$

Moreover, defining

$$\mathcal{E}^k := \mathcal{E}_{\theta^k R}(\pi_k), \quad \mathcal{L}^k := \mathcal{L}_{\theta^k R}(\rho_0^k, \rho_1^k) \quad \text{and} \quad \mathcal{H}^k := \mathcal{H}_{\theta^k R}(c_k), \quad (2.4.96)$$

we have

$$\mathcal{E}^k \leq \theta^{2\beta} \mathcal{E}^{k-1} + C_\beta (\mathcal{L}^{k-1} + \mathcal{H}^{k-1}), \quad (2.4.97)$$

$$|B_k - \mathbb{I}|^2 + \frac{1}{(\theta^{k-1}R)^2} |b_k|^2 \lesssim \mathcal{E}^{k-1} + \mathcal{L}^{k-1}, \quad (2.4.98)$$

$$|D_k - \mathbb{I}|^2 + |\gamma_k - 1|^2 \lesssim \mathcal{H}^{k-1}. \quad (2.4.99)$$

Arguing as for (2.4.92) and (2.4.93), we get the existence of a constant $C_1(d, \alpha) > 0$ such that for all $1 \leq k \leq K$

$$\mathcal{L}^k \leq \mathcal{L}^{k-1} \quad \text{and} \quad \mathcal{H}^k \leq \theta^{2\alpha} \left(1 + C_1 \left((\mathcal{E}^{k-1})^{1/2} + (\mathcal{L}^{k-1})^{1/2} + (\mathcal{H}^{k-1})^{1/2} \right) \right) \mathcal{H}^{k-1}. \quad (2.4.100)$$

STEP 4 (Smallness at any step of the iteration). Let us show by induction that there exists a constant $C_2(d, \beta)$ such that for all $K \in \mathbb{N}$

$$\mathcal{E}^K \leq C_2 (\mathcal{E}_R + \mathcal{L}_R + \mathcal{H}_R), \quad (2.4.101)$$

$$\mathcal{H}^K \leq \mathcal{H}^{K-1}. \quad (2.4.102)$$

This will show, together with (2.4.95) and the estimate

$$\mathcal{L}^K \leq \mathcal{L}_R \quad \text{for all } K \in \mathbb{N} \quad (2.4.103)$$

coming from (2.4.100), that we can keep on iterating Proposition 2.4.4.

We set

$$C_2 := \max \left\{ \theta^{2\beta}, \frac{C_\beta}{1 - \theta^{2\beta}} \right\}. \quad (2.4.104)$$

By (2.4.90) and (2.4.93), (2.4.101) and (2.4.102) hold for $K = 1$ since $C_2 \geq \max\{\theta^{2\beta}, C_\beta\}$ and provided $\mathcal{E}_R + \mathcal{L}_R + \mathcal{H}_R$ is small enough.

Assume now (2.4.101) and (2.4.102) hold for all $1 \leq k \leq K - 1$. By induction hypothesis, we have

$$\mathcal{E}^{K-1} \leq C_2 (\mathcal{E}_R + \mathcal{L}_R + \mathcal{H}_R) \quad \text{and} \quad \mathcal{H}^{K-1} \leq \mathcal{H}_R.$$

Combining these two estimates with (2.4.100) for $k = K$ and assuming that

$$\theta^{2\alpha} \left(1 + C_1 \left(C_2^{1/2} (\mathcal{E}_R^{1/2} + \mathcal{L}_R^{1/2} + \mathcal{H}_R^{1/2}) + \mathcal{L}_R^{1/2} + \mathcal{H}_R^{1/2} \right) \right) \leq 1,$$

which is possible by (2.4.85), we get (2.4.102). Similarly, by (2.4.97) with $k = K$ we get, using $C_2 \geq \frac{C_\beta}{1-\theta^{2\beta}}$ by (2.4.104)

$$\mathcal{E}^K \leq \left(\theta^{2\beta} C_2 + C_\beta \right) (\mathcal{E}_R + \mathcal{L}_R + \mathcal{H}_R) \leq C_2 (\mathcal{E}_R + \mathcal{L}_R + \mathcal{H}_R).$$

This concludes the induction.

STEP 5 (Campanato estimate). Assuming (2.4.85), we show (2.4.86).

Letting

$$\begin{aligned} \bar{\gamma}_k &:= \gamma_k \dots \gamma_1, & \bar{B}_k &:= B_1 \dots B_k, & \bar{D}_k &:= B_k D_k \dots B_1 D_1, \\ \bar{b}_k &:= \sum_{i=1}^k (\gamma_i B_i D_i) \dots (\gamma_i B_i D_i) b_i & \text{and} & & \bar{Q}_k(x, y) &:= (\bar{B}_k^{-1} x, \bar{\gamma}_k \bar{D}_k y - \bar{b}_k), \end{aligned}$$

we see that, recalling (2.4.94),

$$\pi_k(x) = \bar{Q}_k \# \pi \quad \text{and} \quad \rho_0^k(x) = \rho_0(\bar{B}_k x). \quad (2.4.105)$$

Moreover, $\det \bar{B}_k = 1$ and from (2.4.98), (2.4.99), (2.4.101), (2.4.102), and (2.4.103), there exists a constant $C_3(d, \alpha, \beta)$ such that,

$$|\bar{B}_k|^2 + |\bar{\gamma}_k|^2 \leq (1 + C_3(\mathcal{E}_R^{1/2} + \mathcal{L}_R^{1/2} + \mathcal{H}_R^{1/2}))^k, \quad (2.4.106)$$

$$\text{and} \quad |\bar{D}_k|^2 \leq (1 + C_3(\mathcal{E}_R^{1/2} + \mathcal{L}_R^{1/2} + \mathcal{H}_R^{1/2}))^{2k}. \quad (2.4.107)$$

In particular, setting

$$\tilde{\theta} := \frac{\theta}{1 + C_3(\mathcal{E}_R^{1/2} + \mathcal{L}_R^{1/2} + \mathcal{H}_R^{1/2})},$$

we obtain the inclusion

$$B_{\frac{1}{2}\tilde{\theta}^k R} \times \mathbb{R}^d \subset \bar{B}_k(B_{\theta^k R}) \times \mathbb{R}^d = \bar{Q}_k^{-1}(B_{\theta^k R} \times \mathbb{R}^d). \quad (2.4.108)$$

Let us now compute, using that $\det \bar{B}_k = 1$, $\|1 - \rho_0^k\|_{L^\infty(B_{\theta^k R})} \lesssim 1$ and $\mathcal{E}^k + 1 \lesssim 1$,

$$\begin{aligned} & \frac{1}{\left(\frac{1}{2}\tilde{\theta}^k R\right)^d} \inf_b \int_{B_{\frac{1}{2}\tilde{\theta}^k R} \times \mathbb{R}^d} |y - b|^2 d\pi \\ & \stackrel{(2.4.108)}{\lesssim} \frac{1}{(\tilde{\theta}^k R)^d} \int_{\bar{Q}_k^{-1}(B_{\theta^k R} \times \mathbb{R}^d)} |y - \bar{\gamma}_k^{-1} \bar{D}_k^{-1} \bar{b}_k|^2 d\pi \\ & \stackrel{(2.4.105)}{=} \frac{1}{(\tilde{\theta}^k R)^d} \int_{B_{\theta^k R} \times \mathbb{R}^d} |\bar{\gamma}_k^{-1} \bar{D}_k^{-1} y|^2 d\pi_k \\ & \lesssim \frac{|\bar{\gamma}_k^{-1} \bar{D}_k^{-1}|^2}{(\tilde{\theta}^k R)^d} \int_{B_{\theta^k R} \times \mathbb{R}^d} |y - x|^2 d\pi_k + \frac{|\bar{\gamma}_k^{-1} \bar{D}_k^{-1}|^2}{(\tilde{\theta}^k R)^d} \int_{B_{\theta^k R}} |x|^2 \rho_0^k(x) dx \\ & \stackrel{(2.4.107) \& (2.4.106)}{\lesssim} \frac{(1 + C_3(\mathcal{E}_R^{1/2} + \mathcal{L}_R^{1/2} + \mathcal{H}_R^{1/2}))^{k(d+3)}}{(\tilde{\theta}^k R)^d} [(\theta^k R)^{d+2} \mathcal{E}^k + (\theta^k R)^{d+2}] \\ & \lesssim (1 + C_3(\mathcal{E}_R^{1/2} + \mathcal{L}_R^{1/2} + \mathcal{H}_R^{1/2}))^{k(d+3)} \theta^{2k} R^2. \end{aligned}$$

Therefore, noticing we may assume that $1 + C_3(\mathcal{E}_R^{1/2} + \mathcal{L}_R^{1/2} + \mathcal{H}_R^{1/2}) \leq \theta^{-\frac{2-2\beta}{d+3+2\beta}}$, we get

$$\frac{1}{\left(\frac{1}{2}\tilde{\theta}^k R\right)^d} \inf_b \int_{B_{\frac{1}{2}\tilde{\theta}^k R} \times \mathbb{R}^d} |y - b|^2 d\pi \lesssim (\tilde{\theta}^k R)^{2\beta} R^{2-2\beta}.$$

STEP 6 (Spt π is contained in the graph of a function T within $B_R \times \mathbb{R}^d$). Exactly as in STEP 7 of the proof of Theorem 2.1.1, (2.4.84) implies the existence of a function $T : B_R \rightarrow Y$ such that

$$(B_R \times \mathbb{R}^d) \cap \text{Spt } \pi \subseteq \text{graph } T. \quad (2.4.109)$$

STEP 7 (Conclusion). By boundedness of ρ_0 , (2.4.84) implies the bound

$$\sup_{0 < r < \frac{R}{2}} \sup_{x_0 \in B_R} \frac{1}{r^{d+2\beta}} \inf_{A,b} \int_{B_r(x_0) \cap B_R} |T(x) - b|^2 dx \lesssim R^{2-2\beta},$$

which by means of Campanato's theory [Cam64] proves that $T \in C^{0,\beta}(B_R)$. \blacksquare

2.4.4 ε -regularity for almost-minimizers

In this section we give a sketch of the proof of Theorem 2.1.18. One of the main differences compared to the situation of Theorem 2.1.1 is that our assumptions do not allow us to prove an L^∞ bound on the displacement (which followed from $(c-)$ monotonicity of Spt π). However, almost-minimality (on all scales) allows us to obtain an L^p bound for arbitrarily large $p < \infty$.

Proposition 2.4.5. *Assume that $\rho_0, \rho_1 \in C^{0,\alpha}$ with $\rho_0(0) = \rho_1(0) = 1$. Let T be an almost-minimizing transport map from $\mu = \rho_0 dx$ to $\nu = \rho_1 dy$ with $\Delta_r \leq 1$. Assume further that T is invertible. Then there exists a radius $R_1 = R_1(\rho_0, \rho_1) > 0$ such that for any $6R \leq R_1$,*

$$\mathcal{E}_{6R}(\pi_T) + R^\alpha ([\rho_0]_{\alpha,6R} + [\rho_1]_{\alpha,6R}) \ll 1 \quad (2.4.110)$$

implies that for any $p < \infty$,

$$\frac{1}{R} \left(\frac{1}{R^d} \int_{B_{2R}} |T(x) - x|^p \mu(dx) + \frac{1}{R^d} \int_{B_{2R}} |T^{-1}(y) - y|^p \nu(dy) \right)^{\frac{1}{p}} \lesssim_p \mathcal{E}_{6R}(\pi_T)^{\frac{1}{d+2}}. \quad (2.4.111)$$

The scale R_1 below which the result holds depends on the global Hölder semi-norms $[\rho_0]_\alpha$ and $[\rho_1]_\alpha$ of the densities and the condition $B_{R_1} \subset \text{Spt } \rho_0 \cap \text{Spt } \rho_1$.

The proof of Proposition 2.4.5 is given in Appendix 2.D. Note that since T^{-1} is also almost-minimizing, the L^p bound for T^{-1} follows from applying Proposition 2.D.1 to T^{-1} . The L^p estimate (for arbitrarily large $p < \infty$) allows us to split the particle trajectories into two groups:

- *good trajectories* that satisfy an L^∞ bound on the displacement, corresponding to starting points in the set

$$\mathcal{G} = \{x \in B_{2R} \cup T^{-1}(B_{2R}) : |T(x) - x| \leq MR\}, \quad (2.4.112)$$

where $M := (\mathcal{E}_{6R}(\pi_T) + \mathcal{D}_{6R}(\mu, \nu))^\alpha$ for some $\alpha \in (0, \frac{1}{d+2})$, that we fix in what follows, and

- *bad trajectories* that are too long, corresponding to

$$\mathcal{B} = \{x \in B_{2R} \cup T^{-1}(B_{2R}) : |T(x) - x| > MR\}. \quad (2.4.113)$$

Due to the L^p bound, the energy carried by bad trajectories is superlinearly small: By definition of $\mathcal{E}_{6R}(\pi_T)$ and M ,

$$\frac{1}{R^d} \mu(\mathcal{B}) \lesssim \frac{1}{M^2} \frac{1}{R^{d+2}} \int_{B_{2R} \cup T^{-1}(B_{2R})} |T(x) - x|^2 \mu(dx) \lesssim \mathcal{E}_{6R}(\pi_T)^{1-2\alpha}, \quad (2.4.114)$$

hence

$$\begin{aligned} \frac{1}{R^{d+2}} \int_{\mathcal{B}} |T(x) - x|^2 \mu(dx) &\leq \left(\frac{1}{R^d} \mu(\mathcal{B}) \right)^{1-\frac{2}{p}} \left(\frac{1}{R^{d+p}} \int_{B_{2R} \cup T^{-1}(B_{2R})} |T(x) - x|^p \mu(dx) \right)^{\frac{2}{p}} \\ &\lesssim \mathcal{E}_{6R}(\pi_T)^{\frac{2}{d+2} + (1-2\alpha)\left(1-\frac{2}{p}\right)}, \end{aligned}$$

from which we see that, given $\alpha \in (0, \frac{1}{d+2})$, we may choose p large enough so that the exponent is larger than 1. In particular, for any $\tau > 0$ we may bound

$$\frac{1}{R^{d+2}} \int_{\mathcal{B}} |T(x) - x|^2 \mu(dx) \leq \tau \mathcal{E}_{6R}(\pi_T), \quad (2.4.115)$$

provided $\mathcal{E}_{6R}(\pi_T) \ll 1$.

Once the bad trajectories have been removed, the good trajectories can be treated as before. More precisely, if we restrict the coupling π_T to the set $\mathcal{G} \times T(\mathcal{G})$, then the resulting coupling is still deterministic and almost-minimizing with respect to quadratic cost (given its own boundary conditions). In particular, since $\mathcal{G} \times T(\mathcal{G}) \subset \#_{2R}$ the *global* estimate

$$\int_{\mathcal{G}} |T(x) - x|^2 \mu(dx) \leq \int_{\mathcal{G}} |\tilde{T}(x) - x|^2 \mu(dx) + \Delta_{2R}$$

for all $\tilde{T}_{\#} \mu|_{\mathcal{G}} = T_{\#} \mu|_{\mathcal{G}} = \nu|_{T(\mathcal{G})}$ holds. This allows for a passage from the Lagrangian to the Eulerian point of view like in Lemma 2.1.12.²⁴ Moreover, the measures $\mu|_{\mathcal{G}}$ and $\nu|_{T(\mathcal{G})}$, as well as the coupling $\pi_T|_{\mathcal{G} \times T(\mathcal{G})}$, satisfy the assumptions of the harmonic approximation Theorem 2.1.13.²⁵ Hence, given $0 < \tau \ll 1$, there exists a threshold $\varepsilon_\tau > 0$, constants

²⁴Using that the optimal coupling between $\mu|_{\mathcal{G}}$ and $\nu|_{T(\mathcal{G})}$ is deterministic, so that one can appeal to almost-minimality within the class of deterministic couplings.

²⁵Note that there is a slight mismatch in the power α in the L^∞ bound between the definition of good trajectories and the setting of [GHO21]. However, one can convince oneself easily that the results of [GHO21] still apply.

$C, C_\tau < \infty$, and a harmonic gradient field $\nabla\Phi$ (defined through (2.1.39)) satisfying (2.1.41), such that (2.1.40) holds for the Eulerian description (ρ, j) of $\pi_T \llcorner_{\mathcal{G} \times T(\mathcal{G})}$, provided $\mathcal{E}_{6R}(\pi_T) + R^{2\alpha} \left([\rho_0]_{\alpha,6R}^2 + [\rho_1]_{\alpha,6R}^2 \right) \leq \varepsilon_\tau$. The harmonic gradient field allows us to define the affine change of coordinates from Lemma 2.1.15 with $B = e^{\frac{A}{2}}$ with $A = \nabla^2\Phi(0)$ and $b = \nabla\Phi(0)$ satisfying (2.1.50) (and $D = \mathbb{I}$) to obtain a new coupling $\widehat{\pi}_T = \pi_{\widehat{T}}$ between the measures $\widehat{\mu}$ and $\widehat{\nu}$ from the (full) coupling π_T .

We can now use the harmonic approximation result Theorem 2.1.13 together with the harmonic estimates (2.1.41) to bound

$$\frac{1}{R^{d+2}} \int_{\#_R \cap (\mathcal{G} \times T(\mathcal{G}))} |y - x - (Ax + b)|^2 d\pi_T \leq \tau \mathcal{E}_{6R}(\pi_T) + C_\tau R^{2\alpha} ([\rho_0]_{\alpha,6R}^2 + [\rho_1]_{\alpha,2R}^2) + \Delta_{2R}.$$

For the bad trajectories we use the estimate (2.4.115) together with the bound

$$\begin{aligned} \int_{\#_R \cap (\mathcal{B} \times T(\mathcal{B}))} |Ax + b|^2 d\pi_T &\lesssim \int_{(B_R \cap \mathcal{B}) \times \mathbb{R}^d} |Ax + b|^2 d\pi_T + |A|^2 \int_{\mathcal{B} \times T(\mathcal{B})} |y - x|^2 d\pi_T \\ &\quad + |A|^2 \int_{\mathcal{B} \times (B_R \cap T(\mathcal{B}))} |y|^2 d\pi_T \\ &\stackrel{(2.4.115)}{\lesssim} (|A|^2 R^2 + |b|^2) \mu(\mathcal{B}) + \tau |A|^2 R^{d+2} \mathcal{E}_{6R}(\pi_T) \\ &\stackrel{(2.4.114) \& (2.1.50)}{\lesssim} R^{d+2} \left(\mathcal{E}_{6R}(\pi_T)^{2-2\alpha} + R^{2\alpha} ([\rho_0]_{\alpha,6R}^2 + [\rho_1]_{\alpha,6R}^2) + \tau \mathcal{E}_{6R}(\pi_T) \right) \end{aligned}$$

to obtain, recalling that $\alpha < \frac{1}{d+2} < \frac{1}{2}$,

$$\frac{1}{R^{d+2}} \int_{\#_R \cap (\mathcal{B} \times T(\mathcal{B}))} |y - x - (Ax + b)|^2 d\pi_T \leq \tau \mathcal{E}_{6R}(\pi_T) + C_\tau R^{2\alpha} ([\rho_0]_{\alpha,6R}^2 + [\rho_1]_{\alpha,6R}^2). \quad (2.4.116)$$

In particular, for $\theta \in (0, 1)$ we can write

$$(\theta R)^{d+2} \mathcal{E}_{\theta R}(\pi_{\widehat{T}}) = \int_{\#_{\theta R}} |\widehat{y} - \widehat{x}|^2 \pi_{\widehat{T}}(d\widehat{x}d\widehat{y}) = |\det B| \int_{Q^{-1}(\#_{\theta R})} |\gamma B^{-*}(y - b) - Bx|^2 \pi_T(dx dy),$$

so that using the identity $\gamma B^{-*}(y - b) - Bx = \gamma B^{-*}(y - x - (Ax + b)) - \gamma B^{-*}(B^* B - \mathbb{I} - A)x + (\gamma - 1)Bx$, the estimate (2.4.116) together with (2.1.41) and (2.1.50) imply that for any $\beta \in (0, 1)$ there exist $0 < \theta \ll 1$ and $C_\beta < \infty$ such that

$$\mathcal{E}_{\theta R}(\pi_{\widehat{T}}) \leq \theta^{2\beta} \mathcal{E}_{6R}(\pi_T) + C_\beta \left(R^{2\alpha} [\rho_0]_{\alpha,6R}^2 + R^{2\alpha} [\rho_1]_{\alpha,6R}^2 + \Delta_{2R} \right),$$

which implies a one-step-improvement result for the case of general almost-minimizers.

It remains to show that the transformed coupling $\pi_{\widehat{T}}$ is still almost-minimal on all small scales in the sense of Definition 2.1.17. To this end, let $r \leq R_1$, $(\widehat{x}_0, \widehat{y}_0) \in \text{Spt } \pi_{\widehat{T}}$, and $\pi_{\widehat{T}_r} \in \Pi(\widehat{\mu}, \widehat{\nu})$ with $\text{Spt}(\pi_{\widehat{T}_r} - \pi_{\widehat{T}}) \subset (B_r(\widehat{x}_0) \times \mathbb{R}^d) \cup (\mathbb{R}^d \times B_r(\widehat{y}_0))$. Then, writing

$\widehat{T}'(\widehat{x}) = \gamma B^{-*}(T'(B^{-1}\widehat{x}) - b)$, where $T'_{\#}\mu = \nu$, one sees that

$$\begin{aligned} \int |\widehat{y} - \widehat{x}|^2 d(\pi_{\widehat{T}} - \pi_{\widehat{T}'}) &= -2 \int (\widehat{T}(\widehat{x}) - \widehat{T}'(\widehat{x})) \cdot \widehat{x} \widehat{\mu}(d\widehat{x}) \\ &= -2|\det B| \int \gamma B^{-*}(T(x) - T'(x)) \cdot Bx \mu(dx) \\ &= -2\gamma|\det B| \int (T(x) - T'(x)) \cdot x \mu(dx) \\ &= \gamma|\det B| \int |y - x|^2 d(\pi_T - \pi_{T'}). \end{aligned}$$

Note that

$$\text{Spt}(\pi_T - \pi_{T'}) = Q^{-1} \text{Spt}(\pi_{\widehat{T}} - \pi_{\widehat{T}'}) \subset (B|_{B|_r}(x_0) \times \mathbb{R}^d) \cup (\mathbb{R}^d \times B_{\frac{|B|}{\gamma}}(y_0)),$$

where $x_0 = B\widehat{x}_0$ and $y_0 = \gamma^{-1}B\widehat{y}_0 + b$. Since π_T is almost-minimizing, it follows that

$$\int |\widehat{y} - \widehat{x}|^2 d(\pi_{\widehat{T}} - \pi_{\widehat{T}'}) \leq \gamma|\det B|(\max\{1, \gamma^{-1}\}|B|r)^{d+2} \Delta_{\max\{1, \gamma^{-1}\}|B|r},$$

hence $\pi_{\widehat{T}}$ is almost-minimizing among deterministic couplings with rate

$$\widehat{\Delta}_r = \gamma|\det B|(\max\{1, \gamma^{-1}\}|B|)^{d+2} \Delta_{\max\{1, \gamma^{-1}\}|B|r}.$$

Assuming that $\Delta_r = Cr^{2\alpha}$, together with the bounds on γ and B from (2.1.50) this gives

$$\widehat{\Delta}_r \leq \left(1 + C(\mathcal{E}_{R_1}^{\frac{1}{2}} + R_1^\alpha[\rho_0]_{\alpha, R_1} + R_1^\alpha[\rho_1]_{\alpha, R_1})\right) \Delta_r,$$

in particular the rate Δ_r exhibits the same behaviour as the Hölder seminorm of $\nabla_{xy}c$ in (2.4.40) and shows that the one-step-improvement can be iterated down to arbitrarily small scales, yielding the $C^{1,\alpha}$ -regularity of T in a ball with radius given by a fraction of R . \blacksquare

2.5 Partial regularity: Proof of Corollary 2.1.4

As a corollary of Theorem 2.1.1, we obtain a variational proof of partial regularity for optimal transport maps proved in [DF15]. The changes of variables used to arrive to a normalized situation are exactly the same as in [DF15] and the argument to derive partial regularity from ε -regularity follows [GO20].

Proof of Corollary 2.1.4. We recall that a classical result in optimal transport states that the optimal map T from ρ_0 to ρ_1 for the cost c and the optimal map T^* from ρ_1 to ρ_0 for the cost $c^*(y, x) := c(x, y)$ are almost everywhere inverse to each other, and are of the form

$$T(x) = c\text{-exp}_x(\nabla u(x)) \quad \text{and} \quad T^{-1}(y) := T^*(y) = c^*\text{-exp}_y(\nabla u^c(y)),$$

where u is a c -convex function and u^c is the c -conjugate of u , see subsection I.2. u and u^c are semi-convex so that by Alexandrov's Theorem, they are twice differentiable almost

everywhere. Therefore, we can find two sets of full measure $X_1 \subseteq X$ and $Y_1 \subseteq Y$ such that for all $(x_0, y_0) \in X_1 \times Y_1$, u is twice differentiable at x_0 , u^c is twice differentiable at y_0 and

$$T^{-1}(T(x_0)) = x_0 \quad \text{and} \quad T(T^{-1}(y_0)) = y_0. \quad (2.5.1)$$

Now let

$$X' := X_1 \cap T^{-1}(Y_1) \quad \text{and} \quad Y' := Y_1 \cap T(X_1). \quad (2.5.2)$$

Because ρ_0 and ρ_1 are bounded and bounded away from zero, T sends sets of measure 0 to sets of measure 0 so that $|X \setminus X'| = |Y \setminus Y'| = 0$. The goal is now to prove that X' and Y' are open sets and that T is a $C^{1,\alpha}$ -diffeomorphism between X' and Y' .

Fix $x_0 \in X'$, then by (2.5.2), $y_0 := T(x_0) \in Y'$. Up to translation, we may assume that $x_0 = y_0 = 0$. Define

$$\begin{aligned} \bar{u}(x) &:= u(x) - u(0) + c(x, 0) - c(0, 0), \\ \bar{c}(x, y) &:= c(x, y) - c(x, 0) - c(0, y) + c(0, 0). \end{aligned} \quad (2.5.3)$$

Then \bar{u} is a \bar{c} -convex function and we have

$$\bar{c}\text{-exp}_x(\nabla \bar{u}(x)) = c\text{-exp}_x(\nabla u(x)), \quad (2.5.4)$$

so that $T(x) = \bar{c}\text{-exp}_x(\nabla \bar{u}(x))$, from which we know that T is the \bar{c} -optimal transport map from ρ_0 to ρ_1 .

By Alexandrov's Theorem, there exist a symmetric matrix A such that

$$\nabla \bar{u}(x) = \nabla \bar{u}(0) + Ax + o(|x|),$$

so that, using that $(p, x) \mapsto c\text{-exp}_x(p)$ is C^1 and setting $M := -\nabla_{xy}\bar{c}(0, 0) = -\nabla_{xy}c(0, 0)$, noticing that by Assumption (C4) M is nondegenerate, a simple computation yields

$$T(x) = M^{-1}Ax + o(|x|).$$

Therefore, we have

$$\frac{1}{R^{d+2}} \int_{B_R} |T(x) - M^{-1}Ax|^2 \rho_0(x) dx \xrightarrow{R \rightarrow 0} 0. \quad (2.5.5)$$

The \bar{c} -convexity of \bar{u} and the fact that $\bar{c}\text{-exp}_x(\nabla \bar{u}(x)) \in \partial_{\bar{c}} \bar{u}(x)$ imply that, see for instance [Fig17, Section 5.3],

$$\nabla^2 \bar{u}(x) + \nabla_{xx} \bar{c}(x, \bar{c}\text{-exp}_x(\nabla \bar{u}(x))) \geq 0,$$

so that, together with (2.5.3), (2.5.4) and the property $T(0) = 0$, the matrix $A = \nabla^2 \bar{u}(0)$ is positive definite. We now make the change of variables $\tilde{x} := A^{\frac{1}{2}}x$ and $\tilde{y} := A^{-\frac{1}{2}}My$ so that

$$\begin{aligned} \tilde{T}(\tilde{x}) &:= A^{-\frac{1}{2}}MT(A^{\frac{1}{2}}\tilde{x}), \\ \tilde{c}(\tilde{x}, \tilde{y}) &:= \bar{c}(A^{-\frac{1}{2}}\tilde{x}, M^{-1}A^{\frac{1}{2}}\tilde{y}). \end{aligned}$$

Defining

$$\tilde{\rho}_0(\tilde{x}) := \det(A^{-\frac{1}{2}})\rho_0(A^{-\frac{1}{2}}\tilde{x}) \quad \text{and} \quad \tilde{\rho}_1(\tilde{y}) := |\det(M^{-1}A^{\frac{1}{2}})|\rho_1(M^{-1}A^{\frac{1}{2}}\tilde{y}),$$

we get that $\tilde{T}_\# \tilde{\rho}_0 = \tilde{\rho}_1$ \tilde{c} -optimally. This may be seen noticing that

$$\tilde{T}(\tilde{x}) = \tilde{c}\text{-exp}_{\tilde{x}}(\nabla \tilde{u}(\tilde{x})),$$

where $\tilde{u}(\tilde{x}) := \bar{u}(A^{-1/2}x)$ is a \tilde{c} -convex function. The cost \tilde{c} satisfies $\nabla_{\tilde{x}\tilde{y}}\tilde{c}(0,0) = -\mathbb{I}$ and by the Monge–Ampère equation

$$|\det \nabla \tilde{T}(x)| = \frac{\tilde{\rho}_0(x)}{\tilde{\rho}_1(\tilde{T}(x))},$$

we obtain $\tilde{\rho}_0(0) = \tilde{\rho}_1(0)$. Up to dividing $\tilde{\rho}_0$ and $\tilde{\rho}_1$ by an equal constant, we may assume that $\tilde{\rho}_0(0) = \tilde{\rho}_1(0) = 1$. Moreover, with this change of variables, (2.5.5) turns into

$$\frac{1}{R^{d+2}} \int_{B_R} |\tilde{T}(\tilde{x}) - \tilde{x}|^2 \tilde{\rho}_0(\tilde{x}) \, d\tilde{x} \xrightarrow{R \rightarrow 0} 0. \quad (2.5.6)$$

Finally, \tilde{c} is still $\mathcal{C}^{2,\alpha}$ and satisfies Assumptions (C2)–(C4) and since ρ_0 and ρ_1 are bounded and bounded away from zero, $\tilde{\rho}_0$ and $\tilde{\rho}_1$ are $\mathcal{C}^{0,\alpha}$, and we have

$$\mathcal{H}_R(\tilde{\rho}_0, \tilde{\rho}_1, \tilde{c}) = R^{2\alpha} \left([\tilde{\rho}_0]_{\alpha,R}^2 + [\tilde{\rho}_1]_{\alpha,R}^2 + [\nabla_{xy}\tilde{c}]_{\alpha,R}^2 \right) \xrightarrow{R \rightarrow 0} 0. \quad (2.5.7)$$

Hence by (2.5.6) and (2.5.7), we may apply Theorem 2.1.1 to obtain that \tilde{T} is $\mathcal{C}^{1,\alpha}$ in a neighborhood of zero. By Remark 2.1.2, we also obtain that \tilde{T}^{-1} is $\mathcal{C}^{1,\alpha}$ in a neighborhood of zero.

Going back to the original map, this means that T is a $\mathcal{C}^{1,\alpha}$ diffeomorphism between a neighborhood U of x_0 and the neighborhood $T(U)$ of $T(x_0)$. In particular, $U \times T(U) \subseteq X' \times Y'$ so that X' and Y' are both open and by (2.5.1), T is a global $\mathcal{C}^{1,\alpha}$ diffeomorphism between X' and Y' . ■

2.A Some technical lemmas

2.A.1 Properties of the support of couplings

The following lemma is an important ingredient in the proofs of our L^∞ bounds on the displacement of couplings with c -monotone support, Proposition 2.1.5 and Lemma 2.2.1:

Lemma 2.A.1. *Let $\pi \in \Pi(\mu, \nu)$ and assume that there exists $R > 0$ such that*

$$\frac{1}{R^{d+2}} \int_{B_{6R} \times \mathbb{R}^d} |y - x|^2 d\pi + \mathcal{D}_{6R} \ll 1. \quad (2.A.1)$$

Then

(i) $(B_{\lambda R} \times B_{2\lambda R}) \cap \text{Spt } \pi \neq \emptyset$ provided $\lambda \gg \left(\frac{1}{R^{d+2}} \int_{B_{6R} \times \mathbb{R}^d} |y - x|^2 d\pi + \mathcal{D}_{6R} \right)^{\frac{1}{d+2}}$.

(ii) For any $x \in B_{5R}$ and $e \in S^{d-1}$ we have that $(S_R(x, e) \times B_{7R}) \cap \text{Spt } \pi \neq \emptyset$, where

$$S_R(x, e) := C(x, e) \cap (B_R(x) \setminus B_{\frac{R}{2}}(x))$$

is the intersection of the annulus $B_R(x) \setminus B_{\frac{R}{2}}(x)$ with the spherical cone $C(x, e)$ of opening angle $\frac{\pi}{2}$ with apex at x and axis along e , see Figure 2.2.

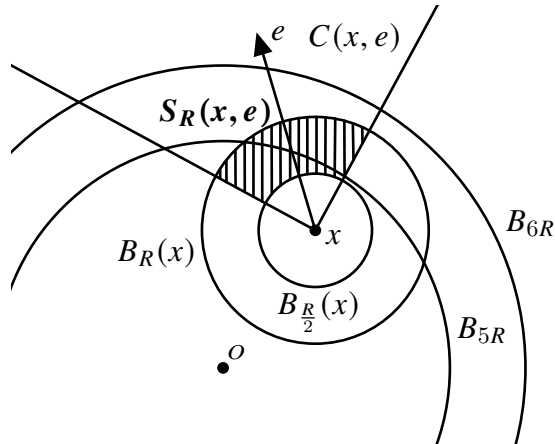


FIGURE 2.2: The definition of $S_R(x, e)$ in Lemma 2.A.1.

Remark 2.A.2. If (2.A.1) is replaced by $\frac{1}{R^{d+2}} \int_{\mathbb{R}^d \times B_{6R}} |y - x|^2 d\pi + \mathcal{D}_{6R} \ll 1$, then the symmetric results hold, namely $(B_{2\lambda R} \times B_{\lambda R}) \cap \text{Spt } \pi \neq \emptyset$ and $(B_{7R} \times S_R(y, e)) \cap \text{Spt } \pi \neq \emptyset$ for all $y \in B_{5R}$ and $e \in S^{d-1}$.

Proof. To lighten the notation in the proof, let us set

$$\mathcal{E}_{6R}^+ := \frac{1}{R^{d+2}} \int_{B_{6R} \times \mathbb{R}^d} |y - x|^2 d\pi. \quad (2.A.2)$$

We start with the more delicate statement (ii). Let $x \in B_{5R}$ and $e \in S^{d-1}$, note that

$$\begin{aligned}\pi(S_R(x, e) \times B_{7R}) &= \pi(S_R(x, e) \times \mathbb{R}^d) - \pi(S_R(x, e) \times B_{7R}^c) \\ &= \mu(S_R(x, e)) - \pi(S_R(x, e) \times B_{7R}^c).\end{aligned}$$

Since $S_R(x, e) \subseteq B_{6R}$, we have

$$\pi(S_R(x, e) \times B_{7R}^c) \lesssim \frac{1}{R^2} \int_{B_{6R} \times B_{7R}^c} \frac{1}{2} |x' - y'|^2 \pi(dx' dy') \stackrel{(2.A.2)}{\lesssim} R^d \mathcal{E}_{6R}^+.$$

To estimate $\mu(S_R(x, e))$ from below, let η be a smooth cut-off function equal to one on a ball of radius $\frac{r}{2}$ and zero outside a concentric ball of radius r satisfying

$$\sup |\eta| + r \sup |\nabla \eta| + r^2 \sup |\nabla^2 \eta| \lesssim 1, \quad (2.A.3)$$

and such that $\text{Spt } \eta \subseteq S_R(x, e) \subseteq B_{6R}$, which is possible provided $r \leq \frac{R}{4}$. Then by (2.A.3)

$$\begin{aligned}\mu(S_R(x, e)) &\geq \int_{B_{6R}} \eta d\mu = \int_{B_{6R}} \eta \kappa_\mu dx + \int_{B_{6R}} \eta (d\mu - \kappa_\mu dx) \\ &\geq \kappa_\mu \left(\frac{r}{2}\right)^d - \left| \int_{B_{6R}} \eta (d\mu - \kappa_\mu dx) \right|.\end{aligned}$$

We now use (2.2.9) with $\zeta = \eta$ to get, by the definition (2.1.9) of \mathcal{D}_{6R} , and since $\kappa_\mu \sim 1$ by (2.A.1), that

$$\left| \int_{B_{6R}} \eta (d\mu - \kappa_\mu dx) \right| \lesssim r^{\frac{d-2}{2}} R^{\frac{d+2}{2}} \mathcal{D}_{6R}^{\frac{1}{2}} + r^{-2} R^{d+2} \mathcal{D}_{6R}.$$

Hence,

$$\frac{\mu(S_R(x, e))}{R^d} \gtrsim \left(\frac{r}{R}\right)^d \left(1 - \left(\left(\frac{r}{R}\right)^{-(d+2)} \mathcal{D}_{6R}\right)^{\frac{1}{2}} - \left(\frac{r}{R}\right)^{-(d+2)} \mathcal{D}_{6R}\right).$$

We may now choose $r = \frac{R}{4}$ so that

$$\frac{\mu(S_R(x, e))}{R^d} \gtrsim 1 - \left(4^{d+2} \mathcal{D}_{6R}\right)^{\frac{1}{2}} - 4^{d+2} \mathcal{D}_{6R},$$

from which we conclude that $\pi(S_R(x, e) \times B_{7R})$ is strictly positive if \mathcal{D}_{6R} and \mathcal{E}_{6R}^+ are small enough.

In order to prove (i) we run a similar argument to obtain

$$\frac{\pi(B_{\lambda R} \times B_{2\lambda R})}{R^d} \gtrsim \lambda^d \left(1 - \left(\lambda^{-(d+2)} (\mathcal{E}_{6R}^+ + \mathcal{D}_{6R})\right)^{\frac{1}{2}} - \lambda^{-(d+2)} (\mathcal{E}_{6R}^+ + \mathcal{D}_{6R})\right).$$

Hence $(B_{\lambda R} \times B_{2\lambda R}) \cap \text{Spt } \pi \neq \emptyset$ provided that $\lambda \gg (\mathcal{E}_{6R}^+ + \mathcal{D}_{6R})^{\frac{1}{d+2}}$. ■

The next lemma, which is quite elementary, relates the support of a measure and the support of its push forward under an affine transformation:

Lemma 2.A.3. *Let γ be a measure on \mathbb{R}^n and set $\tilde{\gamma} := F_{\#}\gamma$, where $F(x) := Ax + b$ with $A \in \mathbb{R}^{n \times n}$ invertible and $b \in \mathbb{R}^n$. Then*

$$\tilde{x} \in \text{Spt } \tilde{\gamma} \quad \Leftrightarrow \quad F^{-1}(\tilde{x}) \in \text{Spt } \gamma.$$

Proof. Let $\tilde{x} \in \text{Spt } \tilde{\gamma}$. Then for all $\varepsilon > 0$

$$\tilde{\gamma}(B_{\varepsilon}(\tilde{x})) = \gamma(F^{-1}(B_{\varepsilon}(\tilde{x}))) > 0.$$

Now

$$\begin{aligned} F^{-1}(B_{\varepsilon}(\tilde{x})) &= \{x : |\tilde{x} - F(x)| < \varepsilon\} = \{x : |(\tilde{x} - b) - Ax| < \varepsilon\} \\ &\subseteq \{x : |A^{-1}(\tilde{x} - b) - x| < |A^{-1}|\varepsilon\} = B_{|A^{-1}|\varepsilon}(F^{-1}(\tilde{x})), \end{aligned}$$

so that for all $\varepsilon' > 0$ we have $\gamma(B_{\varepsilon'}(F^{-1}(\tilde{x}))) > 0$. The other implication follows analogously. \blacksquare

2.A.2 Bound on \mathcal{D}_R

In this subsection we show how the quantity $\mathcal{D}_R(\rho_0, \rho_1)$ can be bounded in terms of the Hölder semi-norms of the densities ρ_0, ρ_1 :

Lemma 2.A.4. *Let $\rho_0, \rho_1 \in C^{0,\alpha}$, $\alpha \in (0, 1)$, be two probability densities with bounded support, and such that $\frac{1}{2} \leq \rho_j \leq 2$, $j = 0, 1$, on their support. If $\rho_0(0) = \rho_1(0) = 1$, then for all $R > 0$ such that $B_R \subseteq \text{Spt } \rho_j$, $j = 0, 1$, we have*

$$\mathcal{D}_R \lesssim R^{2\alpha} \left([\rho_0]_{\alpha,R}^2 + [\rho_1]_{\alpha,R}^2 \right). \quad (2.A.4)$$

Proof. By the definition (2.1.10) of $\kappa_j := \kappa_{\rho_j}$ and using $\rho_j(0) = 1$, Jensen's inequality implies

$$|\kappa_j - 1|^2 = |\kappa_j - \rho_j(0)|^2 \leq \frac{1}{|B_R|} \int_{B_R} |\rho_j(x) - \rho_j(0)|^2 dx \leq [\rho_j]_{\alpha,R}^2 R^{2\alpha}.$$

If $R > 0$ is such that $B_R \subseteq \text{Spt } \rho_j$, the assumption that ρ_j is bounded away from zero on its support implies that $\kappa_j \gtrsim 1$. The claimed inequality (2.A.4) then follows with Lemma 2.A.5. \blacksquare

Lemma 2.A.5. *Let $\rho \in C^{0,\alpha}$, $\alpha \in (0, 1)$, be a density with bounded support, and such that $\frac{1}{2} \leq \rho \leq 2$ on its support. Then*

$$\frac{1}{R^{d+2}} W_{B_R}^2(\rho, \kappa) \lesssim R^{2\alpha} [\rho]_{\alpha,R}^2.$$

Proof. By the assumptions on ρ , the 2-Wasserstein distance between $\rho \, dx$ and $\kappa \, dx$ restricted to B_R can be bounded by

$$W_{B_R}^2(\rho, \kappa) \lesssim \|\nabla\Psi\|_{L^2(B_R)}^2,$$

where Ψ is the mean-zero solution of the Neumann problem

$$\begin{cases} -\Delta\Psi = \rho - \kappa & \text{in } B_R, \\ \nabla\Psi \cdot \nu = 0 & \text{on } \partial B_R, \end{cases} \quad (2.A.5)$$

see [San15, Theorem 5.34]. Global Schauder theory²⁶ for (2.A.5) implies that

$$\frac{1}{R^2} \sup_{B_R} |\nabla\Psi|^2 \lesssim \|\rho - \kappa\|_{C^0(B_R)}^2 + R^{2\alpha} [\rho - \kappa]_{\alpha, R}^2 \lesssim R^{2\alpha} [\rho]_{\alpha, R}^2,$$

so that

$$W_{B_R}^2(\rho, \kappa) \lesssim \|\nabla\Psi\|_{L^2(B_R)}^2 \lesssim R^d \sup_{B_R} |\nabla\Psi|^2 \lesssim R^{d+2} R^{2\alpha} [\rho]_{\alpha, R}^2.$$

■

2.A.3 A property of semi-convex functions

Lemma 2.A.6. *Let $C \in \mathbb{R}$ and assume that $v : \mathbb{R}^d \rightarrow \mathbb{R}$ is C -semi-convex, that is, $x \mapsto v(x) + \frac{C}{2}|x|^2$ is convex. Then*

$$\|\nabla v\|_{L^\infty(B_1)} \lesssim \|v\|_{C^0(B_3)}^{\frac{1}{2}} \max\{\|v\|_{C^0(B_3)}^{\frac{1}{2}}, C^{\frac{1}{2}}\}. \quad (2.A.6)$$

Proof. Notice first that by semi-convexity, the gradient of v exists a.e. Convexity of $x \mapsto v(x) + \frac{C}{2}|x|^2$ implies that for a.e. y and all x ,

$$\nabla v(y) \cdot (x - y) \leq v(x) - v(y) + \frac{C}{2}|x - y|^2.$$

In particular, for a.e. $y \in B_1$ and every $x \in B_3$ lying in a cone of opening angle $\frac{2\pi}{3}$ with apex at y and axis along $\nabla v(y)$, so that $\nabla v(y) \cdot (x - y) \geq \frac{1}{2}|\nabla v(y)||x - y|$, we have

$$|\nabla v(y)| \lesssim \frac{\|v\|_{C^0(B_3)}}{|x - y|} + C|x - y|.$$

Optimizing in $|x - y|$ then gives (2.A.6). ■

²⁶See for instance [Tro87, Theorem 3.16] for details.

2.B The change of coordinates Lemma 2.1.15

Proof of Lemma 2.1.15. First, we clearly have $\widehat{\rho}_0(0) = \widehat{\rho}_1(0) = 1$ and $\nabla_{\widehat{x}\widehat{y}}\widehat{c}(0,0) = -\mathbb{I}$. It is also easy to check that $\widehat{\pi} \in \Pi(\widehat{\rho}_0, \widehat{\rho}_1)$. Let us now compute

$$\begin{aligned} \int \widehat{c}(\widehat{x}, \widehat{y}) \widehat{\pi}(d\widehat{x}d\widehat{y}) &= \frac{\gamma|\det B|}{\rho_0(0)} \int c(Q^{-1}(\widehat{x}, \widehat{y})) (Q_{\#}\pi)(d\widehat{x}d\widehat{y}) \\ &= \frac{\gamma|\det B|}{\rho_0(0)} \int c(x, y) \pi(dxdy). \end{aligned}$$

Thus, \widehat{c} -optimality of $\widehat{\pi}$ is equivalent to c -optimality of π . Indeed, if $\widehat{\pi}$ is not optimal for the cost \widehat{c} then one can find a coupling $\widehat{\sigma} \in \Pi(\widehat{\rho}_0, \widehat{\rho}_1)$ such that $\int \widehat{c}(\widehat{x}, \widehat{y}) \widehat{\sigma}(d\widehat{x}d\widehat{y}) < \int \widehat{c}(\widehat{x}, \widehat{y}) \widehat{\pi}(d\widehat{x}d\widehat{y})$. Defining now $\sigma := \rho_0(0)|\det B|^{-1} (Q^{-1})_{\#}\widehat{\sigma} \in \Pi(\rho_0, \rho_1)$, the previous computation would yield

$$\int c(x, y) \sigma(dxdy) < \int c(x, y) \pi(dxdy),$$

a contradiction. ■

Remark 2.B.1. It is also possible to prove the \widehat{c} -optimality of $\widehat{\pi}$ by showing that $\text{Spt } \widehat{\pi}$ is \widehat{c} -cyclically monotone, which characterizes optimality (see for instance [San15, Theorem 1.49]). This property readily follows from Lemma 2.A.3 and the c -cyclical monotonicity of $\text{Spt } \pi$.

2.C Some aspects of Campanato's theory

Lemma 2.C.1. *Let $R > 0$, $\frac{1}{2} \leq \rho_0 \leq 2$ on B_R , and assume that the coupling $\pi \in \Pi(\rho_0, \rho_1)$ satisfies (2.4.31) for $\alpha \in (0, 1)$. For $r < \frac{R}{2}$ let $A_r = A_r(x_0, \pi) \in \mathbb{R}^{d \times d}$ and $b_r = b_r(x_0, \pi) \in \mathbb{R}^d$ be the (unique) minimizers of*

$$\inf_{A, b} \int_{(B_r(x_0) \cap B_R) \times \mathbb{R}^d} |y - (Ax + b)|^2 d\pi.$$

Then there exist $A_0 \in \mathbb{R}^{d \times d}$ and $b_0 \in \mathbb{R}^d$ such that $A_r \rightarrow A_0$ and $b_r \rightarrow b_0$ uniformly in x_0 and the estimates (2.4.62) hold.

Proof of Lemma 2.C.1. We only give the proof for A_r , as the one for b_r is analogous. Without loss of generality we may assume that r is small enough such that $B_r(x_0) \subset B_R$. For further reference in the proof, let us mention here the following estimates obtained by equivalence of the L^∞ and L^2 norms in the set of polynomials of degree one: Let $P(x) = Ax + b$ for some $A \in \mathbb{R}^{d \times d}$ and $b \in \mathbb{R}^d$, then for any $x_0 \in \mathbb{R}^d$ and $r > 0$ there holds

$$|A|^2 \lesssim \frac{1}{r^{d+2}} \int_{B_r(x_0)} |P(x)|^2 dx, \quad \text{and} \quad |b|^2 \lesssim \frac{1}{r^d} \int_{B_r(x_0)} |P(x)|^2 dx. \quad (2.C.1)$$

STEP 1 Define

$$P_r(x) := A_r x + b_r. \quad (2.C.2)$$

We claim that for any $k \in \mathbb{N}_0$ there holds

$$\int_{B_{r2^{-k-1}}(x_0)} |P_{r2^{-k}}(x) - P_{r2^{-k-1}}(x)|^2 \rho_0(x) dx \lesssim [[\pi]]_\alpha^2 2^{-k(d+2+2\alpha)} r^{d+2+2\alpha}. \quad (2.C.3)$$

Indeed, since $\pi \in \Pi(\rho_0, \rho_1)$ and $B_{r2^{-k-1}}(x_0) \subset B_{r2^{-k}}(x_0)$, we may estimate

$$\begin{aligned} \int_{B_{r2^{-k-1}}(x_0)} |P_{r2^{-k}}(x) - P_{r2^{-k-1}}(x)|^2 \rho_0(x) dx &= \int_{B_{r2^{-k-1}}(x_0) \times \mathbb{R}^d} |P_{r2^{-k}}(x) - P_{r2^{-k-1}}(x)|^2 d\pi \\ &\lesssim \int_{B_{r2^{-k}}(x_0) \times \mathbb{R}^d} |y - P_{r2^{-k}}(x)|^2 d\pi + \int_{B_{r2^{-k-1}}(x_0) \times \mathbb{R}^d} |y - P_{r2^{-k-1}}(x)|^2 d\pi, \end{aligned}$$

so that by the definition of A_r and b_r as minimizers and the definition (2.4.31) of $[[\pi]]_\alpha$, the bound (2.C.3) easily follows.

STEP 2 We claim that for any $i \in \mathbb{N}$,

$$|A_r - A_{r2^{-i}}| \lesssim [[\pi]]_\alpha \sum_{k=0}^{i-1} 2^{-k\alpha} r^\alpha. \quad (2.C.4)$$

Indeed, writing the difference as a telescopic sum, we may apply (2.C.1) to the polynomial $P_{r2^{-k}} - P_{r2^{-k-1}}$ to obtain

$$|A_r - A_{r2^{-i}}| \leq \sum_{k=0}^{i-1} |A_{r2^{-k}} - A_{r2^{-k-1}}| \lesssim \sum_{k=0}^{i-1} \left(2^{(k+1)(d+2)} \frac{1}{r^{d+2}} \int_{B_{r2^{-k-1}}(x_0)} |P_{r2^{-k}} - P_{r2^{-k-1}}|^2 dx \right)^{\frac{1}{2}},$$

so using that $\frac{1}{2} \leq \rho_0 \leq 2$ on $B_{r2^{-k-1}}(x_0)$, the claim follows with (2.C.3).

STEP 3 We show next that the sequence $\{A_{r2^{-i}}\}_{i \in \mathbb{N}}$ converges as $i \rightarrow \infty$ to a limit A_0 independent of r .

Indeed, for any $i > j$ we may use (2.C.4) to estimate

$$|A_{r2^{-j}} - A_{r2^{-i}}| \lesssim [[\pi]]_\alpha \sum_{k=j}^{i-1} 2^{-k\alpha} r^\alpha \xrightarrow{i,j \rightarrow \infty} 0,$$

since the series $\sum_{k=0}^{\infty} 2^{-k\alpha}$ converges. Hence, the sequence $\{A_{r2^{-i}}\}_{i \in \mathbb{N}}$ is Cauchy and there exists $A_0 \in \mathbb{R}^{d \times d}$ such that $A_{r2^{-i}} \rightarrow A_0$ as $i \rightarrow \infty$.

To see the independence of the limit of r , let $0 < r < \rho < \frac{R}{2}$ be small enough. Then applying (2.C.1) to the function $P_{r2^{-j}} - P_{\rho 2^{-j}}$ for $j \in \mathbb{N}$ gives

$$|A_{r2^{-j}} - A_{\rho 2^{-j}}|^2 \lesssim 2^{j(d+2)} r^{-d-2} \int_{B_{r2^{-j}}(x_0)} |P_{r2^{-j}}(x) - P_{\rho 2^{-j}}(x)|^2 dx,$$

which can be bounded similarly to STEP 1 to yield

$$|A_{r2^{-j}} - A_{\rho 2^{-j}}|^2 \lesssim 2^{-2j\alpha} r^{2\alpha} [[\pi]]_\alpha^2 \left(1 + \left(\frac{\rho}{r} \right)^{d+2+2\alpha} \right) \xrightarrow{j \rightarrow \infty} 0.$$

The claimed inequality (2.4.62) now follows easily by letting $i \rightarrow \infty$ in (2.C.4). ■

2.D An L^p bound on the displacement for almost-minimizing transport maps

In this section we give a proof of the interior L^p estimate for arbitrary $p < \infty$ on the displacement of almost-minimizing transport maps:

Proposition 2.D.1. *Assume μ has a $C^{0,\alpha}$ density f satisfying $f(0) = 1$. Let T be an almost-minimizing transport map from μ to ν in the sense of (2.1.54) with a rate function $\Delta_r \leq 1$. Then there exists $R_1 > 0$ such that, if*

$$\frac{1}{R_1^{d+2}} \int_{B_{R_1}} |T(x) - x|^2 \mu(dx) + R_1^\alpha [f]_{\alpha, R_1} \ll 1, \quad (2.D.1)$$

then for any $p < \infty$,

$$\left(\int_{B_{\frac{R_1}{4}}} |T(x) - x|^p \mu(dx) \right)^{\frac{1}{p}} \lesssim_p R_1^{\frac{d}{p}} \left(\int_{B_{R_1}} |T(x) - x|^2 \mu(dx) \right)^{\frac{1}{d+2}}. \quad (2.D.2)$$

Proof. STEP 1 (A smooth transport map from μ to the Lebesgue measure). We fix an open ball B_0 centered at 0 such that $B_0 \subset \text{Spt } f$. Let S be the Dacorogna-Moser transport map from $\mu|_{B_0}$ to $\frac{\mu(B_0)}{|B_0|} dx|_{B_0}$, which is $C^{1,\alpha}$ on B_0 , see for example [Vil09, Chapter 1]. By an affine change of variables in the target space, we may define for any $x_0 \in X$ a map S_{x_0} with the same regularity as S that pushes forward $\mu|_{B_0}$ to $f(x_0) dx|_{\tilde{B}_0}$, where \tilde{B}_0 is the modified target space under this affine transformation, such that

$$S_{x_0}(x_0) = x_0 \quad \text{and} \quad DS_{x_0}(x_0) = \mathbb{I}.$$

In particular, for all $x \in B_0$,

$$|S_{x_0}(x) - x| \leq [DS_{x_0}]_\alpha |x - x_0|^{1+\alpha} \leq C|x - x_0|^{1+\alpha},$$

where C depends on S and therefore on f through the global Hölder semi-norm $[f]_\alpha$. Moreover, since S^{-1} has the same regularity as S , $S_{x_0}^{-1}$ is $C^{1,\alpha}$ and for all $y \in B_R(x_0)$,

$$|S_{x_0}^{-1}(y) - y| \leq C|y - x_0|^{1+\alpha}. \quad (2.D.3)$$

Letting $R_1 > 0$ small enough so that first, $B_{R_1} \subset B_0 \subset \text{Spt } f$, and second, $CR_1^\alpha \leq 1$, we may therefore assume, writing $S_{x_0}(x) - x_0 = S_{x_0}(x) - x + x - x_0$ and $S_{x_0}^{-1}(y) - x_0 = S_{x_0}^{-1}(y) - y + y - x_0$, that for all $x_0 \in B_{R_1}$,

$$|S_{x_0}(x) - x_0| \leq 2|x - x_0| \quad \text{and} \quad |S_{x_0}^{-1}(y) - x_0| \leq 2|y - x_0|, \quad \text{for all } x, y \in B_{R_1}. \quad (2.D.4)$$

We note that R_1 depends on f only through $[f]_\alpha$. In view of (2.D.1) and the condition $f(0) = 1$, we may assume that

$$\frac{1}{2} \leq f \leq 2 \quad \text{on} \quad B_{R_1}. \quad (2.D.5)$$

STEP 2 (Use of almost-minimality at scale R). Let us denote the displacement by $u := T - \text{Id}$. We claim that if R_1 is small enough, for any ball $B := B_{12R}(x_0) \subset B_{R_1}$ and any set $A \subset B_R(x_0)$, we have²⁷

$$\int_A |u| \lesssim R^{d+1} + |A| \int_B |u|. \quad (2.D.6)$$

To see this, we start from the following pointwise identity, that holds for any map Φ ,

$$\begin{aligned} & (T(x) - x) \cdot (\Phi(x) - x) + (T(\Phi(x)) - \Phi(x)) \cdot (x - \Phi(x)) \\ &= \frac{1}{2} (|T(x) - x|^2 + |T(\Phi(x)) - \Phi(x)|^2) - \frac{1}{2} (|T(\Phi(x)) - x|^2 + |T(x) - \Phi(x)|^2) + |\Phi(x) - x|^2, \end{aligned} \quad (2.D.7)$$

which is (a deformation of) a standard identity used to derive (c -)monotonicity of an optimal transport map. As usual in optimal transportation, such a monotonicity property follows from considering a competitor that swaps points in the target space. However, because of our assumption of almost-minimality, we cannot really work with a pointwise argument but have to consider small sets A of positive measure. We will do this by applying (2.D.7) with a map Φ that swaps some parts of the graph of T . Indeed, assuming Φ is a μ -preserving involution such that $\Phi = \text{Id}$ outside of a set of diameter of order R , we may integrate (2.D.7) with respect to μ , so that the identities

$$\begin{aligned} \int |T(\Phi(x)) - \Phi(x)|^2 \mu(dx) &= \int |T(x) - x|^2 \mu(dx), \\ \int |T(x) - \Phi(x)|^2 \mu(dx) &= \int |\tilde{T}(x) - x|^2 \mu(dx) \\ \text{and } \int (T(\Phi(x)) - \Phi(x)) \cdot (x - \Phi(x)) \mu(dx) &= \int (T(x) - x) \cdot (\Phi(x) - x) \mu(dx), \end{aligned}$$

combined with almost-minimality of T at a scale of order R , yield

$$\begin{aligned} \int (T(x) - x) \cdot (\Phi(x) - x) \mu(dx) &\lesssim R^{d+2} + \int |\Phi(x) - x|^2 \mu(dx) \\ &\lesssim R^{d+2} + R^2 \mu(B) \stackrel{(2.D.5)}{\lesssim} R^{d+2}. \end{aligned} \quad (2.D.8)$$

To implement this, we give ourselves a direction $e \in B_2 \setminus B_1$, swap the (images through S of the)²⁸ sets

$$A_+ := \{x \in A \mid u(x) \cdot e \geq 0\} \quad \text{and} \quad A_- := A \setminus A_+,$$

with their translates along e or $-e$, and average over all directions. For this we first define a map ϕ via

$$\phi(x) := \begin{cases} x + 2Re & \text{for } x \in S(A_+), \\ x - 2Re & \text{for } x \in S(A_-), \\ x - 2Re & \text{for } x \in S(A_+) + 2Re, \\ x + 2Re & \text{for } x \in S(A_-) - 2Re, \\ x & \text{otherwise.} \end{cases} \quad (2.D.9)$$

²⁷We use the notation \int for the average of a function on a set, with respect to the Lebesgue measure.

²⁸In the rest of STEP 2, we drop the index x_0 for the map S_{x_0} . Now, S is the map such that $S(x_0) = x_0$ and $DS(x_0) = \mathbb{I}$.

Notice that ϕ is well-defined because, in view of (2.D.4), we have $S(A) \subset B_{2R}(x_0)$. Furthermore, we see that $\phi = \text{Id}$ outside of $B_{6R}(x_0)$, and that ϕ swaps $S(A_+)$ and its translate $S(A_+) + 2Re$ and swaps $S(A_-)$ and its translate $S(A_-) - 2Re$, so that ϕ is an involution; in particular, ϕ preserves the Lebesgue measure. Combining the second point with the inclusion $B_{6R}(x_0) \subset S_{x_0}(B)$ from (2.D.4), we get that $\Phi := S^{-1} \circ \phi \circ S$ is identity outside of B . From the third point, we see that Φ is also an involution and $\Phi_{\#}\mu = \mu$, and it follows that the map $\tilde{T} := T \circ \Phi$ satisfies $\tilde{T}_{\#}\mu = \nu$ and $\tilde{T} = T$ outside of B . Thus, (2.D.8) holds.

Now, recalling that $u = T - \text{Id}$,

$$\int (T(x) - x) \cdot (\Phi(x) - x) \mu(dx) = \int u(S^{-1}(y)) \cdot (S^{-1}(\phi(y)) - S^{-1}(y)) f(x_0) dy,$$

so using the fact that $\phi = \text{Id}$ everywhere except on $S(A)$ and $(S(A_+) + 2Re) \cup (S(A_-) - 2Re)$, we have two terms to estimate. From the identity

$$\int_{S(A)} u(S^{-1}(y)) \cdot (\phi(y) - y) f(x_0) dy = 2R \int_A |u \cdot e| d\mu,$$

and (2.D.3), we first have

$$\begin{aligned} & \left| \int_{S(A)} u(S^{-1}(y)) \cdot (S^{-1}(\phi(y)) - S^{-1}(y)) f(x_0) dy - 2R \int_A |u \cdot e| d\mu \right| \\ & \leq C \int_{S(A)} |u(S^{-1}(y))| (|\phi(y) - x_0|^{1+\alpha} + |y - x_0|^{1+\alpha}) f(x_0) dy \\ & \leq CR^{1+\alpha} \int_A |u| d\mu. \end{aligned} \quad (2.D.10)$$

We then estimate, using (2.D.3) and (2.D.5),

$$\begin{aligned} & \left| \int_{(S(A_+) + 2Re) \cup (S(A_-) - 2Re)} u(S^{-1}(y)) \cdot (S^{-1}(\phi(y)) - S^{-1}(y)) f(x_0) dy \right| \\ & \lesssim R \int_{(S(A_+) + 2Re) \cup (S(A_-) - 2Re)} |u(S^{-1}(y))| dy, \end{aligned} \quad (2.D.11)$$

so that combining (2.D.8), (2.D.10) and (2.D.11) gives

$$\int_A |u \cdot e| d\mu - CR^\alpha \int_A |u| d\mu \lesssim R^{d+1} + \int_{(S(A_+) + 2Re) \cup (S(A_-) - 2Re)} |u(S^{-1}(y))| dy. \quad (2.D.12)$$

It remains to integrate (2.D.12) with respect to $e \in B_2 \setminus B_1$, using that

$$\int_{B_2 \setminus B_1} |u(x) \cdot e| de \gtrsim |u(x)| \quad (2.D.13)$$

and the equivalence

$$x \in (S(A_+) + 2Re) \cup (S(A_-) - 2Re) \iff e \in \frac{1}{2R}(x - S(A_+)) \cup \frac{1}{2R}(x + S(A_-)),$$

which gives

$$\int_{B_2 \setminus B_1} 1_{(S(A_+) + 2Re) \cup (S(A_-) - 2Re)}(x) \, de \leq \left| \frac{1}{2R}(x - S(A_+)) \cup \frac{1}{2R}(x + S(A_-)) \right| \lesssim \frac{|S(A)|}{|B|}. \quad (2.D.14)$$

From (2.D.13), (2.D.14) and the inclusion

$$(S(A_+) + 2Re) \cup (S(A_-) - 2Re) \subset B_{6R}(x_0),$$

we obtain

$$\begin{aligned} \int_{B_2 \setminus B_1} \int_{(S(A_+) + 2Re) \cup (S(A_-) - 2Re)} |u(S^{-1}(y))| \, dy \, de &\stackrel{(2.D.5)}{\lesssim} \frac{|S(A)|}{|B|} \int_{B_{6R}(x_0)} |u(S^{-1}(y))| f(x_0) \, dy \\ &= \frac{|S(A)|}{|B|} \int_{S^{-1}(B_{6R}(x_0))} |u| \, d\mu \\ &\stackrel{(2.D.4)}{\leq} \frac{|S(A)|}{|B|} \int_B |u| \, d\mu. \end{aligned}$$

Finally, the transport condition $|\det DS| = \frac{f}{f(x_0)}$ ensures that we have $|S(A)| \lesssim |A|$, so that (2.D.12) yields

$$(1 - CR^\alpha) \int_A |u| \, d\mu \lesssim R^{d+1} + \frac{|A|}{|B|} \int_B |u| \, d\mu.$$

Choosing R_1 so that $CR_1^\alpha \leq \frac{1}{2}$ and using (2.D.5), we obtain (2.D.6). We note again that the choice of R_1 depends on f only through $[f]_\alpha$.

STEP 3 We claim that for any set $A \subset B_{\frac{R_1}{2}}$ with $\text{diam } A \leq R \ll R_1$, for any $\beta > 0$, we have

$$\int_A |u| \lesssim_\beta R^{d+1} + \frac{|A|}{R^\beta} \left(\int_{B_{R_1}} |u|^2 \right)^{\frac{1+\beta}{d+2}}. \quad (2.D.15)$$

STEP 3.A. Let us show that for all $\beta > 0$ and for any two balls $B_r(x_0) \subset B_R(x_0) \subset B_{R_1}$,

$$\int_{B_r(x_0)} |u| \lesssim_\beta \left(\frac{R}{r} \right)^\beta \left(R + \int_{B_R(x_0)} |u| \right). \quad (2.D.16)$$

To this end, we momentarily introduce

$$\Lambda := \sup_{r \leq \rho \leq R} \rho^\beta \int_{B_\rho(x_0)} |u|.$$

Applying STEP 2 to the sets $A := B_\rho(x_0)$ and $B := B_{M\rho}(x_0)$, we obtain, provided $12\rho \leq M\rho \leq R$,

$$\int_{B_\rho(x_0)} |u| \lesssim M^{d+1} \rho + \int_{B_{M\rho}} |u|,$$

so that, provided $r \leq \rho \leq \frac{R}{M}$,

$$\rho^\beta \int_{B_\rho(x_0)} |u| \lesssim M^{d-\beta} R^{1+\beta} + \frac{\Lambda}{M^\beta}.$$

Fixing a large enough $M \sim 1$, this yields in the range $r \leq \rho \leq \frac{R}{M}$,

$$\rho^\beta \int_{B_\rho(x_0)} |u| - \frac{\Lambda}{2} \lesssim R^{1+\beta}. \quad (2.D.17)$$

In the remaining range $\frac{R}{M} \leq \rho \leq R$, we have

$$\rho^\beta \int_{B_\rho(x_0)} |u| \lesssim_\beta R^\beta \int_{B_R(x_0)} |u|,$$

so that (2.D.17) turns into

$$\frac{\Lambda}{2} \lesssim R^{1+\beta} + R^\beta \int_{B_R(x_0)} |u|,$$

which is (2.D.16).

STEP 3.B. We prove (2.D.15).

By applying Cauchy-Schwarz to (2.D.16), we obtain

$$\int_{B_r(x_0)} |u| \lesssim \left(\frac{R}{r}\right)^\beta \left(R + \frac{1}{R^{\frac{d}{2}}} \int_{B_{R_1}} |u|^2\right),$$

so that, optimizing in R by choosing $R = \left(\int_{B_{R_1}} |u|^2\right)^{\frac{1}{d+2}} \ll R_1$ by (2.D.1), we get, provided $B_r(x_0) \subset B_{R_1}$,

$$r^\beta \int_{B_r(x_0)} |u| \lesssim \left(\int_{B_{R_1}} |u|^2\right)^{\frac{1+\beta}{d+2}}.$$

In combination with STEP 2, this yields (2.D.15).

STEP 4 (Conclusion). We now have all the ingredients to prove the estimate (2.D.2).

Given a threshold $M < \infty$ and a ball $B \subset B_{\frac{R_1}{2}}$ of radius $R \ll R_1$, we apply STEP 3 to $A := \{|u| > M\} \cap B$, to the effect that

$$MR^\beta \gg \left(\int_{B_{R_1}} |u|^2\right)^{\frac{1+\beta}{d+2}} \implies M|A| \lesssim R^{d+1}. \quad (2.D.18)$$

Provided $M \geq \left(\int_{B_{R_1}} |u|^2\right)^{\frac{1}{d+2}}$, thanks to (2.D.1), there are radii $R \ll R_1$ for which the statement on the left-hand side of (2.D.18) holds. Hence, by covering $B_{\frac{R_1}{4}}$ by these balls, we obtain

$$M \left| \{|u| \geq M\} \cap B_{\frac{R_1}{4}} \right| \lesssim R_1^d \left(\frac{1}{M} \left(\int_{B_{R_1}} |u|^2\right)^{\frac{1+\beta}{d+2}} \right)^{\frac{1}{\beta}}. \quad (2.D.19)$$

Note that (2.D.19) is trivially satisfied for $M \leq \left(\int_{B_{R_1}} |u|^2 \right)^{\frac{1}{d+2}}$, so that with $p := 1 + \frac{1}{\beta}$,

$$M \left| \{ |u| \geq M \} \cap B_{\frac{R_1}{4}} \right|^{\frac{1}{p}} \lesssim R_1^{\frac{d}{p}} \left(\int_{B_{R_1}} |u|^2 \right)^{\frac{1}{d+2}}$$

holds for all M . This amounts to an estimate in the weak L^p norm of u on $B_{\frac{R_1}{4}}$. Because of (2.D.1), we trivially have

$$\left(\int_{B_{\frac{R_1}{4}}} |u|^2 \right)^{\frac{1}{2}} \lesssim R_1^{\frac{d}{2}} \left(\int_{B_{R_1}} |u|^2 \right)^{\frac{1}{d+2}},$$

so by interpolation, we obtain (2.D.2). ■

CHAPITRE 3

Vitesse de convergence optimale vers la loi circulaire pour l'ensemble de Ginibre

Sommaire

3.1	Introduction	125
3.2	Estimation de la variance du nombre de points	126
3.2.1	Rectangle dans $D(0, \sqrt{N} - \sqrt{\ln N})$	129
3.2.2	Disque de rayon inférieur à $\sqrt{N} - \sqrt{\ln N}$	132
3.2.3	Arc sur le bord du disque	134
3.2.4	Intersection d'un rectangle avec un disque	135
3.3	Optimal matching dans un carré	138
3.3.1	Estimation de la distance	138
3.3.2	Concentration : preuve du Théorème IV.3	142
3.4	Optimal matching dans un disque	145
3.4.1	Construction des mesures	145
3.4.2	Estimation de la distance	148
3.5	Optimal matching dans un anneau	153
3.6	Vitesse de convergence pour la loi du cercle	157
3.A	Lemmes techniques	159

3.1 Introduction

Soit M_N une matrice aléatoire complexe à entrées gaussiennes indépendantes, soient $\lambda_1, \dots, \lambda_N$ ses valeurs propres distinctes (l'ensemble des matrices ayant N valeurs propres distinctes est

de mesure pleine dans $\mathbb{C}^{N \times N}$). Le N -uplet $(\lambda_1, \dots, \lambda_N)$ a pour densité

$$\varphi_N(z_1, \dots, z_N) := \frac{1}{Z_N} \exp\left(-\sum_{i=1}^N |z_i|^2\right) \prod_{1 \leq i < j \leq N} |z_i - z_j|^2$$

par rapport à la mesure de Lebesgue sur \mathbb{C}^N , où

$$Z_N := \pi^N \prod_{k=1}^N k!.$$

On notera \mathbb{P} la mesure de probabilité sur $\mathbb{C}^{N \times N}$ associée à la densité φ_N et \mathbb{E} l'espérance associée.¹ Soit

$$\mu_N = \sum_{k=1}^N \delta_{\lambda_k}$$

la mesure spectrale empirique et $\nu := \frac{1}{\pi} \mathbf{1}_{\mathbb{D}_N}(z) dz$ avec $\mathbb{D}_N := D(0, \sqrt{N})$. Le but du chapitre est de prouver le Théorème III.1, que l'on redonne ici sous une forme un peu différente.

Théorème 3.1.1. *Il existe une constante $C < \infty$ telle que pour tout $N \geq 2$*

$$\mathbb{E}W_2^2(\mu_N, \nu) \leq CN.$$

Dans la suite, on utilise à nouveau la notation

$$W_A(\alpha, \beta) := W_2(\alpha|_A, \beta|_A), \tag{3.1.1}$$

pour tout ensemble A et toutes mesures α et β .

Le chapitre est organisé de la façon suivante : dans la section 3.2, on estime la variance du nombre de points dans plusieurs types de domaines qui vont nous être utiles par la suite. La section 3.3 contient l'estimation de la distance W_2 dans un carré, ainsi que la preuve du Théorème IV.3. On n'utilise pas ce résultat pour prouver le Théorème 3.1.1 mais la stratégie est la même que celle utilisée dans les sections 3.4 et 3.5, on l'expose donc d'abord dans ce cadre. Pour finir, la section 3.6 contient la preuve du Théorème 3.1.1.

3.2 Estimation de la variance du nombre de points

Le processus de points défini par μ_N a une structure déterminantale, en particulier les fonctions de k -corrélation s'écrivent ([Meh04, Ch. 15])

$$\begin{aligned} \varphi_{N,k}(z_1, \dots, z_k) &:= \int_{\mathbb{C}^{N-k}} \varphi(z_1, \dots, z_N) dz_{k+1} \cdots dz_N \\ &= \frac{(N-k)!}{\pi^k N!} \exp\left(-\sum_{i=1}^k |z_i|^2\right) \det [K(z_i, z_j)]_{1 \leq i, j \leq k}, \end{aligned}$$

¹Ces deux quantités dépendent bien sûr de N mais on ne rappelle pas cette dépendance dans la notation par souci de simplicité d'écriture.

où

$$K(z, z') := \sum_{k=0}^{N-1} \frac{(z\bar{z}')^k}{k!}.$$

Cette propriété permet en particulier d'obtenir une formule pour la variance de la variable aléatoire $\mu_N(A)$ où A est un ensemble mesurable. De plus, on sait ([HKPV06, Th. 7]) que le nombre de points d'un processus déterminantal dans un ensemble A , ici $\mu_N(A)$, suit la même loi qu'une somme de variables de Bernoulli dont les paramètres sont les valeurs propres de l'opérateur \mathcal{K} de $L^2(A)$ défini par

$$\mathcal{K}f(z) := \int_A K(z, w)f(w) dw.$$

Ceci permet d'obtenir l'inégalité de concentration ([MM14, Prop. 4])

$$\mathbb{P}(|\mu_N(A) - \mathbb{E}\mu_N(A)| \geq t) \leq \exp\left(-\min\left\{\frac{t^2}{4 \operatorname{Var} \mu_N(A)}, \frac{t}{2}\right\}\right), \quad (3.2.1)$$

valable pour tout $t \geq 0$ et tout $A \subset \mathbb{D}_N$.

Soit $A \subset \mathbb{C}$ un ensemble mesurable, on se propose d'estimer la variance de $\mu_N(A)$, c'est-à-dire les fluctuations du nombre de valeurs propres se trouvant dans A . On obtient d'abord l'espérance de cette variable aléatoire,

$$\begin{aligned} \mathbb{E}\mu_N(A) &= \int_{\mathbb{C}^N} \sum_{k=1}^N \chi_A(z_k) \varphi_N(z_1, \dots, z_N) dz_1 \cdots dz_N = N \int_A \varphi_{N,1}(z) dz \\ &= \frac{1}{\pi} \int_A e^{-|z|^2} K(z, z) dz = \frac{|A|}{\pi} - \frac{1}{\pi} \int_A e^{-|z|^2} \sum_{k=N}^{\infty} \frac{|z|^{2k}}{k!} dz, \end{aligned} \quad (3.2.2)$$

ce qui permet d'en donner une estimation qui nous sera utile.

Lemme 3.2.1. *Soit $A \subset D(0, \sqrt{N} - \sqrt{\ln N})$,*

$$\frac{|A|}{\pi} - \frac{e}{N^{\frac{3}{2}}} \leq \mathbb{E}\mu_N(A) \leq \frac{|A|}{\pi}.$$

Démonstration. Soit $r = 1 - \varepsilon_N = 1 - \sqrt{\frac{\ln N}{N}}$, on a

$$\begin{aligned} \frac{1}{\pi} \int_A e^{-|z|^2} \sum_{k=N}^{\infty} \frac{|z|^{2k}}{k!} dz &\leq \frac{1}{\pi} \int_{r\mathbb{D}_N} e^{-|z|^2} \sum_{k=N}^{\infty} \frac{|z|^{2k}}{k!} dz \\ &\leq \frac{1}{\pi} \int_{r\mathbb{D}_N} e^{-|z|^2} \left(\frac{e|z|^2}{N}\right)^N dz \\ &= \left(\frac{e}{N}\right)^N \int_0^{r^2 N} e^{-s} s^N ds \\ &= \left(\frac{e}{N}\right)^N N! e^{-r^2 N} \sum_{k=N+1}^{\infty} \frac{(r^2 N)^k}{k!} \\ &\leq e\sqrt{N} e^{N(1+2\ln r - r^2)} = e\sqrt{N} e^{N(2\ln(1-\varepsilon_N) + 2\varepsilon_N - \varepsilon_N^2)}, \end{aligned}$$

ce qui donne le résultat en utilisant l'inégalité $\ln(1 - \varepsilon) \leq -\varepsilon - \frac{\varepsilon^2}{2}$ pour $\varepsilon \in (0; 1)$. ■

Calculons maintenant

$$\begin{aligned} \mathbb{E}|\mu_N(A)|^2 &= \int_{\mathbb{C}^N} \sum_{i,j=1}^N \chi_A(z_i)\chi_A(z_j)\varphi_N(z_1, \dots, z_N) dz_1 \cdots dz_N \\ &= N \int_A \varphi_{N,1}(z) dz + N(N-1) \int_{A \times A} \varphi_{N,2}(z_1, z_2) dz_1 dz_2. \end{aligned}$$

Or

$$\begin{aligned} N(N-1) \int_{A \times A} \varphi_{N,2}(z_1, z_2) dz_1 dz_2 &= \frac{1}{\pi^2} \int_{A \times A} e^{-|z_1|^2} e^{-|z_2|^2} (K(z_1, z_1)K(z_2, z_2) - K(z_1, z_2)K(z_2, z_1)) dz_1 dz_2 \\ &= \left(\frac{1}{\pi} \int_A e^{-|z|^2} K(z, z) dz \right)^2 - \frac{1}{\pi^2} \int_{A \times A} e^{-|z_1|^2} e^{-|z_2|^2} |K(z_1, z_2)|^2 dz_1 dz_2, \end{aligned}$$

ce qui donne, en reconnaissant grâce à (3.2.2) l'espérance au carré de $\mu_N(A)$ dans le premier terme de droite,

$$\text{Var}(\mu_N(A)) = N \int_A \varphi_{N,1}(z) dz - \frac{1}{\pi^2} \int_{A \times A} e^{-|z_1|^2} e^{-|z_2|^2} |K(z_1, z_2)|^2 dz_1 dz_2.$$

Maintenant,

$$\begin{aligned} \int_{A \times A} e^{-|z_1|^2} e^{-|z_2|^2} |K(z_1, z_2)|^2 dz_1 dz_2 &= \int_{A \times \mathbb{C}} e^{-|z_1|^2} e^{-|z_2|^2} |K(z_1, z_2)|^2 dz_1 dz_2 \\ &\quad - \int_{A \times A^c} e^{-|z_1|^2} e^{-|z_2|^2} |K(z_1, z_2)|^2 dz_1 dz_2, \end{aligned}$$

et

$$\begin{aligned} \int_{A \times \mathbb{C}} e^{-|z_1|^2} e^{-|z_2|^2} |K(z_1, z_2)|^2 dz_1 dz_2 &= \int_{A \times \mathbb{C}} e^{-|z_1|^2} e^{-|z_2|^2} K(z_1, z_1)K(z_2, z_2) dz_1 dz_2 \\ &\quad - \int_{A \times \mathbb{C}} e^{-|z_1|^2} e^{-|z_2|^2} \det [K(z_i, z_j)]_{1 \leq i, j \leq 2} dz_1 dz_2 \\ &= N\pi \int_A e^{-|z|^2} K(z, z) dz - N(N-1)\pi^2 \int_{A \times \mathbb{C}} \varphi_{N,2}(z_1, z_2) dz_1 dz_2 \\ &= N^2\pi^2 \int_A \varphi_{N,1}(z) dz - N(N-1)\pi^2 \int_A \varphi_{N,1}(z) dz \\ &= N\pi^2 \int_A \varphi_{N,1}(z) dz, \end{aligned}$$

d'où finalement

$$\text{Var}(\mu_N(A)) = \frac{1}{\pi^2} \int_{A \times A^c} e^{-(|z_1|^2 + |z_2|^2)} |K(z_1, z_2)|^2 dz_1 dz_2. \quad (3.2.3)$$

C'est à partir de cette formule que l'on va maintenant estimer la variance du nombre de points dans quelques domaines particuliers.

3.2.1 Rectangle dans $D(0, \sqrt{N} - \sqrt{\ln N})$

On considère tout d'abord un ensemble quelconque de la forme $\sqrt{N}A$ où $A \subset \subset \mathbb{D}$. On a d'après (3.2.3)

$$\begin{aligned} \text{Var}(\mu_N(\sqrt{N}A)) &= \frac{1}{\pi^2} \int_{(\sqrt{N}A) \times (\sqrt{N}A)^c} e^{-(|z_1|^2 + |z_2|^2)} |K(z_1, z_2)|^2 dz_1 dz_2 \\ &= \frac{N^2}{\pi^2} \int_{A \times A^c} e^{-N(|z_1|^2 + |z_2|^2)} \left| \sum_{k=0}^{N-1} \frac{(Nz_1 z_2^*)^k}{k!} \right|^2 dz_1 dz_2. \end{aligned}$$

On sépare cette intégrale en trois parties.

- **Cas $|z_1 z_2| > 1$.** Grâce au Lemme 3.A.3, on a

$$\begin{aligned} I_1 &:= \int_{(A \times A^c) \cap \{|z_1 z_2| > 1\}} e^{-N(|z_1|^2 + |z_2|^2)} \left| \sum_{k=0}^{N-1} \frac{(Nz_1 z_2^*)^k}{k!} \right|^2 dz_1 dz_2 \\ &\leq \frac{N}{2\pi} \int_{(A \times A^c) \cap \{|z_1 z_2| > 1\}} e^{-N(|z_1|^2 + |z_2|^2)} e^{2N|z_1 z_2|^{2N}} dz_1 dz_2. \end{aligned}$$

En utilisant l'inclusion $(A \times A^c) \cap \{|z_1 z_2| > 1\} \subset \{|z_1| < 1 - \varepsilon, |z_2| > \frac{1}{1-\varepsilon}\}$ avec $\varepsilon := d(A, \partial\mathbb{D}) > 0$, on obtient

$$I_1 \leq \frac{N}{2\pi} \left(\int_{|z| < 1-\varepsilon} e^{N(1+\ln|z|^2 - |z|^2)} dz \right) \left(\int_{|z| > \frac{1}{1-\varepsilon}} e^{N(1+\ln|z|^2 - |z|^2)} dz \right). \quad (3.2.4)$$

Par concavité de $f : x \mapsto 1 + \ln x - x$, si $x < (1 - \varepsilon)^2$,

$$f(x) \leq \frac{2\varepsilon - \varepsilon^2}{(1 - \varepsilon)^2} x + 2 \ln(1 - \varepsilon), \quad (3.2.5)$$

et si $x > \frac{1}{(1-\varepsilon)^2}$,

$$f(x) \leq (-2\varepsilon + \varepsilon^2)x - 2 \ln(1 - \varepsilon).$$

L'inégalité (3.2.4) devient

$$\begin{aligned} I_1 &\leq \frac{N}{2\pi} \left(\int_{|z| < 1-\varepsilon} \exp \left[\frac{2\varepsilon - \varepsilon^2}{(1 - \varepsilon)^2} N|z|^2 \right] dz \right) \left(\int_{|z| > \frac{1}{1-\varepsilon}} \exp \left[(-2\varepsilon + \varepsilon^2) N|z|^2 \right] dz \right) \\ &= \frac{N}{2\pi} \cdot \frac{\pi(1 - \varepsilon)^2}{(2\varepsilon - \varepsilon^2)N} (\exp \left[(2\varepsilon - \varepsilon^2)N \right] - 1) \cdot \frac{\pi}{(2\varepsilon - \varepsilon^2)N} \exp \left[-\frac{(2\varepsilon - \varepsilon^2)N}{(1 - \varepsilon)^2} \right] \\ &= \frac{\pi(1 - \varepsilon)^2}{2(2\varepsilon - \varepsilon^2)^2 N} \exp \left[-\frac{(2\varepsilon - \varepsilon^2)^2 N}{(1 - \varepsilon)^2} \right] \\ &= \frac{\pi}{8\varepsilon^2 N} \frac{(1 - \varepsilon)^2}{(1 - \frac{\varepsilon}{2})^2} \exp \left[-4\varepsilon^2 N \frac{(1 - \frac{\varepsilon}{2})^2}{(1 - \varepsilon)^2} \right] \\ &\leq \frac{\pi e^{-4\varepsilon^2 N}}{8N\varepsilon^2} \leq \frac{\pi}{8N^4 \ln N}, \end{aligned} \quad (3.2.6)$$

si $\varepsilon \geq \sqrt{\frac{\ln N}{N}}$.

- **Cas** $|z_1 z_2| < 1$. Avec l'inclusion grossière $(A \times A^c) \cap \{|z_1 z_2| < 1\} \subset \{|z_1| < 1 - \varepsilon\} \times \mathbb{C}$ et la décomposition $\mathbb{C} = \{|z_2| < \frac{1}{1-\varepsilon}\} \cup \{|z_2| > \frac{1}{1-\varepsilon}\}$, on obtient grâce au Lemme 3.A.4,

$$I_2 := \int_{(A \times A^c) \cap \{|z_1 z_2| < 1\}} e^{-N(|z_1|^2 + |z_2|^2)} \left| \sum_{k=N}^{\infty} \frac{(N z_1 z_2^*)^k}{k!} \right|^2 dz_1 dz_2$$

$$\leq \int_{|z| < 1-\varepsilon} e^{N(1+\ln|z|^2 - |z|^2)} dz \int_{|z| < \frac{1}{1-\varepsilon}} e^{N(1+\ln|z|^2 - |z|^2)} dz \quad (3.2.7)$$

$$+ \int_{|z| < 1-\varepsilon} e^{N(1+\ln|z|^2 - |z|^2)} dz \int_{|z| > \frac{1}{1-\varepsilon}} e^{N(1+\ln|z|^2 - |z|^2)} dz. \quad (3.2.8)$$

Le terme (3.2.8) a déjà été estimé plus haut et nous donne l'estimation

$$\int_{|z| < 1-\varepsilon} e^{N(1+\ln|z|^2 - |z|^2)} dz \int_{|z| > \frac{1}{1-\varepsilon}} e^{N(1+\ln|z|^2 - |z|^2)} dz \leq \frac{\pi^2}{4N^2\varepsilon}. \quad (3.2.9)$$

Pour le terme (3.2.7), on réutilise l'inégalité de concavité (3.2.5) pour la première intégrale et on majore l'exponentielle par 1 dans la deuxième intégrale puisque $1 + \ln x - x \leq 0$ pour tout $x \in (0, 1)$, puis on utilise l'inégalité $\ln(1 - \varepsilon) \leq -\varepsilon - \frac{\varepsilon^2}{2}$ valable pour tout $\varepsilon \in (0, 1)$ et l'estimation $2\varepsilon - \varepsilon^2 \geq \varepsilon$, pour obtenir

$$\begin{aligned} & \int_{|z| < 1-\varepsilon} e^{N(1+\ln|z|^2 - |z|^2)} dz \int_{|z| < \frac{1}{1-\varepsilon}} e^{N(1+\ln|z|^2 - |z|^2)} dz \\ & \leq \int_{|z| < 1-\varepsilon} \exp\left[\frac{2\varepsilon - \varepsilon^2}{(1-\varepsilon)^2} N|z|^2\right] dz \cdot \frac{\pi}{(1-\varepsilon)^2} e^{2N \ln(1-\varepsilon)} \\ & = \frac{\pi^2 e^{2N \ln(1-\varepsilon)}}{(2\varepsilon - \varepsilon^2)N} \left(e^{(2\varepsilon - \varepsilon^2)N} - 1\right) \\ & \leq \frac{\pi^2 e^{-2\varepsilon^2 N}}{N\varepsilon}. \end{aligned} \quad (3.2.10)$$

La combinaison de (3.2.9) et (3.2.10) donne finalement

$$I_2 \leq \frac{\pi^2 e^{-2\varepsilon^2 N}}{N\varepsilon} + \frac{\pi^2 e^{-4\varepsilon^2 N}}{4N^2\varepsilon^2} \leq \frac{\pi^2}{N^{\frac{5}{2}}\sqrt{\ln N}} + \frac{\pi^2}{4N^5 \ln N}, \quad (3.2.11)$$

si $\varepsilon \geq \sqrt{\frac{\ln N}{N}}$. On voit ici que pour avoir $I_2 \lesssim \frac{1}{N^2}$, ce qui sera nécessaire dans la suite, il faut que $\varepsilon \geq C\sqrt{\frac{\ln N}{N}}$ avec $C > \frac{\sqrt{3}}{2}$.

- **Conclusion.** Il reste alors à étudier

$$\begin{aligned} I_3 & := \int_{(A \times A^c) \cap \{|z_1 z_2| < 1\}} e^{-N(|z_1|^2 + |z_2|^2)} \left| e^{N z_1 z_2^*} \right|^2 dz_1 dz_2 \\ & = \int_{(A \times A^c) \cap \{|z_1 z_2| < 1\}} e^{-N|z_1 - z_2|^2} dz_1 dz_2 \\ & \leq \int_{A \times A^c} e^{-N|z_1 - z_2|^2} dz_1 dz_2. \end{aligned} \quad (3.2.12)$$

A partir d'ici, considérons le cas d'un domaine A rectangulaire. Par invariance par translation de $|z_1 - z_2|^2$, on se ramène au domaine

$$A := \left[-\frac{l_1}{2}; \frac{l_1}{2}\right] \times \left[-\frac{l_2}{2}; \frac{l_2}{2}\right] \quad \text{avec} \quad \frac{1}{\sqrt{N}} \leq l_1, l_2 \leq \frac{\sqrt{2}}{2}.$$

Pour se ramener à la dimension 1, on écrit $z_1 = x_1 + iy_1$ et $z_2 = x_2 + iy_2$ et on obtient, avec $Q := \left[-\frac{1}{2}; \frac{1}{2}\right]^2$,

$$\begin{aligned} & \int_{A \times A^c} e^{-N|z_1 - z_2|^2} dz_1 dz_2 \\ &= l_1^2 l_2^2 \int_{Q \times Q^c} e^{-l_1^2 N(x_1 - x_2)^2} e^{-l_2^2 N(y_1 - y_2)^2} dz_1 dz_2 \\ &\leq l_1^2 l_2^2 \left(\int_{-\frac{1}{2}}^{\frac{1}{2}} \int_{[-\frac{1}{2}; \frac{1}{2}]^c} e^{-l_1^2 N(x_1 - x_2)^2} dx_1 dx_2 \right) \left(\int_{-\frac{1}{2}}^{\frac{1}{2}} \int_{\mathbb{R}} e^{-l_2^2 N(y_1 - y_2)^2} dy_1 dy_2 \right) \\ &\quad + l_1^2 l_2^2 \left(\int_{-\frac{1}{2}}^{\frac{1}{2}} \int_{[-\frac{1}{2}; \frac{1}{2}]^c} e^{-l_2^2 N(y_1 - y_2)^2} dy_1 dy_2 \right) \left(\int_{-\frac{1}{2}}^{\frac{1}{2}} \int_{\mathbb{R}} e^{-l_1^2 N(x_1 - x_2)^2} dx_1 dx_2 \right). \end{aligned}$$

Il y a donc 2 termes à estimer, le plus simple étant (on note l pour l_1 ou l_2)

$$\int_{-\frac{1}{2}}^{\frac{1}{2}} \int_{\mathbb{R}} e^{-l^2 N(y_1 - y_2)^2} dy_1 dy_2 = \sqrt{\frac{\pi}{l^2 N}}. \quad (3.2.13)$$

Pour le deuxième terme, on a pour $\alpha \leq 1$

$$\begin{aligned} \int_{-\frac{1}{2}}^{\frac{1}{2}} \int_{[-\frac{1}{2}; \frac{1}{2}]^c} e^{-l^2 N(x_1 - x_2)^2} dx_1 dx_2 &= 2 \int_{-\frac{1}{2}}^{\frac{1}{2}} \int_{\frac{1}{2}}^{+\infty} e^{-l^2 N(x_1 - x_2)^2} dx_1 dx_2 \\ &= 2 \int_{-\frac{1}{2}}^{\frac{1}{2} - \alpha} \int_{\frac{1}{2}}^{+\infty} e^{-l^2 N(x_1 - x_2)^2} dx_1 dx_2 \end{aligned} \quad (3.2.14)$$

$$+ 2 \int_{-\frac{1}{2}}^{\frac{1}{2}} \int_{\frac{1}{2}}^{\frac{1}{2} + \alpha} e^{-l^2 N(x_1 - x_2)^2} dx_1 dx_2 \quad (3.2.15)$$

$$+ 2 \int_{-\frac{1}{2} - \alpha}^{\frac{1}{2}} \int_{\frac{1}{2} + \alpha}^{+\infty} e^{-l^2 N(x_1 - x_2)^2} dx_1 dx_2. \quad (3.2.16)$$

L'intégrale (3.2.14) est estimée grâce au Lemme 3.A.5 :

$$\begin{aligned} \int_{-\frac{1}{2}}^{\frac{1}{2} - \alpha} \int_{\frac{1}{2}}^{+\infty} e^{-l^2 N(x_1 - x_2)^2} dx_1 dx_2 &\leq \frac{1}{2l^2 N} \int_{-\frac{1}{2}}^{\frac{1}{2} - \alpha} \frac{e^{-l^2 N(\frac{1}{2} - x_2)^2}}{\frac{1}{2} - x_2} dx_2 \\ &\leq \frac{1}{2l^2 N \alpha} \int_{-\infty}^{\frac{1}{2} - \alpha} e^{-l^2 N(\frac{1}{2} - x)^2} dx \\ &\leq \frac{e^{-l^2 N \alpha^2}}{(2l^2 N \alpha)^2}. \end{aligned}$$

En estimant l'intégrale (3.2.16) de la même façon et en bornant l'intégrale (3.2.15) par α^2 , on obtient

$$\int_{-\frac{1}{2}}^{\frac{1}{2}} \int_{[-\frac{1}{2}; \frac{1}{2}]^c} e^{-l^2 N(x_1 - x_2)^2} dx_1 dx_2 \leq \frac{e^{-l^2 N \alpha^2}}{(Nl^2 \alpha)^2} + 2\alpha^2. \quad (3.2.17)$$

Prenons $\alpha := \frac{1}{\sqrt{l^2 N}}$, ce qui donne, en combinant (3.2.12), (3.2.13) et (3.2.17),

$$I_3 \leq (2 + e^{-1}) \sqrt{\pi} \frac{l_1 + l_2}{N^{\frac{3}{2}}}. \quad (3.2.18)$$

Il reste à remarquer que $\text{Var } \mu_N(\sqrt{N}A) \leq \frac{N^2}{\pi^2} (I_1 + 2I_2 + 2I_3)$ et combiner (3.2.6), (3.2.11) et (3.2.18) en utilisant $\sqrt{N}(l_1 + l_2) \geq 2$ pour obtenir le résultat suivant.

Lemme 3.2.2. *Soit $N \geq 2$. Soit $R \subset D(0, \sqrt{N} - \sqrt{\ln N})$ un rectangle de côtés $l_1, l_2 \geq 1$, alors le nombre $\mu_N(R)$ de valeurs propres de M_N dans R vérifie*

$$\text{Var}(\mu_N(R)) \leq \left(\frac{4 + 2e^{-1}}{\pi^{3/2}} + \frac{1}{\sqrt{N \ln N}} + \frac{1}{16\pi N^2 \ln N} + \frac{1}{4N^3 \ln N} \right) (l_1 + l_2) \leq 2(l_1 + l_2).$$

De façon équivalente, le nombre $\mu_N(\sqrt{N}R)$ de valeurs propres de $\frac{1}{\sqrt{N}}M_N$ dans un rectangle $R \subset D(0, 1 - \varepsilon_N)$, où $\varepsilon_N := \sqrt{\frac{\ln N}{N}}$, de côtés $l_1, l_2 \geq \frac{1}{\sqrt{N}}$ vérifie

$$\text{Var}(\mu_N(\sqrt{N}R)) \leq 2\sqrt{N}(l_1 + l_2).$$

Dans la suite, on notera toujours

$$\varepsilon_N := \sqrt{\frac{\ln N}{N}}. \quad (3.2.19)$$

3.2.2 Disque de rayon inférieur à $\sqrt{N} - \sqrt{\ln N}$

Ensuite, soit $D := D(0, R)$, avec $1 \leq R \leq \sqrt{N} - \sqrt{\ln N}$. On utilise la même stratégie que dans [MM14, Prop. 10] pour estimer la variance du nombre de points dans D . Ainsi, par passage

aux coordonnées polaires dans (3.2.3),

$$\begin{aligned}
 \text{Var } \mu_N(D) &= \frac{1}{\pi^2} \int_{|z_1| \leq R} \int_{|z_2| > R} e^{-(|z_1|^2 + |z_2|^2)} \left| \sum_{k=0}^{N-1} \frac{(z_1 z_2^*)^k}{k!} \right|^2 dz_1 dz_2 \\
 &= \frac{1}{\pi^2} \sum_{k=0}^{N-1} \sum_{l=0}^{N-1} \frac{1}{k!l!} \int_0^R e^{-r^2} r^{k+l+1} dr \int_R^\infty e^{-r^2} r^{k+l+1} dr \int_0^{2\pi} e^{i(k-l)\varphi} d\varphi \int_0^{2\pi} e^{i(l-k)\varphi} d\varphi \\
 &= 4 \sum_{k=0}^{N-1} \frac{1}{(k!)^2} \int_0^R e^{-r^2} r^{2k+1} dr \int_R^\infty e^{-r^2} r^{2k+1} dr \\
 &= \sum_{k=0}^{N-1} \frac{1}{(k!)^2} \int_0^{R^2} e^{-s} s^k ds \int_{R^2}^\infty e^{-s} s^k ds \\
 &\leq \sum_{k=0}^{\beta} \frac{1}{k!} \int_{\beta}^\infty e^{-s} s^k ds + \sum_{k=\beta+1}^{N-1} \frac{1}{k!} \int_0^{\beta+1} e^{-s} s^k ds, \tag{3.2.20}
 \end{aligned}$$

où $\beta := \lfloor R^2 \rfloor$ et où on a utilisé $\int_0^a e^{-s} s^k ds \leq k!$ et $\int_a^\infty e^{-s} s^k ds \leq k!$ pour tout a . Estimons d'abord la première somme de (3.2.20), en utilisant le Lemme 3.A.1 pour estimer la factorielle d'un nombre :

$$\begin{aligned}
 \sum_{k=0}^{\beta} \frac{1}{k!} \int_{\beta}^\infty e^{-s} s^k ds &= \sum_{k=0}^{\beta} e^{-\beta} \sum_{l=0}^k \frac{\beta^l}{l!} = e^{-\beta} \sum_{l=0}^{\beta} \frac{\beta^l}{l!} (\beta - l + 1) \\
 &\leq 1 + e^{-\beta} \left(\sum_{l=0}^{\beta} \frac{\beta^{l+1}}{l!} - \sum_{l=1}^{\beta} \frac{\beta^l}{(l-1)!} \right) = 1 + e^{-\beta} \frac{\beta^{\beta+1}}{\beta!} \\
 &\leq 1 + \sqrt{\frac{\beta}{2\pi}}.
 \end{aligned}$$

De façon similaire, la deuxième somme de (3.2.20) est estimée par

$$\begin{aligned}
 \sum_{k=\beta+1}^{N-1} \frac{1}{k!} \int_0^{\beta+1} e^{-s} s^k ds &= \sum_{k=\beta+1}^{N-1} e^{-\beta-1} \sum_{l=k+1}^{\infty} \frac{(\beta+1)^l}{l!} \\
 &\leq e^{-\beta-1} \sum_{l=\beta+2}^{\infty} \frac{(\beta+1)^l}{l!} (l - \beta - 1) \\
 &= e^{-\beta-1} \frac{(\beta+1)^{\beta+2}}{(\beta+1)!} \leq \sqrt{\frac{\beta+1}{2\pi}},
 \end{aligned}$$

ce qui donne le résultat suivant :

Lemme 3.2.3. *Si $1 \leq R \leq \sqrt{N} - \sqrt{\ln N}$ alors*

$$\text{Var } \mu_N(D(0, R)) \leq 2R. \tag{3.2.21}$$

3.2.3 Arc sur le bord du disque

Considérons maintenant un arc d'anneau chevauchant le bord de \mathbb{D}_N . On rappelle que \mathbb{D}_N est le disque de rayon \sqrt{N} centré en 0 et la définition (3.2.19) de ε_N . Soit

$$B' := \left\{ z \in \mathbb{C} \mid \sqrt{N}(1 - \varepsilon_N) < |z| < \sqrt{N}(1 + 3\varepsilon_N) \text{ et } 0 < \arg z < \theta \right\}.$$

On a, avec le même passage en polaire que dans la sous-section précédente,

$$\text{Var } \mu_N(B') = J_1 + J_2 + J_3, \quad (3.2.22)$$

où

$$\begin{aligned} J_1 &= \frac{1}{\pi^2} \sum_{k=0}^{N-1} \sum_{l=0}^{N-1} \frac{1}{k!l!} \int_{\sqrt{N}(1-\varepsilon_N)}^{\sqrt{N}(1+3\varepsilon_N)} e^{-r^2} r^{k+l+1} dr \int_0^{\sqrt{N}(1-\varepsilon_N)} e^{-r^2} r^{k+l+1} dr \int_0^\theta e^{i(k-l)\varphi} d\varphi \int_0^{2\pi} e^{i(l-k)\varphi} d\varphi, \\ J_2 &= \frac{1}{\pi^2} \sum_{k=0}^{N-1} \sum_{l=0}^{N-1} \frac{1}{k!l!} \int_{\sqrt{N}(1-\varepsilon_N)}^{\sqrt{N}(1+3\varepsilon_N)} e^{-r^2} r^{k+l+1} dr \int_{\sqrt{N}(1+3\varepsilon_N)}^\infty e^{-r^2} r^{k+l+1} dr \int_0^\theta e^{i(k-l)\varphi} d\varphi \int_0^{2\pi} e^{i(l-k)\varphi} d\varphi, \\ J_3 &= \frac{1}{\pi^2} \sum_{k=0}^{N-1} \sum_{l=0}^{N-1} \frac{1}{k!l!} \int_{\sqrt{N}(1-\varepsilon_N)}^{\sqrt{N}(1+3\varepsilon_N)} e^{-r^2} r^{k+l+1} dr \int_{\sqrt{N}(1-\varepsilon_N)}^{\sqrt{N}(1+3\varepsilon_N)} e^{-r^2} r^{k+l+1} dr \int_0^\theta e^{i(k-l)\varphi} d\varphi \int_\theta^{2\pi} e^{i(l-k)\varphi} d\varphi. \end{aligned}$$

En majorant $\int_{\sqrt{N}(1-\varepsilon_N)}^{\sqrt{N}(1+3\varepsilon_N)} e^{-r^2} r^{k+l+1} dr$ par $\int_{\sqrt{N}(1-\varepsilon_N)}^\infty e^{-r^2} r^{k+l+1} dr$, le calcul précédent avec ici $R_N := \sqrt{N}(1 - \varepsilon_N)$ donne

$$J_1 \leq \frac{\theta}{2\pi} \left(1 + \frac{\sqrt{N}(1 - \varepsilon_N)}{\sqrt{2\pi}} \right) \leq \frac{\theta}{2\pi} \sqrt{N}.$$

De même, en notant $\gamma := N(1 + 3\varepsilon_N)^2$, on obtient

$$\begin{aligned} J_2 &\leq \frac{\theta}{2\pi} \sum_{k=0}^{N-1} \frac{1}{k!} \int_\gamma^\infty e^{-r} r^k dr = \frac{\theta}{2\pi} \sum_{k=0}^{N-1} e^{-\gamma} \sum_{l=0}^k \frac{\gamma^l}{l!} \\ &\leq \frac{\theta}{2\pi} \sum_{l=0}^{N-1} e^{-\gamma} \frac{\gamma^l}{l!} (\gamma - l) = \frac{\theta N}{2\pi} e^{-\gamma} \frac{\gamma^N}{N!} \\ &\leq \frac{\theta \sqrt{N}}{(2\pi)^{\frac{3}{2}}} \exp \left[N \left(1 - \frac{\gamma}{N} + \ln \frac{\gamma}{N} \right) \right] \\ &\leq \frac{\theta \sqrt{N}}{(2\pi)^{\frac{3}{2}}} \exp \left[-\frac{N}{2} \left(1 - \frac{\gamma}{N} \right)^2 \right] \\ &\leq \frac{\theta \sqrt{N}}{(2\pi)^{\frac{3}{2}}} e^{-18N\varepsilon_N^2} \leq \frac{\theta}{(2\pi)^{\frac{3}{2}}} \cdot \frac{\sqrt{N}}{N^{18}}. \end{aligned}$$

Estimons maintenant J_3 . En remarquant que si $k \neq l$,

$$\int_0^\theta e^{i(k-l)\varphi} d\varphi \int_\theta^{2\pi} e^{i(l-k)\varphi} d\varphi = - \left| \int_0^\theta e^{i(k-l)\varphi} d\varphi \right|^2 \leq 0,$$

on a

$$J_3 \leq \frac{\theta}{2\pi} \left(1 - \frac{\theta}{2\pi}\right) \sum_{k=0}^{N-1} \left(\frac{1}{k!} \int_{N(1-\varepsilon_N)^2}^{N(1+3\varepsilon_N)^2} e^{-s} s^k ds \right)^2 \leq \frac{\theta}{2\pi} \left(1 - \frac{\theta}{2\pi}\right) \sum_{k=0}^{N-1} \frac{1}{k!} \int_{N(1-\varepsilon_N)^2}^{N(1+3\varepsilon_N)^2} e^{-s} s^k ds.$$

Sur $(0, +\infty)$, la fonction $s \mapsto e^{-s} s^k$ atteint un unique maximum global en $s = k$ donc

$$\int_{N(1-\varepsilon_N)^2}^{N(1+3\varepsilon_N)^2} e^{-s} s^k ds \leq \begin{cases} 8N\varepsilon_N(1+\varepsilon_N)e^{-N(1-\varepsilon_N)^2}(N(1-\varepsilon_N)^2)^k & \text{si } k \leq N(1-\varepsilon_N)^2, \\ k! & \text{si } N(1-\varepsilon_N)^2 < k \leq N-1. \end{cases}$$

D'où

$$J_3 \leq \frac{\theta}{2\pi} \left(1 - \frac{\theta}{2\pi}\right) \left[8N\varepsilon_N(1+\varepsilon_N)e^{-N(1-\varepsilon_N)^2} \sum_{k=0}^{\lfloor N(1-\varepsilon_N)^2 \rfloor} \frac{N^k(1-\varepsilon_N)^{2k}}{k!} + 2N\varepsilon_N \right] \leq 2\theta N\varepsilon_N.$$

On obtient finalement par (3.2.22)

$$\text{Var } \mu_N(B') \leq \left(\frac{1}{2\pi\sqrt{\ln N}} + \frac{1}{(2\pi)^{\frac{3}{2}} N^{18}\sqrt{\ln N}} + 2 \right) \theta N\varepsilon_N \leq 3\theta N\varepsilon_N,$$

ce qui donne, avec le résultat $|B'| = 4\theta N\varepsilon_N(1+\varepsilon_N)$, l'estimation

$$\text{Var } \mu_N(B') \leq \frac{3}{4(1+\varepsilon_N)} |B'| \leq |B'|.$$

Lemme 3.2.4. *Si $B' := \{z \in \mathbb{C} \mid \sqrt{N}(1-\varepsilon_N) < |z| < \sqrt{N}(1+3\varepsilon_N) \text{ et } 0 < \arg z < \theta\}$, alors*

$$\text{Var } \mu_N(B') \leq |B'|.$$

3.2.4 Intersection d'un rectangle avec un disque

Considérons enfin l'intersection d'un rectangle du quadrant $\{\Re \geq 0, \Im \geq 0\}$ avec un cercle centré en 0 de rayon inférieur à $\sqrt{N} - \sqrt{\ln N}$. Plus précisément, soit $A \subset \mathbb{D}$ défini comme suit :

$$A := Q \cap D(0, r) \quad \text{avec} \quad Q = [a; a+L] \times [b; b+H], \quad (3.2.23)$$

où $a \geq 0$, $b \geq 0$, $r \leq 1 - \varepsilon_N$, $L := \sqrt{r^2 - b^2} - a$ la longueur et $H := \sqrt{r^2 - a^2} - b$ la hauteur de Q , tels que $L, H \geq \frac{1}{\sqrt{N}}$. On renvoie à la Figure 3.1 pour une illustration de cet ensemble. Par la formule (3.2.3),

$$\begin{aligned} \text{Var } (\mu_N(\sqrt{N}A)) &= \frac{N^2}{\pi^2} \int_{A \times A^c} e^{-N(|z_1|^2 + |z_2|^2)} \left| \sum_{k=0}^{N-1} \frac{(Nz_1 z_2^*)^k}{k!} \right|^2 dz_1 dz_2 \\ &= \frac{N^2}{\pi^2} \int_{A \times Q^c} e^{-N(|z_1|^2 + |z_2|^2)} \left| \sum_{k=0}^{N-1} \frac{(Nz_1 z_2^*)^k}{k!} \right|^2 dz_1 dz_2 \end{aligned} \quad (3.2.24)$$

$$+ \frac{N^2}{\pi^2} \int_{A \times (A^c \cap Q)} e^{-N(|z_1|^2 + |z_2|^2)} \left| \sum_{k=0}^{N-1} \frac{(Nz_1 z_2^*)^k}{k!} \right|^2 dz_1 dz_2. \quad (3.2.25)$$

L'intégrale (3.2.24) permet de se ramener au cas traité dans la sous-section 3.2.1, en utilisant les inclusions $A \times Q^c \subset A \times A^c$ pour les deux premières intégrales et $A \times Q^c \subset Q \times Q^c$ pour la troisième et en remarquant que l'estimation de $\int_{Q \times Q^c} e^{-N|z_1 - z_2|^2}$ en (3.2.18) est valable pour n'importe quel rectangle Q de \mathbb{C} :

$$\begin{aligned}
 & \int_{A \times Q^c} e^{-N(|z_1|^2 + |z_2|^2)} \left| \sum_{k=0}^{N-1} \frac{(Nz_1 z_2^*)^k}{k!} \right|^2 dz_1 dz_2 \\
 & \leq \int_{(A \times Q^c) \cap \{|z_1 z_2| > 1\}} e^{-N(|z_1|^2 + |z_2|^2)} \left| \sum_{k=0}^{N-1} \frac{(Nz_1 z_2^*)^k}{k!} \right|^2 dz_1 dz_2 \\
 & \quad + 2 \int_{(A \times Q^c) \cap \{|z_1 z_2| < 1\}} e^{-N(|z_1|^2 + |z_2|^2)} \left| \sum_{k=N}^{\infty} \frac{(Nz_1 z_2^*)^k}{k!} \right|^2 dz_1 dz_2 \\
 & \quad + 2 \int_{A \times Q^c} e^{-N|z_1 - z_2|^2} dz_1 dz_2 \\
 & \lesssim \frac{1}{N^4 \log N} + \frac{1}{N^{\frac{5}{2}} \sqrt{\log N}} + \frac{L + H}{N^{\frac{3}{2}}}. \tag{3.2.26}
 \end{aligned}$$

Pour l'intégrale (3.2.25), introduisons le plus petit angle θ tel que Q soit inclus dans un radiant d'angle θ et la distance $p := d(a + ib, \partial D(0, r))$, voir la Figure 3.1. Supposons ici $L \leq H$ (ce qui correspond au cas $a \geq b$), on a alors $p \leq L$ ainsi que $\theta r \leq \frac{\sqrt{2}\pi}{4} H$ (en effet, si h est la longueur de la corde : $h = 2r \sin \frac{\theta}{2} \geq 2r \frac{2\theta}{\pi}$ car $0 \leq \frac{\theta}{2} \leq \frac{\pi}{4}$, donc $\theta r \leq \frac{\pi}{4} h = \frac{\pi}{4} \sqrt{L^2 + H^2} \leq \frac{\sqrt{2}\pi}{4} H$), ce qui donne :

$$\theta r p \leq \frac{\sqrt{2}\pi}{4} L H \leq \frac{\sqrt{2}\pi}{2} |A|. \tag{3.2.27}$$

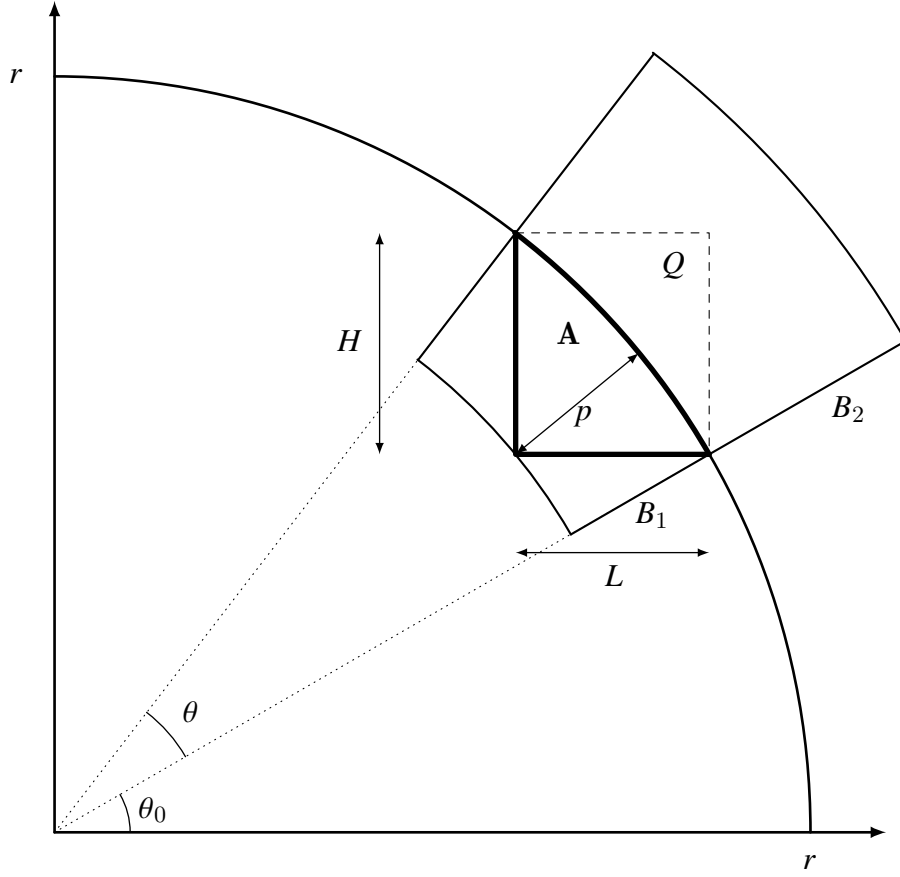
On définit les ensembles

$$\begin{aligned}
 & B_1 := \{r - p \leq |z| \leq r, \theta_0 \leq \arg(z) \leq \theta_0 + \theta\} \\
 \text{et } & B_2 := \{r \leq |z| \leq r + \sqrt{2}H\},
 \end{aligned}$$

de façon à ce que

$$A \times (A^c \cap Q) \subset B_1 \times B_2.$$

On a alors


 FIGURE 3.1: Définition de A dans (3.2.23).

$$\begin{aligned}
 & \int_{A \times (A^c \cap Q)} e^{-N(|z_1|^2 + |z_2|^2)} \left| \sum_{k=0}^{N-1} \frac{(N z_1 z_2^*)^k}{k!} \right|^2 dz_1 dz_2 \\
 & \leq \int_{B_1 \times B_2} e^{-N(|z_1|^2 + |z_2|^2)} \left| \sum_{k=0}^{N-1} \frac{(N z_1 z_2^*)^k}{k!} \right|^2 dz_1 dz_2 \\
 & = \sum_{k,l=0}^{N-1} \frac{N^{k+l}}{k!l!} \int_{r-p}^r e^{-Nu^2} u^{k+l+1} du \int_{\theta_0}^{\theta_0+\theta} e^{i(k-l)\varphi} d\varphi \int_r^{r+\sqrt{2}H} e^{-Nu^2} u^{k+l+1} du \int_0^{2\pi} e^{i(l-k)\varphi} d\varphi \\
 & = 2\pi\theta \sum_{k=0}^{N-1} \frac{N^{2k}}{(k!)^2} \int_{r-p}^r e^{-Nu^2} u^{2k+1} du \int_r^{r+\sqrt{2}H} e^{-Nu^2} u^{2k+1} du \\
 & = \frac{\pi\theta}{2N^2} \sum_{k=0}^{N-1} \frac{1}{(k!)^2} \int_{N(r-p)^2}^{Nr^2} e^{-s} s^k ds \int_{Nr^2}^{N(r+\sqrt{2}H)^2} e^{-s} s^k ds \\
 & \leq \frac{\pi\theta}{2N^2} \sum_{k=0}^{N-1} \frac{1}{k!} \int_{N(r-p)^2}^{Nr^2} e^{-s} s^k ds.
 \end{aligned}$$

Or, sur $(0; +\infty)$ la fonction $s \mapsto e^{-s}s^k$ atteint un unique maximum en $s = k$ donc

$$\int_{N(r-p)^2}^{Nr^2} e^{-s}s^k ds \leq \begin{cases} 2pNre^{-N(r-p)^2}(N(r-p)^2)^k & \text{si } k \leq N(r-p)^2, \\ k! & \text{si } N(r-p)^2 < k < Nr^2, \\ 2pNre^{-Nr^2}(Nr^2)^k & \text{si } k \geq Nr^2. \end{cases}$$

Ceci donne alors

$$\begin{aligned} \frac{N^2}{\pi^2} \int_{A \times (A^c \cap Q)} e^{-N(|z_1|^2 + |z_2|^2)} \left| \sum_{k=0}^{N-1} \frac{(Nz_1 z_2^*)^k}{k!} \right|^2 dz_1 dz_2 &\leq \frac{\theta}{2\pi} \cdot 6pNr \\ &\stackrel{(3.2.27)}{\leq} \frac{3\sqrt{2}}{2} N|A|. \end{aligned} \quad (3.2.28)$$

En insérant (3.2.26) et (3.2.28) dans (3.2.24) et (3.2.25), on obtient le lemme suivant.

Lemme 3.2.5. *Soit A défini en (3.2.23). Supposons que $L, H \geq \frac{1}{\sqrt{N}}$ alors*

$$\text{Var}(\mu_N(\sqrt{N}A)) \leq 2\sqrt{N}(L + H) + \frac{3\sqrt{2}}{2} N|A| \leq 11N|A|.$$

3.3 Optimal matching dans un carré

3.3.1 Estimation de la distance

Soit $Q \subset D(0, \sqrt{N} - \sqrt{\ln N})$ un carré de côté $L = 2^K l_0$ avec $l_0 \in [1; 2)$ et $K \in \mathbb{N}$, et notons ici $\mu := \mu_N$. On rappelle la notation (3.1.1). On cherche à estimer la valeur moyenne de

$$W_Q^2 \left(\mu, \frac{\mu(Q)}{|Q|} \right).$$

La stratégie que l'on va utiliser m'a été suggérée par F. Otto.

ETAPE 1 (Construction d'une suite finie de mesures). On découpe Q en quatre carrés disjoints Q_1, Q_2, Q_3 et Q_4 chacun de côté $l_{K-1} := 2^{K-1} l_0$ et on définit la mesure

$$\nu_{K-1} := \sum_{i=1}^4 \frac{\mu(Q_i)}{|Q_i|} dz|_{Q_i} = \sum_{i=1}^4 \frac{\mu(Q_i)}{l_{K-1}^2} dz|_{Q_i}.$$

Cette mesure est simplement constante sur chacun des sous-carrés, chaque constante étant choisie pour que ν_{K-1} donne la même masse que μ à chaque sous-carré. De la même façon, on découpe chaque Q_i en quatre carrés disjoints $Q_{i,1}, Q_{i,2}, Q_{i,3}$ et $Q_{i,4}$ chacun de côté $l_{K-2} := 2^{K-2} l_0$ et on définit la mesure

$$\nu_{K-2} := \sum_{i=1}^4 \sum_{j=1}^4 \frac{\mu(Q_{i,j})}{l_{K-2}^2} dz|_{Q_{i,j}}.$$

On continue de cette manière jusqu'à des carrés de côté l_0 et on définit ainsi les mesures ν_{K-1} , ν_{K-2}, \dots, ν_0 . En particulier

$$\nu_k := \sum_{i_1=1}^4 \cdots \sum_{i_{K-k}=1}^4 \frac{\mu(Q_{i_1, \dots, i_{K-k}})}{l_k^2} dz|_{Q_{i_1, \dots, i_{K-k}}}.$$

On définit finalement $\nu_K := \frac{\mu(Q)}{|Q|} dz|_Q$.

ETAPE 2 (Utilisation des propriétés de W_2). Par inégalité triangulaire, on a

$$\left(\mathbb{E}W_Q^2 \left(\mu, \frac{\mu(Q)}{|Q|} \right) \right)^{\frac{1}{2}} \leq \left(\mathbb{E}W_2^2(\mu|_Q, \nu_0) \right)^{\frac{1}{2}} + \sum_{k=1}^K \left(\mathbb{E}W_2^2(\nu_{k-1}, \nu_k) \right)^{\frac{1}{2}}. \quad (3.3.1)$$

Puisque par construction, on a pour tout k et tous $1 \leq i_1 \leq 4, \dots, 1 \leq i_{K-k} \leq 4$,

$$\nu_{k-1}(Q_{i_1, \dots, i_{K-k}}) = \nu_k(Q_{i_1, \dots, i_{K-k}}),$$

la fonction

$$T_k := \sum_{i_1=1}^4 \cdots \sum_{i_{K-k}=1}^4 T_{i_1, \dots, i_{K-k}}|_{Q_{i_1, \dots, i_{K-k}}},$$

où $T_{i_1, \dots, i_{K-k}}$ est le transport optimal de $\nu_{k-1}|_{Q_{i_1, \dots, i_{K-k}}}$ à $\nu_k|_{Q_{i_1, \dots, i_{K-k}}}$, transporte ν_{k-1} sur ν_k . D'où

$$\mathbb{E}W_2^2(\nu_{k-1}, \nu_k) \leq \sum_{i_1=1}^4 \cdots \sum_{i_{K-k}=1}^4 \mathbb{E}W_{Q_{i_1, \dots, i_{K-k}}}^2(\nu_{k-1}, \nu_k). \quad (3.3.2)$$

De la même façon

$$\mathbb{E}W_2^2(\mu|_Q, \nu_0) \leq \sum_{i_1=1}^4 \cdots \sum_{i_K=1}^4 \mathbb{E}W_{Q_{i_1, \dots, i_K}}^2(\mu, \nu_0). \quad (3.3.3)$$

Les termes de la somme dans (3.3.3) peuvent être estimés grossièrement par

$$\mathbb{E}W_{Q_{i_1, \dots, i_K}}^2(\mu, \nu_0) \leq (\sqrt{2}l_0)^2 \mathbb{E}\mu(Q_{i_1, \dots, i_K}) \leq \frac{2}{\pi} l_0^4 \leq \frac{8}{\pi} l_0^2,$$

ce qui donne l'estimation

$$\mathbb{E}W_2^2(\mu|_Q, \nu_0) \leq \frac{8}{\pi} L^2. \quad (3.3.4)$$

ETAPE 3 (Estimation de la distance entre ν_{k-1} et ν_k dans un carré de taille l_k). Il reste maintenant à estimer $\mathbb{E}W_{Q'}^2(\nu_{k-1}, \nu_k)$, où l'on note $Q' := Q_{i_1, \dots, i_{K-k}}$ et $l := \frac{l_k}{2}$. Q' est donc un carré de côté $2l$ découpé en quatre carrés disjoints Q'_i chacun de côté l . En notant

$$\chi_{\leq} := \chi_{\{\exists i, \mu(Q'_i) \leq \frac{l^2}{2\pi}\}} \quad \text{et} \quad \chi_{>} := \chi_{\{\mu(Q'_i) > \frac{l^2}{2\pi}, \forall i\}},$$

on peut écrire

$$\mathbb{E}W_{Q'}^2(\nu_{k-1}, \nu_k) = \mathbb{E} \left[W_{Q'}^2(\nu_{k-1}, \nu_k) \chi_{>} \right] + \mathbb{E} \left[W_{Q'}^2(\nu_{k-1}, \nu_k) \chi_{\leq} \right]. \quad (3.3.5)$$

ETAPE 3.A. (Premier terme dans (3.3.5)). Introduisons la fonction φ qui vérifie

$$\begin{cases} -\Delta\varphi = \nu_k - \nu_{k-1} & \text{sur } \mathcal{Q}', \\ \nabla\varphi \cdot n = 0 & \text{sur } \partial\mathcal{Q}'. \end{cases}$$

Si pour tout i , on a $\frac{\mu(\mathcal{Q}'_i)}{l^2} > \frac{1}{2\pi}$, alors $\frac{\mu(\mathcal{Q}')}{4l^2} \geq \frac{1}{2\pi}$ et on a d'après (II.7)

$$W_{\mathcal{Q}'}^2(\nu_{k-1}, \nu_k) \leq 2\pi \int_{\mathcal{Q}'} |\nabla\varphi|^2 \leq 8\pi l^2 \int_{\mathcal{Q}'} |\nu_k - \nu_{k-1}|^2 = 8\pi l^4 \sum_{i=1}^4 (c - c_i)^2, \quad (3.3.6)$$

où $c := \frac{\nu_k(\mathcal{Q}')}{|\mathcal{Q}'|} = \frac{\mu(\mathcal{Q}')}{4l^2}$ et $c_i := \frac{\nu_{k-1}(\mathcal{Q}'_i)}{|\mathcal{Q}'_i|} = \frac{\mu(\mathcal{Q}'_i)}{l^2}$, ce qui donne

$$\begin{aligned} \mathbb{E} \left[W_{\mathcal{Q}'}^2(\nu_{k-1}, \nu_k) \chi_{>} \right] &\leq \frac{\pi}{2} \sum_{i=1}^4 \mathbb{E} \left[(\mu(\mathcal{Q}') - 4\mu(\mathcal{Q}'_i))^2 \right] \\ &\leq \pi \mathbb{E} \left[\left(\mu(\mathcal{Q}') - \frac{4l^2}{\pi} \right)^2 \right] + 16\pi \sum_{i=1}^4 \mathbb{E} \left[\left(\mu(\mathcal{Q}'_i) - \frac{l^2}{\pi} \right)^2 \right] \\ &\leq 2\pi \text{Var}(\mu(\mathcal{Q}')) + 2\pi \left(\mathbb{E}\mu(\mathcal{Q}') - \frac{4l^2}{\pi} \right)^2 \\ &\quad + 32\pi \sum_{i=1}^4 \left(\text{Var}(\mu(\mathcal{Q}'_i)) + \left(\mathbb{E}\mu(\mathcal{Q}'_i) - \frac{l^2}{\pi} \right)^2 \right). \end{aligned}$$

En utilisant le Lemme 3.2.1, on obtient²

$$\begin{aligned} \mathbb{E} \left[W_{\mathcal{Q}'}^2(\nu_{k-1}, \nu_k) \chi_{>} \right] &\lesssim \frac{1}{N^3} + \text{Var}(\mu(\mathcal{Q}')) + \sum_{i=1}^4 \text{Var}(\mu(\mathcal{Q}'_i)) \\ &\lesssim l + \text{Var}(\mu(\mathcal{Q}')) + \sum_{i=1}^4 \text{Var}(\mu(\mathcal{Q}'_i)). \end{aligned}$$

D'après le Lemme 3.2.2, on a donc

$$\mathbb{E} \left[W_{\mathcal{Q}'}^2(\nu_{k-1}, \nu_k) \chi_{>} \right] \lesssim l. \quad (3.3.7)$$

ETAPE 3.B. (Deuxième terme dans (3.3.5)). Ecrivons

$$\begin{aligned} \mathbb{E} \left[W_{\mathcal{Q}'}^2(\nu_{k-1}, \nu_k) \chi_{\leq} \right] &= \mathbb{E} \left[W_{\mathcal{Q}'}^2(\nu_{k-1}, \nu_k) \chi_{\leq} \chi_{\{\mu(\mathcal{Q}') \leq \frac{8l^2}{\pi}\}} \right] \\ &\quad + \sum_{j=2}^{\infty} \mathbb{E} \left[W_{\mathcal{Q}'}^2(\nu_{k-1}, \nu_k) \chi_{\leq} \chi_{\{\frac{2j^2 l^2}{\pi} < \mu(\mathcal{Q}') \leq \frac{2(j+1)^2 l^2}{\pi}\}} \right]. \end{aligned} \quad (3.3.8)$$

²On ne cherche dorénavant plus à expliciter les constantes mais il est possible de le faire.

Si N est suffisamment grand³ pour que $\frac{e}{N^{3/2}} \leq \frac{l_0^2}{4\pi}$, de façon à ce qu'on ait $\frac{l^2}{\pi} - \mathbb{E}\mu(Q'_i) \leq \frac{l^2}{4\pi}$ d'après le Lemme 3.2.1, on a pour tout $i \in \{1, 4\}$

$$\begin{aligned} \mathbb{P}\left(\mu(Q'_i) \leq \frac{l^2}{2\pi}\right) &= \mathbb{P}\left(\mu(Q'_i) - \mathbb{E}\mu(Q'_i) \leq \frac{l^2}{\pi} - \mathbb{E}\mu(Q'_i) - \frac{l^2}{2\pi}\right) \\ &\leq \mathbb{P}\left(\mu(Q'_i) - \mathbb{E}\mu(Q'_i) \leq -\frac{l^2}{4\pi}\right) \\ &\leq \frac{16\pi^2 \text{Var}(\mu(Q'_i))}{l^4}. \end{aligned}$$

D'où, en estimant grossièrement la distance W_2 et en utilisant le Lemme 3.2.2,

$$\mathbb{E}\left[W_{Q'}^2(\nu_{k-1}, \nu_k) \chi_{\leq \chi_{\{\mu(Q') \leq \frac{8l^2}{\pi}\}}}\right] \leq (\sqrt{2}l)^2 \frac{8l^2}{\pi} \mathbb{P}\left(\exists i, \mu(Q'_i) \leq \frac{l^2}{2\pi}\right) \quad (3.3.9)$$

$$\lesssim \sum_{i=1}^4 \text{Var}(\mu(Q'_i)) \lesssim l. \quad (3.3.10)$$

De plus, en utilisant à nouveau le Lemme 3.2.1, pour tout $j \geq 2$,

$$\mathbb{P}\left(\mu(Q') \geq \frac{2j^2 l^2}{\pi}\right) \leq \mathbb{P}\left(\mu(Q') - \mathbb{E}\mu(Q') \geq \frac{2(j^2 - 2)l^2}{\pi}\right) \leq \frac{\pi^2 \text{Var}(\mu(Q'))}{4(j^2 - 2)^2 l^4},$$

de sorte que

$$\begin{aligned} \mathbb{E}\left[W_{Q'}^2(\nu_{k-1}, \nu_k) \chi_{\leq \chi_{\{\frac{2j^2 l^2}{\pi} < \mu(Q') \leq \frac{2(j+1)^2 l^2}{\pi}\}}}\right] &\leq (\sqrt{2}l)^2 \cdot \frac{2(j+1)^2 l^2}{\pi} \cdot \frac{\pi^2 \text{Var}(\mu(Q'))}{4(j^2 - 2)^2 l^4} \\ &= \pi \text{Var}(\mu(Q')) \frac{(j+1)^2}{(j^2 - 2)^2} \\ &\lesssim l \frac{(j+1)^2}{(j^2 - 2)^2}. \end{aligned} \quad (3.3.11)$$

En injectant (3.3.9) et (3.3.11) dans (3.3.8), on obtient

$$\mathbb{E}\left[W_{Q'}^2(\nu_{k-1}, \nu_k) \chi_{\leq}\right] \lesssim l, \quad (3.3.12)$$

car $\sum_{j=2}^{\infty} \frac{(j+1)^2}{(j^2-2)^2} < \infty$.

ETAPE 4 (Conclusion). En combinant (3.3.12) et (3.3.7), on obtient finalement

$$\mathbb{E}W_{Q'}^2(\nu_{k-1}, \nu_k) \lesssim l. \quad (3.3.13)$$

³ $N \geq 11$ marche, sinon il suffit de choisir l_0 plus grand : $\frac{4\pi e}{N^{3/2}} \leq 13$ car $N \geq 2$. L'estimation voulue dans le cas $L < 13$ est immédiate.

Il reste maintenant à remonter les calculs. Avec (3.3.13), (3.3.2) devient

$$\mathbb{E}W_2^2(\nu_{k-1}, \nu_k) \lesssim \left(\frac{L}{l_k}\right)^2 l_{k-1} \lesssim \frac{L^2}{l_k},$$

ce qui, combiné avec (3.3.4) donne finalement dans (3.3.1)

$$\left(\mathbb{E}W_Q^2\left(\mu, \frac{\mu(Q)}{|Q|}\right)\right)^{\frac{1}{2}} \lesssim L + L \sum_{k=1}^K \frac{1}{\sqrt{l_k}} \lesssim L.$$

On a donc démontré le résultat suivant :

Théorème 3.3.1. *Soit $Q \subset D(0, \sqrt{N} - \sqrt{\ln N})$ un carré alors*

$$\mathbb{E}W_Q^2\left(\mu_N, \frac{\mu_N(Q)}{|Q|}\right) \lesssim |Q|.$$

3.3.2 Concentration : preuve du Théorème IV.3

On donne ici la preuve du Théorème IV.3, c'est-à-dire que l'on prouve une inégalité de concentration de la forme

$$\mathbb{P}\left(W_Q\left(\mu_N, \frac{\mu_N(Q)}{|Q|}\right) \geq t\right) \lesssim \exp\left(-c \frac{t^2}{L^2}\right),$$

pour tout $t \gg L\sqrt{\ln L}$. Pour cela, on va réutiliser la construction précédente, avec les mesures ν_0, \dots, ν_K sur les domaines carrés $Q_{i_1, \dots, i_{K-k}}$. Ceci permet de transformer

$$\mathbb{P}\left(W_Q\left(\mu_N, \frac{\mu_N(Q)}{|Q|}\right) \geq t\right)$$

en une somme de probabilités de déviation pour les mesures ν_k dans chaque sous-domaine $Q_{i_1, \dots, i_{K-k}}$. Ces probabilités de déviation se réécrivent alors en probabilités de déviation du nombre de points dans le sous-domaine, pour lesquelles on peut utiliser l'inégalité de concentration (3.2.1). La différence est que l'on ne va pas descendre jusqu'à l'échelle microscopique $l_0 \sim 1$ car cela ne nous donnerait rien d'intéressant. On va en fait devoir choisir une échelle minimale l_0 mésoscopique dépendant de t pour optimiser les estimations qu'on obtient.

Démonstration du Théorème IV.3. A nouveau, on note $\mu := \mu_N$. Soit $t \geq L$.

ETAPE 1 (Cas où $t \gg L^2$). Remarquons tout d'abord que si $t \geq \frac{2}{\sqrt{\pi}}L^2$ alors $\frac{t^2}{2L^2} - \frac{L^2}{\pi} \geq \frac{t^2}{4L^2}$ ce qui donne grâce au Lemme 3.2.2

$$\begin{aligned} \mathbb{P}\left(W_Q\left(\mu, \frac{\mu(Q)}{|Q|}\right) \geq t\right) &\leq \mathbb{P}\left(\mu(Q) \geq \frac{t^2}{2L^2}\right) \\ &\leq \mathbb{P}\left(\mu(Q) - \mathbb{E}\mu(Q) \geq \frac{t^2}{4L^2}\right) \\ &\leq \exp\left(-\min\left\{\frac{t^4}{64L^4 \text{Var } \mu(Q)}, \frac{t^2}{8L^2}\right\}\right) \\ &\leq \exp\left(-\min\left\{\frac{t^4}{256L^5}, \frac{t^2}{8L^2}\right\}\right). \end{aligned}$$

Puisque $t \geq \frac{2}{\sqrt{\pi}}L^2$, on a $\frac{t^2}{256L^3} \geq \frac{L}{64\pi} \geq \frac{1}{64\pi}$ donc finalement

$$\mathbb{P}\left(W_Q\left(\mu, \frac{\mu(Q)}{|Q|}\right) \geq t\right) \leq \exp\left(-\frac{t^2}{64\pi L^2}\right) \quad \text{si } t \geq \frac{2}{\sqrt{\pi}}L^2. \quad (3.3.14)$$

Cette inégalité n'est probablement pas optimale. En particulier, dans le cas où le carré est macroscopique, c'est-à-dire $L \sim \sqrt{N}$, [CHM18] suggère une inégalité de la forme

$$\mathbb{P}\left(W_Q\left(\mu, \frac{\mu(Q)}{|Q|}\right) \geq t\right) \leq \exp\left(-\frac{ct^2}{L}\right) \quad \text{si } t \gg L^{3/2}\sqrt{\log L}.$$

Cependant, les auteurs obtiennent cette inégalité de concentration uniquement pour la mesure empirique globale et sa limite ν .

ETAPE 2 (Utilisation de la construction précédente). Fixons maintenant $L \leq t < \frac{2}{\sqrt{\pi}}L^2$. L'idée est de reproduire la construction précédente mais sans descendre jusqu'à l'échelle microscopique. Soit donc $1 \leq l_0 := l_0(t) \leq L$ à fixer plus tard, tel que $\frac{L}{l_0} = 2^K$ pour un certain $K \in \mathbb{N}$. Définissons également une famille $(\alpha_k)_k \in (0; 1)^{K+1}$, vérifiant $\sum_{k=0}^K \alpha_k = 1$, à préciser plus tard. On a alors, par inégalité triangulaire puis admissibilité du transport carré par carré comme dans (3.3.2) et (3.3.3),

$$\begin{aligned} \mathbb{P}\left(W_Q\left(\mu, \frac{\mu(Q)}{|Q|}\right) \geq t\right) &\leq \mathbb{P}\left(W_2(\mu \lfloor_Q, \nu_0) + \sum_{k=1}^K W_2(\nu_{k-1}, \nu_k) \geq t\right) \\ &\leq \mathbb{P}\left(W_2(\mu \lfloor_Q, \nu_0) \geq \alpha_0 t\right) + \sum_{k=1}^K \mathbb{P}\left(W_2(\nu_{k-1}, \nu_k) \geq \alpha_k t\right) \\ &= \mathbb{P}\left(W_2^2(\mu \lfloor_Q, \nu_0) \geq \alpha_0^2 t^2\right) + \sum_{k=1}^K \mathbb{P}\left(W_2^2(\nu_{k-1}, \nu_k) \geq \alpha_k^2 t^2\right) \\ &\leq \sum_{i_1=1}^4 \cdots \sum_{i_K=1}^4 \mathbb{P}\left(W_{Q_{i_1, \dots, i_K}}^2(\mu, \nu_0) \geq \frac{\alpha_0^2 t^2}{L^2} l_0^2\right) \\ &\quad + \sum_{k=0}^K \sum_{i_1=1}^4 \cdots \sum_{i_{K-k}=1}^4 \mathbb{P}\left(W_{Q_{i_1, \dots, i_{K-k}}}^2(\nu_{k-1}, \nu_k) \geq \frac{\alpha_k^2 t^2}{L^2} l_k^2\right). \end{aligned}$$

Posons alors $s_k := \frac{\alpha_k^2 t^2}{L^2} l_k^2$ et notons $Q' := Q_{i_1, \dots, i_{K-k}}$ un carré générique de côté $2l := l_k$ (la

dépendance en k sera claire dans le contexte). On a d'abord

$$\begin{aligned}
\mathbb{P}\left(W_{\mathcal{Q}'}^2(\mu, \nu_0) \geq s_0\right) &\leq \mathbb{P}\left(\mu(\mathcal{Q}') \geq \frac{s_0}{2l_0^2}\right) \leq \mathbb{P}\left(\mu(\mathcal{Q}') - \mathbb{E}\mu(\mathcal{Q}') \geq \frac{s_0 - \frac{2l_0^4}{\pi}}{2l_0^2}\right) \\
&\stackrel{(3.2.1)}{\leq} \exp\left(-\min\left\{\frac{\left(s_0 - \frac{2l_0^4}{\pi}\right)^2}{16l_0^4 \text{Var } \mu(\mathcal{Q}')}, \frac{s_0 - \frac{2l_0^4}{\pi}}{4l_0^2}\right\}\right) \\
&\leq \exp\left(-\min\left\{\frac{s_0^2}{256l_0^5}, \frac{s_0}{8l_0^2}\right\}\right), \tag{3.3.15}
\end{aligned}$$

où on a supposé que $s_0 > \frac{4l_0^4}{\pi}$, ce qui équivaut à $l_0 < \frac{\sqrt{\pi} \alpha_0 t}{2L}$, pour avoir $s_0 - \frac{2l_0^4}{\pi} > \frac{s_0}{2}$. Définissons maintenant l'ensemble

$$\mathcal{A} := \left\{ \forall i, \quad \mu(\mathcal{Q}'_i) \geq \frac{|\mathcal{Q}'_i|}{2\pi} = \frac{l^2}{2\pi} \right\},$$

qui permet d'écrire

$$\mathbb{P}\left(W_{\mathcal{Q}'}^2(\nu_{k-1}, \nu_k) \geq s_k\right) \leq \mathbb{P}\left(\left\{W_{\mathcal{Q}'}^2(\nu_{k-1}, \nu_k) \geq s_k\right\} \cap \mathcal{A}\right) + \mathbb{P}(\mathcal{A}^c). \tag{3.3.16}$$

Tout d'abord,

$$\begin{aligned}
\mathbb{P}(\mathcal{A}^c) &\leq \sum_{i=1}^4 \mathbb{P}\left(\mu(\mathcal{Q}'_i) \leq \frac{l^2}{2\pi}\right) \leq \sum_{i=1}^4 \mathbb{P}\left(\mu(\mathcal{Q}'_i) - \mathbb{E}\mu(\mathcal{Q}'_i) \leq -\frac{l^2}{4\pi}\right) \\
&\stackrel{(3.2.1)}{\leq} \sum_{i=1}^4 \exp\left(-\min\left\{\frac{l^4}{16\pi^2 \text{Var } \mu(\mathcal{Q}'_i)}, \frac{l^2}{8\pi}\right\}\right) \\
&\leq 4 \exp\left(-\min\left\{\frac{l^3}{64\pi^2}, \frac{l^2}{8\pi}\right\}\right). \tag{3.3.17}
\end{aligned}$$

Ensuite, grâce à l'inégalité (3.3.6) et les définitions de c et c_i ,

$$\begin{aligned}
\mathbb{P}\left(\left\{W_{\mathcal{Q}'}^2(\nu_{k-1}, \nu_k) \geq s_k\right\} \cap \mathcal{A}\right) &\leq \mathbb{P}\left(\frac{\pi}{2} \sum_{i=1}^4 (\mu(\mathcal{Q}') - 4\mu(\mathcal{Q}'_i))^2 \geq s_k\right) \\
&\leq \sum_{i=1}^4 \mathbb{P}\left(|\mu(\mathcal{Q}') - 4\mu(\mathcal{Q}'_i)| \geq \sqrt{\frac{s_k}{2\pi}}\right) \\
&\leq 4\mathbb{P}\left(|\mu(\mathcal{Q}') - \mathbb{E}\mu(\mathcal{Q}')| \geq \frac{1}{3}\sqrt{\frac{s_k}{2\pi}}\right) \\
&\quad + \sum_{i=1}^4 \mathbb{P}\left(|\mu(\mathcal{Q}'_i) - \mathbb{E}\mu(\mathcal{Q}'_i)| \geq \frac{1}{12}\sqrt{\frac{s_k}{2\pi}}\right) \\
&\quad + \sum_{i=1}^4 \mathbb{P}\left(|\mathbb{E}\mu(\mathcal{Q}') - 4\mathbb{E}\mu(\mathcal{Q}'_i)| \geq \frac{1}{12}\sqrt{\frac{s_k}{2\pi}}\right).
\end{aligned}$$

D'après le Lemme 3.2.1, on a $|\mathbb{E}\mu(Q') - 4\mathbb{E}\mu(Q'_i)| \leq \frac{5e}{N^{3/2}}$. Supposons alors que pour tout k , $\frac{1}{12}\sqrt{\frac{s_k}{2\pi}} > \frac{5e}{N^{3/2}}$, c'est-à-dire $\frac{\alpha_0 t}{L} l_0 > \frac{60e}{N^{3/2}}$, ce qui est vrai puisque $t \geq L$ et $l_0 \geq 1$. On obtient l'existence d'une constante $c > 0$ telle que

$$\begin{aligned} \mathbb{P}\left(\left\{W_{Q'}^2(\nu_{k-1}, \nu_k) \geq s_k\right\} \cap \mathcal{A}\right) &\leq 4 \exp\left(-\frac{1}{6} \min\left\{\frac{s_k}{12\pi \text{Var } \mu(Q')}, \sqrt{\frac{s_k}{2\pi}}\right\}\right) \\ &\quad + \sum_{i=1}^4 \exp\left(-\frac{1}{24} \min\left\{\frac{s_k}{48\pi \text{Var } \mu(Q'_i)}, \sqrt{\frac{s_k}{2\pi}}\right\}\right) \\ &\leq 8 \exp\left(-c \min\left\{\frac{s_k}{l}, \sqrt{s_k}\right\}\right). \end{aligned} \quad (3.3.18)$$

En rappelant qu'ici $l = l_{k-1} = \frac{l_k}{2}$, on obtient en combinant (3.3.15), (3.3.17) et (3.3.18) dans (3.3.16)

$$\begin{aligned} \mathbb{P}\left(W_Q\left(\mu, \frac{\mu(Q)}{|Q|}\right) \geq t\right) &\lesssim 4^K \exp\left(-c \min\left\{\frac{\alpha_0^4 t^4}{L^4 l_0}, \frac{\alpha_0^2 t^2}{L^2}\right\}\right) \\ &\quad + \sum_{k=1}^K 4^{K-k} \left(\exp\left(-\min\{l_{k-1}^3, l_{k-1}^2\}\right) + \exp\left(-c \min\left\{\frac{\alpha_k^2 t^2}{L^2} l_k, \frac{\alpha_k t}{L} l_k\right\}\right)\right). \end{aligned}$$

Il est maintenant temps de choisir les α_k . Il apparaît que le meilleur choix est

$$\alpha_0 := \frac{1}{\sum_{k=0}^K \frac{1}{\sqrt{2}^k}} \geq \frac{1}{2} \quad \text{and} \quad \forall k, \alpha_k := \frac{\alpha_0}{\sqrt{2}^k},$$

ce qui correspond d'ailleurs à l'estimation obtenue dans la sous-section précédente. On fixe également $l_0 := \frac{\sqrt{\pi}\alpha_0 t}{2L}$. Ceci est possible tant que $l_0 < L$, ce qui est vrai car $t < \frac{2}{\sqrt{\pi}}L^2$, et $l_0 \geq 1$, ce qui est vrai tant que $t \geq \frac{2}{\sqrt{\pi}\alpha_0}L$. Le terme dominant dans la somme est alors le terme $k = 1$, ce qui nous donne en utilisant l'égalité $4^K = \left(\frac{L}{l_0}\right)^2$ et en remplaçant l_0

$$\mathbb{P}\left(W_Q\left(\mu, \frac{\mu(Q)}{|Q|}\right) \geq t\right) \lesssim \frac{L^4}{t^2} \exp\left(-\min\left\{c_1 \frac{t^3}{L^3}, c_2 \frac{t^2}{L^2}\right\}\right) \lesssim \exp\left(-c \frac{t^2}{L^2} + C \ln\left(\frac{L^4}{t^2}\right)\right).$$

Il reste à remarquer que si $t \gg L\sqrt{\ln L}$ alors $\ln\left(\frac{L^4}{t^2}\right)$ est négligeable devant $\frac{t^2}{L^2}$. ■

3.4 Optimal matching dans un disque

3.4.1 Construction des mesures

Soit $R \leq \sqrt{N} - \sqrt{\ln N}$ et $D = D(0, R)$ et notons à nouveau $\mu := \mu_N$, on veut estimer

$$\mathbb{E}W_D^2\left(\mu, \frac{\mu(D)}{|D|}\right).$$

Pour cela, soit $\mathcal{Q} := [-R; R]^2$ et découpons \mathcal{Q} en 4 carrés disjoints de côté R : notons \mathcal{Q}_1 , respectivement $\mathcal{Q}_2, \mathcal{Q}_3, \mathcal{Q}_4$, le carré correspondant à $\mathfrak{X} \geq 0$ et $\mathfrak{Y} \geq 0$, respectivement $\mathfrak{X} \geq 0$ et $\mathfrak{Y} \leq 0$, $\mathfrak{X} \leq 0$ et $\mathfrak{Y} \leq 0$, $\mathfrak{X} \leq 0$ et $\mathfrak{Y} \geq 0$. Soit $K \in \mathbb{N}$ tel que $\theta_0 R := \frac{\pi}{2^{K+1}} R \in [1; 2)$. Définissons les ensembles $A_{0,0,K}^p := \mathcal{Q}_p \cap D$ pour $1 \leq p \leq 4$ et les mesures

$$\nu_{K+1} := \frac{\mu(D)}{|D|} dz|_D \quad \text{et} \quad \nu_K := \sum_{p=1}^4 \frac{\mu(A_{0,0,K}^p)}{|A_{0,0,K}^p|} dz|_{A_{0,0,K}^p}.$$

La construction que l'on va présenter sera symétrique par rapport aux axes (Ox) et (Oy) , on notera avec un exposant $p \in \{1, 2, 3, 4\}$ les domaines dans \mathcal{Q}_p définis à partir de ceux dans \mathcal{Q}_1 .

Soit $\theta_{K-1} := \frac{\pi}{4}$ et pour $i \in \{0, 1\}, j \in \{0, 1\}$,

$$\begin{aligned} \mathcal{Q}_{i,j,K-1} &:= [R \cos((i+1)\theta_{K-1}); R \cos(i\theta_{K-1})] \times [R \sin(j\theta_{K-1}); R \sin((j+1)\theta_{K-1})], \\ A_{i,j,K-1} &:= \mathcal{Q}_{i,j,K-1} \cap D. \end{aligned}$$

On définit alors

$$\nu_{K-1} := \sum_{p=1}^4 \sum_{i=0}^1 \sum_{j=0}^1 \frac{\mu(A_{i,j,K-1}^p)}{|A_{i,j,K-1}^p|} dz|_{A_{i,j,K-1}^p},$$

avec la convention $\frac{\mu(A_{i,j,K-1}^p)}{|A_{i,j,K-1}^p|} = 0$ si $|A_{i,j,K-1}^p| = 0$. On continue cette méthode jusqu'à des angles θ_0 . Plus précisément, à l'étape k , on définit $\theta_{K-k} := \frac{\pi}{2^{k+1}}$ et pour tous $0 \leq i \leq 2^k - 1, 0 \leq j \leq 2^k - 1$,

$$\nu_{K-k} := \sum_{p=1}^4 \sum_{i=0}^{2^k-1} \sum_{j=0}^{2^k-1} \frac{\mu(A_{i,j,K-k}^p)}{|A_{i,j,K-k}^p|} dz|_{A_{i,j,K-k}^p},$$

avec la convention $\frac{\mu(A_{i,j,K-k}^p)}{|A_{i,j,K-k}^p|} = 0$ si $|A_{i,j,K-k}^p| = 0$, où

$$\begin{aligned} \mathcal{Q}_{i,j,K-k} &:= [R \cos((i+1)\theta_{K-k}); R \cos(i\theta_{K-k})] \times [R \sin(j\theta_{K-k}); R \sin((j+1)\theta_{K-k})], \\ A_{i,j,K-k} &:= \mathcal{Q}_{i,j,K-k} \cap D. \end{aligned}$$

Rassemblons quelques propriétés utiles :

Lemme 3.4.1. (i) $A_{i,j,k} = \mathcal{Q}_{i,j,k} \Leftrightarrow i > j$ et $\mathcal{Q}_{i,j,k} \cap \partial D \neq \emptyset \Leftrightarrow i = j$.

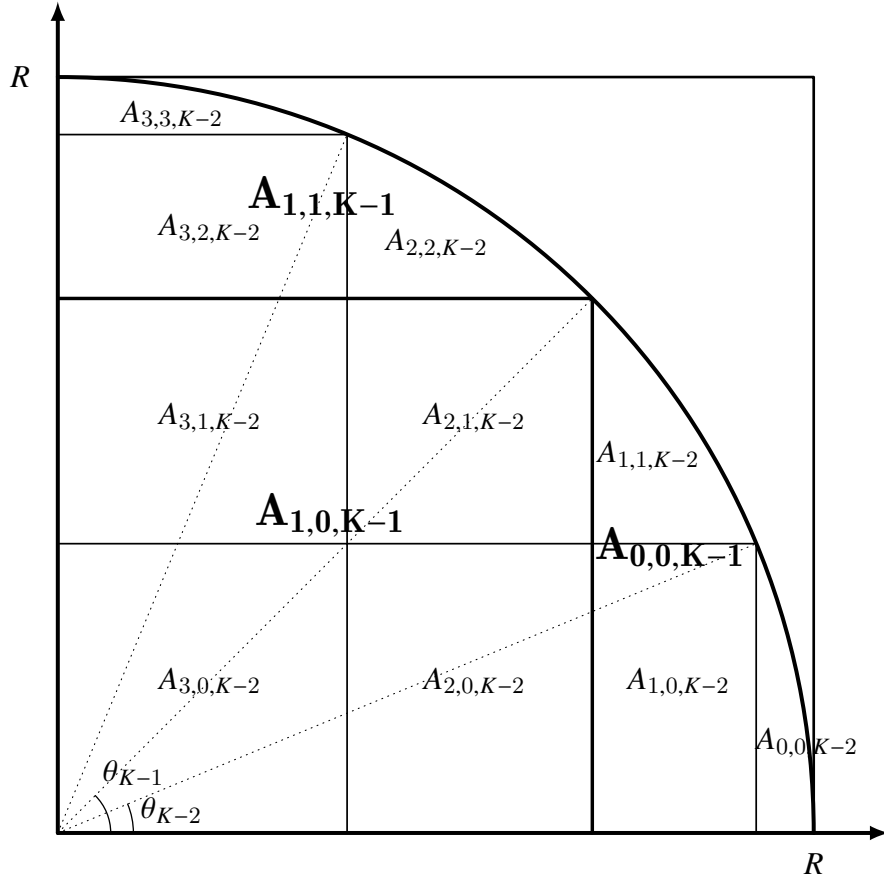
(ii) En notant L la longueur et H la hauteur,

$$L(\mathcal{Q}_{i,j,k}) = 2R \sin \frac{\theta_k}{2} \sin \left(\frac{2i+1}{2} \theta_k \right) \quad \text{et} \quad H(\mathcal{Q}_{i,j,k}) = 2R \sin \frac{\theta_k}{2} \cos \left(\frac{2j+1}{2} \theta_k \right),$$

et si $i \geq j$, $L(A_{i,j,k}) = L(\mathcal{Q}_{i,j,k})$ et $H(A_{i,j,k}) = H(\mathcal{Q}_{i,j,k})$.

(iii) $L(\mathcal{Q}_{i,j,k}) \leq H(\mathcal{Q}_{i,j,k}) \Leftrightarrow i + j \leq 2^{K-k}$.

(iv) $\mathcal{Q}_{i,j,k} = \mathcal{Q}_{2i,2j,k-1} \sqcup \mathcal{Q}_{2i,2j+1,k-1} \sqcup \mathcal{Q}_{2i+1,2j,k-1} \sqcup \mathcal{Q}_{2i+1,2j+1,k-1}$ où \sqcup dénote une union disjointe.


 FIGURE 3.2: Construction des domaines $A_{i,j,k}$.

$$(v) |Q_{i,j,k}| = 4R^2 \sin^2 \frac{\theta_k}{2} \sin \left(\frac{2i+1}{2} \theta_k \right) \cos \left(\frac{2j+1}{2} \theta_k \right) \text{ et si } i \geq j, \frac{|Q_{i,j,k}|}{2} \leq |A_{i,j,k}| \leq |Q_{i,j,k}|.$$

(vi) $Q_{i,j,k}$ est le symétrique de $Q_{2^{K-k}-j-1, 2^{K-k}-i-1, k}$ par rapport à la droite $y = x$. En particulier,

$$L(A_{i,j,k}) = H(A_{2^{K-k}-j-1, 2^{K-k}-i-1, k}) \text{ et } H(A_{i,j,k}) = L(A_{2^{K-k}-j-1, 2^{K-k}-i-1, k}).$$

On déduit des Lemmes 3.2.2 et 3.2.5 les estimations suivantes de variance : si $i > j$ alors

$$\text{Var } \mu(A_{i,j,k}) \leq 2(L(A_{i,j,k}) + H(A_{i,j,k})), \quad (3.4.1)$$

et si $i = j$ avec $L(A_{i,i,k}) \geq 1$ et $H(A_{i,i,k}) \geq 1$,

$$\text{Var } \mu(A_{i,i,k}) \leq 11|A_{i,i,k}|. \quad (3.4.2)$$

Enfin, on donne une estimation du rapport de surface entre un rectangle de la construction et un de ses sous-rectangles.

Lemme 3.4.2. Pour tout $1 \leq k \leq K$, tous $0 \leq i \leq 2^{K-k} - 1$, $0 \leq j \leq 2^{K-k} - 1$ et tout $\varepsilon \in \{0, 1\}^2$,

$$1 \leq \frac{|Q_{i,j,k}|}{|Q_{2^{i+\varepsilon_1}, 2^{j+\varepsilon_2}, k-1}|} \leq \pi^4.$$

Démonstration. Par croissance de sin et décroissance de cos sur $[0; \frac{\pi}{2}]$,

$$\begin{aligned} \frac{|Q_{i,j,k}|}{|Q_{2i+\varepsilon_1, 2j+\varepsilon_2, k-1}|} &\leq \frac{|Q_{i,j,k}|}{|Q_{2i, 2j+1, k-1}|} \\ &= \frac{\sin^2\left(\frac{\theta_k}{2}\right)}{\sin^2\left(\frac{\theta_k}{4}\right)} \cdot \frac{\sin\left(\frac{2i+1}{2}\theta_k\right) \cos\left(\frac{2j+1}{2}\theta_k\right)}{\sin\left(\frac{4i+1}{4}\theta_k\right) \cos\left(\frac{4j+3}{4}\theta_k\right)}. \end{aligned}$$

Or $\cos\left(\frac{2j+1}{2}\theta_k\right) = \sin\left(\frac{\pi}{2} - \frac{2j+1}{2}\theta_k\right) = \sin\left((2^{K-k} - j - \frac{1}{2})\frac{\theta_k}{2}\right)$ et sur $[0; \frac{\pi}{2}]$, $x \geq \sin x \geq \frac{2}{\pi}x$, d'où

$$\begin{aligned} \frac{|Q_{i,j,k}|}{|Q_{2i+\varepsilon_1, 2j+\varepsilon_2, k-1}|} &\leq \frac{\pi^4}{4} \cdot \frac{i + \frac{1}{2}}{i + \frac{1}{4}} \cdot \frac{2^{K-k} - j - \frac{1}{2}}{2^{K-k} - j - \frac{3}{4}} \\ &= \frac{\pi^4}{4} \left(1 + \frac{\frac{1}{4}}{i + \frac{1}{4}}\right) \left(1 + \frac{\frac{1}{4}}{2^{K-k} - j - \frac{3}{4}}\right) \leq \pi^4, \end{aligned}$$

car $i \geq 0$ et $j \leq 2^{K-k} - 1$. ■

3.4.2 Estimation de la distance

Avec la construction que l'on vient de présenter, on va estimer

$$\mathbb{E}W_D^2\left(\mu, \frac{\mu(D)}{|D|}\right)$$

en utilisant la même stratégie que dans le cas du carré, dans la section 3.3. Tout d'abord, par inégalité triangulaire, on a

$$\left(\mathbb{E}W_D^2\left(\mu, \frac{\mu(D)}{|D|}\right)\right)^{\frac{1}{2}} \leq \left(\mathbb{E}W_2^2(\mu|_D, \nu_0)\right)^{\frac{1}{2}} + \sum_{k=1}^{K+1} \left(\mathbb{E}W_2^2(\nu_{k-1}, \nu_k)\right)^{\frac{1}{2}}. \quad (3.4.3)$$

ETAPE 1 (Distance de μ à ν_0). Pour tous $0 \leq i \leq 2^K - 1$ et $0 \leq j \leq 2^K - 1$, on a $\nu_0(A_{i,j,0}) = \mu(A_{i,j,0})$ donc

$$\begin{aligned} \mathbb{E}W_2^2(\mu|_D, \nu_0) &\leq \sum_{p=1}^4 \sum_{i=0}^{2^K-1} \sum_{j=0}^{2^K-1} \mathbb{E}W_{A_{i,j,0}}^2(\mu, \nu_0) \\ &= 4 \sum_{i=0}^{2^K-1} \sum_{j=0}^{2^K-1} \mathbb{E}W_{A_{i,j,0}}^2(\mu, \nu_0). \end{aligned} \quad (3.4.4)$$

Or

$$W_{A_{i,j,0}}^2(\mu, \nu_0) \leq \left(\sqrt{2} \max\{L(A_{i,j,0}), H(A_{i,j,0})\}\right)^2 \mu(A_{i,j,0}),$$

donc en utilisant le Lemme 3.4.1 (ii) et en bornant $\cos\left(\frac{2j+1}{2}\theta_0\right)$ et $\sin\left(\frac{2i+1}{2}\theta_0\right)$ par 1 et $\sin\frac{\theta_0}{2}$ par $\frac{\theta_0}{2}$, on obtient

$$\mathbb{E}W_{A_{i,j,0}}^2(\mu, \nu_0) \leq \frac{2}{\pi}(\theta_0 R)^4 \leq \frac{32}{\pi},$$

ce qui fait que (3.4.4) devient, en rappelant que $2^{K+1} \leq \pi R$,

$$\mathbb{E}W_2^2(\mu|_D, \nu_0) \leq 32\pi R^2. \quad (3.4.5)$$

ETAPE 2 (Distance de ν_K à ν_{K+1}). En suivant la même stratégie que dans le cas du carré dans la section précédente, on montre que

$$\mathbb{E}W_2^2(\nu_K, \nu_{K+1}) \lesssim R^2. \quad (3.4.6)$$

En effet, en définissant φ solution du problème de Poisson correspondant, on a grâce à (II.7),

$$\begin{aligned} \mathbb{E} \left[W_2^2(\nu_K, \nu_{K+1}) \chi_{\left\{ \frac{\mu_N(Q_p \cap D)}{|Q_p \cap D|} > \frac{1}{2\pi}, \forall i \right\}} \right] &\leq 2\pi \mathbb{E} \int_D |\nabla \varphi|^2 \leq 2\pi |D| \mathbb{E} \int_D |\nu_{K+1} - \nu_K|^2 \\ &= \frac{\pi}{2} \sum_{p=1}^4 \mathbb{E} \left[(\mu(D) - 4\mu(Q_p \cap D))^2 \right] \\ &\leq 4\pi \text{Var } \mu(D) + 16\pi \sum_{p=1}^4 \text{Var } \mu(Q_p \cap D) \\ &\stackrel{(3.2.21) \& (3.4.2)}{\leq} 8\pi R + 16\pi \cdot 11\pi R^2 \lesssim R^2. \end{aligned} \quad (3.4.7)$$

De plus, en notant $\chi_{\leq} := \chi_{\left\{ \exists i, \frac{\mu(Q_p \cap D)}{|Q_p \cap D|} \leq \frac{1}{2\pi} \right\}}$,

$$\begin{aligned} \mathbb{E} \left[W_2^2(\nu_K, \nu_{K+1}) \chi_{\leq} \right] &= \mathbb{E} \left[W_2^2(\nu_K, \nu_{K+1}) \chi_{\leq} \chi_{\left\{ \mu(D) \leq \frac{4|D|}{\pi} \right\}} \right] \\ &\quad + \sum_{s=2}^{\infty} \mathbb{E} \left[W_2^2(\nu_K, \nu_{K+1}) \chi_{\leq} \chi_{\left\{ \frac{s^2|D|}{\pi} < \mu(D) \leq \frac{(s+1)^2|D|}{\pi} \right\}} \right] \\ &\leq 2R^2 \frac{|D|}{\pi} \left(4\mathbb{P} \left(\exists p, \frac{\mu(Q_p \cap D)}{|Q_p \cap D|} \leq \frac{1}{2\pi} \right) + \sum_{l=2}^{\infty} (s+1)^2 \mathbb{P} \left(\frac{s^2|D|}{\pi} < \mu(D) \right) \right), \end{aligned}$$

ce qui, avec les estimations

$$\begin{aligned} \mathbb{P} \left(\frac{\mu(Q_p \cap D)}{|Q_p \cap D|} \leq \frac{1}{2\pi} \right) &\leq \frac{16\pi^2 \text{Var } \mu(Q_p \cap D)}{|Q_p \cap D|^2} \\ \text{et } \mathbb{P} \left(\frac{s^2|D|}{\pi} < \mu(D) \right) &\leq \frac{\pi^2}{(s^2 - 1)^2} \cdot \frac{\text{Var } \mu(D)}{|D|^2}, \end{aligned}$$

ainsi que (3.2.21) et (3.4.2), donne

$$\mathbb{E} \left[W_2^2(\nu_K, \nu_{K+1}) \chi_{\leq} \right] \lesssim R^2. \quad (3.4.8)$$

En combinant (3.4.7) et (3.4.8), on obtient (3.4.6).

ETAPE 3 (Distance de ν_{k-1} à ν_k). Soit maintenant $1 \leq k \leq K$. Pour tous $0 \leq i \leq 2^{K-k} - 1$ et $0 \leq j \leq 2^{K-k} - 1$, on a $\nu_k(A_{i,j,k}) = \nu_{k-1}(A_{i,j,k})$ donc

$$\begin{aligned} \mathbb{E}W_2^2(\nu_{k-1}, \nu_k) &\leq \sum_{p=1}^4 \sum_{i=0}^{2^{K-k}-1} \sum_{j=0}^{2^{K-k}-1} \mathbb{E}W_{A_{i,j,k}}^2(\nu_{k-1}, \nu_k) \\ &= 4 \sum_{i=0}^{2^{K-k}-1} \sum_{j=0}^{i-1} \mathbb{E}W_{Q_{i,j,k}}^2(\nu_{k-1}, \nu_k) + 4 \sum_{i=0}^{2^{K-k}-1} \mathbb{E}W_{A_{i,i,k}}^2(\nu_{k-1}, \nu_k), \end{aligned} \quad (3.4.9)$$

car si $j > i$, $A_{i,j,k} = \emptyset$.

ETAPE 3.A. (Domaines dont le bord est trop court). On commence par traiter un cas particulier : les domaines $A_{i,j,k}$ (non vides) tels que $L(A_{i,j,k}) < 1$ ou $H(A_{i,j,k}) < 1$. Commençons par en estimer le nombre :

$$\begin{aligned} L(A_{i,j,k}) &= 2R \sin\left(\frac{\theta_k}{2}\right) \sin\left(\frac{2i+1}{2}\theta_k\right) \\ &\geq (2i+1) \frac{2R\theta_k^2}{\pi^2} \geq (2i+1) \frac{4^k}{2^k \pi}, \end{aligned}$$

d'où

$$L(A_{i,j,k}) < 1 \implies 2i+1 < 2^{K-2k} \pi \implies i < 2^{K-2k-1} \pi.$$

En particulier, ce n'est pas possible si $k \geq \frac{K}{2}$. Par symétrie,

$$H(A_{i,j,k}) < 1 \implies 2^{K-k+1} - 2j - 1 < 2^{K-2k} \pi \implies 2(2^{K-k} - j) - 1 < 2^{K-2k} \pi.$$

Pour ces domaines, on a l'estimation

$$\mathbb{E}W_{A_{i,j,k}}^2(\nu_{k-1}, \nu_k) \leq 2 \max\{L(A_{i,j,k}), H(A_{i,j,k})\}^2 \frac{|A_{i,j,k}|}{\pi} \leq \frac{2}{\pi} \max\{L(A_{i,j,k}), H(A_{i,j,k})\}^3.$$

Or, en utilisant à nouveau le Lemme 3.4.1 (ii),

$$\max\{L(A_{i,j,k}), H(A_{i,j,k})\} \leq R\theta_k \leq \frac{2^{K+2}}{\pi} \cdot \frac{\pi}{2^{K-k+1}} = 2^{k+1},$$

donc

$$\mathbb{E}W_{A_{i,j,k}}^2(\nu_{k-1}, \nu_k) \leq 2^{3k} \frac{16}{\pi}.$$

On a donc

$$\begin{aligned} \sum_{\substack{L(A_{i,j,k}) < 1 \\ \text{ou } H(A_{i,j,k}) < 1}} \mathbb{E}W_{A_{i,j,k}}^2(\nu_{k-1}, \nu_k) &\leq 2 \sum_{i=0}^{2^{K-2k}} \sum_{j=0}^i \mathbb{E}W_{A_{i,j,k}}^2(\nu_{k-1}, \nu_k) \\ &\leq 2^{3k} \frac{32}{\pi} \cdot 2^{2K-4k} \leq 16 \cdot 2^{K-k} R. \end{aligned} \quad (3.4.10)$$

Dans la suite de l'ETAPE 3, on suppose implicitement que les domaines que l'on considère ont des côtés plus grands que 1, de façon à ce qu'on puisse utiliser les estimations de variance obtenues dans les Lemmes 3.2.2 et 3.2.5.

ETAPE 3.B. (Cas d'un rectangle). Soit $i > j$, on note $Q := Q_{i,j,k}$ et pour $\varepsilon \in \{0,1\}^2$, $Q_\varepsilon := Q_{2i+\varepsilon_1, 2j+\varepsilon_2, k-1}$ ses 4 sous-rectangles. Comme précédemment, en notant φ la solution du problème de Poisson

$$\begin{cases} -\Delta\varphi = v_k - v_{k-1} & \text{sur } Q, \\ \nabla\varphi \cdot n = 0 & \text{sur } \partial Q, \end{cases}$$

on a, en notant $\chi_\geq := \chi_{\{\frac{\mu(Q_\varepsilon)}{|Q_\varepsilon|} \geq \frac{1}{2\pi}, \forall \varepsilon \in \{0,1\}^2\}}$,

$$\begin{aligned} \mathbb{E} \left[W_Q^2(v_{k-1}, v_k) \chi_\geq \right] &\leq 2\pi \mathbb{E} \int_Q |\nabla\varphi|^2 \leq 2\pi |Q| \mathbb{E} \int_Q |v_k - v_{k-1}|^2 \\ &= 2\pi |Q| \sum_{\varepsilon \in \{0,1\}^2} |Q_\varepsilon| \mathbb{E} \left[\left(\frac{\mu(Q)}{|Q|} - \frac{\mu(Q_\varepsilon)}{|Q_\varepsilon|} \right)^2 \right] \\ &\lesssim |Q| \sum_{\varepsilon \in \{0,1\}^2} |Q_\varepsilon| \frac{\text{Var } \mu(Q)}{|Q|^2} + |Q| \sum_{\varepsilon \in \{0,1\}^2} \frac{\text{Var } \mu(Q_\varepsilon)}{|Q_\varepsilon|} \\ &\quad + |Q| \sum_{\varepsilon \in \{0,1\}^2} |Q_\varepsilon| \left(\frac{\mathbb{E}\mu(Q)}{|Q|} - \frac{\mathbb{E}\mu(Q_\varepsilon)}{|Q_\varepsilon|} \right)^2. \end{aligned}$$

On utilise alors l'égalité $\sum_{\varepsilon \in \{0,1\}^2} |Q_\varepsilon| = |Q|$ et les Lemmes 3.2.1 et 3.4.2 pour obtenir

$$\mathbb{E} \left[W_Q^2(v_{k-1}, v_k) \chi_\geq \right] \lesssim \text{Var } \mu(Q) + \sum_{\varepsilon \in \{0,1\}^2} \text{Var } \mu(Q_\varepsilon) + \frac{1}{N^3}. \quad (3.4.11)$$

Notons maintenant $\chi_\leq := \chi_{\{\exists \varepsilon \in \{0,1\}^2, \frac{\mu(Q_\varepsilon)}{|Q_\varepsilon|} \leq \frac{1}{2\pi}\}}$,

$$\begin{aligned} \mathbb{E} \left[W_Q^2(v_{k-1}, v_k) \chi_\leq \right] &= \mathbb{E} \left[W_Q^2(v_{k-1}, v_k) \chi_\leq \chi_{\{\mu(Q) \leq \frac{4|Q|}{\pi}\}} \right] + \sum_{s=2}^{\infty} \mathbb{E} \left[W_Q^2(v_{k-1}, v_k) \chi_\leq \chi_{\{\frac{s^2|Q|}{\pi} \leq \mu(Q) \leq \frac{(s+1)^2|Q|}{\pi}\}} \right] \\ &\leq (\sqrt{2} \max\{L(Q), H(Q)\})^2 \frac{4|Q|}{\pi} \mathbb{P} \left(\exists \varepsilon \in \{0,1\}^2, \frac{\mu(Q_\varepsilon)}{|Q_\varepsilon|} \leq \frac{1}{2\pi} \right) \\ &\quad + (\sqrt{2} \max\{L(Q), H(Q)\})^2 \frac{|Q|}{\pi} \sum_{s=2}^{\infty} (s+1)^2 \mathbb{P} \left(\mu(Q) \geq \frac{s^2|Q|}{\pi} \right). \end{aligned} \quad (3.4.12)$$

Or, si N est suffisamment grand pour que $\frac{e}{N^{\frac{3}{2}}} \leq \frac{|Q_\varepsilon|}{4\pi}$ ($|Q_\varepsilon| \geq 1$ puisqu'on a supposé que les côtés sont plus grands que 1), on a pour $\varepsilon \in \{0,1\}^2$ en utilisant le Lemme 3.2.1,

$$\mathbb{P} \left(\frac{\mu(Q_\varepsilon)}{|Q_\varepsilon|} \leq \frac{1}{2\pi} \right) \leq \frac{16\pi^2 \text{Var } \mu(Q_\varepsilon)}{|Q_\varepsilon|^2}.$$

De la même façon, pour tout $s \geq 2$,

$$\mathbb{P} \left(\mu(Q) \geq \frac{s^2|Q|}{\pi} \right) \leq \frac{\pi^2 \text{Var } \mu(Q)}{|Q|^2 (s^2 - 1)^2},$$

donc, en utilisant à nouveau le Lemme 3.4.2 et le fait que $\sum_{s=2}^{\infty} \frac{(s+1)^2}{(s^2-1)^2} < \infty$, (3.4.12) devient

$$\begin{aligned} \mathbb{E} \left[W_Q^2(\nu_{k-1}, \nu_k) \chi_{\leq} \right] &\lesssim \frac{\max\{L(Q), H(Q)\}^2}{|Q|} \left(\sum_{\varepsilon \in \{0,1\}^2} \text{Var } \mu(Q_\varepsilon) + \text{Var } \mu(Q) \right) \\ &\lesssim \frac{\max\{L(Q), H(Q)\}}{\min\{L(Q), H(Q)\}} \left(\sum_{\varepsilon \in \{0,1\}^2} \text{Var } \mu(Q_\varepsilon) + \text{Var } \mu_N(Q) \right). \end{aligned} \quad (3.4.13)$$

En combinant (3.4.11) et (3.4.13), on obtient finalement

$$\mathbb{E} W_Q^2(\nu_{k-1}, \nu_k) \lesssim \frac{\max\{L(Q), H(Q)\}}{\min\{L(Q), H(Q)\}} \left(\text{Var } \mu(Q) + \sum_{\varepsilon \in \{0,1\}^2} \text{Var } \mu(Q_\varepsilon) \right),$$

ce qui donne grâce au Lemme 3.2.2,

$$\mathbb{E} W_Q^2(\nu_{k-1}, \nu_k) \lesssim \frac{\max\{L(Q), H(Q)\}^2}{\min\{L(Q), H(Q)\}}. \quad (3.4.14)$$

Grâce à la symétrie précisée dans le Lemme 3.4.1 (vi), on s'intéresse maintenant exclusivement au cas $i + j \leq 2^{K-k}$, c'est-à-dire au cas où $L(Q_{i,j,k}) \leq H(Q_{i,j,k})$. On a alors

$$\frac{H(Q_{i,j,k})^2}{L(Q_{i,j,k})} = 2R \sin\left(\frac{\theta_k}{2}\right) \frac{\cos^2\left(\frac{2j+1}{2}\theta_k\right)}{\sin\left(\frac{2i+1}{2}\theta_k\right)} \leq \frac{\pi R}{2i+1}. \quad (3.4.15)$$

Il reste à sommer sur tous les rectangles, ce qui donne

$$\begin{aligned} \sum_{i=0}^{2^{K-k}-1} \sum_{j=0}^{i-1} \mathbb{E} W_{R_{i,j,k}}^2(\nu_{k-1}, \nu_k) &\leq 2 \sum_{i=0}^{2^{K-k}-1} \sum_{j=0}^{\min\{i-1, 2^{K-k}-i\}} \mathbb{E} W_{Q_{i,j,k}}^2(\nu_{k-1}, \nu_k) \\ &\stackrel{(3.4.14)}{\lesssim} \sum_{i=0}^{2^{K-k}-1} \sum_{j=0}^{\min\{i-1, 2^{K-k}-i\}} \frac{H(Q_{i,j,k})^2}{L(Q_{i,j,k})} \\ &\stackrel{(3.4.15)}{\lesssim} R \sum_{i=0}^{2^{K-k}-1} \sum_{j=0}^{i-1} \frac{1}{2i+1} \\ &= R \sum_{i=0}^{2^{K-k}-1} \frac{i}{2i+1} \leq 2^{K-k-1} R. \end{aligned} \quad (3.4.16)$$

ETAPE 3.C. (Cas d'un domaine touchant le cercle). Soit $0 \leq i \leq 2^{K-k} - 1$, on note ici $A := A_{i,i,k}$ et $Q_A := Q_{2i+1, 2i, k-1}$ le sous-rectangle contenu dans A , ainsi que $A_\varepsilon := A_{2i+\varepsilon, 2i+\varepsilon, k-1}$ pour $\varepsilon \in \{0, 1\}$. Les calculs précédents s'appliquent, grâce en particulier au Lemme 3.4.1 (v), jusqu'à obtenir

$$\mathbb{E} W_A^2(\nu_{k-1}, \nu_k) \lesssim \frac{\max\{L(A), H(A)\}}{\min\{L(A), H(A)\}} \left(\text{Var } \mu(A) + \text{Var } \mu(Q_A) + \sum_{\varepsilon \in \{0,1\}} \text{Var } \mu(A_\varepsilon) \right),$$

ce qui donne, en utilisant (3.4.2),

$$\mathbb{E}W_A^2(\nu_{k-1}, \nu_k) \lesssim \frac{\max\{L(A), H(A)\}}{\min\{L(A), H(A)\}} |A| \leq \max\{L(A), H(A)\}^2. \quad (3.4.17)$$

Or $\max\{L(A), H(A)\}^2 \leq R^2 \theta_k^2 \lesssim \frac{R^2}{4^{K-k}}$, donc en sommant les contributions de la forme (3.4.17), on a

$$\sum_{i=0}^{2^{K-k}-1} \mathbb{E}W_{A_{i,i,k}}^2(\nu_{k-1}, \nu_k) \lesssim \frac{R^2}{2^{K-k}} \lesssim 2^k R. \quad (3.4.18)$$

ETAPE 4 (Estimation finale). En injectant les estimations obtenues dans (3.4.10), (3.4.16) et (3.4.18) dans (3.4.9), on trouve

$$\mathbb{E}W_2^2(\nu_{k-1}, \nu_k) \lesssim (2^k + 2^{K-k})R,$$

ce qui donne

$$\sum_{k=1}^K (\mathbb{E}W_2^2(\nu_{k-1}, \nu_k))^{\frac{1}{2}} \lesssim R,$$

et finalement, avec (3.4.5) et (3.4.6), (3.4.3) devient

$$\mathbb{E}W_D^2\left(\mu, \frac{\mu(D)}{|D|}\right) \lesssim R^2.$$

On a donc prouvé l'estimation suivante.

Théorème 3.4.1. *Soit $D := D(0, R_N)$ avec $R_N \leq \sqrt{N} - \sqrt{\ln N}$ alors*

$$\mathbb{E}W_D^2\left(\mu_N, \frac{\mu_N(D)}{|D|}\right) \lesssim R_N^2.$$

3.5 Optimal matching dans un anneau

On définit $\mathbb{D}' := D(0, \sqrt{N}(1 - \varepsilon_N))$ en rappelant (3.2.19), $A := \mathbb{D}_N \setminus \mathbb{D}'$, $\tilde{\mu} := \mu|_{\mathbb{C} \setminus \mathbb{D}'}$ où on note encore $\mu := \mu_N$, et $\kappa := \frac{\mu(\mathbb{C} \setminus \mathbb{D}')}{|A|}$. On veut estimer

$$\mathbb{E}W_2^2(\tilde{\mu}, \kappa dz|_A).$$

Le domaine de restriction de μ_N est ici non borné, on a donc besoin d'un résultat a priori précisant le comportement du rayon spectral de notre matrice aléatoire. D'après [Rid03, Preuve du Th. 2], pour tout $\delta > 2$, on a

$$\mathbb{P}\left(\frac{\rho(M_N)}{\sqrt{N}} > 1 + \delta \varepsilon_N\right) \leq N^{-\frac{\delta}{2}}, \quad (3.5.1)$$

où $\rho(M_N)$ est le rayon spectral de M_N . Ceci suggère la décomposition suivante

$$\mathbb{E}W_2^2(\tilde{\mu}, \kappa dz|_A) = \mathbb{E} \left[W_2^2(\tilde{\mu}, \kappa dz|_A) \chi_{\{\rho(M_N) \leq \sqrt{N}(1+3\varepsilon_N)\}} \right] \quad (3.5.2)$$

$$+ \sum_{k=3}^{\infty} \mathbb{E} \left[W_2^2(\tilde{\mu}, \kappa dz|_A) \chi_{\{\sqrt{N}(1+k\varepsilon_N) < \rho(M_N) \leq \sqrt{N}(1+(k+1)\varepsilon_N)\}} \right]. \quad (3.5.3)$$

Chaque terme de la somme dans (3.5.3) peut être estimé de la façon suivante, en supposant $1 + (k+1)\varepsilon_N \leq k$ et en notant $\chi_k := \chi_{\{\sqrt{N}(1+k\varepsilon_N) < \rho(M_N) \leq \sqrt{N}(1+(k+1)\varepsilon_N)\}}$,

$$\begin{aligned} \mathbb{E} \left[W_2^2(\tilde{\mu}, \kappa dz|_A) \chi_k \right] &\leq \mathbb{E} \left[4k^2 N \mu_N(\mathbb{C} \setminus \mathbb{D}') \chi_k \right] \\ &\leq 4k^2 N^2 \mathbb{P} \left(\frac{\rho(M_N)}{\sqrt{N}} > 1 + k\varepsilon_N \right) \leq 4N^2 \frac{k^2}{N^{\frac{k}{2}}}. \end{aligned}$$

Ainsi, (3.5.3) devient

$$\sum_{k=3}^{\infty} \mathbb{E} \left[W_2^2(\tilde{\mu}, \kappa dz|_A) \chi_k \right] \lesssim \sqrt{N}. \quad (3.5.4)$$

Il reste donc à estimer (3.5.2), qui correspond à l'espérance de la distance d'appariement dans le cas où toutes les valeurs propres se trouvent dans $D(0, \sqrt{N}(1+3\varepsilon_N))$, c'est-à-dire la situation où $\tilde{\mu} = \mu|_B$ et $\kappa = \frac{\mu(B)}{|A|}$, où

$$B := D(0, \sqrt{N}(1+3\varepsilon_N)) \setminus \mathbb{D}'.$$

D'où, en notant $\chi_\rho := \chi_{\{\rho(M_N) \leq \sqrt{N}(1+3\varepsilon_N)\}}$,

$$\mathbb{E} \left[W_2^2(\tilde{\mu}, \kappa dz|_A) \chi_\rho \right] = \mathbb{E} \left[W_2^2 \left(\mu_N|_B, \frac{\mu_N(B)}{|A|} dz|_A \right) \chi_\rho \right]. \quad (3.5.5)$$

Pour estimer (3.5.5), la stratégie est à nouveau la même que dans les cas du carré et du disque. On découpe A et B en deux arcs d'anneau, chacun d'angle $\theta_{K-1} := \pi$,

$$\begin{aligned} B_1 &:= \left\{ z \in \mathbb{C} \mid \sqrt{N}(1-\varepsilon_N) < |z| < \sqrt{N}(1+3\varepsilon_N) \text{ et } \arg(z) \in]0, \pi[\right\}, \quad A_1 := B_1 \cap \mathbb{D}_N; \\ B_2 &:= \left\{ z \in \mathbb{C} \mid \sqrt{N}(1-\varepsilon_N) < |z| < \sqrt{N}(1+3\varepsilon_N) \text{ et } \arg(z) \in]\pi, 2\pi[\right\}, \quad A_2 := B_2 \cap \mathbb{D}_N. \end{aligned}$$

Définissons $\nu_{K+1} := \kappa dz|_A$ et

$$\nu_K := \sum_{i=1}^2 \frac{\mu(B_i)}{|A_i|} dz|_{A_i}.$$

On découpe chacun de ces arcs en deux sous-arcs $B_{1,1}, B_{1,2}, B_{2,1}, B_{2,2}$ et $A_{1,1}, A_{1,2}, A_{2,1}, A_{2,2}$, chacun de ces sous-arcs correspondant à un angle $\theta_{K-2} := \frac{\pi}{2}$. La mesure associée est

$$\nu_{K-1} := \sum_{i=1}^2 \sum_{j=1}^2 \frac{\mu(B_{i,j})}{|A_{i,j}|} dz|_{A_{i,j}}.$$

On continue ces découpages jusqu'à des arcs d'angle $\theta_0 \in [\varepsilon_N; 2\varepsilon_N[$ choisi de façon à ce qu'il existe $K \in \mathbb{N}$ tel que $2\pi = 2^K\theta_0$. Ce faisant, on a défini les mesures ν_{K-2}, \dots, ν_1 , la mesure ν_k étant constante sur des arcs d'angle $\theta_{k-1} := 2^{k-1}\theta_0$. Comme précédemment, l'inégalité triangulaire donne

$$\mathbb{E} \left[W_2^2 \left(\mu|_B, \frac{\mu(B)}{|A|} dz|_A \right) \chi_\rho \right]^{1/2} \leq \mathbb{E} \left[W_2^2(\mu|_B, \nu_1) \chi_\rho \right]^{1/2} + \sum_{k=1}^K \mathbb{E} \left[W_2^2(\nu_k, \nu_{k+1}) \chi_\rho \right]^{1/2}.$$

L'admissibilité du transport arc par arc donne ensuite

$$\mathbb{E} \left[W_2^2(\mu|_B, \nu_1) \chi_\rho \right] \leq \sum_{i_1=1}^2 \cdots \sum_{i_K=1}^2 \mathbb{E} \left[W_2^2(\mu|_{B_{i_1, \dots, i_K}}, \nu_1|_{A_{i_1, \dots, i_K}}) \chi_\rho \right] \quad (3.5.6)$$

$$\text{et } \mathbb{E} \left[W_2^2(\nu_k, \nu_{k+1}) \chi_\rho \right] \leq \sum_{i_1=1}^2 \cdots \sum_{i_{K-k}=1}^2 \mathbb{E} \left[W_{A_{i_1, \dots, i_{K-k}}}^2(\nu_k, \nu_{k+1}) \chi_\rho \right]. \quad (3.5.7)$$

Pour (3.5.6), on estime grossièrement le transport par

$$W_2^2(\mu|_{B_{i_1, \dots, i_K}}, \nu_1|_{A_{i_1, \dots, i_K}}) \leq (2\sqrt{N}\varepsilon_N)^2 \mu(B_{i_1, \dots, i_K})$$

de sorte que, par isotropie radiale de la distribution des valeurs propres et en utilisant le Lemme 3.2.1,

$$\begin{aligned} \mathbb{E} \left[W_2^2(\mu|_{B_{i_1, \dots, i_K}}, \nu_1|_{A_{i_1, \dots, i_K}}) \chi_\rho \right] &\leq 4N\varepsilon_N^2 \mathbb{E} \mu(B_{i_1, \dots, i_K}) = \frac{4N\varepsilon_N^2}{2^K} \mathbb{E} \mu(B) \\ &\leq \frac{4N\varepsilon_N^2}{2^K} \mathbb{E} \mu(\mathbb{C} \setminus \mathbb{D}') \\ &\leq \frac{4N\varepsilon_N^2}{2^K} \left(N - \left(\frac{|\mathbb{D}'|}{\pi} - \frac{e}{N^{3/2}} \right) \right) \\ &= \frac{4N\varepsilon_N^2}{2^K} \left(2N\varepsilon_N - \varepsilon_N^2 + \frac{e}{N^{3/2}} \right) \\ &\lesssim \frac{N^2\varepsilon_N^3}{2^K}, \end{aligned}$$

et finalement

$$\mathbb{E} \left[W_2^2(\mu|_B, \nu_1) \chi_\rho \right] \lesssim N^2\varepsilon_N^3. \quad (3.5.8)$$

On se tourne maintenant vers (3.5.7). Notons $B' := B_{i_1, \dots, i_{K-k}}$ et $A' := A_{i_1, \dots, i_{K-k}}$ des arcs d'angle θ_k et B'_1, B'_2, A'_1, A'_2 leurs deux sous-arcs chacun d'angle θ_{k-1} . On veut estimer

$$\mathbb{E} [W_{A'}^2(\nu_k, \nu_{k+1}) \chi_\rho].$$

Puisque $\rho(M_N) \leq \sqrt{N}(1 + 3\varepsilon_N)$, on a par isotropie, pour $i \in \{1, 2\}$

$$\mathbb{E} \mu(B'_i) = \frac{\theta_{k-1}}{2\pi} \mathbb{E} \mu(B) = \frac{\theta_{k-1}}{2\pi} \mathbb{E} \mu(\mathbb{C} \setminus \mathbb{D}') = \frac{\theta_{k-1}}{2\pi} (N - \mathbb{E} \mu(\mathbb{D}')),$$

donc d'après le Lemme 3.2.1,

$$\mathbb{E}\mu(B'_i) \geq \frac{\theta_{k-1}}{2\pi} \left(N - \frac{|\mathbb{D}'|}{\pi} \right) = \frac{\theta_{k-1}}{2\pi} \frac{|A|}{\pi} = \frac{|A'_i|}{\pi}.$$

On peut donc estimer, grâce au Lemme 3.2.4,

$$\begin{aligned} \mathbb{P} \left(\mu(B'_i) \leq \frac{|A'_i|}{2\pi} \right) &= \mathbb{P} \left(\mu(B'_i) - \mathbb{E}\mu(B'_i) \leq \frac{|A'_i|}{\pi} - \mathbb{E}\mu(B'_i) - \frac{|A'_i|}{2\pi} \right) \\ &\leq \mathbb{P} \left(\mu(B'_i) - \mathbb{E}\mu(B'_i) \leq -\frac{|A'_i|}{2\pi} \right) \\ &\leq \frac{4\pi^2 \text{Var}(\mu(B'_i))}{|A'_i|^2} \leq \frac{4\pi^2 |B'_i|}{|A'_i|^2} \lesssim \frac{1}{|A'_i|}. \end{aligned} \quad (3.5.9)$$

On va également avoir besoin de l'estimation suivante, pour tout $s \geq 2$,

$$\begin{aligned} \mathbb{P} \left(\mu(B') \geq \frac{s^2 |B'|}{\pi} \right) &\leq \mathbb{P} \left(\mu(B') - \mathbb{E}\mu(B') \geq \frac{s^2 - 1}{\pi} |B'| \right) \\ &\leq \frac{\pi^2 \text{Var} \mu(B')}{(s^2 - 1)^2 |B'|^2} \leq \frac{\pi^2}{(s^2 - 1)^2 |B'|}. \end{aligned} \quad (3.5.10)$$

On peut alors estimer

$$\begin{aligned} \mathbb{E} \left[W_{A'}^2(\nu_k, \nu_{k+1}) \chi_\rho \chi_{\{\exists i, \mu_N(B'_i) \leq \frac{|A'_i|}{2\pi}\}} \right] &\leq (2\theta_k \sqrt{N})^2 \left(\mathbb{E} \left[\mu(B') \chi_{\{\exists i, \mu(B'_i) \leq \frac{|A'_i|}{2\pi}\}} \chi_{\{\mu(B') \leq \frac{4|B'|}{\pi}\}} \right] \right. \\ &\quad \left. + \sum_{s=2}^{\infty} \mathbb{E} \left[\mu(B') \chi_{\{\frac{s^2 |B'|}{\pi} < \mu(B') \leq \frac{(s+1)^2 |B'|}{\pi}\}} \right] \right) \\ &\stackrel{(3.5.9) \& (3.5.10)}{\lesssim} \theta_k^2 N \left(|B'| \sum_{i=1}^4 \frac{1}{|A'_i|} + \sum_{s=2}^{\infty} \frac{(s+1)^2}{(s^2 - 1)^2} \right) \lesssim \theta_k^2 N. \end{aligned} \quad (3.5.11)$$

Si maintenant $\mu(B'_i) \geq \frac{|A'_i|}{2\pi}$ pour tout $i \in \{1, 2\}$, alors $\mu(B') \geq \frac{|A'|}{2\pi}$ et

$$W_{A'}^2(\nu_k, \nu_{k+1}) \leq 2\pi \int_{A'} |\nabla \varphi|^2 \leq 2\pi |A'| \int_{A'} |\nu_{k+1} - \nu_k|^2 = \pi |A'|^2 \sum_{i=1}^2 (c - c_i)^2,$$

où $c := \frac{\nu_{k+1}(B')}{|A'|} = \frac{\mu(B')}{|A'|}$, $c_i := \frac{\nu_k(B'_i)}{|A'_i|} = 2 \frac{\mu(B'_i)}{|A'|}$ et φ est solution de

$$\begin{cases} \Delta \varphi = \nu_{k+1} - \nu_k & \text{sur } A', \\ \nabla \varphi \cdot n = 0 & \text{sur } \partial A'. \end{cases}$$

On obtient donc, par le Lemme 3.2.4,

$$\begin{aligned} \mathbb{E} \left[W_{A'}^2(\nu_k, \nu_{k+1}) \chi_\rho \chi_{\{\mu(B'_i) \geq \frac{|A'_i|}{2\pi}, \forall i\}} \right] &\leq \pi \sum_{i=1}^2 \mathbb{E} \left[(\mu(B') - 2\mu(B'_i))^2 \right] \\ &\leq 2\pi \text{Var}(\mu(B')) + 2\pi \sum_{i=1}^2 \text{Var}(\mu(B'_i)) \\ &\leq 4\pi |B'| \lesssim |A'|, \end{aligned} \quad (3.5.12)$$

où on a utilisé le fait que $\mathbb{E}\mu(B') = 2\mathbb{E}\mu(B'_i)$ par isotropie. En combinant (3.5.11) et (3.5.12), on a finalement

$$\mathbb{E} \left[W_{A'}^2(v_k, v_{k+1}) \chi_\rho \right] \lesssim \theta_k^2 N + \theta_k N \varepsilon_N \lesssim \theta_k^2 N,$$

car $\theta_k \geq \varepsilon_N$. On remonte maintenant les calculs :

$$\mathbb{E} \left[W_2^2(v_k, v_{k+1}) \chi_\rho \right] \lesssim 2^{K-k} \theta_k^2 N = 2\pi \theta_k N,$$

et donc (3.5.2) devient

$$\begin{aligned} \mathbb{E} \left[W_2^2(\tilde{\mu}, \kappa \, dz|_A) \chi_\rho \right]^{1/2} &\lesssim N \varepsilon_N^{3/2} + \sum_{k=1}^K \sqrt{\theta_k} \sqrt{N} = N \varepsilon_N^{3/2} + \sqrt{N} \sum_{k=1}^K \sqrt{2^k} \sqrt{\theta_0} \\ &\leq N \varepsilon_N^{3/2} + \sqrt{N} \sqrt{2^{K+1} \theta_0} \lesssim N \varepsilon_N^{3/2} + \sqrt{N} \lesssim \sqrt{N}. \end{aligned} \quad (3.5.13)$$

La réunion de (3.5.4) et (3.5.13) donne ainsi le résultat suivant.

Théorème 3.5.1. *Soit $\mathbb{D}' := D(0, \sqrt{N}(1 - \varepsilon_N))$ et $A := \mathbb{D}_N \setminus \mathbb{D}'$, alors*

$$\mathbb{E} W_2^2 \left(\mu_N|_{\mathbb{C} \setminus \mathbb{D}'}, \frac{\mu_N(\mathbb{C} \setminus \mathbb{D}')}{|A|} dz|_A \right) \lesssim N.$$

3.6 Vitesse de convergence pour la loi du cercle

On est maintenant prêt à donner la preuve du Théorème 3.1.1.

Démonstration du Théorème 3.1.1. On rappelle que $\mathbb{D}_N := D(0, \sqrt{N})$ et $\mathbb{D}' := D(0, \sqrt{N}(1 - \varepsilon_N))$ où $\varepsilon_N := \sqrt{\frac{\ln N}{N}}$. Définissons la mesure

$$\nu_N := \frac{\mu_N(\mathbb{D}')}{|\mathbb{D}'|} dz|_{\mathbb{D}'} + \frac{\mu_N(\mathbb{C} \setminus \mathbb{D}')}{|\mathbb{D}_N \setminus \mathbb{D}'|} dz|_{\mathbb{D}_N \setminus \mathbb{D}'}$$

Par inégalité triangulaire, il suffit d'estimer $\mathbb{E} W_2^2(\mu_N, \nu_N)$ et $\mathbb{E} W_2^2(\nu_N, \nu)$. En transportant de façon optimale $\mu_N|_{\mathbb{D}'}$ sur $\nu_N|_{\mathbb{D}'}$ et $\mu_N|_{\mathbb{C} \setminus \mathbb{D}'}$ sur $\nu_N|_{\mathbb{D}_N \setminus \mathbb{D}'}$, on obtient grâce aux Théorèmes 3.4.1 et 3.5.1 l'estimation

$$\mathbb{E} W_2^2(\mu_N, \nu_N) \lesssim N.$$

Il reste enfin à estimer $\mathbb{E} W_2^2(\nu_N, \nu)$. Si $\frac{\mu_N(\mathbb{D}')}{|\mathbb{D}'|} \geq \frac{1}{2\pi}$ et $\frac{\mu_N(\mathbb{C} \setminus \mathbb{D}')}{|\mathbb{D}_N \setminus \mathbb{D}'|} \geq \frac{1}{2\pi}$ alors

$$\begin{aligned} W_2^2(\nu_N, \nu) &\leq 2\pi \int_{\mathbb{D}_N} |\nabla \phi|^2 \leq 2\pi^2 N \int_{\mathbb{D}_N} |\nu - \nu_N|^2 \\ &= 2\pi^2 N \left[\int_{\mathbb{D}'} \left(\frac{1}{\pi} - \frac{\mu_N(\mathbb{D}')}{|\mathbb{D}'|} \right)^2 + \int_{\mathbb{D}_N \setminus \mathbb{D}'} \left(\frac{1}{\pi} - \frac{\mu_N(\mathbb{C} \setminus \mathbb{D}')}{|\mathbb{D}_N \setminus \mathbb{D}'|} \right)^2 \right], \end{aligned}$$

où ϕ est solution de

$$\begin{cases} \Delta \phi = \nu - \nu_N & \text{sur } \mathbb{D}_N, \\ \nabla \phi \cdot n = 0 & \text{sur } \partial \mathbb{D}_N. \end{cases}$$

Donc, en notant $\chi_{\geq} := \chi_{\left\{\frac{\mu_N(\mathbb{D}')}{|\mathbb{D}'|} \geq \frac{1}{2\pi} \text{ et } \frac{\mu_N(\mathbb{C} \setminus \mathbb{D}')}{|\mathbb{D}_N \setminus \mathbb{D}'|} \geq \frac{1}{2\pi}\right\}}$,

$$\begin{aligned} \mathbb{E} \left[W_2^2(\nu_N, \nu) \chi_{\geq} \right] &\leq 2\pi^2 N \left(\frac{1}{|\mathbb{D}'|} \mathbb{E} \left[\left(\frac{|\mathbb{D}'|}{\pi} - \mu_N(\mathbb{D}') \right)^2 \right] + \frac{1}{|\mathbb{D}_N \setminus \mathbb{D}'|} \mathbb{E} \left[\left(\frac{|\mathbb{D}_N \setminus \mathbb{D}'|}{\pi} - \mu_N(\mathbb{C} \setminus \mathbb{D}') \right)^2 \right] \right) \\ &\lesssim N \left(\frac{1}{|\mathbb{D}'|} \text{Var } \mu_N(\mathbb{D}') + \frac{1}{|\mathbb{D}_N \setminus \mathbb{D}'|} \text{Var } \mu_N(\mathbb{D}') \right) \\ &\lesssim N \left(\frac{1}{N} + \frac{1}{N\varepsilon_N} \right) \sqrt{N} \lesssim \frac{N}{\sqrt{\ln N}}. \end{aligned}$$

Maintenant,

$$\begin{aligned} \mathbb{P} \left(\mu_N(\mathbb{D}') \leq \frac{|\mathbb{D}'|}{2\pi} \right) &= \mathbb{P} \left(\mu_N(\mathbb{D}') - \mathbb{E}\mu_N(\mathbb{D}') \leq \frac{|\mathbb{D}'|}{\pi} - \mathbb{E}\mu_N(\mathbb{D}') - \frac{|\mathbb{D}'|}{2\pi} \right) \\ &\leq \mathbb{P} \left(\mu_N(\mathbb{D}') - \mathbb{E}\mu_N(\mathbb{D}') \leq \frac{e}{N^{3/2}} - \frac{|\mathbb{D}'|}{2\pi} \right) \\ &\leq \mathbb{P} \left(\mu_N(\mathbb{D}') - \mathbb{E}\mu_N(\mathbb{D}') \leq -\frac{|\mathbb{D}'|}{4\pi} \right) \\ &\leq \frac{16\pi^2 \text{Var } \mu_N(\mathbb{D}')}{|\mathbb{D}'|^2} \lesssim \frac{1}{N^{3/2}}. \end{aligned}$$

La même stratégie pour estimer $\mathbb{P} \left(\mu_N(\mathbb{C} \setminus \mathbb{D}') \leq \frac{|\mathbb{D}_N \setminus \mathbb{D}'|}{2\pi} \right)$ ne donne que du $\frac{1}{\sqrt{N \ln N}}$, ce qui ne sera pas suffisant. On utilise alors l'inégalité de concentration (3.2.1) pour obtenir

$$\begin{aligned} \mathbb{P} \left(\mu_N(\mathbb{C} \setminus \mathbb{D}') \leq \frac{|\mathbb{D}_N \setminus \mathbb{D}'|}{2\pi} \right) &= \mathbb{P} \left(\mu_N(\mathbb{D}') \geq \frac{N}{2} + \frac{|\mathbb{D}'|}{2\pi} \right) \\ &\leq \mathbb{P} \left(\mu_N(\mathbb{D}') - \mathbb{E}\mu_N(\mathbb{D}') \geq \frac{N}{2} - \frac{|\mathbb{D}'|}{2\pi} \right) \\ &= \mathbb{P} \left(\mu_N(\mathbb{D}') - \mathbb{E}\mu_N(\mathbb{D}') \geq N\varepsilon_N \left(1 - \frac{\varepsilon_N}{2} \right) \right) \\ &\leq \mathbb{P} \left(\mu_N(\mathbb{D}') - \mathbb{E}\mu_N(\mathbb{D}') \geq \frac{N\varepsilon_N}{2} \right) \\ &\leq \exp \left(- \min \left\{ \frac{N^2 \varepsilon_N^2}{64 \text{Var } \mu_N(\mathbb{D}')} , \frac{N\varepsilon}{2} \right\} \right) \\ &\leq \exp \left(-c\sqrt{N \ln N} \right). \end{aligned}$$

Avec une estimation grossière du transport, on obtient donc

$$\mathbb{E} \left[W_2^2(\nu_N, \nu) \chi_{\left\{ \frac{\mu_N(\mathbb{D}')}{|\mathbb{D}'|} \leq \frac{1}{2\pi} \text{ ou } \frac{\mu_N(\mathbb{C} \setminus \mathbb{D}')}{|\mathbb{D}_N \setminus \mathbb{D}'|} \leq \frac{1}{2\pi} \right\}} \right] \lesssim N^2 \left(\frac{1}{N^{3/2}} + e^{-c\sqrt{N \ln N}} \right) \lesssim \sqrt{N},$$

ce qui conclut la preuve. ■

3.A Lemmes techniques

On rassemble dans cet appendice quelques lemmes techniques élémentaires qui nous sont utiles dans ce chapitre.

Lemme 3.A.1. *Pour tout $n \geq 1$,*

$$\sqrt{2\pi n} \left(\frac{n}{e}\right)^n \leq n! \leq e\sqrt{n} \left(\frac{n}{e}\right)^n.$$

Les constantes n'ont pas d'importance pour l'utilité qu'on fait de ce résultat. Cette version est prouvée par exemple dans [MM14, Lem. 7]

Le lemme suivant se prouve facilement par récurrence.

Lemme 3.A.2. *Pour tous $k \in \mathbb{N}$ et $a > 0$,*

$$\frac{1}{k!} \int_a^\infty e^{-s} s^k ds = e^{-a} \sum_{l=0}^k \frac{a^l}{l!} \quad \text{et} \quad \frac{1}{k!} \int_0^a e^{-s} s^k ds = e^{-a} \sum_{l=k+1}^\infty \frac{a^l}{l!}.$$

Les deux lemmes suivants permettent d'estimer les sommes partielles et le reste de la série exponentielle.

Lemme 3.A.3. *Si $r \geq 1$, on a*

$$\sum_{k=0}^{n-1} \frac{(nr)^k}{k!} \leq (er)^n \sqrt{\frac{n}{2\pi}}.$$

Démonstration. Soit $r \geq 1$,

$$\sum_{k=0}^{n-1} \frac{(nr)^k}{k!} \leq \frac{(nr)^n}{(n-1)!} \cdot \frac{1}{nr - n + 1} \leq \frac{n^{n+1} r^n}{n^n e^{-n} \sqrt{2\pi n}} \leq (er)^n \sqrt{\frac{n}{2\pi}}.$$

■

Lemme 3.A.4. *Si $r < 1$,*

$$\sum_{k=n}^\infty \frac{(nr)^k}{k!} \leq (er)^n.$$

Démonstration. Soit X une variable aléatoire suivant la loi de Poisson de paramètre nr . Alors pour tout $t > 0$,

$$e^{-nr} \sum_{k=n}^\infty \frac{(nr)^k}{k!} = \mathbb{P}(X \geq n) = \mathbb{P}\left(e^{tX} \geq e^{tn}\right) \leq e^{-tn} \mathbb{E}e^{tX} = e^{tn} e^{(e^t-1)nr},$$

donc, en choisissant $t := -\ln r > 0$,

$$\sum_{k=n}^\infty \frac{(nr)^k}{k!} \leq (er)^n.$$

■

Enfin, on utilise les versions quantitatives suivantes de la méthode de Laplace.

Lemme 3.A.5. *Pour tous $a \in \mathbb{R}$ et $x_0 < a$, pour tout $t > 0$,*

$$\int_a^{+\infty} e^{-t(x-x_0)^2} dx \leq \frac{e^{-t(a-x_0)^2}}{2t(a-x_0)}.$$

Pour tous $a \in \mathbb{R}$ et $x_0 > a$, pour tout $t > 0$,

$$\int_{-\infty}^a e^{-t(x-x_0)^2} dx \leq \frac{e^{-t(x_0-a)^2}}{2t(x_0-a)}.$$

Démonstration. Soit $x_0 < a$, $t > 0$,

$$\int_a^{+\infty} e^{-t(x-x_0)^2} dx = \frac{1}{2\sqrt{t}} \int_{t(a-x_0)^2}^{+\infty} e^{-u} \frac{du}{\sqrt{u}} \leq \frac{1}{2t(a-x_0)} \int_{t(a-x_0)^2}^{+\infty} e^{-u} du = \frac{e^{-t(a-x_0)^2}}{2t(a-x_0)}.$$

La deuxième intégrale se ramène à la première par le changement de variables $x \mapsto -x$. ■

Références bibliographiques

- [AKT84] M. AJTAI, J. KOMLÓS, and G. TUSNÁDY, On optimal matchings. *Combinatorica* **4** (1984), 259–264. 29
- [Ale58] A. D. ALEXANDROV, Dirichlet’s problem for the equation $\det \|z_{ij}\| = \varphi(z_1, \dots, z_n, z, x_1, \dots, x_n)$ I. (Russian). *Vestnik Leningrad. Univ. Ser. Mat. Meh. Astr.* **13** (1958), no. 1, 5–24. Zbl 0114.30202. 19
- [Alm76] F. J. ALMGREN JR., *Existence and regularity almost everywhere of solutions to elliptic variational problems with constraints*. Memoirs of the American Mathematical Society **4** (1976), no. 165. American Mathematical Society, Providence, RI, 1976. ISBN: 978-0-8218-1865-7. 35
- [Amb03] L. AMBROSIO, Lecture Notes on Optimal Transport Problems. In *Mathematical Aspects of Evolving Interfaces*, Springer Verlag, Lecture Notes in Mathematics, **1812** (2003), 1–52. 17
- [AFP00] L. AMBROSIO, N. FUSCO, and D. PALLARA, *Functions of bounded variation and free discontinuity problems*. Oxford Mathematical Monographs, The Clarendon Press, Oxford University Press, New-York, 2000. ISBN: 0-19-850254-1. 14
- [AGS08] L. AMBROSIO, N. GIGLI, and G. SAVARÉ, *Gradient flows in metric spaces and in the space of probability measures, Second Edition*. Lectures in Mathematics ETH Zürich, Birkhäuser Verlag, Basel, 2008. 8, 14
- [AGT19] L. AMBROSIO, F. GLAUDO, and D. TREVISAN, On the optimal map in the 2-dimensional random matching problem. *Discrete Contin. Dyn. Syst.* **39** (2019), no. 12, 7291–7308. 29
- [AP03] L. AMBROSIO and A. PRATELLI, Existence and stability results in the L^1 theory of optimal transportation. *Optimal transportation and applications. Lectures Notes in Mathematics* **1813** (2003). 16

- [AST19] L. AMBROSIO, F. STRA, and D. TREVISAN, A PDE approach to a 2-dimensional matching problem. *Probab. Theory Relat. Fields* **173** (2019), 433–477. 29
- [AGZ10] G. W. ANDERSON, A. GUIONNET, and O. ZEITOUNI, *An introduction to random matrices*. Cambridge University Press, 2010. ISBN: 978-0-521-19452-5. 30
- [BÉ85] D. BAKRY and M. ÉMERY, Diffusions hypercontractives. *Séminaire de Probabilités, XIX, 1983/84*, 177–206, Lecture Notes in Math., 1123, Springer, Berlin (1985). 27
- [BB13] F. BARTHE and C. BORDENAVE, Combinatorial optimization over two random point sets. *Séminaires de Probabilités XLV*, 483–535. 29
- [BJ21] T. BECK and D. JERISON, The Friedland-Hayman inequality and Caffarelli’s contraction theorem. *Preprint* (2021). arXiv:2102.00571. 47
- [BGMS09] M. BEIGLBÖCK, M. GOLDSTERN, G. MARESCH, and W. SCHACHERMAYER, Optimal and better transport plans. *J. Funct. Anal.* **256** (2009), 1907–1927. 16
- [BB99] J.-D. BENAMOU and Y. BRENIER, A numerical method for the optimal time-continuous mass transport problem and related problems. In *Monge-Ampère equation: applications to geometry and optimization (Deerfield Beach, FL, 1997)*, 1–11, *Contemporary Mathematics* **226**, American Mathematical Society, Providence, RI, 1999. 13, 14, 65
- [BCCNP15] J.-D. BENAMOU, G. CARLIER, M. CUTURI, L. NENNA, and G. PEYRÉ, Iterative Bregman projections for regularized transportation problems. *SIAM J. Sci. Comput.* **37** (2015), no.2, A1111–A1138. 48
- [BC10] S. BIANCHINI and L. CARAVENNA, On optimality of c -cyclically monotone transference plans. *C. R. Acad. Sci. Paris, Ser. I* **348** (2010), 613–618. 16
- [BL16] S. BOBKOV and M. LEDOUX, *One-dimensional empirical measures, order statistics, and Kantorovich transport distances*. Memoirs Amer. Math. Soc. **261**, American Mathematical Society, Providence, RI, 2019. ISBN: 978-1-4704-3650-6. 28
- [Bom82] E. BOMBIERI, Regularity theory for almost minimal currents. *Arch. Ration. Mech. Anal.* **78** (1982), 99–130. 35
- [BC12] C. BORDENAVE and D. CHAFAI, Around the circular law. *Probab. Surveys* **9** (2012), 1–89. 31
- [BJR07] G. BOUCHITTÉ, C. JIMENEZ, and M. RAJESH, A new L^∞ estimate in optimal mass transport. *Proc. Amer. Math. Soc.* **135** (2007), no. 11, 3525–3535. 35
- [Bre87] Y. BRENIER, Décomposition polaire et réarrangement monotone des champs de vecteurs. *C. R. Acad. Sci. Paris Sér. I. Math.* **305** (1987), no. 19, 805–808. MR 0923203. 12

- [Caf90] L. A. CAFFARELLI, Interior $W^{2,p}$ estimates for solutions of the Monge-Ampère equation. *Ann. of Math. (2)*, **131** (1990), no.1, 135–150. 21
- [Caf91] L. A. CAFFARELLI, Some regularity of solutions to Monge-Ampère equations. *Comm. Pure Appl. Math.* **44** (1991), 965–969. 20
- [Caf92a] L. A. CAFFARELLI, The regularity of mappings with a convex potential. *J. Am. Math. Soc.* **5** (1992), no. 1, 99–104. 21, 23
- [Caf92b] L. A. CAFFARELLI, Boundary regularity of maps with convex potentials. *Comm. Pure Appl. Math.* **45** (1992), no. 9, 1141–1151. 21
- [Caf96] L. A. CAFFARELLI, Allocation maps with general cost functions. In P. Marcellini et al, editor, *Partial Differential Equations and Applications* **177** in Lecture Notes in Pure and Appl. Math. (1996), 29–35. 17
- [Caf00] L. A. CAFFARELLI, Monotonicity properties of optimal transportation and the FKG and related inequalities. *Comm. Math. Phys.* **214** (2000), no. 3, 547–563. 26, 41
- [Caf02] L. A. CAFFARELLI, Erratum: “Monotonicity of optimal transportation and the FKG and related inequalities” [*Comm. Math. Phys.* **214** (2000), no. 3, 547–563; MR1800860 (2002c:60029)]. *Comm. Math. Phys.* **225** (2002), no. 2, 449–450. 26, 41
- [CFM02] L. A. CAFFARELLI, M. FELDMAN, and R. J. MCCANN, Constructing optimal maps for Monge’s transport problem as a limit of strictly convex costs. *J. Amer. Math. Soc.* **15** (2002), no. 1, 1–26. 17
- [Cam64] S. CAMPANATO, Proprietà di una famiglia di spazi funzionali. *Ann. Sc. norm. super. Pisa – Cl. sci. (Serie 3)* **18** (1964), no. 1, 137–160. 24, 96, 98, 107
- [CLPS14] S. CARACCIOLO, C. LUCIBELLO, G. PARISI, and G. SICURO, Scaling hypothesis for the Euclidean bipartite matching problem. *Phys. Rev. E* **90** 012118 (2014). 29
- [CDPS17] G. CARLIER, V. DUVAL, G. PEYRÉ, and B. SCHMITZER, Convergence of entropic schemes for optimal transport and gradient flows. *SIAM J. Math. Anal.* **49** (2017), no. 2, 1385–1418. 33, 44
- [CHM18] D. CHAFAI, A. HARDY, and M. MAÏDA, Concentration for Coulomb gases and Coulomb transport inequalities. *J. Funct. Anal.*, **275** (2018), no. 16, 1447–1483. 32, 39, 143
- [CF15] S. CHEN and A. FIGALLI, Boundary ε -regularity in optimal transportation. *Adv. Math.* **273** (2015), 540–567. 24, 26
- [CF16] S. CHEN and A. FIGALLI, Stability results on the smoothness of optimal transport maps with general costs. *J. Math. Pures Appl.* **106** (2016), no. 2, 280–295. 26

- [CF17] S. CHEN and A. FIGALLI, Partial $W^{2,p}$ regularity for optimal transport maps. *J. Funct. Anal.* **272** (2017), no. 11, 4588–4605. 26
- [CF03] G.-Q. CHEN and H. FRID, Extended Divergence-Measure Fields and the Euler Equations for Gas Dynamics. *Comm. Math. Phys.* **236** (2003), 251–280. 65
- [CGP16] Y. CHEN, T. GEORGIU, and M. PAVON, Entropic and displacement interpolation: a computational approach using the Hilbert metric. *SIAM J. Math. Analysis* **76** (2016), no. 6, 2375–2396. 48
- [CJLPR15] O. CHODOSH, V. JAIN, M. LINDSEY, L. PANCHEV, and Y. A. RUBINSTEIN, On discontinuity of planar optimal maps. *J. Topol. Anal.* **7** (2015), no. 2, 239–260. 20
- [CFJ17] M. COLOMBO, A. FIGALLI, and Y. JHAVERI, Lipschitz changes of variables between perturbations of log-concave measures. *Ann. Sc. Norm. Super. Pisa Cl. Sci. (5)* **17** (2017), no. 4, 1491–1519. 27
- [Cor02] D. CORDERO-ERAUSQUIN, Some applications of mass transport to Gaussian-type inequalities. *Arch. Ration. Mech. Anal.* **161** (2002), no. 3, 257–269. 27
- [CF20] T. A. COURTADE and M. FATHI, Stability of the Bakry-Émery theorem on \mathbb{R}^n . *J. Funct. Anal.* **279** (2020), no. 2, 108523. 27
- [CFP18] T. A. COURTADE, M. FATHI, and A. PANANJADY, Quantitative stability of the entropy power inequality. *IEEE Trans. Inform. Theory* **64** (2018), no. 8, 5691–5703. 27
- [Csi75] I. CSISZÁR, I -divergence geometry of probability distributions and minimization problems. *Ann. Probab.* **3** (1975), 146–158. 33, 45
- [Cut13] M. CUTURI, Sinkhorn distances: lightspeed computation of optimal transport. *Proceedings of the 26th International Conference on Neural Information Processing Systems* **2** (2013), 2292–2300. 10, 48
- [Dal12] S. DALLAPORTA, Eigenvalue variance bounds for Wigner and covariance random matrices. *Random Matrices Theory Appl.* **1** (2012), no. 3. 30
- [DF15] G. DE PHILIPPIS and A. FIGALLI, Partial regularity for optimal transport maps. *Publ. Math. Inst. Hautes Études Sci.* **121** (2015), 81–112. 25, 26, 34, 35, 36, 59, 60, 61, 70, 101, 110
- [DF17] G. DE PHILIPPIS and A. FIGALLI, Rigidity and stability of Caffarelli’s log-concave perturbation theorem. *Nonlinear Anal. Theory Methods Appl.* **154** (2017), 59–70. 27, 47
- [DY95] V. DOBRIĆ and J. YUKICH, Asymptotics for transportation cost in high dimensions. *J. Theoret. Probab.* **8** (1995), 97–118. 29

- [EP19] M. ESSID and M. PAVON, Traversing the Schrödinger Bridge Strait: Robert Fortet’s Marvelous Proof Redux. *J. Optim. Theory Appl.* **181** (2019), no. 1, 23–60. 47
- [EG15] L. C. EVANS and R. F. GARIEPY, *Measure theory and fine properties of functions*. Revised edition. Textbooks in Mathematics. CRC Press, Boca Raton, FL, 2015. ISBN: 978-1-4822-4238-6. MR 3409135. 60
- [FGP20] M. FATHI, N. GOZLAN, and M. PROD’HOMME, A proof of the Caffarelli contraction theorem via entropic regularization. *Calc. Var. PDE* **59**, 96 (2020). 5
- [Fig10] A. FIGALLI, Regularity properties of optimal maps between nonconvex domains in the plane. *Comm. Partial Differential Equations* **35** (2010), no. 3, 465–479. 23
- [Fig17] A. FIGALLI, *The Monge-Ampère Equation and Its Applications*. Zurich Lectures in Advanced Mathematics. European Mathematical Society (EMS), Zürich, 2017. ISBN: 978-3-03719-170-5. 19, 111
- [FK10] A. FIGALLI and Y.-H. KIM, Partial regularity of Brenier solutions of the Monge-Ampère equation. *Discrete Cont. Dyn. Syst. (Series A)* **28** (2010), 559–565. 23
- [For40] R. FORTET, Résolution d’un système d’équations de M. Schrödinger. *J. Math. Pures Appl.* **19** (1940), 83–105. 47, 48
- [FG15] N. FOURNIER and A. GUILLIN, On the rate of convergence in Wasserstein distance of the empirical measure. *Probab. Theory Relat. Fields* **162** (2015), 707–738. 30
- [FL89] J. FRANKLIN and J. LORENZ, On the scaling of multidimensional matrices. *Linear Algebra Appl.* **114/115** (1989), 717–735. 48
- [Gal16] A. GALICHON, *Optimal Transport Methods in Economics*. Princeton University Press, 2016. ISBN: 978-0-691-17276-7. 10
- [GM95] W. GANGBO and R. J. MCCANN, Optimal maps in Monge’s mass transport problem. *C. R. Acad. Sci. Paris Sér. I Math.* **321** (1995), no. 12, 1653–1658. MR MR1367824. 17
- [Gen19] A. GENEVAY, Entropy-Regularized Optimal Transport for Machine Learning. Ph.D. Thesis, Université Paris-Dauphine, 2019. 48
- [Gig12] N. GIGLI, *Second order analysis on $(\mathcal{P}_2(M), W_2)$* . Memoirs of the American Mathematical Society **216**, no. 1018, American Mathematical Society, Providence, RI, 2012. ISBN: 978-0-821-85309-2. 15
- [Gin65] J. GINIBRE, Statistical ensembles of complex, quaternion, and real matrices. *J. Mathematical Phys.* **6** (1965), 440–449. 31

- [Gol20] M. GOLDMAN, An ε -regularity result for optimal transport maps between continuous densities. *Atti della Accademia Nazionale dei Lincei. Rendiconti Lincei. Matematica e Applicazioni* **31** (2020), no. 4, 971–979. 24, 37, 60, 100, 101
- [GHO18] M. GOLDMAN, M. HUESMANN, and F. OTTO, A large-scale regularity theory for the Monge-Ampère equation with rough data and application to the optimal matching problem. *Preprint* (2018). arXiv:1808.09250. 39
- [GHO21] M. GOLDMAN, M. HUESMANN, and F. OTTO, Quantitative linearization results for the Monge-Ampère equation. *Comm. Pure Appl. Math.* (2021), <https://doi.org/10.1002/cpa.21994>. 22, 24, 34, 35, 36, 37, 39, 59, 61, 62, 66, 67, 73, 76, 78, 79, 81, 82, 108
- [GO20] M. GOLDMAN and F. OTTO, A variational proof of partial regularity for optimal transportation maps. *Ann. Sci. de l'Ecole Norm. Supérieure. Quatrième Série* **53** (2020), no. 5, 1209–1233. 5, 6, 23, 24, 25, 26, 34, 35, 36, 37, 61, 66, 70, 81, 82, 83, 110
- [GT21] M. GOLDMAN and D. TREVISAN, Convergence of asymptotic costs for random Euclidean matching problems. *Probability and Mathematical Physics* **2** (2021), no. 2, 121–142. 29
- [GJa20] F. GÖTZE and J. JALOWY, Rate of convergence to the circular law via smoothing inequalities for log-potentials. *Random Matrices: Theory and Application*, page 2150026, (2020). 31
- [GT05] F. GÖTZE and A. N. TIKHOMIROV, The rate of convergence for spectra of GUE and LUE matrix ensembles. *Cent. Eur. J. Math.* **3** (2005), 666–705. 30
- [GT14] F. GÖTZE and A. N. TIKHOMIROV, Rate of convergence of the empirical spectral distribution function to the semi-circular law. *Preprint* (2014). arXiv:1407.2780. 30
- [GT16] F. GÖTZE and A. N. TIKHOMIROV, Optimal bounds for convergence of expected spectral distributions to the semi-circular law. *Probab. Theory Relat. Fields* **165** (2016), 163–233. 30
- [GJu20] N. GOZLAN and N. JUILLET, On a mixture of Brenier and Strassen theorems. *Proc. London Math. Soc.* **120** (2020), no. 3, 434–463. 34, 42, 52, 53
- [GMRS17] N. GOZLAN, M. MADIMAN, C. ROBERTO, and P.-M. SAMSON, Deviation inequalities for convex functions motivated by the Talagrand conjecture. *Zap. Nauchn. Sem. S.-Peterburg. Otdel. Mat. Inst. Steklov. (POMI)* **457** (2017), Veroyatnost' i Statistika. 25, 168–182. MR 3723580. 51
- [Gro75] L. GROSS, Logarithmic Sobolev inequalities. *Amer. J. Math.* **97** (1975), no. 4, 1061–1083. 27

- [GS14] A. GUIONNET and D. SHLYAKHTENKO, Free monotone transport. *Invent. Math.* **197** (2014), no. 3, 613–661. 47
- [GM21] C. E. GUTIÉRREZ and A. MONTANARI, L^∞ estimates in optimal transport for non quadratic costs. *Preprint* (2021). arXiv:2106.02899. 35
- [Har99] G. HARGÉ, A particular case of correlation inequality for the Gaussian measure. *Ann. Probab.* **27** (1999), no. 4, 1939–1951. 27
- [Har01] G. HARGÉ, Inequalities for the Gaussian measure and an application to Wiener space. *C. R. Acad. Sci. Paris Sér. I Math.* **333** (2001), no. 8, 791–794. 27
- [HL93] J.-B. HIRIART-URRUTY and C. LEMARÉCHAL, *Convex analysis and minimization algorithms. I.* Grundlehren der Mathematischen Wissenschaften **305**, Springer Verlag, Berlin, 1993. ISBN: 978-3-662-02796-7. 56
- [HL01] J.-B. HIRIART-URRUTY and C. LEMARÉCHAL, *Fundamentals of convex analysis.* Grundlehren Text Editions, Springer Verlag, Berlin, 2001. ISBN: 978-3-642-56468-0. 53
- [HKPV06] J. B. HOUGH, M. KRISHNAPUR, Y. PERES, and B. VIRÁG, Determinantal processes and independence. *Probab. Surv.* **3** (2006), 206–229. 127
- [Hwa84] C.-R. HWANG, A brief survey on the spectral radius and the spectral distribution of large random matrices with i.i.d. entries. Random matrices and their applications (Brunswick, Maine, 1984), *Contemp. Math.* **50**, Amer. Math. Soc., Providence, RI, 1986, 145–152. 31
- [Jha19] Y. JHAVERI, On the (in)stability of the identity map in optimal transportation. *Calc. Var. PDE* **58** (2019), no. 3. 21
- [KTT07] A. A. KHOLOPOV, A. N. TIKHOMIROV, and D. A. TIMUSHEV, On the accuracy of the approximation of the GOE spectrum by the semi-circular law. *Teor. Veroyatn. Primen.* **52** (2007), no. 1, 180–185; translation in *Theory Probab. Appl.* **52** (2008), no. 1, 171–177. 30
- [KM12] Y.-H. KIM and E. MILMAN, A generalization of Caffarelli’s contraction theorem via (reverse) heat flow. *Math. Ann.* **354** (2012), no. 3, 827–862. 32
- [Kla13] B. KLARTAG, Poincaré inequalities and moment maps. *Ann. Fac. Sci. Toulouse Math. (6)* **22** (2013), no. 1, 1–41. 47
- [Kol01] A. V. KOLESNIKOV, On Diffusion Semigroups Preserving the Log-Concavity. *J. Funct. Anal.* **186** (2001), no. 1, 196–205. 34
- [Kol10] A. V. KOLESNIKOV, Global Hölder estimates for optimal transportation. *Mat. Zametki* **88** (2010), no. 5, 708–728. 27, 47
- [Kol13] A. V. KOLESNIKOV, On Sobolev regularity of mass transport and transportation inequalities. *Theory Probab. Appl.* **57** (2013), no. 2, 243–264. 27, 32, 41, 47

- [KPTSR17] S. KOLOURI, S.-R. PARK, M. THORPE, D. SLEPČEV, and G. K. ROHDE, Optimal Mass Transport: Signal processing and machine-learning applications. *IEEE Signal Processing Magazine* **34** (2017), no. 4, 43–59. 10
- [Led99] M. LEDOUX, *Concentration of measure and logarithmic Sobolev inequalities*. Séminaire de Probabilités, XXXIII, 120–216, Lecture Notes in Math., 1709, Springer, Berlin, 1999. 27
- [Led19] M. LEDOUX, On optimal matching of Gaussian samples. *J. Math. Sci.* **238** (2019), no. 4, 495–522. 29
- [LZ19] M. LEDOUX and J.-X. ZHU, On optimal matching of Gaussian samples III. *Preprint* (2019). arXiv:1911.07579. 29
- [Léo12] C. LÉONARD, From the Schrödinger problem to the Monge-Kantorovich problem. *J. Funct. Anal.* **262** (2012), no. 4, 1879–1920. 33, 42, 43
- [Léo14] C. LÉONARD, A survey of the Schrödinger problem and some of its connections with optimal transport. *Discrete Contin. Dyn. Syst.* **34** (2014), no. 4, 1533–1574. 33, 42
- [Léo19] C. LÉONARD, Revisiting Fortet’s proof of existence of a solution to the Schrödinger system. *Preprint* (2019). arXiv:1904.13211. 47
- [Loe09] G. LOEPER, On the regularity of solutions of optimal transportation problems. *Acta Math.* **202** (2009), no. 2, 241–283. 25
- [Lot08] J. LOTT, Some geometric calculations on Wasserstein space. *Comm. Math. Phys.* **277** (2008), 423–437. 15
- [LV09] J. LOTT and C. VILLANI, Ricci curvature for metric-measure spaces via optimal transport. *Ann. of Math. (2)* **169** (2009), no. 3, 903–991. 10
- [MTW05] X.-N. MA, N. S. TRUDINGER, and X.-J. WANG, Regularity of potential functions of the optimal transportation problem. *Arch. Ration. Mech. Anal.* **177** (2005), no. 2, 151–183. 17, 25
- [McC95] R. J. MCCANN, Existence and uniqueness of monotone measure-preserving maps. *Duke Math. J.* **82** (1995), no. 2, 309–323. 12
- [McC97] R. J. MCCANN, A convexity principle for interacting gases. *Adv. Math.* **128** (1997), no. 1, 153–179. 19
- [MM13] E. S. MECKES and M. W. MECKES, Concentration and convergence rates for spectral measures of random matrices. *Probab. Theory Relat. Fields* **156** (2013), 145–164. 30
- [MM14] E. S. MECKES and M. W. MECKES, A rate of convergence for the circular law for the complex Ginibre ensemble. *Ann. Fac. Sci. Toulouse* **24** (2014), no. 1, 93–117. 31, 38, 127, 132, 159

- [Meh67] M. L. MEHTA, *Random matrices and the statistical theory of energy levels*. Acad. Press, 1967. ISBN: 978-1-4832-3258-4. 31
- [Meh04] M. L. MEHTA, *Random matrices*, Third ed. Pure and Applied Mathematics **142**, Acad. Press, 2004. ISBN: 978-0-12-088409-4. 30, 126
- [Mik04] T. MIKAMI, Monge’s problem with a quadratic cost by the zero-noise limit of h -path processes. *Probab. Theory Relat. Fields* **129** (2004), no. 2, 245–260. 33, 42, 43
- [Mil09] E. MILMAN, On the role of convexity in isoperimetry, spectral gap and concentration. *Invent. Math.* **177** (2009), no. 1, 1–43. 47
- [Mil18] E. MILMAN, Spectral estimates, contractions and hypercontractivity. *J. Spectr. Theory* **8** (2018), no. 2, 669–714. 27
- [MO21] T. MIURA and F. OTTO, Sharp boundary ε -regularity of optimal transport maps. *Adv. Math.* **381** (2021), Article ID 107603, 65 pp. 24, 34, 37
- [NW21] M. NUTZ and J. WIESEL, Entropic optimal transport: convergence of potentials. *Preprint* (2021) arXiv:2104.11720. 34
- [OT20] S. OHTA and A. TAKATSU, Equality in the logarithmic Sobolev inequality. *Manuscripta Math.* **162** (2020), 271–282. 27
- [Ott01] F. OTTO, The geometry of dissipative evolution equations: the porous medium equation. *Commun. Partial. Differ. Equ.* **26** (2001), 101–174. 15
- [OPR21] F. OTTO, M. PROD’HOMME, and T. RIED, Variational approach to regularity of optimal transport maps: general cost functions. To appear in: *Ann. PDE* (2021). 5
- [OV00] F. OTTO and C. VILLANI, Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality. *J. Funct. Anal.* **173** (2000), no. 2, 361–400. 22
- [Pey16] R. PEYRE, Comparison between W_2 distance and \dot{H}^{-1} norm, and localisation of Wasserstein distance. *Preprint* (2016). arXiv:1104.4631. 22
- [PC19] G. PEYRÉ and M. CUTURI, *Computational Optimal Transport: With Applications to Data Science*. Foundations and Trends in Machine Learning, **11** (2019), no. 5–6, 355–607. 10
- [Pra08] A. PRATELLI, On the sufficiency of c -cyclical monotonicity for optimality of transport plans. *Math. Z.* **258** (2008), 677–690. 16
- [Pré73] A. PRÉKOPA, On logarithmic concave measures and functions. *Acta Sci. Math. (Szeged)* **34** (1973), 335–343. MR 0404557. 48, 52

- [Rid03] B. RIDER, A limit theorem at the edge of a non-Hermitian random matrix ensemble. *J. Phys. A: Math. Gen.* **36** (2003), 3401–3409. 38, 39, 153
- [Roc66] R. T. ROCKAFELLAR, Characterization of the subdifferentials of convex functions. *Pac. J. Math.* **17** (1966), 497–510. 11
- [Roc70] R. T. ROCKAFELLAR, *Convex Analysis*. Princeton Mathematical Series **28**, Princeton University Press, Princeton, N.J., 1970. ISBN: 978-0-691-01586-6. 56
- [Rüs96] L. RÜSCHENDORF, On c -optimal random variables. *Stat. Prob. Lett.* **27** (1996), 267–270. 16
- [San15] F. SANTAMBROGIO, *Optimal transport for applied mathematicians*. Calculus of Variations, PDEs, and Modeling. Progress in Nonlinear Differential Equations and their Applications **87**, Birkhäuser, Cham, 2015. ISBN: 978-3-319-20827-5. 7, 10, 14, 17, 22, 65, 116, 117
- [ST09] W. SCHACHERMAYER and J. TEICHMANN Characterization of optimal transport plans for the Monge–Kantorovich problem. *Proc. Amer. Math. Soc.* **137** (2009), no. 2, 519–529. 16
- [SS82] R. SCHOEN and L. SIMON, A New Proof of the Regularity Theorem for Rectifiable Currents which Minimize Parametric Elliptic Functionals. *Indiana Univ. Math. J.* **31** (1982), no. 3, 415–434. 23
- [Sch32] E. SCHRÖDINGER, Sur la théorie relativiste de l'électron et l'interprétation de la mécanique quantique. *Ann. Inst. H. Poincaré* **2** (1932), no. 4, 269–310. MR 1508000. 45
- [Stu06a] K.-T. STURM, On the geometry of metric measure spaces I. *Acta Math.* **196** (2006), no. 1, 65–131. 10
- [Stu06b] K.-T. STURM, On the geometry of metric measure spaces II. *Acta Math.* **196** (2006), no. 1, 133–177. 10
- [Sud79] V. N. SUDAKOV, Geometric problems in the theory of infinite-dimensional probability distributions. *Proc. Steklov Inst. Math.*, **141** (1979), 1–178. MR MR0530375 17
- [Tal92] M. TALAGRAND, Matching random samples in many dimensions. *Ann. Appl. Probab.* **2** (1992), no. 4, 846–856. 29
- [Tal18] M. TALAGRAND, Scaling and non-standard matching theorems. *C. R. Math.* **356** (2018), no. 6, 692–695. 29
- [TV10] T. TAO and V. VU, with an appendix by M. KRISHNAPUR, Random matrices: universality of ESDs and the circular law. *Ann. Probab.* **38** (2010), no. 5, 2023–2065. 31

- [Tro87] G. M. TROIANIELLO, *Elliptic differential equations and obstacle problems*. The University Series in Mathematics, Plenum Press, New York, 1987. ISBN 0-306-42448-7. 116
- [TW01] N. S. TRUDINGER and X.-J. WANG, On the Monge mass transfer problem. *Calc. Var. PDE* **13** (2001), 19–31. 17
- [Urb07] J. URBAS, A remark on minimal Lagrangian diffeomorphisms and the Monge-Ampère equation. *B. Aust. Math. Soc.* **76** (2007), no. 2, 215–218. 20
- [Vil03] C. VILLANI, *Topics in optimal transportation*. Graduate Studies in Mathematics **58**, American Mathematical Society, Providence, RI, 2003. ISBN: 978-0-8218-3312-4. 8, 14, 52, 65
- [Vil09] C. VILLANI, *Optimal transport: old and new*. Grundlehren der Mathematischen Wissenschaften **338**, Springer-Verlag, Berlin, 2009. ISBN: 978-3-540-71049-3. 15, 19, 42, 52, 54, 55, 60, 119
- [Wig55] E. P. WIGNER, Characteristic vectors of bordered matrices with infinite dimensions. *Ann. Math.* **62** (1955), no. 3, 548–564. 30
- [Wig58] E. P. WIGNER, On the distribution of the roots of certain symmetric matrices. *Ann. Math.* **67** (1958), no. 2, 325–327. 30
- [Wol97] J. WOLFSON, Minimal Lagrangian diffeomorphisms and the Monge-Ampère equation. *J. Differ. Geom.* **46** (1997), no. 2, 335–373. 20
- [Yuk92] J. YUKICH, Some generalizations of the Euclidean two-sample matching problem. *Probability in Banach spaces* **8**, *Prog. Probab.* **30** (1992), Birkhäuser, 55–66. 29